

UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Chemie und Metabolomics
der Universität zu Lübeck
Direktor: Prof. Dr. Thomas Peters

Virus Receptor Identification via Click Chemistry and Fragment-based Drug Discovery using NMR Spectroscopy

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Naturwissenschaften

vorgelegt von
Tobias Schöne
aus Schwerin

Lübeck, 2018

1. Berichterstatter: Prof. Dr. rer. nat. Thomas Peters

2. Berichterstatter: Prof. Dr. rer. nat. Lars Redecke

Tag der mündlichen Prüfung: 19. Dezember 2018

Zum Druck genehmigt. Lübeck, den 3. Januar 2019

To my family, my friends, and Marlene,
for their love and encouragement.

Abstract

For more than 60 years, human beings have attempted to combat viruses by using antiviral agents. Due to the rapid progress of biomolecular and biophysical methodologies, we are beginning to understand the complex interaction mechanisms between viruses and their host organisms. Based on this knowledge, researchers have developed a variety of antiviral therapies against some of the known human pathogens, for example, against the life-threatening human immunodeficiency viruses (HIV) or hepatitis C viruses (HCV). The first steps of the rational development of antiviral agents involve the identification and characterization of essential macromolecules, which permit replication of viruses in susceptible cells. Following these steps, structure-based drug design is nowadays widely applied to create selective drug molecules that, e.g., block viral entry or inhibit enzymatic activity. In this thesis, two aspects of pharmaceutical research are addressed, including a strategy for the identification of cellular virus receptors as well as fragment-based drug discovery based on nuclear magnetic resonance (NMR) spectroscopy.

The first part of this thesis introduces a novel method to identify glycosylated cellular receptors for enveloped viruses. This strategy involves four steps: At first, glycoproteins of living cells are metabolically labeled by the use of monosaccharide precursors that carry alkyne groups as bioorthogonal chemical functionalities. In a second step, viral glycoproteins that contain glycans with azido groups are recombinantly synthesized. Afterwards, a copper-catalyzed alkyne-azido cycloaddition (*click chemistry*) allows the formation of a covalent linkage between viral and cellular glycan structures upon incubation of the cells with viral glycoproteins. Finally, viral proteins that are covalently linked to their specific receptors are purified by the use of affinity tags and analyzed by immunoblotting or mass spectrometry. As a proof of concept, the method was tested employing the ectodomain of the E_2 protein of bovine viral diarrhea viruses (BVDV) and susceptible MDBK cells, which express the BVDV receptor CD46. Viral E_2 glycoproteins with azido groups in *N*-glycans were recombinantly synthesized by a human cell line (HEK293T cells). Immunofluorescence with E_2-antibodies strongly indicated the formation of a covalent linkage between azido group-bearing E_2 proteins and alkyne group-bearing glycoconjugates of MDBK cells induced through *click chemistry*. Moreover, the first immunoblot of putative E_2-receptor complexes is presented, suggesting the feasibility of receptor identification using this method.

The second part of this thesis describes the screening of fragment libraries employing saturation transfer difference (STD) NMR experiments as an initial read-out. The aim was to identify novel structural motifs with binding affinity for the West Nile virus (WNV) protease NS2B-NS3^{pro} as well as for dimers of the protruding (P) domain of murine (MNV-1) and human Norovirus strains (Saga and MI001). Hit validation was achieved either by protease activity assays (WNV) or surface plasmon resonance (SPR) spectroscopy experiments (MNV-1). STD NMR competition experiments allowed the identification of secondary site binders and potential allosteric leads. Protein-observed NMR methods involving chemical shift perturbation (CSP) experiments of isotopically labeled NS2B-NS3^{pro} and Saga P dimers led to the discovery of secondary binding sites. In the case of WNV protease, structure-activity relationships (SAR) were investigated, revealing critical pharmacophores of inhibitory fragments. Furthermore, molecular docking served as an important tool to guide the interpretation of experimental data. Several docking models of the WNV protease in complex with hits that share a common structural motif suggested a critical role of a carboxyl group for inhibitory efficacy, which points towards the catalytic triad of NS2B-NS3^{pro}. In the case of Norovirus P dimers, “hot spots” for fragment binding were discovered through so-called “blind” virtual docking, targeting practically all possible binding sites in parallel. Chemical shift perturbation mapping of Saga P dimers in complex with hits was directly compared to binding pocket predictions based on molecular docking algorithms as implemented in Glide. Docking models, which were selected based on the *Emodel* scoring function, were in good agreement with the majority of CSP data. This knowledge was used to predict binding sites for fragments that inhibited MNV-1 infection *in vitro*, which provided structural models for further studies.

Zusammenfassung

Schon mehr als 60 Jahre versucht die Menschheit, Viren mithilfe antiviraler Wirkstoffe zu bekämpfen. Durch die rasante Entwicklung molekularer und biophysikalischer Methoden beginnen wir, die komplexen Interaktionsmechanismen zwischen Viren und ihren Wirtsorganismen zu entschlüsseln. Eine Vielzahl antiviraler Therapien wurden auf Grundlage dieses gesammelten Wissens gegen einige der bekannten humanen Erreger entwickelt, wie zum Beispiel gegen das lebensbedrohliche humane Immundefizienz-Virus (HIV) oder das Hepatitis C Virus (HCV). Bei der Entwicklung antiviraler Medikamente werden zunächst strukturelle Makromoleküle identifiziert und charakterisiert, die für eine Virusinfektion essentiell sind. Mithilfe strukturbasierter Verfahren werden daraufhin Wirkstoffmoleküle entworfen, die zum Beispiel den Viruseintritt in die Wirtszelle verhindern oder virale Enzyme inhibieren. In dieser Arbeit werden zwei Aspekte der Wirkstoffforschung beleuchtet: zum einen Strategien zur Identifizierung viraler Zellrezeptoren und zum anderen das Screening von Fragmenten mittels Kernspinresonanzspektroskopie.

Im ersten Kapitel wird eine neue Methode zur Identifizierung von glycosylierten Zellrezeptoren behüllter Viren vorgestellt. Diese Strategie beinhaltet vier Schritte: Als erstes werden synthetische Monosaccharide, die bioorthogonale Alkingruppen tragen, mithilfe der zellulären Glycolysierungsmaschinerie in Glycoproteine lebender Zellen eingeschleust. Parallel werden virale Glycoproteine rekombinant hergestellt, deren Glycanketten mit Azidogruppen ausgestattet sind. Im dritten Schritt werden die modifizierten Wirtszellen mit viralen Glycoproteinen inkubiert. Durch eine mit Kupfer katalysierte Azido-Alkin-Cycloadditon (*Klickchemie*) werden virale Glycoproteine spezifisch mit glycosylierten Virusrezeptoren über die Glycanketten kovalent verknüpft. Anschließend werden die gebildeten Komplexe gereinigt und mithilfe immunologischer oder massenspektrometrischer Verfahren analysiert. Zur Etablierung dieser Methode wurde ein bekanntes Modellsystem bestehend aus MDBK-Zellen, welche den zellulären Rezeptor boviner Virusdiarrhoe-Viren (BVDV) exprimieren, und der löslichen Domäne von BVDV E_2 Proteinen verwendet. Dafür wurden mit Azidogruppen markierte E_2 Glycoproteine (E_2-Az) mittels einer humanen Zelllinie (HEK293T-Zellen) rekombinant hergestellt. Darüber hinaus lieferten Immunfluoreszenzanalysen einen Hinweis darauf, dass die Bildung kovalenter Bindungen zwischen E_2-Az und Alkingruppen der MDBK-Zellen mithilfe von *Klickchemie* realisiert werden kann. Außerdem deuteten Analysen mittels Western Blot auf eine erfolgreiche Isolierung möglicher E_2-Rezeptorkomplexe hin.

Der zweite Teil der Arbeit beschreibt das Screening von Fragmentbibliotheken auf Basis von Sättigungstransfer-Differenz NMR-Experimenten. Ziel war es, Fragmente zu identifizieren, die entweder mit der Westnilvirus (WNV)-Protease NS2B-NS3^{pro} oder mit Dimeren der sogenannten P Domäne von murinen (MNV-1) oder humanen Noroviren (Saga und MI001) wechselwirken. Zudem wurden positive Hits mittels Proteaseaktivitätsassays (WNV) oder Oberflächen-plasmonenresonanzspektroskopie (MNV-1) validiert. Des Weiteren erlaubten STD-NMR-Kompetitionsexperimente die Identifizierung von sekundär bindenden Fragmenten und potentiell allosterisch wirkenden Inhibitoren. Basierend auf 2D-NMR-Methoden von isotopenmarkierter WNV Protease oder Saga P Dimeren wurden sekundäre Bindungstaschen von einigen Liganden ermittelt. Bei der WNV Protease gaben Struktur-Wirkung-Beziehungen wichtige Aufschlüsse über funktionelle Gruppen, die für die Inhibierung von NS2B-NS3^{pro} essentiell sind. Darüber hinaus stellte sich molekulares Docking als wichtiges Instrument zur Interpretation experimenteller Daten heraus. Zum Beispiel wiesen Dockingmodelle der WNV Protease mit Analoga darauf hin, dass eine Carboxylgruppe, die in Richtung katalytischer Triade von NS2B-NS3^{pro} zeigt, vorhanden sein muss, damit Fragmente inhibitorisch wirken können. Im Fall von Norovirus P Dimeren wurden durch das virtuelle Screening von Fragmenten wichtige sekundäre Bindungstaschen aufgedeckt. Außerdem wurde ein direkter Vergleich zwischen experimentellen 2D-NMR-Daten und *in silico* Vorhersagen von Bindungstaschen gezogen. Bindungsmodelle, die auf Grundlage der *Emodel-Scoring*-Funktion ausgewählt wurden, zeigten eine gute Übereinstimmung mit experimentellen CSP-Daten. Diese Erkenntnis wurde verwendet, um Bindungsmodelle für Fragmente mit inhibitorischer Wirkung gegen MNV-1 zu erstellen.

Acknowledgements

I would like to express my deepest gratitude to Prof. Dr. Thomas Peters for the unlimited support during the last seven years at the University of Lübeck and for offering the possibility to work on exciting projects involving antiviral drug discovery. Moreover, I would like to thank for numerous suggestions, fruitful discussions, and the enthusiastic scientific environment including many stimulating conferences and symposia during my Ph.D. period.

Specials thanks are devoted to Prof. Dr. Norbert Tautz and his team from the Institute of Virology and Cell Biology, including Dr. Olaf Isken, Dr. Danilo Dubrau, and Barbara Bruhn for their assistance regarding the projects 'Viral receptor identification', which would not have been possible without them.

I very much thank Prof. Dr. Rolf Hilgenfeld and Dr. Linlin Zhang from the Institute of Biochemistry for their scientific support concerning the West Nile Virus protease, also for the opportunity to solve crystal structures at the German Electron Synchrotron Center (DESY, Hamburg).

For the supply of murine Norovirus proteins and viral stocks, I would like to thank Prof. Dr. Stefan Taube and Henrik Wegener from the Institute of Virology and Cell Biology. Also long-lasting discussions and valuable suggestions made this Ph.D. thesis conceivable.

I would like to acknowledge Prof. Dr. Jeanette Erdmann, Dr. Stephanie Tennstedt, and Michaela Müller (M.Sc.) from the Institute for Cardiogenetics for providing access to the Docking Software Glide.

In particular, I would like to thank Dr. Alvaro Mallagaray as a colleague and as a friend. He provided not only valuable chemicals and biomolecules but also helped me in many concerns including the correction of this manuscript. Furthermore, his inspiring and encouraging personality made him to an integral part of this Ph.D. thesis.

Also many colleagues and former members of the Institute of Chemistry at the University of Lübeck deserve my heartful thanks for contributing to this work. Dr. Hannelore Peters taught me in a pleasant way how to manage lab work. Specials thanks are devoted to Dr. Thorsten Biet for the help to understand the principles of NMR spectroscopy and for his assistance in practice. I thank PD Dr. Thomas Weimar and Thies Köhli for the support with the operation of the Biacore instrument. I owe sincere thanks to my fellow compaigners, Dr. Friedemann Flügge and Barbara Richer (M.Sc.), for their fruitful scientific discussions and for their contribution to an enjoyable atmosphere. Many others had been contributed to this thesis or had been part of my scientific education, including Richarda Marschall (M.Sc.), Fabienne Irrgang (B.Sc.), Christoph Müller-Hermes (B.Sc.), Robert Creutznacher (M.Sc.), Dr. Mailin Döpkens, Dr. Lena Lisbeth Grimm, Dr. Rosemarie Pulz, Dr. Karsten Seeger, and Dr. Christoph Rademacher.

I am in debt to the SILVER project, DZIF and Virocarb for financial support of my Ph.D. thesis on the antiviral drug discovery.

Last but not least, I owe my heartful thanks to my family and my love Marlene for their immeasurable encouragement, love, patience, and faith in me.

Content

Abstract 6
Zusammenfassung 8
Acknowledgements 10
List of Figures 17
List of Tables 20
Abbreviations and Symbols 21
Chapter 1 | Identification of Virus Receptors using Click Chemistry 25
1 Receptor Identification of Viruses using Click Chemistry 26
1.1 Introduction 26
1.1.1 The Search of Cellular Virus Receptors 26
1.1.2 Metabolic Labeling of Glycoconjugates using Modified Monosaccharides 27
1.1.3 The Bovine Viral Diarrhea Virus and its Receptor 28
1.2 Objectives 30
1.3 Material and Methods 31
1.3.1 Cell Culture 31
General Procedures 31
Production of MDBK-Az or MDBK-Alk cells 31
Production of E_2A and E_2A-Az Protein 31
1.3.2 Protein Analytics 32
Spectroscopic Protein Determination 32
SDS Polyacrylamide Gel Electrophoresis 33
Western Blotting 33
1.3.3 Indirect Immunofluorescence 34
1.3.4 Click Chemistry 34
1.3.5 Enrichment of CD46 Molecules using Click Chemistry 35
1.4 Results 37
1.4.1 Incorporation of Modified Monosaccharide Precursors 37
1.4.2 Production of the Modified BVDV Protein E_2A-Az 40
1.4.3 Click Reaction of E_2A-Az and MDBK-Alk Cells 42
1.4.4 Purification of CD46 Molecules and E_2A-Az-receptor Complexes 43
1.5 Discussion 46
1.6 Summary and Outlook 50

Chapter 2 | Fragment Screening Targeting Viral Proteins based on NMR Spectroscopy 53
1 Introduction - Fragment-based Lead Discovery 54
1.1 Nuclear Magnetic Resonance Spectroscopy in Drug Design 54
1.2 Complementary Validation Methods 56
1.3 *In silico* Docking 57
2 The West Nile Virus Protease 58
2.1 Introduction 58
2.1.1 The West Nile Virus – a Global Threat 58
2.1.2 Pathogenesis and Flaviviral Life Cycle 58
2.1.3 The WNV Protease – a Pivotal Drug Target 59
2.2 Objectives 61
2.3 Material and Methods 62
2.3.1 Production of the WNV Protease NS2B-NS3^{pro} 62
Unlabeled WNV Protease 62
Uniformly [^{2}H,^{15}N]-labeled WNV Protease 62
^{13}C-methyl-labeled WNV Protease 63
2.3.2 Protease Activity Assay 65
Salt and pH Dependency 65
Fragment Screening 65
Determination of IC_{50} values 66
Fluorogenic Competition Assay – Dixon Plot 66
2.3.3 Maybridge Libraries CR2006 and TS2017 67
2.3.4 Chemoinformatics – Database of Fragments 71
Tanimoto Coefficient – a Measure of Similarity 71
Molecular Descriptors 71
2.3.5 Descriptive Statistics of Fragments based on Molecular Descriptors 74
SxG Matrices 74
Principal Component Analysis 74
Boxplots and Histograms 75
2.3.6 Prediction of Fragment Mixtures 76
2.3.7 Nuclear Magnetic Resonance Spectroscopy 77
Proton Reference Spectra of Fragments 77
STD NMR Screening 77
STD NMR Competition Assays 78
STD NMR Titration 78

2D methyl-TROSY NMR Experiments 79
$^{1}H,^{15}N$-TROSY-HSQC NMR Experiments 79
2.3.8 Ligand Docking 80
2.4 Results 81
2.4.1 Fragment Screening based on STD NMR Experiments 81
2.4.2 Validation of the Binding Mode of CR112 and Analogs 85
STD NMR Titration and Fluorogenic Competition Assay 85
Identification of Binding Pockets using Chemical Shift Mapping 86
Structure-Activity Relationships and Molecular Docking 90
2.5 Discussion 95
2.5.1 Protein Biosynthesis 95
2.5.2 Evaluation of Activity and Binding Assays 95
2.5.3 Identification of Binding Sites 99
2.5.4 Hit-to-lead Optimization 100
2.6 Summary and Outlook 103
3 The Major Capsid Protein of Murine and Human Noroviruses 105
3.1 Introduction 105
3.1.1 Norovirus Genome Organization 105
3.1.2 Norovirus Capsid and Cell-entry 105
3.1.3 Current Drug Candidates against Noroviruses 107
3.2 Objectives 108
3.3 Material and Methods 109
3.3.1 Production and Purification of Norovirus P Dimers 109
Unlabeled MNV-1 P Dimers, Saga P Dimers, and MI001 P Dimers 109
Uniformly $[^{2}H,^{15}N]$-labeled Saga P Dimers 111
3.3.2 Protein Analytics – Dynamic Light Scattering 113
3.3.3 Lead Discovery based on NMR Spectroscopy 114
Fragment Screening based on STD NMR 114
Competition Assays using STD NMR 115
Chemical Shift Mapping of U-$[^{2}H,^{15}N]$-labeled Saga P Dimers 116
T_2 Determination using Spin-echo NMR Experiments 117
3.3.4 Surface Plasmon Resonance Spectroscopy 117
Adsorption to the Chip Matrix 117
Direct Immobilization of MNV-1 P Dimers using Amine Coupling 118
Validation of Positive Hits from STD NMR Screening 118

3.3.5 MNV-1 Inhibition Assay 119
3.3.6 Virtual Screening and Molecular Docking 121
3.4 Results 122
3.4.1 Fragment Screening 122
3.4.2 Dynamic Light Scattering – Aggregation of MNV-1 P Dimers 125
3.4.3 Hallmarks of Hits for Norovirus P Dimers 127
3.4.4 Evaluation of Hits for MNV-1 P Dimers 131
Surface Plasmon Resonance Spectroscopy 132
STD NMR Competition Experiments 133
Virus Inhibition Assays 135
3.4.5 Discovery of Novel Hits for Saga P Dimers 137
STD NMR Screening of the Maybridge TS2017 Library 137
STD NMR Competition Experiments 138
Chemical Shift Perturbation Experiments 139
3.4.6 Virtual Docking 145
Assessment of Binding Site Prediction for Saga P Dimers 145
Prediction of Binding Pockets of MNV-1 P Dimers 155
3.5 Discussion 159
3.5.1 Fragment Library Design 159
Molecular Size and Shape 159
Reactive Compounds and Promiscuous Binders 161
Structural and Chemical Diversity 163
Solubility in Solution 164
3.5.2 Fragment Screening 167
Aggregation of MNV-1 P Dimers 167
Hallmarks of Positive Hits 168
3.5.3 Inhibitors of Murine Noroviruses 168
3.5.4 Hits of Saga P Dimers 170
3.5.5 Virtual Screening 171
3.6 Summary and Outlook 174

Appendix 178
A1 Sequences and Vector Maps 178
Amino Acid Sequence of Bovine eE_2 + Protein A tag 178
Sequence of Bovine CD46 (Uniprot Q6VE48) 178
WNV Protease 179
MNV-1 P Dimer 180
Saga P Dimer 181
MI001 P Dimer 182
Protein Sequence Alignment using the ExPAsy Web Tool 183
A2 Maybridge Library, Fragment Cluster and STD NMR/CSP hits 185
Maybridge CR2006 Library 185
CR2006 Fragments: Scaffold Cluster S10 – Complex Structures 195
Maybridge TS2017 Library 196
Dendrogram of combined Maybridge CR2006/TS2017 Library 201
Mixtures for Fragment Screening 202
Compounds showing direct Irradiation 203
List of STD NMR hits for WNV protease 206
Twelve Most Potent Inhibitors of NS2B-$NS3^{pro}$ 208
List of STD NMR hits for Norovirus Screening Targets 209
Saga P Dimers - Comparison of STD NMR Competition Assays and CSP Experiments 213
Dendrogram of Fragments that Bind to MNV-1 P Dimers 214
Dendrogram of Fragments that Bind to Saga P Dimers 216
Dendrogram of Fragments that Bind to MI001 P Dimers 218
A3 2D Projections of 3D Plots 219
PCA – STD NMR Hits of Norovirus P Dimers 219
PCA – Soluble and Poorly Soluble Fragments 220
3D Histograms- STD NMR Hits of Norovirus P dimers 221
3D Histograms – Fragments of the Maybridge Libraries 222
A4 Python, Matlab, and Maestro Scripts 223
calcChemParam_Library.py 223
Matlab: Dendrogram 228
Matlab: Principal component analysis 228
grepPeaksfrom_XML_to_EXCEL.py 229
makeTanimotoMatrix.py 230
poolCompounds_2014.py 232

SMILES Definition according to Hann et al. 239
makeBinsforScreening_2017.py 240
Solubility_Boxplot_and_Histogram.py 247
Boxplot_and_Histogram_STD_Hits.py 249
searchParameter.py 252
3D_histo _for_A_D_Q.py 253
3D_histo _plot_for_A_vdW_Q.py 256
Correlation_Analysis_Heatmap.py 260
A5 NMR Pulse Programs and AU Programs 261
Zgesgp 261
Stddiffesgp.3 263
AU program: Stdproc_2 265
Trosyetf3gpsi.2 267
Metr-hmqcph.tb 272
A6 SPR Experiments 275
Adsorption Experiments 275
Sensograms 275
Biacore Methods 277
A7 Chemical Shift Perturbation Experiments 278
Assignment of $^{1}H,^{15}N$-TROSY-HSQC Spectrum of NS2B-NS3^{pro} based on BMRB Entry 11053 278
$^{1}H,^{15}N$-TROSY-HSQC spectra of NS2B-NS3^{pro} in Complex with Different Fragments 279
Assignment of U-[$^{2}H,^{15}N$]-labeled Saga P Dimer and TROSY-HSQC Spectra 284
A8 Docking and Virtual Screening 304
WNV Protease Screening 304
Norovirus P Dimers Screening 305
A9 Lists of Chemicals, Materials, and Equipments 307
Bibliography 312
Curriculum Vitae 325

List of Figures

Figure 1.1 | Metabolic labeling of glycan structures using chemically modified monosaccharide precursors 28
Figure 1.2 | Structural model of the interaction of E_2^{BVDV} and its viral receptor $CD46^{bov}$ 29
Figure 1.3 | Virus receptor identification using metabolic labeling combined with *click chemistry* 30
Figure 1.4 | Detection of glycoconjugates of MDBK cells using metabolic labeling and *click chemistry* 38
Figure 1.5 | Time course of the click reaction using THPTA as catalyst 38
Figure 1.6 | Optimization of the click reaction using BTTAA as a catalyst 39
Figure 1.7 | Labelling of surface glycoproteins of MDBK and Huh7 cells using modified monosaccharides 40
Figure 1.8 | Production and purification of the viral protein E_2 fused to a protein A tag 41
Figure 1.9 | Evaluation of the biological activity of the BVDV ectodomain E_2A and E_2A-Az by virus titer determination 41
Figure 1.10 | Covalent coupling of E_2A-Az on MDBK-Alk cells induced by click chemistry 42
Figure 1.11 | E_2A-Az concentration-dependent azido-alkyne cycloaddition via glycan chains 43
Figure 1.12 | First indications of the enrichment of CD46 molecules using ligand-specific *click chemistry* 44
Figure 1.13 | Validation of the covalent coupling of E_2A-Az with glycan structures of MDBK-Alk cells 47
Figure 1.14 | Evaluation of purifying E2A-receptor complexes from MDBK cells 48
Figure 1.1 | Schematic representation of the STD NMR experiments 55
Figure 1.2 | Illustration of complementary methodologies used in this study to confirm positive hits 56
Figure 2.1 | Organization and polyprotein processing of WNV proteins 59
Figure 2.2 | Two putative conformational states of the WNV protease NS2B-$NS3^{pro}$ 60
Figure 2.3 | (A) Hydrogen-deuterium exchange of α-2-keto-3-H_2-4-^{13}C-butyrate and (B) racemic mixture of α-ketoisovalerate (3-methyl-^{13}C; 3,4,4,4-D_4) 63
Figure 2.4 | Biosynthesis of ^{13}C-methyl-ILV-labeled NS2B-$NS3^{pro}$ 64
Figure 2.5 | Schematic representation of substrate processing by NS2B-$NS3^{pro}$ in pocket P1' 66
Figure 2.6 | SxG distribution matrix of the Maybridge CR2006 fragment library 67
Figure 2.7 | Dendrogram of 432 fragments of the Maybridge CR2006 library 68
Figure 2.8 | SXG distribution matrix of the Maybridge TS2017 fragment library 69
Figure 2.9 | Dendrogram of 238 fragments of the Maybridge TS2017 library 70
Figure 2.10 | SxG performance matrix of WNV protease binders identified by STD NMR 82
Figure 2.11 | Characterization of positive hits derived from the WNV protease screening based on STD NMR 83
Figure 2.12 | Fragment screening and STD NMR competition assay 84
Figure 2.13 | Fragment CR112 is a non-competitive inhibitor of the WNV-protease NS2B-$NS3^{pro}$ 85
Figure 2.14 | Chemical shift perturbations of ^{13}C-methyl-labeled NS2B-$NS3^{pro}$ upon ligand binding 86
Figure 2.15 | Production of the WNP protease in M9+/H_2O medium vs. M9+/D_2O medium 87
Figure 2.16 | Observation of a pH artefact during a chemical shift perturbation (CSP) analysis 88
Figure 2.17 | Chemical shift perturbation of the WNV protease upon ligand binding 89
Figure 2.18 | Structure-activity-relationship (SAR) of structurally similar fragments, which bind to NS2B-$NS3^{pro}$ 90
Figure 2.19 | Docking of fragments from the SAR analysis into corresponding putative binding sites 92
Figure 2.20 | Promiscuous binding of CR402 illustrated by computational docking 93
Figure 2.21 | Influence of buffer conditions on the protease activity 96
Figure 2.22 | Quenching of AMC fluorescence in the presence of different fragments 98
Figure 2.23 | Distribution of ILV residues throughout the WNV protease NS2B-$NS3^{pro}$ 100
Figure 2.24 | Docking of CR112 analogs that carry a carboxyl group into a secondary binding site of NS2B-$NS3^{pro}$ 101
Figure 3.1 | The Norovirus capsid and crystal structures of P dimers of two different Genotypes 106
Figure 3.2 | Biosynthesis and purification of MNV-1 P dimers 110
Figure 3.3 | Growth curve of E.coli strain BL21 (DE3) expressing the P domain of humane Norovirus strain GII.4 Saga 111
Figure 3.4 | Biosynthesis and purification of Saga P dimers 112
Figure 3.5 | Overview of the screening process 122
Figure 3.6 | Illustration of the STD NMR-based fragment screening for one mixture 123

Figure 3.7 | STD NMR-based fragment screening against different Norovirus P dimers 123
Figure 3.8 | Estimation of the molecular weight of P dimers using 1D 1-1-echo NMR experiments 124
Figure 3.9 | Quality control of MNV-1 P dimer production by dynamic light scattering 125
Figure 3.10 | Time course of MNV-1 P dimer aggregation at high ionic strength 126
Figure 3.11 | Aggregation of MNV-1 P dimers analyzed by dynamic light scattering 127
Figure 3.12 | SxG matrices of fragments, which bind to different Norovirus P dimers 128
Figure 3.13 | Comparison of binders and non-binders of Norovirus targets using principal component analysis 129
Figure 3.14 | 1D molecular descriptors of hits identified by STD NMR 130
Figure 3.15 | 3D histograms of informative scalar descriptors of STD NMR positive hits 131
Figure 3.16 | Validation of STD NMR hits of MNV-1 P dimers by using SPR 132
Figure 3.17 | Descriptive statistics of SPR hits of MNV-1 P dimers 133
Figure 3.18 | STD NMR competition experiments reveal 19 putative competitors 134
Figure 3.19 | MNV-1 inhibition screen of fragments 135
Figure 3.20 | Cell viability test 136
Figure 3.21 | Characterization of STD NMR hits of Saga P dimers 137
Figure 3.22 | Analysis of competition for fucose binding sites of the Saga P dimer using STD NMR as read-out 138
Figure 3.23 | Chemical shift perturbation (CSP) upon binding of CR065 140
Figure 3.24 | Chemical shift perturbation (CSP) of fragments, which bind to Saga P dimers 143
Figure 3.25 | Evaluation of chemical shift perturbation (CSP) experiments 144
Figure 3.26 | Binding pocket prediction based on chemical shift perturbation experiments 145
Figure 3.27 | Virtual screening of the Maybridge fragment libraries CR2006 and TS2017 targeting the Saga P dimer 146
Figure 3.28 | Docking of α-L-fucose into different receptor grids of the Saga P dimer 147
Figure 3.29 | Docking of CR065 into putative binding sites of the Saga P dimer 148
Figure 3.30 | Comparison of chemical shift perturbation experiments and docking results 149
Figure 3.31 | Virtual screening of fragments of the Maybridge libraries CR2006 and TS2017 against MNV-1 P dimers 155
Figure 3.32 | Docking of 3'-sialyllactose and CR120 into different binding sites of the MNV-1 P dimer 156
Figure 3.33 | 2D ligand interaction diagrams of fragments docked into several binding sites of the MNV-1 P dimer 157
Figure 3.34 | Distribution of molecular shapes in the fragment libraries CR2006 and TS2017 160
Figure 3.35 | Property distribution of fragment libraries 164
Figure 3.36 | Principal component analysis of structural and chemophysiological properties of fragments 165
Figure 3.37 |The partition coefficient log P is a reasonable predictor of solubility 166
Figure 3.38 | Docking of H type 2 trisaccharide into the canonical glycan binding site of the Saga P dimer 173

Appendix

Figure A1 | Dendrogram of 657 water-soluble fragments 201
Figure A2 | Fragments exhibiting the highest inhibitory potency against the WNV protease 208
Figure A3 | STD NMR Hits of MNV-1 P dimer 214
Figure A4 | Positive hits from SPR screening 215
Figure A5 | Dendrogram of CR2006 fragments binding to Saga P dimers 216
Figure A6 | Dendrogram of TS2017 fragments binding to Saga P dimers 217
Figure A7 | Dendrogram of CR2006 fragments binding to MI001 P dimers 218
Figure A8 | Comparison of Hits based on the principal component analysis 219
Figure A9 | PCA of the Solubility of Fragments 220
Figure A10 | Comparison of hits from the Norovirus screening based on structural properties 221
Figure A11 | Fragment selection for a library design 222
Figure A12 | Adsorption of MNV-1 P dimers to the chip matrix depending on pH and protein concentration 275
Figure A13 | SPR sensogram of the immobilization of MNV-1 using the amine coupling methods 275
Figure A14 | SPR Sensogram of 3'-sialyllactose titration 276
Figure A15 | Exemplary SPR sensograms of different fragments 276
Figure A16 | Assignment of $^{1}H,^{15}N$-TROSY-HSQC spectrum of NS2B-NS3^{pro} 278
Figure A17 | Chemical shift perturbations of uniformly ^{15}N-labeled NS2B-NS3^{pro} upon binding of CR112 279
Figure A18 | Chemical shift perturbations of uniformly ^{15}N-labeled NS2B-NS3^{pro} upon binding of CR113 or CR114 280
Figure A19 | CR081 and CR069 (Figure A18 continued) 281
Figure A20 | CR082 and CR402 (Figure A18 continued) 282
Figure A21 | CR116 and CR053 (Figure A18 continued) 283
Figure A22 | Assignment of uniformly [$^{2}H,^{15}N$]-labeled Saga P dimer 284
Figure A23 | Chemical shift perturbations upon pH differences 285
Figure A24 | Chemical shift perturbation upon different sodium chloride concentration 286
Figure A25 | Chemical shift perturbations of uniformly ^{15}N-labeled Saga P dimers upon binding of CR001 and CR025 287
Figure A26 | CR039 and CR054 (Figure A25 continued) 288
Figure A27 | CR055 and CR065 (Figure A25 continued) 289
Figure A28 | CR066 and CR070 (Figure A25 continued) 290
Figure A29 | CR100 and CR106 (Figure A25 continued) 291
Figure A30 | CR120 and CR136 (Figure A25 continued) 292
Figure A31 | CR196 and CR215 (Figure A25 continued) 293
Figure A32 | CR251 and CR254 (Figure A25 continued) 294
Figure A33 | CR311 and CR365 (Figure A25 continued) 295
Figure A34 | CR367 and CR396 (Figure A25 continued) 296
Figure A35 | CR440 and TS012 (Figure A25 continued) 297
Figure A36 | TS026 and TS039 (Figure A25 continued) 298
Figure A37 | TS040 and TS050 (Figure A25 continued) 299
Figure A38 | TS074 and TS098 (Figure A25 continued) 300
Figure A39 | TS124 and TS153 (Figure A25 continued) 301
Figure A40 | TS162 and TS188 (Figure A25 continued) 302
Figure A41 | TS189 and TS207 (Figure A25 continued) 303
Figure A42 | Docking of SAR fragments into binding pocket Q167 304
Figure A43 | Receptor grids of Saga P dimer used in the virtual screening 305
Figure A44 | Receptor grids of MNV-1 P dimer used in the virtual screening 306

List of Tables

Table 1.1 | Composition of SDS-PAGE gels 33
Table 1.2 | Composition of 3-fold SDS buffer 33
Table 2.1 | Composition of the optimized minimal medium M9+ 64
Table 2.2 | Lipinski's Ro5 parameter of the Maybridge library CR2006 73
Table 2.3 | Lipinski's Ro5 parameter of the Maybridge library TS2017 73
Table 2.4 | Structure-activity relationship of mainly TS2017 fragments 94
Table 3.1 | Composition of the minimal medium M9/D_2O 113
Table 3.2 | Summary of MNV-1 inhibitors 136
Table 3.3 | Structure-activity relationship of MNV-1 inhibitors 136
Table 3.4 | Comparison of empirically determined binding sites and molecular docking 154
Table 3.5 | Prediction of fragment binding sites of MNV-1 P dimers based on docking 157
Table 3.6 | List of favorable chemical groups (adapted from REF[292,293]) 161

Appendix

Table A1 | Maybridge library CR2006 – fragment mixtures 202
Table A2 | Maybridge library TS2017– fragment mixtures 202
Table A3 | List of positive hits from STD NMR screeing targeting the WNV protease 206
Table A4 | List of positive hits from STD NMR screeing targeting NoV P dimers 209
Table A5 | List of positive hits from STD NMR screeing targeting Saga P dimers 212
Table A6 | Evaluation of STD NMR competition assays of Saga P dimers 213
Table A7 | Position and dimensions of receptor grids of the Saga P dimer 305
Table A8 | Position and dimensions of receptor grids of the MNV-1 P dimer 306
Table A9 | Dihedral angles of 3'-sialyllactose docked to the MNV-1 P dimer 306

Abbreviations and Symbols

A_{280}	Molar absorption at 280 nm
Ac_4GalNAlk	per-*O*-acetylated *N*-(pent-4-yn-1-one-yl)-galactosamine/
	Synonym**:** per-*O*-acetylated *N*-alkynoacetylgalactosamine
Ac_4GalNAz	per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-galactosamine
	Synonym**:** per-*O*-acetylated *N*-azidoacetylgalactosamine
Ac_4GlcNAlk	per-*O*-acetylated *N*-(pent-4-yn-1-one-yl)-glucosamine
	Synonym**:** per-*O*-acetylated *N*-alkynoacetylglucosamine
Ac_4GlcNAz	per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-glucosamine
	Synonym**:** per-*O*-acetylated *N*-azidoacetylglucosamine
A_{c4}ManNAlk	per-*O*-acetylated *N*-(pent-4-yn-1-one-yl)-mannosamine
	Synonym**:** per-*O*-acetylated *N*-alkyneacetylmannosamine
Ac_4ManNAz	per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-mannosamine
	Synonym**:** per-*O*-acetylated *N*-azidoacetylmannosamine
Alk	Alkyne/alkyno group
Az	Azido group
BSA	Bovine Serum Albumin
BVDV	Bovine Viral Diarrhea Viruses
CD46	Cluster of Differentiation 46
CHO cells	Chinese Hamster Ovary cells
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSP	Chemical Shift Perturbation
CuAAC	Copper-catalyzed Azido-Alkyne Cycloaddition
DENV	Dengue Viruses
DLS	Dynamic Light Scattering
E_2	Envelope protein E_2
ECL	Enhanced chemiluminescence
ER	Endoplasmic reticulum
FCS	Fetal Calf Serum
Fuc	Fucopyranose
g	Gravity
Gal	Galactopyranose
GalNAc	*N*-acetylgalactosamine (pyranose)
Glc	Glucopyranose
GlcNAc	*N*-acetylglucosamine (pyranose)
HBGA	Human Blood Group Antigen
H-bond A	Hydrogen bond Acceptor
H-bond D	Hydrogen bond Donor

HCV	Hepatitis C Viruses
HEK293T cells	Transfected Human Embryonic Kidney Cells
Hela cells	Human epithelial cells
HIV	Human Immunodeficiency Viruses
HSQC	Heteronuclear single quantum correlation
HTS	High-throughput
Huh7 cells	Human hepato cellular carcinoma cells
IC_{50}	Half maximal inhibitory concentration
IF	Immunofluorescence
IgG	Immunoglobulin G
K_D	Dissociation constant
kDa	Kilo Dalton, units of the molecular weight, 1 Da = 1 g/mol
K_i	Inhibition constant
LE	Ligand efficiency
LLE	Lipophilic ligand efficiency
log D	distribution coefficient
log P	partition coefficient
ManNAc	*N*-acetylmannosamine (pyranose)
MDBK cells	Madin-Darby Bovine kidney epithelial cells
MNV-1	Murine Norovirus 1
mol	Molar = mol/L
MR	Molecular refractivity
MS	Mass Spectrometry
MSA	molecular surface area
MW	molecular weight in Dalton
NMR	Nuclear Magnetic Resonance spectroscopy
NoV	Norovirus
NS	Non-structural
OD	Optical Density
ORF	Open Reading Frame
P dimer	Dimer of the protruding domain of NoV
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered Saline
PCA	Principal Component Analysis
PDB	Protein Data Base
pH	potential of hydrogen measured in H_2O samples
pH*	potential of hydrogen measured in D_2O samples
PNGase F	Peptide-N-Glycosidase F
ppm	Parts per million

ProtA	Protein A
PSA	Polar surface area
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
RU	Response Units, 1 RU = 1 pg/mm^2
SAR	Structure-Activity Relationship
SARS	Severe Acute Respiratory Syndrome
SDS PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophorese
SMILES	Simplified Molecular Input Line Entry System
SPR	Surface Plasmon Resonance spectroscopy
SR-B1	Scavenger receptor class B type 1
STD	Saturation Transfer Difference
SxG	Scaffold x Group
T_1	Longitudinal relaxation time
T_2	Transversal relaxation time
TAN	Tanimoto coefficient
TROSY	Transverse Relaxation-Optimized Spectroscopy
vdW	Van der Waals
VLP	Virus-like Particle
VP1	Major Capsid Protein of Caliciviruses (Viral Protein 1)
WASA	water accessible surface area
WB	Western Blotting
WNV	West Nile Virus
α-	Anti-
τ_c	Correlation time

Amino acids

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartate	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamate	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Chapter 1 | Identification of Virus Receptors using Click Chemistry

1 Receptor Identification of Viruses using Click Chemistry

1.1 Introduction

In general, the life cycle of a common eukaryotic virus involves five stages: cell attachment, cell entry, viral replication, maturation, and release[1–3]. In order to allow a successful infection, viruses exploit fundamental cellular processes to invade their cellular hosts. The interplay between viruses and their cellular receptors largely defines the virus entry pathways (e.g., endocytosis and membrane fusion). In many studies, researchers have demonstrated that initial encounters between a virus and its host cell are mediated through cellular attachment factors, such as heparan sulfate[4–6], histo blood group antigens[7–11], or sialic acid-containing glycoconjugates[12–16]. In a second step, viruses interact with specific receptors that drive the reactions leading to cell entry. During the last decades, various cell surface components have been identified for several viruses (cf. REF[3] and reference therein). Many viruses exploit only one cell receptor to allow infection, e.g., intercellular adhesion molecule 1 (ICAM-1) for Rhinovirus A/B or CD155 for poliovirus. In other cases, more than one molecular receptor has been identified, e.g., angiotensin-converting enzyme (ACE) or liver-SIGN (L-SIGN) for SARS coronavirus. Moreover, some well-studied viruses, such as HIV (human immunodeficiency virus) or HCV (hepatitis C virus) require engagement with at least two distinct plasma membrane components in parallel. HIV, for example, binds to the primary receptor CD4 and CCR5 or CXCR4 as a coreceptor in a strain-specific manner. The interaction with the coreceptor mediates structural changes of HIV envelope proteins that promote membrane fusion and hence cell entry. In the case of HCV, coexpression of four proteins, CD81, SR-B1, claudin-1, and occludin, is essential to confer permissivity to non-susceptible cells for HCV entry. However, it is believed that claudin-1 and occludin may act as a regulator for CD81 and/ or SR-B1 rather than binding viruses directly[3]. The use of multiple receptors has also been reported for other viruses, including adenoviruses[17], rotaviruses[18], picornaviruses[19], and herpesviruses[20]. Nevertheless, relatively little is known about the stoichiometry or assemble mechanisms towards multimeric receptor complexes.

1.1.1 The Search of Cellular Virus Receptors

A critical first step in developing novel and effective entry inhibitors is the identification of the cellular receptor involved in virus-host cell interactions. In general, there are three different strategies: the immunological, the genetic, and the chemical approach. In the first approach, neutralizing monoclonal antibodies are screened against antigens exposed by a susceptible host cell[21]. Those antibodies that are capable of reducing the viral infectivity are chosen to conduct immunoprecipitation followed by amino acid sequencing of the enriched antigen. Currently, direct co-precipitation with viral proteins has also been described[22]. Another semi-targeted approach is the enrichment of detergent-solubilized membrane glycoconjugates by lentil lectin chromatographic purification and subsequent virus overlay protein blot assays. The putative glycoprotein candidates are proteolyzed by trypsin and subjected to mass spectrometry to identify

peptides that are present in a data base[23]. In both cases, however, viral inhibition assays are required in the presence of specific antibodies to validate putative virus receptors.
On the genetic level, two screening approaches have been employed so far. Briefly, either non-susceptible receptor-negative cells are equipped with virus receptors by using cDNA expression libraries from receptor-bearing cells[24] or several genes encoding for putative receptors are successively removed using, for example, a genome-wide CRISPR/Cas9 guide RNA library[25]. In the end, viral infectivity assays are subsequently conducted to identify promising candidates.
Alternative chemical strategies involve the formation of a covalent linkage between viral components and cellular receptors. One recent method is a ligand-based, receptor capture technology that uses a chemoproteomic reagent termed TRICEPS[26]. This reagent consists of three chemical moieties, including an N-hydroxysuccinimide ester for the coupling to amino groups of ligands (e.g. viral proteins), a trifluoroacetyl-protected hydrazine moiety to capture oxidized aldehyde-containing glycoprotein receptors, and a biotin group for affinity purification. Another approach using diazirine photo-cross-linking has also been proposed for target identification[27]. In this study, an alternative method has been explored, which is based on metabolic labeling of glycan moieties of both viral proteins and cellular receptors followed by so-called *click chemistry* (see below). As a proof of principle, the ectodomain of the E_2 protein of bovine viral diarrhea viruses (BVDV) and its known viral receptor CD46 have been utilized.

1.1.2 Metabolic Labeling of Glycoconjugates using Modified Monosaccharides

About three decades ago, it has been demonstrated for the first time that unphysiological mannosamine derivatives are incorporated into glycan moieties in living rats[28]. Since then various unnatural monosaccharide precursors have been designed and utilized to study, for example, virus-receptor interactions[29–31], cellular processes[32,33], and to probe glycoproteins[34–37].
Later on, Bertozzi and co-workers have established a bioorthogonal chemical reporter strategy in 2007, which involves metabolic labeling of glycans of cultured cells or living animals with synthetic monosaccharide precursors that are functionalized with an azido group (cf. REF[38] and references therein). This bioorthogonal functional group can react with specific reagents, such as phosphines and alkynes. Hence, metabolic labeling followed by chemical labeling permits direct detection of glycoconjugates and glycan-specific enrichment using functionalized fluorophores or affinity tags, respectively. In this present work, glucosamine, galactosamine, and mannosamine derivatives bearing an azido[38] or alkyne[39] group were utilized to incorporate biorthogonal chemical functionalities into glycoconjugates (see **Figure 1.1**). For the metabolic labeling of cultured cells, per-*O*-acetylated monosaccharide precursors that are simply added to the cell culture medium passively diffuse across the cell membrane. Following deacetylation by cytosolic esterases, relatively promiscuous biosynthetic pathways allow the incorporation into cytosolic and transmembrane glycoproteins as well as other glycoconjugates, such as glycolipids[38].
GalNAz/Alk (see **Figure 1.1**) is thought to selectively label nuclear and cytoplasmic glycoproteins via β-*O*-glycosylation. This precursor may also be incorporated into mucin-type *O*-linked glycans

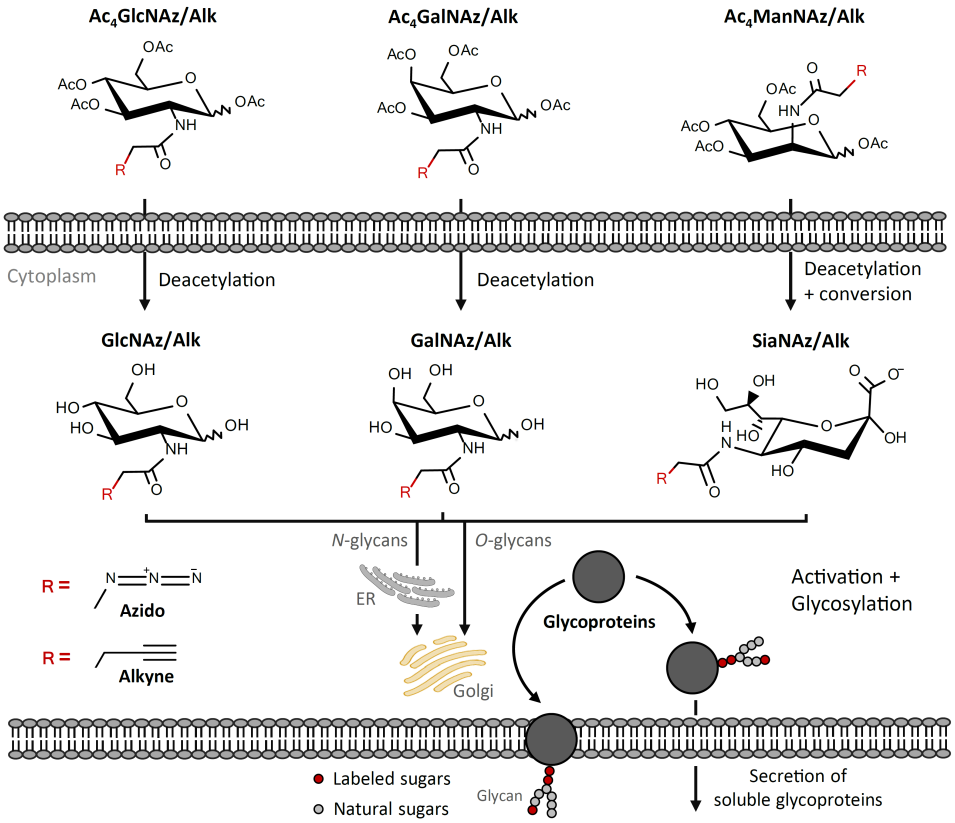


Figure 1.1 | Metabolic labeling of glycan structures using chemically modified monosaccharide precursors
The per-*O*-Acetylated (Ac) sugars carrying either an alkyne (Alk) or azido (Az) group passively diffuse across the cell membrane. After deacetylation by promiscuous intracellular esterases, and conversion in the case of *N*-acetyl-mannosamine derivatives towards sialic acid derivatives, modified sugars are activated and incorporated into glycoconjugates through the cellular machinery including the endoplasmic reticulum (ER) and the Golgi apparatus. The following sugars were used throughout this study: per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-glucosamine (Ac_4GlcNAz), per-*O*-acetylated *N*-(pent-4-yn-1-one-yl)-glucosamine (Ac_4GlcNAlk), per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-galactosamine (Ac_4GalNAz), per-*O*-acetylated *N*-(pent-4-yn-1-one-yl)-galactosamine (Ac_4GalNAlk), per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-mannosamine (Ac_4ManNAz), and per-*O*-acetylated *N*-(pent-4-yn-1-one-yl)-mannosamine (Ac_4ManNAlk).

as a substitute for GalNAc[40]. Moreover, it has been reported that GalNAz/Alk could be partially interconverted to GlcNAz/Alk *in vivo,* which in turn can be theoretically part of the pentasaccharide core, $GlcNAc_2Man_3$, of *N*-linked glycans. GalNAz/Alk may also be incorporated into certain glycolipids and as a terminal modification of rare *N*-linked glycans[38]. Moreover, it has been demonstrated that ManNAz is converted metabolically to SiaNAz[41] (N-azidoacetyl sialic acid) whose natural counterpart is a frequent terminal monosaccharide found on glycoproteins and glycolipids (e.g., gangliosides)[42].

1.1.3 The Bovine Viral Diarrhea Virus and its Receptor

The animal pathogen BVDV belongs to the family *Flaviviridae* along with, e.g. human HCV. The enveloped RNA viruses consist of three envelope glycoproteins, namely E^{rns}, E_1, and E_2 (cf. REF[43] and reference therein). All three glycoproteins form disulfide-linked dimers in the membrane of BVDV: the E^{rns} homodimer, the E_1-E_2 heterodimer, and E_2 homodimer. The major glycoprotein E_2

is the most important structural protein as it interacts with cell surface receptors and harbors the neutralizing antibody epitope. Whereas the E_1-E_2 heterodimer is essential for cell entry and infectivity. In addition, E^{rns} has been proposed to serve as an accessory protein that interacts with other cell surface molecules, such as heparin sulfate[43,44].

Altered glycosylation patterns of E^{rns}, E_1, and E_2 affect the virulence of the closely related swine fever virus. The absence of specific N-linked glycan sites leads to virus attenuation[43]. In the case of BVDV, it has been reported that the loss of two glycan sites resulted in a reduced capacity of recombinant E_2 proteins to inhibit virus infection[45], demonstrating the importance of glycans for receptor binding. The bovine CD46 is the primary cellular receptor of BVDV[21], but unknown coreceptors/cofactors seem to be also important for complete viral infectivity[21,46]. The low-density lipoprotein receptor is one candidate[47].There are indications that BVDV enters the host cell through clathrin-mediated endocytosis and CD46 is not significantly internalized[48,49]. Therefore postattachment internalization is required[50]. A structural model of E_2^{BVDV}-$CD46^{bov}$ binding is presented in **Figure 1.2**. The real binding mode and membrane fusion mechanism remain elusive[48]. However, an unknown coreceptor and large rearrangements are assumed, as E_1-E_2 heterodimers must reach the cellular membrane to initiate the fusion process. During this process, it is very likely that glycan chains come across, which would be necessary for an efficient click reaction.

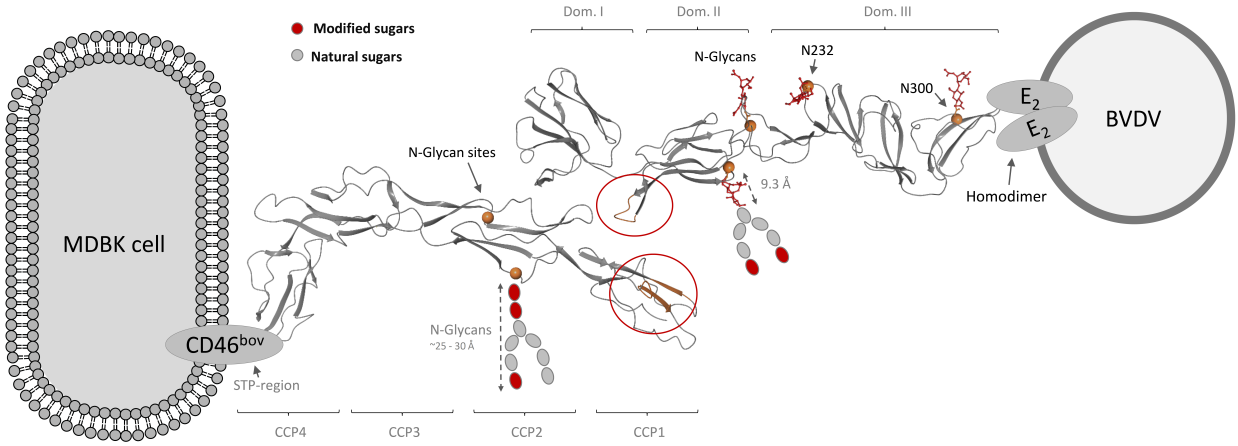


Figure 1.2 | Structural model of the interaction of E_2^{BVDV} and its viral receptor $CD46^{bov}$
The bovine CD46 receptor consists of four single complement control protein modules (CCPs). Differential splicing (apparent molecular weight of ~50 kDa and ~90 kDa) determines the degree of *O*-glycosylation in the STP region (rich in serine, threonine, and proline)[44]. The homology model comprising CPP1 to CPP4 was built using Swiss model[51] based on human CD46. The *N*-glycan sites are represented by orange spheres based on UniProt Q6VE48 and REF[21]. The putative binding site for BVDV including the peptide sequences $E_{66}QIV_{69}$ and $G_{82}QVLAL_{87}$ (orange) are highlighted (red circles). In the case of the E_2 protein (PDB: 4JNT) of BVDV containing domain I to III, an exposed β-hairpin is proposed as interaction site[48]. *N*-glycans are represented in red stick models. Only the first two monosaccharides (GlcNAc-GlcNAc) are resolved. Gray and red spheres illustrate a typical core of *N*-glycans. The glycans linked to residues N232 and N300 seem to be an essential regulator of E_2 protein function[45]. The ectodomains of E_2 forms disulfide-bridged homodimers (in the viral membrane and in solution). For simplicity, domain I to III of the second E_2 molecule are not shown. (MDBK, Madin-Darby Bovine kidney epithelial cells used in this study; BVDV, bovine viral diarrhea virus)

1.2 Objectives

In this work, I aimed to develop a novel strategy for virus receptor identification by using metabolic labeling and *click chemistry* as illustrated in **Figure 1.3**. As a proof of principle**,** the ectodomain of the E2 protein of bovine viral diarrhea viruses (BVDV) and MDBK cells bearing the cellular receptor $CD46^{bov}$ were utilized.

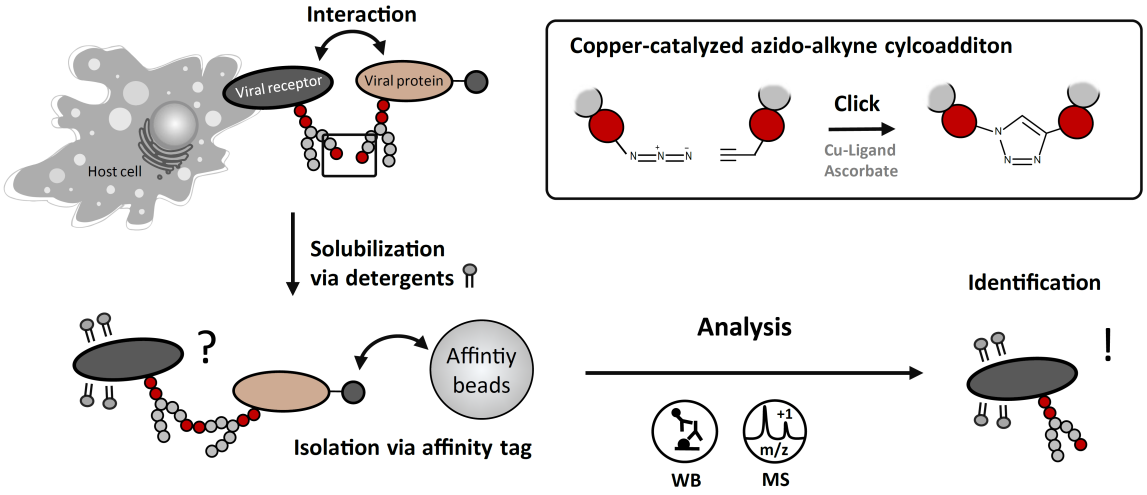


Figure 1.3 | Virus receptor identification using metabolic labeling combined with *click chemistry*
The affinity-tagged viral glycoprotein bearing, e.g. alkyne groups interacts with its cellular virus receptor that in turn carries azido groups or *vice versa*. Due to these specific interactions, the biorthogonal groups are in close contact, which promotes the copper-catalyzed azido-alkyne [4+2] cycloaddition (*click chemistry*). After the formation of a covalent linkage, the complex is purified and analyzed, for example, via western blotting (WB) or mass spectrometry (MS) following trypsin digestion. A proteomic database query leads to the identification of the viral receptor(s).

The following questions had to be answered.

- Is the incorporation of synthetic monosaccharide precursors into glycoconjugates of different cell types (MDBK and Huh7) sufficient?
- Which conditions are required to achieve an optimal click reaction?
- Is it feasible to purify biotinylated CD46-Alk molecules from cell lysate using streptavidin beads? Which detergents are suitable?

- Is it possible to synthesize Protein A-tagged E_2 proteins that carry azido groups (E_2A-Az)?
- Are these E_2A-Az proteins still biologically active and capable to inhibit BVDV infection?
- Which protocol is the best to enrich E_2A-Az proteins using IgG beads in a small-scale?
- How efficient is the covalent linkage of E_2A-Az and its receptor(s) that carry alkyne groups?
- Is the enrichment of E_2A-CD46 complexes using IgG beads sufficient to detect CD46 molecules by western blotting or mass spectrometry?

1.3 Material and Methods

1.3.1 Cell Culture

General Procedures

In general, cells were cultured at 37 °C and 5% CO_2 in either in 10 cm plates, T175 flasks or 6-/12-/24-well plates. The MDBK line (Madin-Darby Bovine kidney epithelial cells) and HEK293T line (transfected human embryonic kidney cells) were cultured in DMEM (Dulbecco's Modified Eagle medium) containing 1% penicillin/streptomycin (pen/strep) and 10% horse serum or 10% fetal calf serum (FCS), respectively. Prof. Dr. Tautz kindly provided the stable cell line HEK293T expressing permanently the protein E_2A. The expression protocol has been established by Susanne Luu in the lab of Abdul Ghafoor Khan at the center for advanced biotechnology and medicine, Rutgers University. The amino acid sequence of E_2A can be found in Appendix A1. Cells were quantified using a Neubauer count chamber. Buffer substances and pH adjusted cell media were sterile-filtered or autoclaved when possible.

Production of MDBK-Az or MDBK-Alk cells

MDBK cells were seeded in 6-/12-well or 10 cm cell culture plates to reach ~60 to 70% confluence, and incubated with an appropriate amount of DMEM (e.g. 10 ml for 10 cm dish) that was adjusted to pH 6.6 in the presence of 200 to 800 µM modified sugars carrying azido (Az) or alkyne groups (Alk) for 4 h – 6 h, if not otherwise stated. Following cell medium change against freshly prepared DMEM (pH 6.6) containing 200 to 800 µM modified sugars, the cells were cultured 16 h before further usage (see below *Click Chemistry*) to allow incorporation of modified sugars into glycan structures exposed on the cell surface. Huh7 cells were equally treated. The exact amount of monosaccharides is given when applied. The modified sugars (Ac_4GalNAz/Alk, Ac_4ManNAz/Alk, and Ac_4GlcNAz/Alk, see **Figure 1.1**) were synthesized in-house by Dr. Alvaro Mallagaray based on REFs[38,39] and 10 mM sterile-filtered stock solutions dissolved in water were stored at -20 °C. After thawing, the stock solutions were heated to 37 °C and sonified for 5 min, if necessary, to allow complete dissolving of acetylated sugars.

Production of E_2A and E_2A-Az Protein

A 1 ml frozen cell suspension of HEK293T cells (in liquid nitrogen) were thawed and diluted with 9 ml pre-heated cell medium (37 °C). Following centrifugation at 700 x g for 2 min, the cell pellet was resuspended in 10 cell medium. The cell culture was enlarged to obtain ten 175T cell culture flasks containing 30 ml cell medium (pH adjusted to 6.7 – 6.8) and approximately $2 \cdot 10^7$ cells each. The medium (~500 ml) containing E_2A proteins was filtered, supplemented with one pill of protease inhibitors (PI) mix (cOmplete, Roche), and stored at -20 °C until subsequent purification (see below). In order to produce E_2A-Az, 200 µM of a mixture of azido sugars (ratio of 7:2:1 Ac_4GalNAz, Ac_4ManNAz, and Ac_4GlcNAz) were periodically added five times for two days. On

the second day, the medium was removed and stored separately at -20 °C (E_2A-Az I). After addition of 30 ml freshly prepared medium (pH 6.7 – 6.8), 200 µM azido sugars were periodically added four times in a period of 1.5 days (E_2A-Az II). Next, the most viable cells were selected (five 175T flasks), divided into ten 175T flasks containing 25 ml standard cell medium, and incubated for one day to allow recovery. Next day, the medium was changed to pH 6.8 medium and 200 µM azido sugars were periodically added six times in a period of 2 days. Along with the supernatant obtained before, the cell medium was filtered, one pill of PI mix was added, and the medium was stored separately at -20 °C (E_2A-Az III).

The cell culture supernatants were separately purified using a 5 ml IgG Sepharose™ 6 Fast Flow column (or IgG beads), which had been washed with 25 ml TST buffer (50 mM Tris pH 7.6, 150 mM NaCl, and 0.05% (v/v) Tween-20. Of note, Tween-20 is incompatible with mass spectrometry), 50 ml Millipore water and 50 ml of 25 mM sodium phosphate buffer pH 7 (washing buffer). The filtered supernatant was passed through the column at a flow rate of 1 ml/min, and the column was rewashed with 60 ml washing buffer. Elution was conducted using 0.1 M glycine/HCl pH 3. Therefore, elution buffer was added at a flow rate of 1 ml/min and 1 ml fractions were collected in tubes containing 500 µl of 0.2 M Tris pH 9 or a volume of 19 ml flow-through was collected in a 50 ml tube containing 1ml of 1 M Tris pH 9 to neutralize the sample. Following 5x buffer exchange against E_2 buffer (25 mM HEPES pH 7.5, 250 NaCl, 5% glycerol), the protein yield was as follows: ~430µg E_2A from 500 ml, ~320 µg E_2A-Az I from 300 ml, ~540 µg E_2A-Az II from 300 ml, and ~1.46 mg E_2A-Az III from 300 ml cell culture supernatant. In total, 2.6 mg/L E_2A-Az can be expected. In a second approach, even 3.7 mg/L E_2A-Az could be obtained (data not shown and not part of this study). After quality control (see below), all E_2A-Az samples were stored either at 4 °C or at -20 °C until usage. Throughout this study, only E_2A-Az III was utilized.

The presence of *N*-glycans in E_2A-Az proteins was confirmed by PNGase F digestion. Therefore, 40µl of 2.4 µM E_2A-Az were first heated to 95 °C for 10 min. Next, 39 µl of 500 µM sodium phosphate buffer (pH 7.5) was added to cool down the protein solution. Finally, 1 µl of PNGase F stock solution (500,000 units/ml, New England BioLabs Inc.) was added and deglycosylation was confirmed after 4 h incubation at 37 °C by SDS PAGE analysis.

1.3.2 Protein Analytics

Spectroscopic Protein Determination

The protein concentration c was determined based on the absorption coefficient using a UV-VIS spectrometer (NanoDrop 1000) according to the Lambert-Beer law $A_{280} = \varepsilon dc$ (d = 1 mm). The absorption coefficient ε was calculated using the ExPASy web tool ProtParam (see REF[52]) (www.web.expasy.org/protparam/) depending on the amino acid sequence. The output parameter (e.g., MW and ε) are listed in Appendix A1.

SDS Polyacrylamide Gel Electrophoresis

To monitor protein biosynthesis and to confirm purity and integrity of proteins, sodium dodecyl sulfate-polyacrylamide gel electrophorese (SDS-PAGE) was performed. Proteins of different molecular weights were separated using SDS PAGE gels according to Laemmli et al.[53] (cf. **Table 1.1**). Approximately, 2- 5 µl protein sample were mixed with 5 – 8 µl SDS buffer (**Table 1.2**). Unless otherwise stated, a 3-fold reducing buffer was used. Following denaturing at 85 °C for 5 min, about 10 µl sample was applied onto SDS PAGE gels placed in electrode buffer (6 g/L Tris, 28.8 g/L glycine, and 1 g/L SDS). To separate proteins, 80 V and 120 – 180 V were applied for 15 min and 40 – 80 min, respectively. A protein marker (Precision Plus Protein™ Dual Xtra Standards, BioRad) indicated successful electrophoresis. To finish, the gel was washed two times with water, stained using a staining reagent (GelCode® Blue Stain Reagent, BioRad) for at least 30 min, and destained for at least 1 h by shaking in water. Alternatively, the gel was subjected to western blotting without prior staining.

Table 1.1 | Composition of SDS-PAGE gels

Ingredient	Separation gel			Stacking gel
	10%	12%	15%	6%
Rotiphorese® Gel 30	3.33 ml	4 ml	5 ml	1 ml
1.5 M Tris, pH 8.8	2.5 ml	2.5 ml	2.5 ml	-
1.5 M Tris, pH 6.8	-	-	-	1.25 ml
Millipore H_2O	3.87 ml	3.2 ml	2.2 ml	2.58 ml
10 % SDS	100 µl	100 µl	100 µl	50 µl
80% glycerol	200 µl	200 µl	200 µl	100 µl
20% APS	50 µl	50 µl	50 µl	25 µl
TEMED	10 µl	10 µl	10 µl	10 µl

Table 1.2 | Composition of 3-fold SDS buffer

Ingredient	Reducing	Non-reducing
Glycerol	1.2 ml	1.2 ml
1.5 M Tris, pH 6.8	1 ml	1 ml
SDS	0.6 g	0.6 g
2-β-Mercaptoethanol	0.9 ml	-
Urea	4.5 g	4.5 g
0.25 M EDTA	50 µl	50 µl
Bromophenol blue	10 mg	10 mg
H_2O	ad 10 ml	ad 10 ml

Western Blotting

Following SDS-PAGE, proteins were blotted onto a nitrocellulose membrane (PALL Life Science BioTrace™) using a BioRad electrophoresis device, following a standard protocol. A chamber was filled with transfer buffer (25 mM Tris, 192 mM glycine, and 20% (v/v) methanol) and a blotting cassette was placed into that chamber along with ice pack. After 1 h at 100 V, the membrane first washed with PBS-T (PBS mixed with 0.05% Tween 20). In order to reduce unspecific binding of antibodies, the membrane was placed in ~10 ml of 3-5% (w/v) BSA in PBST in a container for 1 h

at room temperature (RT) or overnight at 4 °C and the blocking buffer was discarded. Next, 10 ml primary antibody solution (1:10 – 1:30 α-$E2^{BVDV}$ or 1:20 α-$CD46^{bov}$ in 2% BSA PBST) was added. The antibodies derived from hybridoma cells were kindly provided by Prof. Dr. Norbert Tautz. The antibody α-$CD46^{bov}$ is originally termed BVDCA17/1/6 (see also REF[21]). Following incubation for 40 to 60 min at RT, while shaking, the membrane was washed three to five times with ~10 ml PBST and the secondary antibody solution (1:2000 α-mouse-HRP, horseradish peroxidase) was added. Finally, the membrane was incubated for 1 h at RT, while shaking, and washed five times with ~10 ml PBST prior to measurement using enhanced chemiluminescence (ECL) by adding 0.6 to 1 ml ECL reagents (1:1 mix of solution A and B, ThermoFisher Scientific) onto the membrane and employing a LAS detector that contains a charge-coupled device (CCD) camera.

1.3.3 Indirect Immunofluorescence

The cells incubated in a 12-well cell culture plate were washed twice with 0.5 to 1 ml PBS and fixed using 4% paraformaldehyde (PFA) in PBS for 20 min at 4 °C. Following three washing steps with PBS, 1 ml blocking buffer was added 3% BSA in PBS, cells were incubated for 10 to 30 min at RT and the blocking buffer was discarded. Next, 500 µl per well of primary antibody (either 1:50 α-$CD46^{bov}$ or 1:30 α-E_2^{BVDV} in 1% BSA PBS) was added. After incubation of 1 h at RT, the cells were washed twice with 1% BSA in PBS, a volume of 500 µl of secondary antibody (1:1000 α-mCy3 (Dianova) or 1:000 α-mA488 (ThermoFisher Scientific) was added, and cells were incubated 1 h at RT in the dark. Prior to imaging using an immunofluorescence (IF) microscope (Zeiss) the cells were washed three times with PBS. IF images were obtained and stored utilizing the software *AxioVision.* Moreover, IF images were equally prepared using *ImageJ*.

1.3.4 Click Chemistry

In order to link biotin derivatives to modified glycan moieties, a copper-mediated azido-alkyne cycloaddition was conducted. If not otherwise stated, the optimized PBS-buffered click solution contained 500 µM 2-(4-((bis((1-*tert*-butyl-1H-1,2,3-triazole-4-yl)methyl)amino)methyl)-1*H*-1,2,3-triazole-1-yl)acetic acid (BTTAA), 250 µM copper sulfate pentahydrate, 200 µM biotin-alkyne or biotin-azide, and 2.5 mM sodium ascorbate (to obtain reduced copper). The synthesis of biotin derivatives was based on REF[26] and REF[54] and is described in REF[54] in detail. The alternative Cu-ligand, tris(hydroxypropyltriazolyl)-methylamine (THPTA) was purchased from ThermoFisher Scientific, and BTTAA was synthesized in-house. The synthesis of BTTAA is based on REF[22] and REF[55] and is described in REF[56] in detail. To ensure efficiency, the click solution was prepared about 10 min before usage.
In the case of blotted proteins onto a nitro cellulose membrane, the click solution contained THPTA instead of BTTAA and the click reaction was conducted for 30 to 60 min at RT. After two washing steps with PBST, 10 ml blocking buffer (3% BSA in PBST, see above) were added onto the membrane and during 30 min incubation, unspecific binding sites were blocked. For the detection of modified glycoproteins comprising azido or alkyne groups, 10 ml of streptavidin-HRP solution

(1:1000 – 1:2000 in 1% BSA PBST) was added and subsequent steps were conducted as outlined above.
In the case of living cells, the adherent cells were washed three times with PBS and the bioorthogonal click reaction with biotin derivatives was performed for 20 min at 4 °C, if not otherwise stated. Afterwards, the cells were washed three times with PBS and fixed with 2% PFA. For detection, cells were washed and incubated with streptavidin Alexa Fluor® 488 conjugate (1:1000 in PBS). Subsequent steps towards imaging were performed as described above.
In a last fluorogenic experiment, MDBK-Alk cells were incubated with E_2A-Az to confirm a specific click reaction between modified glycans of E_2A-Az and the putative receptor CD46-Alk. MDBK-Alk cells were seeded in 24-well plates (~80-90% confluent). First 150 µl of 1% BSA in PBS were added to each well. After 10 min incubation at 4 °C, 300 nM E_2A-Az were added to the blocking buffer. After 10 min incubation at 4 °C, 150 µl of an ice-cold 2-fold click solution (without biotin-derivatives) were added. The click reaction was accomplished after 15 min at 4 °C. To remove non-covalently bound E_2A-Az, the cells were washed five times with PBS. Huh7 cells served as negative control for specificity. Indirect immunofluorescence was conduced as described above. This procedure was repeated with decreasing amount of E_2A-Az (0.3 nM – 80 nM) to obtain a value for half saturation of α-E_2^{BVDV}/α-mA488 fluorescence (see above).

1.3.5 Enrichment of CD46 Molecules using Click Chemistry

In a pilot test, approximately $1.2 \cdot 10^7$ MDBK-Alk or MDBK-Az cells were prepared for each sample. The click reaction on the cell surface was conducted for 30 min as described above. Untreated MDBK cells served as negative control. Following biotinylation, the modified glycoproteins were enriched using streptavidin agarose (strep) beads (Pierce®).
Therefore, the cells were washed using PBS and collected employing a cell scraper followed by centrifugation (700 *x g*, 8 min at 4°C). The cells were lysed in 1 ml lysis buffer (50 mM Tris/HCl, pH 8.5, 10 mM EDTA, 150 mM NaCl, protease inhibitor mix, supplemented with 1% SDS, 1% desoxycholate, 0.5% N-octyl-β-glucopyranoside (L1) or 1% desoxycholate (L2) or 2% N-octyl-β-glucopyranoside (L3) or 1% desoxycholate, 2% N-octyl-β-glucopyranoside (L4). Cells were incubated for 20 min at 4°C and lysis was completed by sonification one ice (pre-cooled MS 73/D sonotrode, 3 pulses à 3 s full power, 1 min pause in between). Next, 100 µl of washed strep beads (~50% suspension) were added to each sample. To allow binding, the samples were incubated for 1 h at 4 °C, while shaking at 1200 rpm. After centrifugation (3 min at 2500 *x g* at RT) and discarding of the supernatant, the beads were washed three times with 1 ml PBS buffer by centrifugation. Finally, modified glycoproteins were eluted in 50 µl non-reducing SDS buffer mixed with 30 mM biotin while incubating and shaking at 1000 rpm at RT for 10 min and heating to 90 °C for 5 min. After centrifugation (5 min at 5000 *x g*), the supernatant of each sample was applied onto a 10% gel. SDS PAGE followed by western blotting was conducted as described above. Subsequently, $CD46^{bov}$ molecules were visualized using indirect enhanced chemiluminescence (see above).
An important aspect when isolating the E_2-CD46 complexes is the integrity of affinity tags in lysis buffer and subsequent sonification. To confirm the functionality of the protein A tag fused to

E_2^{BVDV}, a 100 µl L4 solution containing 500 nM E_2A-Az was incubated in the presence of 20 µl IgG beads (50% suspension) for 1 h at 4 °C while shaking at 1200 rpm. Following sedimentation of beads by centrifugation (5 min at 5000 *x g*), the supernatant was kept for analysis and the beads were washed two times with 1 ml of 25 mM sodium phosphate buffer pH 7. To elute bound E_2A-Az molecules, 100 µl IgG elution buffer (0.1 M glycine/HCl pH 3 stored at RT) was added and beads were incubated for 10 min at RT or 37 °C while shaking at 1200 rpm. After centrifugation (5 min at 5000 x g at RT), the eluate and other samples were analyzed by western blotting (see above).

In a final step, covalent E_2A-Az/receptor-Alk complexes were enriched using IgG beads in a batch method after click reaction. Therefore, ~$3 \cdot 10^7$ MDBK-Alk cells (see previous sections) cultured in a 15 cm plate were washed three times with PBS and incubated in 12 ml 1% BSA PBS for 10 min at 4 °C. Next, 100 nM E_2A-Az were added. Following incubation for 30 min at 4 °C, while shaking at 200 rpm to cover the whole surface of adherent cells, 6 ml of 3-fold freshly prepared, ice-cold click solution was added. The click reaction was stopped after 15 min at 4 °C while shaking at 200 rpm. Afterwards, the cells were washed two times with 15 ml ice-cold PBS. Cells were finally resuspended using a scraper in 10 ml ice-cold PBS and collected by centrifugation for 10 min at 2000 *x g* at 4 °C. The cell pellet was stored at -60 °C until usage. Next, cells were thawed and 1 ml L4 lysis buffer (see above) was added. After 15 min incubation at 4 °C while shaking (1200 rpm) and sonification (see above), the lysate was centrifuged for 10 min at 8000 *x g* at 4 °C. The supernatant was mixed with 100 µl IgG beads (~50% suspension) followed by one washing step with 1 ml of 25 mM sodium phosphate pH 7 (NaP) supplemented with 10% L4 buffer and two washing steps with 1 ml NaP. Elution was conducted by adding 100 µl IgG elution buffer and incubating 15 min at 37 °C while shaking at 1200 rpm. The eluate was analyzed via western blotting using α-$CD46^{bov}$ as primary antibody as described above under non-reducing conditions.

1.4 Results

This work aimed to explore the possibility of isolating a viral protein-receptor complex following metabolic labeling of glycoconjugates and *click chemistry*. At first, labeling of glycoconjugates of MDBK cells by the use of per-*O*-acetylated *N*-(2-azidoethan-1-one-yl)-galactosamine (Ac_4GalNAz) is presented. Next, the optimization of the click reaction and the incorporation of a panel of unnatural monosaccharide precursors in glycan structures of MDBK and Huh7 cells is shown. Subsequently, indications of the CuAAC-mediated formation of covalent linkages between E_2A-Az proteins and glycoconjugates bearing alkyne groups of MDBK cells are based on indirect immunofluorescence. Finally, the first hint is provided that the purification of E_2A-receptor complexes is feasible.

1.4.1 Incorporation of Modified Monosaccharide Precursors

In the initial approach, increasing amounts of Ac_4GalNAz were added to the cell culture medium of ~$2 \cdot 10^6$ MDBK cells. It has been demonstrated that Ac_4GalNAz undergoes deacetylation at higher pH values of about 8 (Mallagaray, 2014, unpublished data). Therefore, we decided to adjust the value of the cell culture medium to pH 6.6 before adding modified monosaccharide precursors. After metabolic labeling of glycoconjugates of MDBK cells and cell lysis, azido group bearing moieties were detected by western blotting using a biotin-alkyne derivative to visualize both cytosolic and membrane proteins. In a second step, the copper-catalyzed azido-alkyne cycloaddition (CuAAC) was conducted directly on adherent MDBK cells to label membrane exposed glycoconjugates. The results of metabolic labeling are shown in **Figure 1.4**, indicating an efficient incorporation of GalNAz into glycan moieties. It appeared that the efficiency can be increased by replenishing monosaccharide precursors. Moreover, the optimal concentration of Ac_4GalNAz was in the range of 200 to 800 µM.

In general, the click solution contains copper (Cu(I)), a Cu(I)-chelating agent, e.g. THPTA, and sodium ascorbate. The water-soluble Cu(I)-chelating agent, termed Cu-ligand throughout this study, is critical for an efficient click reaction on living cells. The Cu-ligand accelerate the copper azido-alkyne cycloaddition and acts as sacrificial reductant for oxidative species produced by cell toxic copper, atmospheric oxygen, and ascorbate. The latter is required to maintain the copper in the active +1 oxidation state. But THPTA is regarded as toxic and therefore incompatible with living cells[57]. It is apparent that the cell membrane is destabilized through click reactions probably due to 1% DMSO, THPTA, or both. Otherwise, DAPI staining of cell nuclei would not have been occurred (see **Figure 1.5**). In order to gauge the CuAAC reaction over time, MDBK cells were incubated with 800 µM Ac_4GalNAz and labeled with biotin-alkyne. After the addition of streptavidin-Alexa Fluor®488, the cells were monitored by the use of immunofluorescence microscopy. According to these data, the relative fluorescence levels off at a reaction time of 30 min (see **Figure 1.5**). Moreover, cell lysis becomes visible after 45 min incubation.

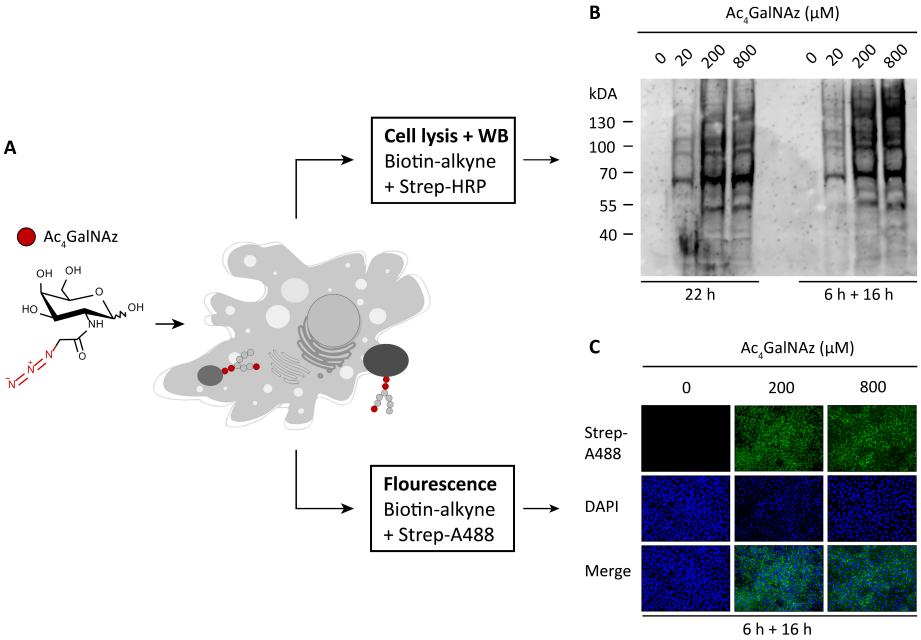


Figure 1.4 | Detection of glycoconjugates of MDBK cells using metabolic labeling and *click chemistry*
(A) MDBK cells were incubated in DMEM (pH 6.6) in 6-well plates with increasing concentrations of tetraacetylated *N*-(2-azidoethan-1-one-yl)-galactosamine ($Ac_4GalNAz$) for up to 22 h. The medium was either retained (22 h) or replaced by freshly prepared medium supplemented with $Ac_4GalNAz$ after 6 h incubation (6 h + 16 h). (B) After cell lysis and western blotting (WB), the copper-mediated azido-alkyne cycloaddition was performed for 1 h at RT using 200 µM biotin-alkyne, 500 µM THPTA, 100 µM $CuSO_4$, and 2.5 mM sodium ascorbate in 10 ml of 100 mM sodium phosphate buffer supplemented with 1% DMSO (pH 7.0). Biotinylated Glycoproteins were visualized using horseradish peroxidase (HRP) coupled to streptavidin (strep). (C) Following incubation with DMEM (pH 6.6) supplemented with $Ac_4GalNAz$, MDBK cells were washed with PBS (three times) and click reaction was conducted on cells for 15 min at 4 °C. For the detection, strep coupled to Alexa Fluor® 488 (A488) was used. The nuclei were stained with 4',6-Diamidin-2-phenylindol (DAPI).

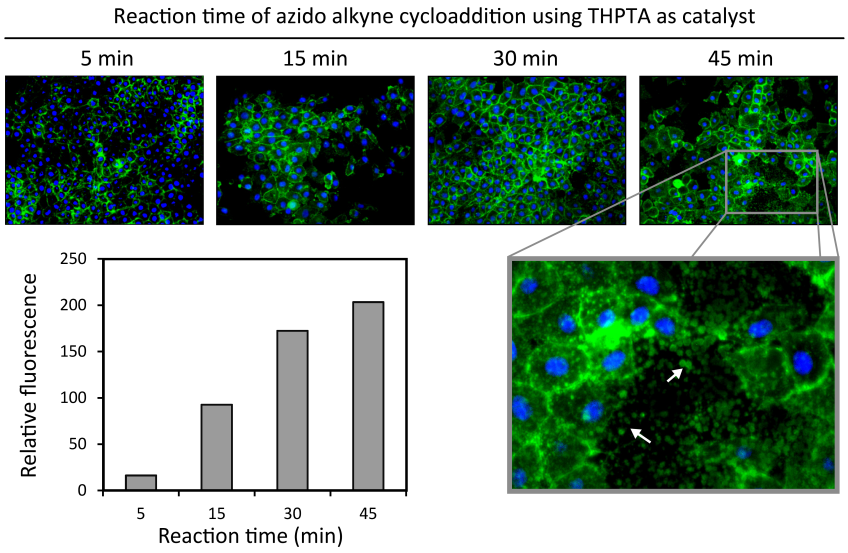


Figure 1.5 | Time course of the click reaction using THPTA as catalyst
MDBK cells were treated with 800 µM $Ac_4GalNAz$ as described in **Figure 1.4** and section **1.3.1**. Successful click reaction between azido groups of surface glycoproteins and biotin-alkyne was monitored over time using streptavidin coupled to Alexa Fluor® 488 as detection reagent (see **Figure 1.4** for composition of the click solution). The nuclei were stained with 4',6-Diamidin-2-phenylindol (DAPI). Cells seem to degrade when the click reaction persist over 30 min as this is seen in the enlarged region (white arrows). ImageJ was used to quantify the relative fluorescence. For the negative controls (MDBK cells incubated without $Ac_4GalNAz$), no fluorescence signal could be detected (not shown).

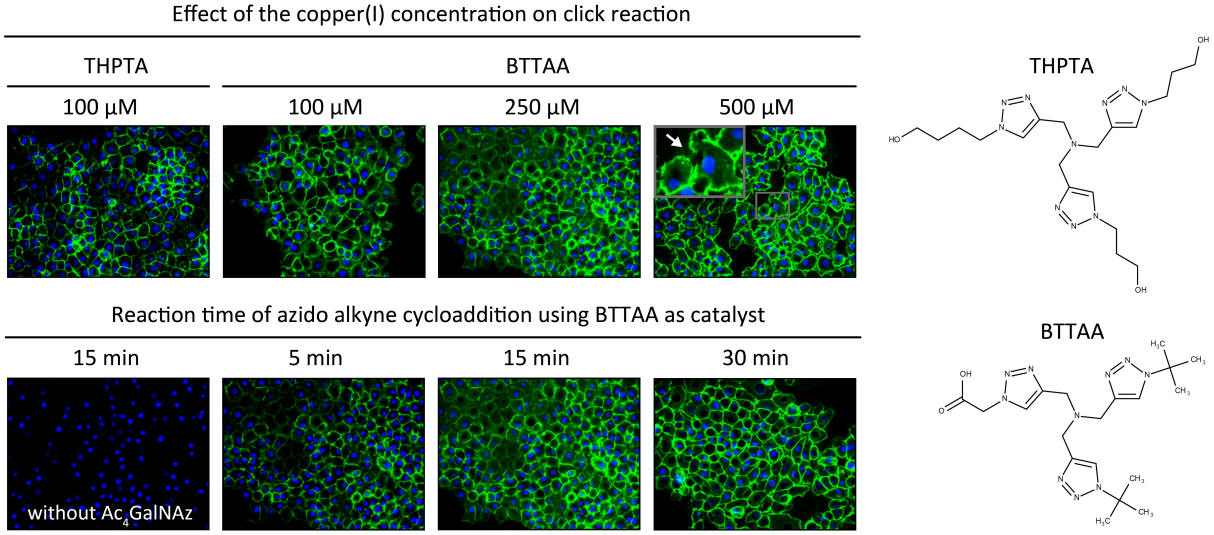


Figure 1.6 | Optimization of the click reaction using BTTAA as a catalyst
After MDBK cells were incubated with 200 µM Ac_4GalNAz in a 12-well plate format, the click reaction was conducted on MDBK cells at 4 °C. Unlike previous conditions (cf. **Figure 1.4**/ **Figure 1.5**), click reagents were dissolved in PBS (pH 7.2) instead of sodium phosphate buffer mixed with 1% DMSO. **Top:** The click reaction was conducted for 15 min at 4 °C at varying copper(I) concentrations in the presence of 500 µM THPTA/ BTTAA, 200 µM biotin-alkyne, and 2.5 mM sodium ascorbate. The reaction efficiency was monitored via detection of Alexa Fluor® 488 coupled to streptavidin. **Bottom:** In order to evaluate the reaction outcome over time, click reagents, as seen above, and 250 µM copper(I) were utilized. The negative control (without supply of Ac_4GalNAz) was equally treated.

During the course of this study, a novel Cu-ligand (BTTAA) has been synthesized in-house, since a 10-fold labeling efficiency of the CuAAC reaction on living cells has been observed by other groups when utilizing BTTAA compared to THPTA[55]. In the current study, however, only a 2-fold to 3-fold enhancement was observed when employing Ac_4GalNAz-treated MDBK cells in combination with BTTAA (see **Figure 1.6**). The labeling efficiency could be enhanced by increasing the $CuSO_4$ concentration from 100 to 250 µM. A higher copper concentration is not beneficial as cell toxicity was prevalent at 500 µM. The optimal click reaction time seems to be in the range of 15 to 20 min.
Following optimization of click reaction, a panel of various synthetic monosaccharide precursors were analyzed to determine an appropriate composition of azido or alkyne sugars. Therefore, MDBK cells were incubated with 600 µM of modified sugars, including Ac_4GlcNAz, Ac_4GlcNAlk, Ac_4GalNAz, Ac_4GalNAlk, Ac_4ManNAz, and Ac_4ManNAz (cf. **Figure 1.1**). The labeling of Huh7 cells was also judged using galactosamine and mannosamine derivatives. In all cases, successful labeling of glycoconjugates could be confirmed (see **Figure 1.7**). Labeling of membrane-associated moieties with Ac_4GlcNAz seemed to be inefficient, which is in accordance with results obtained by Bertozzi and co-workers[38]. Furthermore, it is apparent that MDBK cells tolerate 600 µM unnatural monosaccharide precursors in most cases as this has been also observed for Vero cells[42]. In contrast, the viability of Huh7 cells appeared to be negatively affected. The tolerance of modified monosaccharide seemed to be dependent on the cell type. This phenomenon has been observed by other groups for Hela cells, CHO cells, or Jurkat cells as well[57].

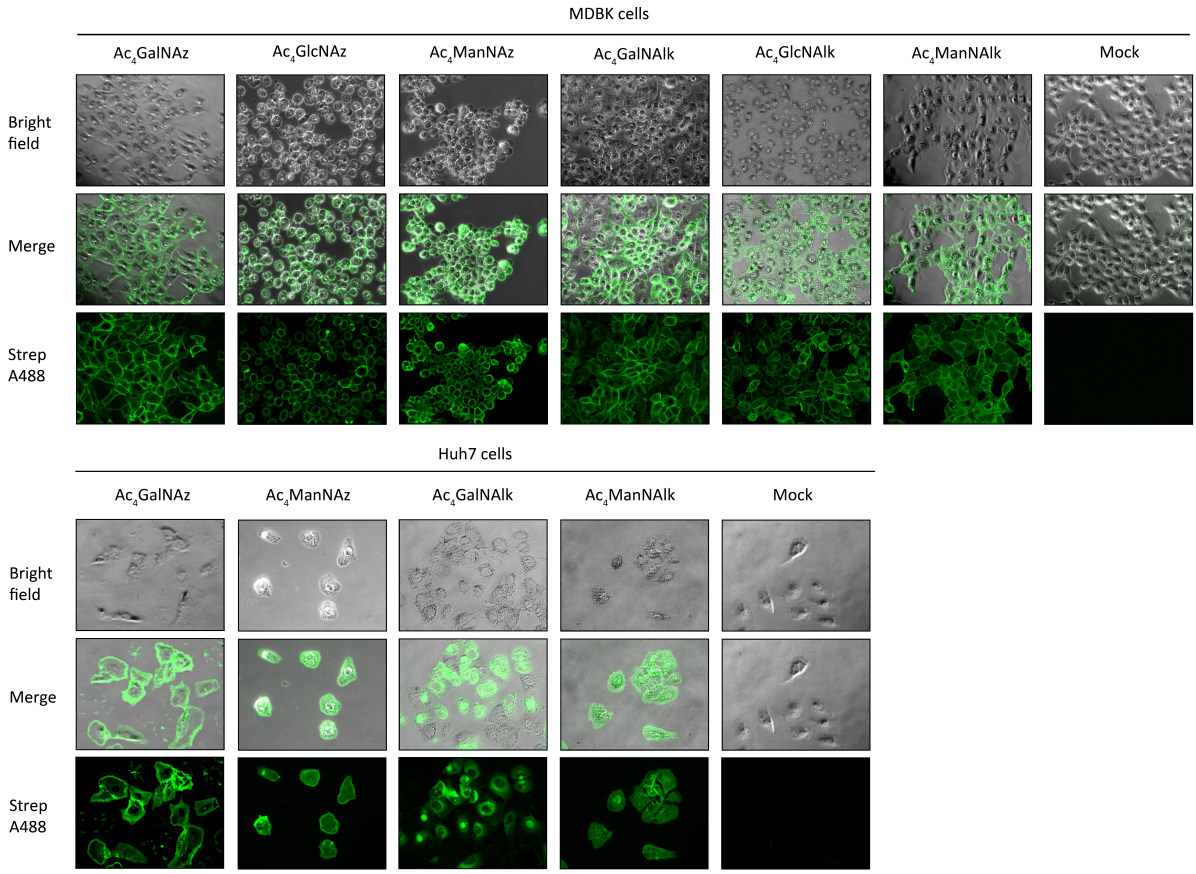


Figure 1.7 |Labelling of surface glycoproteins of MDBK and Huh7 cells using modified monosaccharides
MDBK cells were incubated in Dulbecco's Modified Eagle's medium (DMEM, pH 6.6) with 400 µM modified sugars carrying azido (Az) or alkyne groups (Alk) for 4 h and following medium change (pH 6.6, 600 µM modified sugars) for 16 h at 37 °C and 5% carbon dioxide. The cells were washed three times with PBS buffer. The click reaction was performed in PBS buffer for 15 min at 4°C using 500 µM BTTAA, 250 µM Cu, 200 µM biotin-linker carrying either an azido group or an alkyne group, and 2.5 mM sodium ascorbate. After washing three times with PBS, the cells were fixed using 2% paraformaldehyde in PBS at 4°C for 20 min, washed again three times with PBS, and incubated with streptavidin, Alexa Fluor® 488 conjugate (1:1000 in PBS) for 40 min at RT in the dark. Finally, the cells were washed with PBS three times and the fluorescence of Alexa Fluor® 488 was visualized by fluorescence microscope.

1.4.2 Production of the Modified BVDV Protein E_2A-Az

In order to evaluate the possibility of identifying viral receptors using *in vivo* glycan engineering and *click chemistry*, a recombinant viral glycoprotein was required, which harbors bioorthogonal functionalities. In this study, the ectodomain of BVDV E_2 protein fused to a protein A-tag and labeled with azido groups (E_2A-Az) was recombinantly synthesized by the use of HEK293T cells, which secrete the viral protein permanently into the cell culture medium. About 3 mg per liter medium could be purified by employing a 5 ml IgG Sepharose™ 6 Fast Flow column. Antibody detection of E_2A and SDS PAGE analysis confirmed a successful purification of both native E_2A and E_2A-Az proteins. The biotinylation of E_2A-Az via CuAAC-mediated *click chemistry* using biotin-alkyne derivatives validated the incorporation of azido groups into *N*-glycan chains. The presence of *N*-glycans was confirmed by PNGase F digestion (see **Figure 1.8**).

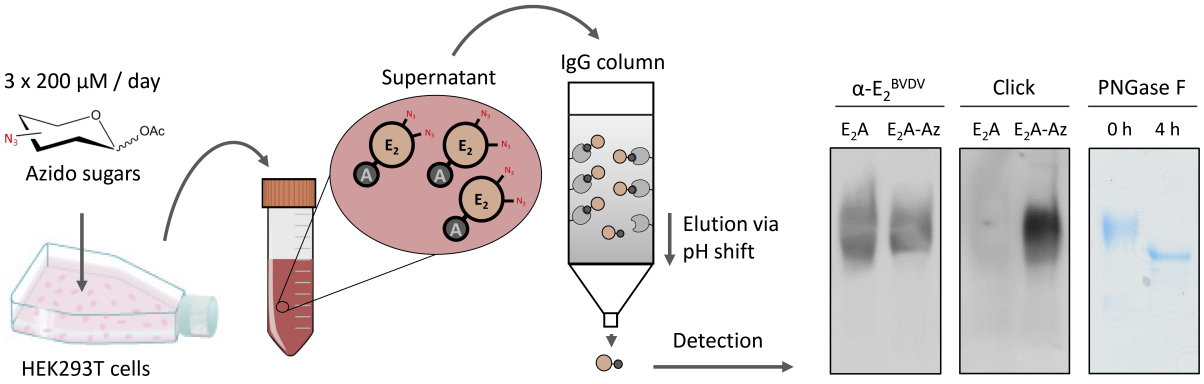


Figure 1.8 | Production and purification of the viral protein E_2 fused to a protein A tag
HEK293T cells were incubated in 30 ml DMEM pH 6.8 (10% FCS and 1% pen/strep) in T175 cell culture flasks. During a time period of 2.5 days, cells were fed periodically seven times with a mixture of azido sugars (7:2:1 Ac_4GalNAz, Ac_4ManNAz, and Ac_4GlcNAz). Filtered supernatant was passed through a 5 ml IgG Sepharose™ 6 Fast Flow column. Elution was performed by addition of 19 ml of 0.1 M glycine/HCl (pH 3.0) into 1 ml Tris buffer (1 M, pH 9). Following buffer exchange against E_2 buffer (25 mM HEPES, 250 mM NaCl, 5% glycerol), E_2A (without modified sugars) and E_2A-Az were blotted onto a nitrocellulose membrane after SDS PAGE under reducing conditions. Visualization was achieved either by indirect antibody detection (α-E_2^{BVDV}) or by streptavidin-HRP bound to biotin after click reaction (click) The click solution contained 500 µM BTTAA, 250 µM Cu(I), 400 µM biotin-alkyne, and 2.5 mM ascorbate dissolved in PBS. The apparent molecular weight of glycosylated E_2A-Az is about 75 kDa. Removal of *N*-glycans by PNGase F digestion of E_2A-Az for 4 h at 37 °C was confirmed by SDS PAGE and Coomassie blue staining.

In a second step, the structural integrity and biological activity should be confirmed. MDBK cells were therefore infected with non-cytopathic (ncp) BVDV (MOI of 5) in the presence and absence of native and modified E2A proteins. The rate of infectivity was qualitatively determined after 18 h infection by staining the BVDV nonstructural protein NS3, showing an evident reduction of infectivity in a dose-dependent manner. Quantitative determination of the virus titer revealed a decrease in $TCID_{50}$ values of about one log level in the presence of both 25 nM E2A and E2A-Az, whilst the addition of bovine serum albumin (BSA) had practically no effect on the virulence of ncp BVDV.

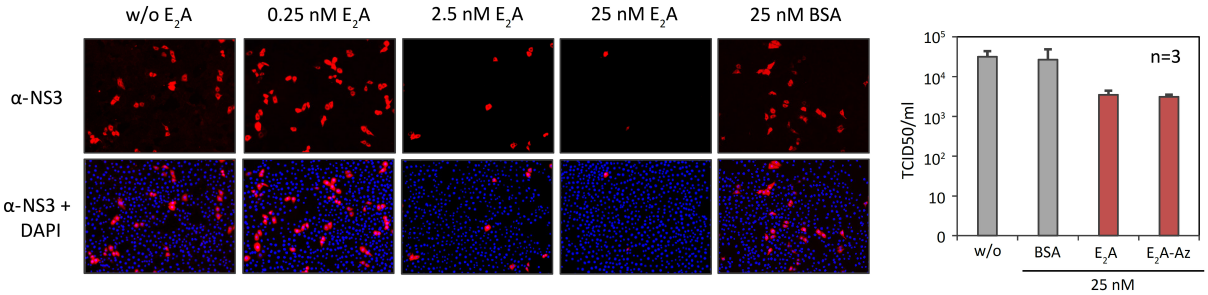


Figure 1.9 | Evaluation of the biological activity of the BVDV ectodomain E_2A and E_2A-Az by virus titer determination
Approximately $6 \cdot 10^5$ MDBK cells per well were infected with ncp BVDV at a MOI of 5 in a 12-well cell culture plate in the absence or presence of increasing amount of the native E_2A protein. The infectivity was judged by immunofluorescence of the BVDV NS3 protein 18 h after infection. BSA served as negative control. On the right side, the virus titer ($TCID_{50}$/ml) of the cell supernatant was determined 18 h after infection. Error bars represent standard deviations of triplicates. The data were obtained in the course of a master's thesis by Ricarda Marshall (Institute of Chemistry, 2016; supervised by me). ($TCID_{50}$, the median of tissue culture infection dose)

1.4.3 Click Reaction of E_2A-Az and MDBK-Alk Cells

A crucial requirement of identifying viral receptors using *click chemistry* is the formation of a covalent linkage between E_2A-Az and surface glycoconjugates on susceptible host cells (i.e. the bovine CD46 receptor and other coreceptors). First, the presence of CD46 molecules on cultured MDBK cells was confirmed by indirect immunofluorescence (**Figure 1.10**, upper panel). Second, the incorporation of modified per-*O*-acetylated alkyne sugars (GalNAlk, ManNAlk, GlcNAlk; 7:2:1 mixture) into glycan structures of MDBK and Huh7 cells (negative control) was verified (**Figure 1.10**, middle panel). In a third experiment, MDBK-Alk cells and Huh7-Alk cells were incubated with 300 nM E_2A-Az and the formation of covalent linkages was performed by *click chemistry* at 4 °C for 15 min. Following rigorous washing with PBS, covalently bound E_2A proteins were

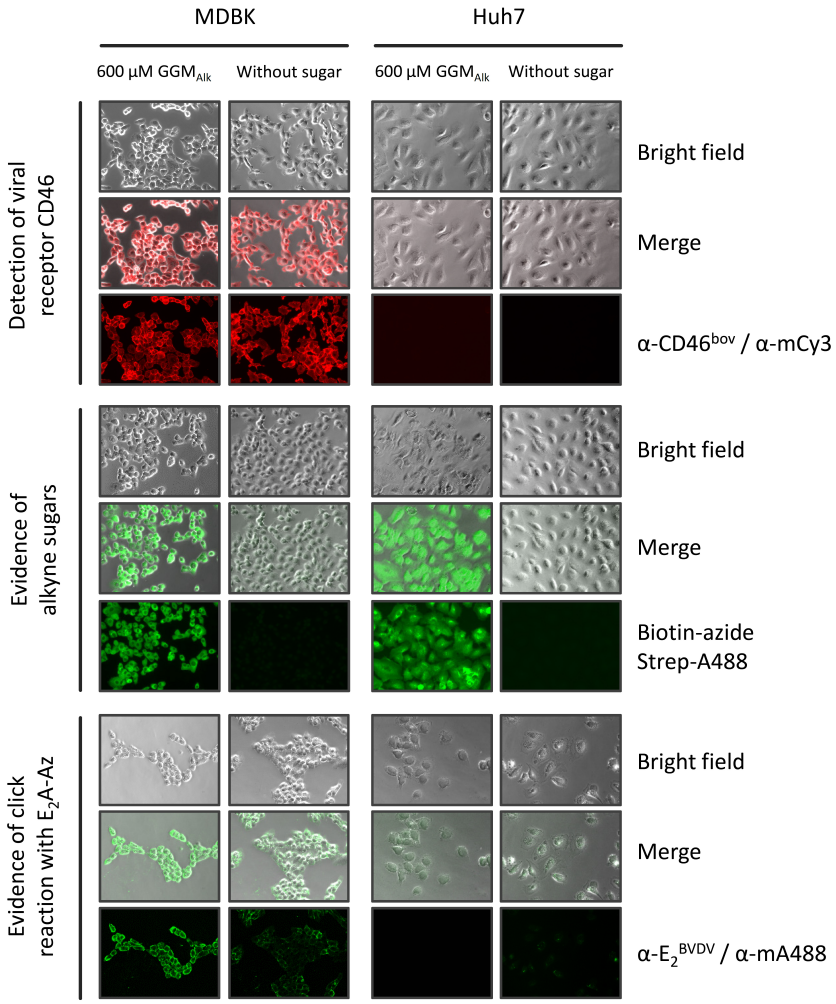


Figure 1.10 | Covalent coupling of E_2A-Az on MDBK-Alk cells induced by click chemistry
MDBK and Huh7 cells were seeded in 12-well plates and cultured in DMEM pH 6.6 (10% PS, 1% pen/strep) in the absence or presence of 600 µM sugar mix (7:2:1 Ac_4GalNAlk, Ac_4ManNAlk, and Ac_4GlcNAlk). First, the presence of the bovine CD46 receptor on the cell surface of MDBK cells was confirmed by indirect immunofluorescence. Second, the incorporation of alkyne sugars was validated after covalent coupling of biotin-azide (200 µM) with surface-exposed glycan structures followed by streptavidin-Alexa488 staining. Third, E_2A-Az proteins were added to cells at a concentration of 300 nM. After click reaction was conducted, non-covalently bound proteins were washed off with PBS five times. Remaining E_2A proteins were detected by indirect immunofluorescence. Therefore primary α-E_2^{BVDV} antibodies and polyclonal anti-mouse A488-coupled secondary antibodies (α-mA488) were used.

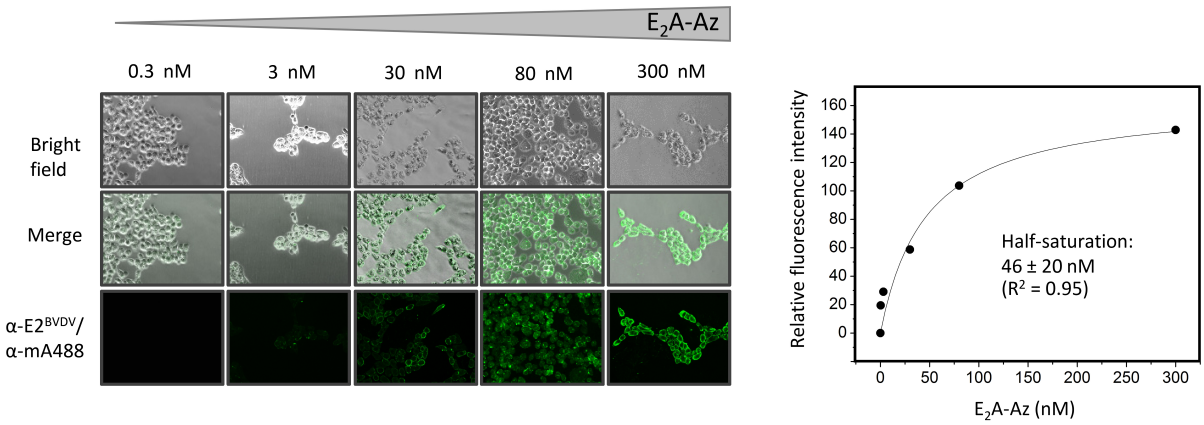


Figure 1.11 | E_2A-Az concentration-dependent azido-alkyne cycloaddition via glycan chains
About $3 \cdot 10^5$ MDBK-Alk cells were incubated with increasing amount of E_2A-Az. Following click reaction, cells were washed five times with PBS and covalently bound E_2A-Az proteins were visualized using indirect immunofluorescence. The relative fluorescence intensity was determined using ImageJ and curve fitting was performed in origin 7.0. The maximal theoretical fluorescence intensity was 4095 units. To normalize fluorescence values, the background fluorescence signal was subtracted for each image.

detected by indirect immunofluorescence using the primary antibody α-E_2^{BVDV}. The results indicate a covalent linkage of E_2A-Az and glycoconjugates that are present on the surface of MDBK cells when incubated with alkyne sugars. Untreated MDBK cells did not show strong fluorescence signals following E_2A-specific antibody staining (**Figure 1.10**, lower panel). Incubating unmodified E_2A proteins on MDBK cells without rigorous washing, however, resulted in clearly detectable immunofluorescence derived from bound E_2A proteins (data not shown). The lack of fluorescence signals regarding Huh7 cells suggests that a specific interaction between E_2A-Az and its receptors (and potential attachment factors) seems to be necessary for a CuAAC-mediated formation of covalent linkages between modified glycoconjugates. The same experiment was conducted with unmodified E_2A proteins on MDBK-Alk cells. This resulted in an almost complete reduction of the fluorescence level after rigorous washing similar to that seen for E_2A-Az incubated on native MDBK cells (data not shown). In order to determine the amount of E_2A-Az that is needed for an efficient CuCCA reaction, a titration was performed (**Figure 1.11)**. It is apparent that full saturation is nearly reached at a concentration of E_2A-Az of 300 nM.

1.4.4 Purification of CD46 Molecules and E_2A-Az-receptor Complexes

The previous experiments provided strong indications that it is feasible to link E_2A-Az proteins onto surface moieties of MDBK-Alk cells. In the last step, three key experiments were conducted by employing a micro-scale purification method using affinity beads: the enrichment of biotinylated CD46 molecules along with other labeled glycoproteins, the purification of E_2A-Az alone, and the enrichment of E_2A-receptor complexes. The experimental design and results are shown in **Figure 1.12 A-C**. In order to select an appropriate lysis buffer that allows both a convenient purification of transmembrane proteins and a downstream analysis by mass spectrometry (e.g. Tween-20 is incompatible), four mixtures of detergents were analyzed. According to a western blot analysis of

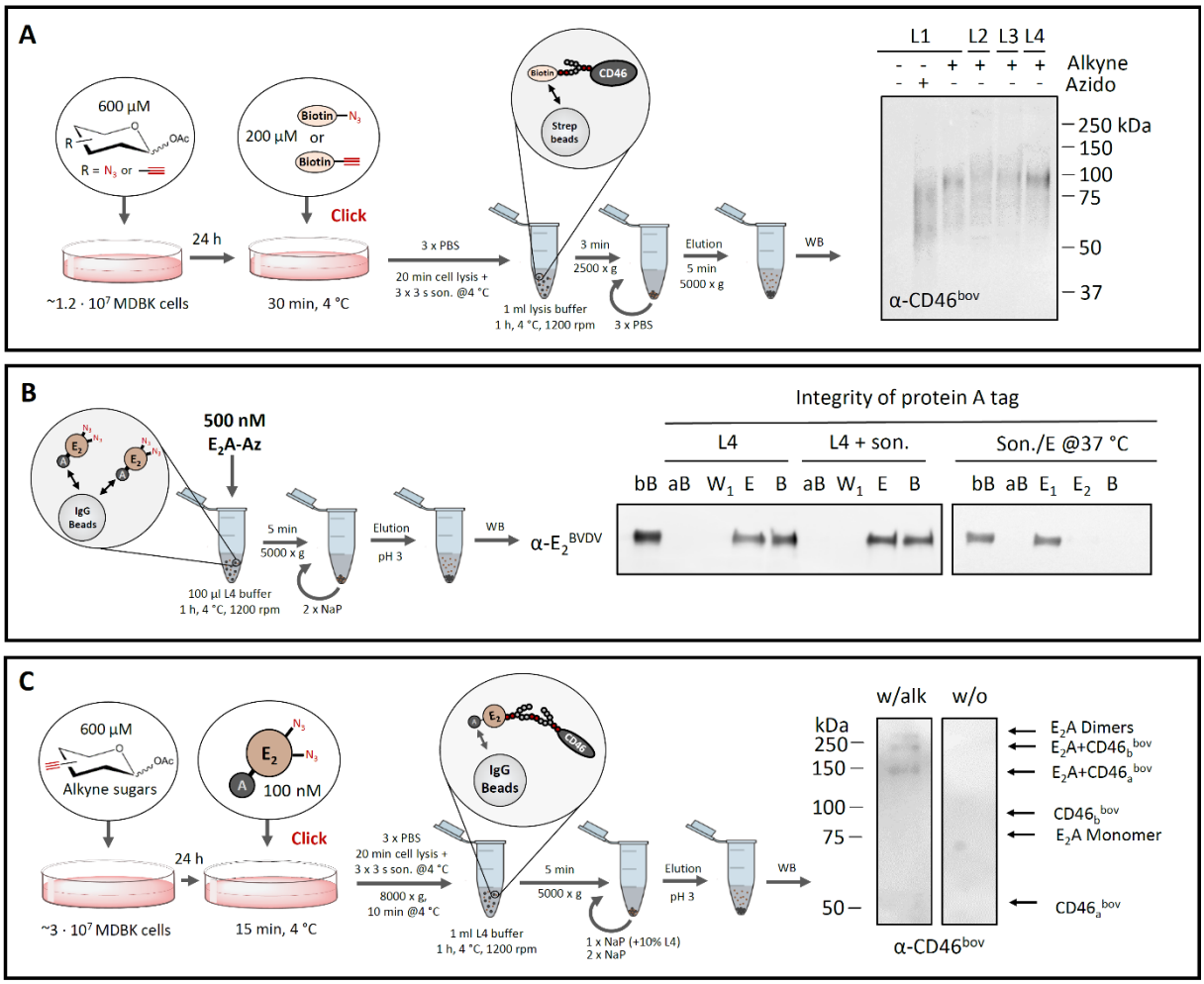


Figure 1.12 | First indications of the enrichment of CD46 molecules using ligand-specific *click chemistry*
(A) MDBK cells were incubated in DMEM pH 6.6 (10% PS, 1% pen/strep) in the absence or presence of 600 µM sugar mix (7:2:1 Ac_4GalNAz/Alk, Ac_4ManNAz/Alk, and Ac_4GlcNAz/Alk). Modified surface exposed glycans were biotinylated using *click chemistry*. Following cell lysis and sonification (son.), biotinylated glycoproteins were enriched by streptavidin agarose beads using a batch method. The different lysis buffer consist of 50 mM Tris/HCl, pH 8.5, 10 mM EDTA, 150 mM NaCl, protease inhibitor mix, and L1 (1% SDS, 1% desoxycholate, 0.5% N-octyl-β-glucopyranoside) or L2 (1% desoxycholate), L3 (2% N-octyl-β-glucopyranoside), or L4 (1% desoxycholate, 2% N-octyl-β-glucopyranoside). After elution with 30 mM biotin, boiling at 90 °C in 50 µl non-reducing SDS buffer, and centrifugation (5 min at 5000 x g), glycoproteins were blotted. CD46 molecules were visualized via indirect enhanced chemiluminescence utilizing α-$CD46^{bov}$ as the primary antibody. **(B)** The integrity of the protein A tag dissolved in L4 buffer was examined. Elution from ~10 µl IgG beads worked best by the addition of 100 µl 0.1 M glycine/HCl (pH 3) at 37 °C instead of 20 °C while shaking for 10 min at 1000 rpm. **(C)** E_2A-Az was added to the culture supernatant of MDBK-Alk cells (w/alk) or on native MDBK cells (w/o). After 30 min incubation at 4 °C, click reaction was conducted on adherent cells. Following washing with PBS and cell lysis, the putative E_2A-$CD46^{bov}$ complex was isolated by the use of 100 µl IgG beads in a batch method. Following elution and centrifugation, the supernatant was analyzed by western blotting (WB). The proposed position of associated proteins and macromolecular complexes are depicted on the right side based on reference blots and apparent molecular weights.

enriched biotinylated glycoproteins after staining with α-CD46^{bov}/ α-mHRP, L1 buffer and L4 buffer were most suitable for the solubilization of CD46 molecules (see **Figure 1.12** for the composition of lysis buffer). L4 buffer contains only non-denaturing detergents (i.e. 1% desoxycholate and 2% N-octyl-β-glucopyranoside), which is reported to ensure full activity of the Protein A (ProtA)-tag[58]. Hence, L4 lysis buffer was selected for further studies. It is striking that one smeared band at 60 kDa, which represents one splice variant of CD46 (cf. REF[21]), is seemingly missing in lysates originated from MDBK-Alk cells when L4 buffer was used, whereas the splice variant with extended STP domain (~90 kDa)[21] is clearly visible. Possible reasons for that are discussed later. In a next step, the enrichment of 500 nM E_2A-Az dissolved in 100 µl L4 buffer with ~10 µl IgG beads was optimized. It turned out that the protein A tag fused to E_2 remains intact even after short pulses of sonification. Moreover, incubating IgG beads that captured E_2A-Az molecules in 100 µl glycine/HCl (pH 3) at 37 °C for 10 min while shaking at 1000 rpm ensured a complete detachment of bound proteins (cf. **Figure 1.12B**). Using these optimized conditions, the possibility of purifying a covalently linked E_2A-receptor complex was explored. Therefore, ~$3 \cdot 10^7$ MDBK-Alk cells were incubated with 100 nM E_2A-Az. Next, cells were incubated for 30 min at 4 °C. Both the incubation time and temperature were chosen to allow the formation of specific interactions between E_2A-Az and CD46-Alk and to eliminate internalization of cell-surface species during the process of attachment[57]. After click reaction was performed for 15 min at 4 °C, solubilized complexes were purified as illustrated in **Figure 1.12C**. The α-CD46^{bov} immunoblotting of purified complexes showed proteins bands that can not be readily explained either by E_2A species alone or by CD46 splice variants. These findings indicate a covalent complex formation of E_2A-Az and CD46-Alk molecules and/or other coreceptors as this is discussed below.

1.5 Discussion

The first task was to demonstrate the incorporation of azido or alkyne sugars into glycoconjugates of living cells as this has been reported by many researchers before[35,36,38,39,42]. In this study, MDBK cells and Huh7 cells were successfully labeled. However, cytotoxic effects were observed for some of the modified monosaccharide precursors. According to the cell morphology of MDBK cells, Ac_4GlcNAz and Ac_4ManNAz seem to reduce the cell viability as cells appeared small and spherical, a typical indication of toxic effects for this cell type. Han and co-workers have interrogated the physiological impact of Ac_4ManNAz on human lung adenocarcinoma cells (A549), suggesting a maximal non-cytotoxic concentration as little as 10 µM. For example, 50 µM Ac_4ManNAz led to the reduction of major cellular functions, such as energy generation capacity, cellular infiltration ability, and channel activity[59]. On the other hand, another study has demonstrated cytotoxic effects with Vero cells when utilizing an Ac_4ManNAz concentration higher than 200 µM. Moreover, this cell type tolerated Ac_4GalNAz up to a concentration of 1000 µM without affecting the cell viability. Another important consideration was the fact that Ac_4GlcNAz is proven to predominantly label cytosolic or nuclear glycoproteins that generally bear β-O-GlcNAc residues[38,60]. Due to these findings, modified monosaccharide precursors were employed in a ratio of 7:2:1 Ac_4GalNAz: Ac_4ManNAz: Ac_4GlcNAz, which allowed the production of biologically active E_2A-Az proteins expressed by HEK293T cells. Since a high labeling efficiency is expected for azido monosaccharide precursors, the total concentration of these precursors may be reduced from 600 µM to 300 µM per day to ensure permanent viability of HEK293T cells. In the case of alkyne sugar derivatives, MDBK cells seem to tolerate all sugar species according to the cell morphology. This is in line with results from Hsu and co-workers that showed reduced toxicity of alkynyl fucose compared to azido fucose[39]. Hence, alkyne sugars were used to label glycoconjugates of MDBK cells, including the glycosylated virus receptor CD46. The presence of CD46 molecules of MDBK-Alk cells could be confirmed by indirect immunofluorescence. However, in order to detect only cell surface glycoconjugates confocal microscopy is required. Moreover, a change in the physiological cell state and hence an alteration of the gene expression of MDBK cells could not be excluded upon treatment of alkyne sugars.

The second prerequisite was the formation of a covalent linkage of E_2A-Az with glycoconjugates of MDBK-Alk cells. Results from indirect immunofluorescence targeting E_2A-Az epitopes gave strong indications of a productive click reaction and hence covalent linkage. Moreover, the lack of fluorescence signals of Huh7-Alk cells treated with E_2A-Az suggested that a specific interaction is required for CuAAC-mediated complex formation. It is well known that the glycosylation pattern differs for various cell culture lines[61,62] and binding of several viruses depends on a proper glycosylation[29–31]. But these facts would not satisfactorily explain the lack of fluorescence signals regarding Huh7 cells, in particular, due to the absence of CD46 molecules.

To validate these findings, the click reaction with MBDK-Alk cells and E_2A-Az proteins was repeated (see **Figure 1.13**). In order to assuredly remove non-covalently bound proteins, the concentration of NaCl in the washing buffer was increased to 600 mM.

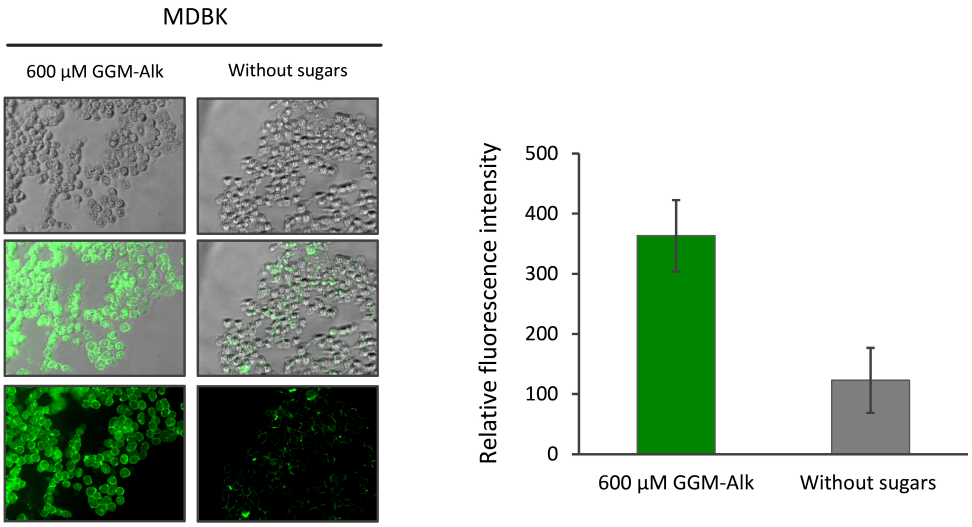


Figure 1.13 | Validation of the covalent coupling of E_2A-Az with glycan structures of MDBK-Alk cells
MDBK cells were seeded in 12-well plates and cultured in DMEM pH 6.6 (10% PS, 1% pen/strep) in the absence or presence of 600 µM sugar mix (GGM-Alk, 7:2:1 Ac_4GalNAlk, Ac_4ManNAlk, and Ac_4GlcNAlk). Incubation of 320 nM E_2A-Az and the click reaction were conducted as this is described in **Figure 1.10**. In order to disrupt non-covalent binding, the cells were washed five times with phosphate-buffer saline containing 600 mM NaCl. Three randomly chosen regions were imaged for each condition and fluorescence intensities were quantified using ImageJ as this is depicted on the right side. Error bars represent the standard deviation of three relative fluorescence values.

The final steps involved the enrichment of MDBK-derived biotinylated glycoproteins and, finally, the purification of E_2A-receptor complexes. It was conspicuous that the band pattern of immunoblotted CD46 molecules differed when comparing bovine CD46-Az-biotin and CD46-Alk-biotin molecules. Only in the case of CD46-Az-biotin, both splice variants (50 to 60 kDa and ~90 kDa) were detectable, which vary by a stretch of 36 amino acids in the STP region. This region seems to determine the degree of *O*-glycosylation[21] as 14 of 16 proposed *O*-glycan sites[63] are located in this STP-rich stretch. None are present in the exposed domains CCP1 to CCP4 (cf. **Figure 1.2** and Appendix A1), which, in contrast, contain at least two *N*-glycan sites.

One reasonable explanation for the missing ~60 kDa splice variant in the case of alkyne group-bearing CD46 molecules would be the fact that azido monosaccharides are tolerated by both the *N*-glycosylation and the *O*-glycosylation cellular machinery, whilst alkynyl sugars predominantly lead to *O*-glycosylation. To my knowledge, there is no comprehensive study, which compared the incorporation mode of all available alkyne or azido monosaccharide precursors. One study has claimed that GNA1$^{-/-}$ yeast *N*-glycans can be labeled with both GlcNAz and GlcNAlk, but the ultimate evidence was only given for GlcNAz by mass spectrometry[64]. In addition, Péanne and co-workers utilized alkyne-tagged sialic acid (SiaNAlk) to investigate the *N*-glycosylation process in ER and Golgi[65].

Sialic acids are not only found in *N*-glycans but also in *O*-glycans. Since it has been demonstrated that the incorporation rate of SiaNAz derived from Ac_4ManNAz into sialylated glycoconjugates is very low in a cell type-dependent manner[34,38], Chang and co-workers compared the efficiencies of metabolic conversion of Ac_4ManNAz and Ac_4ManNAlk to glycoconjugate-bound SiaNAz and SiaNAlk, respectively. Indeed, the incorporation rate of Ac_4ManNAlk into sialic acid bearing glycoconjugates of several cell lines was significantly higher compared to Ac_4ManNAz[66]. This may

explain why the band of the 90 kDa splice variant was more intense in the case of MDBK-Alk derived glycoproteins compared to MDBK-Az derived glycoproteins. However, more experiments are required to shed light on the different labeling rates and pathways of both azido and alkyne sugar derivatives. Moreover, it is not predictable whether a different metabolic labeling pattern affects the CuCAA-mediated formation of the covalent linkage of E_2A-Az and CD46-Alk molecules. However, the usage of SiaNAlk precursors may lead to efficient labeling of *N*-glycans of CD46 receptors.

Finally, first hints were provided that purifying E_2A-receptor complexes is feasible using IgG affinity beads following chemical linkage of E_2A-Az and surface-exposed glycoconjugates from MDBK-Alk cells. CD46-specific western blotting of eluates revealed two bands that may represent E_2A-CD46 complexes. Since the ProtA-tag fused to E_2 proteins also interacts with IgG antibodies, other E_2A-receptor complexes are plausible as well. A comprehensive analysis of the purification of E_2A-receptor complexes is given in **Figure 1.14**. It is apparent that CD46 molecules are present in the first washing fraction W_1 in the case of MDBK-Alk cells. Of note, the washing buffer W_1 contained 10% L4 lysis buffer and hence 0.1% desoxycholate (DOC) and 0.2% *N*-octyl-β-glucopyranoside (NOG). Since E_2A-Az proteins predominantly form dimers in aqueous solution under non-reducing conditions, it is imaginable that the interaction of E_2A-Az dimers with CD46 receptors lead to clustering of CD46 receptors on the cell surface. Subsequent cell lysis and solubilization of membrane proteins may result in micelles or other membrane-like complexes (e.g. liposomes), which contain more than one CD46 molecule. Due to the presence of detergents in W_1 buffer, various CD46 molecules may be solubilized, which is in turn detected in the supernatant of IgG beads. It may also simply be an impurity originated from the supernatant. To solubilize all membrane proteins in an efficient manner, the detergent-to-protein ratio should be ten or higher[67,68]. Unfortunately, the total protein concentration of the lysate of $3 \cdot 10^7$ MDBK cells has not been determined. However, the critical micelle concentration (CMC) was exceeded for both detergents

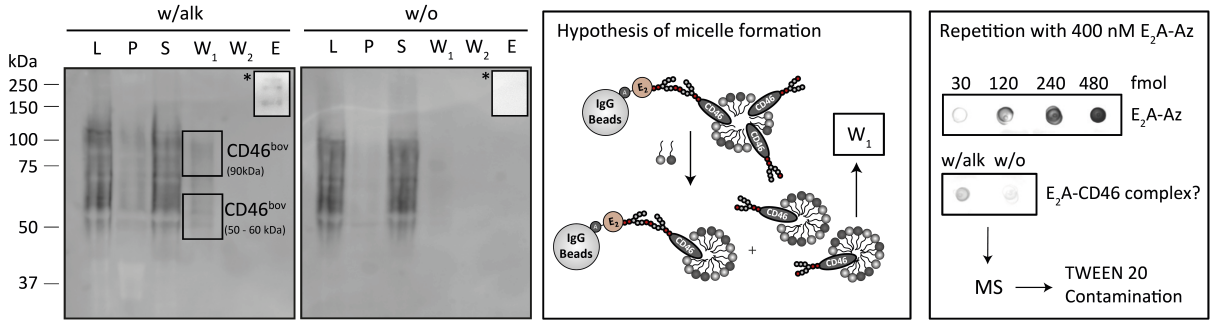


Figure 1.14 | Evaluation of purifying E2A-receptor complexes from MDBK cells
On the left side, the α-CD46^{bov}-probed western blots show samples originated from purification steps, which are illustrated in **Figure 1.12** (w/alk, MDBK-Alk cells; w/o MDBK cells). The lysate (L), pellet (P), and supernatant (S) after adding IgG beads followed by centrifugation show a smeared band, indicating either unspecific antibody binding or different glycosylation pattern of CD46 molecules. The first washing buffer (W_1) contained 10% L4 lysis buffer for which a typical band pattern is visible representing CD46 splicing species, which is missing in the negative control. One possible explanation is depicted beside and described in the main text. In lane E of IgG eluates, the contrast of a particular region was increased to show weak bands that may represent E_2A-CD46 complexes (asterisks). On the right side, α-E_2^{BVDV}-probed dot blots represent samples with increasing amount of E_2A-Az and IgG eluates originated from lysed MDBK-Alk cells (w/Alk) or native MDBK cells (w/o), which were incubated with 400 nM E_2A-Az. Samples were subjected to mass spectroscopy (MS), but a contamination with TWEEN 20 impeded MS analyses.

(i.e. 1%/24 mM DOC; 8 mM CMC and 2%/68 mM NOG; 24 mM CMC), which is compulsory to permit solubilization of transmembrane proteins. To increase the amount of E_2A-receptor complexes, the aforementioned experiment was repeated as described in **Figure 1.12C,** but the concentration of E_2A-Az had been increased from 100 nM to 400 nM. Moreover, a volume of 4 ml L4 lysis buffer was utilized to increase the fraction of solubilized membrane proteins. The presence of E_2A proteins/complexes in the eluate of IgG beads originated from MDBK-Alk cells was confirmed by α-E_2^{BVDV}-treated dot blots using a nitrocellulose membrane (**Figure 1.14**, right). A reference dot blot of E_2A-Az with defined quantities indicated a total amount of E_2A-receptor complexes in the range of 30 to 120 fmol. Considering that cell receptors for viruses are often relatively sparse, ranging from 10^3 to 10^5 per cell[69], the anticipated amount of CD46 molecules of $3 \cdot 10^7$ MDBK cells would be in the range of ~50 fmol to ~5000 fmol, which is in agreement with the obtained results. However, the fraction of covalently bound E_2A-Az is considered to be low (data not shown). Due to the limited amount of material, the entire samples that contained E_2A-Az protein and/or E_2A-receptor complexes were subjected to mass spectrometry analyses, but traces of TWEEN-20 that was present in the storage buffer of IgG beads hampered the identification of potential BVDV receptors, in spite of rigorous washing of beads before usage.

In conclusion, the detection limit of immunoblotting in combination with the limited amount of E_2A-receptor complexes did not allow an unambiguous proof of the purification of CD46 molecules using metabolic glycan labeling followed by *click chemistry*. There may be several reasons, which would result in a reduced efficacy of the click reaction between E_2A-Az and its alkyne-labeled receptors. First, the labeling pattern of CD46 glycans, in particular, the suspected lack of modified *N*-glycans in MDBK-Alk cells (see above). Second, the hampered dimerization of E_2A proteins through one intermolecular disulfide bridge, in part because of the ProtA-tag (data not shown), and especially due to the presence of the reducing agent, ascorbate, which is present in the click solution. The latter may also lead to destabilization of E_2 during the click reaction as 16 cysteines are involved in intramolecular disulfide bridges that are supposed to be essential for the structural integrity[48]. Consequently, binding of E_2A-Az to cellular receptors may be hindered under reducing conditions. The amount of sodium ascorbate may be reduced in future studies. Moreover, the temperature may be increased from 4 to 37 °C to accelerate the click reaction and, in parallel, to reduce the reaction time to five minutes. Third, the amount of lysis buffer was seemingly to less to solubilize all transmembrane protein complexes. The total protein concentration of lysed MDBK cells should be determined using the Bicinchoninic acid (BCA) protein assay method. It is compatible with certain amounts of detergents[67], which ensures a detergent-protein ratio of at least ten. Fourth, it has been reported that glycans can be removed by sonifiction (2 x 5 min at 45 °C) at high pH[70]. Yet, it is doubtful that short pulses of 3 s sonification would influence glycan stability, as the glycosylation pattern of E_2A-Az was seemingly unaffected upon sonification (cf. **Figure 1.12B**). Taken together, optimizing both the formation and the purification of covalently linked E_2A-receptor complexes may be challenging, but increasing the amount of cellular material and viral protein in combination with highly sensitive mass spectrometry may provide the proof of principle that virus receptor identification is achievable by the use of metabolic glycan labeling and CuAAC-mediated *click chemistry*.

1.6 Summary and Outlook

The first crucial step of the viral life cycle is the attachment of viruses to cell surface components. Identifying virus attachment factors and receptors is necessary to fully understand the mechanisms of virus infection and to develop suitable antiviral therapies. This study aimed to evaluate the possibility of virus receptor identification by using metabolic labeling of glycan moieties of both viral glycoproteins bearing azido groups and cell receptors bearing alkyne groups. This is followed by the formation of covalent linkages of glycan chains by a copper-catalyzed azido-alkyne cycloaddition (CuAAC), which is also termed *click chemistry*.

For the labeling of glycan structures, the following per-*O*-acetylated monosaccharide precursors were utilized: *N*-(2-azidoethan-1-one-yl)-glucosamine (Ac_4GlcNAz), *N*-(pent-4-yn-1-one-yl)-glucosamine (Ac_4GlcNAlk), *N*-(2-azidoethan-1-one-yl)-galactosamine (Ac_4GalNAz), *N*-(pent-4-yn-1-one-yl)-galactosamine (Ac_4GalNAlk), *N*-(2-azidoethan-1-one-yl)-mannosamine (Ac_4ManNAz), and *N*-(pent-4-yn-1-one-yl)-mannosamine (Ac_4ManNAk). The incorporation of synthesized monosaccharide precursors into MDBK cells and Huh7 cells was confirmed by biotin-labeling based on *click chemistry* and subsequent detection by streptavidin-Alexa Fluor® 488 fluorescence. In order to optimize the click reaction, the copper-ligand BTTAA was synthesized and several reaction conditions were examined, including reaction time and copper (Cu) concentrations. Best results were obtained after 15 min incubation at 4 °C with a reaction mix that contained 500 µM BTTAA, 250 µM $CuSO_4$, 200 µM biotin-alkyne or biotin-azide, and 2.5 mM sodium ascorbate.

In addition, the ProtA-tagged ectodomain of the structural protein E_2 originated from bovine viral diarrhea viruses (BVDV) was recombinantly synthesized using HEK293T cells. The production of azido group-containing E_2A was established by adding per-*O*-acetylated synthetic monosaccharide derivatives (Ac_4GlcNAz, Ac_4GalNAz, and Ac_4ManNAz) into the cell culture medium for several days. Following purification using IgG affinity chromatography, the biological activity was validated by inhibiting BVDV infectivity.

Furthermore, indirect immunofluorescence targeting BVDV E_2 epitopes strongly suggested a specific formation of covalent linkages of E_2A-Az proteins with alkyne-bearing glycoconjugates of MDBK cells. In contrast, alkyne-bearing Huh7 cells did not show any covalent coupling with E_2A-Az when the click reaction was performed.

Next, the enrichment of biotinylated glycoproteins derived from MDBK cells was confirmed by immunoblotting and detection of the viral receptor CD46. Moreover, the purification of E_2A-Az molecules was optimized using a micro-scale batch method. This led to preliminary results from western and dot blotting, indicating a successful purification of covalently linked E2A-receptor complexes. Since the presence of CD46 molecules has not been proven by mass spectrometry (MS) yet, the ultimate proof of principle is still lacking. However, this present study provides a suitable protocol for the metabolic labeling of living cells and for the purification of E_2A-receptor complexes. Moreover, valuable hints are offered on how to optimize this protocol. The labeling rate of *N*-glycans of CD46 molecules may be increased by supplementing alkyne-tagged sialic acid to the mixture of monosaccharide precursors. The time of click reaction may be reduced to 5 min and, simultaneously, the temperature may be increased to 37 °C to raise the efficiency of the click reaction between azido groups of E_2A and alkyne groups of glycoconjugates of MDBK cells.

Furthermore, a sufficient amount of detergents should be experimentally determined to ensure complete solubilization of membrane proteins. This can be validated by removal of all membrane-like structures using centrifugation at 100,000 *x g* for 1 h at 4 °C[67]. A comprehensive detergent screening[71] may be conducted to increase the amount of E_2A-receptor complexes in solution, including other MS-compatible detergents such as N-dodecyl-β-D-maltoside (DDM)[71], Maltose-neopentyl glycol[72] (MNG) or RapiGest surfactant[26] besides desoxycholate (DOC) or *N*-octyl-β-glucopyranoside (NOG). It is recommended to not use Triton X-100, NP-40 or TWEEN-20 in any step of preparation or purification, because small impurities will hamper the analysis via MS.
Following successful purification of E_2A-receptor complexes using IgG affinity beads, potential receptor molecules may be released by PNGase F digestion or treatment with endoglycosidase H[73]. Once BVDV cell receptors have been identified via MS, receptor capturing using *click chemistry* should be repeated with CD46-deficient MDBK cells to identify unspecific interactions. Moreover, metabolic labeling may also be applied on entire viruses harboring glycoproteins, as this has been demonstrated for measles viruses[42]. In the case of BVDV comprising three different glycoproteins, more than one coreceptor may be discovered. Finally, this methodology could be utilized also for other enveloped viruses, e.g. West Nile viruses, Dengue viruses or Zika viruses.

Chapter 2 | Fragment Screening Targeting Viral Proteins based on NMR Spectroscopy

1 Introduction - Fragment-based Lead Discovery

Twenty years ago, high-throughput screening (HTS) of millions of fairly large compounds (~500 Da) was the dominant lead-discovery approach. Nowadays, fragment-based lead discovery and rational drug design are becoming more and more important, in particular, due to the development of highly sensitive techniques to detect low-affinity ligands in the micromolar to millimolar range. The most powerful and popular methods are nuclear magnetic resonance (NMR) spectroscopy and surface plasmon resonance (SPR) spectroscopy. There are several advantages of fragment screening compared to conventional HTS[74]. It is assumed that low-molecular-weight compounds (<300 Da) sample a higher fraction of the biologically relevant chemical space. Hence, much fewer compounds (~1000 molecular with less than 20 non-hydrogen atoms) are required to identify positive hits for a given drug target. Another reason why it is beneficial to screen smaller molecules is the fact that small molecules tend to engage in fewer but high-quality interactions. In contrast, as molecules become larger and more complex, they have more possible interactions and hence the likelihood increases to form unfavorable contacts or steric clashes with the receptor binding site. This concept designated as 'molecular complexity' was first introduced by Hann and co-workers in 2001[75]. Finally, starting the chemical optimization pipeline with low-molecular-mass compounds is likely to produce a lead structure with better pharmaceutical and drug-like properties. Many fragment-based drug design (FBDD) campaigns yielded drugs that have entered clinical phases[76]. The first FBDD-derived drug was vemurafenib, which was approved by the food and drug administration in 2011[77], demonstrating the usefulness of fragment-based approaches.

1.1 Nuclear Magnetic Resonance Spectroscopy in Drug Design

Besides SPR[78,79] and crystallography[80], NMR[81–86] methods are tremendously suitable to validate and characterize weak fragment-receptor interactions at the atomic level[81,87]. One significant benefit of NMR techniques is the fact that binding studies are carried out in solution under nearly physiological conditions. Moreover, NMR provides a variety of experimental formats and hence allows to shed light onto different aspects of molecular interactions, starting with a simple yes/no answer regarding binding, followed by affinity measurements towards the determination of the binding mode that includes structural information and/or molecular dynamics. In general, there are two different types of strategies: ligand-observed or protein-observed NMR experiments.

In order to screen hundreds or thousands of molecules, the two most widespread ligand-observed applications are STD NMR (**S**aturation **T**ransfer **D**ifference)[87] and waterLOGSY (water **L**igand-**O**bserved by **G**radient Spectroscop**Y**)[88]. Both methods take advantage of a long-lasting reservoir of magnetization (stored in the protein or bulk water) that is imprinted onto bound ligands, which in turn, by exchange, is accumulated in solution where it is detected. These experiments can be conducted even at very low fractions of bound ligands[88], thus, only small amounts of the protein of interest are required (≤20 μM).

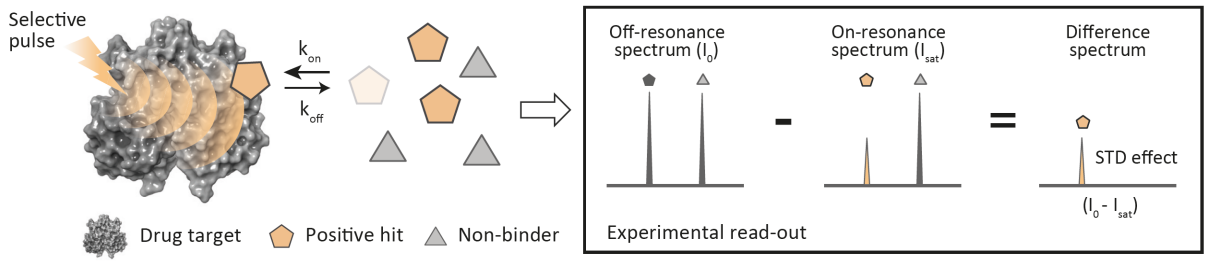


Figure 1.1 | Schematic representation of the STD NMR experiments
In a so-called on-resonance experiment, the protein is selectively saturated by a train of lower-power Gaussian-shaped pulses in which the frequency is set near protein signals (0 to -4 ppm) without directly affecting the ligands. The saturation is spread throughout the protein via spin-diffusion and transferred to bound ligands through cross-relaxation, which is based on the nuclear Overhauser effect (NOE). In contrast, free ligands do not experience any change in their magnetization. Due to the long-lasting saturation of the protein, many ligands can bind and dissociate during acquisition, which increases the fraction of saturated ligands in solution where they are detected. In a control experiment, the so-called off-resonance experiment, the frequency of selective pulses is set far away (e.g. 80 ppm) from any protein or ligand signals to obtain a normal 1D spectrum. By subtracting the on-resonance spectrum from the off-resonance spectrum, binding ligands can be readily detected. (k_{on} = association rate, k_{off} = dissociation rate).

WaterLOGSY is often applied for small targets or macromolecules with low proton density such as RNA, where spin diffusion is inefficient. STD NMR is very beneficial when working with larger proteins (> 50 kDa) or even with oligomers such as virus-like particles (~10 MDa)[11,89–91]. In this study, STD NMR has been utilized as an initial read-out to screen fragments against viral drug targets, including the West Nile virus (WNV) protease NS2B-NS3^{pro} and the P domain of the major capsid protein VP1 of several Noroviruses (NoV) strains. The main advantage of the STD NMR method is that several fragments that are present in a mixture can be screened in parallel without deconvolution. Binding fragments are readily detected by so-called STD effects. **Figure 1.1** illustrates the principle of STD NMR experiments.

For the validation of positive hits from STD NMR screening, numerous complementary NMR techniques and biophysical or biochemical methods are available, and some were used in this study (see below). Chemical shift perturbation (CSP) experiments are very reliable, provided that pH and salt content of the assay buffer remain constant during experiments. One advantage of this protein-observed method is not only that effects on the protein are directly detected, but the integrity of the protein can be confirmed at the same time. On the other hand, isotope-labeling (e.g. ^{15}N or ^{13}C) of the protein is required, rendering this NMR method rather expansive. In particular for larger proteins (>30 kDa), this is often combined with perdeuteration of the protein of interest, as replacing of non-exchangeable protons by deuterons leads to a reduction of relaxation rates and therefore to higher resolution in e.g. heteronuclear single quantum correlation (HSQC) spectra. Moreover, the usage of other labeling strategies such as ^{13}C-methyl ILV-labelling[92,93] has dramatically raised the size of macromolecules that can be studied by protein-observed NMR methods. In addition, so-called TROSY-based experiments[94] (**T**ransverse **R**elaxation-**O**ptimized **S**pectroscop**Y**) improve the spectral resolution even further, which can be critical to probe weak interactions of small molecules. Briefly, TROSY suppresses the fast transverse relaxation (T_2) of ^{1}H, ^{15}N, or ^{13}C nuclei, which is caused by chemical shift anisotropy (CSA) and dipole-dipole (DD) relaxation in a protein-size-

dependent manner (T_2 decreases with molecular size). In a system of two coupled spins (e.g. NH), J-coupling causes splitting into four multiplet components in the case of large biomolecules. The interference between the aforementioned relaxation mechanisms leads to different relaxation rates for the four multiplet components. The TROSY pulse sequence selects exclusively one component with lower transverse relaxation rate ($R_2 = 1/T_2$), thus giving rise to only one narrow NMR signal (instead of four). Using perdeuterated isotopically labeled proteins (100 – 200 µM) in combination with TROSY allows for the determination of the dissociation constant of a given ligand within less than one day of measurement time[95]. Additionally, it is possible to probe binding sites and (with a bit of computational effort) the orientation of the bound ligand, provided that both a crystal structure and chemical shift assignment are available[95].

1.2 Complementary Validation Methods

In this work, besides NMR experiments several other methodologies were utilized to confirm fragment binding to the protein of interest and to reduce the number of suitable hits in terms of affinity or biological activity. These methodologies can be divided into binding assays (SPR) and functional assays (protease activity and virus inhibition assays) as depicted in **Figure 1.2**.
SPR spectroscopy has emerged as an alternative biophysical technique for fragment screening (see REF[79] and references therein). This technique is based on the so-called surface plasmon effect[96]. In a nutshell, proteins or other biomolecules are immobilized onto a dextran matrix by using e.g. amine coupling methods. This matrix is in turn coupled to a ca. 50 nm thin gold film placed on a glass layer of a biosensor chip. Incident light directed onto the gold film induces so-called surface plasmon polaritons or evanescent waves along a metal-dielectric interface under conditions of total internal reflection. These oscillations that reach into the medium with lower refractive index (i.e. the medium of the flow cell) are very sensitive to any change at the boundary such as adsorption of molecules to the conducting surface. Hence, following injection of potential ligands into the flow cell at continuous flow, binding events are detected in real-time due to the change of refractive index coupled with the change in reflective light. This change is detected by a photo-detector array and is proportional to the bound mass.

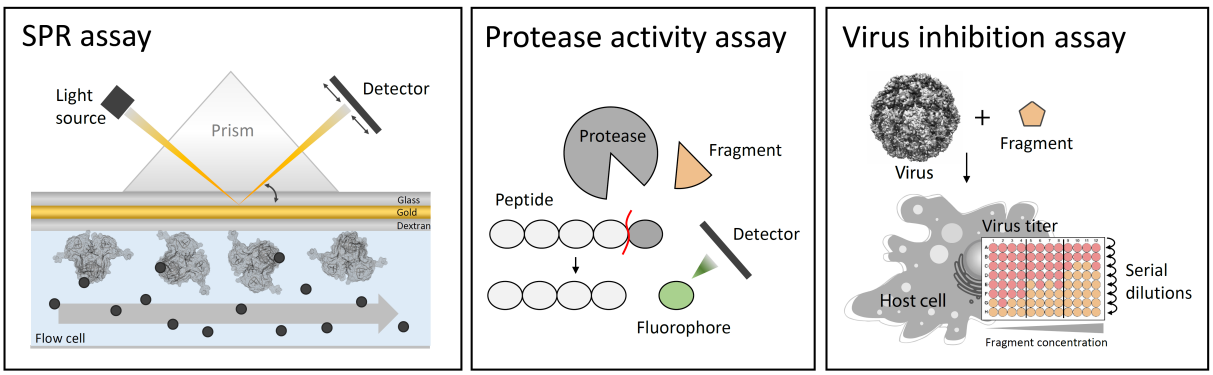


Figure 1.2 | Illustration of complementary methodologies used in this study to confirm positive hits

In the lead-discovery process, functional assays are inevitable when starting hit-to-lead optimization. However, some of these assays are prone to give false positive results. Therefore, a combination of several read-out methods should be always applied. In this study, the proteolytic activity of the WNV protease NS2B-NS3^{pro} was determined by monitoring the removal of the fluorophore 7-amino-4-methylcoumarin (AMC) from the peptidic substrate as illustrated in **Figure 1.2**. By using this assay format, not only binding is confirmed but also the inhibitory efficacy can be assessed, rendering this method useful in the search for potent lead structures[97].
Finally, another fundamental component of the pipeline of drug development is the cell-based assay, because many aspects such as toxicological effects or off-target interactions have to be excluded in the earliest stages of drug discovery[98]. In this study, the capability of murine noroviruses to infect susceptible cells was investigated by the use of endpoint titration in the presence of potential low-molecular-weight inhibitors.

1.3 *In silico* Docking

Nowadays, computer-aided drug design is becoming a vital part of lead discovery and rational drug design[99–103]. Provided that three-dimensional data of the protein of interest are available, virtual screening and molecular docking can guide the interpretation of experimental results and *vice versa*. A myriad of docking algorithms such as FlexX, X-Score, AutoDock, BLEEP, Glide and many more evolved over the past two decades. Along with the increasing amount of structural data of protein-ligand complexes with high resolution (≤2 Å), the accuracy of these docking methods is getting more reliable[99]. In general, these algorithms have two steps in common: First, ligands are placed into the respective binding pocket and, secondly, several poses are ranked against each other for each given ligand. In order to predict the binding free energy between a ligand and its protein target, many so-called scoring functions have been developed, which are based on three strategies: force field methods, empirical scoring functions or knowledge-based potentials (see REF[101,103] and reference therein). In principle, all these algorithms account for inter- and intramolecular contributions[103], including conformational forces such as torsional angles etc., van der Waals interactions, π-π stacking, hydrogen bonding, and electrostatic interactions. Some algorithms also include desolvation effects[103] of both the ligand and protein or simulate flexibility of receptor sites (induced fit docking[104]). In this study, Glide as implemented in Schrödiner's Maestro was utilized to obtain structural models of protein-ligand complexes. The extra precision scoring function and rigid receptor docking were used to provide starting points for rational drug design.

2 The West Nile Virus Protease

2.1 Introduction

2.1.1 The West Nile Virus – a Global Threat

Along with other human pathogenic RNA viruses, e.g. Hepatitis C virus, Yellow Fever virus, Dengue virus, and Zika virus, West Nile virus (WNV) belongs to the family *Flaviviridae*[105]. The mosquito-borne flavivirus was discovered in 1937 in Uganda[106]. Ever since WNV has become a global threat. During the last decades, WNV has spread from Africa and the Middle East into the southwest of Europe, the entire American continent, many parts of Asia, and Australia. It is assumed that the natural reservoir hosts of WNV are wild birds (e.g. blackbirds) based on the high viremia after WNV infection in these animals. The disease is transmitted to people almost exclusively via mosquito bites since WNV uses arthropods as vectors[106–108].
In most cases the WNV infection is asymptomatic. Some infected individuals show mild flu-like symptoms, including fever, headaches, and muscle pain. However, in immunodeficient patients, WNV may cause fatal neurological diseases, such as encephalitis, meningitis, or polio-like paralytic syndrome[109–112]. Moreover, the consequences of a persistent infection have not been entirely investigated[113,114].

2.1.2 Pathogenesis and Flaviviral Life Cycle

It is believed that during infection WNV initially replicates in dendritic cells and keratinocytes, and subsequently spreads throughout the whole organism via lymph nodes and blood circulation[110,115]. The molecular basis of this tissue tropism is not well understood[116,117].
The mature flaviviral particle is about 50 nm in diameter. The particles contain viral capsid proteins that are associated with the positive-sense single-stranded RNA genome that is surrounded by a lipid bilayer, which embeds 180 copies of two glycoproteins[1]. Following receptor-mediated endocytosis and acidification of endosomal vesicles, the viral RNA genome is released into the cytoplasm. Utilizing the cellular machinery, the ~11 kb genome is then translated into a single polyprotein (**Figure 2.1**). Cellular proteases and the WNV two-component protease NS2B-NS3^{pro} cleave the genomic product into three structural proteins (envelope, E; membrane precursor, prM; capsid, C) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5)[1]. The latter proteins (except NS2B-NS3^{pro} and NS4B) are directly involved in viral RNA replication located in an intracellular membranous web[118]. The immature virus particles assemble at the surface of the endoplasmic reticulum. During exocytosis along the trans-Golgi network, the viral particles are proteolytically processed by furin proteases, releasing infectious virions into the extracellular space.

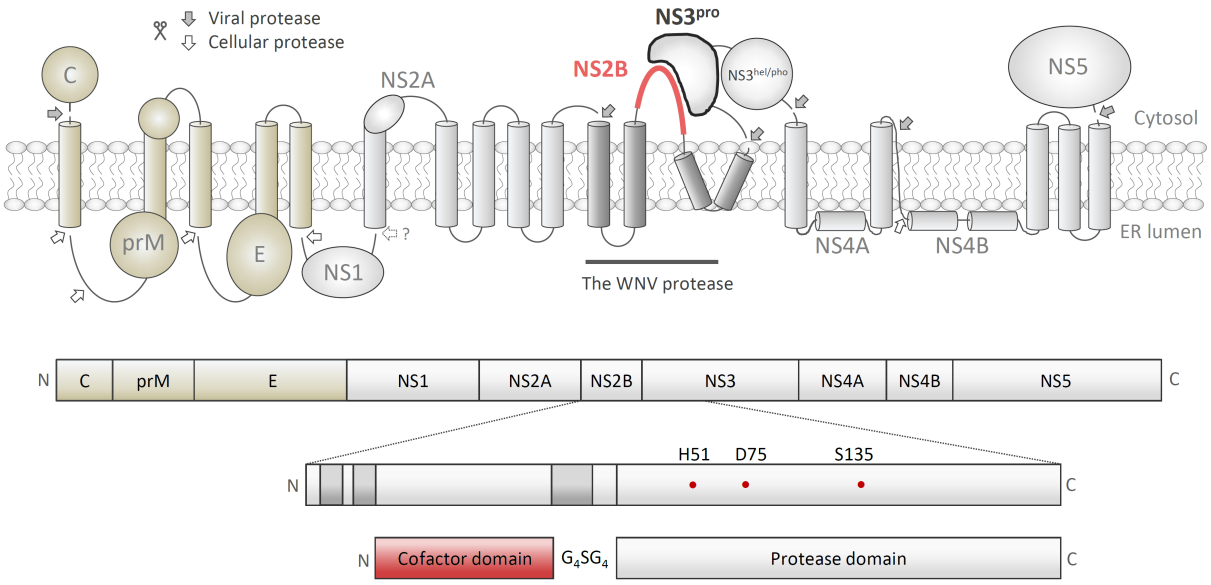


Figure 2.1 | Organization and polyprotein processing of WNV proteins
The scheme above represents the putative topology of flaviviral proteins after proteolytic cleavage via both the WNV protease NS2B-NS3^{pro} and cellular proteases (arrows) at the membrane of the endoplasmic reticulum (ER). Structural proteins and non-structural proteins are depicted in beige and gray, respectively. The lower panel shows the linear organization of the polyprotein. In order to enable biochemical studies of the WNV protease in an aqueous environment, Nall et al.[119] have designed a water-soluble fusion protein. Therefore, the essential part of the cofactor domain of NS2B (red) was linked to the NS3 protease domain (NS3^{pro}), which embeds the catalytic triad (His51-Asp75-Ser135). The membrane-associated domain (dark gray) was replaced by the nonapeptide GGGGSGGGG. (The scheme is modified and adapted from REF[120] and REF[105])

2.1.3 The WNV Protease – a Pivotal Drug Target

Despite the emerging global threat of WNV, no specific antiviral therapy for humans that prevents or treats WNV infection is currently available. The non-structural protein NS3 plays a critical role in several steps of the virus life cycle, including polyprotein processing, RNA replication, and formation of virus particles[118,121,122], therefore it is considered as a central drug target.
It is well established that the N-terminal part of the non-structural protein 3 (NS3) is essential for polyprotein processing. The multifunctional NS3 shows protease activity (NS3^{pro}) that strongly depends on the presence of a cofactor domain consisting of ~40 residues of the membrane-associated NS2B protein[119]. Besides the protease activity, NS3 contains other functionalities, such as a nucleotide triphosphatase, as an RNA triphosphatase, or as a helicase[121,123,124]. Due to its essential role in the viral life cycle, WNV NS3^{pro} has been considered as a promising drug target for more than a decade. In the past, researchers have utilized a two-component fusion protein NS2B-NS3^{pro} that has NS2B linked to NS3^{pro} via a nonapeptide sequence (G_4SG_4), conferring protease activity to the construct. NS3^{pro} comprises a chymotrypsin-like structural motif that consists of two six-stranded beta-barrels. The center of NS3^{pro} encompasses the catalytic triad (His 51, Asp 75, Ser 135) and Gly 133, which is part of the oxyanion hole[119,125,126]. Comprehensive investigations, including crystallography[126–129], small-angle X-ray scattering[130], and NMR studies[131,132], suggest that especially the β-hairpin of the NS2B cofactor is highly flexible and exists as an ensemble of

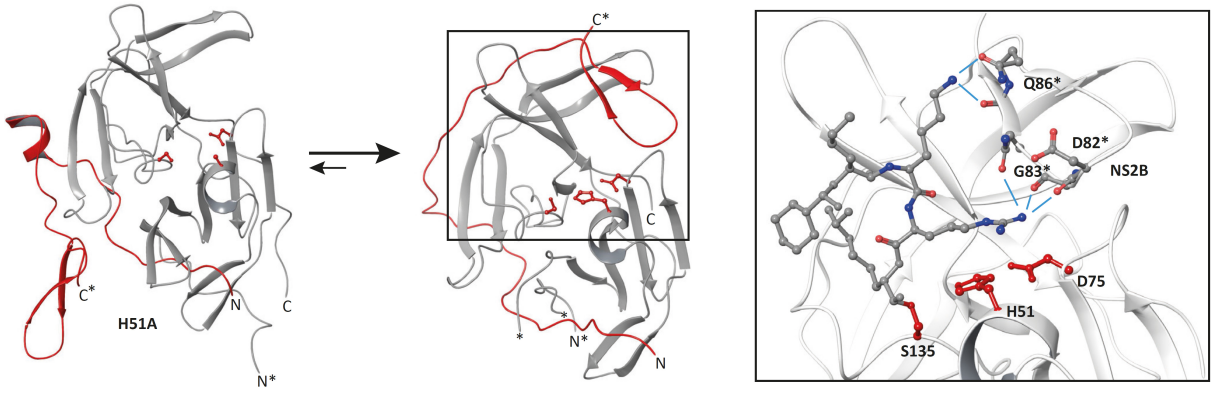


Figure 2.2 | Two putative conformational states of the WNV protease NS2B-NS3^{pro}
Two crystal structures of NS2B-NS3^{pro} are shown as ribbon models representing the so-called open conformation (left, PDB: 2GGV) and the closed/active conformation (right, PDB: 2FP7). The protease domain NS3^{pro} embracing the catalytic triad (H51, D75, and S135) and the essential cofactor domain NS2B are colored in gray and red, respectively. The region between C* and N* and between the asterisks in NS3^{pro} is not resolved due to low electron density. NS2B is critical for protease activity, since in the closed conformation residues of the β-hairpin (e.g. D82*, G83*, and Q86*) are likely involved in substrate and inhibitor binding. On the right side, the covalently bound peptidic inhibitor N-benzoyl-L-norleucyl-Lys-Arg-Arg-aldehyde is depicted. Blue lines indicate hydrogen bonding. (The scheme is adapted and modified from REF[129,131])

conformations in the absence of a substrate or an active site inhibitor. It is believed that the binding of a ligand to the active site stabilizes a distinct conformational state[133]. In this so-called closed conformation, the C-terminal part of NS2B wraps around NS3^{pro} and stabilizes substrate binding as further substantiated by mutagenesis studies[134,135].

During the last years, many peptidic[105,136,137], peptidomimetic[126,128,138–141], and non-peptidic inhibitors[142–150], which are directed against the catalytically active site, have been developed. Some of these inhibitors were considerably potent[137,139] but to date, none of them has reached preclinical trials. Due to the characteristics of the active binding site competitive inhibitors mostly possess positive charges, making passage through membranes an issue. Other inhibitors such as highly reactive aldehyde derivatives exhibit low specificity as well as little stability in an aqueous environment. Therefore, researchers have started to search for secondary binding sites by using e.g. high throughput screening (HTS) combined with rational drug design[151–154], or virtual screening[155]. Albeit some of these inhibitors investigated showed improved drug-like properties[154,155], promising lead structures are still not available. As an alternative, fragment-based drug design targeting secondary binding sites may circumvent some of the drawbacks of HTS[74,76,156–158]. As one advantage, fragment-based screening can identify small and even shallow binding pockets as parts of larger binding sites, allowing the synthesis of highly selective inhibitors via e.g. fragment linking[159,160]. In this study, fragment-based screening of the Ro5 Maybridge library was conducted by the use of saturation transfer difference NMR spectroscopy (STD NMR)[87] as an initial read-out[81,90,161–167].

2.2 Objectives

The goal was to identify novel and potent WNV protease inhibitors targeting secondary binding sites. Therefore, the Maybridge library CR2006 was screened during my master's thesis (Schöne, 2014), employing STD NMR as initial read-out followed by protease activity assays and STD NMR-based competition experiments with the known active site binder fragment **A**. In this study, the search for potential allosteric inhibitors was followed up that involved following steps:

- Find chemical descriptors beyond the Tanimoto coefficient that characterize binders and non-binders, which may guide the screening process.
- A validation of STD NMR competition assays had to be conducted. Therefore, this assay format was repeated for fragments, which were selected based on their inhibitory efficacy.
- Perform experiments that involve ligand-based and protein-observed NMR methods in order to identify fragments that interact with secondary binding sites.
- Fluorogenic competition assays (e.g. Dixon plot) may identify potential non- or uncompetitive inhibitors.
- Identify molecular key features that are important for the inhibition of the WNV protease. Thus, perform structure-activity relationship analysis (SAR) for particular fragments.
- Moreover, molecular docking of important fragments will be beneficial to guide the optimization of promising hits towards lead structures.

The following questions had to be addressed to guide the process of rational drug design:

- Does the water-soluble two-component protease sufficiently reflect the *in-vivo* membrane-associated state?
- Is it also important to address the question of whether assay buffers with different pH values would affect the outcome of a fragment screening?
- What are the potential causes of artefacts that may increase the rate of false positive hits?

2.3 Material and Methods

2.3.1 Production of the WNV Protease NS2B-NS3^{pro}

Unlabeled WNV Protease

The unlabeled WNV protease was prepared during my master's thesis (Schöne, 2014) and the biosynthesis is described in detail elsewhere[168]. Briefly, the E. coli strain BL21(DE3) was transformed with a plasmid pET22b, containing the sequence for the C-terminally His-tagged fusion protein NS2B(46-96)-Gly4-Ser-Gly4-NS3^{pro} (cf. Appendix A1). A point mutation K96A has been inserted to reduce auto-cleavage[169]. The plasmid was kindly provided by Dr. Noaki Sakai (Institute of Biochemistry, Prof. Rolf Hilgenfeld).
Next, bacterial cells were cultured in 2 L of 2 x YT^{amp} medium (16g tryptone, 10 g yeast extract, 5 g NaCl, and 100 µg/ml ampicillin) at 37 °C and protein over-expression was induced by IPTG addition and a temperature shift to 16 °C. Following 18 h incubation, cells were collected via centrifugation for 30 min (4200 x g, 4 °C) and lysed in 120 ml ice-cold lysis buffer (50 mM Tris/HCl, 5% (v/v) glycerol, pH 8.5), using a precooled French Press at 1200 psi (two passes).
Finally, NS2B-NS3pro was purified employing metal-affinity chromatography (5 ml Ni-NTA Superflow, Qiagen, equilibrated with lysis buffer) by applying sterile-filtered supernatant onto the column. The bound WNV protease was eluted with an imidazole gradient of 0 – 250 mM in lysis buffer. Fractions containing NS2B-NS3^{pro} were concentrated to 2 - 5 ml, using an Amicon ultra centrifugal filter (10 kDa cut-off, Millipore). Further purification was conducted employing size exclusion chromatography (SEC) (HiLoad® 26/600 Superdex® 75 prep grade, GE Healthcare) with SEC buffer (25 mM Tris/HCl, 5% (v/v) glycerol, pH 8.5).
The WNV protease was stored in SEC buffer at -20 °C. For NMR experiments, the buffer was exchanged several times against NMR buffer (25mM sodium phosphate in D_2O, pH* 7.05), using a 10 kDa cut-off filter device (Millipore).

Uniformly [^{2}H,^{15}N]-labeled WNV Protease

First, the viability of protein biosynthesis for uniformly ^{15}N-labeled, non-perdeuterated WNV protease was confirmed using M9 minimal medium in H_2O. The protocol is basically the same as for the {U-[^{2}H,^{15}N]}-labeling of NS2B-NS3^{pro} as described below: Approximately 3 ml 2 x YT^{amp} medium were inoculated with glycerol stock containing transformed BL21(DE3) (see above). Unless otherwise stated, bacterial cells were incubated at 37 °C in a shaking incubator (200 rpm) shielded from light. Bacterial cell growth was monitored by measuring the optical density at a wavelength of 600 nm (OD_{600}) employing a Jasco V550 photometer. Following 6 h incubation, 100 µl of bacterial cell suspension was diluted with 2 ml M9 minimal medium in D_2O (MM9+ modified from REF[92,93]; see **Table 2.1**). The volume was increased to 12 ml MM9+ after 4 h. Next day, the cell suspension was diluted with 500 ml MM9+ to yield an OD_{600} of ~0.1 After 6 h bacterial cell proliferation an OD_{600} of 0.9 units was reached, and protein expression was induced by addition of 1 mM IPTG followed by incubation for 42 h at 17 °C while shaking at 240 rpm.

Following protein biosynthesis, the uniformly {U-[^{2}H,^{15}N]}-labeled WNV protease (~34 mg) was isolated from the bacterial pellet (4.8 g) employing French press lysis, metal-affinity chromatography, and size exclusion chromatography as described above. Using a 10 kDa cut-off filter device (Millipore), the buffer was exchanged several times against sodium phosphate buffer (25 mM sodium phosphate in H_2O, pH 7.0). SDS-PAGE (cf. **1.3.1**) and NMR spectroscopy were used to confirm purity and structural integrity.

^{13}C-methyl-labeled WNV Protease

Selective labeling is suitable especially for high-molecular-weight proteins. However, even for relatively small proteins (<30 kDa), this strategy can have advantages since the number of NMR signals in a ^{13}C methyl-TROSY-HMQC spectrum is dramatically reduced when compared to a ^{1}H-^{15}N TROSY-HSQC spectrum of the same ^{15}N-labeled protein.

The preparation of the {U-[^{2}H,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Valγ-[$^{13}CH_3$]}-labeled WNV protease followed established protocols[169]. Briefly, the precursor alpha-ketoisovalerate (Na^+) (3-methyl-^{13}C; 3,4,4,4-D_4) (Cambridge Isotopes) and 2-keto-3-D_2-4-^{13}C-butyrate (see above) were used to selectively introduce ^{13}C-labeled methyl groups into the amino acid residues isoleucine, and leucine along with valine, respectively. Protein expression was performed in optimized M9+ minimal medium (see **Table 2.1**). First, transformed BL21(DE3) cells (see above) were cultured in 2 ml 2 x YT^{amp} medium for 6 h at 37 °C. Next, 2 ml M9+ medium was inoculated with 100 µl cell suspension. After 4 h, the volume was increased to 20 ml M9+ medium and cells were incubated overnight at 37 °C. The cell suspension was diluted with 240 ml M9+ medium. Once the cell density was high enough (OD_{600} ~ 0.4 – 0.5), 60 ml of ILV precursor in M9+ (without Silantes medium) were added. Following 1 h incubation, protein expression was induced by addition of 1 mM IPTG. The OD_{600} at this point was 0.73 units. The cells were cultured at 17 °C for 42 h (or at 18 °C for 92 h as described in section **2.4.2**), while shaking (240 rpm). Protein purification was conducted as described above, yielding ~16 mg protein. The purity was verified by SDS-PAGE (cf. **Figure 2.4** and section **1.3.2**) and NMR spectroscopy.

Deuteration of 2-keto-3-H_2-4-13C-butyrate

The protons of the methylene group at position 3 of 2-keto-3-H_2-4-^{13}C-butyrate (CortecNet) were replaced by deuterium atoms. Therefore, 60 mg was dissolved in 200 ml D_2O (99.8%), and 0.57 g KH_2PO_4 was added. The solution was incubated at pH 10.5 overnight at 45 °C in a sealed round flask punctured by a cannula and placed in an oil bath while stirring. Next, 1.2 g NaH_2PO_4 (2xH_2O) were added, and the pH was adjusted to 7.4. Sterile-filtered 20 ml aliquots were stored at -20 °C.

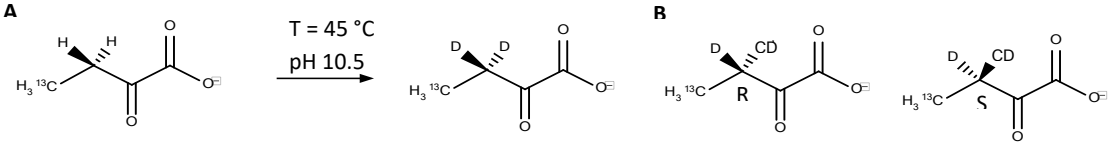


Figure 2.3 | (A) Hydrogen-deuterium exchange of α-2-keto-3-H_2-4-^{13}C-butyrate and (B) racemic mixture of α-ketoisovalerate (3-methyl-^{13}C; 3,4,4,4-D_4)

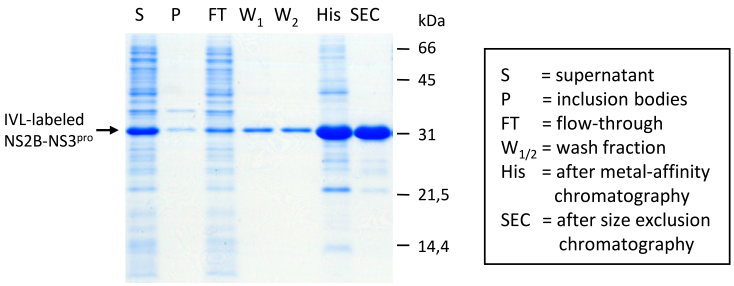


Figure 2.4 | Biosynthesis of ^{13}C-methyl-ILV-labeled NS2B-NS3^{pro}

Table 2.1 | Composition of the optimized minimal medium M9+

All Substances were dissolved in D_2O. The final concentration is listed beside each ingredient. To promote hydrogen-deuterium exchange, D_2O samples containing growth substances, salts, trace minerals, ampicillin, or precursor stock solution were lyophilized from D_2O twice and finally dissolved in D_2O. The pH* of M9+ was adjusted to 7.2 when required. Prior to inoculation, the minimal medium was sterile-filtered. The ILV precursor is missing in case of uniform ^{15}N-labeling.

Ingredient	^{15}N-labeling/ ^{13}C-ILV-labeling*	Comments
Na_2HPO_4	50 mM	Growth substances were prepared as a 10-fold stock solution just before usage. The composition was adapted from REF[170].
KH_2PO_4	50 mM	
$MgSO_4$	4 mM	
Na_2SO_4	5 mM	
$^{15}NH_4Cl$	50 mM	
Deuterated glucose-D_7	2 mg/ml	
$CaCl_2$	20 µM	Salts and trace elements were prepared as a 1000-fold stock solution and stored at RT.
$FeCl_3$	100 µM	
$MnCl_2$	10 µM	
$ZnSO_4$	10 µM	
$CoCl_2$	2 µM	
$CuCl_2$	2 µM	
$NiSO_4$	2 µM	
Na_2SeO_3	2 µM	
Na_2MoO_4	2 µM	
H_3BO_3	2 µM	
Thiamine (B1)	20 µg/ml	B1 was prepared as 20 mg/ml solution.
Riboflavin (B2)	0.1 µg/ml	All other vitamins were prepared as a 100-fold stock solution. Therefore, the powder was weighed and directly dissolved in D_2O. Due to degradation upon light exposure, the stock solution was always kept in the dark. The stock solution was stored at -20 °C.
Nicotinamide (B3)	1 µg/ml	
Choline chloride (B4)	1 µg/ml	
D-Pantothenate (B5)	1 µg/ml	
Pyridoxal hydrochloride (B6)	1 µg/ml	
D-biotin (B7/B8/H)	1 µg/ml	
Folate (B9)	1 µg/ml	
Cobalamin (B12)	1 µg/ml	
Myo-inositol	2 µg/ml	
Ampicillin	100 µg/ml	Unstable at 37 °C (half-life <24 h)[119]
Eco E.coli-OD2 DN	20%	Silantes ancillary medium
3-^{13}C-alpha-ketoisovalerate*	0.6 mg/ml	Leu/Val precursor
2-keto-3-D_2-4-^{13}C-butyrate*	0.3 mg/ml	Ile precursor
D_2O (99.85%)	Ad final volume	

2.3.2 Protease Activity Assay

The proteolytic activity of the WNV protease NS2B-NS3^{pro} was determined by monitoring the removal of the fluorophore 7-amino-4-methylcoumarin (AMC) from the peptidic substrate N-benzyl-norleucine-lysine-lysine-arginine-AMC (Bz-Nle-K-K-R-AMC) (**Figure 2.5**). Unless otherwise stated, assays were performed on black 96-well plates in 50 µl SEC buffer per well at 37 °C, employing a photometer (Biotek, Flx800). The software Gen5.1.11 (Biotek Instruments Inc) was used for data processing. The excitation wavelength was set at 360 nm, and AMC fluorescence was detected at 460 nm using an optical filter every 30 s during a period of 0.5 – 1 h. Prior acquisition of each data point, the plate was shaken automatically for 5 s. The detector sensitivity was set to 80.

Salt and pH Dependency

The investigation of proteolytic activity upon different salt concentration was performed during my master's thesis[133] (Schöne, 2014). Each well contained 25 nM WNV protease and 20 µM peptidic substrate in SEC buffer with varying amount of NaCl (0, 5, 10, 15, 20, 25, and 30 mM).To explore the influence of pH variation, the cleavage of 40 µM substrate was monitored in the presence of 25 nM protease in SEC buffer at different pH (7.4 – 8.7).

Fragment Screening

Positive hits from STD-NMR screening were dissolved in SEC buffer (see above) yielding a final fragment concentration of 2 mM. Since the protease activity assay is highly pH sensitive[119,133] (see results **2.5.2**, page 95 ff.), the pH of a stock solution of each fragment was measured and adjusted to pH 8.5 when required. For measurement, each well contained 900 mM fragment, 25 nM WNV protease, and 40 µM substrate in a total volume of 50 ml SEC buffer. Following 10 min incubation of protease and compound at 37 °C, the substrate was added immediately and fluorescence of AMC was monitored at 460 nm for 60 min (see above). A mixture of fragment and substrate without protease served as blank value for each fragment. Moreover, fragments were analyzed spectroscopically to determine the absorption maximum, because interference with AMC fluorescence was expected for some fragments. The concentration of fragments was 900 µM. Some particular fragments were further analyzed by detecting the fluorescence of 4 µM AMC in the presence and absence of 900 µM fragment.

Determination of IC_{50} values

The half-maximal inhibitory concentrations (IC_{50}) were determined using a fluorophore-based protease activity assay as described above during my master's thesis[133] (Schöne, 2014). Serial dilution gave fragment concentrations ranging from 1.4 to 860 µM yielding sigmoidal inhibition curves containing seven data points each. All measurements were done in triplicate. Curve fitting was performed using the Boltzmann model implemented in Origin 7.0 representing the sigmoidal curve:

$$\alpha = \alpha_{min} + \frac{\alpha_{max} - \alpha_{min}}{1+\exp\left(\frac{c-IC_{50}}{dc}\right)} \qquad \text{(I)}$$

with α being the relative protease activity and c being the ligand concentration.

Fluorogenic Competition Assay – Dixon Plot

Alternatively to NMR spectroscopy (see below), the inhibition mechanism of fragment **CR112** and **CR081** was explored using a fluorogenic competition assay. The proteolytic cleavage of Bz-Nle-K-K-R-AMC was monitored for 20 min as described above, applying 50 nM WNP protease. The initial rate v^{-1} $(RFU/min)^{-1}$ (RFU, relative fluorescence units) of AMC-peptide conversion was plotted against increasing fragment concentration at different amount of substrate (8, 12, and 24 µM) gave a family of intersecting lines. When lines intersect on or below the x-axis, non-competitive inhibitions can be assumed.

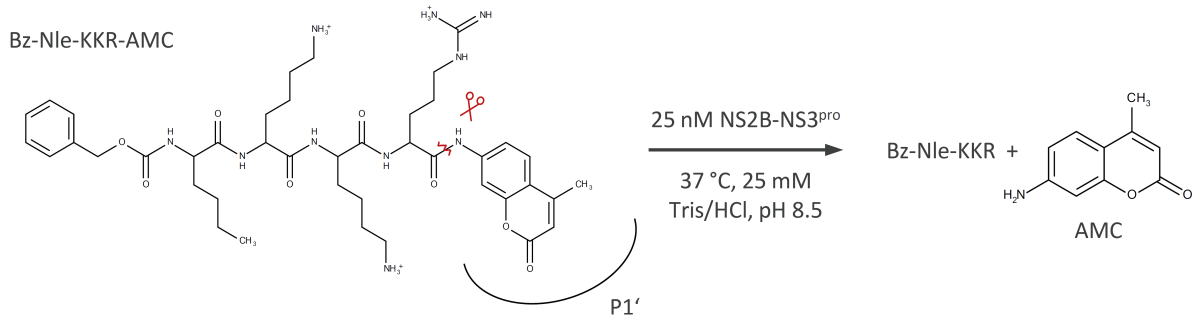


Figure 2.5 | Schematic representation of substrate processing by NS2B-NS3^{pro} in pocket P1'

2.3.3 Maybridge Libraries CR2006 and TS2017

The Maybridge CR2006 library was purchased from Thermo Fisher Scientific in 2006 by Dr. Christoph Rademacher. These fragments have been utilized in several studies[90,165]. Since samples containing 1 – 2 mM fragment in NMR buffer (25 mM sodium phosphate, D_20, pH* 7.05) were stored many years at -20 °C, both the integrity and the solubility of fragments were confirmed using NMR spectroscopy prior to STD NMR screening (section **3.3.3**). Approximately 83% of these 500 fragments were water-soluble at a concentration of at least 1 mM in NMR buffer. In order to get insight into the chemical structures, a scaffold-matrix (SxG) distribution matrix and a dendrogram of water-soluble members of the CR2006 library are depicted in **Figure 2.6** and **Figure 2.7**.

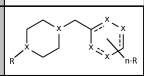
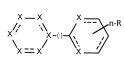
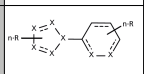
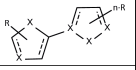
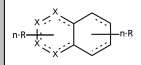
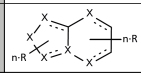
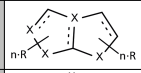
		G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	
	R = $(CH_2)_m$ +	COOH	COO-CH_3 or COO-CH_2CH_3	CO-CH_3 (R''-CO-R'')	OH	O-CH_3	NH_2	$NHCH_3$ or $N(CH_3)_2$	Halogens	Other groups	Only CH_3 or without any group	Σ
S1		1.4	1.0	0.4	3.9	0.2	4.7	1.4	1.6	0.2	0.2	15.2
S2		1.8	0.2	0.2	1.4		2.5	1.0	0.2	0.2		7.6
S3		4.5	0.2	0.4	5.5		7.2	1.6	0.6	0.4	0.6	21.1
S4		1.2	0.6		1.0		1.0					3.9
S5		3.1	0.8	0.8	1.6		2.3		0.2	0.4		9.2
S6		7.0	2.3	1.6	4.1	0.6	4.3	1.8	0.2	0.8		22.8
S7		0.6	0.8		0.4			0.2				2.1
S8			1.6	0.4	0.4	0.4	1.8		1.0	0.2		6.0
S9		0.2	2.3		0.4		1.4	0.2				4.5
S10	Other structures	1.4	0.2	1.4	1.4		1.2		0.6	0.4	0.8	7.6
	Σ	21.4	10.1	5.3	20.3	1.2	26.5	6.4	4.5	2.7	1.6	100

Figure 2.6 | SxG distribution matrix of the Maybridge CR2006 fragment library
The numbers in the matrix represent the ratio of subgroups in %. The skeletal structural formulas represent motifs that are present in the CR2006 library. The majority of 432 water-soluble fragments are aromatic molecules with several heteroatoms. The symbol X represents carbon, nitrogen, oxygen, or sulfur atoms. The scaffold cluster S10 resembles more complex molecular structure compared to the other clusters. In addition, most fragments are functionalized by chemical groups (R, n = 0, 1, 2, or 3 and m = 0, 1, or 2). These are carboxyl groups (G1), carboxylic ester groups (G2), other carbonyl groups, including acetyl groups, ketones, or amides (G3; R'' = ring system), hydroxyl groups (G4), methoxy groups (G5), primary (G6) and secondary or tertiary (G7) amino groups, halogens (G8), and other diverse groups that are underrepresented in the fragment library (G9, S-CH_3, SO_2-R, SO_2-NH_2, CN, NH-NH_2, CO-NH-NH_2, CO-NH_2, R-NH-CO-CH_3). Some fragments carry only methyl groups or lack functional groups (G10). Moreover, 55 fragments exhibit two functional groups. Here, the second group was included. For simplification, methyl groups as a second (78 fragments) or third group (16 fragments) were ignored.

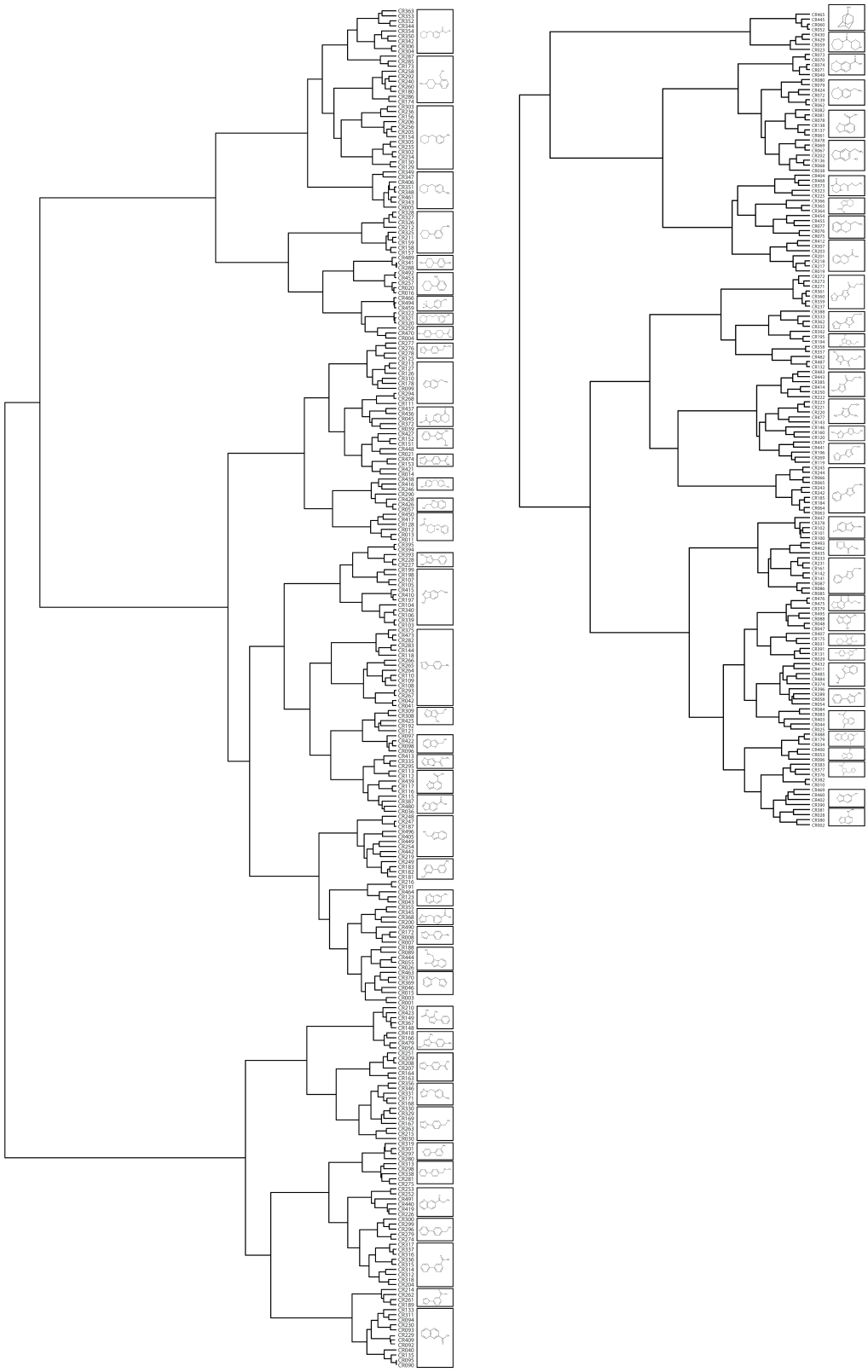


Figure 2.7 | Dendrogram of 432 fragments of the Maybridge CR2006 library
The dendrogram was created using the Matlab script *dendrogram.m* (cf. A4) based on the pairwise Tanimoto coefficients[171], which was calculated by using Open Babel implemented in the python script *makeTanimotoMatrix.py* (for details see Appendix A4). Each cluster comprises structurally similar fragments. Representative fragments are depicted beside each cluster. The compound ID (e.g., CR001) can be searched in the pdf-version of this figure using the strg-F keyboard combination.

In order to increase the diversity of chemical scaffolds, the Maybridge Library CR2006 has been extended in 2017. These fragments were purchased from Thermo Fischer Scientific and essentially selected based on chemical scaffolds and chemical groups. The SxG distribution matrix and the dendrogram for the Maybridge library TS2017 are shown in **Figure 2.8** and **Figure 2.9**.
The compounds were delivered in pure form as powder or liquid and stored at 4 °C. Fragments were weighted ~0.5 mg each and dissolved in 25 mM sodium phosphate buffer in D_2O (pH* 7.05) mixed with 50 µM TSP-d_4 to yield 2 mM final concentration. Samples were shaken for 1 h at 30 °C. To confirm solubility and stability of fragments, ^1H reference spectra of fragments at a concentration of 1 mM were acquired as described in section **2.3.7**. For some fragments (~15%), aggregates or particles have been observed. These samples were shaken overnight at 40 °C or ultra sonification was conducted for 30 min at RT to promote solubilization. About 95% were soluble at a fragment concentration of 1 mM. All stock solutions and NMR samples were stored at -20 °C.

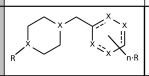
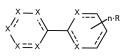
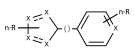
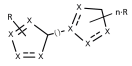
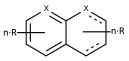
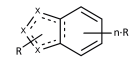
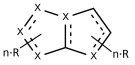
		G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	
	R = $(CH_2)_m$ +	COOH	COO-CH_3 or COO-CH_2CH_3	CO-R' R' = ring system or methyl group	OH	O-CH_3	NH_2	$NHCH_3$ or $N(CH_3)_2$	Halogens	Other groups	Only CH_3 or without any group	Σ
S1												
S2		0.8			2.9		2.4	0.4			1.2	7.8
S3					0.4		4.1	0.4				4.9
S4		5.7	3.7	5.7	3.7		3.7	0.8	0.8	12.7	4.5	41.2
S5		1.6		0.4	1.6	0.4	0.8			0.0		4.9
S6		4.1		0.4	2.4	0.4	3.3	0.4	3.3	2.9	1.2	18.4
S7		2.9	2.0	0.4	0.4		0.4			0.8		6.9
S8		1.2	0.4		0.8		0.4			0.8		3.7
S9		0.4	1.6	0.4	0.4		1.2		0.4	2.0		6.5
S10	Other structures	0.8	0.4	1.6		0.4			2.0		0.4	5.7
	Σ	17.6	8.2	9.0	12.7	1.2	16.3	2.0	6.5	19.2	7.3	100

Figure 2.8 | SXG distribution matrix of the Maybridge TS2017 fragment library
The description for scaffolds (S1 – S10) and chemical groups (G1 – G8) is stated in **Figure 2.6**. Chemical groups of category G9 are the following: S-H, S-CH_3, S-CH_2CH_3, RCSR, SO_2CH_3, NHC(=S)$NHNH_2$, NO_2, CH_2NHOH, C(=NH)NHOH, and ring systems. Some fragments carry only methyl groups or lacking functional groups (G10). Moreover, 26 fragments exhibit two functional groups. The second group was also included (except methyl groups, 34 fragments).

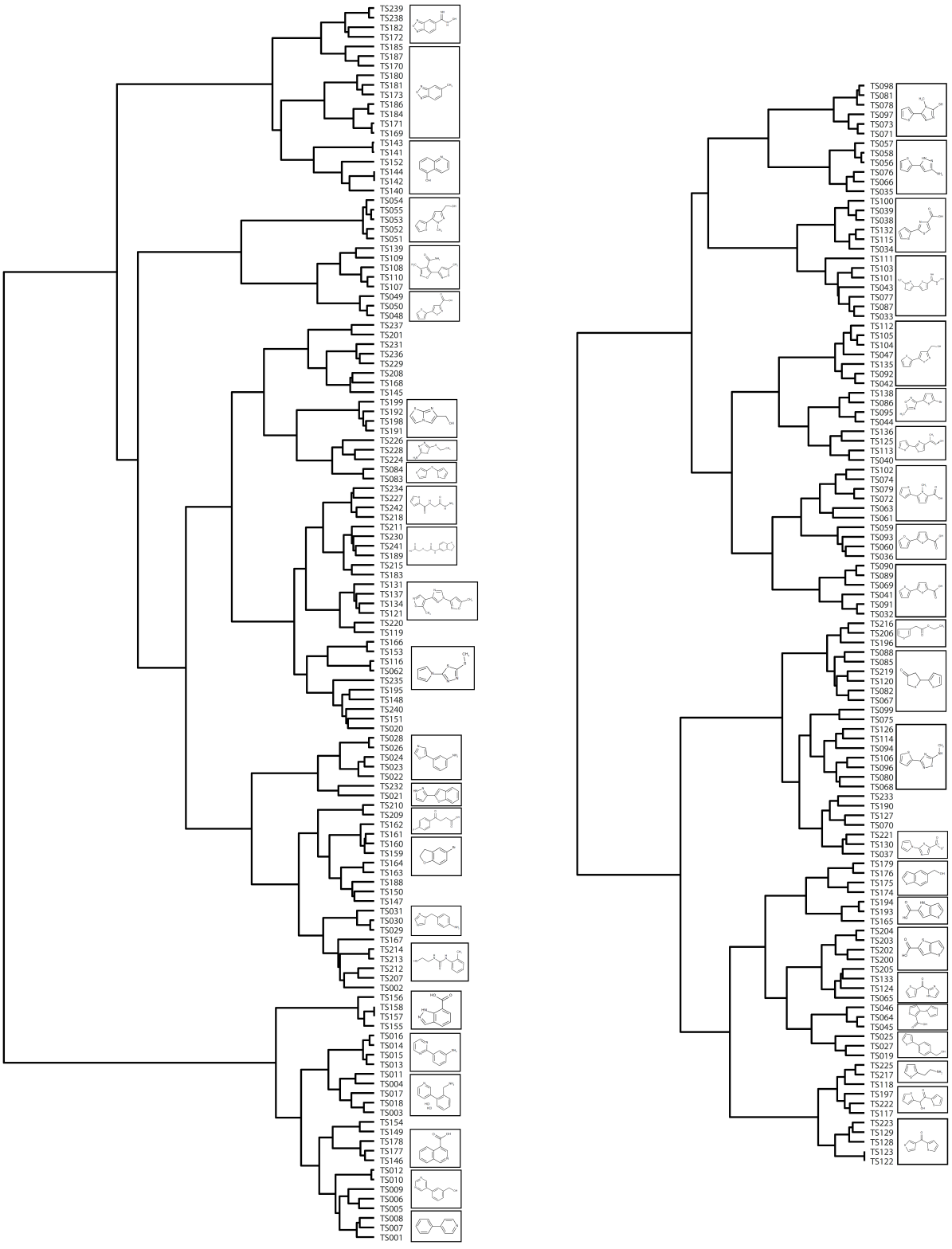


Figure 2.9 | Dendrogram of 238 fragments of the Maybridge TS2017 library
Clustering based on the Tanimoto coefficients (TAN) was conducted by utilizing the Matlab script *dendrogram.m* (cf. A4). The TAN was calculated using Open Babel implemented in the python script *makeTanimotoMatrix.py* (for details see Appendix A4). In order to find similar compounds, the compound ID (e.g., TS001) can be searched in the pdf-version of this figure (available on request). Typical chemical structures are depicted beside each cluster.

2.3.4 Chemoinformatics – Database of Fragments

As stated above, fragments of the Maybridge libraries were purchased from Thermo Fischer Scientific. The whole Maybridge collection comprises 30,000 small compounds (MW< 300). The complete set can be downloaded from the website www.maybridge.com as sdf-file. The software *Instant JChem* (ChemAxon) along with the excel-add-in *JChem for Excel* were used for data handling. For similarity searching within 670 fragments used in this study, a pairwise Tanimoto coefficient (TAN)[171,172] was calculated. Moreover, 23 parameters have been collected for each fragment using the ChemAxon's calculator as described below.

Tanimoto Coefficient – a Measure of Similarity

First, chemical structures have been converted to SMILEs strings (www.daylight.com). This simplified molecular input line entry system (SMILES) allows for the linear representation in which molecules are specified using ASCII characters representing atom and bond symbols (e.g. NC1=CC=C(Cl)N=N1). Next, these SMILES stored in an excel sheet were subjected to the chemical toolbox *Open Babel* (www.openbabel.org) utilized in the python script *makeTanimotoMatrix.py* (cf. A4). If not otherwise stated, all python scripts were written in *Python 3.6* using *PyCharm* as an integrated development environment on Windows 10. Basically, the python script converts SMILE codes from the excel sheet to .smi-files. Based on the fingerprint format (-ofpt) FP2, the output of the command line *"obabel fragment1.smi fragment2.smi -ofpt >temp.txt"* is the pairwise TAN stored as .txt-file. The path-based fingerprint FP2 indexes small molecules based on linear segments of up to seven atoms (cf. Open Babel documentation release 2.3.1). First, a molecule structure is fragmented into substructures of length 1-7 atoms. Single fragments of C, N, and O are ignored. A fragment is terminated when the atoms form a ring. These substructures are compared, and distinct entities are assigned a hash number from 0 to 1020. Afterwards, two 1024 bit vectors representing two different molecules are compared by counting matching entries. Finally, the TAN is the quotient of matching entries and the length of the one-bit vector. These value can range between 0 (high diversity) and 1 (high similarity).

Molecular Descriptors

In order to get insight into hallmarks of small molecules that interact with a specific macromolecule, hence, modulating its biological activity (e.g., inhibition of an enzyme), the concept Quantitative Structure-Activity Relationship (QSAR) was founded by Corwin Hansch in 1962[173,174]. Chemical descriptors are the bases of QSAR modeling, and myriad molecular descriptors have been proposed[173,175], reflecting various levels of chemical structure representation. These levels range from scalar descriptors such as molecular weight, a one-dimensional representation such as molecular formula or SMILES code, graph-based two-dimensional (2D) descriptors, geometrical descriptors (3D), and even higher levels taking into account mutual orientation and time-dependent dynamics of molecules.

The Lipinski's rule-of-five (Ro5) that evaluates druglikeness[176] represents one famous application of scalar features. It implies that orally active drugs have no more than one violation of the following criteria: < 10 hydrogen bond acceptor, < 5 hydrogen bond donors, < 5 rotatable bonds, molecular weight < 500 Da, and an octanol-water partition coefficient log P < 5. However, there are many exceptions to Lipinski's rule.
Numerous scalar descriptors can be used to describe a molecule, including feature counts (e.g., the number of rotatable bonds), topological indices (e.g., the polar surface area), and physicochemical properties (e.g., solubility in aqueous solution). For this study, 23 parameters were collected for each fragment using the ChemAxon's calculator. This command line program was executed using the python script *calcChemParam_Library.py* (cf. Appendix A4). Several of these scalar descriptors correlates as described in section **3.4.3** (e.g., MW and vdW volume). All parameter used in this study are summarized below:

Feature counts are the number of:

- Hydrogen bond acceptors
- Hydrogen bond donors
- Total charge
- Rotatable bonds

Topological and shape indices:

- The number of rings and ring systems
- Molecular weight (MW) in Da
- Molecular surface area in $Å^2$
- Van der Waals (vdW) volume in $Å^3$
- Water accessible surface area (WASA) in $Å^2$
  using the radius of the solvent (1.4 Å in the case of water) [177,178]
- solvent accessible surface area of all atoms with
  positive partial charge (ASA+) or negative partial charge (ASA-)
- solvent accessible surface area of all hydrophobic atoms ($|qi|<0.125$)
  or all polar atoms ($|qi|>0.125$) ($|qi|$ is the absolute value of the partial charge of the atom)

Physico-chemical properties:

- Octanol-water partition coefficient log P as a measure of molecular hydrophobicity[179,180]
- Distribution coefficient log D at pH 7.0 for molecules with ionizable groups
- Intrinsic solubility (log S_0) and pH-dependent solubility log S (at pH 7.0)
- Molar refractivity (MR), a measure of total polarizability describing the propensity of a molecule to be involved in van der Waals interactions based on London dispersive forces[179]

Parameters of the Maybridge library CR2006 and TS2017 that dictate Lipinski's rule-of-five (Ro5) are listed in **Table 2.2** and **Table 2.3**, respectively. Almost all fragments comply with this rule of thumb.

Table 2.2 | Lipinski's Ro5 parameter of the Maybridge library CR2006

Parameter	Minimum	Maximum	Mean	Standard deviation
MW (Da)	94.1	291.4	185.0	26.5
Log P	-5.7	2.6	-0.4	1.8
TPSA ($Å^2$)[a]	12.0	102.4	44.9	13.7
H-bond acceptors	0	5	2.3	1.0
H-bond donors	0	2	0.8	0.6
Rotatable bonds	0	4	1.8	0.9

[a]Total polar surface area included in "Rule of Three"[181]

Table 2.3 | Lipinski's Ro5 parameter of the Maybridge library TS2017

Parameter	Minimum	Maximum	Mean	Standard deviation
MW (Da)	118.1	295.8	189.9	30.7
Log P	-3.5	4.3	0.8	1.8
TPSA ($Å^2$) [a]	9.2	97.4	45.7	17.9
H-bond acceptors	0	6	2.1	0.9
H-bond donors	0	3	0.7	0.7
Rotatable bonds	0	6	1.8	1.2

[a]Total polar surface area included in "Rule of Three"[181]

2.3.5 Descriptive Statistics of Fragments based on Molecular Descriptors

In order to get insights into hallmarks of binding fragments, several approaches have been applied. First, SxG matrices served as systematic representations of fragment libraries utilized in this study. Secondly, several scalar descriptors had been collected for each fragment as described in section **2.3.4**. Principal component analysis (PCA) was utilized to extract the most informative parameters that may define, for example, binders and non-binders. In a third step, the distributions of selected parameters were analyzed for STD NMR hits using boxplots and histograms.

SxG Matrices

The SxG distribution matrices of the fragment libraries CR2006 and TS2017 are introduced in section **2.3.3**. Fragments were manually allocated into the scaffold cluster S1 – S10 and group cluster G1 – G10. A detailed description is given in **Figure 2.6** and **Figure 2.8**. In addition, SxG performance matrices were utilized to picture structural properties of positive hits from STD NMR screening. The percentage of STD NMR hits per subcluster (e.g., S1/G1) was normalized to yield the so-called performance factor *PF*, which is defined for each subcluster as follows:

$$PF = \frac{\frac{number\ of\ hits\ per\ subcluster}{number\ of\ all\ hits}}{\frac{number\ of\ all\ fragments\ per\ subcluster}{number\ of\ all\ fragments}} \qquad \text{(II)}$$

A small value below one indicates that only a small proportion of this subcluster interacts with the target screened. It has to be noted that the performance factor is biased for subclusters with a small number of fragments.

Principal Component Analysis

The ***n*-*by*-*p*** data matrix ***M*** with ***n*** fragments and ***p*** =16 descriptors was subjected to PCA using Matlab, which by default uses the singular value decomposition (SVD) algorithm. The Matlab script can be found in Appendix A4. Since the parameters used are not on the same scale and have not the same units, the data were normalized using z-scoring, resulting in a standard normal distribution centered at zero for each parameter:

$$z = \frac{x - \mu}{\sigma} \qquad \text{(III)}$$

where μ is the mean and σ is the standard deviation for the respective parameter set. The Matlab function *[coeff, score, latent, tsquared, explained, mu] = pca(normalized **M**)*, which is implemented in Matlab, returns the principal component coefficients (also referred to as loadings) as ***p***-by-***p*** matrix, the principal component scores as ***n*-*by*-*p*** matrix, the ordered principal component variance (eigenvalues of the covariance matrix of *normalized **M*** stored in *latent*) for each parameter, the Hotelling's T-squared statistic for each fragment (not used), the percentage of the total variance

explained by each principal component stored in *explained*, and the estimated mean μ of each parameter (not used as μ equals zero).
Giving a detailed explanation of SVD is beyond the scope of this thesis. In principle, this statistical method includes three mutually compatible steps: (1) It identifies dimensions (or components) along which data points exhibit the most variation, (2) it transforms correlated variables into a set of linearly uncorrelated ones based on eigenvectors and eigenvalues, and (3) it simultaneously reduces the number of dimensions. In a nutshell, PCA exposes the most meaningful information from a given data set.
To visualize the reduced data set, the principal component scores (referred to as transformed variables) of the first three principal components were plotted using a 3D scatter plot. Data points were colored depending on the category (e.g., binders vs. non-binders). In addition, a biplot was used to find parameters that may explain the underlying dispersion of respective categories. The biplot represents the whole data set as points (principal component scores) and parameters as vectors (principal component coefficients). The direction and length of the vectors indicate how each variable contributes to the three principal components.

Boxplots and Histograms

In descriptive statistics, boxplots and histograms are used to describe the distribution of continuous data (e.g., MW) and discrete data (e.g., the number of rings), respectively, in terms of central tendency and dispersion. A boxplot visualizes several quartiles of a ranked set of data values. The bottom and top of the box are the first quartile (Q1, 25th percentile) and the third quartile (Q3, 75th percentile) enclosing 50% of all data values (the interquartile range, IQR). A line inside the box represents the median. The box is extended using lines referred to as whiskers which ends are defined as Q1-1.5·IQR and Q3 +1.5·IQR, respectively. Outliers are depicted as small circles. In order to assess the inner structure of continuous data, histograms can also be applied by so-called data binning. Therefore, values in a given small interval are combined and counted.
Several python scripts had been written for this purpose (cf. Appendix A4). For example, the python script *searchParameter.py* gets a list of positive hits for different targets stored in an EXCEL-file, collects all values of hits for a given parameter (e.g., charge) from the fragment database, and the script *Boxplot_and_Histogram_STD_Hits.py* returns a histogram or boxplot, depending on the operator's choice. A paired t-test and a chi-square test of homogeneity have been implemented for continuous data and discrete data, respectively.
Moreover, 3D histograms were utilized to represent fragment classes as a combination of three informative parameters, including van der Waals volume, total charge, and the number of hydrogen bond acceptors or donors.

2.3.6 Prediction of Fragment Mixtures

The advantage of STD NMR screening is the capability to screen up to 30 low molecular weight compounds in parallel. A python script *poolCompounds.py* was written by Dr. Christoph Rademacher in 2008 to provide fragment mixures[90] based on following rules:

(1) A unique signal per fragment is available for hit identification.
(2) Exclude chemical reactions between fragments that are present in one mixture.
(3) Fragments of one mixture exhibit the same number of rotatable bonds.

The most important requirement for a screening employing fragment mixtures is the unambiguous identification of fragments within on mixture based on ^{1}H NMR signals. Therefore, ^{1}H reference spectra of soluble fragments were acquired as described in the next section. Next, chemical shifts were extracted using the Bruker software *TopSpin 3.1*. The relative minimum intensity was set to 1 by executing the command "*1 MI 1"* upon all spectra using the command "*multicmd*". Then, peak picking (command: *pp*) was conducted with following parameters: Left picking limit of 12 ppm, right picking limit of -2 ppm, relative intensity of reference peak of 15, minimal intensity of 1, maximal intensity of 30, and detection sensitivity of 1. The peak list of chemical shifts was stored as XML-file (relative path: ./pdata/1). Finally, all peak lists were merged into a single EXCEL-file employing the python script *grepPeaksfrom_XML_to_EXCEL.py.* Signals between 4.3 ppm and 5.3 ppm were excluded due to the water signal and water suppression.

To meet rule 2, fragments were categorized into acids, bases, neutrals, nucleophiles, and electrophiles by using SMILES code filtering rules according to Hann et al.[182] (cf. Appendix A4). Only fragments with compatible properties were mixed, for example, nucleophilic bases and nucleophilic neutral fragments (for more details see REF[90]).

In this thesis, the third requirement was replaced by the rule that one mixture contains only structurally diverse fragments. To fulfill this condition, the TAN was used to compare chemical structures as described above. If the first two conditions were met, and the TAN was lower than 0.55, compounds were merged into the same mixture yielding 15 fragments per mixture on average. The modified python script *poolCompounds_2014.py* including the method *getTanimoto* was written using *Python 2.7* (see Appendix A4). Note that this script was rewritten in *Python 3.6* and some features were added as described in section **3.3.3** (fragment screening against NoV P dimers).

2.3.7 Nuclear Magnetic Resonance Spectroscopy

Unless otherwise stated, all NMR experiments were performed at 25 °C in 25 mM sodium phosphate buffer (D_2O, pH* 7.05) in 3 mm NMR tubes (final volume 160 – 180 µl), using a Bruker Avance III 500 MHz spectrometer equipped with a TCI CryoProbe™. For data processing, a Bruker TopSpin 3.1 software was used.

Proton Reference Spectra of Fragments

Proton (^1H) spectra of soluble fragments were acquired at a fragment concentration of 1 mM in 5 mm NMR tubes in the presence of 50 µM trimethylsilyl-2,2',3,3'-tetradeuteropropionic acid (TSP-d_4) as reference substance. In case of fragments from the Maybridge library CR2006, ^1H spectra were recorded without water suppression using the pulse program *zg30* with standard settings. For TS2017 fragments, excitation sculpting implemented in the pulse program *zgesgp* was used to suppress the residual water resonance signal. All ^1H spectra were referenced to TSP-d_4 (δ = 0.00 ppm). Moreover, an exponential window function, a line broadening of 2 Hz (command: *lb 2; efp*) and an automated polynomial baseline correction (command: *absn*) was applied. The following setting was applied: 8 scans, 2.3 s acquisition time, 32 k data points, 2 s relaxation delay, and 14 ppm spectral width.

STD NMR Screening

In order to identify ligands for NS2B-$NS3^{pro}$, only the Maybridge library CR2006 has been screened as part of my master's thesis[133] (Schöne, 2014), because TS2017 fragments had not been purchased at this time. The following steps were performed:

(1) Proton NMR spectra were collected with 64 scans for each fragment mixture to confirm chemical integrity.
(2) STD NMR spectra were acquired in the absence and presence of 5 µM WNV protease using the standard pseudo-2D pulse sequence, stddiffesgp.3 (cf. Appendix A5 and REF[87]).
(3) Potential WNV protease binders were identified visually using ^1H spectra of individual fragments as a reference.

The following parameters were used to acquire STD NMR spectra. FIDs were acquired with 24.5 k data points, 1024 scans, and an acquisition time of 1.7 s. A spinlock filter of 10 ms was applied to remove protein background. A train of selective Gaussian-shaped pulses (Gaus1.1000) of 50 ms duration was used to saturate protein signals. The relaxation delay was set at 4 s and a saturation time of 2 s at a power level of 45 dB (0.3 µW) was applied. This corresponds to a 529° flip angle and $\gamma B1(max)/2\pi$ = 71 Hz. The on- and off-resonance frequencies were set at -1 ppm and 80 ppm, respectively. For fragments showing direct irradiation (observed in ~10% of all fragments) in the absence of protease, intensities resulting from direct irradiation effects were subtracted from intensities of STD effects in the presence of protease. When the difference was zero or negative, these fragments were excluded from further analysis.

STD NMR Competition Assays

The design of STD NMR competition assays is described elsewhere (see REF[133,168]). Briefly, positive hits from STD NMR screening were clustered into eight groups based on pairwise TAN. At least two of the most potent fragments per cluster were subjected to STD-NMR competition experiments using the known active site fragment **A**[144] as reporter ligand. Out of 36 fragments, 14 competitively binding fragments were identified, and 22 fragments not competing with **A**.
Interestingly, 10 of these 22 fragments showed a notable increase of STD intensities in the presence of compound **A** suggesting cooperative binding (cf. REF[168]). In this thesis, the findings were confirmed for particular fragments. Therefore, competition experiments were repeated with **CR112** and **CR081** pooled with one or two other positive hits using samples containing 10 µM of protease, 200 µM of each fragment, and 800 µM of fragment **A**. A sample of each fragment mixture in the presence of **A** with no protease present served as negative control.
In order to further minimize artefacts, the saturation level was changed from 43 dB to 48 dB (0.15 µW), corresponding to a flip angle of 374° and γB1(max)/2π of 50 Hz. ^{1}H NMR spectra of each fragment mixture ensured that at least two isolated resonance lines of each component are detectable. STD NMR spectra were recorded with the following settings: 8.8 µs hard pulse, 24.5 k data points, 512 transients, spinlock filter set at 10 ms, 4.5 s relaxation delay, 4 s saturation time, 1.7 s acquisition time, -1.4 ppm on-resonance and 80 ppm off-resonance frequency. Under these conditions, artefacts from direct irradiation were practically absent.

STD NMR Titration

In order to confirm that **CR112** and **CR081** interact with binding sites different to **A**, STD amplification factors[87] were calculated as a function of increasing amount of **A** (30 – 600 µM). Therefore, samples containing 3 mM or 4 mM of fragments **CR112** or **CR081**, respectively, in the presence of 5 µM WNV protease were prepared. Fragment **A** was then titrated into both samples and STD-NMR spectra were recorded (see above, latter settings). STD amplification factors[87] were calculated using the following formula:

$$STD_{amp} = ligand\ excess \cdot \frac{I_0 - I_{sat}}{I_0} \qquad \text{(IV)}$$

with I_0 being the intensity of a proton of **A** in a spectrum without saturation and I_{sat} being the intensity of the same proton of **A** in a spectrum with saturation. Direct irradiation of proton resonance signals of **A** was observed at concentrations above 100 µM, likely due to self-aggregation of **A** leading to considerably larger line widths. The contribution from direct irradiation was approximately 20%.

Consequently, intensities I_{Sat} were corrected by subtracting corresponding intensities from direct irradiation in the absence of protease. Titration curves were fitted to a Langmuir binding isotherm:

$$STD_{amp} = STD_{amp,max} \cdot \frac{[L]}{K_D+[L]} \quad \text{(V)}$$

with *[L]* being the ligand concentration and K_D being the dissociation constant. Curve fitting was performed using Origin 7.0.

2D methyl-TROSY NMR Experiments

Since protein binding sites almost always comprise hydrophobic amino acid residues, side-chain methyl groups are perfectly suited as reporters for ligand binding events[159]. In this study, isoleucine, leucine, and valine side chains (ILV) of a uniformly perdeuterated ({U-^{2}H}) WNV protease NS2B-NS3^{pro} were equipped with ^{13}C-methyl-labels as described above (see **2.3.1**). After protein purification, the SEC buffer was exchanged several times against NMR buffer (see above) at 4 °C, using a 10 kDa cut-off filter device (Millipore). In addition, stock solutions of fragment **A** (16 mM) and a saturated solution of **CR112** were prepared in NMR buffer and the pH* was adjusted accordingly (see abbreviations for the definition of pH*). Each NMR sample contained 50 µM or 100 µM protein and 50 µM TSP-d_4 as an internal reference standard. The quality of isotopic labeling was judged by standard 1D NMR spectra. To observe chemical shift perturbations upon ligand binding, methyl-TROSY spectra of {U-[^{2}H,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Valγ-[$^{13}CH_3$]}-labeled protease in the absence and presence of saturating amount of **A** were acquired using the pulse program *metr-hmqcph_new.tb,* which has been established in a former study (see REF[183] and Appendix A5). Next, **CR112** was titrated into samples containing protease and **A**. Fragment **CR112** is less soluble in water, hence, undissolved remains of a nominally 20 mM sample were removed via centrifugation (17 000 x *g*, 5 min). The concentration of **CR112** in 2D methyl-TROSY experiments was estimated as 10 mM using TSP-d_4 as a reference.
The following setting was applied in all methyl-TROSY experiments: 16-48 scans, relaxation delay of 1.5 s, sweep width of 6 ppm x 24 ppm, O1P of 0.5 ppm x 17.5 ppm, and 792 x 150 data points for the respective dimension. The digital FID resolution was 7.6 Hz/point and 40.2 Hz/point for the ^{1}H and for the ^{13}C dimension, respectively.

^{1}H,^{15}N-TROSY-HSQC NMR Experiments

Stock solutions of fragment **A** (16 mM) and saturated solutions of **CR112** and **CR081** were prepared in phosphate buffer (25 mM sodium phosphate buffer in H_2O, pH 7.0) as described above. Since chemical shift perturbation experiments often suffer from pH artefacts, the pH was adjusted exactly to pH 7.0 for each sample. In order to reveal the binding modes of **CR112** and **CR081** as well as for structurally similar fragments, ^{1}H,^{15}N-TROSY-HSQC NMR experiments were performed in 25 mM sodium phosphate buffer in 90% H_2O/ 10% D_2O (pH 7.0).

First, 2D spectra were recorded of a sample containing uniformly [^{2}H,^{15}N]-labeled NS2B-NS3^{pro} (~130 µM) in the absence and in the presence of 1.6 mM fragment **A** employing the pulse program *trosyetf3gpsi*[184,185] as implemented in TopSpin 3.1. The parameters were set as follows: 48 scans, relaxation delay of 1 s, sweep width of 16 ppm x 35 ppm, O1P of 4.7 ppm x 120 ppm, and 2048 x 256 data points for the respective dimension. The digital resolution was 7.3 Hz/point and 15.8 Hz/point for the ^{1}H and for the ^{15}N dimension, respectively.
In a next step, ^{1}H,^{15}N-TROSY-HSQC spectra of NS2B-NS3^{pro} were obtained in the presence of circa 10 mM of **CR112**, and in the presence of both **CR112** (~10 mM) and fragment **A** (1.6 mM). For a structure-binding relationship analysis, this procedure was repeated with other fragments. The concentration of fragments ranged from 4 to 6 mM (**CR116** and **CR402**) and from 8 to 12 mM (**CR053**, **CR069**, **CR081**, **CR082**, **CR113**, and **CR114**) depending on the solubility of fragments. The concentration of fragments was determined by integration of ^{1}H NMR signals using TSP-d_4 as a reference. Chemical shift perturbations Δδ of backbone amide protons of the protease upon addition of fragments were given as Euclidian distances:

$$\Delta\delta = \sqrt{{}^{1}/_{2}\ (\Delta\delta(\,{}^{1}H)^2 + 0.2 \cdot \Delta\delta({}^{15}N)^2)} \qquad \text{(VI)}$$

with $\Delta\delta\ (^{1}H)$ and $\Delta\delta\ (^{15}N)$ being the chemical shift perturbations in the respective dimension. The scaling factor of 0.2 is arbitrary and accounts for the larger chemical shift dispersion in the ^{15}N dimension. Data were analyzed using the software CcpNmr Analysis 2.4.2.

2.3.8 Ligand Docking

Rigid receptor docking was performed using Glide[186] as implemented in Maestro 11 (2017; Schrödinger Inc.). Crystal structure data of WNV NS2B-NS3^{pro} in complex with the inhibitor 3,4-dichlorophenylacetyl-Lys-Lys-GCMA (PDB: 2YOL[128]) were chosen for docking experiments. The protein preparation wizard from Maestro was used to process the crystal structure data. This included several steps: Bond orders were assigned, missing hydrogen bonds were added, and water molecules beyond 5 Å were deleted. Moreover, the protonation states of the amino acid residues Asp, Glu, and His were altered in order to optimize the hydrogen bond network using the PROPKA program[187,188]. The latter is part of Maestro, which predicts pK_a values of protein residues. The pH was set to 7.0. Finally, initial energy minimization employing the OPLS_2005 force field was performed. Ligands were built with Maestro and energy-minimized using the LigPrep tool with default program settings. The quadratic receptor grid was centered with respect to corresponding amino acid residues in the binding pocket. The length of the receptor grid was set to 10 Å each edge. For ligand docking, no other constraints were used. The Glide extra precision scoring function[103], which represents the relative free binding energies (kcal/mol^{-1}), was used to rank docking poses after post-docking minimization (with full ligand flexibility).

2.4 Results

In the search for novel structural motifs with binding affinity for the WNV protease NS2B-NS3^{pro}, the Maybridge library CR2006 was screened by the use of STD NMR spectroscopy as initial read-out. Fluorogenic-based protease activity assays were performed to rank fragments according to their ability to inhibit the enzymatic cleavage of a peptidic substrate. The best fragments were subjected to STD NMR-based competition experiments to find compounds that interact with secondary binding sites (see **Figure 2.12**). Subsequently, the binding modes of two compounds (**CR112** and **CR081**) and analogs that share a common structural motif were elucidated using a competition protease assay format as well as ligand-observed and protein-observed NMR experiments. In addition, structural models were built by employing *in silico* docking to guide the hit selection as well as the hit-to-lead optimization.

2.4.1 Fragment Screening based on STD NMR Experiments

Ligand-observed STD NMR spectroscopy allows for the detection of low-molecular-weight compounds (<300 Da) with binding affinities in the micromolar to the millimolar range without deconvolution of fragment mixtures. In order to identify binders for the WNV protease NS2-NS3^{pro}, mixtures containing 14 fragments on average had been screened using STD NMR, yielding 126 positive hits with STD effects ranging from 0.3% to 9.4%. In ~10% of all ^{1}H signals, direct irradiation was observed in the absence of protein. However, the magnitude of these artefacts was significantly lower as compared to the STD effects derived from the saturation transfer via cross-relaxation with protons of the protein.

In an attempt to find important scaffolds and chemical groups promoting binding for NS2B-NS3^{pro}, a SxG performance matrix has been created (see **Figure 2.10**). The values in the matrix represent so-called performance factors for each subcluster. A value close to one means that the percentage of this subcluster in relation to all positive hits is not significantly different from the percentage of fragments of this subcluster compared to all fragments present in the Maybridge library CR2006. In other words, there is no clear preference or rejection for this type of binding motif. It is important to mention that the performance factor is biased towards higher values for subclusters that contain only a few fragments. The scaffold clusters S1, S3, and S6 account for ~60% of fragments. The majority of compounds of cluster S1 (15.2%) can be characterized by a phenyl-substituent that is attached either to a flexible morpholine, piperazine, or piperidine ring via two single bonds. Scaffold cluster S3 (21.1%) comprises predominantly phenyl rings linked to a 5-membered heterocycle. Cluster S6 (22.8%) contains functionalized ortho-fused heterocyclic compounds, where a five-ring is attached to a six-ring (e.g. derivatives of benzofuran). No clear preference or rejection can be observed for these scaffold clusters with respect to their molecular shape. However, fragments of these scaffold clusters equipped with alkylated carboxyl, hydroxyl, or amino groups are more frequently classified as hits. Fragments from cluster S2 and S4 are classified as frequent hitters as ~50% to ~60 % of compounds, which share these motifs, interact with NS2B-NS3^{pro}, albeit these scaffolds are underrepresented (~10%) compared to other members of the fragment library CR2006

(cf. **Figure 2.6**). Cluster S2 contains biphenyl derivatives as well as analogs of fragments from cluster S1 with only one single bond between two rings. Fragments with two five-membered heterocycles linked by a single covalent bond are allocated in cluster S4. In the case of cluster S2, especially hydroxyl groups, amino groups, and halogens seem to promote binding. However, carboxyl groups and secondary or tertiary amino groups are also accepted, whilst carboxyl and ester groups are preferred over hydroxyl and amino groups for fragments from cluster S4.

For very small fragments (cluster S8 and S9), only compounds with ester, methoxy, and amino groups have been identified as positive hits during the STD NMR-based screening. The other chemical groups seem to hamper binding to the WNV protease.

From the perspective of chemical groups, about two third of positive hits have either a carboxyl group (~20%), a hydroxyl group (~16%), or a primary amino group (~29%), reflecting approximately the same distribution as members of the Maybridge library CR2006. The percentage of ester, methoxy, and alkylated amino groups is increased for positive hits, which indicates a preference for these functionalities being identified as positive hits. In contrast, halogen-containing compounds and fragments with uncommon chemical groups (e.g. cyano groups) often have been classified as non-binders based on the absence of STD effects.

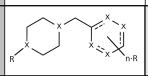
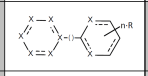
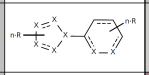
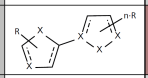
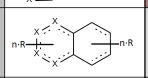
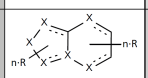
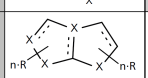
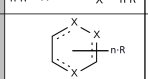
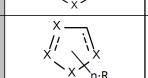
		G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	
	R = $(CH_2)_m$ +	COOH	COO-CH_3 or COO-CH_2CH_3	CO-CH_3 (R''-CO-R'')	OH	O-CH_3	NH_2	$NHCH_3$ or $N(CH_3)_2$	Halogens	Other groups	Only CH_3 or without any group	SP
S1		0.5	2.0	0.0	0.5	3.3	0.6	1.4	0.0	0.0	0.0	0.7
S2		1.1	0.0	0.0	2.4		2.2	0.7	3.3	0.0		1.6
S3		0.9	3.3		1.0		1.1	1.2	0.0	0.0	1.1	1.0
S4		2.2	3.3		1.3		1.3					1.9
S5		0.7	3.3	1.6	0.8		1.2		0.0	1.6		1.2
S6		0.9	1.8	1.2	0.5	3.3	0.9	1.8	0.0	0.0		1.0
S7			0.8		3.3			3.3				1.3
S8			0.8	0.0	0.0	1.6	0.7		0.0	0.0		0.6
S9		0.0	1.5		0.0		0.9	0.0				1.0
S10	Complex structures	0.5	3.3	0.0	0.0		1.1		1.1	0.0	1.1	0.5
	GP	0.9	1.7	0.6	0.8	2.7	1.1	1.4	0.3	0.3	0.8	

Figure 2.10 | SxG performance matrix of WNV protease binders identified by STD NMR
A detailed description of scaffolds and groups are given in **Figure 2.6**. Each number reflects the performance factor of a specific subcluster as this is described elsewhere (section **2.3.5**). In this case, a value of 3.3 indicates that all fragments from this subcluster had been identified as binders. (SP, scaffold performance; GP, group performance; *X* = C, N, O, or S atoms; R = chemical groups, n = 0, 1, 2, or 3 and m = 0, 1, or 2).

The inspection of scaffolds and chemical groups provides a fuzzy pharmacophore picture of positive hits from the STD NMR screening, albeit some tendencies are notable. In order to find more general properties (e.g. total charge) that may promote or hamper binding to NS2B-NS3^{pro}, a principal component analysis (PCA) of 16 scalar molecular descriptors was conducted (see section **2.3.5** and **Figure 2.11**). None of the parameters that have been analyzed seem to determine alone whether or not a fragment interacts with the WNV protease. Therefore, so-called 3D histograms were used to illustrate the combination of three parameters, revealing that about one-quarter of positives hits are natural compounds with two hydrogen bond acceptors and a van der Waals volume between 160 – 180 Å^3.

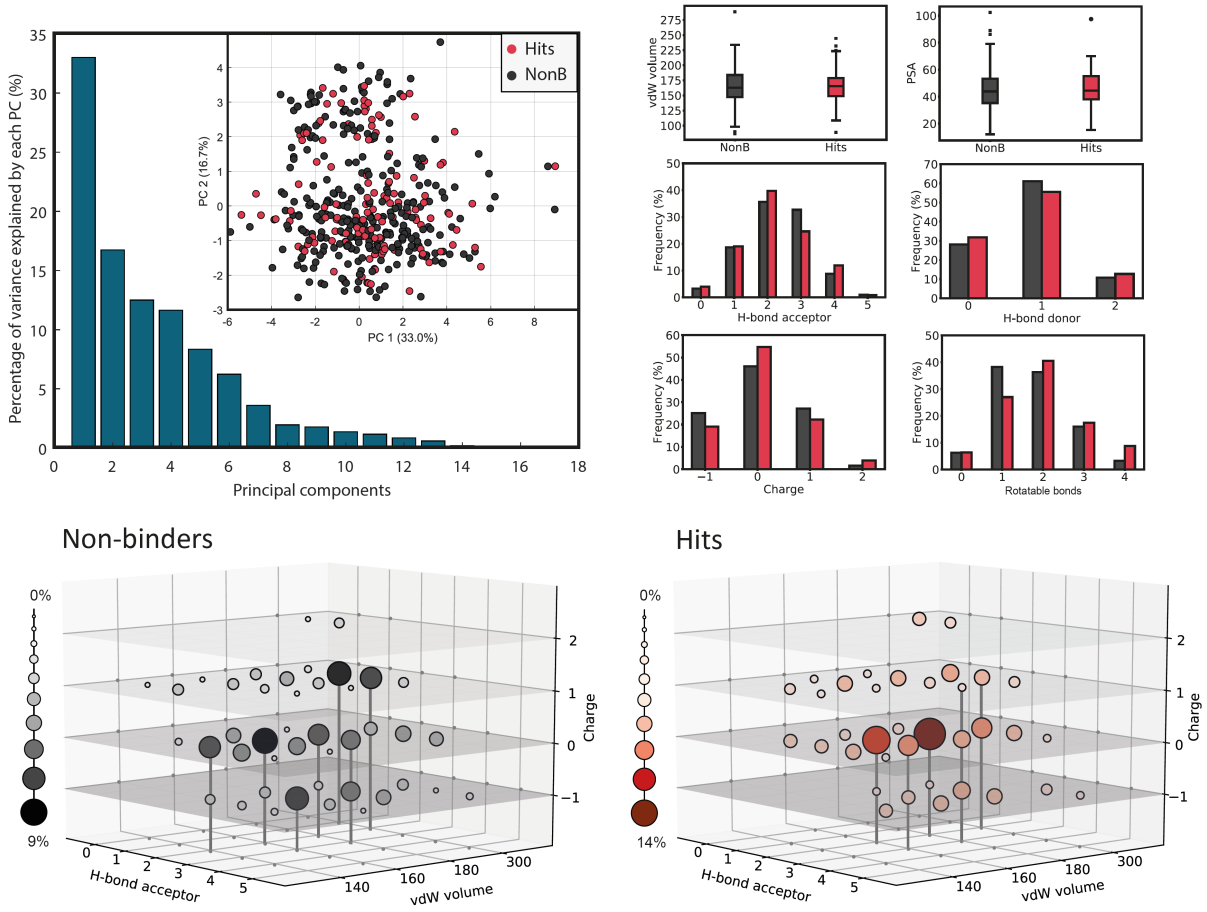


Figure 2.11 | Characterization of positive hits derived from the WNV protease screening based on STD NMR
In total, 16 chemical parameters had been used for the principal component analysis, including molecular weight, molecular surface area, van der Waals (vdW) volume, molecular refractivity, water accessible surface area (WASA), the ratio of hydrophobic and polar WASA, polar surface area (PSA in Å^2), the number of ring system, the number of rings, the number of rotatable bonds, total charge, the number of hydrogen-bond acceptors (H-A) or donors (H-D), the partition coefficient log P, and the distribution coefficient log D. The first two components explain ~50% of variance of the given data set. No clear differences of binders (hits) and non-binders (NonB) are notable in terms of their distribution along the principal components (PC). On the right side, boxplots and histograms are depicted for representative scalar descriptors. The p-values of statistical tests used (cf. section **2.3.5**) are always above 0.3, indicating no significant differences between these distributions. In the lower panel, a combination of three informative descriptors is shown as 3D histograms for non-binders and binders. These plots describe the distribution of fragments characterized by the number of hydrogen bond (H-bond) acceptors, van der Waals (vdW) volume, and total charge. Fragments were equally allocated into four classes of vdW volumes: <140, < 160, <180, and <300 Å^3.

In order to validate positive hits from the STD NMR screening, fluorogenic protease activity assays were performed, yielding 83 compounds with inhibitory properties. None of the molecular properties regarding the chemical scaffolds, functional groups or scalar descriptors could be classified as a single driving force for protease inhibition. In a next step, the competition between the active site binder termed as fragment **A**, and the best 36 hits according to their inhibitory potency were analyzed using STD NMR as read-out. This was achieved by adding of **A** to a mixture of NS2B-NS3^{pro} with respective hits and observing the changes in STD effects. A detailed description can be found in REF[133] and REF[168]. Briefly, 14 out of 36 fragments showed a notable decrease of the saturation transfer (~25% or more) in the presence of **A**, indicating binding to the active site. Of the remaining 22 fragments, 12 were unaffected and 10 fragments showed an increased STD effect in the presence of **A**. The latter group comprises fragment **CR081** and **CR112,** which are the most potent inhibitors compared to the remainder of all fragments, with an IC_{50} of ~500 µM. Exemplary STD NMR spectra of both fragments in the absence and presence of **A** are depicted in **Figure 2.12C**, showing clearly that fragment **A** bound to the active site of NS2B-NS3^{pro} promotes binding of **CR081** and **CR112**.

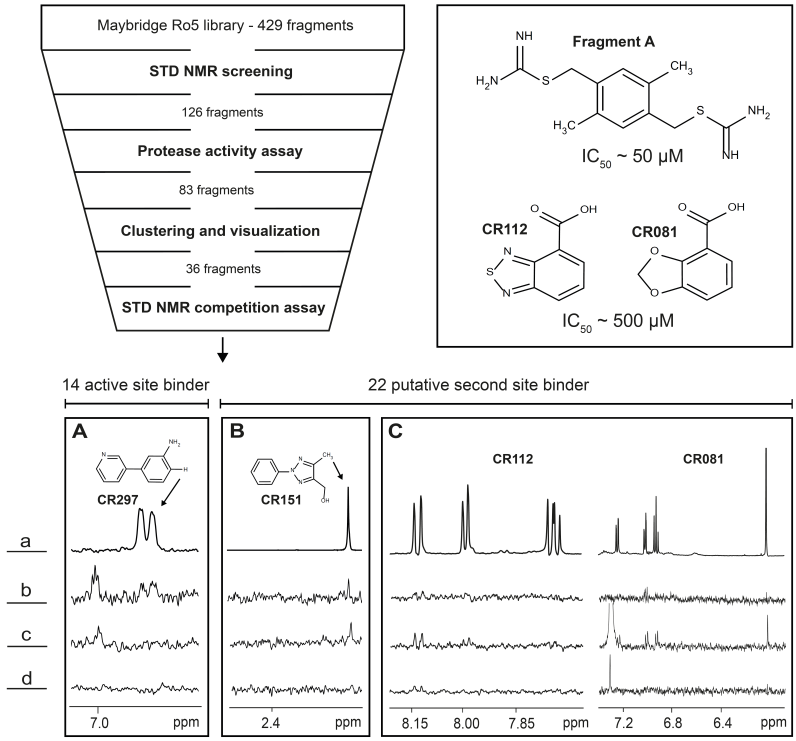


Figure 2.12 | Fragment screening and STD NMR competition assay
The Maybridge Ro5 library CR2006 was screened for fragments, which bind to WNV protease NS2B-NS3^{pro}, by employing saturation transfer difference NMR spectroscopy (STD NMR). In a second step, biological active fragments have been identified using protease activity assays. Finally, STD NMR competition assays were used to dissect competitive binders from second-site binders. Fragment **A** served as an active site binder in STD NMR competition assays. Fragments **CR112** and **CR081** are the most potent secondary site binders. By the addition of **A** in a 20-fold excess over fragment concentration, the STD effect of fragments either decreased more than 25% (A), remained constant (B) or increased more than 25% (C)**.** Bottom: The spectra were recorded with 512 scans (A-B) or 1024 scans (C) at a ^{1}H frequency of 500 MHz. Representative regions of reference spectra (a) and STD difference spectra (b-d) are shown for certain proton resonances of **CR297**, **CR151**, **CR112** or **CR081**. (b) STD difference spectra of samples containing protease and corresponding fragments. (c) Same as (b) supplemented with fragment **A**. (d) Same as (c) without protease (negative control). The signals at 7.0 ppm (A, b and c) or 7.3 ppm (C, c and d) belongs to fragment **CR262** and fragment **A**, respectively. This figure was adapted from REF[168].

2.4.2 Validation of the Binding Mode of CR112 and Analogs

STD NMR Titration and Fluorogenic Competition Assay

For the validation of STD NMR competition assays, the dissociation constant K_D of **A** was determined by STD titration experiments in the absence and presence of **CR081** or **CR112**. These experiments substantiated that **CR081** and **CR112** bind to secondary binding sites, with almost no effect on the interaction of **A** with the WNV protease (**Figure 2.13**, left). To further elucidate the binding mechanism of **CR112** was subjected to a fluorogenic competition assay. It has long been practice to determine the binding mode by the method of Dixon[189], in which the reciprocal velocity of an enzyme-catalyzed reaction is plotted against the inhibitor concentration at increasing amount of a substrate (**Figure 2.13**). In the case of **CR112**, the extrapolated lines at different substrate concentrations intersect at a single point on the x-axis, strongly suggesting non-competitive inhibition, with an inhibition constant K_i of ~500 µM, where K_i represents the dissociation constant of the enzyme-inhibitor (EI) complex. Of note, it is characteristic for non-competitive inhibitors that IC_{50} equals K_i (cf. REF[190]).

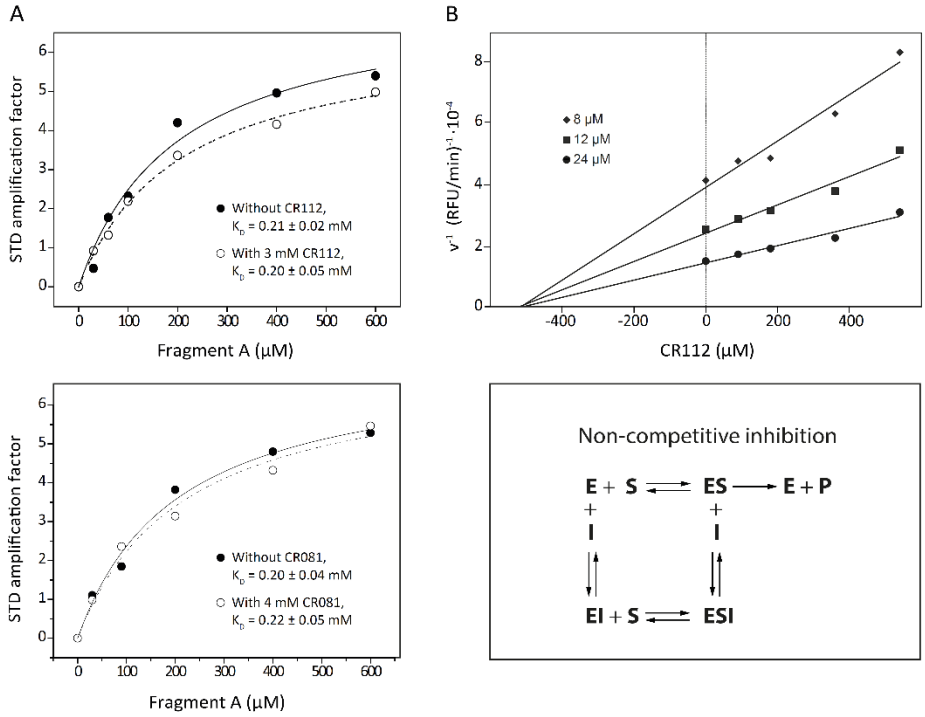


Figure 2.13 | Fragment CR112 is a non-competitive inhibitor of the WNV-protease NS2B-NS3^{pro}
(A) STD NMR titration of fragment **A** in the presence of 1 µM WNV protease. The dissociation constant of **A** is not significantly altered upon addition of 3 mM **CR112** or 4 mM **CR081**. (B) Dixon plot of 1/v versus increasing concentration of **CR112** at different substrate concentration (8, 12 and 24 µM) represents a non-competitive inhibition mode, with an inhibition constant K_i of ~ 500 µM. The scheme of the presumed inhibition mode is depicted below. This figure was adapted from REF[168].

Identification of Binding Pockets using Chemical Shift Mapping

The strategy of chemical shift mapping by using heteronuclear NMR spectroscopy has been widely applied to identify binding sites of a protein of interest. First, {U-[^{2}H,^{15}N]; Ileδ1-[^{13}CH$_3$]; Leu,Valγ-[^{13}CH$_3$]}-labeled WNV protease expressed in M9+/D$_2$O minimal medium[93] was subjected to chemical shift perturbation experiments, thereby exploiting the high resolution and sensitivity of ^{13}C-labeled methyl groups. The acquisition of one methyl-TROSY spectrum took less than one hour. Chemical shifts changed substantially when fragment **A** was added at saturating amounts to the protein sample as described in previous studies[131,183], whilst **CR112** did not perturb the ^{1}H,^{13}C-correlation spectrum when titrated into an NMR sample containing NS2B-NS3^{pro} in complex with **A**. Noteworthy, the ^{13}C-methyl-labeling was incomplete as side chains of leucine residues were absent in the methyl-TROSY spectra (**Figure 2.14**, left), probably due to a long incubation time of ~92 h along and the use of 20% (v/v) Silantes medium containing perdeuterated amino acids. Thereby, the direct ^{13}C-labeled precursor of leucine had been depleted during protein expression and replaced by unlabeled leucine present in the Silantes medium.

To evade this circumstance, the duration of protein biosynthesis was reduced to 42 h, yielding all expected cross peaks in the methyl-TROSY NMR spectrum (**Figure 2.14**, right). In addition, two unassigned NMR peaks with low signal intensities became visible due to isotopic scrambling[191]. Only for these signals, small chemical shift changes could be observed upon binding of **CR112**.

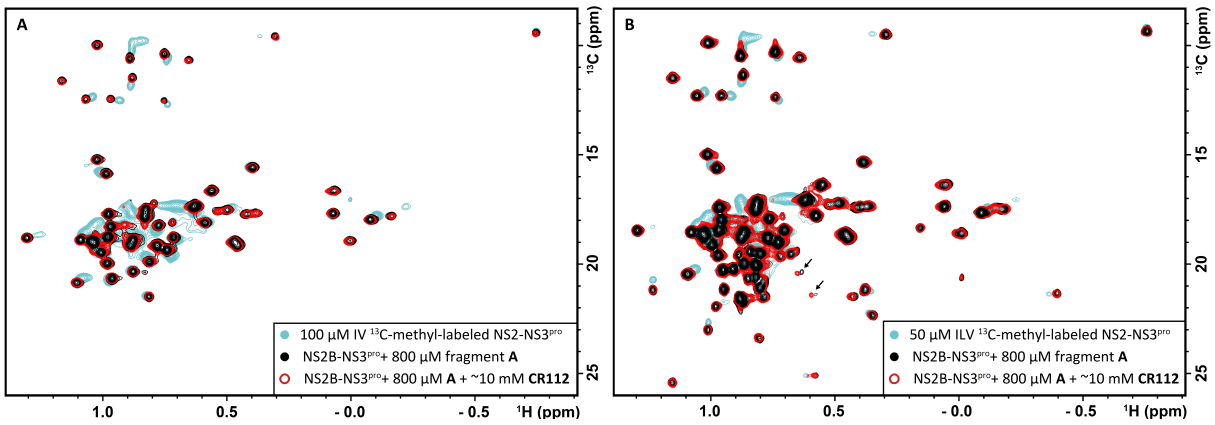


Figure 2.14 | Chemical shift perturbations of ^{13}C-methyl-labeled NS2B-NS3^{pro} upon ligand binding
Protein biosynthesis of WNV protease NS2B-NS3^{pro} was conducted in M9+/D$_2$O medium supplemented with ^{13}C-methyl-labeled isoleucine and leucine/valine precursor (for details see section **2.3.1**). (A) Surprisingly, the ^{13}C-methyl-label is missing in side chains of leucine residues when the protein expression was extended to 92 h at 18 °C. While fragment **A** greatly perturbs the methyl-TROSY spectrum, the addition of **CR112** has almost no effect on NMR signals. The chemical shift differences are below the digital resolution of 7.6 Hz for the proton dimension. (B) When expressed for 42 h at 17 °C, isotopically labeled leucine residues are detectable in the methyl-TROSY spectrum. But significant chemical shift perturbations upon **CR112** binding are still absent, except for two unassigned NMR signals (black arrows). The assignment of methyl groups can be found in the Biological Magnetic Resonance Bank (BMRB entry 11053) as well as in REF[132].

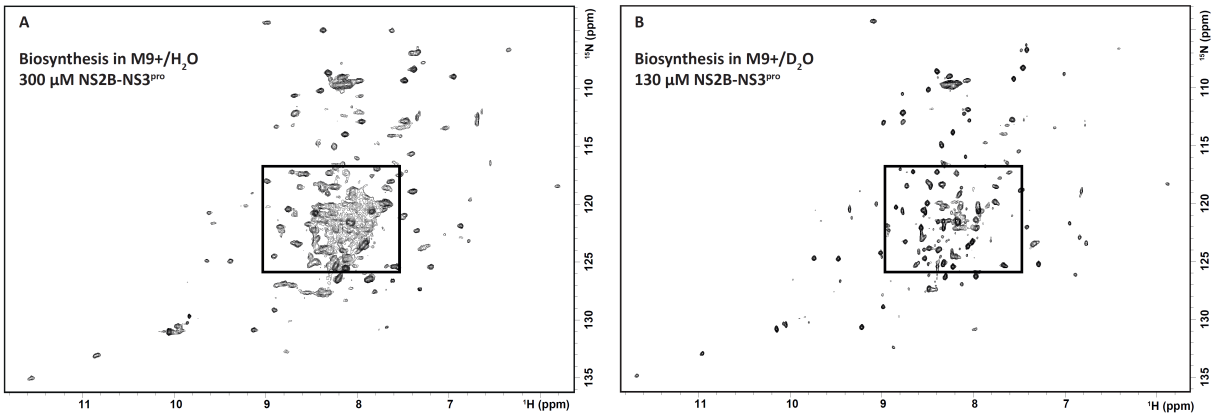


Figure 2.15 | Production of the WNP protease in M9+/H_2O medium vs. M9+/D_2O medium
Perdeuteration of uniformly ^{15}N-labeled NS2B-NS3^{pro} improves the ^{1}H,^{15}N-TROSY-HSQC spectrum (B) in terms of signal overlapping (black box). The spectra (A/B) were recorded with the same NMR settings as described in section **2.3.7**.

Although ^{13}C-methyl labeling of isoleucine, leucine, and valines gave a good distribution of sensitive probes throughout NS2B-NS3^{pro}, a few "blind spots" were still present. Therefore, another strategy using ^{15}N-labeling of the protein has been explored. First of all, protein expression of non-deuterated ^{15}N-labeled WNV protease in M9+/H_2O medium had been tested. The ^{1}H,^{15}N-TROSY-HSQC spectrum matches quite well the ^{15}N-HSQC spectrum shown in a previous study[132]. Despite the utilization of TROSY-effects, the ^{1}H,^{15}N-correlation NMR spectrum suffered from signal overlapping, rendering it difficult to detect small changes in chemical shifts as this was expected for small low-affinity binders such as **CR112**. As a consequence, protein expression was performed in D_2O-based minimal medium with perdeuterated glucose to obtain uniformly [^{2}H,^{15}N]-labeled NS2B-NS3^{pro}, providing high-resolution 2D NMR spectra (see **Figure 2.15** and following). In a first step, fragment **CR112** was titrated into an NMR sample containing isotopically labeled NS2B-NS3^{pro}, leading to notable chemical shift changes. However, it turned out that these changes were caused by a small variation of the pH value by less than 0.2 units (**Figure 2.16**). From this point, the pH of stock solutions of fragments was adjusted as accurately as possible to keep pH values of NMR samples constant.

Under strict control of sample conditions, a putative second binding site for **CR112** could be identified. A backbone assignment of U-[^{2}H,^{15}N]-labeled WNV protease in complex with **A** is available (BMRB Entry 11053; REF[131]) and served as a basis for the identification of binding sites. An assigned the ^{1}H,^{15}N-TROSY-HSQC spectrum used in this study can be found in Appendix A7 (see **Figure A16**). A chemical shift titration of NS2B-NS3^{pro} with fragment **A** at increasing concentration resulted in vast spectral changes associated with an enormous spatial reorganization of the WNV protease, including the β-hairpin of NS2B as described in a previous study[132] (cf. **Figure 2.17)**. The addition of **CR112** to NS2B-NS3^{pro} in complex with **A** yielded chemical shift

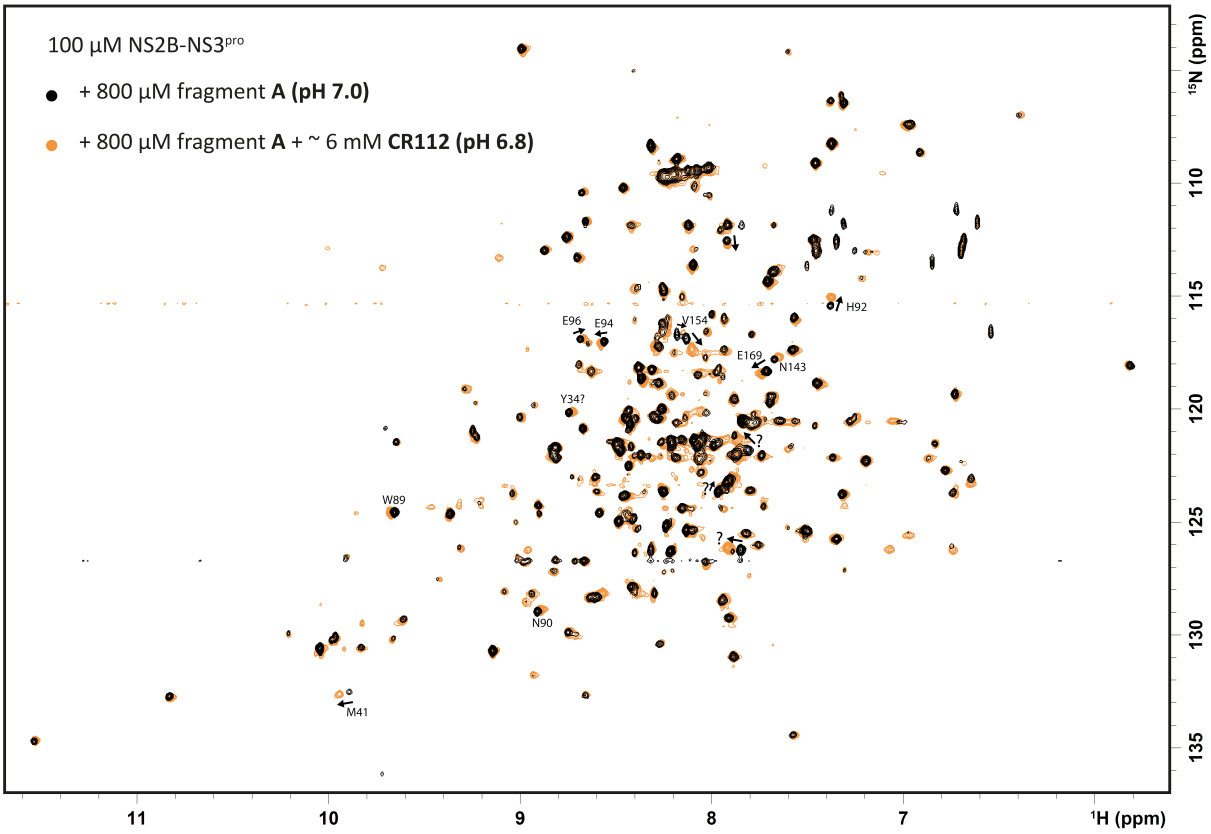


Figure 2.16 | Observation of a pH artefact during a chemical shift perturbation (CSP) analysis
In a first CSP experiment of U-[^{2}H,^{15}N]-labeled NS2BNS3^{pro} in the presence of fragments, a relatively small pH variation of about 0.2 units led to misinterpretation and wrong prediction of binding sites. Therefore, the pH of stock solutions was adjusted accurately before any further CSP experiment using a microvolume pH electrode (Orion Star A221).

perturbations of more than 0.03 ppm for the backbone NH signals of Leu 31 of NS3^{pro} and of Met 52* of NS2B (the asterisk indicates NS2B), whilst signals that had experienced large chemical shifts upon binding of **A** (e.g. Glu 101, **Figure 2.17**) were not affected. Additionally, four of the unassigned NH signals showed perturbations of greater than 0.03 ppm. Two of these signals are most likely Trp side chain NH resonances, whereas one of these two peaks may belong to Trp 53*. The other two NH signals likely belong to Met 59 and Ser 60, respectively, since these residues are not assigned and in contact with Met 52* and Trp 53* via backbone hydrogen bonds.
Interestingly, the secondary site, approximately 12 to 14 Å apart from the active site, is present in both, the protease in complex with **A** and the apo form. The same NMR signals are affected upon binding of **CR112** to the free enzyme and to the protease in complex with **A** (cf. Appendix A7, **Figure A17**). These NMR findings substantiate the conclusions derived from the fluorogenic protease activity assay that had identified fragment **CR112** as a non-competitive inhibitor.

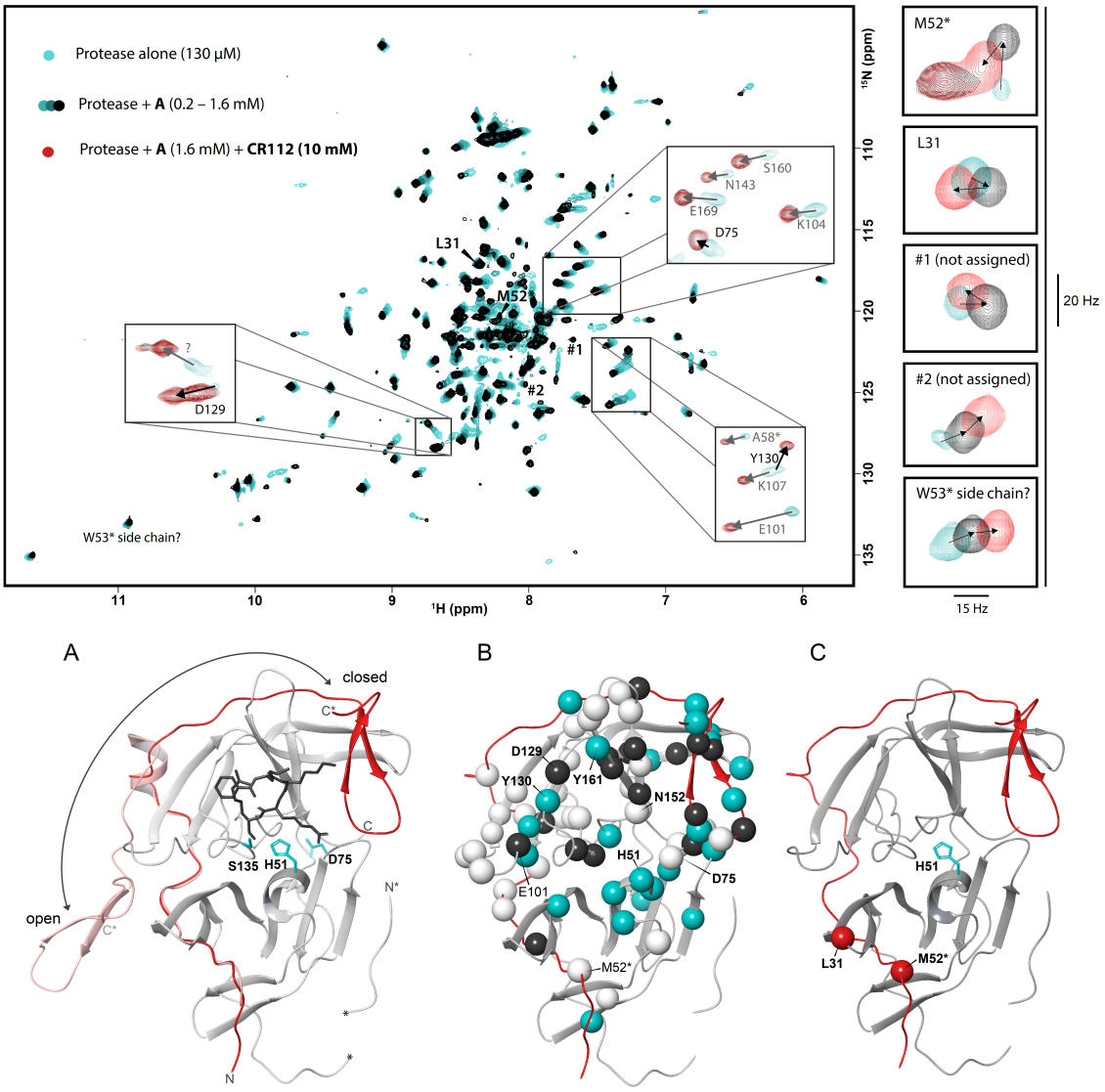


Figure 2.17 | Chemical shift perturbation of the WNV protease upon ligand binding
Top) Superposition of $^{1}H,^{15}N$-TROSY-HSQC spectra (7.3 Hz/ ^{1}H and 15.8 Hz/ ^{15}N resolution) of 130 µM u-$^{2}H,^{15}N$-labeled NS2B-NS3^{pro} titrated with increasing concentration of fragment **A** (turquoise to black) and in complex with both, 1.6 mM fragment **A** and ~10 mM fragment **CR112** (red). The ^{15}N backbone assignment was performed based on the BMRB entry 11053. Side chain nitrogen atoms and 50 out of 246 residues are not assigned (Prolines, the G_4SG_4-linker and residues 3-16, 19, 59, 60, 87, 131 – 133, and 151 of NS3^{pro}). Interaction of **A** with active site residues H51, D75, D129, Y130, N152, and Y161 (see REF[145]) results in an enormous rearrangement of the $^{1}H,^{15}N$-TROSY-HSQC spectrum. In the presence of **CR112**, however, only M52* (of NS2B), L31 and some of the unassigned resonances (signal #1, #2 and W53* side chain?) show significant chemical shift perturbations whereas resonance signals that shift greatly upon binding of **A** remain unaffected (e.g. E101, D129, or E169). Bottom) Representation of crystal structures of NS2B-NS3^{pro}. (A) Superposition of open (NS2B part, PDB 2GGV) and closed (PDB 2YOL) conformation of NS2B-NS3^{pro}. The Cofactor domain of NS2B (red) is critical for protease activity since in the closed conformation the β-hairpin is part of the S2 pocket. A peptidomimetic inhibitor 3,4-dichlorophenylacetyl-Lys-Lys-GCMA bound to the active site harboring the catalytic triad (H51, D75, and S135) is shown as a black stick model. The region between C* and N* and between the stars in NS3^{pro} is not resolved due to low electron density. (B) shows backbone nitrogen atoms associated with chemical shift perturbations of NH signals upon binding of fragment **A** as balls with good agreement with former 2D NMR studies of the WNV protease with **A**[168]. White balls: Resonances are broadened or overlapped in the absence of **A**. Black balls: Euclidian distances of more than 0.1 ppm. Turquoise balls: Chemical shift perturbations between 0.05 and 0.1 ppm. (C) Chemical shift perturbations observed upon binding of **CR112**. These perturbations are observed in both cases for binding to the WNV protease alone, or when saturated with **A** (see Appendix A7 and **Figure A17**). For color-coding, only chemical shift differences of spectra recorded in the presence of **A** and **CR112** were considered. The signals of M52* (of NS2B) and L31 shift 0.040 ppm and 0.034 ppm, respectively. The NH signal of H51 shows a chemical shift difference of 0.016 ppm. It is reasonable that signal #1 and #2 belongs to M59 and S60, respectively, since these residues are not assigned and in contact with M52* and W53* via backbone hydrogen bonds. The figure was adapted from REF[168].

Structure-Activity Relationships and Molecular Docking

To pave the way for lead optimization, structurally closely related fragments from the Maybridge library CR2006 were subjected to protease activity assays and ^{15}N-HSQC-TROSY based chemical shift perturbation experiments (corresponding spectra are shown in Appendix A7, **Figure A18 - Figure A21)**. Not all of these fragments had been identified as positive hits during the screening using STD NMR, because some were classified as insoluble fragments (**CR114** and **CR402**) and, hence, these fragments had not been included into the fragment mixtures. For the other fragments no unambiguous STD NMR signals could be detected (**CR053**, **CR069**, **CR082**, and **CR113**). One critical element of the structure-activity relationship (SAR) of these analogs seems to be the position of the carboxyl group. The constitutional isomer **CR113** loses inhibitory potency as compared to **CR112 (**see **Figure 2.18**), even though they most likely bind to the same protein binding site according to CSP experiments. Comparison of fragments **CR069** and **CR081** suggests that the annealed five-membered ring system is also important for successful inhibition. The exchange of one of the oxygen atoms with a methylene group in the 1,3-benzodioxole ring system of **CR081** dramatically decreases inhibitory potency.

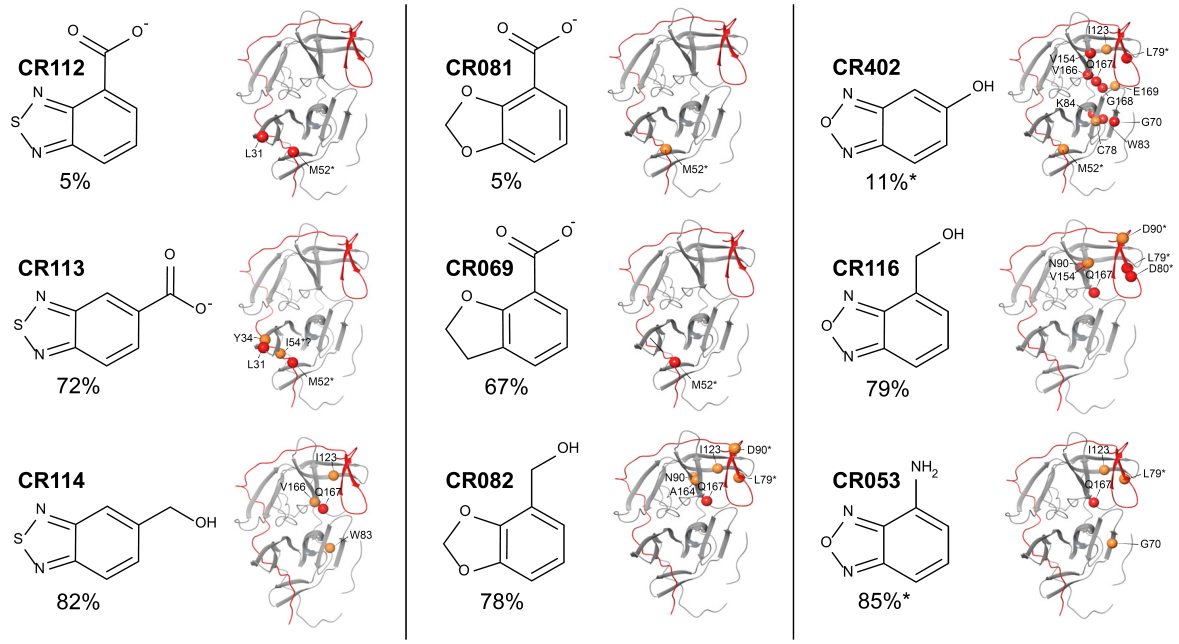


Figure 2.18 | Structure-activity-relationship (SAR) of structurally similar fragments, which bind to NS2B-NS3^{pro}
Chemical shift perturbations were plotted onto the crystal structure of the WNV protease. The backbone nitrogen atoms of amino acids with spectral changes are drawn as balls and are colored to identify the residues with, respectively, little and large perturbations between spectra recorded with and without fragment. For color-coding, chemical shift differences of spectra recorded with fragment **A** alone, and in combination with shown fragments were considered. Red: resonances shifted more than 0.03 ppm. Orange: resonance signals showed a chemical shift difference between 0.02 and 0.03 ppm. *N.B.: For compounds **CR053** and **CR402,** there is an overlap of absorption wavelengths with the excitation wavelength for the fluorophore AMC used in the activity assay. Therefore, the % values of these fragments were corrected as this is seen in **Figure 2.22**. The relative protease activity of 25 nM NS2B-NS3^{pro} supplemented with 900 µM fragment and 40 µM peptidic substrate is depicted below each chemical structure in percentage. The proteolytic activity of the WNV protease mixed with 40 µM substrate was set to 100%. The average standard deviation of duplicate measurements was ~5%. The figure was adapted from REF[168]

Since crystallization experiments with low-molecular-weight compounds had not been succeeded as experienced by other researchers[154], instead molecular docking was performed using Glide as implemented in Maestro 11 (Schrödinger, release 2017-2) to obtain structural models for **CR112** and analogs, which bind to the closed form of NS2B-NS3^{pro} (PDB 2YOL). Fragments were docked into corresponding binding sites in which chemical shift changes had been observed. In addition, fragment **A** was docked into the active site, yielding a pose that is in reasonable agreement with previous NMR data[145]. A docking score of -4.90 was obtained representing the approximation of the free binding energy in kcal · mol^{-1}. For fragments **CR112** and **CR081,** only a poor docking score of -2.69 and -3.00 was obtained, respectively, for the best docking pose. The corresponding analogs **CR113** and **CR069**, however, gave even lower scores, with -2.11 and -2.61 (see **Figure 2.19**). Note that a decrease of 0.65 units corresponds approximately to a 3-fold increase of the dissociation constant (K_D) according to the relation $\Delta G = RT\ ln\ K_D$, with R being the gas constant and T being the temperature.

A detailed docking model of **CR112** bound to the secondary binding site of NS2B-NS3^{pro} is illustrated in **Figure 2.19**. This model explains the observed chemical shift perturbations for Met 52*, His 51, and Trp 53* quite well. It is reasonable to assume that Leu 31 is probably indirectly affected, as it is positioned at the tip of a flexible β-loop. Moreover, docking models depicted in **Figure 2.19** along with SAR studies suggest that inhibition by binding to the secondary site is caused by a subtle perturbation of the conformation of the loop containing His 51, which is part of the catalytic triad. One requirement seems to be the right orientation of the carboxyl group pointing towards His 51 as seen for **CR081** and **CR112**.

Noteworthy, all non-charged analogs subjected to CSP experiments (**CR053**, **CR082**, **CR114**, **CR116**, and **CR402**) seem to bind, relative to the binding pocket comprising Met 52*, on the opposite side of the WNV protease. Furthermore, **CR402** seems to interact promiscuously with three different binding sites represented by Met 52*, Trp 83, and Gln 167 (cf. **Figure 2.20**). The latter two may also serve as binding pockets for **CR053** and **CR114**.

Except for fragment **CR402**, none of the other fragments affecting backbone NH chemical shifts of amino acids of these "alternative" secondary sites due to binding to NS2BNS3^{pro} led to notable protease inhibition. Based on the docking model for **CR402**, the inhibition mechanism could involve two binding events: (1) Interaction with Met 53* leading to malfunction of the catalytic triad as seen for **CR112** or **CR081** and (2) binding to the pocket composed of Glu 74, Lys 88, Ile 123, and Asn 167 of NS3^{pro} as well as Leu 79* and Phe 85* of NS2B. In particular, the interaction of **CR402** with residues of NS2B may be necessary for the inhibitory effect, since the β-hairpin of NS2B is crucial for an efficient catalysis.

During the course of this study additional 238 compounds, termed as TS2017 fragments, have been purchased from Thermo Fisher Scientific. Besides fragments sharing the chemical motifs 2,1,3-benzothiadiazole, 1,3-benzodioxole, or 2,1,3-benzoxadiazole, several other nine-membered ortho-fused ring systems were screened in terms of inhibitory potency. In total, 28 fragments were tested. The results are listed in **Table 2.4**. The majority of these fragments was ineffective. For the remainder of these fragments, only moderate inhibition was obtained as this has been seen, for example, for **CR069** or **CR113**.

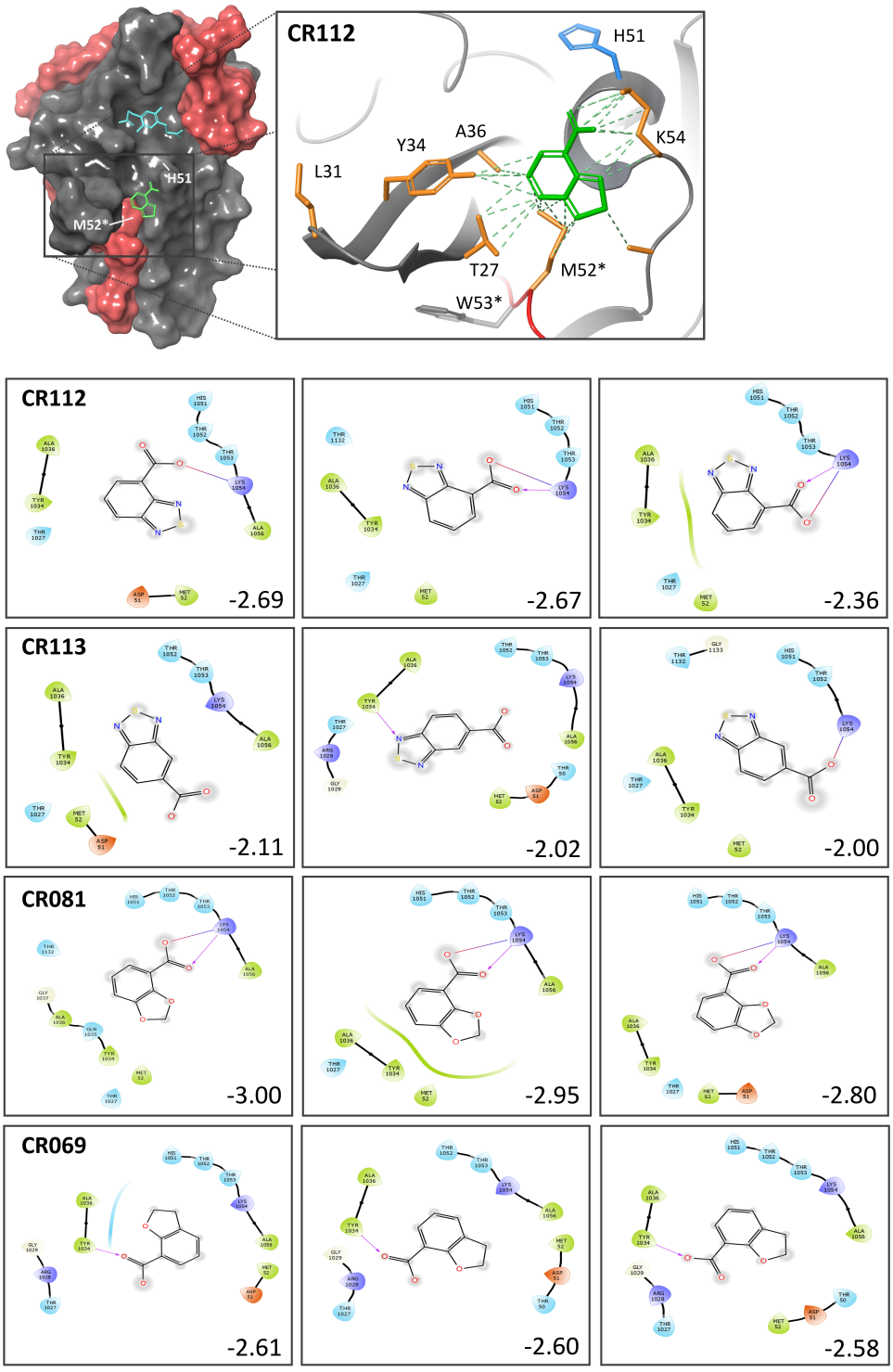
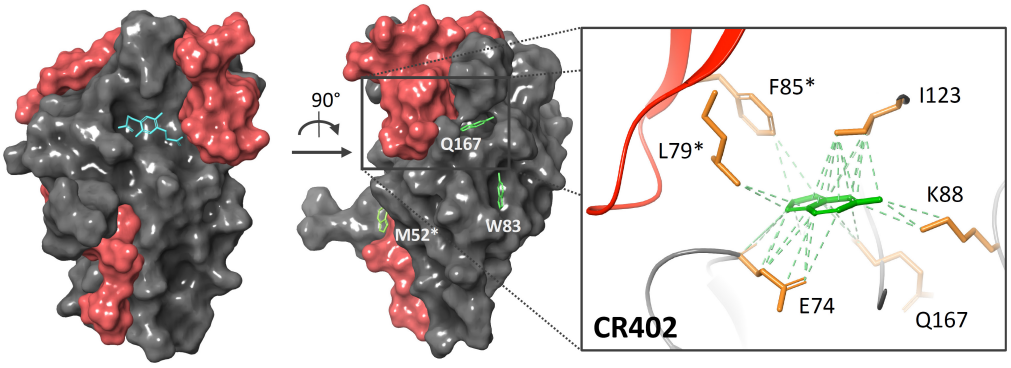
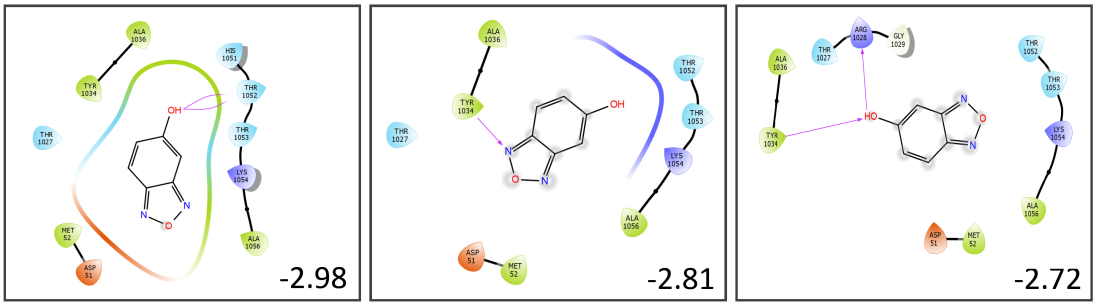


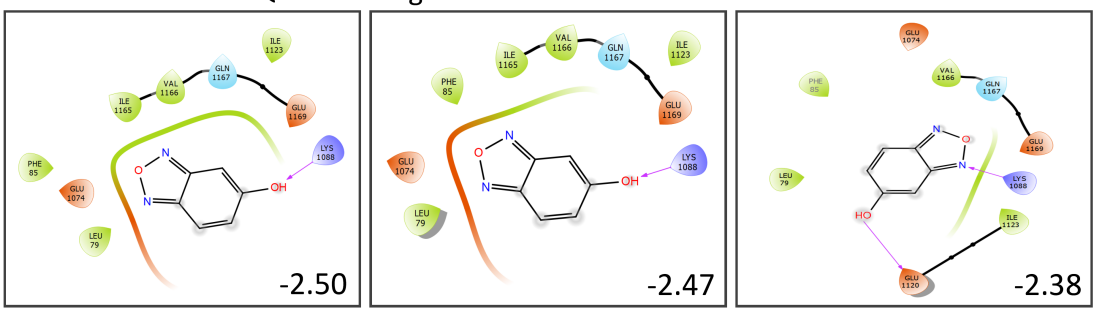
Figure 2.19 | Docking of fragments from the SAR analysis into corresponding putative binding sites
Protease activity assays and NMR data suggest that fragment **CR112** (green) follows a non-competitive inhibition mechanism mediated by the interaction with M52* (NS2B, red) corresponding to the chemical shift perturbation observed for the NH signal of M52*. For orientation, fragment **A** (turquoise) was docked into the active binding site of $NS3^{pro}$ (gray). The best three docking poses of **CR112** and analogs that interact with M53* are depicted. The ligand interaction diagrams were created using Maestro 11. The docking score (right bottom corner) represents the Gibbs free energy in kcal · mol^{-1}. The numeration of residues associated with $NS3^{pro}$ is modified (e.g. His 1051 = His 51) due to the numeration in the crystal structure.



CR402 docked into M52* binding site



CR402 docked into Q167 binding site



CR402 docked into W83 binding site

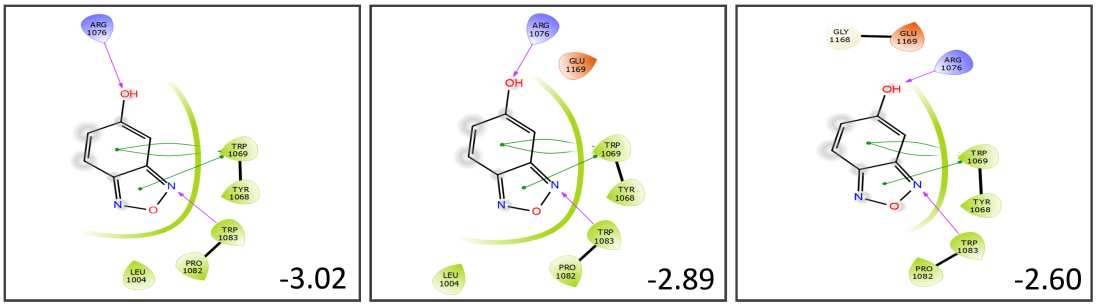
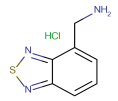
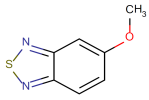
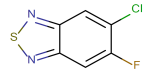
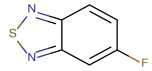
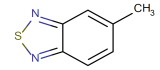
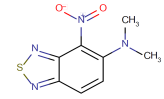
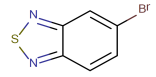
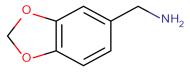
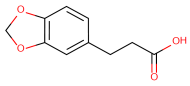
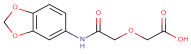
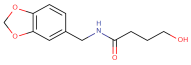
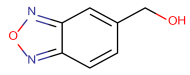
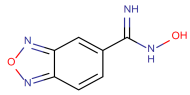
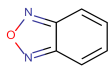
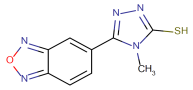
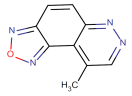
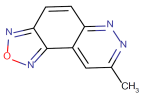
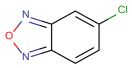
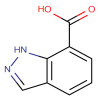
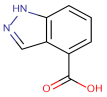
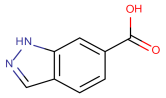
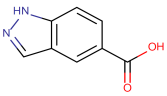
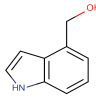
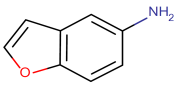
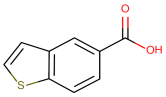


Figure 2.20 | Promiscuous binding of CR402 illustrated by computational docking
According to chemical shift perturbation using $^{2}H,^{15}N$-labeled WNV protease, fragment **CR402** seems to interact with three different binding sites represented by M52*, W83, and Q167. The best three docking poses per binding site are depicted. The docking score (right bottom corner) represents the Gibbs free energy in kcal · mol^{-1}. The numeration of residues associated with NS3^{pro} is modified (e.g. His 1051 = His 51) due to the numeration in the crystal structure. The interaction diagrams of **CR053**, **CR082**, **CR114**, and **CR116** can be found in Appendix A8.

Table 2.4 | Structure-activity relationships of TS2017 fragments

Compound ID	Structure	Protease activity	Std (n=2)
TS170		77%	8%
TS169		92%	2%
TS185		95%	10%
TS186		96%	4%
TS184		98%	9%
TS187		104%	21%
TS183*		109%	4%
TS171		112%	11%
TS164		77%	2%
CR019		86%	1%
TS189		95%	2%
TS188		97%	5%
TS172		79%	13%
TS182		98%	6%
TS180		100%	8%

Compound ID	Structure	Protease activity	Std (n=2)
TS235*		101%	2%
TS239		102%	20%
TS238		112%	0.4%
TS181		115%	2%
TS177		67%	6%
TS156		71%	3%
TS157		73%	1%
TS154		78%	2%
TS158		79%	12%
TS155		80%	6%
TS178		83%	10%
TS168		98%	4%
TS174		100%	9%

* normalized relative protease activity due to spectroscopic profile (for details see section **1.1**)

Std = standard deviation of two measurements

2.5 Discussion

Important aspects of the fragment libraries in terms of library design and handling will be addressed at the end of the next topic (see discussion **3.5.1** on page 159). In the following, the design and the production of the two-component WNV protease as well as the influence of buffer conditions on binding and activity assays, are briefly discussed. Moreover, strategies that avoid artefacts are highlighted. Positive hits from STD NMR screening will be assessed in general. Approaches for the identification of binding sites using ligand-observed and protein-observed NMR techniques followed by *in silico* docking are also discussed. In the end, suggestions are made on how to optimize initial hits towards a potent lead structure based on the gathered knowledge.

2.5.1 Protein Biosynthesis

Prior to the STD NMR screening and subsequent experiments, expression of NS2B-NS3^{pro} was conducted in *E.coli* cells following established protocols[119,126,192]. A detailed discussion of the biosynthesis is given in REF[133]. Briefly, approximately 44 mg per liter cell culture medium of unlabeled NS2B-NS3^{pro} could be obtained using nickel-affinity chromatography followed by gel filtration. In the case of isotope labeling, even ~77 mg of U-[^{2}H,^{15}N]-labeled protease per liter minimal medium (M9+/D_2O) and ~16 mg {U-[^{2}H,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Valγ-[$^{13}CH_3$]}-labeled NS2B-NS3^{pro} from 300 ml M9+/D_2O containing ILV precursors could be isolated. The purity of protein samples was judged to be at least ~95% according to SDS PAGE analyses. Since marginal auto-proteolysis had been observed, protein samples were stored at -70 °C until usage. To sum up, the optimized expression protocols gave decent yields of catalytically active protease sufficient to work with.

2.5.2 Evaluation of Activity and Binding Assays

In this study, a two-component protease construct was utilized to screen for potential allosteric inhibitors. The fusion construct has been designed by linking the central hydrophilic part of NS2B consisting of 45 amino acids to the N-terminus of NS3^{pro} via a flexible G4SG4 linker[119]. Whereby, the linker replaces a hydrophobic C-terminal part of NS2B, which is normally associated with the ER membrane. Moreover, it has been proposed that a hydrophobic loop (Leu 31 – Ser 33) of NS3^{pro} may serve as an anchoring site to stabilize NS3^{pro} on the membrane, as this region interacts with detergent micelles according to NMR experiments[193]. The question arises whether the water-soluble two-component protease reflects the *in vivo* membrane-associated state, and secondly, whether the construct used is a suitable drug target. On the one hand, the artificial character of the molecular model system becomes visible when considering the fact that the highest proteolytic activity for NS2B-NS3^{pro} has been found in buffer with pH values between 8 and 9 containing up to 30% (v/v) glycerol[119]. However, in this study, a 25 mM Tris buffer adjusted to pH 8.5 with only 5% glycerol was sufficient to obtain a reasonable activity **(**see **Figure 2.21A)**. Furthermore, only a relatively small amount of salt (30 mM) resulted in a loss of ~70% enzymatic activity **(Figure 2.21B)**. These conditions appear to be far away from being physiological.

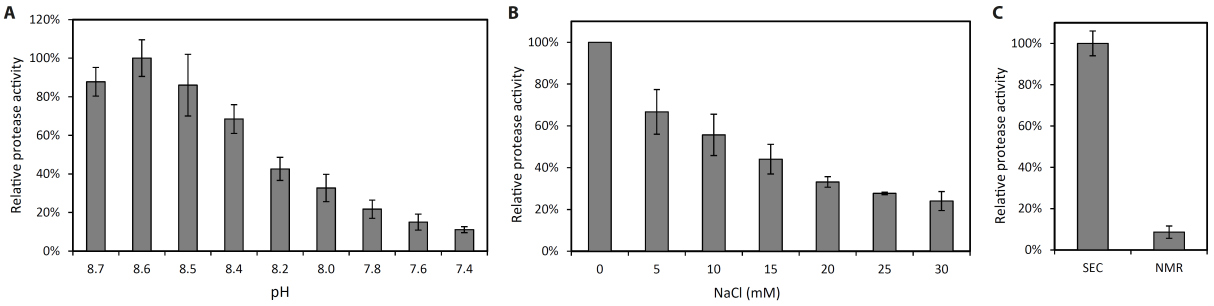


Figure 2.21 | Influence of buffer conditions on the protease activity
The relative protease activity was determined using fluorogenic cleavage assays conducted in 25 mM Tris buffer with 5% glycerol. The effect of different pH values (A), increasing amount of sodium chloride at pH 8.5 (B) and a comparison of SEC buffer (25 mM Tris, pH 8.5, 5% glycerol) and NMR buffer (25 mM sodium phosphate, pH 7.0) is depicted. The error bars represent the standard deviation of two measurements. This figure was partially adapted from REF[133].

On the other hand, the exact cytosolic environment near the ER membrane, including molecular crowding[194], is not known so far. The flexible linker may also contribute to the "artificial behavior" of the WNV protease. For the homologous DENV protease, an unlinked co-expressed protease construct was subjected to activity assays and NMR experiments[195]. This unlinked protease complex was catalytically active at neutral pH and in the absence of glycerol. The presence of salt in the assay buffer still inhibited NS2B-NS3^{pro} activity. Furthermore, experiments based on paramagnetic relaxation enhancement suggested that the essential NS2B cofactor region of a unlinked DENV protease complex exists predominantly in a conformation similar to that of the linked protease in the so-called closed conformation. This holds true even in the absence of inhibitors. Hence, the authors concluded that the unlinked construct provides a more realistic model for biochemical assays or inhibitor screening. Since no data exist for an unlinked WNV protease construct, it remains unpredictable whether the findings for the DENV protease also holds true for the WNV protease.

The STD NMR screening was conducted in 25 mM sodium phosphate buffer (pH ~7.0), although only ~10% of the maximal protease activity had been obtained under these conditions. Nall et al. was one of the first who explored the pH dependency of the WNV protease. But no satisfactory explanation for the pH dependency was presented. One possible explanation may involve two mechanisms: Sodium ions may compete with positively charged amino acid residues of the peptidic substrate (Lys or Arg) in the active site and, secondly, the equilibrium of protonation states of critical amino acid residues (e.g. His) of the protease may be altered upon pH variations. The latter aspect also affects STD effects of low-molecular-weight compounds that bind to NS2B-NS3^{pro} as saturation transfers from a receptor to its ligand also depends on the density of protons within the corresponding binding pocket[81,196]. Moreover, it is assumed that the major advantage of fragment screening is the fact that only fragments, which make high-quality contacts with the receptor, exhibit affinities that are high enough so that these fragment are detected as positive hits[74]. If only one of these few interactions (e.g. a hydrogen bonding) is missing due to an altered protonation state, some fragments may fail to bind with sufficient affinity, which would have specifically bound at different pH values. However, the influence of altered protonation states was expected to be low.

Using the STD NMR settings as described above (cf. section **2.3.7**) combined with a threshold of 0.3% attenuation of signal intensity, a hit rate of about ~29% was obtained, which is comparable with hit rates for medium-sized proteins (<40 kDa) observed in other studies[197]. In order to identify important properties that promote binding to the WNV protease, positive hits were visually clustered based on their structural scaffolds and chemical groups, yielding the so-called SxG matrices (see **Figure 2.10**). In Addition, principal component analysis and descriptive statistics were utilized. In conclusion, these analyses revealed no single parameter that defines binders or non-binders respectively. A slight preference for biphenyl derivatives and aromatic fragments characterized by two five-membered rings that are separated by one bond could be observed. This may be explained by findings of Rademacher et al. (cf. REF[89]) that fragments with symmetric and isolated aromatic protons show higher STD effects due to longer T_1 relaxation times, hence STD signals are better detectable as they are far above the noise level of STD NMR difference spectra. On the contrary, the best 20 binders according to STD effects are from other scaffold clusters (except **CR360**). Nonetheless, all of these fragments also exhibit symmetric or isolated aromatic protons. Even though SxG matrices are useful to provide a condensed and simplified representation of positive hits, a selection of promising lead structures is not possible based on this kind of analysis as the diversity of chemical structures is still too high. The combination of abstract scalar descriptors can deliver a more defined picture depending on the parameter chosen. In the case of the WNV protease, neutral medium-sized fragments (140 – 180 $Å^3$) with two hydrogen bond acceptors and one hydrogen bond donor made up ~20% of positive hits (n = 25), whereby only two of these fragments were found among the best 20 hits according to their STD effects.

It is well established, that saturation transfer from a protein target to a ligand critically depends on several factors, with the dissociation rate of a ligand being of critical importance. For aromatic low-molecular-weight compounds as used in this study, it is assumed that dissociation rates are high, and thus STD intensities qualitatively reflect binding affinities. However, besides the underlying kinetics of the exchange process, the composition of corresponding binding pockets and the ligand's longitudinal relaxation also determine the observed STD effect[89]. Neither a ranking of fragments exclusively based on STD effects nor a selection depending on specific scalar descriptors revealed important scaffolds with distinct molecular properties to derive promising lead structures. Therefore, a validation of positive hits by the use of complementary methods was inevitable.

To extend selection criteria, protease activity assays were performed. No simple correlation was observed between STD effects and the ability to inhibit the WNV protease, suggesting that a substantial fraction of positive hits interact with sites remote from the active site. On the other hand, the increased pH of 8.5 of the assay buffer may alter binding sites as described above, and hence influence the interaction behavior of interacting fragments.

At this point, two potential causes of artefacts often encountered during functional fluorogenic assays shell be discussed. Small molecule non-specific inhibitors, which act by aggregation in solution and undesirable precipitation in aqueous assay buffers, are major issues in drug discovery research[198]. Detergents are often used to evade this problem. Ezgimen and co-workers evaluated the effects of commonly used detergents, including Triton X-100, Tween 20, Nonidet-20 (NP-20), Brij-35, and CHAPS, on the enzymatic activity of WNV protease[198]. The study's results indicated that

nonionic detergents such as Triton X-100, Tween 20, and NP-40 were unsuitable for the purpose of discrimination of specific versus promiscuous inhibitors. According to the authors, the zwitterion detergent, CHAPS, had no significant effect on protease activity, and hence could be used to distinguish false positive hits from "true" hits. In contrast, it has been shown by another group that CHAPS directly interacts with amino acid residues (Thr 52 and Thr 53) in proximity to the active site of the WNV protease NS2B-NS3^{pro} (cf. REF[137]). For this reason, none of the proposed detergents were utilized in this present study. Moreover, no aggregation-driven chemical shift change has been observed in 1D proton NMR spectra for the most potent fragment **CR112** at concentrations up to 6 mM (data not shown) as this was seen for other fragments (e.g. **CR030**; cf. REF[89]). Another reason for obtaining false positive hits derived from fluorogenic activity assays is a phenomenon termed quenching, which refers to any process that decreases the fluorescence intensity of a given substance (here: AMC). The fragments **CR053**, **CR402**, **TS183**, and **TS235** were identified to quench the fluorescence of the fluorophore AMC, which is monitored during protease activity experiments (**Figure 2.22**). A simple indicator for these fluorogenic compounds is a yellowish color when present in dissolved form. In conclusion, it is advisable to put efforts into the identification of "bad actors" using NMR and UV/VIS spectroscopy prior to any activity assay.

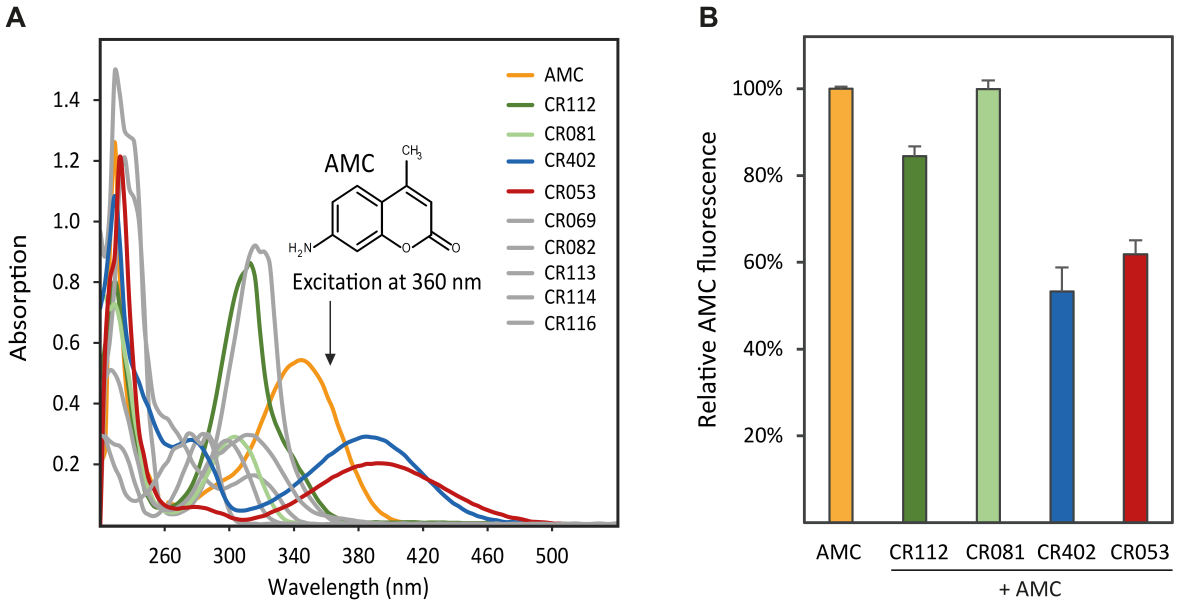


Figure 2.22 | Quenching of AMC fluorescence in the presence of different fragments
(A) Absorption spectra of fragments from structure-activity relationship studies. Each sample contained 900 µM fragment in buffer B (25 mM Tris/HCl, 5% (v/v) glycerol, pH 8.5). In a protease activity assay, the cleaved fluorophore 7-amino-4-methylcoumarin (AMC) is excited by monochromatic light at a wavelength of 360 nm. The emission of light is detected at 460 nm. (B) Fragment **CR402** and **CR053** interfere significantly with the protease activity assay due to the extinction of light at 360 nm. Each sample contained 900 µM fragment in the presence of 4 µM AMC in buffer B. The relative fluorescence of 4 µM AMC was set to 100% (n=2). This figure was adapted from REF[168].

2.5.3 Identification of Binding Sites

A significant contribution, in order to narrow down the choice of potential allosteric inhibitors, came from competition assays using STD NMR and fragment **A** that served as active site binder. A comprehensive discussion, including a detailed comparison of chemical structures, can be found in REF[133] and REF[168]. Briefly, the addition of **A** in a 20-fold excess over fragment concentration caused a notable decrease of the STD effect (≥ 25%) of only 14 out of 36 fragments analyzed. No obvious common scaffold could be identified, albeit nine of these compounds are positively charged. Similar to **A**, these nine fragments contain an amino group (primary to tertiary) connected to a hydrophobic part of the molecule by one to three bonds. It is assumed that a reduction of STD effects, directly indicates competitive binding. This may not being true for fragments, which preferentially interact with binding sites of NS2B-NS3^{pro} that are exclusively present in the "open" conformation, since fragment **A** seems to shift the equilibrium towards the closed state. In the case of the fragments **CR120**, **CR297**, and **CR351** it can be safely assumed based on their relatively high similarity with **A** that these fragments act by competitive inhibition, with estimated IC_{50} values ≤ 900 μM.

Interestingly, ten fragments of the remaining 22 inhibitors experienced a significant enhancement of saturation transfer (≥ 25%), whilst twelve fragments were unaffected. Opposed to above-mentioned potential active site binders, eight out of ten fragments showing enhanced STD effects in the presence of **A** are negatively charged (i.e. **CR081**, **CR112**, **CR115**, **CR121**, **CR192**, **CR204**, **CR395**, and **CR485**). This raises the question whether the salt formation of **A** and corresponding fragments may cause artefacts in STD NMR spectra. Fortunately, only marginal direct irradiation was detected in negative controls comprising a mixture of fragments and **A**. Therefore, an enhancement of STD intensities can be safely interpreted as being due to increased affinities to secondary binding pockets. In order to substantiate these findings, **CR112** had been chosen for further analyses. Both a fluorogenic competition assay using Dixon's method[189] and ligand-observed titration of **A** in the presence of **CR112** confirmed the hypothesis of a non-competitive inhibition mode. The final proof that **CR112** interacts with a binding site remote from the active site came from $^{1}H,^{15}N$-TROSY-HSQC-based chemical shift perturbation experiments under stringent pH control. The underlying principle is that a modulation of the chemical environment around an NMR-active nucleus (here: $^{1}H/^{15}N$), e.g. upon ligand binding, directly lead to chemical shift changes of corresponding NMR signals. Aromatic compounds are beneficial for this type of analysis, as aromatic ring currents induced by a magnetic field can lead to large chemical shift perturbations. A phenomenon referred to as magnetic anisotropy effect[95]. Unexpectedly, the magnitude of chemical shift changes was relatively low ($\Delta\delta \leq 0.04$ ppm) in the case of **CR112**, even at saturating amount (~95% saturation). In contrast, larger chemical shifts were observed ($\Delta\delta_{max}$ ~ 0.2 ppm for Asn 167) for other fragments (e.g. **CR402,** see Appendix A7, **Figure A20**). It is important to note that larger chemical shifts derived from a single-point titration at saturating amounts, though, do not necessarily imply higher affinities. It is rather the orientation and position of the aromatic ring that determines the magnitude of chemical shifts. In the case of **CR402** for example, the docking model suggests that the benzene ring is located about 3 to 4 Å above the backbone NH of Asn 167 almost in a vertical position (see **Figure 2.20**).

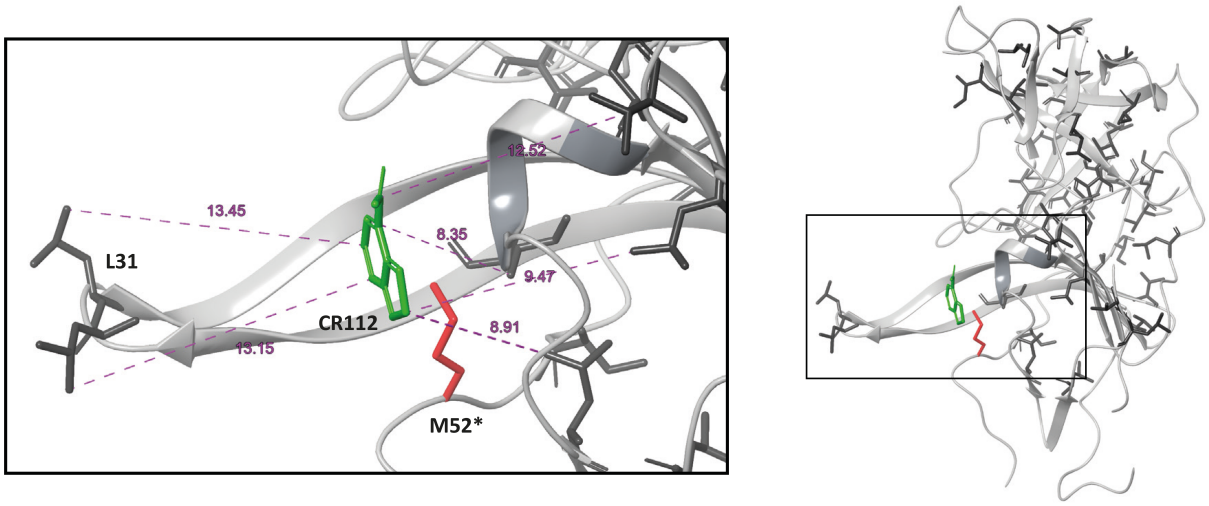


Figure 2.23 | Distribution of ILV residues throughout the WNV protease NS2B-NS3^{pro}
The fragment **CR112** (colored in green) was docked into the putative binding site according to $^{1}H,^{15}N$-TROSY-HSQC-based chemical shift perturbation experiments. Isoleucines, leucines, and valines are colored in black. The distances indicated with dashed purple lines between **CR112** and surrounding ILV residues are given in Å.

Note that the binding model of **CR112** derived from *in silico* docking would explain why ^{13}C-methyl ILV-labels of NS2B-NS3^{pro} were not affected upon binding of **CR112**, as the closest isotopic probe was more than 8 Å away (see **Figure 2.23**). However, it remains unclear why the backbone NH signal of Leu 31 exhibited a significant chemical shift change (~0.03 ppm), whilst the methyl group carrying a ^{13}C-label was not being perturbed. It is possible that small changes of the β-sheet induces marginal conformational changes in the β-turn (about 2 Å indicated by alignment of several crystal structures, PDB codes: 2FP7, 2YOL, 2GGV, 2IJO, and 3E90; not shown), which affects only the chemical environment of the buried backbone nitrogen atom of Leu 31, whereas the solvent-exposed methyl group is insensitive to perturbations.

2.5.4 Hit-to-lead Optimization

Analogs of **CR112** were subjected to protease activity assays to identify steric and electronic features that are critical to block the biological response of NS2B-NS3^{pro}. It appears that both the presence and the right position of a carboxyl group attached to the scaffolds 2,1,3-benzothiadiazole or 1,3-benzodioxole is of critical importance to ensure inhibitory potency associated with binding to Met 52*. Moreover, two symmetric polar atoms (i.e. N or O) at the ortho-fused five ring are also required. The docking models of **CR069**, **CR081**, **CR082**, and **CR112** suggested that a specific interaction of the carboxyl group and Lys 54 is achieved only for **CR081** and **CR112**, conferring notable inhibitory potency to these fragments. Surprisingly, a conversion of a carboxyl group to a hydroxyl group (e.g. **CR114**) resulted in another binding site, and hence to reduced inhibitory effects. These neutral fragments seem to bind preferentially to another secondary binding represented by Asn 167. Here, the nature of fragment-based lead discovery comes to light. Since only a few interactions are formed between a low-molecular-weight compound and its receptor, a small structural change (e.g. **CR113** to **CR114**) can lead to a completely different binding mode.

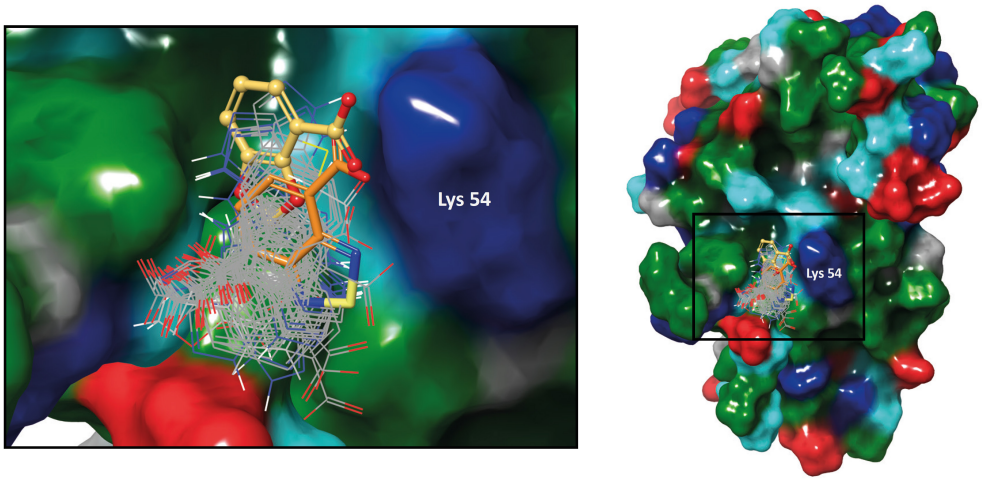


Figure 2.24 | Docking of CR112 analogs that carry a carboxyl group into a secondary binding site of NS2B-NS3^{pro}
Eight fragments from **Table 2.4** carrying a carboxyl group (except **TS189**) were docked by the use of Glide as implemented in Maestro 11. For each fragment (carbon atoms in gray), one to seven poses with docking scores ranging from -2.0 to -4.6 kcal mol^{-1} are depicted. The carbon atoms of **CR081** and **CR112** are colored in yellow and orange, respectively. Nitrogen atoms of fragments are blue, associated hydrogen atoms are white, oxygen atoms are red, and sulfur are light yellow. The amino acid residues are colored using Maestro according to their properties: Negatively charged = red, positively charged = dark blue, polar = light blue, non-polar = green, and glycines = gray.

Docking models of additional eight fragments carrying a carboxyl group (see **Figure 2.24**) substantiate the aforementioned hypothesis, assuming that these fragments may also interact with Met 52*. None of these fragments were as potent as **CR081** or **CR112** (cf. **Table 2.4**). In the majority of poses, the carboxyl group is positioned opposite to Lys 54 as this was seen for **CR069**. In other cases, the carboxyl groups do not match with that of **CR112**, highlighting the small degrees of freedom in terms of the position of a carboxyl group when considering the development of a potent lead structure. In conclusion, only modifications of the benzene ring of **CR081** or **CR112** at position five, six, or seven may lead to optimized compounds.

Besides **CR081** and **CR112**, the compound 2,1,3-benzoxadiazol-5-ol (**CR402**) also exhibits a relatively high inhibitory potency as compared to the remainder of all fragments analyzed. According to CSP experiments, at least three binding sites are affected upon binding of **CR402**. This promiscuous binding behavior may be the main cause of protease inhibition as described above. On the other hand, assuming that **CR402** binds preferentially to a binding site comprising Asn 167, perturbation of the essential cofactor domain NS2B may also explain the inhibitory effect of **CR402**. The docking of analogs that seem to bind to the same interaction site gave reasonable binding models (see Appendix A8, **Figure A42**). However, neither docking scores nor docking poses can explain why solely **CR402** function as a potent protease inhibitor by binding to this particular secondary binding site. In this example, but also in the above-mentioned example, results derived from docking should be treated with caution. At this stage, it provides only a computer-aided working hypothesis. Two facts hamper the interpretation of docking outcomes: the rigid receptor approximation utilized in Glide and the small size of molecules that have been docked. The rigid receptor approximation is an issue particularly when fragments are docked into sites, which comprise parts of NS2B, due to its proposed flexibility. To ensure reliable docking results

despite the low molecular weight of ligands, a so-called extra precision (XP) scoring function[103] was used to rank diverse poses, involving terms that are common to other empirical scoring functions: an atom-atom pair score that rewards interactions between lipophilic atoms on the protein and ligand, a term favoring protein-ligand hydrogen bonds, and an entropic penalty according to the number of rotatable bonds in the ligand. One unique feature is the fact that XP also imposes desolvation penalties for embedded polar and charged groups of both the protein and ligand. Besides others, these terms cover the main aspects that dictate a binding event. However, the authors[103] admitted that empirical corrections of parameterizations were made in the case of smaller ligands based on limited data, and hence improvements may still be necessary[103].

Based on protease activity data, optimization of **CR402** appears to be a major challenge as well. Removing (**TS180**) or substituting (**TS181**, **TS182**, **TS235**) the hydroxyl group resulted in complete elimination of inhibitory potency. But also a subtle change towards a more flexible hydroxyl group (**TS172**) led to a virtually inactive inhibitor. Therefore, only the motif 2,1,3-benzoxadiazol-5-ol can serve as a promising core structure for further modifications. Without any doubts, systematic evaluation of synthetic options will be immense work, and hence quite a challenge.

The development of novel WNV protease inhibitors based on the fragments **CR081**, **CR112**, or **CR402** may involve two strategies: either fragment growing or fragment linking. In the following, critical aspects that may guide the hit-to-lead optimization are summarized. In the first stages of optimization, researchers should focus on the enthalpic component, as the efficient interaction of compounds with the receptor's interface is one requirement to obtain high-affinity binders (cf. REF[74,158,199–201]). For this purpose, the concept of ligand efficiency has been developed by Hopkins and co-workers[202]. This measure has widely been used in drug research. LE represents the average binding energy per non-hydrogen atom (LE = 1.37/number of heavy atom · (-log IC50)). In order to follow a successful optimization trail, the LE of the novel lead structure should be always greater than 0.3 after every optimization step. In other words, adding ~75 ± 10 Da (corresponds to ~5-6 non-hydrogen atoms) to the fragment should result in an increased potency of one log unit (cf. REF[201]). Moreover, an article of Higueruelo et al. has revealed that synthetic molecules bound to proteins engage in less polar contacts than biological complexes and possess a higher portion of unmatched heteroatoms at the interface than the natural ligands. In conclusion, both the density and quality of polar contacts should be enhanced in the early state of the hit-to-lead process while molecules remain relatively small[200]. Yet, to identify unmatched heteroatoms a reliable binding model is indispensable. Finally, besides hydrogen-bonding properties, in particular, few H-bond donors[203], a definite lipophilicity seems to be a key feature in drug development, as these index has remained virtually constant in oral drugs over time[199,204]. Hence, the lipophilic ligand efficiency (LLE) is an informative measure, combining in vitro potency and lipophilicity (LLE = -Log IC50 – cLogP or LogD)[199]. During the evolution of a lead structure, LLE should remain above ~5 units. Note that increasing the binding affinity of new therapeutic agents has tended to result in greater lipophilicity, a property that substantially reduces the chances of a successful drug development pipeline[199].

2.6 Summary and Outlook

Following successful biosynthesis of the two-component WNV protease NS2B-NS3^{pro} using established bacterial expression systems, 429 water-soluble (≥ 1 mM), low-molecular-weight compounds (< 500 Da) of the Maybridge CR2006 library were screened by the use of saturation transfer difference (STD) NMR experiments as initial read-out. By using Python-based *in silico* preparations of fragment mixtures and optimized NMR-settings 126 positive hits with STD effects ranging from 0.3% to 9.4% could be identified within less than one week of measurement time. In order to find important scaffolds with distinct molecular properties for the development of promising lead structures, descriptive statistics (e.g. SxG matrix) and principal component analysis of scalar molecular descriptors were performed. Neither a ranking of fragments exclusively based on STD effects, nor a detailed search for common molecular hallmarks, such as scaffolds, chemical groups, or feature counts (e.g. H-bond acceptors), shed light on the underlying structural principle of binding fragments.

To narrow down the choice of positive compounds, fluorogenic protease activity assays were conducted. Approximately 85% of these fragments were capable of reducing the enzymatic activity of NS2B-NS3^{pro}, whereby 12 compounds exhibited an IC_{50} less or equal to 900 µM (see Appendix A2; **Figure A2** and REF[133,168]). In the search for potentially allosteric inhibitors, STD-NMR competition experiments using the known active site fragment **A** as reporter ligand yielded 14 competitively binding fragments, and 22 fragments not competing with **A**. Interestingly, 10 of these 22 fragments showed a notable increase of STD intensities in the presence of compound **A** suggesting cooperative binding.

Subsequently, the most promising non-competitive inhibitors **CR081** and **CR112** (IC_{50} ~ 500 µM) sharing a structural motif were further analyzed by employing protein-observed chemical shift perturbation (CSP) studies to identify matching protein binding sites. Whilst binding of **CR112** to {Ileδ1-[$^{13}CH_3$]; Leu,Valγ-[$^{13}CH_3$]}-labeled NS2B-NS3^{pro} remained undetected due to the presence of "blind spots", chemical shift changes of ^{1}H,^{15}N-TROSY-HSQC spectra were clearly notable when **CR112** was titrated into a sample containing uniformly [^{2}H,^{15}N]-labeled WNV protease. However, this was only accomplished by controlling the pH as accurately as possible. These data suggest that **CR112** interacts with Met 52* of NS2B. Thus, we have identified a secondary binding site about 10 Å adjacent to the active site. During the course of structure-activity relationship analyses, at least one additional potentially allosteric binding site represented by Asn 167 has been discovered. The fragment **CR402** seems to interact preferentially with this binding site, conferring inhibitory potency to this fragment. However, the presence of **CR402** also caused chemical shift perturbations of NH signals of Met 52* and Trp 83, which indicated promiscuous binding.

To guide the interpretation of experimental results from STD-NMR, CSP experiments, and structure-activity relationships, **CR112** and analogs were docked into corresponding binding sites of NS2B-NS3^{pro} using a crystal structure of the closed conformation. These docking models indicate that the position of a carboxyl group of **CR112** within the binding pocket pointing towards His 51 is of critical importance. Therefore, it is assumed that inhibition by binding to this second site is caused by subtle perturbations of the conformation of the loop containing His 51, which is part of

the catalytic triad. For fragment **CR402**, molecular docking was only of limited use. The interaction with Leu 79* of NS2B may be essential for the inhibition of the WNV protease. Site-directed mutagenesis may answer this question.

Optimizing **CR081**, **CR112**, or **CR402** will be quite a challenge, as none of the 34 closely related analogs possess comparable inhibitory potency, rendering the corridor of possible chemical modifications relatively small. Nevertheless, these fragments are valuable starting points for the design of potent allosteric WNV protease inhibitors. Fragment growing at positions five or six of the benzoic acid ring or phenyl ring will probably provide promising lead structures, while keeping in mind to increase the density and quality of polar contacts and to aim that LE and LEE remain greater than 3 and 5, respectively, as described above. To pave the way for rationally optimized inhibitors, a more constrained structural model based on experimental data will be necessary.

Since crystallization was unsuccessful so far despite huge efforts, other strategies have to be conceived. A study of Zhang and co-workers provides one possible orientation using the so-called STD-NMR/Docking/CORCEMA-ST method[205]. In this strategy, a complete relaxation and conformational exchange matrix analysis of saturation transfer (CORCEMA-ST)[206,207] is applied on high-quality STD NMR data to either select for the best docking pose within one binding site (see also REF[208,209]), or to identify the best corresponding binding site when dealing with multiple binding pockets[210]. Nonetheless, assumptions have to be made about saturated residues of the protein and corresponding correlation times. Moreover, rigid receptor approximations must not be ignored. Future molecular modeling may involve induced fit docking as implemented in Maestro to account for the flexibility of NS2B and other parts of the WNV protease followed by molecular dynamics simulations. Moreover, the orientation of a fragment bound to the receptor site can be in principle determined by mapping its solvent accessibility using waterLOGSY-based NMR experiments[211]. In particular for small aromatic molecules, epitope mapping by the use of STD NMR has often been difficult to interpret[89,211] due to different longitudinal relaxation rates between aliphatic and aromatic protons (cf REF[89]). Hence, waterLOGSY may be more satisfactory. Yet, in the case of **CR112,** it is almost impossible to determine the orientation due to its small spin system**.** Alternatively, virtual screening of the ZINC database using a filter for carboxyl group-containing compounds that match with the position of **CR081** or **CR112** may be beneficial.

Moreover, the binding modes of the remainder of potent fragments that presumably do not bind in the active site of NS2B-NS2^{pro} (i.e. **CR025**, **CR178**, **CR204**, **CR317,** and **CR448**) may be explored. Besides **CR081**, **CR112** or **CR402**, these fragments may be also linked to the active site binder **A** to obtained novel bidentate WNV protease inhibitors.

Despite all considerations, in the end, it still remains uncertain whether or not the binding pockets of the fused NS2B-NS3pro WNV protease reflects the *in vivo* equivalent in the exact same manner. To confirm this, cell-based infection assays should be performed before any optimization efforts. Moreover, the site of action should be taken into account, as immunodeficient patients represent a major target group of antiviral therapy, where WNV likely cause fatal neurological diseases[109–111]. Therefore, optimized lead structures should not only fulfill ADME requirements but also should be able to cross the blood-brain barrier[212].

3 The Major Capsid Protein of Murine and Human Noroviruses

3.1 Introduction

Noroviruses (NoV) belong to the family *Calciviridea*. Up to now, seven genogroups (GI to GVII) have been described based on the sequence identity of the major structural protein VP1. Members of the genogroups GI, GII, and GIV cause diseases in humans, whilst members of GIII infect cows and swines[213], GV is found rodents (rats and mice)[214,215], and GVI to GVII in dogs[216]. Besides rotaviruses, human noroviruses are the leading cause of a non-bacterial epidemic outbreak of gastroenteritis across all ages, in particular in children[217], with GII.4 being the predominant genotype. Symptoms of NoV infection are rapid onset of watery diarrhea and vomiting often combined with abdominal cramps, headaches, myalgia (muscle pain) and fatigue[218,219]. Although the disease is usually self-limiting, human NoV cause several million symptomatic infections worldwide[220] and no effective vaccines, or specific therapies are available yet to combat human NoV infection.

3.1.1 Norovirus Genome Organization

Noroviruses have a linear, single-stranded, positive-sense RNA genome ranging from 7.5 to 7.7 kb in length, which is flanked by highly conserved untranslated regions. The overall genome organization is conserved among all known members of the *Calciviridea* family. The polycistronic genome comprises at least three open reading frames (ORF) [221]. The first ORF encodes a large polyprotein that is co- and posttranslationally processed mainly by the virus-encoded NS6 protease. For murine NoV (MNV), host caspases may also be vital[221]. Nonstructural proteins include a nucleotide triphosphatase, the VPg protein covalently linked to the 5’ end of the genomic RNA, a 3C-like cysteine protease, and a RNA-dependent RNA polymerase. The second ORF and third ORF encode the major capsid protein VP1 and the minor structural protein VP2, respectively, whereas the latter is thought to be involved in genome packaging[222]. Only for MNV, a fourth ORF has been reported that encodes the nonstructural protein VF1 (virulence factor 1). It is believed that this protrein acts as an antagonist of the innate immune response[223].

3.1.2 Norovirus Capsid and Cell-entry

During virus maturation, 180 copies of the major capsid protein VP1 assemble, resulting in the non-enveloped icosahedral virions. It has been well described that dimers of the protruding (P) domain of VP1 of human NoV interact with histo-blood group antigens (HBGAs) as initial attachment factors in a strain-dependent manner[10,224]. HBGAs are found in mucosal secretions and on epithelial cells[225]. Moreover, it has been demonstrated that human NoV utilizes HBGA,which are produced by enteric bacteria, to attach and productively infect B cells[226]. In contrast, recent *in vitro* studies with intestinal epithelial cells suggested a HBGAs-independent binding of the human NoV strain

GII.6 Ueno 7k[227]. Besides HBGAs, other potential cofactors that promote or inhibit infectivity has been proposed (e.g., heparan sulfate or breast milk glycans)[228].

Nowadays, the interaction of several types of HBGAs with truncated dimers, so-called P dimers, of GII.4 human NoV has been thoroughly studied by the use of crystallography[224] and NMR methods[10,11,90]. These studies propose a cooperative multistep binding process, which involves two to four glycan binding sites. Due to the lack of efficient cell culture systems or small animal models for human NoV, MNV serve as role model because it is the most prevalent virus in research mice[15,229]. The infection of murine macrophages[15] by MNV depends on α2,3-ganglioside-linked terminal sialic acid moieties (e.g., GD1 or GM3). Binding to glycolipids and glycoproteins in a strain-dependent manner has been also reported. The region of the glycan binding site that was mapped based on reverse genetics is topologically similar to the HBGA-binding sites of human NoV[230]. Recently, the two surface molecules CD300If and CD300Id were identified as functional receptors for murine NoV[25]. Moreover, it is assumed that the cell entry is mediated through a non-clathrin-, non-caveolae-, dynamin- and cholesterol-dependent pathway[231] after attachment to the cell surface. Despite the different binding specificities of different NoV due to antigenic variability, the folding pattern of the P2 domain is conserved in all *Caliciviridea* P domain structures, which comprise six antiparallel β-sheets that form a β-barrel-like motif (see **Figure 3.1**).

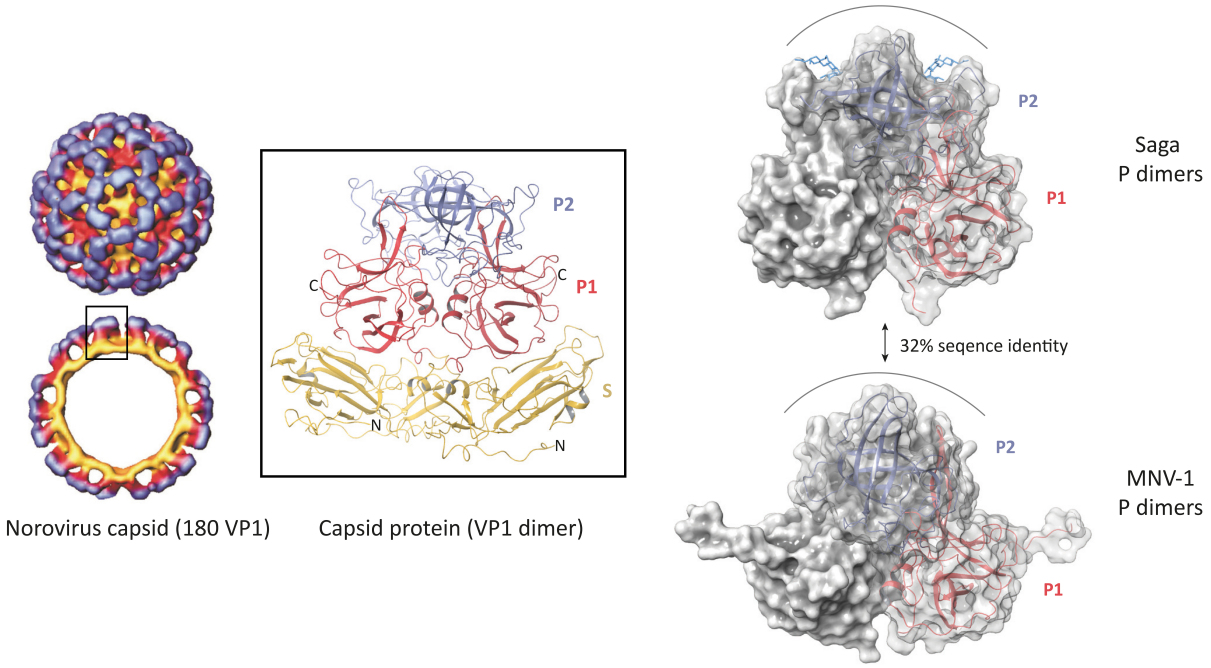


Figure 3.1 | The Norovirus capsid and crystal structures of P dimers of two different Genotypes
The Norovirus capsid comprises 90 dimers of the major capsid protein VP1 (PDB: 1IHM), which is divided into the shell (S) domain and the protruding (P) domain consisting of the subdomains P1 and P2. On the right side, the crystal structures of P dimers of the human NoV strain GII.4 Saga (PDB: 4WKZ) in complex with H type 2 trisaccharide and murine NoV strain GV.1 MNV-1.CW3 (PDB: 3LQ6) are depicted. The glycan binding site is located at the P2 subdomain (gray curved line). Albeit the amino acid sequence identity of Saga P dimers and MNV-1 P dimers is low (32%), the topological homology is relatively high (maximum root mean square deviation = 2.04 Å).

3.1.3 Current Drug Candidates against Noroviruses

The genetic and antigenic diversity of NoV and the rapid emergence of new strains hamper the development of effective vaccines or antiviral agents. Besides non-structural proteins that served as NoV drug targets, many researchers have focused on the major capsid protein VP1 to design antiviral agents that specifically block viral cell entry[225,232]. In the first approach ten years ago, a HTS library with 5000 compounds was screened against VA387 or Norwalk VLPs using saliva-based enzyme-linked immune assays, yielding 14 medium-sized compounds with effective doses in the low micromolar range[233]. Natural substances such as citrate[9], Tannic acid[234] present in Chinese herbs or human milk oligosaccharides[235] have also been identified as potential entry inhibitors.

Despite the high mutation rate within the subdomain P2 of the major capsid protein as result of immune evasion, the fucose binding pockets of GII.4 P dimers seem to be highly conserved across a broad range of NoV GII.4 strains[224]. Therefore, the 430-membered Maybridge library (termed CR2006 library throughout this study) was screened against Ast6139 VLPs employing STD and spin-lock filtered NMR experiments. Subsequently, competitive hits binding to the HBGA-binding site were utilized to design potent polymeric inhibitor, with an IC_{50} value of below 1 μM[90,218]. Inspired by this work, Zhang and co-workers targeted the P domain HBGA-site of the human NoV strain GII.4 VA387 by virtual screening of >2 million molecules generated from the ZINC database using a multistage docking protocol. Of 160 hit molecules, several compounds sharing a cyclopenta[a]dimethylphenanthren scaffold exhibited IC_{50} values below 10 μM according to saliva-based assays[236]. However, none of these inhibitors have reached clinical phases so far.

Monoclonal antibodies[237–239] or single-chain antibody fragments, so-called nanobodies[240,241], has been likewise proposed to counter NoV infections. Interestingly, some of these neutralizing nanobodies that inhibit the binding of GII.10 VLPs to HBGAs were found to interact with the P1 subdomain distant from the HBGA binding pockets[241]. This region comprises the so-called (N)ERK motif (N310, E316, R484, and K493 according to GII.4 numbering), a motif that is conserved among GII NoV and could function as conformational regulator through an allosteric effect[241,242]. Targeting this less variable region may be a promising approach towards virus inhibition.

3.2 Objectives

Due to the high variability of the P2 subdomain, in particular, residues surrounding the glycan binding site[224], the search for novel potential NoV inhibitors was extended towards secondary site binders.

- In order to complement existing compounds with binding affinity for NoV, the Maybridge fragment library was screened against NoV P dimers of two different genogroups (GII, Saga/M001 and GV, MNV-1) by employing saturation transfer difference NMR spectroscopy (STD NMR)[87] as initial read-out[81].
- Similar to the previous section, scaffold-group (SxG)-matrices and principal component analysis of molecular descriptors were applied to identify important structural or chemical properties.
- In order to find competitive or allosteric fragments, STD NMR-based competition assays were performed. Natural ligands that interact with the respective glycan binding sites were used as competitors.
- Following hit identification using STD NMR, a validation of binding fragments had to be conducted using complementary methods. Therefore, hits were subjected to surface plasmon resonance spectroscopy and protein-observed NMR methods.
- To confirm inhibition, MNV-1 infection assays were performed to guide the selection of leads, which interact with MNV-1 P dimers.
- Virtual screening/docking was employed to guide the interpretation of experimental data.

- In addition, important aspects of fragment library design are highlighted to reduce the rate of false positive hits.
- Moreover, NMR-based methods will be presented to identify so-called “bad actors” that may lead to promiscuous binding and/or unspecific inhibition.

3.3 Material and Methods

3.3.1 Production and Purification of Norovirus P Dimers

Unlabeled MNV-1 P Dimers, Saga P Dimers, and MI001 P Dimers

For STD NMR and SPR experiments, the protruding (P) domain of the capsid protein VP1 of murine Noroviruses (Isolate: MNV-1.CW1, GV/MNV1/2002/USA; GenBank: AY228235.2, K296E mutation) was used as screening target. Two P domains form stable dimers in aqueous solution. In total, three different batches of MNV-1 P dimer were used throughout the STD NMR screening and SPR analysis. The first MNV-1 P dimer batch was delivered by Henrik Wegener (Institute of Cell Biology and Virology, Prof. Dr. Stefan Taube). This sample was stored in 25 mM sodium phosphate buffer (pH 7.3) supplemented with 300 mM sodium chloride (NaCl) for several months at 4 °C. Due to aggregation problems of the protein as described later (section **3.4.2**), two additional protein samples were freshly purified from a frozen bacterial pellet delivered by Henrik Wegener before each application (i.e., STD NMR or SPR). Therefore, *E.coli* BL21(DE3) cells transformed by a pFL-His6-MOCR plasmid that comprises the sequence for the MNV-1 P domain (cf. Appendix A1) were collected from 1 L TB medium. Transformation of bacterial cells and protein expression essentially followed established protocols (cf. section **2.3.1**).

To purify the fusion protein containing the P domain equipped with a polyhistidine epitope tag (His-tag) and a solubility enhancing MOCR-tag (MOCR, monomeric overcome classical restriction), the bacterial pellet was suspended in ~180 ml ice-cold lysis buffer containing 25 mM sodium phosphate (pH 7.3) and 300 mM NaCl (first sample) or 150 mM NaCl (second sample), supplemented with one tablet of a protease inhibitor mix (cOmplete™ Protease Inhibitor Cocktail, Roche). If not otherwise stated, all buffer solutions were filtered using a 0.2 µm sterile filter, degassed and stored at 4 °C. Bacterial cells were lysed using mechanical pressure (precooled French Press, 1200 psi, two passes). Cell debris and unbroken cells were removed via centrifugation (30 000 x g, 60 min) utilizing an SA-600 fixed angle rotor along with a Sorvall RC 5B Plus centrifuge. The supernatant was filtered through a 0.45 µm Millipore filter and the lysate was directly applied on a 5 ml HisTrap FF column (GE, Healthcare Life Science; equilibrated with lysis buffer), employing a chromatography pump system (ÄKTA™ pure) at a flow rate of maximal 2 ml/min. The bound His-tagged protein was eluted by an imidazole gradient of 0 - 300 mM imidazole in lysis buffer. Fractions containing the fusion protein were concentrated to 15 ml using an Amicon ultra centrifugal filter (10 kDa cut-off, Millipore).

In order to remove the His-tag and MOCR-tag, the fusion protein was cleaved by utilizing a TEV (Tobacco Etch Virus) protease, which was provided by Henrik Wegener. Therefore, cleavage was performed overnight at 4 °C in 15 ml elution buffer supplemented with 3% (v/v) glycerol and 2 mM dithiothreitol (DTT) at a molar protease-protein ratio of 1:20. During incubation, the solution was dialyzed using a 10 kDa dialyze cassette (Slide-A-Lyzer™ dialysis cassette, Thermo Fisher Scientific) against 2 L lysis buffer containing 0.3 mM DTT to remove residual imidazole. Next day,

the dialysis buffer was replaced by freshly prepared lysis buffer without DTT. After 4 h incubation at 4 °C, the His-tagged TEV protease was removed by metal-affinity chromatography. The flow-through containing MNV-1 P dimer was concentrated to ~10 ml and purified via size exclusion chromatography (SEC) (HiLoad® 26/600 Superdex® 75 prep grade, GE Healthcare). Therefore, the SEC column was washed with 310 ml Millipore water and then equilibrated with 310 ml lysis buffer (flow rate: 1.5 ml/min). Subsequently, the protein solution was injected onto the SEC column, and pure MNV-1 P dimer was eluted with lysis buffer at a flow rate of 0.8 ml/min. For NMR experiments, the buffer was exchanged several times against NMR buffer (25 mM sodium phosphate in D2O, pH* 7.05) by centrifugation at 2000 x *g* using an Amicon ultra centrifugal filter (10 kDa cut-off, Millipore). The process of purification of MNV-1 P dimer was monitored via SDS PAGE analysis (for details see section **1.3.1**).

Protein samples containing MI001 P dimer (Hu/GII.4/MI001/2011/US GenBank: KC631814.1) or Saga P dimer (details below) were provided by Henrik Wegener and Dr. Alvaro Mallagaray, respectively. The protein biosynthesis of these proteins is described elsewhere[10,243,244]. The plasmid vectors and amino acid sequences of Norovirus P dimers are shown in Appendix A1. After buffer exchange, all protein samples were stored in NMR buffer (D_2O) at 4 °C.

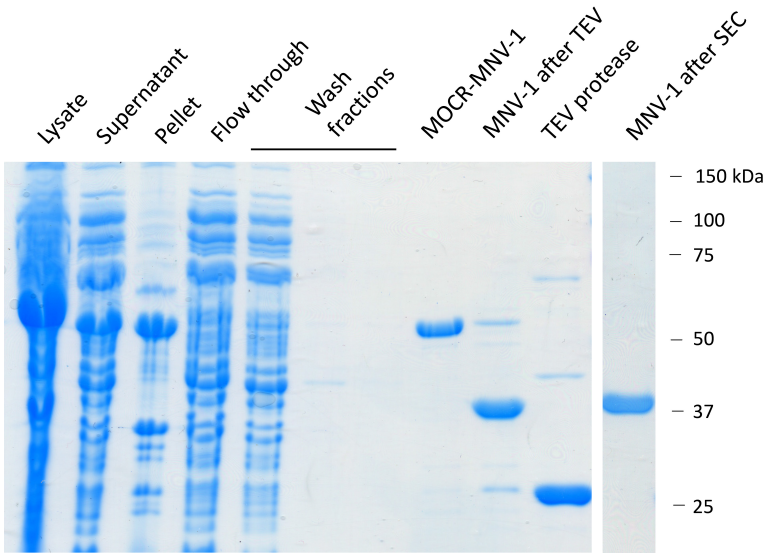


Figure 3.2 | Biosynthesis and purification of MNV-1 P dimers
After each purification step, 80 µl protein samples were kept until analysis via SDS PAGE at -20 °C. The lanes in the Coomassie blue stained SDS gel (10% bisacrylamide/acrylamide 30:1) represents the lysate after French press application, supernatant and corresponding pellet after centrifugation, flow through and wash fractions of metal-affinity chromatography, fusion protein His-MOCR-MNV-1, cleaved P domain of MNV-1, TEV protease used for cleavage, and purified P domain of MNV-1 after size exclusion chromatography (SEC).

Uniformly [^{2}H,^{15}N]-labeled Saga P Dimers

For 2D protein NMR experiments, uniformly [^{2}H,^{15}N]-labeled P dimers were required. Therefore, an optimized protocol for the expression of Saga P dimers in minimal medium (M9/D_2O) was used[10,93]. Transformed BL21(DE3) cells comprising a vector pCWΔlac-MBP-2xHis, which contains the sequence of the protruding domain (P domain) of Norovirus strain GII.4 Saga (Norovirus Hu/GII-4/Saga1/2006/JP, GenBank: AB447456.2), were provided by Dr. Alvaro Mallagaray as glycerol stock stored at -70 °C. In this study, a truncated P domain was utilized missing the last 10 amino acids (GNGTGRRRAL) at the C-terminus.

The exact composition of M9/D_2O is shown in **Table 3.1** (see below). Basically, the media resembles a 100 mM phosphate buffer in D_2O containing essential vitamins, magnesium sulfate along with other salts, ^{15}N-labeled ammonium sulfate as a nitrogen source, and perdeuterated glucose as carbon source. Prior to protein expression, 40 ml LB^{Amp} medium were inoculated with a bacterial glycerol stock. If not otherwise stated, bacterial cells were cultured at 37 °C in a shaking incubator (200 rpm). Bacterial cell growth was monitored by measurement of the optical density at a wavelength of 600 nm (OD_{600}) employing a Jasco V550 photometer (cf. **Figure 3.3**). Following incubation overnight, 350 µl of a cell suspension was collected via centrifugation at RT for 5 min at 1200 x *g*. The bacterial pellet was resuspended in 40 ml M9 to gain an OD_{600} of 0.1 units. Following 4.5 h incubation, the cell suspension (OD_{600} = 0.3) of 33 ml was centrifuged at RT for 15 min at 1200 x *g*. The pellet was again resuspended in 100 ml M9 (OD_{600} ~ 0.1). After 2.5 h incubation, once an OD_{600} of 0.4 was reached, the volume of cell suspension was increased to 200 ml M9, and 1.5 h later to 1 L M9. Prior to induction of protein overexpression at OD_{600} of 0.9 by addition of 1 mM IPTG, the bacterial cells were cooled down at 6 °C for 15 min. After IPTG induction, the cell suspension was distributed into four 2 L incubation flasks covered with aluminum foil to protect vitamins from light exposure. After ~56 h at 16 °C while shaking (220 rpm), the bacterial cells were harvested by centrifugation at 5000 x *g* for 30 min at 4 °C. The pellet was stored at -70 °C.

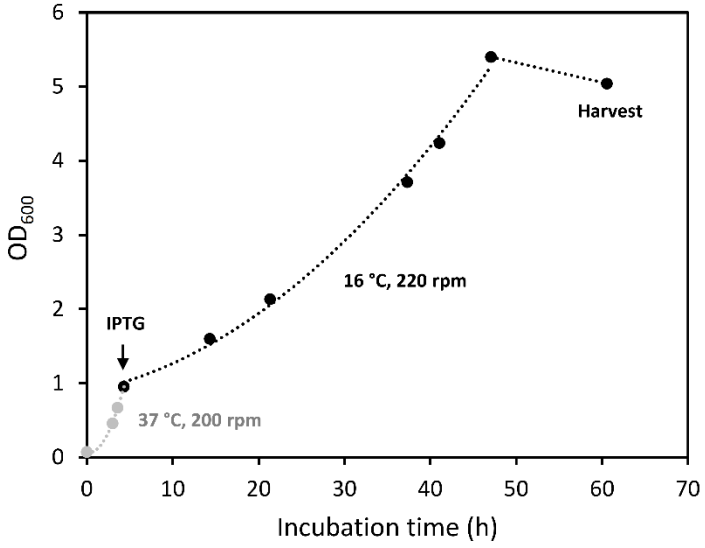


Figure 3.3 | Growth curve of E.coli strain BL21 (DE3) expressing the P domain of humane Norovirus strain GII.4 Saga
The bacterial growth was monitored of a 1 L cell suspension by measuring the optical density at a wavelength of 600 nm (OD_{600}) using a Jasco V550 photometer. After IPTG induction, cells were incubated at 16 °C to promote proper protein expression.

Purification and cleavage of Saga P dimers were similar to the isolation of MNV-1 P dimers with following alterations: The frozen bacterial pellet was resuspended in 40 ml of ice-cold PBS buffer containing 0.1 µg/ml aprotinin, 0.1 µg/ml leupeptin, 6 µg/ml lysozyme, and 60 mU/ml benzonase. Cell lysis was conducted using mechanical pressure (precooled French Press, 1200 psi, two passes). Next, the lysate suspension was centrifuged at 125 000 x *g* for 1 h at 4 °C. The supernatant containing soluble His-tagged fusion proteins was filtered using a 0.22 µm filter device. Prior to the addition of about 20 ml nickel-charged agarose beads (Ni-beads as a 50% suspension), the Ni-beads were washed with 10 mM imidazole in PBS several times. The suspension was rotated for 1 h at 4 °C and poured into an empty column. After collecting the flow through by gravity chromatography, the Ni-beads were successively washed with 35 ml PBS buffer containing 10 mM, 20 mM, and 50 mM imidazole, respectively. The fusion protein was eluted by adding 40 ml of 250 mM imidazole in PBS buffer. After dialysis against PBS containing 10 mM imidazole, the fusion protein was cleaved using the HRV (human rhinovirus) 3C protease at a protease-protein ratio of 1:100 for 16 h at 4 °C in the presence of 40 mM DTT. The HRV 3C protease was provided by Henrik Wegener. The MBP and the His-tagged HRV 3C protease were removed using Ni-beads, and the cleaved P domain was harvested in the flow through. Further purification was performed using size exclusion chromatography (see above) with 20 mM sodium phosphate buffer (pH 7.3). Fractions containing Saga P domain were collected, pooled, and concentrated using a 10 kDa cut-off filter device (Amicon ultra centrifugal filter, Millipore). All steps of the biosynthesis of Saga P dimers were monitored by SDS PAGE analysis (**Figure 3.4**).

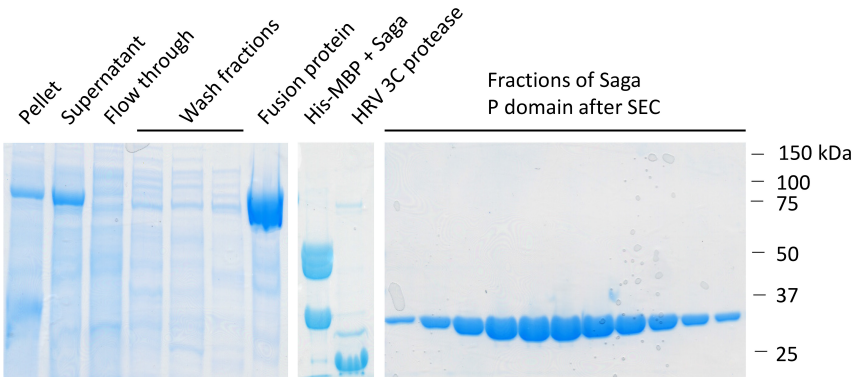


Figure 3.4 | Biosynthesis and purification of Saga P dimers
Protein expression was conducted in *E.coli* BL21 (DE3) cells cultured in minimal medium. After cell lysis, cell debris (pellet) were removed by ultracentrifugation. The supernatant was passed through a metal-affinity chromatography column to collect the fusion protein comprising two His-tags, a maltose binding protein (MBP), and the P domain of the human Norovirus strain GII.4 Saga. After cleavage by utilizing the 3C protease of human Rhinoviruses (HRV), the His-tagged MBP and other impurities were removed using metal-affinity chromatography and size exclusion chromatography (SEC).

Table 3.1 | Composition of the minimal medium M9/D_2O
All Substances were dissolved in D_2O. The final concentration is listed beside each ingredient. To promote hydrogen-deuterium exchange, D_2O samples containing growth substances, salts, trace minerals, and ampicillin were lyophilized twice in D_2O and finally dissolved in D_2O. Prior to inoculation, the minimal medium was sterile-filtered.

Ingredient	Final concentration	Comments
Na_2HPO_4	73.3 mM	Growth substances were prepared based on an optimized protocol (established by Robert Creutznacher, 2017) * equal to ~2g/L
KH_2PO_4	26.6 mM	
$MgSO_4$	4 mM	
$^{15}NH_4Cl$	35 mM*	
Deuterated glucose-D_7	3 g/L	
$CaCl_2$	100 µM	Salts were prepared as 1000-fold stock solution and stored at -20 °C.
$MgCl_2$	1 mM	
Thiamine (B1)	20 µg/ml	B1 was prepared as 20 mg/ml solution.
Riboflavin (B2)	0.1 µg/ml	All other vitamins were prepared as a 100-fold stock solution. Therefore, the powder was weighed and directly dissolved in D_2O. Due to degradation upon light exposure, the stock solution was always kept in the dark. The stock solution was stored at -20 °C.
Nicotinamide (B3)	1 µg/ml	
Choline chloride (B4)	1 µg/ml	
D-Pantothenate (B5)	1 µg/ml	
Pyridoxal hydrochloride (B6)	1 µg/ml	
D-biotin (B7/B8/H)	1 µg/ml	
Folate (B9)	1 µg/ml	
Cobalamin (B12)	1 µg/ml	
Myo-inositol	2µg/ml	
Ampicillin	100 µg/ml	Unstable at 37 °C (half-life <24 h)[170]
Eco E.coli-OD2 DN	24%	Silantes ancillary medium
D_2O (99.85%)	Ad final volume	

3.3.2 Protein Analytics – Dynamic Light Scattering

It is known that depending on the design of constructs P dimer preparations may contain varying amounts of monomers[10]. Moreover, it has been reported that P dimers from human Norovirus strain VA387 (GII.4) can form different types of oligomers, also referred to as P particles[245–247]. A cysteine-containing peptide fused to either C-termini of the P domain stabilizes oligomerization through inter-subunit disulfide bonding[248]. However, the macromolecular assembly has also been observed even without C-terminal modification[249]. In contrast, a deletion of the C-terminal arginine cluster prevents P particle formation[250].

In order to judge the integrity of both human and murine P dimers, dynamic light scattering (DLS) analyses under different conditions were conducted. Prior to measurement, protein samples containing ~100 µM P dimers were centrifuged at 17 000 x *g* for 5 min at RT. A black quartz cuvette was filled with 20 µl protein sample. After 1-2 min equilibration, the analysis was performed at RT, and 10 scans of 20 sec duration were accumulated per sample employing a SpectroSize™ 300 device (XtalConcepts). For each condition, measurements were repeated at least two times to confirm reproducibility.

3.3.3 Lead Discovery based on NMR Spectroscopy

In the search for lead structures for novel NoV entry inhibitors, the fragment libraries CR2006 and TS2017 were screened, employing STD NMR spectroscopy. In principle, STD NMR experiments were performed as described in the previous chapter (cf. section **2.3.7**). First, the purity and solubility of novel fragments were analyzed by acquiring ^{1}H NMR spectra. Soluble fragments were distributed *in silico* in smaller compound mixtures and screened against Norovirus protruding (P) domain of one murine strain (MNV-1) and two human strains (Saga and MI001). Following hit identification using STD NMR, positive hits binding to MNV-1 were subjected to surface plasmon resonance (SPR) experiments to classify potent fragments. In order to predict putative binding regions of both murine and human P dimers, STD NMR-based competition assays were conducted. The minimal motifs of putative natural ligands severed as competitors, namely 3'-sialyllactose and methyl α-L-fucopyranoside for MNV-1 and Saga, respectively. Promising candidates were further analyzed with respect to their inhibitory effect against MNV-1 virions. For Saga P dimers, protein-observed experiments employing isotopically U-[^{2}H,^{15}N]-labeled P dimers were performed to explore putative binding sites. This was further substantiated by virtual docking using Glide implemented in Maestro 11.

Fragment Screening based on STD NMR

Prior to fragment screening, the solubility of fragments was tested by observation and standard 1D NMR. Subsequently, the fragments were allocated into fragment mixtures. For this screening, the python script *makeBinsforScreening_2017.py* was written to guarantee that at least one NMR signal per fragment is isolated within a certain range (e.g. 0.05 ppm). In the former python script *poolCompounds_2014.py*, signal overlapping was still an issue in some cases, when combining more than ten fragments. Moreover, a new method (noPeakOverlapp) has been implemented to allow the identification of all NMR signals per fragment within a fragment mixture. If desired, this method has to be set manually. The stringent rule that fragments are classified as electrophiles and nucleophiles etc. according to the SMILES code definitions of Hann et al. (cf. REF[90,182]) was also modified. To prevent chemical reactions, amines and esters/acids encoded by "NH" or "COO", respectively, are not allowed to be in the same mixture. The chemical groups are listed in the database stored as excel-file (see section **2.3.4**). Fragment mixtures are listed in Appendix A2.

[90]STD NMR-based screening was conducted in 180 µl NMR buffer II (25 mM sodium phosphate buffer/ D_2O, pH* ~ 7.05, and 100 µM DSS-d_6). Each sample contained ~3 µM P dimer (MNV-1, Saga, or MI001) and 9 to 20 fragments. The concentration was 125 µM for a single fragment. In the case of MNV-1 P dimers and MI001 P dimers, only members from the Maybridge library CR2006 were screened, yielding 28 fragment mixtures. In order to increase the number of hits for Saga P dimers, the Maybridge library TS2017 was also screened.

The following acquisition parameters were applied: 32 k data points, 1024 transients, 2.7 s acquisition time, 10 ms spinlock filter, 6 s relaxation delay, and 2.5 s saturation time at a power level of 46 dB (0.25 µW). This corresponds to a 471° flip angle and $\gamma B1(max)/2\pi$ = 63.6 Hz. The on- and off-resonance frequencies were set at -1.8 ppm and 80 ppm, respectively. To split the STD

pseudo-2D datasets into a reference spectrum and difference spectrum, the command *stdproc_2* was used (for AU program see Appendix A5). Following phase correction and referencing using TSP-d_4 (command: *sref* or *.cal*), an exponential window function with line broadening of 2 Hz (command: *lb 2; efp*) was applied, for difference spectra, the phases (0^{th} and 1^{st} order correction) were copied from the according to reference spectrum, and Fourier transformation was conducted as described above.
Binding fragments were identified visually using 1D spectra of single fragments as a reference (see **Figure 3.6**). Moreover, the STD effect for each isolated NMR signal was determined by scaling the reference spectrum manually until it matches the signal height in the difference spectrum. The STD effect is defined as:

$$STD_{effect} = \frac{I_0 - I_{sat}}{I_0} \quad \text{(VII)}$$

with I_0 being the intensity of a proton of a fragment in a spectrum without saturation and I_{sat} being the intensity of the same proton of a spectrum acquired with saturation. When direct irradiation occurred (cf. section **2.3.7**), only fragments with positive STD effect, after subtracting the STD effect of the negative control, were considered as hits.

Competition Assays using STD NMR

In order to narrow down the number of attractive fragments, the competition between positive hits from the STD NMR screen and a natural binding motif of P dimers were analyzed. In the case of MNV-1 P dimers, 3'-sialyllactose served as a competitor. In total, 12 mixtures with only three to seven fragments were screened to guarantee that at least two isolated signals of a fragment could be unambiguously monitored. First, STD NMR spectra of samples containing fragments at a concentration of 200 µM each mixed with 5 µM MNV-1 P dimers were acquired in NMR buffer II (see above). Next, the same samples were measured again in the presence of 8 mM 3'-sialyllactose to monitor the reduction of STD effect for each fragment.
Methyl α-L-fucopyranoside has been identified to be the natural, minimal binding motif of human Norovirus P dimers (cf. REF[90] and REF[10]). Therefore, methyl α-L-fucopyranoside was also utilized in this study employing Saga P dimers as a molecular target. The Competition assays were conducted as described above. The concentration of methyl α-L-fucopyranoside was 8 mM and 32 mM for CR2006 fragments and TS2017 fragments, respectively. Acquisition parameters for STD NMR experiments are described above.

Chemical Shift Mapping of U-[^{2}H,^{15}N]-labeled Saga P Dimers

In order to find binding sites for 29 positives hits from the STD NMR competition screen, uniformly [^{2}H,^{15}N]-labeled Saga P dimers were subjected to chemical shift perturbation (CSP) experiments. Moreover, six fragments from the virtual screening (**CR025**, **CR070**, **CR055**, **CR365**, **TS188**, and **TS189**; see below) exhibiting high docking scores were tested as these compounds may bind near the fucose binding site. In the first step, both a 1D NMR spectrum and a ^{1}H,^{15}N-TROSY-HSQC reference spectrum of 60 µM Saga P dimers in 160 µl CSP buffer (75 mM sodium phosphate, pH 7.3, 8% (v/v) D_2O, 100 mM NaCl, 300 µM imidazole, and 100 µM DSS-d_6) were acquired.
Next, the spectral perturbation of isotopically labeled Saga P dimers was observed upon fragment binding. Therefore, fragments were weighed and dissolved in CSP buffer (see above) to yield 12 – 20 mM stock solutions. The pH was adjusted when required. In a next step, 1D NMR spectra and ^{1}H,^{15}N-TROSY-HSQC spectra of samples containing 60 µM Saga P dimers mixed with a single fragment stock solution were recorded. Since some fragments were less soluble in water, undissolved remains were removed via centrifugation (17 000 x *g*, 5 min) prior to measurement. The concentration of fragments was determined by integration of ^{1}H NMR signals, using DSS-d_6 as a reference. To confirm pH invariability, imidazole served as pH reporter as the ^{1}H signals of imidazole are highly sensitive to pH variation. A pH titration with imidazole was conducted by Christoph Müller-Hermes (2017, unpublished data) and is presented in Appendix A7 (**Figure A23**). Using the Boltzmann model implemented in Origin 7.0, data fitting yielded a sigmoidal function:

$$\delta = 7.76045 + \frac{8.68989 - 7.76045}{1+\exp\left(\frac{pH-7.40974}{0.4738}\right)} \quad \text{(VIII)}$$

$$pH = 0.4738 \cdot \ln\left(\frac{8.68989 - 7.76045}{\delta - 7.76045} - 1\right) + 7.40974 \quad \text{(IX)}$$

with δ being the chemical shift (in ppm) of one ^{1}H NMR signal of imidazole and *pH* being the calculated pH of a NMR sample. Proton NMR spectra were acquired with water suppression employing the pulse program *zgesgp*. The parameters were set as follows: 16 scans, 1.8 s acquisition time, 22 k data points, 6 s relaxation delay, and 12 ppm spectral width. TROSY-HSQC experiments were conducted using the pulse program *trosyeft3gpsi.2* with setting as follows: 24 scans, P1 of ~9.24 µs (calibrated), and 1.5 s relaxation delay. The following parameters were set for the indirect (^{1}H) and direct dimension (^{15}N), respectively: O1P of 4.69 ppm and 117.50 ppm, 16 ppm and 35 ppm sweep width, and 2048 and 430 data points, yielding a FID resolution of 7.8 Hz and 8.84 Hz. Data were analyzed using the software CcpNmr Analysis 2.4.2. The Euclidean distances of chemical shift differences were calculated employing **Equation (VI)** (cf. section **2.3.7**). Finally, chemical shifts were mapped on the crystal of Saga P dimers using the Python script *python2Maestro_plotCSP.py*. It has to be noted that for the first five NMR samples, a protein stock kindly provided by Robert Creutznacher was used. To reuse U-[^{2}H,^{15}N]-labeled Saga P dimers, 5 – 16 NMR samples were combined, and fragments and other substances were removed by a 6-fold buffer exchange using a 10 kDa cut-off filter device. The integrity of recycled Saga P dimers was confirmed via acquiring ^{1}H,^{15}N-TROSY-HSQC reference spectra.

T_2 Determination using Spin-echo NMR Experiments

Determining the transverse relaxation time (T_2) of amide protons of the backbone using 1-1 echo 1D NMR experiments[251] allows for a rough estimation of the molecular weight of a protein of interest[252]. The NMR buffer was exchanged prior to each measurement. Several spectra of 100 µM unlabeled Norovirus P dimers (MNV-1 and Saga) in 25 mM sodium phosphate 90%H_2O/10%D_2O (pH 7.3) were recorded with several delays Δ (0.1, 0.2, 0.4, 0.8, 1.6, and 3 ms). The ratio of intensities in the amide region of the spectrum was determined by manual scaling for four different pairs (e.g., Δ_A = 0.1 ms and Δ_B = 3 ms). The delay differences multiplied by two were plotted against the natural logarithm of each intensity ratio. The slope of the linear regression line represents T_2 (in ms) in analogy to **Equation (X)**:

$$T_2 = \frac{2(\Delta_A - \Delta_B)}{ln(\frac{I_A}{I_B})} \quad \text{(X)}$$

with I being the intensity of the respective delays Δ. The molecular weight (in kDa) was estimated based on the rotational correlation time τ_c (in ns) using the following relations:

$$\tau_c\,[ns] \approx \frac{1}{5T_2\,[s]} \quad \text{(XI)}$$

$$MW\,[kDa] \approx 2 \cdot \tau_c\,[ns] \quad at\ 20\,°C \quad \text{(XII)}$$

with τ_c being the time taken for the molecule to rotate by roughly one radian about an axis.

3.3.4 Surface Plasmon Resonance Spectroscopy

Prior to competition assays with MNV-1 P dimers, positive hits from the STD NMR screening were validated using SPR spectroscopy. All experiments were performed on a Biacore 3000 instrument at RT (GE Healthcare) following general instrument handling stated in the Biacore 3000 handbook. All buffer solutions were sterile-filtered and degassed before use.

Adsorption to the Chip Matrix

First, the Biacore methods *sanitize* and *desorb* were conducted according to the Biacore manual instruction and the system was flushed with Millipore water for at least two hours. Next, a Biacore sensor chip CM5 (GE Healthcare) providing a 100 nm carboxy-methylated dextran matrix was docked and primed with running buffer (25 mM sodium phosphate, pH 7.3, 10 mM NaCl, and 0.01% sodium azide). To confirm that the negatively charged matrix adsorbs MNV-1 P dimers, as this is a requirement for amine coupling (see below), freshly prepared protein solutions were diluted with 10 mM sodium acetate at varying pH values (4 – 4.3) to yield a final concentration of

500 µg/ml. Sensograms were acquired by injection of 10 µl protein solution at a flow rate of 5 µl/min. This particular pH range was tested because the theoretical isoelectric point (pI) of MNV-1 P dimers is calculated as 4.67 pH units according to the *ExPASy ProtParam* web tool[52]. The net charge of MNV-1 P dimers is, hence, positive at pH values lower than the pI.
In a second optimization step, an appropriate protein concentration of protein stock was determined. Therefore, samples with increasing amount of protein (50 – 500 µg/ml) were diluted with 10 mM sodium acetate buffer to yield a final pH of 4.1 units. Finally, 10 µl of these protein samples were injected at a flow rate of 5 µl/min, and the maximal Biacore response was monitored for each protein concentration. The results can be found in Appendix A6 (**Figure A12**).

Direct Immobilization of MNV-1 P Dimers using Amine Coupling

The chip surface was normalized using 70% glycerol as described in the Biacore user manual. Following equilibration with running buffer, the carboxy-methylated surface was activated by injection of 120 µl of a freshly prepared mixture containing 0.05 M N-hydroxysuccinimide (NHS) and 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) at a flow rate of 10 µl/ min. In order to immobilize the protein, a solution of freshly purified MNV-1 P dimers (~7 mg/ml) were diluted with 10 mM sodium acetate to reach a final concentration of ~400 µg/ml and a final pH of 4.1. Immobilization was conducted at a flow rate of 5 µl/min with several injections of 5 – 15 µl until intended response level was reached (~6500 RU). Remaining activated groups were deactivated by injection of 120 µl of 1 M ethanolamine (pH 8.5) at a flow rate of 10 µl/min. An exemplary SPR sensorgram can be found in Appendix A6 (see **Figure A13)**. The reference flow cell was only activated and deactivated.

Validation of Positive Hits from STD NMR Screening

The integrity of immobilized MNV-1 P dimers was verified by injection (mode: *QUICKINJECT*) of 3'-sialyllactose dissolved in running buffer (pH adjusted) at increasing concentrations (3 – 20 mM). After 2 min injections at a flow rate of 10 µl/min, the maximal RU levels (cf. **Figure A14**) at the end of injection were monitored and plotted against the injected analyte concentration. The data were fitted using non-linear regression assuming a model that superimposes one-site binding as a primary binding event with one subsequent cooperative binding step. Therefore, equation (XIII) was implemented in Origin 8:

$$RU = \frac{RU_{max,1} \cdot [L]}{K_{D1} + [L]} + \frac{RU_{max,2} \cdot [L]^h}{{K_{D2}}^h + [L]^h} \qquad \text{(XIII)}$$

with K_{D1} and K_{D2} being the microscopic dissociation constants, RU_{max} being the maximum change of the response for each binding event, and h being the Hill coefficient.
In the next step, titration experiments of two fragments (**CR273** and **CR432**) were conducted to estimate a suitable fragment concentration for the SPR screen. Fragments were weighed and dissolved in running buffer to reach a final concentration of 1 mM. Serial dilution gave fragment

concentrations ranging from 20 to 300 µM. The difference SPR responses at the end of 10 µl injections (flow rate: 10 µl/min) were plotted against increasing fragment concentrations yielding Langmuir binding isotherms:

$$RU = RU_{max} \cdot \frac{[L]}{K_D + [L]} \quad \text{(XIV)}$$

with *[L]* being the ligand concentration and with K_D being the dissociation constant. Data were fitted using non-linear regression implemented in Origin 7.0.
For the classification of fragments with sufficient binding affinity to MNV-1 P dimers, 168 fragments were screened at a concentration of 200 µM in running buffer (see above). Therefore, sensorgrams were recorded after injection of 10 µl of the sample (flow rate of 10 µl/min). A Biacore method was used to automate SPR measurements. The flow cells were washed with running buffer after each measurement for 2 min. An exemplary Biacore method can be found in Appendix A6. In order to judge the robustness of the binding assay, a reporter fragment (**CR432**) was injected every 10 to 15 fragments.

3.3.5 MNV-1 Inhibition Assay

Particular fragments whose binding to MNV-1 P dimers had been influenced upon binding of 3'-sialyllactose according to STD NMR experiments and several analogs from the TS2017 library were further analyzed using functional virus inhibition assays. Depending on the solubility, solutions containing 3 to 12 mM fragment were prepared in PBS buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, and 137 mM NaCl; pH 7.4). The pH was adjusted when required, except for **CR385** as the solubility decreases dramatically above pH 6.8. In a 96-well format, approximately $1 \cdot 10^5$ infectious murine Noroviruses (strain: MNV-1.CW3, Genbank: EF014462) diluted in 300 µl PBS buffer were incubated in the absence and presence of 0.5 to 2 mM fragment for 1 h at 37 °C. The virus samples were stored at -80 °C until measurement of the virus titer by endpoint dilution.
Determination of $TCID_{50}$ values, which represent the median of tissue culture infective dose, was conducted in 96-well tissue culture plates by simple observation of the percentage of cell death for each virus dilution. Therefore, ~$1 \cdot 10^4$ BV2 cells (murine microglial cells[253]) were seeded in each well containing 160 µl DMEM medium supplemented with penicillin/streptomycin (50 000 U), 2 mM L-glutamine, 0.1 mM non-essential amino acids, and 5% fetal calf serum. Next, 40 µl of each virus sample was transferred into four wells, resulting in an initial 1:5 dilution. Using a multi-channel pipette, a volume of 40 µl per well was transferred to the next row, leading to 8 rows with serial 1:5 dilutions. For the last row, a volume of 40 µl per well was discarded. In total, three fragments per 96-well plate could be analyzed in parallel. Subsequently, BV2 cells were incubated for 3 to 5 days at 37 °C, 5%, and 96% relative humidity. The virus stock and BV2 cells were kindly provided by Henrik Wegener.

Finally, all wells were classified based on developing a cytopathic effect judged by the color of the cell medium. Reddish medium indicates reduced cell proliferation and, hence, infectious cells (positive wells). The $TCID_{50}$ per ml was calculated based on the Reed-Muench-algorithm[254,255]. The formula has been simplified based on logarithmic identities:

$$TCID_{50}/ml = \frac{DF^{\Delta ratios}}{V \cdot D} \quad \text{(XV)}$$

$$\Delta ratios = \frac{R_{row\ above\ 0.5} - 0.5}{R_{row\ above\ 0.5} - R_{next\ row}} \quad \text{(XVI)}$$

$$R_{row\ above\ 0.5} = \frac{\sum_{row\ at\ D+} (wells_{pos})}{\sum_{row\ at\ D} (wells_{pos}) + \sum_{row\ at\ D} (wells_{neg})} \quad \text{(XVII)}$$

$$R_{next\ row} = \frac{\sum_{row+1\ at\ D} (wells_{pos})}{\sum_{row+1\ at\ D} (wells_{pos}) + \sum_{row+1\ at\ D} (wells_{neg})} \quad \text{(XVIII)}$$

with *DF* being the dilution factor (e.g., 5), *V* being the volume in ml per well, *D* being the dilution showing a mortality greater or equal 50% (e.g., $1/5^6$), and $R_{row\ above\ 0.5}$ being the ratio of cumulative positive wells at dilution *D*. Positive wells and negative wells are included at higher dilutions and lower dilutions, respectively. $R_{next\ row}$ is equal to $R_{row\ above\ 0.5}$ but is calculated for the next dilution (e.g., $1/5^7$). The $TCID_{50}$ calculator draft was created by Shi-Hsia Hwa (2014) and is available on request.

To validate inhibition, and to estimate IC_{50} values for **CR054**, **CR120**, **CR196**, **CR294**, and **TS039**, end point dilution assays were repeated at varying fragment concentrations. These measurements were done in triplicate if not otherwise stated. Moreover, WST1-based colorimetric assays (Roche) were performed in a 96-well plate format following manufacturer's instructions to evaluate cell viability when treated with fragments. Approximately $1 \cdot 10^4$ BV2 cells per well were incubated in 100 µl DMEM medium (see above) for 24 h using a cell incubator at 37 °C in the presence of the respective fragment at increasing concentration. Medium without cells served as reference. The number of cells was estimated using a hemocytometer (referred to as Neubauer counting chamber). The determination of cell activity involves the addition of 10 µl of WST-1 reagent (stored at -20 °C) followed by detection of absorbance of formazan, which is the product of WST-1 cleavage by the mitochondrial succinate-tetrazolium-reductase system. The absorbance was measured at 450 nm after 10 min incubation at 37 °C using a microplate photometer. The reference wavelength was set at 645 nm. The measurements were done in duplicate. The median of lethal concentration LC_{50} was estimated by data fitting using the Boltzmann model implemented in Origin 7.0:

$$\alpha = \alpha_{min} + \frac{\alpha_{max} - \alpha_{min}}{1 + \exp\left(\frac{c - LC_{50}}{dc}\right)} \quad \text{(XIX)}$$

with α being the relative cell activity and *c* being the fragment concentration.

3.3.6 Virtual Screening and Molecular Docking

In order to evaluate NMR-based binding studies, including STD NMR screening and chemical shift perturbation (CSP) experiments, a virtual screening was conducted using Glide as implemented with Maestro 11 (2017; Schrödinger Inc.) as this is described in section **2.3.8**.
Virtual screening of the fragment libraries CR2006 and TS2017 involved four steps: (1) Chemical structures were transferred from the excel-file to Maestro as sdf-file. Next, the LigPrep tool was utilized in Maestro to obtain minimized 3D structures using the optimized OPLS3 force field[256]. This step includes generating possible protonation states at pH 7 and creating tautomers, theoretically yielding up to 32 combinations.
(2) The protein preparation wizard transforms crystal structure data of NoV P dimers (Saga: PDB 4WZK and MNV-1: PDB 3LQ6) into an energy-minimized version based on the OPLS3 force field. During this process, bond orders are assigned, water molecules are removed beyond 5 Å and with less than three H-bonds to "non-waters". Moreover, missing side chains and missing protons are added. In the case of Saga P dimers, the side chains of Glu 372, Gln 523, Asn 310, and Glu 315 were automatically included. Moreover, the most populated orientation of alternative side chains had been chosen for Arg 287 and Arg 484. Finally, protonation states of the residues Asp, Glu, and His were modified to optimize the hydrogen bond network using the PROPKA program[187,188], which is implemented in Maestro (cf. section **2.3.8**).
(3) Receptor grids were generated to cover three different binding sites of Saga P dimers that had been identified during CSP experiments. The positions and sizes of respective grids can be found in Appendix A8 (**Figure A43** and **Figure A44**; **Table A7** and **Table A8**). In the case of MNV-1 P dimers, the inner core was omitted by the use of spherical exclusion volumes.
(4) By using the virtual screening workflow, several docking sites can be screened in parallel. In the first screening round, all fragments from the Maybridge libraries were docked into three rigid receptor grids (for Saga P dimers: grid 1, 2a, and 3) accounting for almost the complete protein surface. Docking was performed with full ligand flexibility. Up to three poses per compound state were generated. Finally, the Glide extra precision scoring function[103,186] was applied to rank docking poses after post-docking minimization. Based on the docking score, the best 30% of all fragments had been kept for further statistical analysis (see above).
Furthermore, fragments from CSP experiments, using ^{15}N-labeled Saga P dimers, and fragments from MNV-1 inhibition assays were individually docked into four binding sites present in Saga P dimers (1,2a/b, and 3) and into three binding sites present in MNV-1 P dimers, respectively. All good scoring states were retained. This step was repeated for MNV-1 P dimers
Several scores were used to judge the quality of binding pocket and pose prediction: The *docking score* and the *GlideScore* are the same, except for compounds exhibiting several protonation states at a given pH. Both scores are empirical scoring functions that approximate the ligand binding free energies. The third measure is the *Emodel*, which determines the best pose for a ligand. It has a more significant weighting of the noncovalent force field components, including electrostatic and van der Waals energies.

3.4 Results

3.4.1 Fragment Screening

After quality control of fragments dissolved in NMR buffer, mixtures of 10 to 20 fragments were prepared and screened against NoV P dimers by the use of STD NMR spectroscopy. In this study, P dimers from different NoV strains, including MNV-1, Saga, and MI001, were utilized (for details see section **3.3.1**). An overview of the screening process is shown in **Figure 3.5**. Due to optimized NMR STD settings, positive hits were readily identified by visual inspection of STD NMR spectra and 1D proton reference spectra. One example of a fragment mixture in the presence and absence of 5 µM MNV-1 P dimers is given in **Figure 3.6**. In the first screening round, protein samples of all three different NoV P dimers were kindly provided by Henrik Wegener and Dr. Alvaro Mallagaray. These protein stocks were stored at 4 °C for several weeks/months in 20 – 25 mM phosphate buffer (pH 7.3) in the presence (MNV-1/MI001) or absence (Saga) of 300 mM sodium chloride (NaCl). The purity and integrity of P domains were confirmed via SDS PAGE prior to fragment screening. Unexpectedly, in the case of anticipated MNV-1 P dimers, 297 positive hits were identified, which showed saturation transfer-mediated signal attenuations of ~10% on average. Whereas only 76 and 60 hits were found for Saga P dimers and MI001 P dimers, respectively, with average STD effects of about 1% (see **Figure 3.7**).

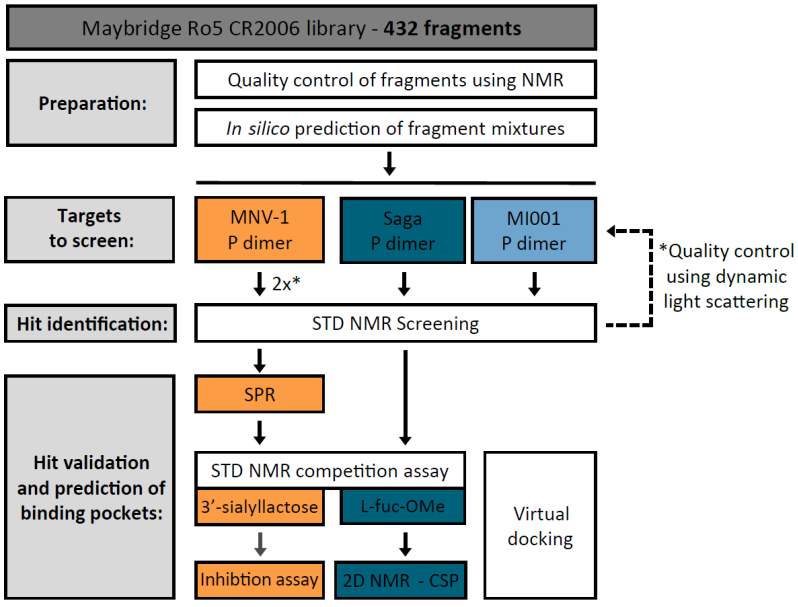


Figure 3.5 | Overview of the screening process
The purity and solubility of fragments were analyzed by ^{1}H NMR. Soluble fragments of the Maybridge library CR2006 (and of the Maybridge library TS2017 for Saga P dimers, 227 soluble members) were distributed *in silico* in smaller compound mixtures and screened against Norovirus protruding (P) domain of one murine strain (MNV-1) and two human strains (Saga and MI001). Following hit identification using saturation transfer difference (STD) NMR spectroscopy, hits binding to MNV-1 were subjected to surface plasmon resonance (SPR) experiments to search for potent fragments. In order to predict putative binding regions of both murine and human P dimers, STD NMR-based competition assays were conducted. The minimal motifs of natural ligands severed as competitors, namely 3'-sialyllactose and methyl α-L-fucopyranoside (L-fuc-OMe) for MNV-1 and Saga, respectively. Promising candidates were further analyzed with respect to their inhibitory effect against MNV-1 virions. For Saga P dimers, chemical shift perturbation (CSP) experiments with isotopically labeled P dimers revealed putative binding sites as substantiated by virtual docking using Glide as implemented in Maestro 11 (Schrödinger release 2017-2).

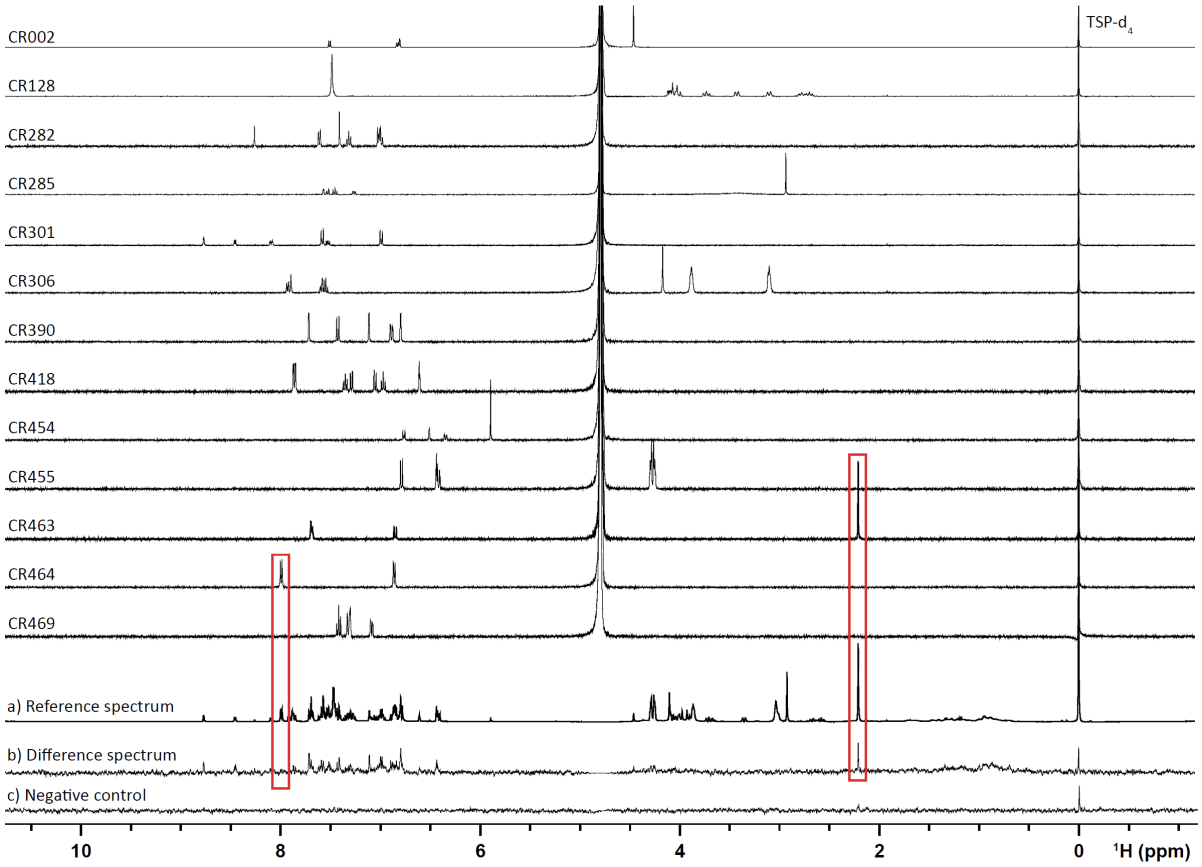


Figure 3.6 | Illustration of the STD NMR-based fragment screening for one mixture
Fragments were allocated into mixtures to allow the identification of at least one NMR signal. The mean of identifiable NMR signals per fragment was two in a reference spectrum of a mixture (a). The 1D NMR spectra of fragments **CR002** to **CR469** are shown above. All spectra were referenced using TSP-d_4 (δ = 0.00 ppm) as an internal reference standard. According to the STD NMR difference spectrum, it is obvious that **CR464** does not interact with MNV-1 P dimers, while **CR463** binds (red boxes). STD NMR spectra were recorded with 1024 scans at 25 °C (for more details see section **3.3.3**).

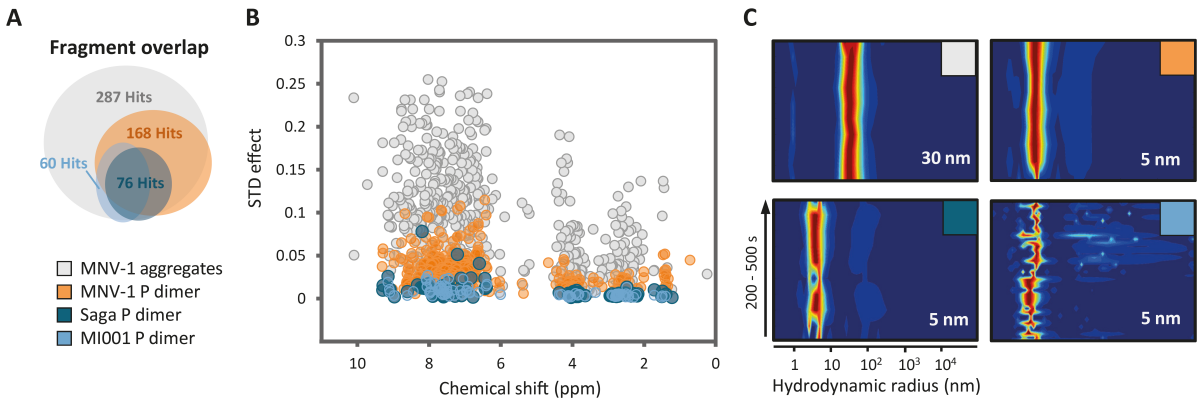


Figure 3.7 | STD NMR-based fragment screening against different Norovirus P dimers
In the first screening procedure, 432 water-soluble fragments were screened against Norovirus P dimers. The protein samples were stored several weeks at 4 °C in 20 – 25 mM phosphate buffer (pH 7.3) in the presence (MNV-1 P dimer aggregates, MI001 P dimer) or absence (Saga P dimer) of 300 mM sodium chloride. The fragment screening had been repeated using freshly prepared MNV-1 P dimer kept in 25 mM sodium phosphate at 4 °C, without sodium chloride. In the case of MNV-1, the number of hits (A), the profile of STD effects of binding fragments (B), and analyses via dynamic light scattering (C) indicate formation of defined aggregates with a hydrodynamic radius of 30 nm on average when stored several month in the presence of a high salt concentration (for details see section **3.3.2** and **3.4.2**).

Since a hit rate of ~66% was undesirable and rather seen for multimeric macromolecules, such as virus-like particles[89,90], freshly expressed MNV-1 P dimers were again subjected to fragment screening using STD NMR as read-out. Prior NMR settings were applied, except the number of scans was increased to 1024. The second screen resulted in 168 positive hits with STD effects of ~3% on average. About 92% of these positive hits had been also found in the first screening round. In order to find reasons for the difference in saturation intensities, dynamic light scattering (DLS) was conducted. Indeed, it turned out that MNV-1 P dimers form aggregates with distinct hydrodynamic radii of ~30 nm when stored at 4 °C in 25 mM sodium phosphate buffer in the presence of 300 mM NaCl. This also explains the high STD effects as this is shown in **Figure 3.7B.** In comparison, the remaining types of P dimers were comparable in size, with a hydrodynamic radius of ~5 nm. Moreover, 1-1 echo NMR experiments were performed to estimate the size of MNV-1 P domain aggregates. In the case of MNV-1 aggregates, an estimated MW of ~64 kDa was obtained, which is in the range of NoV P dimers. For Saga P dimers a MW of ~69 kDa was estimated. When comparing the dispersion of 1D NMR spectra (cf. **Figure 3.8**), it is apparent that Saga P dimers are properly folded, whereas the NMR spectrum of MNV-1 aggregates indicates, to some extent, partially unfolded proteins, which is discussed later (section **3.5.2**).

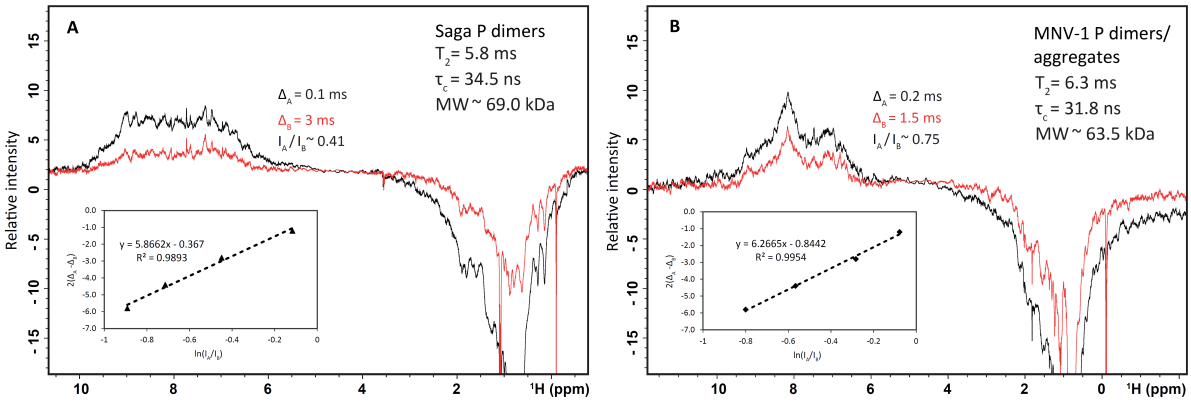


Figure 3.8 | Estimation of the molecular weight of P dimers using 1D 1-1-echo NMR experiments
Each NMR sample contained ~100 µM protein in 25 mM sodium phosphate buffer (pH 7.3) and 10% D_2O. Both samples, Saga P dimers (A) and MNV-1 P dimers/aggregates (B), were treated equally. The transverse relaxation time (T_2) was estimated from the regression line derived by plotting $2(\Delta_A - \Delta_B)$ against ln (I_A/I_B), where Δ_A and Δ_B are different delays and I_A/I_B is the scaling factor (for details see section **3.3.3**). The correlation time τ_c and molecular weight (MW) was estimated using equation **(XI)** and **(XII)**, respectively.

3.4.2 Dynamic Light Scattering – Aggregation of MNV-1 P Dimers

To further investigate the aggregation process of MNV-1 P dimers, protein samples were kept during purification and analyzed by DLS (**Figure 3.9**). After nickel-affinity chromatography, the hydrodynamic radius of ~265 nm of the His-tagged fusion protein MOCR-MNV-1 P domain (MW ~ 50 kDa) was unusually high, indicating oligomerization. The MOCR-tag was removed by complete cleavage with the TEV protease leaving MNV-1 P dimers and aggregates with a hydrodynamic radius of <10 nm and 30-60 nm, respectively. These aggregates could be easily erased by size exclusion chromatography. Since NMR data were acquired at room temperature (RT) occasionally in an automated manner, measuring several NMR samples in a row, the samples of MNV-1 P dimers were monitored for one week at RT. According to DLS measurements, the integrity of P dimers can be guaranteed at least up to three days. However, aggregates were observed on day seven.

In parallel, 100 µM freshly prepared MNV-1 P dimers were stored at 4 °C in 25 mM sodium phosphate (pH 7.3) in the presence or absence of 300 mM NaCl and DLS heat maps were acquired over a period of 36 days. The results are shown in **Figure 3.10**. A slow transition towards defined aggregates appears to start at day seven in the presence of 300 mM NaCl. On the contrary, no aggregation could be detected after 36 days under conditions in which salt was absent.

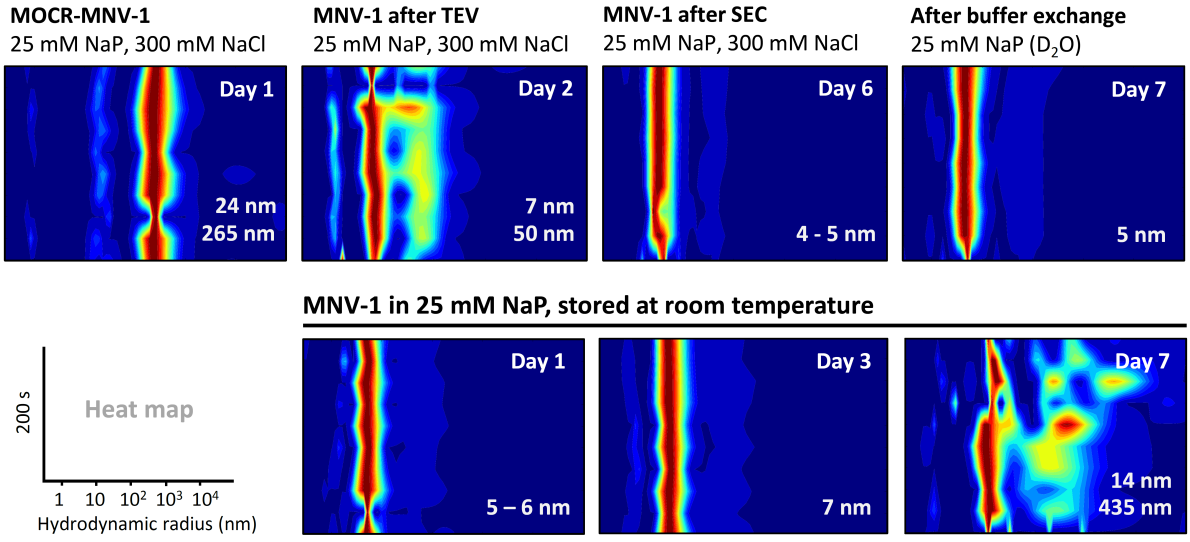


Figure 3.9 | Quality control of MNV-1 P dimer production by dynamic light scattering
Dynamic light scattering of 100 µM protein samples stored at 4 °C in sodium phosphate buffer (NaP, pH 7.3) was conducted after each purification step, including biosynthesis of the fusion protein MOCR-MNV-1, cleavage using TEV proteases, removal of impurities via size exclusion chromatography, and buffer exchange with a 10 kDa cut-off filter (for details see section **3.3.1**). The heat maps reveal a hydrodynamic radius of 5 nm for MNV-1 P dimers on average when freshly purified. At room temperature, MNV-1 P dimers stored in phosphate buffer without salt seem to be invariable up to three days.

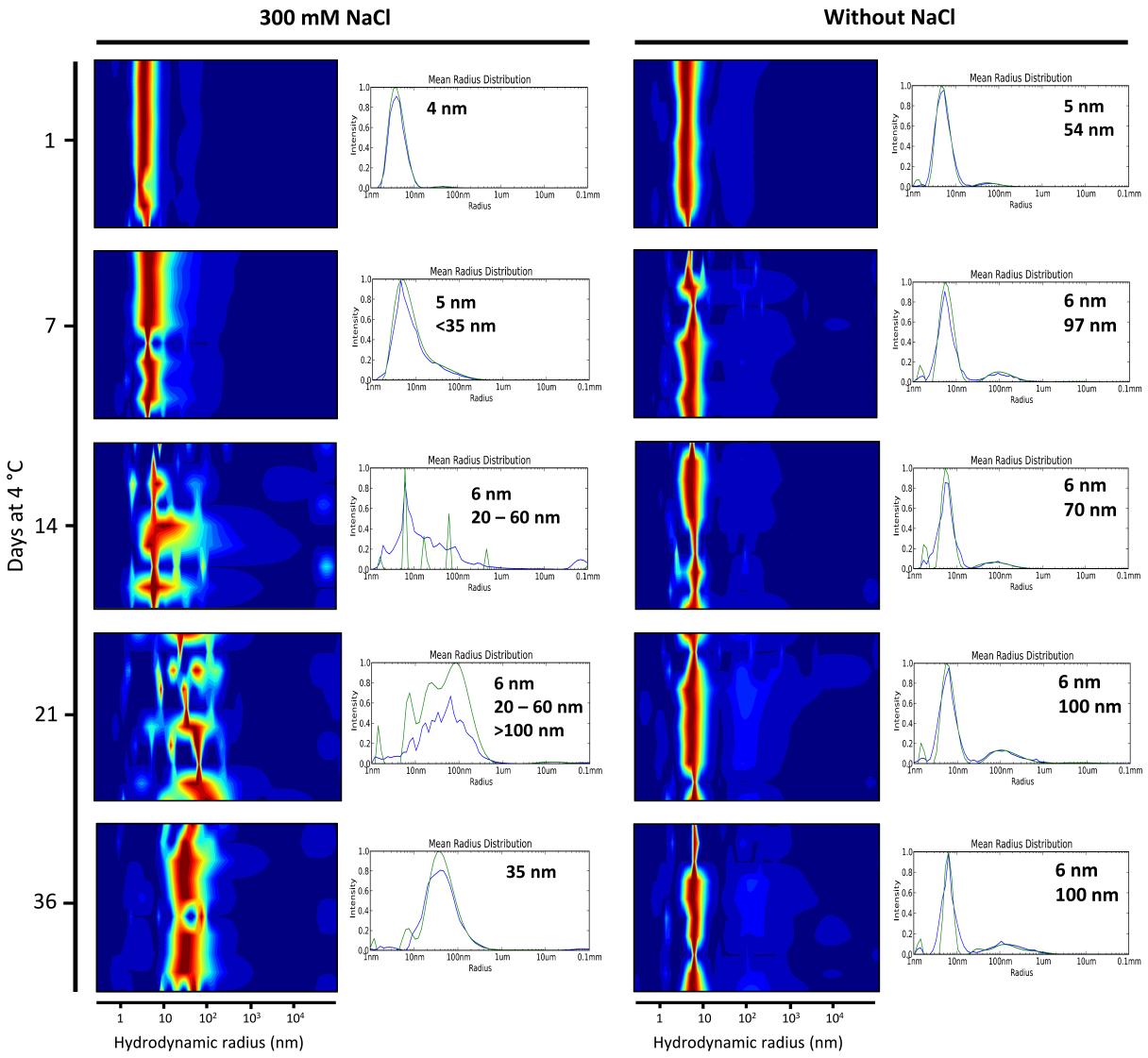


Figure 3.10 | Time course of MNV-1 P dimer aggregation at high ionic strength
Protein samples with 100 µM MNV-1 P dimers stored at 4 °C in 25 mM sodium phosphate (pH 7.3) in the presence or absence of sodium chloride (NaCl) were analyzed by the use of dynamic light scattering over a time period of 36 days. The heat maps and mean radius distribution diagrams were created by the DLS software supplied along with the SpectroSize™300 machine. The maxima of the radius distribution are depicted in each diagram. Almost no aggregation was detected under low salt conditions.

A second protein stock of MNV-1 P dimers had been prepared for SPR experiments as described below. It is believed that in a SPR assay the presence of salt may reduce unspecific binding, which is mediated by electrostatic attraction between ligands and their receptor. Therefore, it was tested whether or not a reduced amount of NaCl still lead to the formation of aggregates (**Figure 3.11**). Moreover, so-called LoBind protein tubes (1.5 ml, Eppendorf) were assessed according to their ability to prevent protein aggregation. It appears that protein aggregation even occurs at 150 mM NaCl when stored at 4 °C. In contrast, the usage of LoBind protein tubes seems to preclude aggregation when salt is absent. When stored at -60 °C the majority of MNV-1 P domains exists as dimers in both cases, albeit a very small fraction tend to form soluble aggregates.

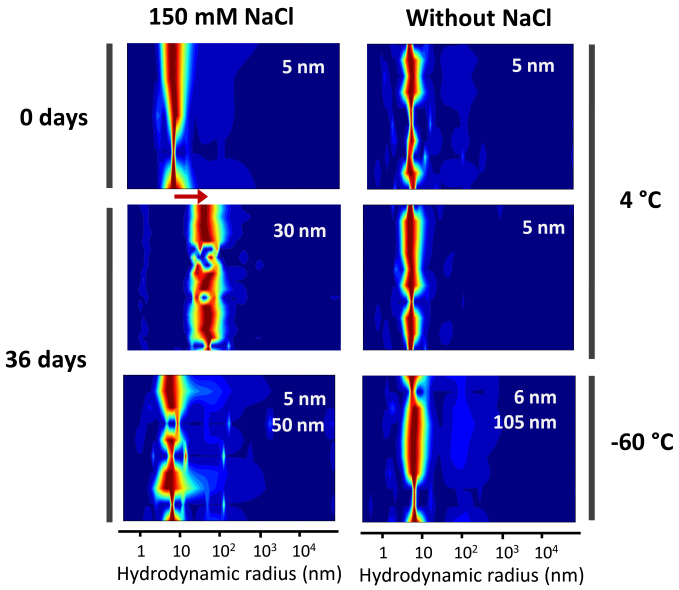


Figure 3.11 | Aggregation of MNV-1 P dimers analyzed by dynamic light scattering
The investigation as described in the main text was repeated under different storage temperatures using less sodium chloride (NaCl) and LoBind protein tubes (purchased from Eppendorf). The hydrodynamic radius (in nm) is presented in the top right corner for each species.

3.4.3 Hallmarks of Hits for Norovirus P Dimers

As already applied for the WNV protease NS2B-NS3^{pro}, SxG matrices were also used to represent positive hits of NoV targets derived from the STD NMR screening (see Figure **3.12**). In total, 30 structurally unrelated hits were found to bind to all three NoV targets. Only one of these fragments carry a carboxyl group and hence being negatively charged at neutral pH. The remainder was assembled of 7 neutral and 22 positively charged compounds. This pattern matches quite well with the general distribution of all positive hits, independently of the type of P dimers. Overall, fragments with higher molecular weights, particularly scaffold clusters S3 and S4, seem to be more frequently identified as positive hits. Whereas compounds consisting of one aromatic five-membered (S9) or six-membered ring (S8) were only accepted if the functional group was either an ester or a primary amino group. In the case of MNV-1 P dimers, methoxy groups were also tolerated. Another common feature of positive hits of all three NoV targets is the fact that only a very small fraction of carboxyl group-containing compounds (G1) were classified as binders. On the contrary, methoxy groups (G5) and amino groups (G6-7) seem to mediate binding to all three types of P dimers, practically no matter which scaffold is presented. In the case of MNV-1 P dimers and Saga P dimers, ester groups (G2) are also highly accepted.
In order to identify common features on the bases of scalar molecular descriptors, a principal component analysis (PCA) was performed using 16 parameters as described in section **2.3.5** and **2.4.1**. Briefly, PCA allows for the extraction of principal components (PCs). The first three to four PCs reflect correlating parameters with high variances, and hence being the most informative ones.

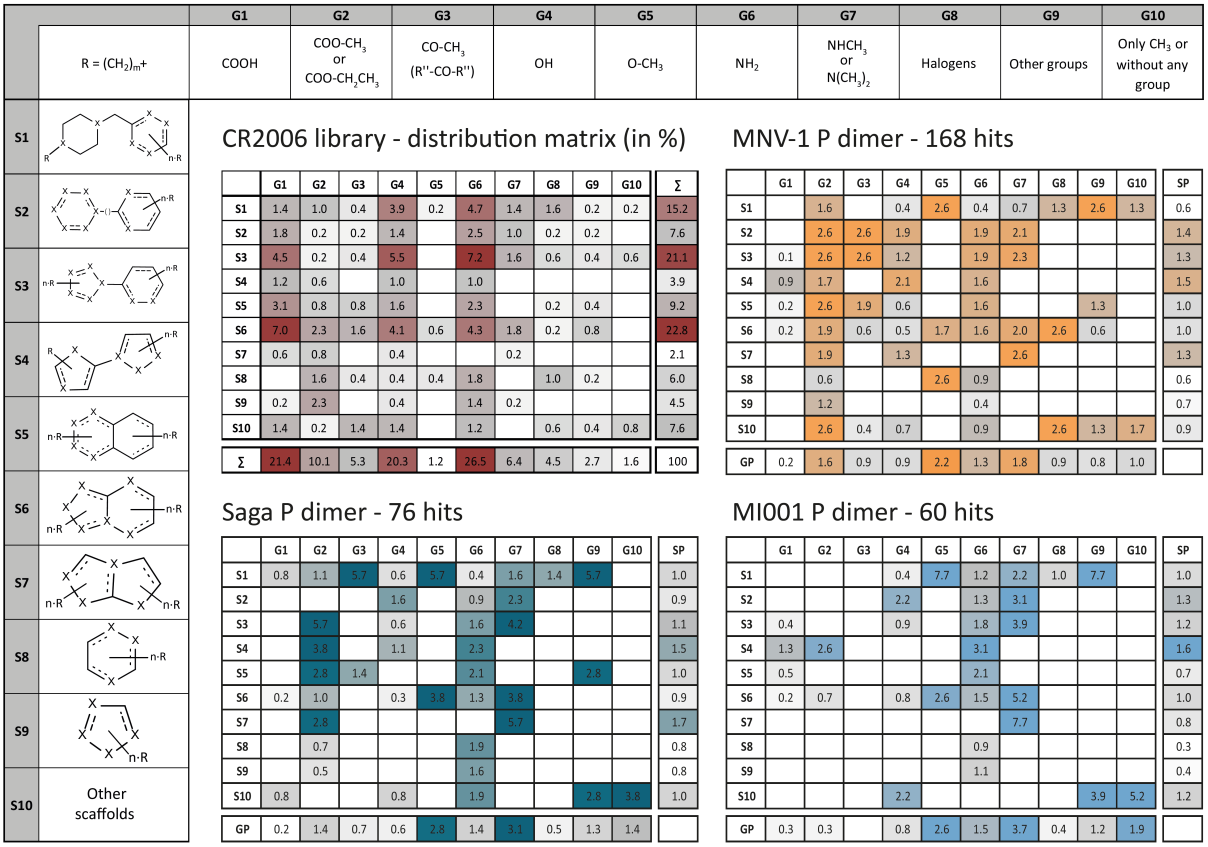


Figure 3.12 | SxG matrices of fragments, which bind to different Norovirus P dimers
The skeletal structural formulas represent typical chemical scaffolds (S) of the Maybridge Ro5 fragment library comprising 432 fragments. The majority of fragments are aromatic molecules containing carbon atoms and several heteroatoms, whereby X symbolizes carbon, nitrogen, oxygen, or sulfur atoms. The scaffold cluster S10 resembles complex molecular structures. In addition, most fragments are functionalized by a number of chemical groups (R, n = 1, 2, or 3 and m = 0, 1, or 2). The functional groups G1 – G10 include carboxyl groups (G1), carboxylic ester groups (G2), other carbonyl groups, including acetyl groups, ketones, or amides (G3; R'' = ring system), hydroxyl groups (G4), methoxy groups (G5), primary (G6) and secondary or tertiary (G7) amino groups, halogens (G8), and other diverse groups that are under-represented in the fragment library (G9, S-CH_3, SO_2-R, SO_2-NH_2, CN, NH-NH_2, CO-NH-NH_2, CO-NH_2, R-NH-CO-CH_3). Some of the fragments carry only methyl groups or no functional group (G10). Moreover, 55 fragments exhibit two functional groups, in which the second chemical group was included. Methyl groups as a second (78 fragments) or third group (16 fragments) were ignored for simplification. The fractions of subclusters of the CR2006 library are presented in % in the top left corner. In contrast, each number of a SxG-matrix of different NoV targets reflects the performance factor of a specific subcluster as this is described elsewhere (**2.3.5**). A small value below one indicates that only a small proportion of this subcluster interacts with the screening target and *vice versa*. (SP, scaffold performance; GP, group performance)

The results shown in **Figure 3.13** are in line with the findings described above. Albeit binders and non-binders are not completely separated as this is seen in the 3D scatter plot of the first three PCs, a tendency of an accumulation is apparent. The major driving force seems to be the molecular charge along with the number of hydrogen bond acceptors and donors. Both the polar surface area (PSA) and the distribution coefficient log D also contribute to the dispersion of binders and non-binders, since these parameters correlate either negatively or positively with the aforementioned scalar descriptors (see **Figure 3.13B/D)**.

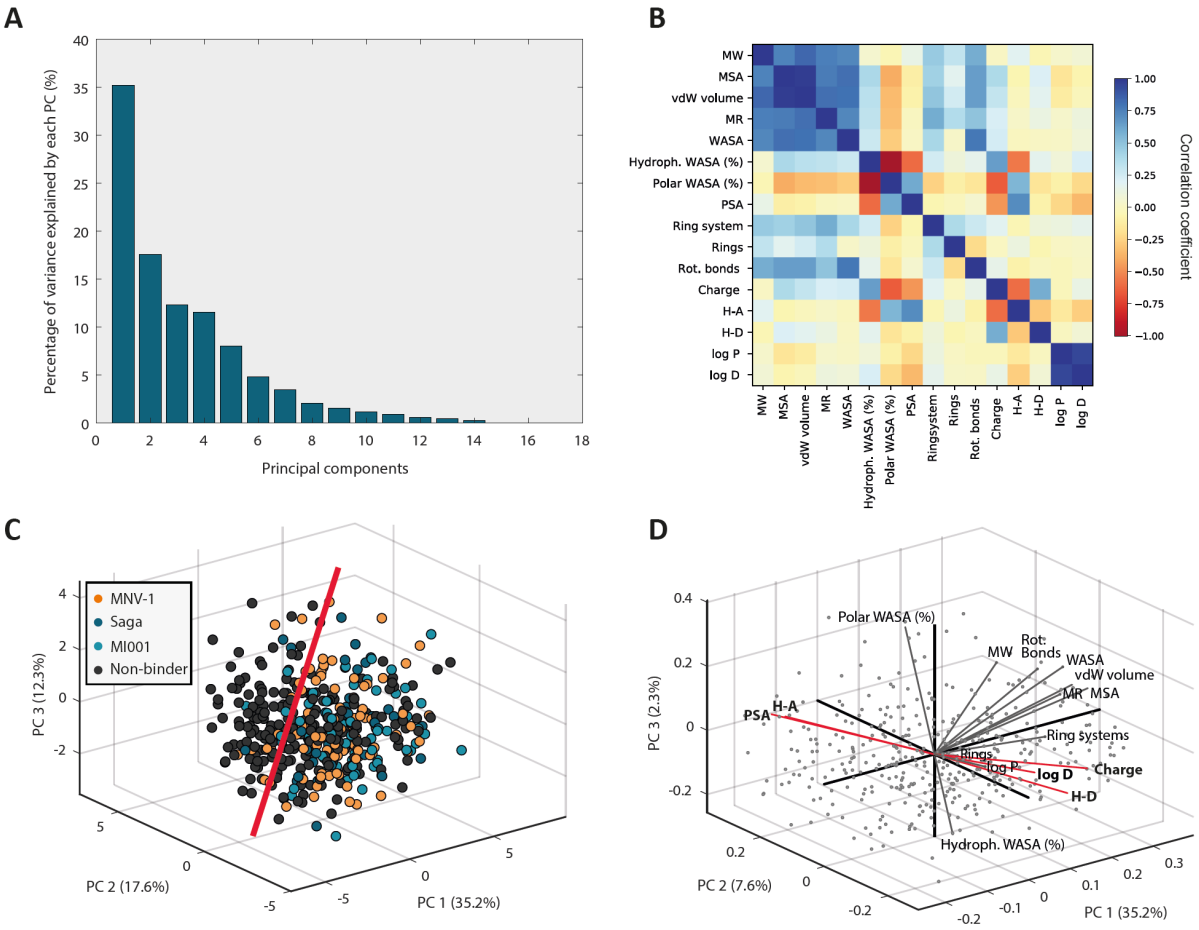


Figure 3.13 | Comparison of binders and non-binders of Norovirus targets using principal component analysis In total, 16 chemical parameters were determined for each fragment of the Maybridge library using the ChemAxon's calculator, a command line program in JChem implemented in the python script *calcChemParam_Library.py* (see Appendix A4). The molecular properties are molecular weight (MW in Da), molecular surface area (MSA in Å^2), van der Waals volume (vdW volume in Å^3), molecular refractivity (MR), water accessible surface area (WASA in Å^2), the ratio of hydrophobic and polar WASA, polar surface area (PSA in Å^2), the number of ring system, the number of rings, the number of rotatable bonds, total charge, the number of H-bond acceptors (H-A) or H-bond donors (H-D), the partition coefficient log P, and the distribution coefficient log D. The function PCA implemented in Matlab (Appendix A4) extracts principal components that combine correlating parameters (e.g., PSA and H-A), which are the most informative variables. **(A)** The first three components explain 65.1% of the variance of the given data set. **(B)** The heat map shows the magnitude of the pairwise Pearson's correlation coefficient for each molecular descriptor. **(C)** The 3D scatter plot displays the distribution of non-binders (black dots) and positive hits for MNV-1 P dimer (orange), Saga P dimer (dark blue), or MI001 P dimer (light blue) along the first three principal components (PC). Fragments that were identified to interact with more than one target are also displayed in orange (ligands of MNV-1, Saga, and MI001 P dimers) or dark blue (ligands of Saga and MI001 P dimers). The majority of colored dots accumulates on the right side of the red line. **(D)** The biplot represents each fragment as gray points and parameters as gray vectors. The distribution of data points is the same as for (C). The direction and length of the vectors indicate how each variable contributes to the three principal components. The red vectors PSA, H-A, log D, H-D, and charge are almost perpendicular to the red line in (C). Therefore, these parameters predominantly determine the dispersion of binders and non-binders. The PC scores are normalized due to the algorithm of Matlab. For a better visualization, 2D projections of 3D plots are presented in Appendix A3.

In order to confirm the results from PCA, the statistical dispersions of selected scalar descriptors were graphically depicted by boxplots or histograms. The comparison of distributions of all water-soluble fragments from the Maybridge library CR2006 and positive hits derived from STD NMR screening substantiate that the total charge, the number of hydrogen-bond acceptors as well as donors, and PSA predominantly determine whether or not a fragment is classified as binder. Note that these descriptors are the only ones with significantly (p-values < 0.05) different distributions for all NoV targets (statistical tests are described in section **2.3.5**). The reasons for that are discussed later.

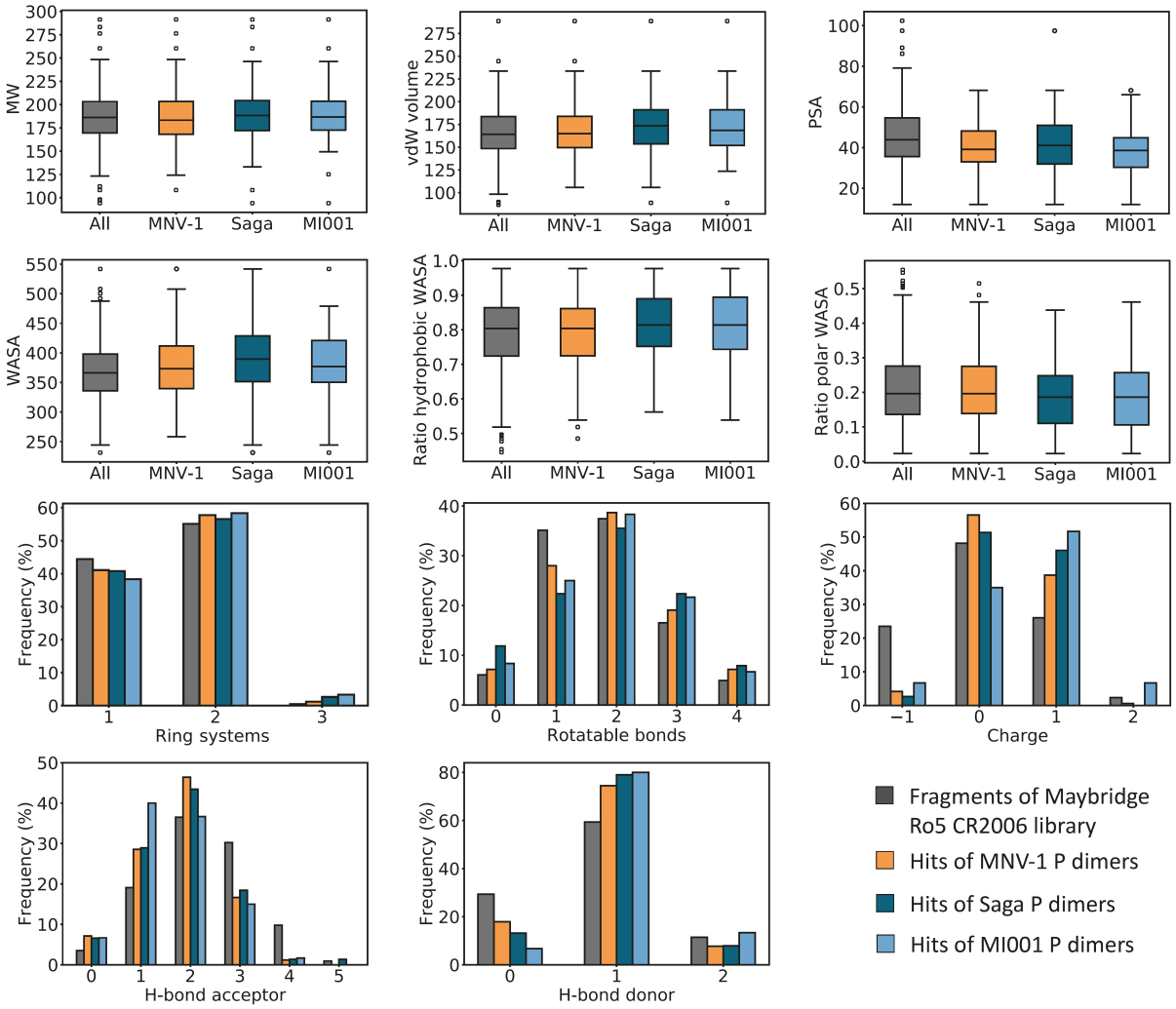


Figure 3.14 | 1D molecular descriptors of hits identified by STD NMR
Box plots and histograms represent the distribution of structural and physicochemical properties of fragments, which were identified to bind to different Norovirus P dimers. The upper box plots show size and surface properties, including molecular weight (MW; Da), van der Waals volume (vdW; Å^3), polar surface area (PSA; Å^2), water accessible surface area (WASA; Å^2), the ratio of hydrophobic WASA/WASA and polar WASA/WASA. The lower histograms display structural or physicochemical properties such as the number of ring systems, number of rotatable bonds, total charge, and the number of hydrogen bond (H-bond) acceptor and donor. All values were determined by employing the ChemAxon *cxcalc* plug-in as described in section **2.3.4**. The boxplots and histograms were created using the Python script *Boxplot_Histogram_STD_Hits.py*, which is available in Appendix A4.

Finally, so-called 3D bubble charts or also termed 3D histograms were utilized to get insights into the underlying structural principles of binders on a general abstract level. Therefore, binned van der Waals volumes were plotted against the number of hydrogen-bond acceptors and the molecular charge, thereby combining one shape parameter with two physicochemical parameters. In the case of MNV-1 P dimers, for example, approximately one third of hits are neutral fragments with two hydrogen bond acceptors and a vdW volume above 140 Å^3. In addition, one fifth of hits are characterized by the following parameters: a van der Waals volume between 160 Å^3 and 300 Å^3, one hydrogen bond acceptor, and one positive charge. The latter fragments were also identified as "frequent hitters" for Saga P dimers and MI001 P dimers.

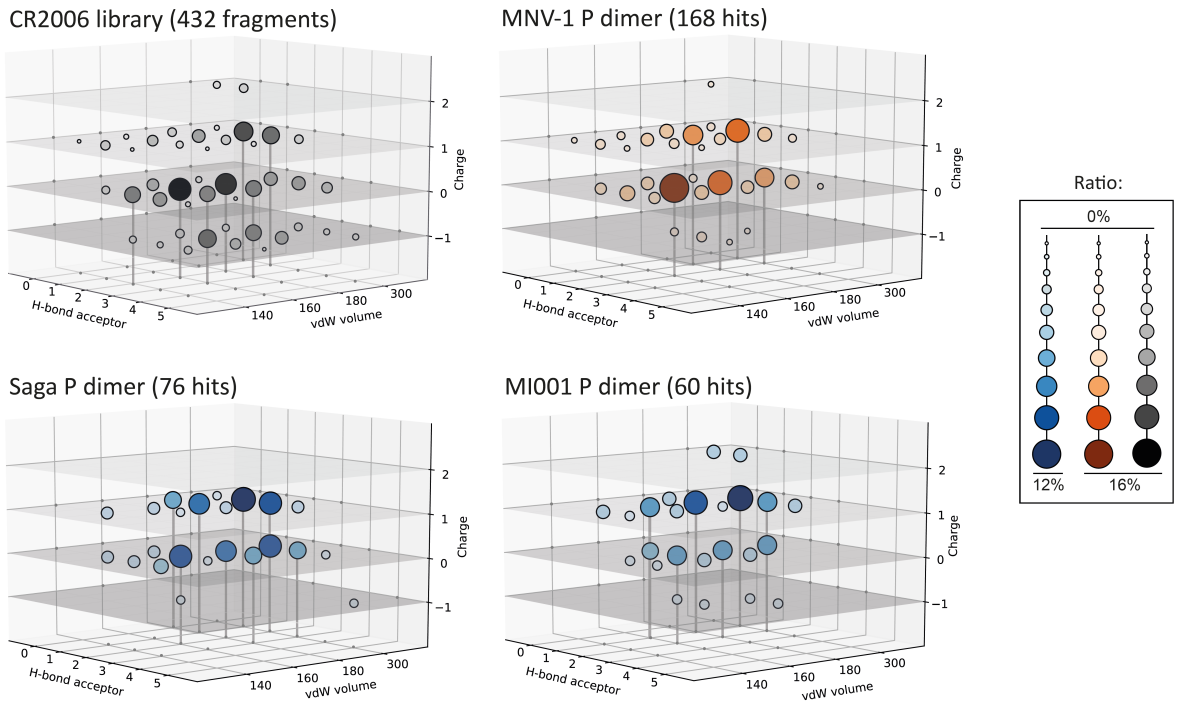


Figure 3.15 | 3D histograms of informative scalar descriptors of STD NMR positive hits
These plots describe the distribution of fragments characterized by the number of hydrogen bond (H-bond) acceptors, van der Waals (vdW) volume, and total charge. Regarding the vdW volumes, fragments were equally divided into four classes: <140, < 160, <180, and <300 Å^3. The plots were created by the python scripts *searchParameter.py* and *3D_histo_for_A_vdW_Q.py* (see Appendix A4).

3.4.4 Evaluation of Hits for MNV-1 P Dimers

Albeit the hit rate of positive hits that interact with MNV-1 P dimers could be reduced from ~66% to ~39% by controlling the dimerization state and avoiding aggregation of P dimers, no structural motif could be defined as a promising lead structure for the development of viral modulators. To restrict the number of possible hits, surface plasmon spectroscopy (SPR) was performed as a second screening method followed by STD NMR competition experiments and virus inhibition assays.

Surface Plasmon Resonance Spectroscopy

Approximately 6500 RU freshly expressed MNV-1 P dimers were immobilized on a Biacore CM5 sensor chip by using the amine coupling method. In order to verify the integrity of immobilized protein with respect to its binding activity, 3'-sialyllactose was injected at increasing concentrations. Non-linear least squares fitting of the titration data yielded a two-step binding curve (see **Figure 3.16**) with two dissociation constants (K_{D1} ~ 3.5 mM and K_{D2} ~ 12.2 mM). A similar curve as already been published in another SPR-based analysis[243] for 3'-sialyllactose interacting with immobilized MI001 P dimers. Furthermore, the two STD NMR hits **CR273** and **CR432** were chosen according to their high STD effects (8 – 9%) for the titration onto immobilized MNV-1 P dimers. A concentration of 200 µM was sufficient to reach saturation with respective response levels of ~6 RU and ~12 RU. It appeared that unspecific interaction dominates the binding process at higher fragment concentrations, as data points resemble a straight line in the range of 1 mM to 4 mM (data not shown). By the use of optimized settings for the SPR screen, 69 hits were verified, which correspond to a final hit rate of ~16%.

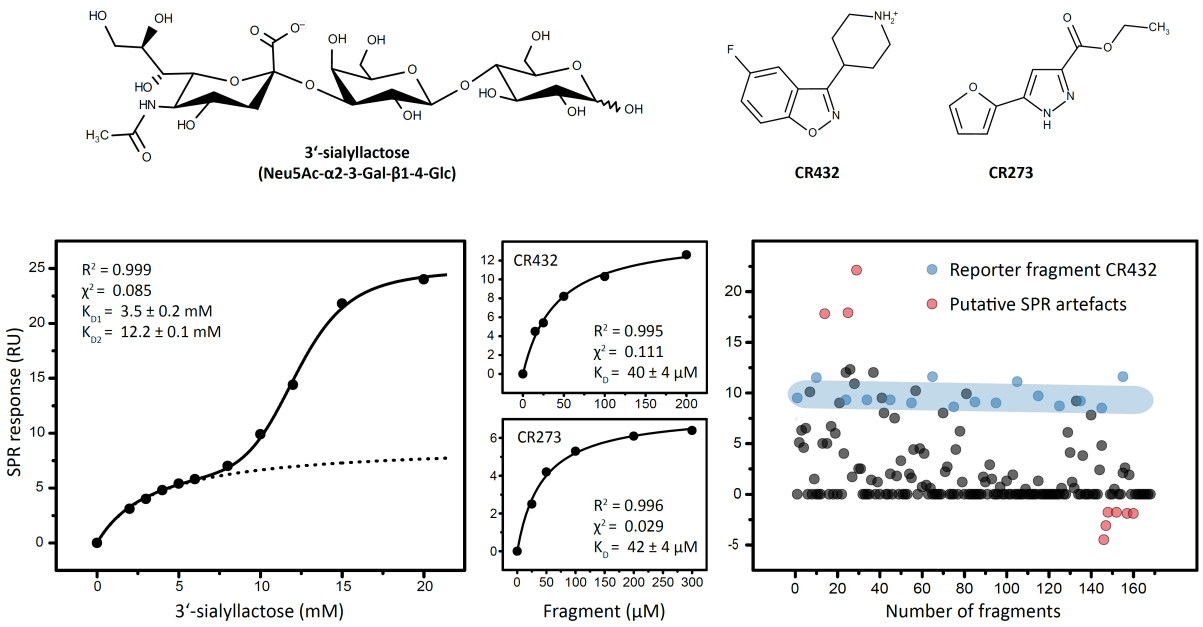


Figure 3.16 | Validation of STD NMR hits of MNV-1 P dimers by using SPR
The SPR response (RU, relative units) is shown as a function of ligand concentration. Prior to the screening of fragments, the integrity of MNV-1 P dimers immobilized onto a Biacore sensor chip CM5 was tested by titrating the natural binding motif 3'-sialyllactose. Additionally, the titration of fragment **CR432** and **CR273** was performed to determine an appropriate fragment concentration for the SPR screening (K_D, dissociation constant). The right scatter plot represents the SPR response of 168 fragments screened at a concentration of 200 µM. During the screening procedure, a reporter fragment (**CR432**) was injected every 10 to 15 fragments to judge the robustness of the binding assay. The blue area indicates the regression line of blue spots. Data points colored in red symbolize putative SPR artefacts.

To further assess results from SPR screening, SxG matrices were created along with boxplots and histograms as this is seen in **Figure 3.17**. It is striking that, besides carboxylated compounds, also fragments from cluster S8 and S9 did not show any SPR response. On the other hand, there was a clear preference for positively charged hits with higher van der Waals volume and concomitant decreased polar surface area. Moreover, the majority of SPR hits exhibited less than three hydrogen-bond acceptors and only one hydrogen-bond donor.

168 STD NMR hits - SxG distribution matrix

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	∑
S1		1.6		1.6	0.5	2.1	1.1	2.1	0.5	0.5	10.1
S2		0.5	0.5	2.7		4.8	2.1				10.6
S3	0.5	0.5	1.1	6.9		13.8	3.7				26.6
S4	1.1	1.1		2.1		1.6					5.9
S5	0.5	2.1	1.6	1.1		3.7			0.5		9.6
S6	1.6	4.3	1.1	2.1	1.1	6.9	3.7	0.5	0.5		21.8
S7		1.6		0.5			0.5				2.7
S8		1.1		0.0	1.1	1.6					3.7
S9		2.7				0.5					3.2
S10		0.5	0.5	1.1		0.5		1.6	0.5	1.1	5.9
∑	3.7	16.0	4.8	18.1	2.7	35.6	11.2	4.3	2.1	1.6	100

69 SPR hits - SxG distribution matrix

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	∑
S1				1.3	1.3	3.8	2.6	1.3	2.6		12.8
S2				2.6		3.8	5.1				11.5
S3		1.3		2.6		17.9	3.8				25.6
S4		1.3		3.8		2.6					7.7
S5			2.6			3.8	1.3				7.7
S6		3.8			2.6	6.4	9.0	1.3			23.1
S7		1.3					1.3				2.6
S8											
S9											
S10			1.3	2.6				2.6		2.6	9.0
∑		7.7	3.8	12.8	3.8	38.5	23.1	5.1	2.6	2.6	100

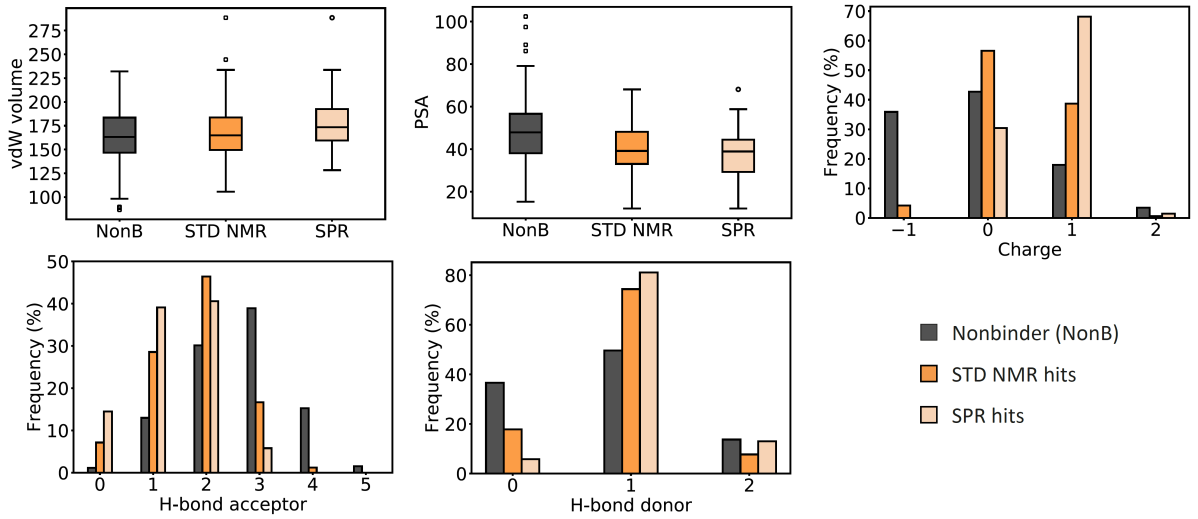


Figure 3.17 | Descriptive statistics of SPR hits of MNV-1 P dimers
The scaffold-group (SxG) distribution matrices are shown for verified hits from STD NMR screening and SPR screening, in which MNV-1 P dimers were utilized as a molecular target. The values (in %) in the matrices reflect the ratio of each subcluster to all 168 STD NMR hits or all 69 SPR hits. The definition for scaffold cluster (S1 – S10) and group cluster (G1 – G10) can be found in section **2.3.5** and **Figure 2.6**. In the lower panel, both boxplots and histograms summarize informative key properties of fragments that were identified as positive hits in the respective assay.

STD NMR Competition Experiments

In a third step, the selection process was guided by STD NMR competition assays using 3'-sialyllactose as a competitor for the glycan binding site of MNV-1 P dimers. In total, 12 mixtures containing three to seven fragments were screened at a concentration of 200 µM each. Therefore, SPR hits were mixed with 5 µM MNV-1 P dimers and the attenuation of NMR signals upon saturation transfer was quantified in the presence or absence of 8 mM 3'sialyllactose. Four NMR signals could be evaluated per fragment on average, yielding 19 putative competitors. The greatest reduction of STD effects was observed for fragment **CR120**, with an average decline of 35%.

Exemplary results for one fragment mix and chemical formulas of presumed competitors are given in **Figure 3.18**. One common feature for almost all competing fragments is a primary or secondary amino group on the one end of the molecule and at least one hydrogen bond acceptor (divalent sulfur included[257]) on the opposite side. Exceptions are the fragments **CR054**, **CR106**, **CR107**, **CR152**, and **CR385**. These can be approximately characterized by detergent-like structural properties, namely a hydrophilic and a hydrophobic region.

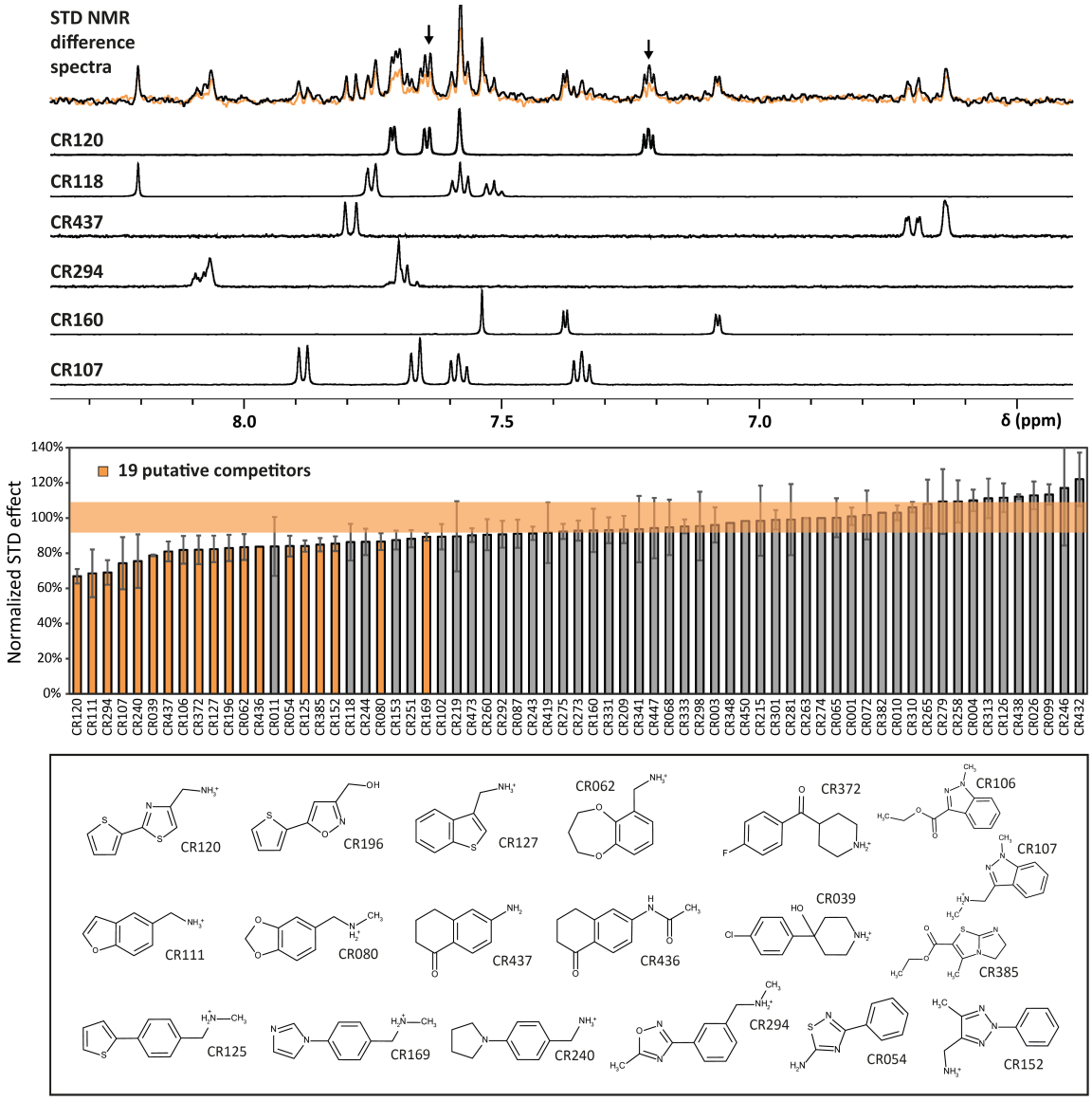


Figure 3.18 | STD NMR competition experiments reveal 19 putative competitors
The competition between SPR positive hits and the competitor 3'-sialyllactose was analyzed by the observation of the reduction of the STD effect of several ^{1}H NMR resonance signals of a specific fragment. In the upper panel, the STD NMR difference spectra are shown for a fragment mixture (**CR120** – **CR107**; 200 µM each) prepared with (orange) or without (black) competing ligand (8 mM). At least two isolated signals of a fragment could be monitored as this is seen in the reference spectrum of **CR120** (arrows). The lower diagram illustrates the amount of reduction of the STD effect of four NMR signals on average for each fragment. The error bars represent the standard deviation. Fragments showing STD effects (including error) below 90% in the presence of 3'-sialyllactose were considered as competing ligands. The threshold (orange bar) was set based on an estimated experimental error of approximately 10%.

Virus Inhibition Assays

It has been shown that the infectivity of murine NoV (MNV-1) can be modulated when treated with small substances, such as L-fucose or ammonium sulfate (Wegener et al., 2017; unpublished data). In this present study, MNV-1 were mixed with selected competitors (see above) and structurally related compounds. The final concentration of fragments was set to 0.5 to 2 mM, depending on the solubility. Following incubation for 1 h at 37 °C, the viral infectivity was quantified using endpoint titration (see section **3.3.5**). The results from infection assays are shown in **Figure 3.19**. In total, seven fragments (out of 28) caused at least a five-fold reduction of virus titer. To confirm these findings, the dose-response of the fragments **CR120** and **CR107** was evaluated. In the case of **CR120**, a concentration of 1 mM was sufficient to decrease the infectivity to ~1% as compared to untreated MNV-1, whereas at 6 mM the $TCID_{50}$ was virtually below the detection limit.

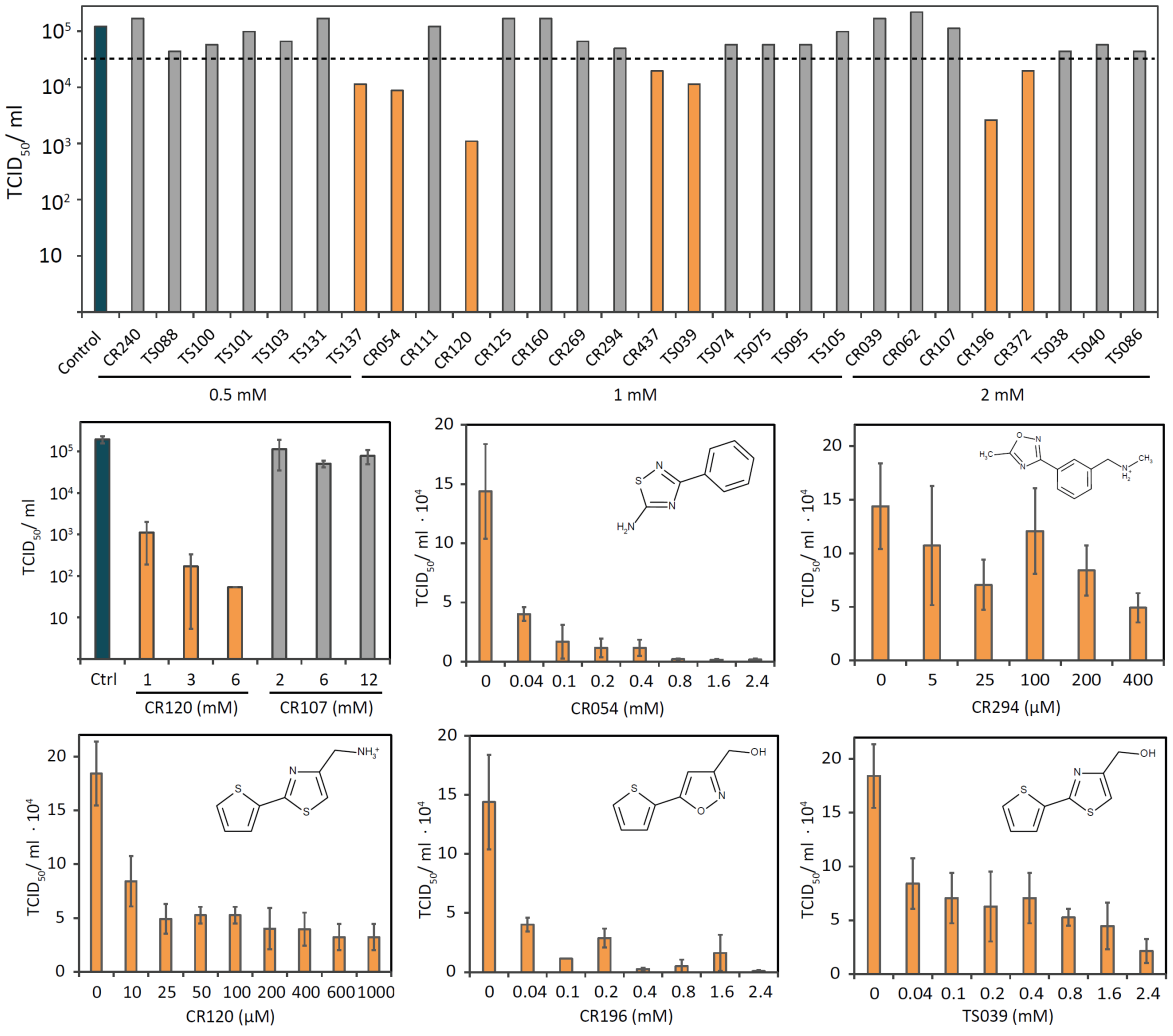


Figure 3.19 | MNV-1 inhibition screen of fragments
In order to identify fragments with inhibitory properties, approximately 10^5 infectious murine Noroviruses (in 300 µl PBS buffer) were incubated in the presence of 0.5 to 2 mM fragment (depending on the solubility) for 1 h at 37 °C. After incubation, the number of infectious virions was determined by endpoint titration ($TCID_{50}$, median of tissue culture infection dose). For promising candidates and analogs, the concentration dependency was evaluated. The fragment **CR107** served as negative control. The error bars represent the standard deviation of a double determination (upper left diagram, **CR120** and **CR107**) or triple determination (for details see section **3.3.5**).

Since **CR120** is positively charged any ionic-strength-dependent inhibition should be ruled out. Therefore, **CR107** (see **Figure 3.18**) served as negative control. Even at 12 mM substantially no unspecific inhibition was detectable. Next, the dose-dependent inhibitory efficacy of the most promising compounds (**CR054**, **CR120**, **CR196,** and **TS039**) was estimated. The fragment **CR294** was also tested due to its similarity with **CR054**. Besides, the cytotoxic effect on cell proliferation was analyzed for highly potent fragments (see **Figure 3.20**). Estimated IC_{50} values in the low micromolar range and LC_{50} values in the millimolar range are given in **Table 3.2**. An analysis of structure-activity relationships suggests that larger groups at the thiazole ring of **CR120** hamper the inhibitory activity of these fragments. The same applies to the oxazole ring of **CR196**. In the case of **CR120** the position of the amino group seems to be important as well (see **Table 3.3; TS074/TS075**), whilst a hydroxyl group appears to be beneficial for **CR196**.

Table 3.2 | Summary of MNV-1 inhibitors

Fragment	Estimated IC_{50} (µM)	Estimated LC_{50} (mM)
CR054	<40	~1.2
CR120	~10	~0.6
CR196	<40	>1.6
CR294	<400	n.d.
TS039	~40	~1.0
TS137	<500	n.d.

n.d. = not determined

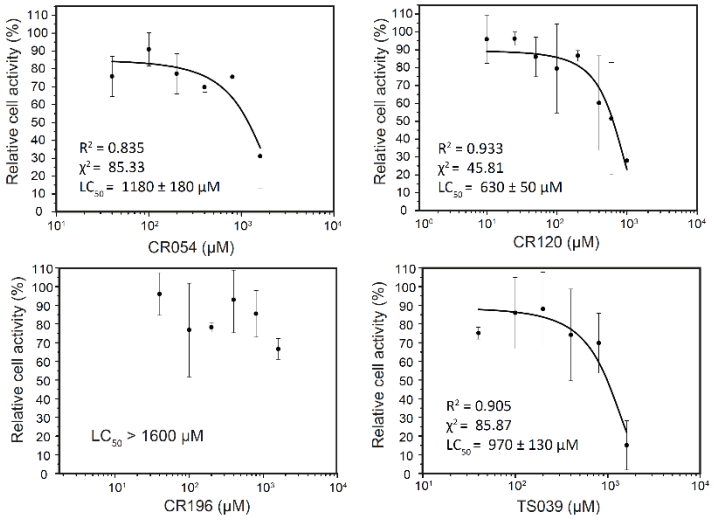
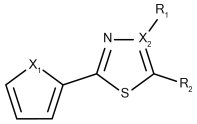
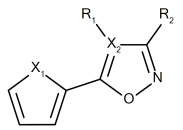


Figure 3.20 | Cell viability test ▶
Approximately 10^4 BV2 cells were incubated in a 96-well plate for 24 h at 37 °C with fragments at increasing concentration. To determine the median of lethal concentration (LC_{50}), a WST1-based colorimetric assay served as read-out.

Table 3.3 | Structure-activity relationship of MNV-1 inhibitors

Scaffold	Compound	c (mM)	X_1	X_2	R_1	R_2	$TCID_{50}$ (%)
	CR120	1	S	C	CH_2NH_2	H	1%
	TS039	2	S	C	CH_2OH	H	9%
	TS038	2	S	C	COOH	H	36%
	TS100	0.5	S	C	$COOCH_2CH_3$	H	47%
	TS103	0.5	S	C	$CH_2CONHNH_2$	H	54%
	TS101	0.5	S	C	CH_3	$CH_2CONHNH_2$	81%
	TS074	1	S	N	-	NH_2	47%
	TS075	1	O	N	-	NH_2	47%
	CR196	2	S	C	H	CH_2OH	2%
	TS086	2	S	N	-	CH_2OH	36%
	CR269	1	O	C	H	CH_2NH_2	54%
	TS105	1	S	C	$CONHNH_2$	CH_3	81%

$TCID_{50}$ (%) = normalized tissue culture infection dose 50; c = final concentration

3.4.5 Discovery of Novel Hits for Saga P Dimers

In the first screening round using STD NMR spectroscopy as a read-out, 76 fragments had been identified to interact with Saga P dimers. To further explore the chemical space of specific binders, the fragment library CR2006 has been extended by 227 water-soluble fragments (see section **2.3.3**), designated as TS2017 library. Afterward to that, STD NMR competition assays were conducted by the use of methyl α-L-fucopyranoside as a competitor. For the identification of corresponding binding pockets, U-[^{2}H,^{15}N]-labeled Saga P dimers were subjected to chemical shift perturbation experiments. Finally, in order to interpret results from epitope mapping, virtual docking was performed as this is described in the last section.

STD NMR Screening of the Maybridge TS2017 Library

The screening of the library TS2017 resulted in 51 additional hits. Interestingly, the fraction of negatively charged compounds remained very small when comparing all hits originated from the libraries CR2006 and TS2017 and non-binders. According to PCA, the net charge seems to be the main factor that hinders binding to Saga P dimers (cf. **Figure 3.21**). The increased number of positive hits has not altered the overall distribution of scalar descriptors. As noted above, neutral or positively charged compounds with slightly higher van der Waals volumes and two or less H-bond acceptors combined with only one H-bond donor appear to be favored.

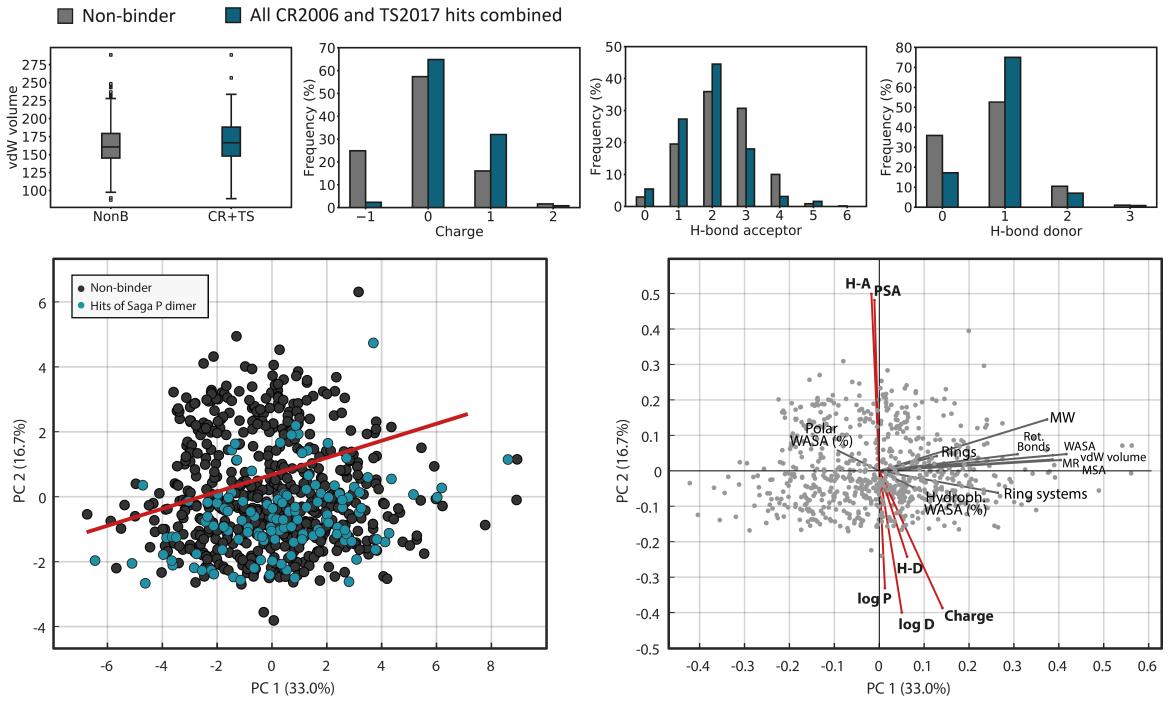


Figure 3.21 | Characterization of STD NMR hits of Saga P dimers
A boxplot and histograms are depicted for four informative molecular parameters, including van der Waals (vdW) volume in Å^{3}, total charge, the number of hydrogen bond (H-bond) acceptors, and the number of H-bond donors. The charge seems to be a strong criterion for fragments interacting with Saga P dimers as this is substantiated by a principal component analysis (for details see section **2.3.5**). The scatter plot displays the distribution of non-binders (black) and positive hits of Saga P dimers (blue) along the first two principal components (PC). The biplot (on the bottom right) indicates that the charge is the main cause of the dispersion depicted in the scatter plot among others.

STD NMR Competition Experiments

The STD NMR competition format enabled the identification of positive hits whose binding is influenced by the presence of methyl α-L-fucopyranoside mediated either through direct competition or through allosterism[163,205,258]. The presence of methyl α-L-fucopyranoside in a 40-fold or 160-fold excess caused a notable reduction of STD effects for 21 fragment. These fragments are depicted in **Figure 3.22**. The majority was practically unaffected. However, a few compounds experienced a significant enhancement of saturation transfer when methyl α-L-fucopyranoside was present (e.g., **CR004**, **CR031**, **CR039**, and **TS019**). The statistical distributions of scalar descriptors of putatively competing or allosteric fragments almost mirror the distributions of all hits derived from STD NMR screening (data not are shown), and hence there are no obvious selection criteria based on one single molecular parameter alone. Moreover, it is not possible to define a distinct pharmacophore model due to the high diversity of structural motifs, indicating either a low selectivity of the corresponding binding pocket or several binding pockets that are involved.

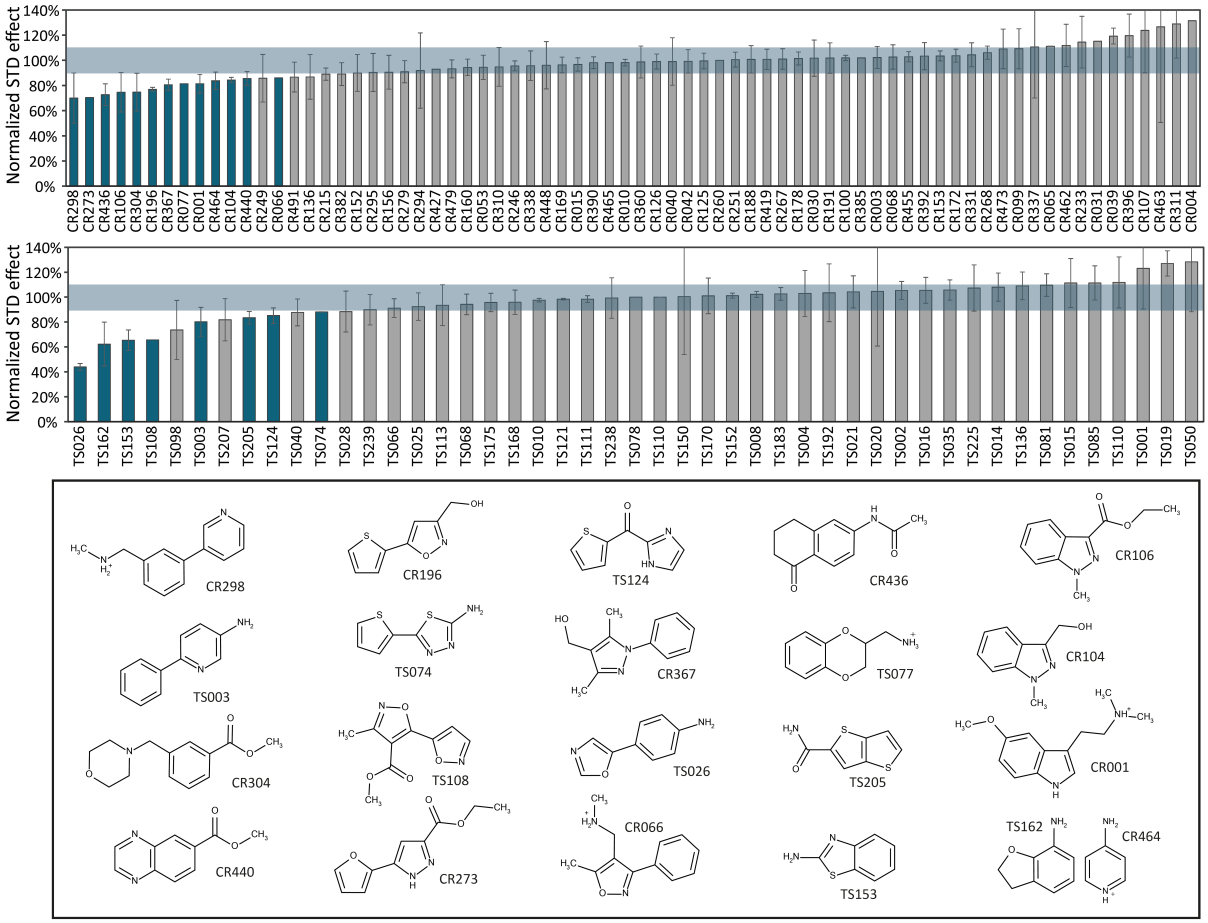


Figure 3.22 | Analysis of competition for fucose binding sites of the Saga P dimer using STD NMR as read-out
The STD effect of 200 µM fragment was monitored in the absence and presence of 8 mM (CR fragments) or 32 mM (TS fragments) methyl α-L-fucopyranoside supplemented with 5 µM Saga P dimers (for details see section **3.3.3**). For each fragment, about three NMR signals could be analyzed. The error bars represent the standard deviation. The lower panel shows 21 fragments whose interaction with Saga P dimers seem to be influenced by fucose binding in a competing manner. The estimated error limit was set to ±10% (transparent bar).

Chemical Shift Perturbation Experiments

The ultimate aim of this study was the identification of novel structural motifs that may guide the development of viral entry inhibitors[90]. Besides that, fragment screening may support the search for allosteric binding pockets[205]. Thanks to the assignment of U-[^{2}H,^{15}N]-labeled Saga P dimers (Mallagaray and Creutznacher, 2017; unpublished data) the binding mode of particular fragments could be evaluated on an atomic level.

In total, 35 fragments were analyzed. This includes 22 hits derived from STD NMR competition assays showing decreased or increased STD effects in the presence of methyl α-L-fucopyranoside. Moreover, three randomly selected hits showing no change in STD effects (**CR100**, **CR125**, and **CR251**) were chosen. In addition, six fragments from the virtual screening (see below) were selected for which docking into or close to the fucose binding site yielded the highest docking scores (i.e., **CR025**, **CR070**, **CR055**, **CR365**, **TS188**, and **TS189**). Three MNV-1 inhibitors (**CR054**, **CR120**, and **TS039**) and one putative non-binder according to STD NMR were also tested. Unfortunately, following fragments were excluded from CSP experiments due to their relatively low solubility in the assay buffer used: **CR273**, **CR298 CR304**, **CR385**, **CR436**, **CR464**, **TS003**, **TS010**, **TS013**, **TS019**, **TS020**, **TS108**, and **TS205**.

Prior to NMR measurement, the pH of 12 to 20 mM saturated fragment stock solutions was adjusted to pH 7.3 to avoid pH artefacts. Under these circumstances high-quality ^{1}H,^{15}N-TROSY-HSQC spectra were obtained. One example is depicted in **Figure 3.23**. Other 2D correlation spectra can be found in Appendix A7. In order to map corresponding interaction sites, the Euclidian distances of NH NMR signals were determined upon fragment binding to Saga P dimers, yielding chemical shift differences of up to 0.2 ppm (see **CR065** and **TS153**). Next, backbone nitrogen atoms in a crystal structure of a P dimer (strain: GII.4 Saga; PDB 4WZK) were represented as balls and colored according to the magnitude of associated chemical shifts as this seen in **Figure 3.23**. For the automatization of this process, the Python script *python2Maestro_plotCSP.py* was utilized (see Appendix A4). The thresholds for color coding (mean of Euclidian distances + x·standard deviation; where x = 1, 2, or 3) were chosen arbitrarily, and calculated for each fragment separately. The second threshold (x = 2) ensured for all fragments that only meaningful chemical shifts were analyzed, as the experimental error of CSP experiments was estimated to be approximately 0.01 ppm (Mallagaray, 2017; personal communication).

Pursuing this experimental and analytical strategy yielded more or less defined binding pockets for each binding fragment. The fragment **CR065** (an analog of **CR066**), for example, seems to interact with the protein interface being 30 to 35 Å apart from the canonical glycan binding site. This binding pocket, which is mainly defined by Leu 272, Leu 278, Leu 306, and Leu 321 of P domain A as well as Glu 236 and Asn 512 of P domain B, also attracts a relatively great number of other fragments, which were analyzed (e.g. **CR001**, **CR066**, **CR251**, **CR367**, and **TS207**). In contrast, some fragments, in particular when negatively charged, seem to disturb several binding sites, as the amino acid residues affected are distributed throughout the protein. This indicates either promiscuous binding, experimental artefacts, or long-range effects. Some examples include the fragments **CR025**, **CR055**, **CR070**, **CR311**, and **CR365**. All results obtained from chemical shift perturbation experiments are summarized in **Figure 3.24**.

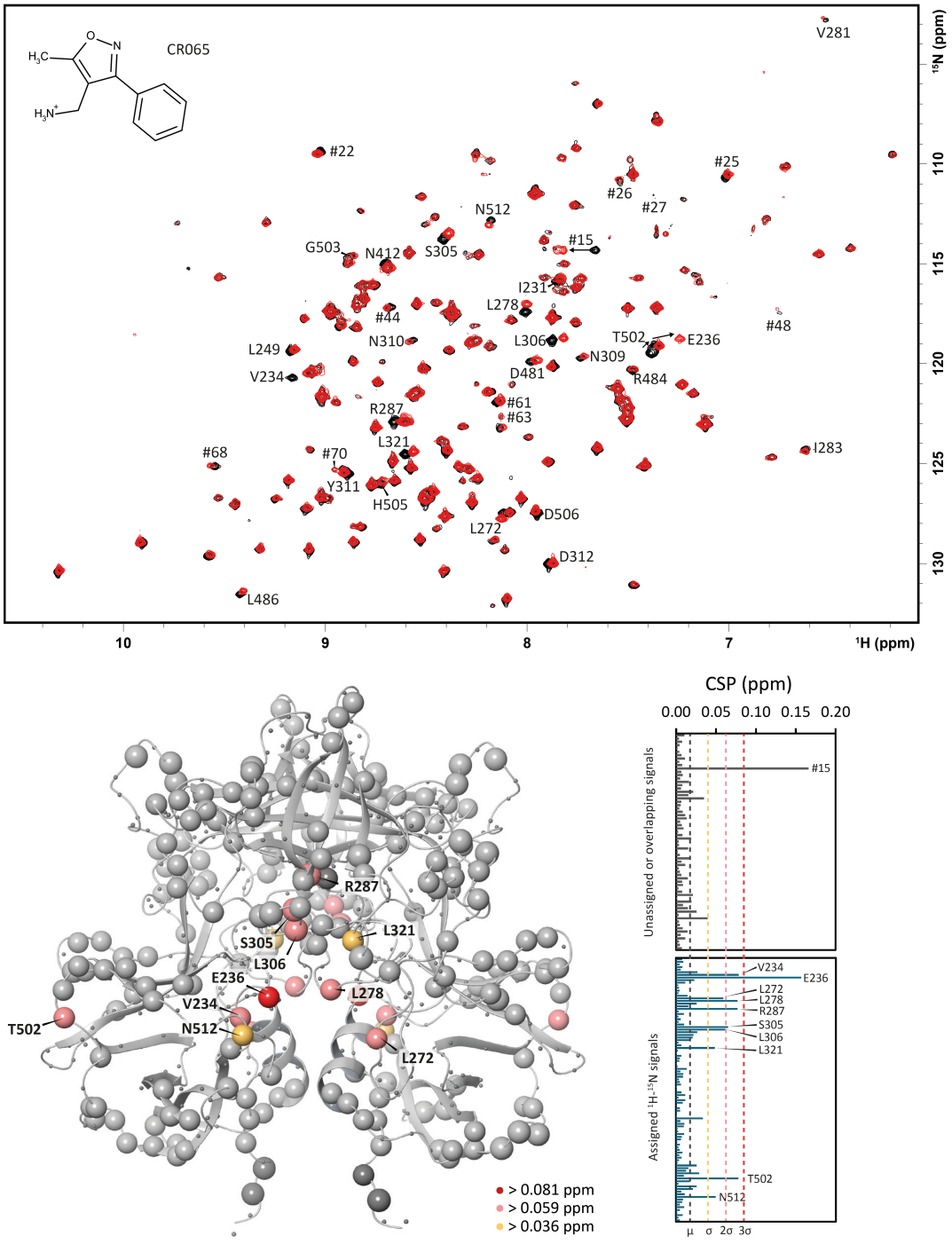


Figure 3.23 | Chemical shift perturbation (CSP) upon binding of CR065
Superposition of $^{1}H,^{15}N$-TROSY-HSQC spectra of 60 µM Saga P dimer in the absence (black) and presence (red) of ~6.4 mM fragment **CR065**. The measurement was conducted in 75 mM sodium phosphate buffer (pH 7.3) and 100 mM sodium chloride. In order to prevent pH artefacts, a constant pH was monitored by ^{1}H signals of 300 µM imidazole as described in section **3.3.3** and shown in Appendix A7/**Figure A23**. NMR signals that shifted more than a Euclidian distance of 0.015 ppm are annotated. The partial backbone assignment is depicted in Appendix A7/**Figure A22**. Unassigned HSQC signals are serially numbered (e.g., #25). A representation of a crystal structure of Saga P dimer (PDB: 4WZK) is depicted below. Backbone nitrogen atoms associated with chemical shift perturbations of NH signals upon binding of **CR065** are represented as balls. Gray balls: Euclidian distance of less than the sum of mean (µ) and one standard deviation (σ) (0.036 ppm); yellow balls: NMR signals shift more than 0.036 ppm; bright red balls: NMR signals shift more than 0.059 ppm (µ+2σ); red balls: chemical shifts of more than 0.081 ppm (µ+3σ). Small gray dots represent amino acids, which were not assigned at the time of writing. The right diagram shows Euclidian distances of 65 unassigned NH signals (gray bars) and 101 assigned NH signals (turquoise bars).

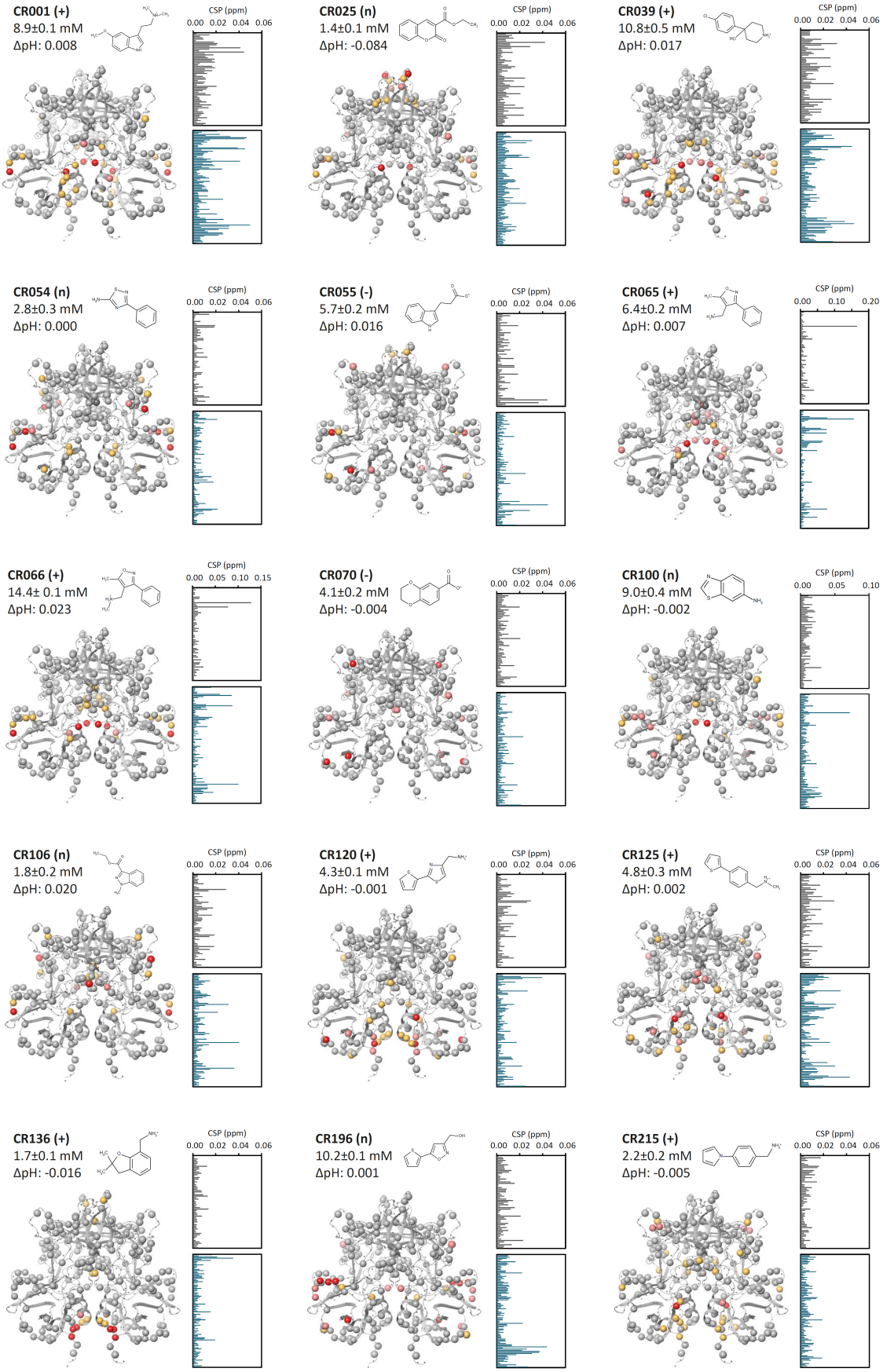


Figure 3.24 | Chemical shift perturbation (CSP) of fragments, which bind to Saga P dimers
This figure extends over three pages. The description is located on page 143.

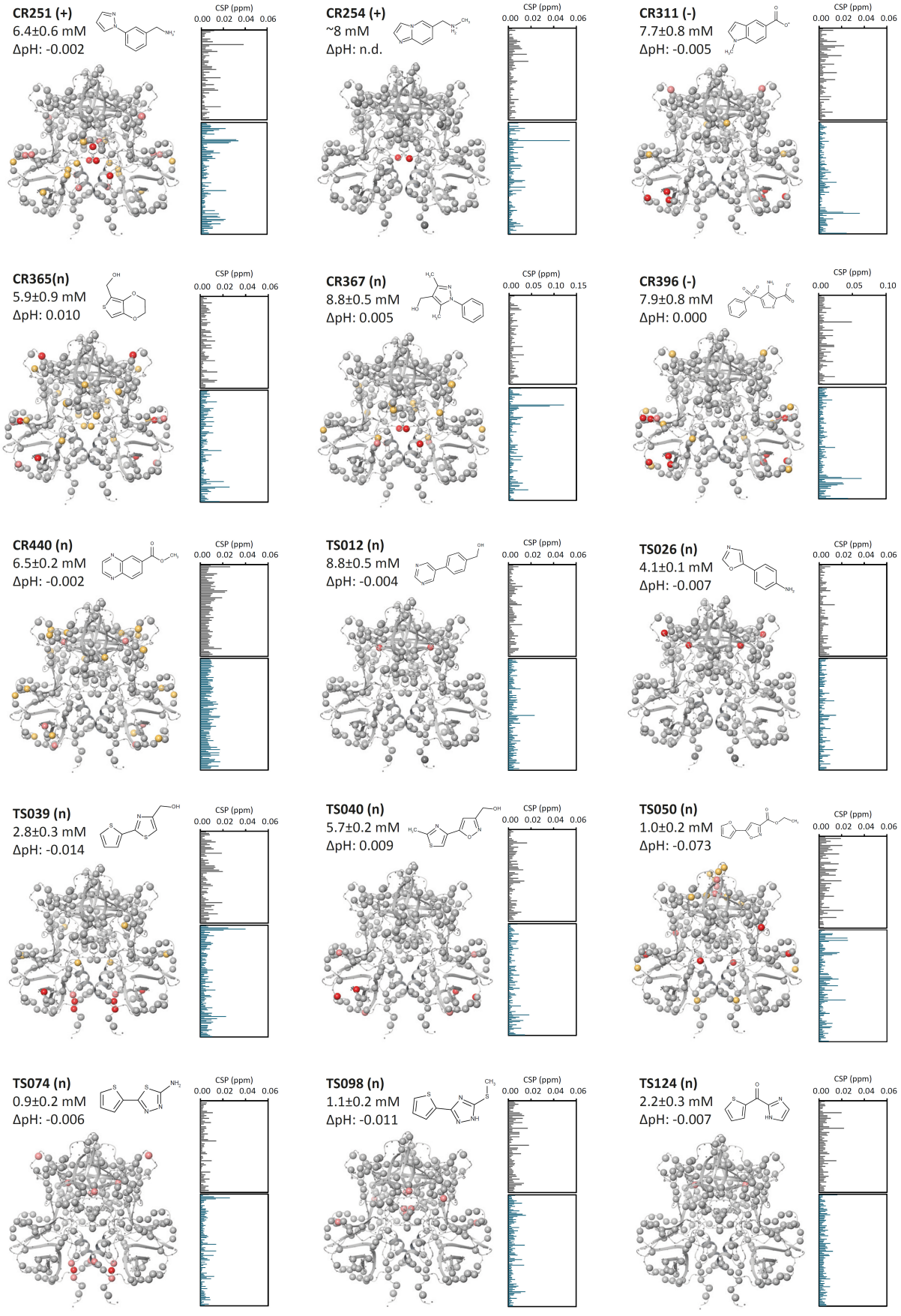


Figure 3.24 | Continued.

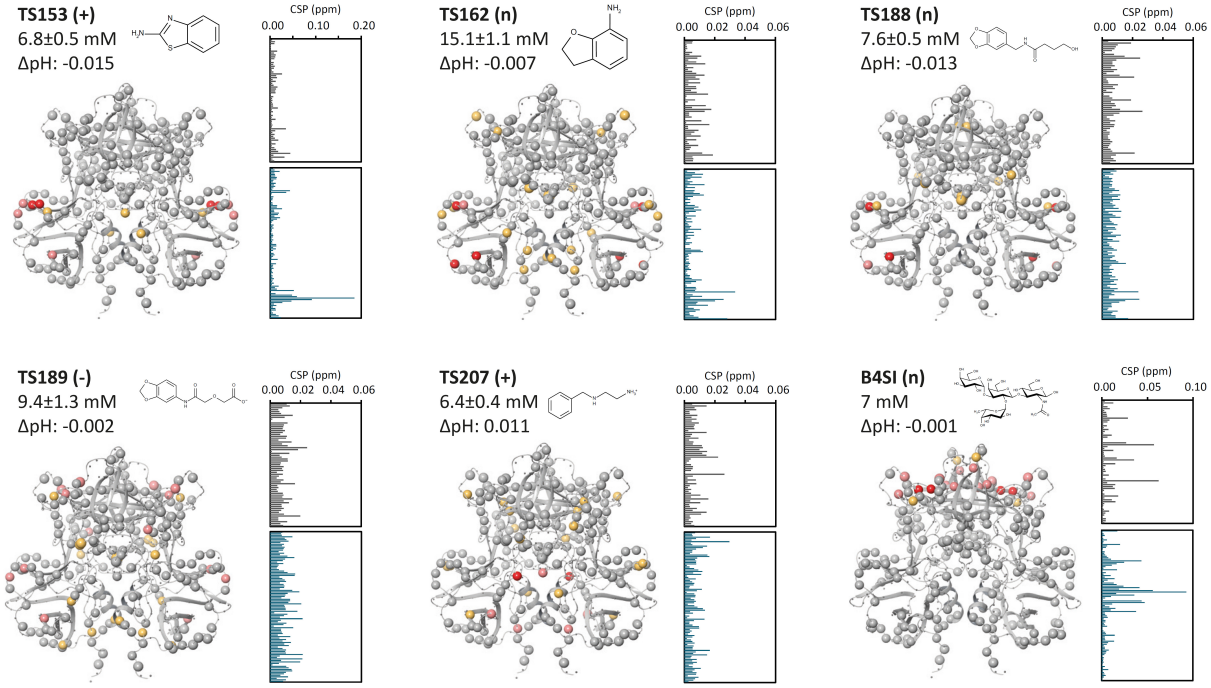


Figure 3.24 | Chemical shift perturbation (CSP) of fragments, which bind to Saga P dimers
The compound ID along with its charge in brackets, the fragment concentration, the pH difference (ΔpH), and the chemical structure are depicted above each crystal structure of the Saga P dimer (PDB: 4WZK). The procedure of CSP experiments and pH control is described elsewhere (**Figure 3.23** and section **3.3.3**). In the crystal structure, colored balls represent backbone nitrogen atoms whose NH signals shift more than the sum of the mean and one (yellow), two (bright red), or three standard deviations (red). The diagram beside each crystal structure shows Euclidian distances of 65 unassigned NH signals (gray bars) and 101 assigned NH signals (turquoise bars). In order to compare the CSP profiles, the maximum Euclidian distance was set to 0.06 ppm, except for **CR065**, **CR066**, **CR100**, **CR367**, **CR396**, and **TS153**. The superimposed ^{1}H,^{15}N- TROSY-HSQC spectra of 60 µM Saga P dimers in the absence and presence of each fragment are shown in Appendix A7. For comparison only, chemical shift perturbations caused by 7 mM B antigen tetrasaccharide type 1 (B4SI) (K_D ~ 100 µM) are also shown. Raw data were kindly provided by Dr. Alvaro Mallagaray.

Surprisingly, only a few fragments caused chemical shift changes of amino acid residues near the canonical fucose binding site (e.g., **CR440** or **TS189**), and furthermore, none of the CSP profiles of these fragments were comparable with that of B antigen tetrasaccharide type 1 that undoubtedly binds exclusively to the glycan binding sites (see **Figure 3.24** last scheme). In order to detect potential pH artefacts, previous CSP data were correlated with spectral changes upon pH shifts (see **Figure 3.25**). This analysis suggested that small pH changes in the CSP assay buffer above 0.05 units can lead to misinterpretation of chemical shift mapping. Therefore, the results of fragments **CR025** and **CR050** have to be interpreted with great care, and thus these fragments should be treated as false positives. The ionic strength is another reason that may cause “dummy” chemical shift perturbations. However, a comparison of CSP profiles derived from fragment binding and from a titration of NaCl (see Appendix A7; **Figure A24**) did not give any indication of ionic-strength-dependent chemical shifts in the case of charged compounds.

When comparing all chemical shift mapping models, it is apparent that some regions are affected more often than others. To identified “hotspots” of perturbations, the Euclidian distances of 33 fragments (**CR025**, **TS050**, and B4SI excluded) were averaged and plotted onto the crystal structure of the Saga P dimer.

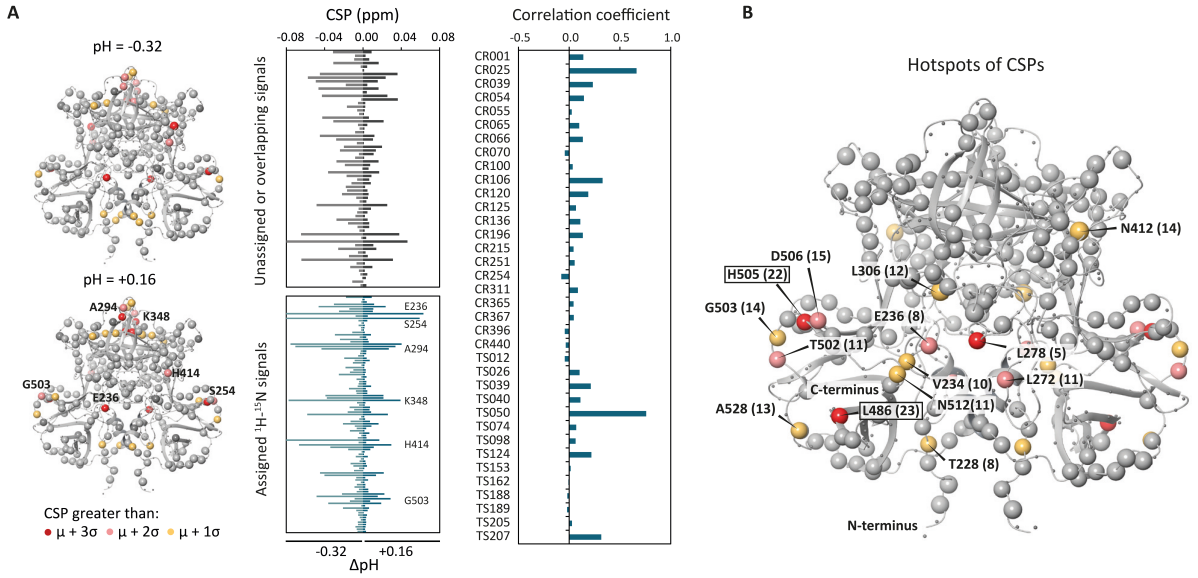


Figure 3.25 | Evaluation of chemical shift perturbation (CSP) experiments
(A) Chemical shift distances of a pH titration with 80 µM Saga P dimers in 100 mM sodium phosphate buffer (Creutznacher et al. 2017, unpublished data) were plotted onto the crystal structure of Saga P dimer (PDB: 4WZK). The reference pH was 7.26. Colored balls represent backbone nitrogen atoms. Thresholds for color-coding are defined by the sum of the mean (µ) and one (yellow), two (bright red), or three standard deviations (σ) (red). The CSP profile besides shows the Euclidian distances of 65 unassigned NH signals (gray bars) and 101 assigned NH signals (turquoise bars). The negative signs for CSPs of ΔpH = -0.32 indicate that NH signals move in opposite directions along a straight line at lower and higher pH values. Superposition of ^{1}H,^{15}N-TROSY-HSQC spectra are depicted in Appendix A7. In order to identify pH artefacts, the CSP profiles of pH shifts were compared to the CSP profiles of fragments by calculating the Pearson's correlation coefficient of Euclidian distances. In the case of fragment **CR025** and **TS050**, the pH difference was -0.083 and -0.073 pH units, respectively, as compared to the reference sample containing only 60 µM of Saga P dimers. (B) The scheme represents the average chemical shift difference for each NH signal of 33 fragments (**CR025** and **TS050** excluded), which were plotted onto the crystal structure of the Saga P dimer. The color code is the same as in (A). The value in brackets beside each amino acid is the number of fragments, at which the NH signal of the corresponding backbone nitrogen atom shifted more than 0.015 ppm.

The greatest averaged perturbations were found for Leu 278, Leu 486, and His 505 (see **Figure 3.25**). No obvious linear relationship between chemical shift changes of the residues and small pH variation was observed (data not shown). Since the residues Leu 486 and His 505 are notably affected in 22 to 23 cases, and not in all cases a direct interaction can be assumed, these region embedding the C-terminus seems to be indirectly perturbed rather than through direct binding. Nevertheless, some fragments may exclusively bind in proximity to that region (e.g., **TS153**).
Figure 3.26 gives an overview of putative binding sites for all fragments analyzed. Suggestions that were made for fragments, which may bind to the glycan binding site (region 1 in **Figure 3.26**), should be treated with care because of their ambiguous CSP pattern. About half of fragments seem to interact with region 2a/b, which comprises well-defined and deep pockets. These cavities represented by Leu 278 or Ile 231 are predominantly defined by hydrophobic and negatively charged residues, whilst the glycan binding region rather contains polar amino acid residues. Moreover, almost every negatively charged residue is accompanied by a positively charged residues, rendering region 1 practically neutral. Interestingly, apart from a few exceptions, fragments that bind to region 2a/b are positively charged. In contrast, fragments that appear to perturb region 3 (e.g., His 505) are mainly neutral or negatively charged.

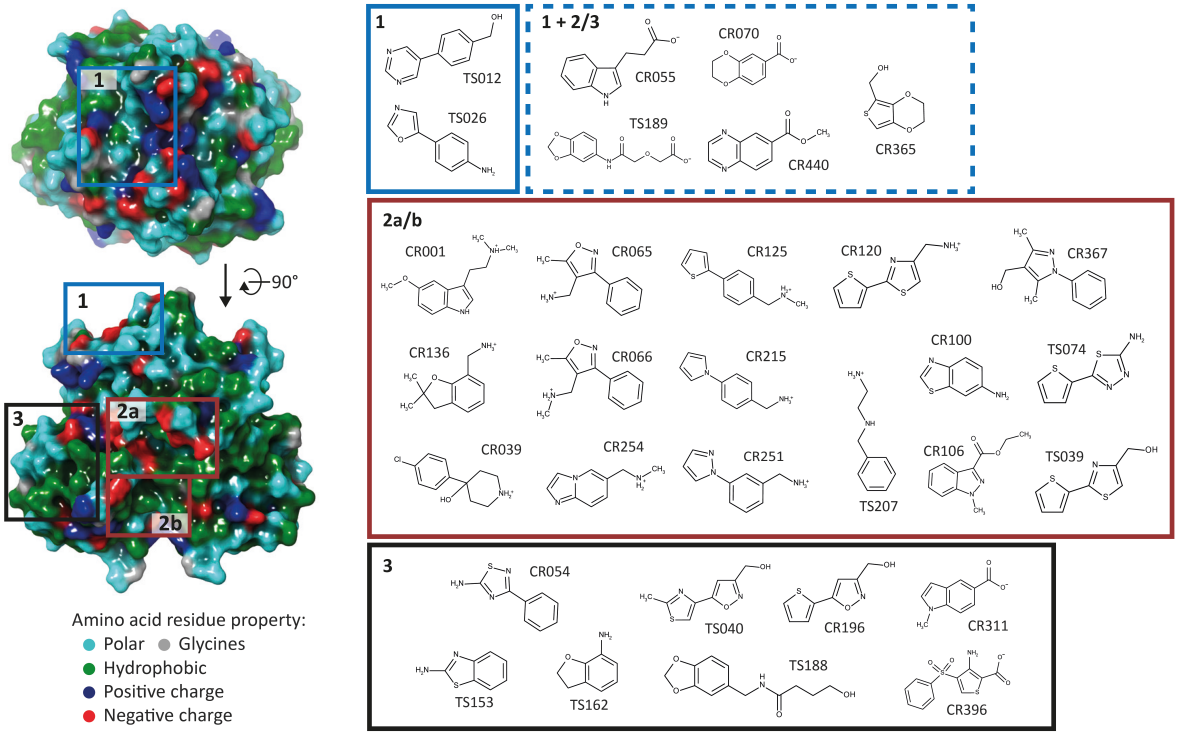


Figure 3.26 | Binding pocket prediction based on chemical shift perturbation experiments
The surface representation of the Saga P dimer (PDB: 4WZK) was created with Maestro 11 after applying the protein preparation wizard. It represents different properties of amino acid residues. In addition, four putative binding regions are highlighted. Fragments that assumedly interact with at least one of these binding sites are depicted on the right side. However, based on CSP experiments the evidence is weak that the upper fragments (**TS012** – **CR365**) bind to region 1 since the chemical shift changes are very small (below 0.015 ppm) or the intensity of shifting NH signals is just above noise level (e.g., V388 for **TS026**). The majority of fragments that seem to bind in pocket 2a or 2b are positively charged, whereas the binding region predominantly embraces hydrophobic and negatively charged amino acid residues.

3.4.6 Virtual Docking

Nowadays, virtual fragment screening is extensively and routinely used in drug discovery research[100,145,210,259–262]. In this study, molecular docking was performed to guide the interpretation of experimental results, and furthermore, to assess the quality of binding pocket prediction for particular fragments. Therefore, fragments from the libraries CR2006 and TS2017 were docked into different binding pockets of the crystal structures of both Saga and MNV-1 P dimers using Glide as implemented in Maestro 11. In the case of Saga P dimers, docking models were compared to data from chemical shift perturbation experiments. Finally, suggestions of binding pockets were made for MNV-1 P dimers based on this gathered knowledge.

Assessment of Binding Site Prediction for Saga P Dimers

Following *in silico* preparation of both crystal structures and fragments, 742 fragments were docked into three different receptor grids in parallel to cover almost all binding pockets of the protein. According to the scoring function used in Glide, the best ~30% of fragments are depicted in **Figure 3.27**. Up to three different energy-minimized conformers per fragment were accepted.

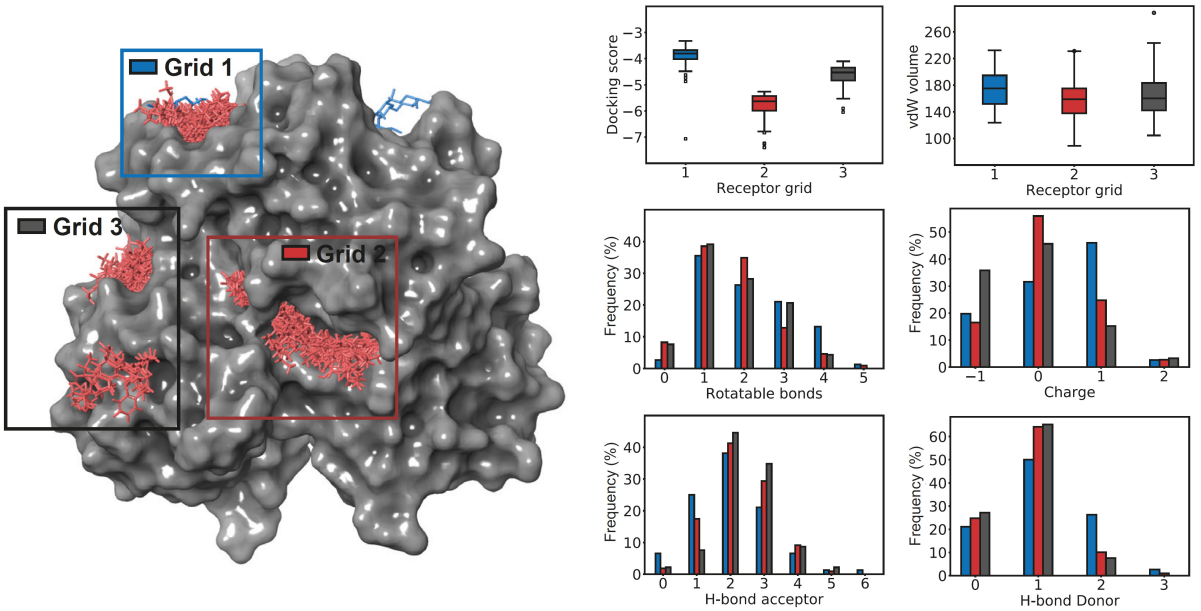


Figure 3.27 | Virtual screening of the Maybridge fragment libraries CR2006 and TS2017 targeting the Saga P dimer
Fragments were docked and scored as described in the material and method section **3.3.6** using Glide as implemented in Maestro 2017. In total, 742 fragments from the Maybridge library CR2006 and TS2017 were subjected to the virtual screening pipeline using the Saga P dimer as a receptor (PDB: 4WZK, resolution 1.49 Å), which is represented as close-up surface. According to the extra precision scoring function, the best 257 fragments (partially including conformers) are displayed in red stick models. Of those, 76 fragments are docked into or adjacent to the binding pocket of H type 2 trisaccharide (blue stick model). Plots on the right side show the distributions of the docking score and scalar descriptors of fragments that were docked into the corresponding grid.

Surprisingly, fragments that were docked into grid 2 (with Leu 278 being the center) exhibited the best docking scores (even without structural water molecules in region 1 or 3; see below), indicating that the cleft provides structural requirements for relatively tight binding fragments. This may explain why so many fragments have been identified to interact with this protein region. According to the docking results, there is no clear preference for the molecular size or flexibility among the three different grids. However, fragments that preferentially bind to the glycan binding site tend to have more rotatable bonds and larger van der Waals volumes. Additionally, grid 1 or 2 seem to accept neutral or positively compounds more often, whilst receptor grid 3 attracts either negatively charged or neutral fragments. In terms of hydrogen bonding, fragments interacting with grid 1 seem to have less hydrogen bond acceptors but on the other hand more hydrogen bond donors. Albeit the opposite picture is true when the proposed structural water molecules that are normally present in the crystal structure had been fully removed before conducting molecular docking as this is discussed later (see section **3.3.6**). In this scenario, grid 1 seems to bind only neutral or negatively charged compounds with more H-bond acceptors and less H-bond donors. This aspect does not account for grid 2 as no structural water molecules are present in this region. By default, Maestro defines water to be structured if more than one H-bond is formed between co-crystallized water and protein residues or bound ligands.

In order to evaluate the scoring function applied for small molecules, α-L-fucose was also docked into the grids 1 to 3. Since it has widely been reported that water molecules often mediate protein-

carbohydrate interactions[263–266], structural water molecules had not been removed. Satisfactory, the best docking score was obtained for α-L-fucose bound to its canonical binding site (cf. **Figure 3.28**). Moreover, 21 fragments from CSP experiments, which enable an adequate prediction of binding sites, were separately docked into regions 1, 2a, 2b, and 3 (cf. **Figure 3.26**), including structural water molecules if present. One example is given in **Figure 3.29** for fragment **CR065** (see below). In this case, the docking model explains the observed chemical shift perturbations, for example, for Glu 236, Leu 272, Leu 278, and Leu 306 quite well. These residues seem to be directly involved in ligand binding. Other residues that apparently interact with **CR065** based on the structural model are unassigned. The chemical shift changes of Arg 287 and Val 234 indicate that this region undergoes conformational alterations upon ligand binding, because these residues are 10 to 14 Å apart but show notable chemical shift changes. Besides this aforementioned binding pocket, another interaction mode is proposed by docking into region 2b. In contrary to the experimental data, this alternative pocket seems to be favored as the *docking score* of -4.68 is about one unit better, and hence a ~five-fold increased affinity should be expected. However, it is important to note that most residues of this second pocket seemingly involved in binding of **CR065** were not assigned at the time of writing. Since several unassigned signals significantly shifted (e.g., #15, #25, or #68; see **Figure 3.23**), the completion of ^{15}N assignment may give evidence for the existence of this alternative pocket for **CR065**. The *docking score* alone would be otherwise an inappropriate predictor for the most realistic docking pose that is in accordance with experimental findings. The so-called *Emodel* is another measure to assess binding poses for a given ligand. In the case of **CR065**, the usage of the *Emodel* scoring function gives a suitable prediction of the binding pocket that is most reasonable based on existing CSP data.

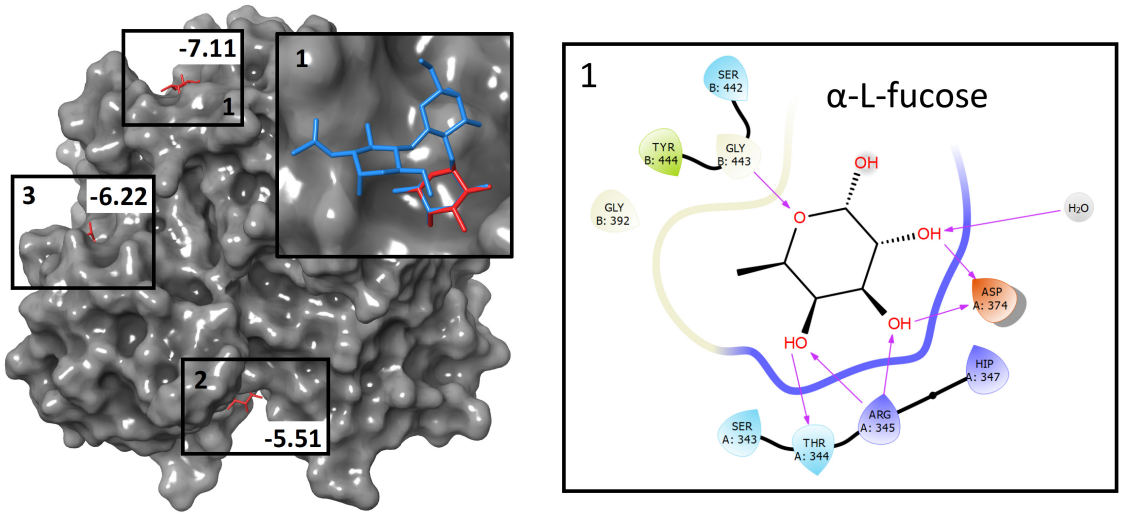


Figure 3.28 | Docking of α-L-fucose into different receptor grids of the Saga P dimer
The best three docking poses are depicted. The numbers represent the corresponding docking scores in kcal mol^{-1}. The best docking score was obtained when docked into the glycan binding site (grid 1) in the presence of structural water molecules. The docked α-L-fucose (red) almost perfectly adopts the position of the fucose unit of H type 2 trisaccharide, which is present in the crystal structure.

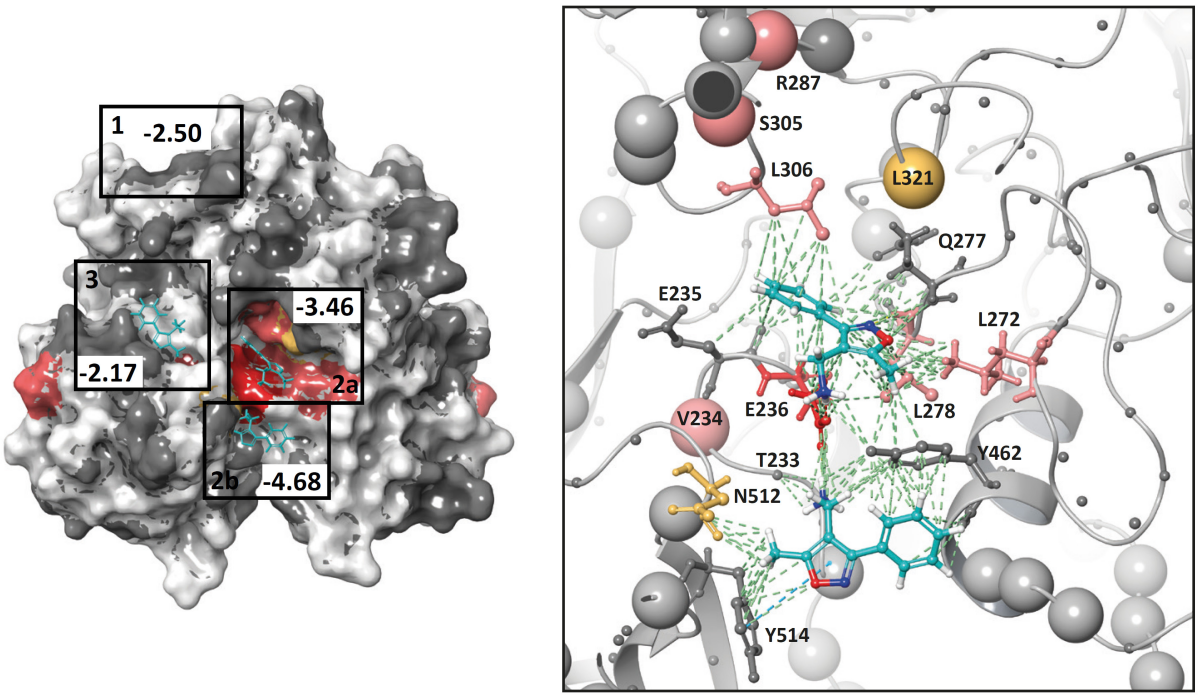


Figure 3.29 | Docking of CR065 into putative binding sites of the Saga P dimer
The *docking scores* for different binding modes are depicted in white boxes. Color-coding of the surface representation is the same as described in **Figure 3.23**. Red reflects large, light red medium, yellow small chemical shift changes, and gray practically no changes. White surfaces represent unassigned residues. On the right side, an enlarged docking model for region 2a and 2b is shown. Amino acid residues that are involved in ligand binding are presented as stick model. For the remainder only backbone nitrogen atoms are predicted as balls. For a better representation, N512 is also shown as a small dot. Small gray dots represent amino acids, which were not assigned at the time of writing, and large gray balls represent amino acids that were practically unaffected upon ligand binding.

The docking results of the remainder of fragments are summarized in **Figure 3.30** and ordered according to the binding region. Moreover, a comprehensive comparison of CSP data and docking results are summarized in **Table 3.4** on page 154. Of note is that the *docking score* and the *Emodel* were similar or even worse for grid 1 and grid 3 when water molecules had been removed prior to docking (data not shown). Overall, a relatively good agreement of docking results and CSP data was found. Whilst the *docking score* provides better predictions for the fragments **CR120** and **CR136**, the best docking poses of **CR001**, **CR065**, **CR066**, or **CR125** based on *Emodel* values explain the experimental findings much better. For the fragments **CR100, CR251**, **CR254**, and **CR367**, both predictors deliver reasonable results in accordance with CSP data. On the other hand in some cases (e.g., **CR196** or **TS162**), none of the ranked poses that had been selected based on the above-mentioned scoring functions would satisfactorily explain the observed chemical shift changes. In conclusion, the combination of chemical shift perturbation experiments and *in silico* docking may guide the selection of fragments that will be further optimized. In about half of the cases, the prediction of the approximate binding region was considerably good, independently of the predictor used. The *Emodel*, though, slightly outperforms the *docking score*. Nonetheless, the evaluation based on both predictors was most valid. Finally, the docking models substantiate results from CSP experiments and from virtual screening, namely that region 2a is most frequently affected probably due to its depth and biochemical characteristics.

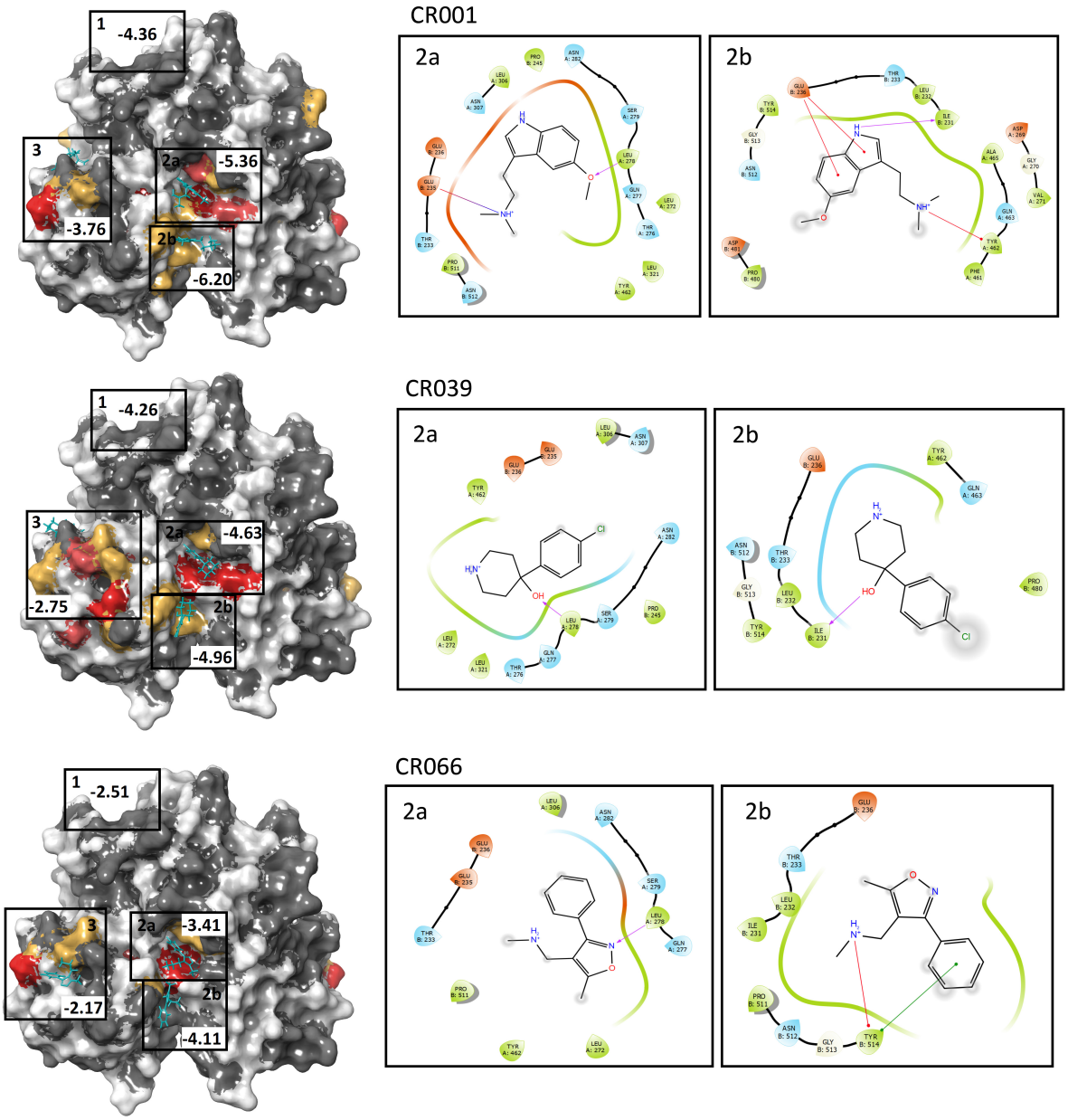


Figure 3.30 | Comparison of chemical shift perturbation experiments and docking results
Docking was performed as described in section **3.3.6.** For simplification, only *docking scores* are depicted. The corresponding values of *Emodel* are stated in **Table 3.4** on page 154. Different colors represent the magnitude of perturbations: red = large, light red = medium, yellow = small and gray = no changes. White represents unassigned residues. Ligand interaction diagrams are depicted for grids with highest docking scores. The color of drops represents the property of the present amino acid. Hydrogen bonding is indicated by purple arrows. Green lines with two dots at the end represent π-π interactions, and red lines denote ionic-π interaction. Solvent-exposed parts of a ligand are highlighted by gray circular shadows.

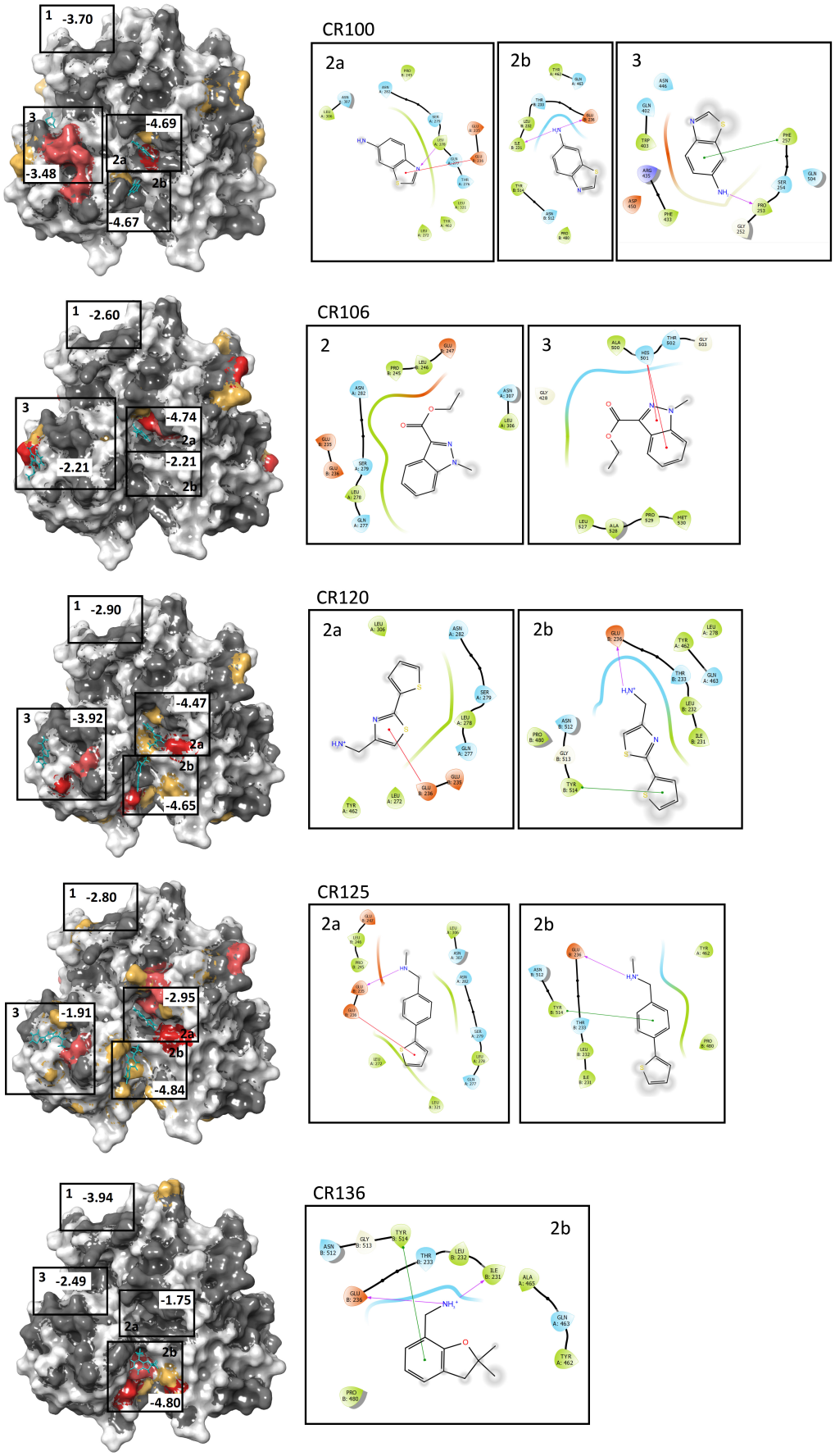


Figure 3.30 | continued.

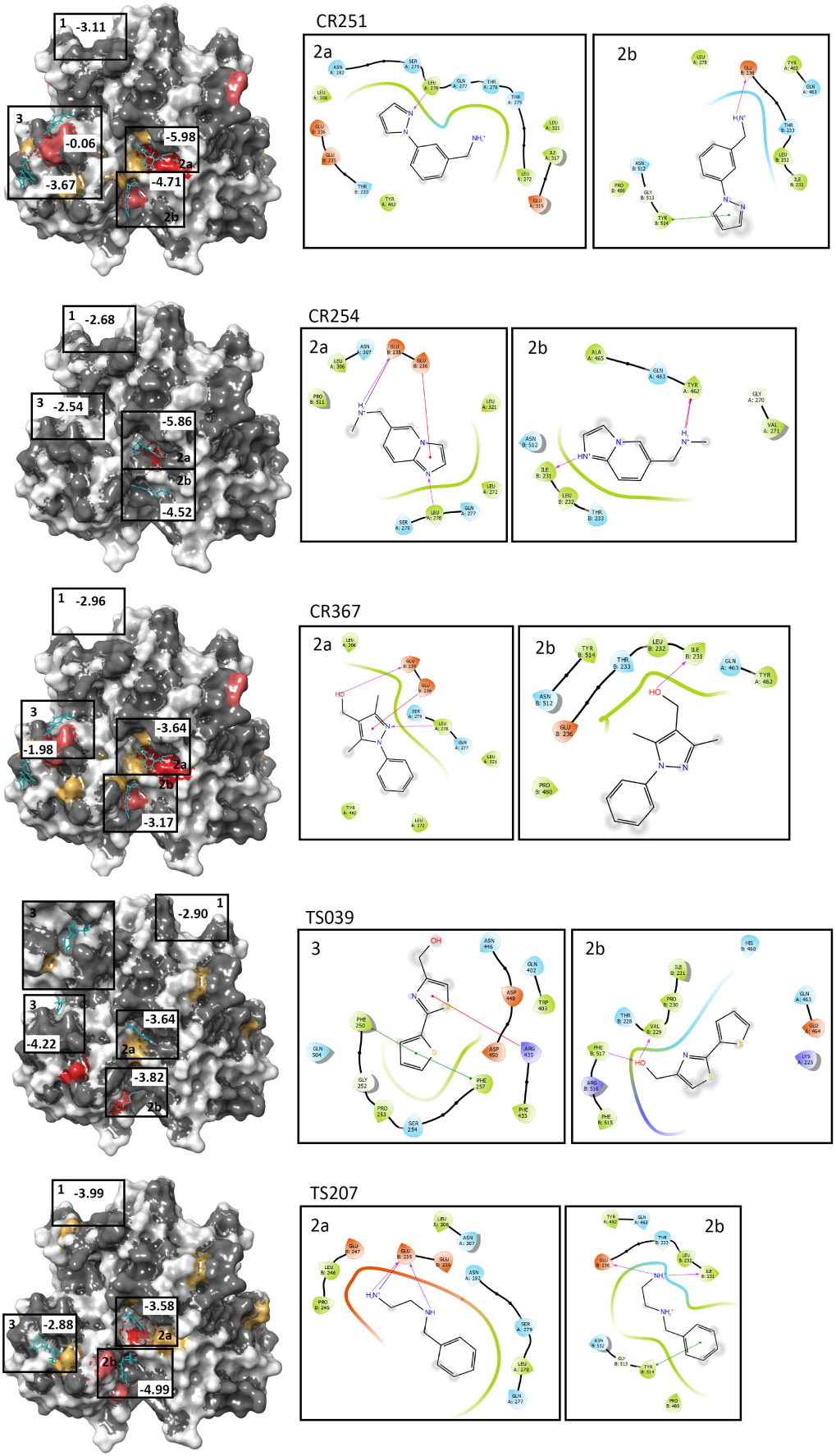


Figure 3.30 | continued.

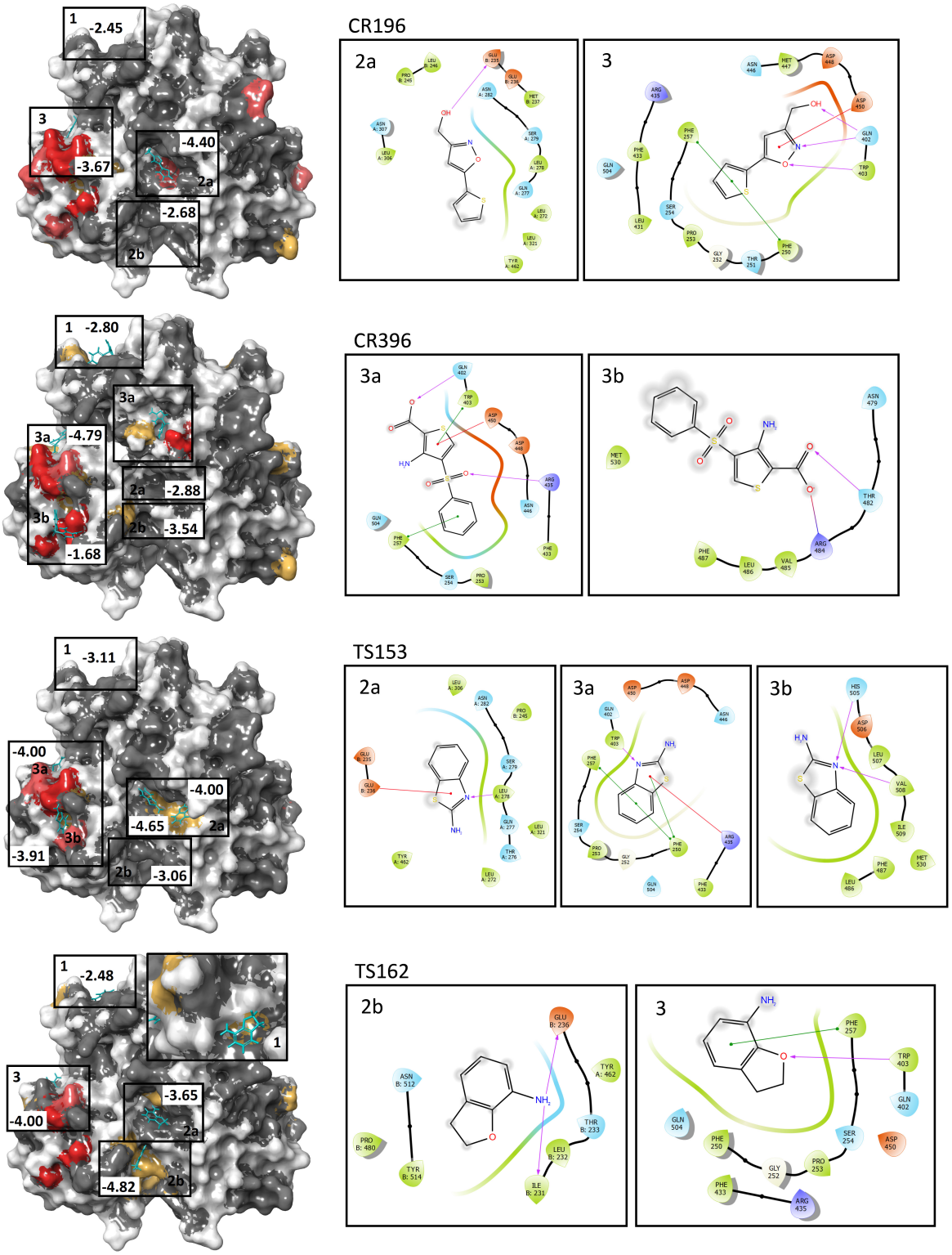


Figure 3.30 | continued.

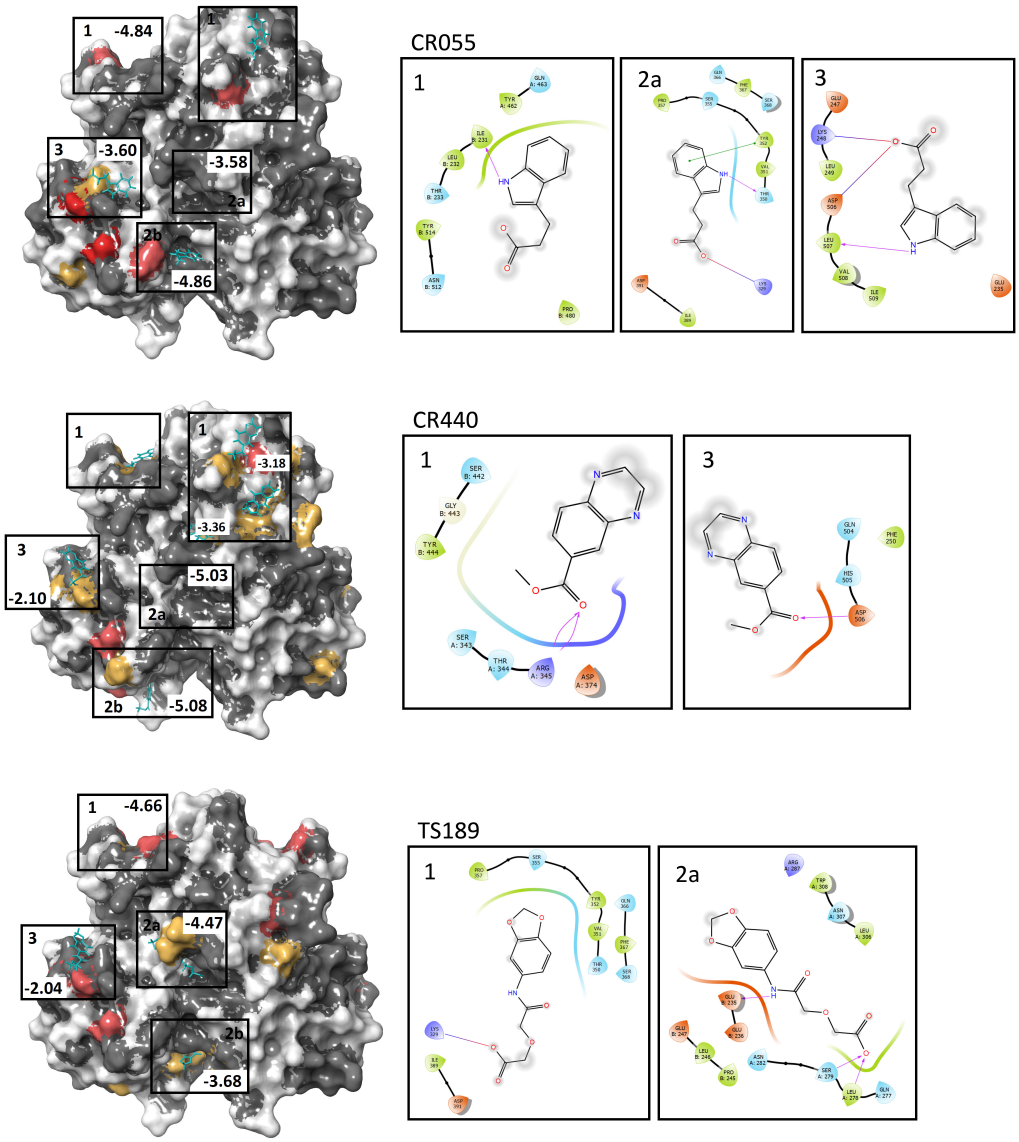


Figure 3.30 | continued.

Table 3.4 | Comparison of empirically determined binding sites (CSP regions) and molecular docking
The second column contains plausible binding regions according to chemical shift perturbation (CSP) experiments or X-ray data (L-fucose) in descending order (most to least probable). For each grid, the *docking score*, *GlideScore*, and *Emodel* values are stated for the best two docking poses per compound. Definitions can be found in section **3.3.6**. For the assessment of the quality of prediction, the following notations are used: -, no experimental evidence; +, poor; ++ acceptable/ second binding site possible; +++ most reasonable. (Score = docking score or GlideScore)

Compound	CSP region	Docking score/ GlideScore/ Emodel												Quality of prediction	
	I II III	Grid 1			Grid 2a			Grid 2b			Grid 3			Score	Emodel
L-fucose	**1** - -	-7.1	-7.1	-40.4	-5.3	-5.3	-30.0	-5.6	-5.6	-30.1	-6.2	-6.2	-27.3	+++	+++
		-7.1	-7.1	-40.6	-4.6	-4.6	-31.9	-5.3	-5.3	-30.7	-6.2	-6.2	-27.5		
CR001	**2a** 2b (3)	-4.4	-4.4	-23.1	-5.3	-5.3	-41.1	-6.2	-6.2	-34.2	-3.7	-3.7	-27.5	++	+++
		-4.0	-4.0	-26.2	-3.4	-3.4	-36.9	-6.2	-6.2	-34.5	-3.4	-3.4	-27.4		
CR039	**2a** 2b (3)	-4.3	-4.3	-28.9	-4.6	-4.6	-32.0	-5.0	-5.0	-30.3	-2.8	-2.8	-20.2	++	+++
		-3.9	-3.9	-31.6	-3.4	-3.4	-33.9	-5.0	-5.0	-30.1	-2.7	-2.7	-21.4		
CR055	**(1)** - -	-4.8	-4.8	-34.7	-3.6	-3.6	-22.4	-4.9	-4.9	-20.4	-3.6	-3.6	-20.2	+	+++
		-4.4	-4.4	-30.9	-3.6	-3.6	-26.9	-4.3	-4.3	-24.6	-3.0	-3.0	-17.1		
CR065	**2a** 2b (3)	-2.5	-2.5	-26.2	-3.5	-3.5	-38.2	-4.7	-4.7	-27.3	-2.2	-2.2	-23.7	+	+++
		-2.3	-2.3	-28.2	-3.2	-3.2	-32.4	-4.6	-4.6	-27.4	0.3	0.3	-22.4		
CR066	**2a** 2b (3)	-2.5	-2.5	-27.7	-3.4	-3.4	-37.3	-4.1	-4.1	-28.0	-2.2	-2.2	-18.1	+	+++
		-2.3	-2.3	-27.9	-3.3	-3.3	-35.6	-4.0	-4.0	-27.8	-1.7	-1.7	-19.0		
CR100	**2a** (3)	-3.7	-3.7	-27.2	-5.3	-5.3	-22.5	-4.7	-4.7	-22.0	-3.5	-3.5	-22.7	+++	+++
		-3.0	-3.0	-22.5	-4.7	-4.7	-27.9	-3.1	-3.1	-20.5	-3.3	-3.3	-21.6		
CR106	**2a** 3 -	-2.6	-2.6	-22.5	-4.7	-4.7	-36.0	-2.2	-2.2	-22.0	-2.2	-2.2	-23.3	++	++
		-2.5	-2.5	-22.4	-4.5	-4.5	-36.1	-2.2	-2.2	-23.8	-2.0	-2.0	-22.0		
CR120	**2a** 2b (3)	-2.9	-2.9	-30.4	-4.5	-4.5	-33.1	-4.7	-4.7	-25.1	-3.9	-3.9	-26.5	+++	++
		-2.7	-2.7	-27.3	-2.3	-2.4	-31.4	-3.6	-3.6	-28.3	-2.4	-2.4	-24.7		
CR125	**2a** 2b -	-2.3	-2.3	-27.1	-2.9	-3.0	-32.7	-4.8	-4.9	-27.3	-2.1	-2.1	-25.0	+	+++
		-2.0	-2.0	-25.2	-2.8	-2.8	-35.7	-4.7	-4.7	-25.2	-1.9	-1.9	-22.1		
CR136	**2b** - -	-2.7	-2.7	-25.1	-3.7	-3.7	-28.2	-4.8	-4.8	-25.2	-2.5	-2.5	-19.1	+++	-
		-2.2	-2.2	-21.2	-2.9	-2.9	-26.2	-4.6	-4.6	-19.7	-2.5	-2.5	-19.9		
CR196	**3** (2a) -	-2.5	-2.5	-25.3	-4.4	-4.4	-32.7	-2.7	-2.7	-30.1	-3.7	-3.7	-31.9	+	+
		-2.5	-2.5	-23.5	-4.2	-4.2	-33.9	-2.4	-2.4	-25.6	-3.7	-3.7	-29.9		
CR251	**2a** (2b) (3)	-3.1	-3.1	-28.0	-6.0	-6.0	-30.2	-4.7	-4.7	-26.3	-3.8	-3.8	-24.7	+++	+++
		-2.9	-2.9	-28.1	-5.4	-5.4	-30.5	-4.4	-4.4	-26.6	-2.4	-2.4	-21.4		
CR254	**2a** - -	-3.9	-4.0	-28.0	-5.9	-5.9	-33.8	-5.1	-5.2	-24.6	-2.5	-4.3	-22.6	+++	+++
		-3.9	-4.0	-23.4	-5.1	-5.2	-31.6	-4.8	-4.8	-25.4	-2.3	-2.3	-21.6		
CR367	**2a** (2b) (3)	-3.0	-3.0	-26.2	-3.7	-3.7	-31.5	-3.5	-3.5	-28.9	-2.0	-2.0	-26.3	+++	+++
		-2.9	-2.9	-28.7	-3.6	-3.6	-36.6	-3.4	-3.4	-30.0	-1.5	-1.5	-21.0		
CR396	**3** (2a) (1)	-2.8	-2.8	-23.9	-2.9	-2.9	-31.9	-3.5	-3.5	-31.6	-4.8	-4.8	-23.6	++	-
		-2.6	-2.6	-24.0	-2.9	-2.9	-31.1	-3.4	-3.4	-30.9	-2.8	-2.8	-21.5		
CR440	**1** (3) -	-3.4	-3.4	-24.5	-5.0	-5.0	-36.4	-5.1	-5.1	-26.3	-2.1	-2.1	-15.8	-	-
		-3.2	-3.2	-26.2				-4.8	-4.8	-23.2	-2.0	-2.0	-16.5		
TS039	**2b** - -	-2.9	-2.9	-24.0	-3.6	-3.6	-29.2	-3.8	-3.8	-31.6	-4.2	-4.2	-31.2	-	-
		-2.8	-2.8	-25.6	-3.5	-3.5	-30.2	-3.5	-3.5	-29.8	-3.0	-3.0	-33.7		
TS153	**3** 2a (2b)	-3.1	-3.1	-22.2	-4.6	-4.6	-28.3	-3.1	-3.1	-21.1	-4.0	-4.0	-21.4	++	++
		-3.1	-3.1	-21.6	-4.1	-4.1	-24.1	-2.8	-2.8	-20.4	-3.7	-3.7	-23.4		
TS162	**3** (2b) -	-2.5	-2.5	-22.9	-3.7	-3.7	-26.3	-4.8	-4.8	-20.4	-4.0	-4.0	-21.7	+	-
		-2.4	-2.4	-21.8				-4.6	-4.6	-20.3	-4.0	-4.0	-24.3		
TS189	**1** 2a 3	-4.7	-4.7	-39.7	-4.5	-4.5	-40.7	-3.7	-3.7	-32.8	-2.0	-2.0	-19.7	++	++
		-4.4	-4.4	-41.4	-4.2	-4.2	-39.7	-3.4	-3.4	-37.1	-2.0	-2.0	-18.5		
TS207	**2a** 2b -	-4.0	-4.2	-27.8	-3.8	-5.0	-32.4	-5.0	-6.3	-27.9	-2.9	-4.1	-31.8	++	++
		-3.9	-4.1	-31.3	-3.6	-3.7	-32.3	-4.9	-5.0	-23.4	-2.6	-2.8	-19.1		

Prediction of Binding Pockets of MNV-1 P Dimers

In order to get structural models that may explain the inhibition mode of the fragments **CR054**, **CR120**, **CR196**, or **TS039**, both virtual screening and molecular docking were conducted for MNV-1 P dimers as well. The same protein sites were targeted as for Saga P dimers, whereby almost all binding pockets were covered in parallel. Results from virtual screening are shown in **Figure 3.31**. The overall *docking score* is more negative and hence better as compared to Saga P dimers. This aspect may explain why the hit rate was more than twice as much. Surprisingly, the best *docking scores* have been found in region 1 and 3. The range of molecular size and flexibility of fragments that had been docked is comparable among all grids. At this point, the general distribution pattern of charges should be highlighted, as it matches the experimental data derived from STD NMR screening (compare **Figure 3.14** and **Figure 3.31**). Fragments docked into grid 1 are rather neutral or positively charged and, additionally, grid 2 predominantly accepts neutral compounds, whilst grid 3 seems to attract positively charged molecules according to molecular docking. Moreover, fragments bound to region 3 tend to have more H-bond acceptors as compared to fragments bound to grid 2 or grid 3, whereby in the latter case two or less H-bond acceptors and only one H-bond donor appear to be favored.

To approximately predict the binding pockets of fragments that showed significant inhibitory efficacy in the MNV-1 inhibition assays, these fragments were docked into the above-mentioned receptor grids along with 3'-sialyllactose. The results are summarized in **Figure 3.32, Figure 3.33**, and **Table 3.5** (see below). Since the conformation of 3'-sialyllactose in the bound state is not known, the docking was performed with full ligand flexibility. A comprehensive study by Taube

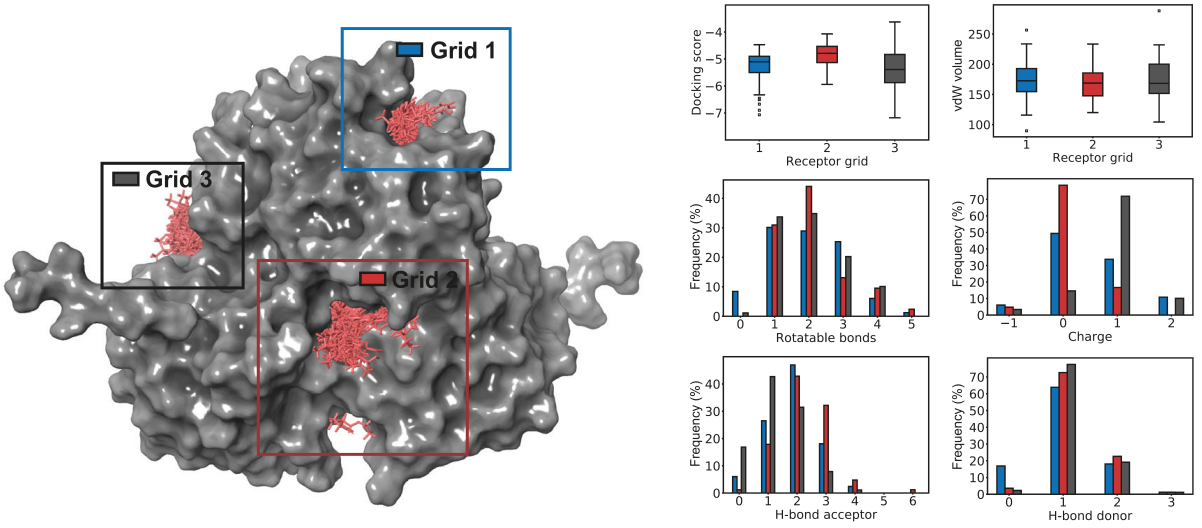


Figure 3.31 | Virtual screening of fragments of the Maybridge libraries CR2006 and TS2017 against MNV-1 P dimers
742 Fragments were docked and scored as described in the material and method section **3.3.6** using Glide implemented in Maestro 2017. According to the extra precision scoring function, the best 238 fragments are displayed in red stick models, in which 83 fragments are docked into or adjacent to the putative binding pocket of GM3 (Grid 1). For grid generation, the inner core of the X-ray structure was excluded by applying spherical exclusion volumes. The size of grids can be found in Appendix A8. (Abbreviations: vdW, van der Waals; H-bond, hydrogen-bond)

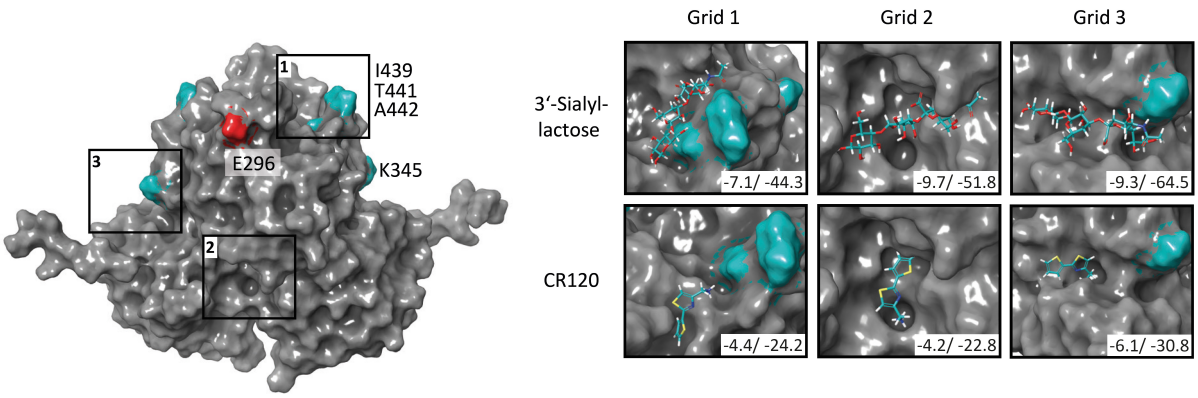


Figure 3.32 | Docking of 3'-sialyllactose and CR120 into different binding sites of the MNV-1 P dimer
P domain dimers of MNV-1.CW1 (PDB 3LQ6) are shown as a 3D-reconstructed surface model. Proposed binding epitopes for sialic acid-containing glycans (see REF[214]) are colored in turquoise. The capsid proteins of MNV-1.CW3 and MNV-1.CW1 differ in only one amino acid, K296E[230] (red). Receptor grids are highlighted as black boxes. On the right side, different docking poses are depicted for three binding sites. Numbers in white boxes represent the *docking score* and *Emodel* values.

and co-workers (see REF[230]) have suggested a conformational glycan binding epitope in the capsid protein of the MNV-1.CW1 strain comprising amino acids Lys 345, Ile 439, Thr 441, and Ala 442 (see **Figure 3.32**) by using mutagenesis and infection assays. Moreover, it has been proposed that the amino acid Glu 296 may also contribute to the host attachment site linked to terminal sialic acid binding (e.g., via ganglioside GD1a or GM3). Unexpectedly, the highest *docking score* was obtained for the best docking pose of 3'-sialyllactose bound to grid 2, which is far away from the supposed glycan binding site. The best *Emodel* values, though, were found for region 3 containing the amino acid Lys 345. Whereas the worst scores were found for grid 1. It is of importance to mention that no structural or conserved water molecule was left after the preparation of the crystal structure using *protein preparation wizard* as implemented in Maestro. Since water molecules often mediate contacts between carbohydrates and its receptor as stated above, and discussed later, the overall docking results would have been different if docking were conducted in the presence of tightly bound water.

In the previous section, molecular docking and protein-observed NMR data were in good agreement, in particular when both predictors, the *docking score* and the *Emodel*, were high and CSP data were solid. In some cases, the *Emodel* could explain the experimental data much better. Therefore, docking of fragment **CR120** strongly suggests that region 3 comprises the "main" binding pocket of **CR120**. In this model, the interaction involves salt bridges or H-bonds between the amino group of **CR120** and Asp 403 as well as Asp 450 of the MNV-1 P domain coupled with a T-shaped π-π stacking of the thiazole ring of **CR120** and Phe 401 (see **Figure 3.33**). Interestingly, assuming that both 3'-sialyllactose and **CR120** interact with this same binding region would readily explain the proposed competition derived from STD NMR experiments.

Binding of the other potent fragments to this pocket 3 would also be proposed if the *Emodel* values are used for binding pocket prediction (see **Table 3.5**). In addition, 2D ligand interaction diagrams were created by Maestro and are depicted in **Figure 3.33**. It is striking that in all cases the

interaction with pocket number 3 seems to be mediated by H-bonding, including Asp 403 and π-interaction with Phe 401. However, these binding models give no obvious indication why fragment **CR294** or **CR107** showed only moderate or no inhibitory potency. Additional methyl groups linked to the rings and to the amino groups may be the main reason. These groups lead to steric clashes with amino acid residues of the MNV-1 P domain as reported by the 3D model in Maestro (data not shown and, by default, not stated in the 2D ligand interaction diagrams).

Table 3.5 | Prediction of fragment binding sites of MNV-1 P dimers based on virtual docking

Compound	Docking score/ GlideScore/ Emodel			Predicted binding site according to	
	Grid 1	Grid 2	Grid 3	Docking score/ Glide Score	Emodel
3'-Sialyllactose	-7.1 -7.1 -44.3 -6.1 -6.1 -46.0	-9.7 -9.7 -51.8 -9.4 -9.4 -49.4	-9.3 -9.3 -64.5 -8.1 -8.1 -60.7	2	3
CR054	-3.1 -3.1 -27.0 -1.1 -1.1 -23.6	-2.8 -2.8 -17.4 -2.6 -2.6 -21.6	-3.1 -3.1 -31.0	1/3	3
CR107	-4.7 -4.7 -26.9 -4.2 -4.2 -27.4	-3.1 -3.1 -22.5 -2.3 -2.3 -23.0	-4.5 -4.5 -31.0 -3.2 -3.2 -31.2	1	3
CR120	-4.4 -4.4 -24.2 -4.4 -4.4 -24.8	-4.2 -4.2 -22.8 -2.3 -2.3 -22.5	-6.1 -6.1 -30.8 -5.2 -5.2 -30.4	3	3
CR196	-3.6 -3.6 -26.2 -3.6 -3.6 -27.4	-4.9 -4.9 -25.4 -4.8 -4.8 -25.3	-2.5 -2.5 -31.3	2	3
CR294	-4.9 -4.9 -25.4 -4.7 -4.7 -24.5	-3.0 -5.1 -25.5 -2.7 -2.7 -25.4	-6.0 -6.0 -40.1 -5.8 -5.8 -37.1	3	3
TS039	-3.3 -3.3 -25.7 -3.1 -3.1 -26.2	-4.2 -4.2 -20.5 -4.0 -4.0 -26.0	-2.7 -2.7 -34.4 -2.7 -2.7 -31.1	2	3

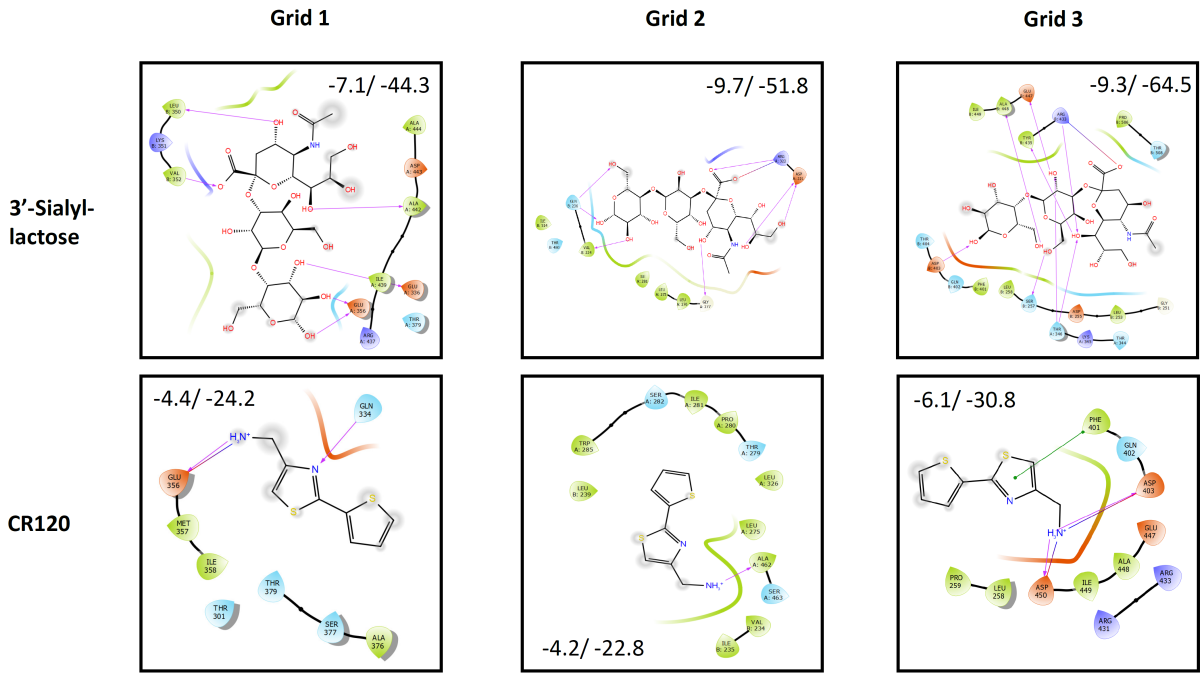


Figure 3.33 | 2D ligand interaction diagrams of fragments docked into several binding sites of the MNV-1 P dimer
Docking was performed as described in section **3.3.6.** Negative numbers represent the *docking score* and *Emodel* values. The colored spheres represent properties of the present amino acids. Hydrogen bonding is indicated by purple arrows. Green straight lines represent π-π interactions, and red-blue lines symbolize electrostatic interactions. Solvent-exposed parts of a ligand are highlighted by gray circular shadows.

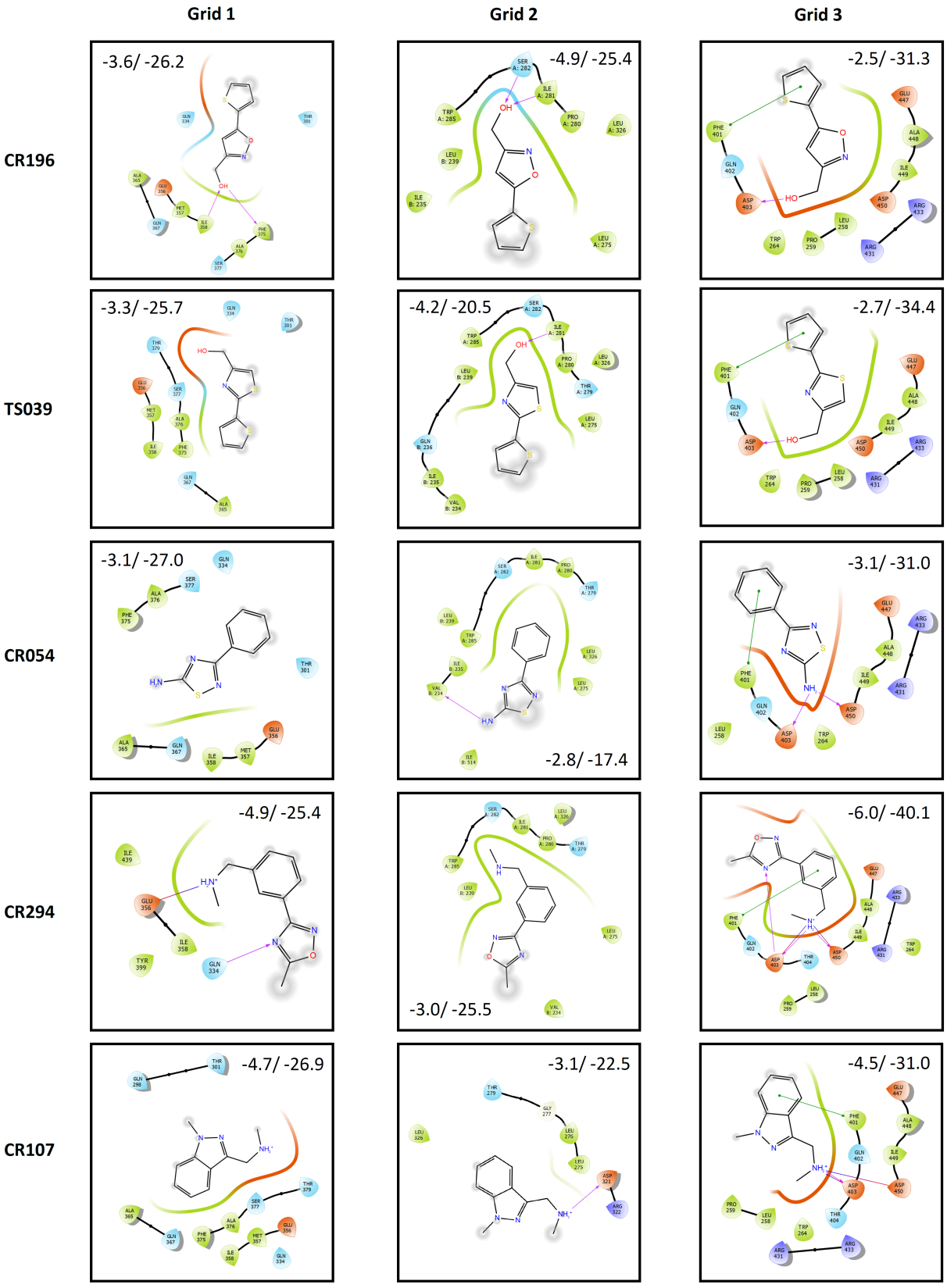


Figure 3.33 | 2D ligand interaction diagrams of fragments docked into several binding sites of the MNV-1 P dimer (Continued)

3.5 Discussion

In the first part, guidelines and precautions for the design of fragment libraries are presented to reduce the number of false positive hits and hence to increase the quality of a fragment screen. Next, a possible reason is given for the aggregation of MNV-1 dimers. Moreover, it is attempted to explain the phenomenon that the fraction of negatively charged compounds is unexpectedly small compared to the remainder of the positive hits derived from STD NMR experiments. Subsequently, several assays that resulted in promising hits for MNV-1 P dimers and Saga P dimers are discussed. And last but not least, the impact of virtual screening and docking on lead discovery is highlighted.

3.5.1 Fragment Library Design

It is believed that well-constructed fragment libraries coupled with comprehensive quality controls heavily increase the success of a fragment screening projects[267–269]. Several factors and guidelines have been debated to be crucial. One prominent guideline established in 2003 implies the "Rule of Three (Ro3)" in which the molecular weight is ≤300 Da, the number of H-bond donors is ≤3, the number of H-bond acceptors is ≤3, cLogP values are ≤3 and total polar surface area (TPSA) ≤ 60 Å^2 (see REF[181]). Although enthusiastically adopted by the scientific community[267], the set of filters to ensure high-quality libraries have been extended during the last decades. General rules are listed below:

(1) Ensure structural diversity to cover a biological relevant chemical space in terms of shape, pharmacophores, and synthetically accessible growth vectors[267].
(2) Thereby, avoid groups known to be associated with a high reactivity[197], or fragments that aggregate in solution, which would lead to false positive data[270],
(3) and confirm that the library contains soluble fragments (here ≥2 mM)[271].

Molecular Size and Shape

Since NMR has shifted the limit of detection from micromolar to the millimolar range, the size and hence the complexity of fragments can be reduced to 140 to 230 Da. Siegal et al. even proposed a lower limit of 150 Da[269]. Therefore, the optimal heavy atom count is assumed to be around twelve. The average molecular weight found in the libraries CR2006 and TS2017 is ~190 Da (±30 Da). Thus, an appropriate size distribution is given. In regard to molecular shape, it has been advocated for increasing the 3D character of fragments, which are included into any fragment library. It is assumed that the occupancy of the pocket volume is higher for less planar fragments, resulting in better affinities. This may be especially beneficial for challenging targets[267,272]. Nevertheless, Fuller et al. have argued that completely flat fragments are often developed into complex compounds towards 3D shapes during optimization process anyway[273]. To assess the flatness of a molecule, the fraction of sp^3 atoms (fsp^3, number of sp^3 hybridized carbons/total carbon count) was considered as a simple descriptor, but this has been rejected as it does not adequately describe the 3D character[268].

Nonetheless, fsp^3 may still be a useful descriptor because the group of Lovering et al. (REF[274]) demonstrated that a high fsp^3 above 0.4 and the presence of chiral centers correlate with success in inhibitors transition from initial hits to approved drugs. It is obvious that this is highly dependent on the drug target of interest and should be treated as a rough guideline. Moreover, a positive correlation between carbon bond saturation (i.e., fsp^3) and solubility has been revealed[274].
More sophisticated shape descriptors are DFP[268] (deviation from the plane), PBF[275] (plane of best fit), or PMI[276] (principle moments of intera). Briefly, the energy-minimized conformation of compounds are determined by using CORINA and 3D coordinates are used to calculate the plane of best fit. The article of Firth et al. (see REF[275]) provides a comprehensive mathematical description, but no web browser-based or command line program is freely available. A rough approximation of the 3D character of fragments originated from the libraries can be made by the use of the ChemAxon software. **Figure 3.34** shows the distribution of four categories from planar fragments towards more 3D-shaped fragments. 3D fragments are often complex high-molecular-weight compounds (e.g., **TS240**). The concept of *molecular complexity*[75] implies that most protein-ligand interactions in crystal structures are highly productive. The more complex the ligand the higher is the probability of unfavorable interactions[267]. But the inherent advantage of the screening of small compounds is believed to be the fact that positive hits predominantly engage in high-quality interactions. Otherwise, they would not been identified as binders[74,262]. Furthermore, about one third of insoluble fragments came from category 3 and ~90% of soluble fragments are flat (see **Figure 3.34**). These two above-mentioned aspects disagree with the dogma "more 3D", at least in our case.

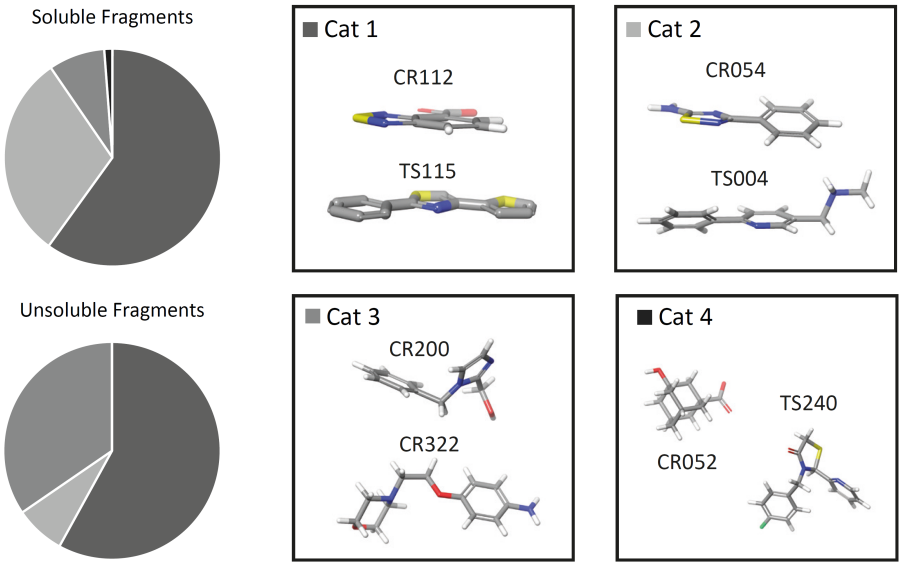


Figure 3.34 | Distribution of molecular shapes in the fragment libraries CR2006 and TS2017
The categories (cat) encompasses very flat molecules and molecules with considerable 3D character. The stick models of energy-minimized conformations were created with Maestro (Schrödinger) by using the ligand preparation tool.

Detailed lists of "unwanted" functionality can be found elsewhere (REF[197,277,278]). Baurin at al. proposed a list of favorable chemical groups to allow rapid chemistry and to increase polarity as well as solubility (see **Table 3.6**). Conveniently, the present libraries CR2006 and TS2017 consists mainly of compounds, which are functionalized with different groups, such as carboxyl, ester, hydroxyl or primary amino groups. To extend the repertoire of functionalities, the library TS2017 also contains more uncommon groups, such as S-CH_3, S-CH_2CH_3, RCSR, SO_2CH_3, NHC(=S)$NHNH_2$, NO_2, CH_2NHOH, or C(=NH)NHOH. Methoxy groups and decorated amino groups are still under-represented. An increase of the number of amide groups (i.e., $CONHCH_3$ or $CON(CH_3)_2$) should also be considered, as these groups mimic the protein backbone. Since many of the TS2017 fragments were collected from the Ro3 Maybridge library 2011 (www.maybridge.com) a "chemical cleanness" should be ensured.

Table 3.6 | List of favorable chemical groups (adapted from REF[197])

R-COOH	R-$COOCH_3$
R-OH	R-OCH_3
R-NH_2	R-$NHCH_3$
R-$CONH_2$	R-$CONHCH_3$ / R-$CON(CH_3)_2$
R-SO_2NH_2	R-SO_2NHMe / R-$SO_2N(CH_3)_2$
R-SCH_3	

Reactive fragments (such as aldehydes or *Michael* acceptors) often result in false positive hits in biochemical or biophysical assays. Therefore, a filter referred to as PAINS (pan assay interference compounds) has been introduced by Baell and co-workers[278,279] (a readily transferable Sybyl line notation is included in the reference). In total, 88 of TS2017 fragments were considered as "PAINS-free" according to the Maybridge's annotations. Following visualization, only the fragments **TS078**, **TS080**, **TS099**, **TS106**, and **TS235** should be rejected as they carry thiol groups. Aliphatic esters (R-CH_2COO-R) should be also examined due to their tendency to undergo hydrolysis at high or low pH[277].

To identify so-called "bad actors" Baell and co-workers tested ~90 000 compounds in six high-throughput (HTS) campaigns. Several problematic structural classes have been identified. These include derivatives of toxoflavin, isothiazolones, curcumin, hydroxyphenyl hydrazones, rhodanines, phenol-sulphonamides, enones, quinones, and catechols, among others (chemical formulas can be found in REF[279]). Fortunately, none of these aforementioned substructures are present in the current fragment library. Based on this study, other substructures should also be pointed out, including alkylated phenylpyrrole derivatives (e.g., **CR479**) or benzofurazan derivatives (e.g., **CR112** or **CR402**) (see REF[278]). For example, ammonium 7-chloro-2,1,3-benzoxadiazole-4-sulfonate is used for fluorescent labeling of protein thiol groups[280,281]. Several derivatives were also described to label other groups as well (e.g., amines, carboxylic acids, etc.). However, a sulfonate or nitro group at position 4 and a halogen at position 7 are required to be chemically active. Since all benzofurazan or benzothiadiazole derivatives present in the CR/TS

library are sparsely functionalized, the reactivity can be considered as very low. Of note is that the WNV protease NS2B-NS3^{pro} contains only one cysteine (residue 78) that is far away from the proposed binding sites of the fragments **CR112** and **CR402**.

To estimate the reactivity for thiol groups, the group of Hajduk has developed an *in silico* predictor[281] based on the so-called *Bayesian classifier*[282] using *Pipeline Pilot 5.1.0*. This publication also provides an excel spreadsheet of 175 substructures along with potentially reactive compounds. Surprisingly, **TS153** is one of these compounds. The results from NMR-based chemical shift perturbation experiments may be explained by the fact that **TS153** is covalently bound to Cys 489 of Saga P dimers. In order to confirm this hypothesis, the assay format ALARM NMR[283] (a La method assay to detect reactive molecules) could be utilized, which is based on DTT-dependent ^{13}C chemical shift changes of the human La antigen in the presence of a compound or fragment mixture. Apparently, any other isotopically labeled cysteine-containing protein can be used as well. In the case of Saga P dimers, binding of **TS153** to U-[^{2}H,^{15}N]-labeled protein should be tested again in the presence of increasing amounts of reducing reagents (e.g., DTT).

Besides covalent modification, interference in assay read-out or formation of aggregates of both fragments and proteins can also lead to a misinterpretation of binding or activity assays. Regarding the WNV protease NS2B-NS3^{pro}, the fragments **CR053**, **CR402**, **TS183**, and **TS235**, for example, have been identified to interfere with the AMC fluorescence. In the case of aggregation, it has been shown that some organic molecules form large colloid-like aggregates that sequester and hence inhibit enzymes or bind unspecifically to other macromolecules[284,285]. Many aggregators are typically hydrophobic, planar, rigid, and decorated with few polar groups[286]. Furthermore, these aggregates are often soluble in aqueous solution or very small and therefore not visually detectable. In order to sense fragment aggregates that can cause unspecific inhibition or promiscuous binding, several strategies have been proposed, such as detergent-dependency assays[198,284] (as described in section **2.5.2**), dynamic light scattering[284], and NMR methods[159,197,286].

Quality control using NMR offers a wide range of applications. NMR can be used, for example, to identify insoluble fragments by observing concentration dependent signal intensities of the ligands[159]. A more sophisticated strategy has been proposed by Baurin et al. or Dalvit et al. (see REF[197,286]). This approach requires the acquisition of a 1D reference spectrum, a WaterLOGSY spectrum or a selective longitudinal relaxation (R_1) filter spectrum of the fragment of interest in the presence of a well-known non-aggregator. Signal changes of the reporter molecule may then indicate an interaction with the surface of aggregates in a non-specific manner.

Alternatively, STD NMR can also be used to detect such behavior. Direct irradiation of a ligand in the absence of a protein may give an indication for aggregating compounds. In total, 50 fragments showed STD effects without protein (see Appendix A2). Some but not all have at least one methyl group with a NMR signal around 1 ppm. This can be one reason for direct irradiation as this is likewise seen for TSP or DSS (δ = 0 ppm). Of those, **CR001**, **CR100**, **CR125**, and **CR396** were investigated using NMR-based protein-observed binding assays. Indeed, the epitope mapping caused by **CR001**, **CR100**, or **CR125** is rather fuzzy. But this is also partially true for **CR039** or **CR106** that did not show any STD NMR artefacts. Nevertheless, a thorough examination of these fragments listed in Appendix A2 should be done in terms of aggregation. Finally, Harner and co-

workers suggested to remove promiscuous fragments that are found to be hits in all screens[159]. Only 14 fragments (**CR003**, **CR004**, **CR065**, **CR100**, **CR104**, **CR169**, **CR178**, **CR188**, **CR191**, **CR251**, **CR273**, **CR279**, **CR310**, and **CR382**) were classified as positive hits in STD NMR-based screens against WNV protease[168], NoV VLPs[90], and NoV P dimers (Saga, MI001, and MNV-1). It is not intuitively apparent whether this is because of a favorable pharmacophore, which promotes binding, covalent modification (see REF[281]) or due to an aggregation-related mode. Though CSP data of **CR065** and **CR251** suggest that at least these fragments behave like "good actors". In contrast, **CR100** or **CR273** should be treated with caution at high concentrations because of direct irradiation combined with a fuzzy CSP profile or unspecific binding above 1 mM as indicated by SPR.

Structural and Chemical Diversity

In order to cover many tracts of the so-called chemical space, a fragment library should consist of several chemical scaffolds functionalized with an acceptable set of chemical groups (see above). To identify gaps in the chemical space, SxG matrices introduced in this study are very useful (see section **2.3.3**). It is anticipated that simpler fragments will allow chemical space to be sampled more effectively[268]. Therefore, the existing complexity and size range of scaffolds is considered to be sufficient. Some chemical groups and molecular properties, however, are still under-represented as seen above. It is important to note that the number of fragments containing halogen atoms should not be increased, as halogen groups often decrease solubility[197], which is in accordance with our results. An illustration of the chemical space is given in **Figure 3.35** in terms of informative molecular descriptors such as molecular weight, charge, and hydrogen bond properties. Based on these distribution plots, a feature filter along with a PAINS filter and a Bayesian filter could be used to select fragments to be purchased. For example, I would search for small or medium-sized fragments (80 – 175 Da) with positively charged groups (e.g., secondary amino groups).

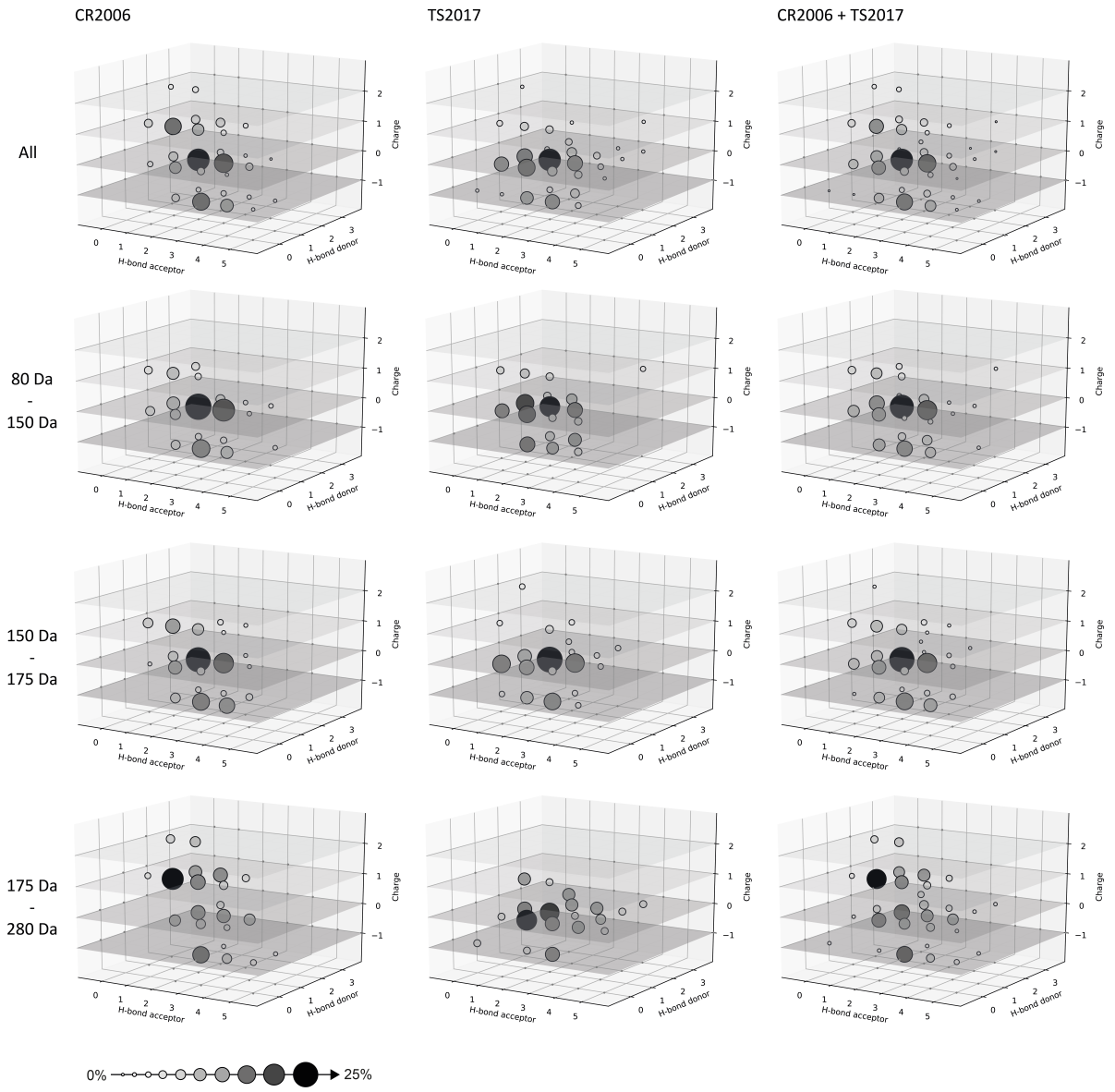


Figure 3.35 | Property distribution of fragment libraries
The 3D distribution plots were created using the Python script *3D_histo_A_D_Q.py*. The molecular properties of fragments were determined using the ChemAxon's calculator implemented in the in-house Python script *calcChemParam_Library.py.* The upper plots show distributions along the number of hydrogen bond (H-bond) acceptors/ donors and the total charge for all fragments of the respective library.

Solubility in Solution

For STD NMR screening, low solubility of about 50 - 200 µM is enough. But in the case of protein-observed NMR methods or crystallography, a high fragment solubility (≥2 mM) is required to ensure sufficient occupancy of the binding pocket, in particular for low-affinity compounds. To identify suitable molecular parameters that correlate with solubility a principal component analysis was conducted. Therefore, fragments from the CR/TS library were classified into four solubility categories (see **Figure 3.36**) based on visual inspection of 2 mM fragment stocks as well as NMR spectra. Beside aforementioned parameters, the PCA also included log S values and intrinsic log S_0

values, which were calculated by the solubility predictor[287,288] as implemented in ChemAxon. This analysis revealed that poorly soluble fragments predominantly cluster along both the log P/log D and log S/log S_0 axis. The distributions of these values are depicted in **Figure 3.37** on the next page, showing that 83% of fragments with log P or log D values below 2 are considerably soluble (≥ 2 mM). Hence, calculated log P values are reasonable predictors to classify soluble, low-molecular-weight compounds. But also log S or log S_0 values are appropriate predictors, albeit there is no clear linear correlation (Pearson's correlation ~ 0.4) between calculated and measured fragment concentrations (data not shown).

In conclusion, the present fragment CR/TS library used in this study delivered reasonable hits with moderate hit rates for the viral drug targets used. Nonetheless, a thorough analysis of fragments from both the CR2006 library and TS2017 library should be considered to sort out all "bad actors" that may increase the rate of false positives. This examination should involve testing of long-term stability in solution at RT, covalent modification, and aggregation behavior, at best under different pH values. Moreover, many computational tools such as PAINS filter or *Bayesian classifier* are available. However, implementing these *in silico* filters was beyond the scope of this thesis. Furthermore, not all nuisance compounds are known a priori, and hence an experimental quality control is still necessary.

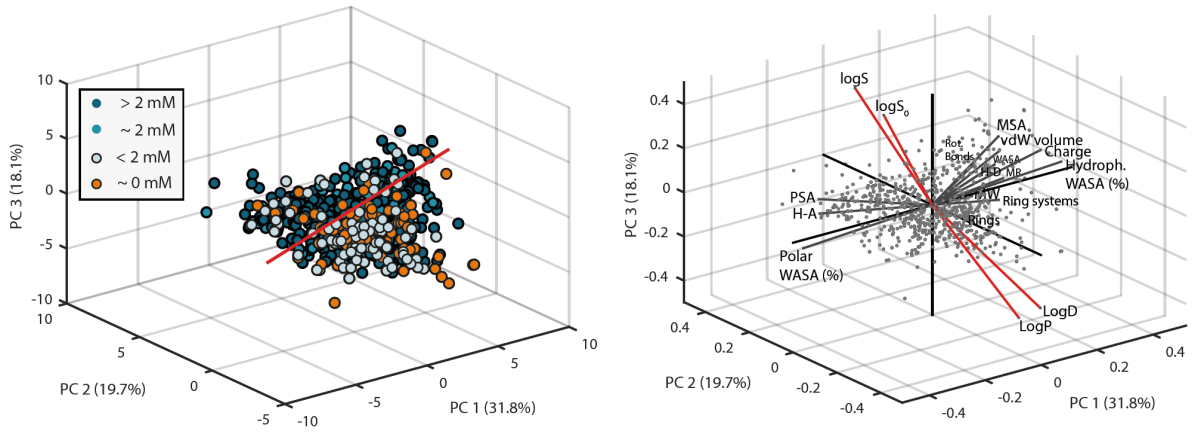


Figure 3.36 | Principal component analysis of structural and chemophysiological properties of fragments
The 3D scatter plot displays the distribution of CR/TS fragments along the first three principal components (PC). Dots are colored according to the solubility classification. On the right side, the same distribution is presented in a so-called biplot. Dots show fragments as gray points and parameters as gray vectors whose direction and length indicate how each variable contributes to the three principal components. (Abbreviations: MW, molecular weight; MSA molecular surface area; vdW, van der Waals; MR, molecular refractivity; WASA, water accessible surface area; and PSA, polar surface area)

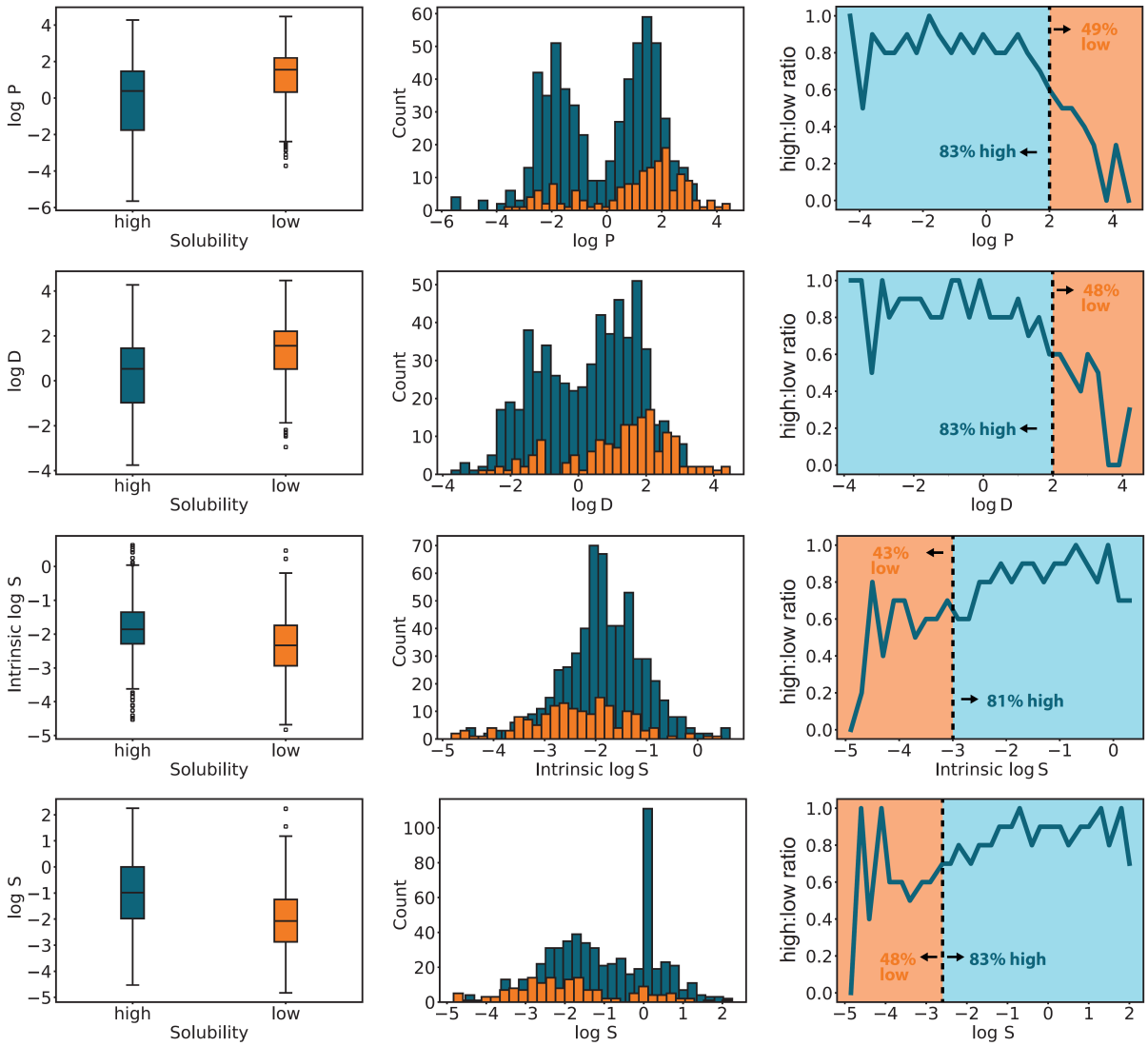


Figure 3.37 |The partition coefficient log P is a reasonable predictor of solubility
Box plots and histograms are shown for different solubility predictors. Solubility was considered as high when absolute no aggregates were detected in a 2 mM stock solution. Moreover, 1D NMR spectra of 1 mM fragment solution served as an indicator for poorly (low = orange) or highly (high = blue) soluble fragments as well. On the right side, the high-to-low ratios of binned values are shown. The threshold was set arbitrarily. These plots were created with the Python script *Solubility_Boxplot_and_Histogram.py*.

3.5.2 Fragment Screening

Aggregation of MNV-1 P Dimers

Following STD NMR screening of the Maybridge library CR2006, an unusually high hit rate of ~66% was found for MNV-1 dimers that were stored several months in 25 mM sodium phosphate buffer (pH 7.3) supplemented with 300 mM NaCl at 4 °C. Interestingly, a similar hit rate combined with high STD effects as observed for long-stored MNV-1 P dimers was obtained by Rademacher and co-workers (cf. REF[89,90]) in a previous fragment screening against Ast6139 virus-like particles (VLP). Although not detectable through conventional SDS PAGE, analysis via dynamic light scattering revealed that aggregation of MNV-1 P dimers might be the reason for this high hit rate. These aggregates derived from the same MNV-1 P domain construct have also been reported by other groups (see REF[289]). Tan and co-workers were the first that demonstrated that the P domain of VA387 strain forms distinct subviral particles[246]. Interestingly, this study revealed that the highly conserved arginine cluster (**RRR**) at the C-terminus of the P domain of human NoV promotes receptor binding and P particle formation[250]. A similar motif can also be found in the P domain of MNV-1 (**R**ML**K**). This RRR-cluster is missing in the constructs of Saga P dimers and MI001 P dimers used in this study. Hence, one reason for particle formation of MNV-1 P dimers could be the presence of the RMLK motif at the exposed C-terminus. On the other hand, 1D NMR spectra revealed partially unfolded proteins, indicating that particles represent unordered aggregates of the P domains. Noteworthy, using 1-1 echo 1D NMR experiments, an estimated molecular weight of about 65 kDa was obtained, which correspond to P dimers. This is in conflict with DLS data. Maybe this apparent molecular weight represents the average of partially unfolded proteins and protein aggregates.
Howsoever, in the absence of NaCl, the state of dimerization could be restored according to DLS measurements. Interestingly, this “low-salt aggregation phenomenon” has been observed in other studies as well[290,291]. In order to find the best condition for the storage of MNV-1 P dimers, different NaCl concentration (0, 150, and 300 mM) were tested. Moreover, two different protein tubes (conventional vs. LoBind tubes) and two different storage temperatures (4 °C vs. -60 °C) were analyzed. The parameters for the best condition were as follows:

(1) 25 mM sodium phosphate buffer (pH 7.3) at 4 °C,
(2) no NaCl,
(3) and the usage of LoBind protein tubes (Eppendorf).

Taken together, by using freshly prepared MNV-1 P dimers stored in suitable protein buffer, the hit rate could be reduced down to ~39%. However, this hit rate was still high and it could not be excluded that aggregation occurred during NMR measurement at RT in the presence of fragments (~2 mM total concentration). Albeit MNV-1 P dimers seemed to be invariable up to three days at RT when stored in phosphate buffer without salt.

Hallmarks of Positive Hits

Only 30 fragments were classified as hits for all three NoV targets. This is not surprising as the amino acid sequence identity of Saga P dimers and MNV-1 P dimers and of Saga P dimers and MI001 P dimers is 32.0% and 92.3%, respectively. Since it is believed that low-molecular-weight fragments engage in only a few but high-quality interactions with the protein pocket, a change of only one amino acid can make the difference. On the basis of molecular descriptors, a principal component analysis indicated that the molecular charge plays a crucial role for fragments that bind to NoV P dimers. Whilst only a small fraction of negatively charged compounds were classified as positive hits, neutral and positively charged fragments were predominantly found to bind P dimers, independently of the molecular scaffold. In the case of Saga P dimers, this could be explained by the fact that binding pocket 2 represented by Leu 278 (see **Figure 3.26**) seems to attract exclusively neutral or positively charged compounds. Second, the architecture of this binding region seems to be beneficial for fragment binding according to the high negative docking scores derived from virtual screening. Third, the density of methyl groups (from Ala, Ile, Leu, Thr, and Val; AILTV) within this pocket is noticeably high as compared, for example, to the binding pocket of fucose. Since it is assumed that protein signals at zero ppm are instantaneously saturated in a STD experiment, AILTV residues ensure efficient transfer of magnetization to bound ligands and hence increasing the likelihood of detecting the binding event by STD NMR[196,206,207]. However, **CR254** has not been detected during STD NMR screening as a positive hit although CSP data suggested binding to pocket 2. This can be explained most likely due to the poor signal-to-noise ratio (~3-4) of the STD NMR difference spectra. According to the crystal structure of the MNV-1 P dimer, AILTV methyl groups are more equally distributed, which may explain the increased STD effects and the high hit rate as compared to Saga P dimers, provided that protein aggregation did not take place. Astonishingly, the overall distribution pattern of molecular charges for binders of MNV-1 P dimers that were identified experimentally was similar to that found for hits derived from virtual screening. Pocket 3, for example, comprises three negatively charged residues (Asp 403, Glu 447, and Asp 450) that seem to attract mainly positively charged compounds, whereas pocket number 1 on the top the MNV-1 P dimer is rather neutral and hence more suitable for neutral compounds. However, this does not satisfactorily clarify why negatively charged fragments appear to be often rejected or not detected. Therefore, the knowledge of preferred binding sites of MNV-1 P dimers is required.

3.5.3 Inhibitors of Murine Noroviruses

Despite the quality control of MNV-1 P dimers by DLS, the number of positive hits from STD NMR screening was still undesirably high (38%). Therefore, surface plasmon resonance spectroscopy was used to validate and to select promising fragments. In order to confirm the structural integrity of MNV-1 P dimers that were immobilized on the chip surface, a titration of 3'-sialyllactose was conducted, yielding two dissociation constants (K_{D1} ~ 3.5 mM and K_{D2} ~12 mM). Interestingly, a step-like binding curve has been also observed for L-fucose titrated onto immobilized Ast6139 VLPs using SPR (see REF[218]). The maximal RU response of 20 mM

3'-sialyllactose was about 4-fold smaller than the theoretical RU_{max}, which was calculated from the chip coverage (6500 RU MNV-1 P dimers), the molecular weight of 3'-sialyllactose (633.5 g/ml), and assuming a valency of two. This may be an indication that only one-quarter of immobilized P dimers had fully accessible glycan binding sites, while the remainder proteins suffered from steric hindrance with the dextran matrix due to random amine coupling. In a second step, an appropriate fragment concentration was determined to ensure RU values that were high enough to detect fragment binding and to concomitantly screen for relative high-affinity binders. Therefore, titrations of **CR273** and **CR432** were performed. These fragments had been chosen based on their high STD effects of 8.1% and 8.7%, respectively. Only two other fragments showed similar or slightly higher STD effects (**CR265** = 8.7% and **CR152** = 10.5%; the median of all hits was 2.3%). Provided that STD effects qualitatively reflect binding affinities for small molecules, it was assumed that these fragments exhibit the highest affinities for MVN-1 P dimers. Indeed, dissociation constants of about 40 µM were obtained, which is remarkable for such small fragments.

In a third step, the usage of the reporter ligand **CR432** confirmed the robustness of SPR screen, which resulted in 69 positive hits at a fragment concentration of 200 µM. It is intriguing that about 70% of SPR-positive fragments were positively charged. Since the carboxymethylated dextran matrix is negatively charged, even after deactivation with ethanolamine, unspecific binding could be assumed. However, adding 10 mM NaCl in running buffer and the usage of a reference cell should reduce the probability of detecting electrostatic-driven false positive hits. Moreover, the average magnitude of STD effects of positive hits from SPR screening was statistically higher ($p<0.05$) compared to fragments without SPR response, with 3.4%±2.1% and 2.4%±1.5%, respectively. Hence, SPR-derived hits seemed to be better binders in terms of affinity. In addition, 12 positively charged fragments were among the best 20 best fragments according to their STD effects. In contrast, eight fragments (**CR053**, **CR098**, **CR191**, **CR218**, **CR266**, **CR269**, **CR276**, and **CR295**) with STD effects of 5 to 7% did not give any SPR response, whereas **CR266** and **CR269** were poorly soluble in running buffer. For the other six fragments, hampered binding to immobilized MNV-1 P dimers may be caused by steric clashes with the dextran matrix.

In a fourth step, competition with 3'-sialyllactose was monitored by STD NMR. As expected for fragments with proposed relatively high affinities (~low µM range), the effect of reduction of STD signals was rather small (maximal 35% signal reduction). It is possible that tight ligands did not show any or only small changes in STD effects in the presence of 8 mM 3'sialyllactose (K_{D1} ~ 3.5 mM; i.e. 70% occupancy of the "first pocket"). To identify other competing or potential allosteric fragments, it is advisable to monitor STD effects of the remaining non-affected hits in the presence of, e.g. **CR120**, with a lower expected K_D (<200 µM).

Finally, a functional virus inhibition assay was used to identify potent fragments as a preliminary read-out. The assay design was chosen because it allows a relatively high throughput in an acceptable short time. Seven fragments caused a reduction of $TCID_{50}$ values more than five times (**CR054**, **CR120**, **CR437**, **CR196**, **CR372**, **TS039**, and **TS137**). Unfortunately, the inhibition mechanism cannot be derived from this kind of study. The inhibition of infectious viruses could have several reasons: aggregation of viral particles during treatment with fragments (as described for nanobodies[241]), partial inactivation through covalent modification (see discussion above[281]),

fragment-induced disassembly of virues[241], direct or indirect competition for glycan binding sites or receptor binding sites that are important for viral attachment[90], or inhibition of cellular processes that are critical for the virus life cycle[292]. STD NMR and SPR experiments, nevertheless, indicated that the interaction of fragments with the P domain of MNV-1 may be essential for virus inhibition. Since finding the ligand-binding site is an important first step in structure-guided drug discovery and crystallization or isotopic labeling of MNV-1 P dimers has failed so far, well-defined pockets that are potentially important for fragment binding were proposed by so-called "blind docking". Therefore, fragments from the CR/TS library were docked into all binding pockets that are present in MNV-1 P dimers in parallel. Three important binding regions could be identified. On the basis of the results from CSP and docking experiments of Saga P dimers, the *Emodel* may serve as an appropriate predictor when comparing docking sites of one ligand. A binding model was proposed, in which pocket 3 served as the main binding site for the MNV-1 inhibitors **CR054**, **CR120**, **CR196**, and **TS039**, whereby the interaction with Phe 401 and Asp 403 seems to be critical.

3.5.4 Hits of Saga P Dimers

During the fragment screening against Saga P dimers, the CR2006 Maybridge library had been extended by 227 water-soluble fragments purchased from Maybridge (ThermoFisher Scientific). In total, 127 hits were identified according to notable STD effects. The majority of hits were neutral or positively charged and exhibited two rings with two or less H-bond acceptors combined with only one H-bond donor. To identify fragments that bind to the canonical glycan binding sites, a competition assay was conducted, using methyl α-L-fucopyranoside as a competitor. Only 21 fragments showed a decreased STD effect (≤30%) in the presence of 8 mM of methyl α-L-fucopyranoside. The effects were rather low, but a complete competition could not be achieved even if 32 mM methyl α-L-fucopyranoside were applied. Studies of Mallagaray et al. suggested that glycan binding to noroviruses is a cooperative multi-step process (see REF[10]). Titration of methyl α-L-fucopyranoside using chemical shift perturbation experiments of Saga P dimers resulted in up to three dissociation constants, with K_{D1} ~ 2.5 mM, K_{D2} ~ 9.0 mM, and K_{D3} ~ 35.0 mM, reflecting four independent binding sites (K_{D4} was experimentally not accessible). According to the law of mass action, the occupancy at 32 mM methyl α-L-fucopyranoside should be ≥80% at least for the two canonical binding pockets of 5 µM Saga P dimers. A fragment at 200 µM that bind rather weakly to the fucose binding site should not give any or only very small STD signals (see also REF[218] on page 89 or REF[10], supplementary material) while competing with 32 mM methyl α-L-fucopyranoside. Otherwise, fragments showing only moderately reduced STD effects in the presence of 32 mM methyl α-L-fucopyranoside must bind sufficiently tight (≤100 µM) or interact additionally with secondary binding sites. None of these assumptions could be readily confirmed by chemical shift perturbation experiments. In the case of the fragments **CR215, CR440**, **TS012**, and **TS026** chemical shift mapping may indicate binding to the glycan binding sites, but the evidences are very weak according to the $^{1}H,^{15}N$-TROSY-HQSC spectra (cf. Appendix A7).
Unfortunately, 13 putative competitors were not soluble enough to accurately adjust the pH of 4 mM stock solutions and hence to ensure "clean" TROSY-HSQC spectra without pH artefacts.

For these fragments, the STD NMR-based competition assay should be repeated, employing HBGAs with higher affinities (e.g., blood group B trisaccharide) to confirm direct competition. Alternatively, CSP experiments could be performed at lower fragment concentrations. Interligand NOE experiments could also be conducted as described in REF[90] for NoV particles to detect fragments that bind in proximity to the fucose binding site.

Since **CR196**, **CR367**, and **TS153,** for example, apparently did not interact with the fucose binding sites but showed a reduction of STD signals in the presence of methyl α-L-fucopyranoside only two assumptions are left (see also Appendix A2, **Table A6**). Either the results from the STD NMR competition assays are artefacts due to experimental errors or the change in STD signals is caused by allosteric effects as outlined in the introductory section[241]. It is conceivable that e.g. pocket 2a is influenced by ligands that bind to pocket 1 and *vice versa*, as the P2 subdomain comprises six antiparallel β-sheets forming a barrel-like structure. The importance of β-sheets for the mechanism of allosteric propagation has been recently described for other proteins (see REF[293,294]). CSP data and the docking model of **CR065**, for instance, suggest that Arg 287 may be part of this allosteric network, albeit the increase of the STD effect for **CR065** was rather low (~11%) after addition of methyl α-L-fucopyranoside. Furthermore, it is not obvious why **CR066**, the analog of **CR065**, showed a reduction of saturation transfer and fragment **CR251** that seemingly also bind to pocket 2 experienced no alteration of its STD effect. An overall comparison between STD effects and CSP data is given Appendix A2, **Table A6**. At this stage of investigation, it is too early to draw any conclusion, even though subtle chemical shift changes of residues of the glycan binding site have been observed for some fragments that apparently interact with region 2 or 3 of the Saga P dimer (e.g., **CR106**, **CR125**, **CR136**, **CR215**, **TS162**, and **TS207**). Thus, further binding studies are required that include high-affinity HBGAs, such as blood group B trisaccharide, to confirm allosteric networks.

In conclusion, the fragments **CR001**, **CR065**, **CR066**, **CR120**, **CR125**, **CR196**, **CR251**, **CR367**, **CR396**, and **TS039** are the most suitable candidates for the investigation of allosterism of Saga P dimers upon HBGAs binding, because they are considerably soluble and, most importantly, robust CSP data indicated interactions with region 2a/b or 3.

3.5.5 Virtual Screening

In this study, molecular docking with Glide was beneficial to guide the interpretation of structure-activity relationships in the case of WNV protease NS2B-NS3^{pro}, to confirm structurally important binding pockets of Saga P dimers, and to propose binding pockets and pharmacophores that may be involved the inhibition of murine Noroviruses.

The usage of other docking algorithms such as AutoDock Vina may be helpful to cross-validate docking models. In contrast, it has been demonstrated that the docking performance of Vina coupled with neural-network scoring functions (NNS) is not statistically different than the performance of Glide. Moreover, the best scoring function is highly system dependent (see REF[295] and cited references).

The question arises whether it is advisable to dock small fragments at all. Docking studies of Kim et al. and Verdonk et al. support the idea of virtual fragment screening, as they demonstrated a strong correlation between the prediction of the binding free energy and experimental-determined binding affinity depending on the molecular weight of ligands[99]. Thereby, the key factor affecting docking performance is ligand efficiency[296] that is thought to be better for low molecular weight compounds[199]. On the contrary, virtual screening of all 659 water-soluble fragments from the CR/TS library targeting Saga P dimers using the extra precision Glide (XP Glide) method[103] yielded only 20% of STD NMR-positive hits (n = 43) when considering only 33% of all docked hits. Concerning MNV-1 P dimers, the coverage of positive hits confirmed by STD NMR was 48% (i.e., 70 of 146 docking hits from 432 CR fragments) and 30% (n = 45) were also found in the SPR screen. Interestingly, the fragments **CR054**, **CR120**, **CR196**, **CR372**, and **CR437,** for which inhibitory efficacy against MNV-1 was confirmed, as well as the tight binders **CR273** and **CR432** (that have not been tested in a virus inhibition assay so far) were among these hits derived from virtual screening. For both cases (Saga or MVN-1 P dimers), no linear correlation between STD effects and docking scores could be derived. Cross-docking enrichment (see above) may increase the rate of STD NMR-positive hits. Nevertheless, in the case of MNV-1 P dimers, it would have been beneficial to do first virtual screening to enrich promising fragments followed by experimental validation. Moreover, so-called "blind" docking has been proven to predict many CSP-derived binding pockets of Saga P dimers based on the *Emodel* pose-selection function, especially when CSP profiles were unambiguous. Therefore, the *Emodel* scoring function should be used to select between protein-ligand complexes of a given ligand and the *docking score*/*GlideScore* function to rank-order fragments to separate ligands that bind strongly from those that do not (see Schrödinger's webpage and REF[186,297]).

Despite the power of existing docking programs, there are still some drawbacks and uncertainties, for example, the identification and treatment (removal or retention) of potential critical structural water molecules, which are embedded in the binding pocket. Although XP Glide accounts for desolvation effects of both the ligand and the protein, even if all water molecules have been removed[103], water molecules that bridge the receptor and ligand by way of H-bonds cannot be readily simulated yet. An illustration of this problem is given in **Figure 3.38**. In the case of 3'-sialyllactose docked into different pockets of the MNV-1 P dimer, the dihedral angle of the β2→3 glycosidic bond that connects NeuNAc and Gal suggests that grid 1 may be the favored binding pocket according to the bioactive conformation as described in REF[9,10] (see also Appendix A9, **Table A9**). Whereas the scoring functions (*docking score* or *Emodel)* would predict grid 2 or grid 3 as a binding site for 3'-sialyllactose. Structural water that mediates hydrogen bonding is certainly critical for glycan binding but probably less important for small aromatic fragments. For the latter other forces seem to be at least equally essential for the process of binding. These forces are van der Waals interactions, π-π stacking, and electrostatic contacts. Regarding Saga P dimers, for example, the pattern of molecular descriptors (in particular the total charge) of fragments that were docked into the glycan binding region was the opposite, since negatively charged fragments were preferred when structural water was missing.

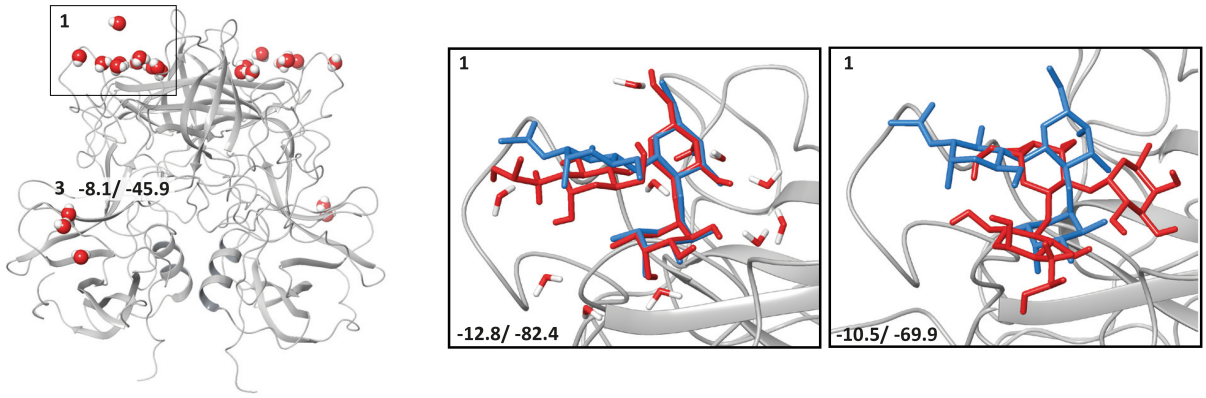


Figure 3.38 | Docking of H type 2 trisaccharide into the canonical glycan binding site of the Saga P dimer
A ribbon model of the crystal structure of Saga P dimer is shown in the presence of co-crystallized structural water based on the algorithm implemented in Maestro 11. H type 2 trisaccharide (red) was docked into all possible sites of the protein in the presence and absence of ordered water molecules. The best docking pose is depicted on the right side according to the *docking score* and *Emodel* (left bottom corner). In the presence of structured water, the pose prediction is highly accurate (the co-crystallized H type 2 trisaccharide is depicted in blue), whilst without water, only the binding pocket prediction is useful when considering the *docking score*/*Emodel* of H type 2 trisaccharide docked into grid 3 (ligand not shown).

This aspect would be in accordance with the fact that citrate, a compound with three negative charges at neutral pH, also binds to the fucose binding site[298]. Moreover, three negatively charged fragments, **CR055**, **CR070**, and **TS189**, from virtual screening (with structural water) that had been chosen because of their binding poses in proximity to the canonical fucose binding site as well as the negatively charged fragment **CR311** caused weak but detectable chemical shift changes of residues of the the glycan binding site. Therefore, it may be recommendable to analyze also fragments, for which docking to the fucose binding site without structural water resulted in appropriate docking scores (e.g., **CR085**, **CR403**, and **TS135**). Albeit these fragments did not show any STD signals during the STD NMR screen. Another aspect that is not yet fully fixed by the docking community is flexibility. Since docking algorithms focus on enthalpic contributions, especially the flexibility of the receptor is ignored in most docking algorithms. However, Schrödinger's Maestro provides an induced fit protocol to account for flexible binding sites. The application of this method probably will enhance the quality of binding pocket and binding pose prediction. In the end, molecular dynamic simulations will further underpin docking models.
Taken together, virtual docking is highly useful to guide experimental data and study design in a rational manner before, during, or after hit identification. Furthermore, it may accelerate the hit-to-lead optimization towards potent drug candidates. But finally, the pure experiment will give the answer whether or not the efforts of lead optimization will pay off.

3.6 Summary and Outlook

In the search of novel motifs with binding affinity for NoV P dimers from different strains, including MNV-1, Saga, and MI001, a screening of 432 fragments of the Maybridge library CR2006 was conducted by employing STD NMR as an initial read-out method. For P dimers of the predominant GII.4 Saga strain, a STD NMR-based screen of the extended 227-membered Maybridge library TS2017 was also performed.

In the case of anticipated MNV-1 dimers, an unusually high hit rate (~66%) led to the assumption of protein aggregation, which has been substantiated by dynamic light scattering (DLS) analysis. Furthermore, DLS experiments revealed that the absence of NaCl in the protein storage buffer prevents aggregation when stored over several weeks at 4 °C. In this regard, the flexible C-terminus of MNV-1 P dimers may be the major cause of aggregation, whether unordered (random coil-like) or structured (P particles) as it comprises the RMLK motif. A similar motif, a conserved arginine cluster (RRR), normally present in the C-terminus of the P domain of human NoV, has been identified to promote receptor binding and P particle formation[250]. This motif was absent in Saga or MI001 P dimers, which did not show aggregation. By controlling the integrity of MNV-1 P dimers, the number of positive hits could be reduced to ~39%, whereas the hit rate of Saga and MI001 P dimers was less than half. The application of SxG-matrices and principal component analysis of scalar molecular descriptors followed by descriptive statistics shed light on the underlying hallmarks of binders and non-binders. The findings revealed that neutral and positively charged compounds are by far more often identified as positive hits than negatively charged fragments, independently of the molecular scaffold.

In a second part, positive hits of MNV-1 P dimers derived from STD NMR experiments were further evaluated. A robust SPR screening setup that had been designed to exclude very weak binders yielded 69 positive hits, which were mostly positively charged, greater in size and exhibited less than three H-bond acceptors combined with only one H-bond donor. Subsequently, STD NMR competition experiments allowed for the identification of 19 to 25 putative competitors (depending on the error threshold specified) by the usage of 3'-sialyllactose as a competitive ligand in a 40-fold excess. Finally, selected putative competitive fragments and analogs were subjected to virus inhibition assays, yielding six potent MNV-1 inhibitors, with estimated IC_{50} values below 500 µM. Interestingly, three of them shared a common structural motif (**CR120**, **CR196**, and **TS039**; estimated $IC_{50} \leq 50$ µM) with fragment **CR160** that has been used earlier to design a multivalent entry inhibitor against the human NoV strain Ast6139/01/Sp[90].

As a highlight, virtual screening of the Maybridge libraries CR2006 and TS2017 directed to all possible binding pockets of MNV-1 P dimers has suggested three important "hot spots" for fragment binding. The pocket represented by Phe 401, Asp 403, and Asp 450 seems to be the most reasonable one that interacts with fragment **CR120** and analogs. This could also explain the structure-activity relationships derived from virus inhibition assays. Finding positive hits from the STD NMR and SPR screen that bind in the same fashion (e.g., **CR251**, **CR292**, **CR348**, or **CR432**) and confirming inhibitory efficacy against MNV-1 will help to underpin the proposed inhibition models. Moreover, it is certainly worth to screen also fragments that showed high STD effects but

no SPR signals (**CR053**, **CR098**, **CR191**, **CR266**, **CR269**, **CR276**, and **CR295**) in terms of inhibitory efficacy. Additionally, the fragments **CR152**, **CR273**, **CR265,** and **CR432** should be subjected to the virus inhibition assay because these fragments showed both high STD effects and high SPR responses. However, the design of inhibition studies should be modified to confirm direct or indirect competition with attachment factors[15,230] and/or the viral receptor[25]. For this purpose, murine virions would be pretreated with a given fragment for an adequate time (e.g., 30 min). Subsequently, susceptible cells would be incubated in the presence of viruses and a given fragment for another short period. All steps have to be conducted at 4 °C to prevent internalization of viruses. Finally, the number of infectious viruses that remain unattached in the supernatant would be quantified by endpoint titration. Additionally, another murine Norovirus strain, e.g. GV/CR3/2005/USA that shares 95% amino acid identity within the capsid region of MNV-1.CW3[230] should be also utilized to validate inhibitory effects.

In order to confirm proposed docking models of fragments bound to the dimer of MNV-1 P domains, mutagenesis can be conducted, thereby replacing Asp 403, for instance, by Asn or Ala. A double mutation, such as F401A and D403A, will further underpin the anticipated binding models. Finally, high-quality crystals of MNV-1 P dimers should be produced. Crystal soaking of potent fragments followed by the determination of crystal structures will provide realistic binding models, which includes the binding pocket and the orientation of the bound ligand. Alternatively, chemical shift perturbation experiments of ^{15}N-labeled MNV-1 P dimers may reveal binding pockets, provided that backbone assignment is available. Furthermore, waterLOGSY-based NMR experiments[211] can be employed to determine the solvent-exposed site of fragments bound to MNV-1 P dimers. The knowledge of the binding mode of single fragments can be utilized either to optimize hits towards valuable leads or to design multivalent entry inhibitors as inspired by studies of Rademacher et al.[90]. The existence of a robust cell culture system that allows the investigation of virus infectivity will be beneficial for this type of research. In the long term, highly potent (multivalent) MNV-1 entry inhibitors can be tested in mice as a role model for human Noroviruses.

In order to provide comprehensive knowledge of compounds that may also inhibit human Noroviruses, the binding mechanisms of positive hits from STD NMR targeting Saga P dimers have been investigated. The combination of STD NMR-based competition assays with methyl α-L-fucopyranoside as a competitor, ^{1}H,^{15}N-TROSY-HSQC-based chemical shift perturbation (CSP) experiments of uniformly [^{2}H,^{15}N]-labeled Saga P dimers, and virtual docking delivered valuable insights of fragment binding. Noteworthy, molecular docking served as an important tool to guide the interpretation of experimental data. The binding site prediction based on the *Emodel* function as implemented in Glide was, in many cases, in good agreement with CSP data. Surprisingly, for none of the putative competitors, which were soluble enough to be analyzed and had experienced a reduction of STD signals upon binding of methyl α-L-fucopyranoside, an interaction with the glycan binding site could be undoubtedly confirmed by chemical shift mapping. In fact, another binding region of the P1 subdomain represented by Leu 278 has been discovered as “hot spot” for fragment binding according to CSP experiments and virtual screening. This cavity, which is 30 to 35 Å apart from the canonical glycan binding site, predominantly comprises hydrophobic and negatively charged residues, leading to the attraction of mainly positively charged compounds.

Fragments that unambiguously interact with this binding pocket (e.g., **CR001**, **CR065**, **CR066**, **CR251**, and **CR367**) can be utilized to sense allosteric networks that may extend from the glycan binding sites to the N-terminus, which is normally linked to the shell domain in Norovirus particles. Other fragments, such as **CR136** or **TS039**, that bind in proximity to the N-terminus according to CSP data may also be used as molecular probes. Many of these aforementioned fragments showed a change in the averaged STD effects when mixed with Saga P dimers upon binding of methyl α-L-fucopyranoside.

In future studies, high-affinity HBGAs, such as blood group B trisaccharide, should be employed to increase the dynamic range of the experimental read-out of STD NMR-based competition assays, and hence to better judge allosteric effects. Furthermore, U-[^{2}H,^{15}N]-labeled Saga P dimers should be subjected to an unfolding/refolding protocol (Mallagaray, 2017) to promote deuterium/proton exchange, in particular for nitrogen atoms in the interface of P dimers. That will increase the number of backbone NH signals and therefore raise the information content of chemical shift perturbation experiments. Titrations of promising hits based on protein-observed NMR methods will further aid the selection of a suitable lead structure. Furthermore, this procedure may elucidate the characteristics of binding modes of potential allosteric compounds.

Finally, virus inhibition assays would be beneficial to guide the selection of potential allosteric inhibitors similar to the murine counterpart. Recently, it has been demonstrated that human Noroviruses can be cultured in B cells[299], but reproducibility seems to be an issue. Alternatively, hemagglutination assays or saliva-based assays can be conducted by using virus-like particles to estimate the inhibitor efficiencies, albeit the accuracy of this assay format is considered to be rather low[218]. In conclusion, NMR experiments delivered fragments that represent valuable starting points for the development of novel Norovirus entry inhibitors. This process may be guided by *in silico* methods to accelerate the process of rational fragment-based drug design.

Appendix

A1 Sequences and Vector Maps

Amino Acid Sequence of Bovine eE_2 + Protein A tag

MDSKGSSQKGSRLLLLLVVSNLLLYQGVVSYPDCKPGFSYAIAKNDEIGPLGATGLTTQWYEYSDGMRLQDSVVEVWCKNGEIKYLIRCGREAR
YLAVLHTRALPTSVVFEKIFDGKEQEDIVEMDDNFEFGLCPCDARPLIRGKFNTTLLNGPAFQMVCPIGWTGTVSCTLANKDTLATIVVRTYKRV
RPFPYRQDCVTQKTIGEDLYDCALGGNWTCVPGDALRYVAGPVESCEWCGYKFLKSEGLPHFPIGKCRLKNESGYRQVDETSCNRNGVAIVPS
GTVKCKIGDTVVQVIAMDDKLGPMPCKPHEIISSEGPVEKTACTFNYTRTLKNKYFEPRDNYFQQYMLKGEYQYWFDLEITDHHRDYASGLEV
LFQ/GPGSGGEAQKLNDSQAPKADAQQNNFNKDQQSAFYEILNMPNLNEAQRNGFIQSLKDDPSQSTNVLGEAKKLNESQAPKADNNFNK
EQQNAFYEILNMPNLNEEQRNGFIQSLKDDPSQSANLLSEAKKLNESQAPKADNKFNKEQQNAFYEILHLPNLNEEQRNGFIQSLKDDPSQSA
NLLAEAKKLNDAQAPKADNKFNKEQQNAFYEILHLPNLTEEQRNGFIQSLKDDAGDYKDHDGDYKDHDIDYKDDDDK

Prolactin signal peptide for secretion

eE_2 protein N-Glycans[48]

PPase cleavage site

Protein A putative N-Glycans

Flag Tag

Ext. coefficient: 74230 M^{-1} cm^{-1}

Molecular weight: 71569.07 Da

Apparent MW: 75 kDa (as monomer)

Sequence of Bovine CD46 (Uniprot Q6VE48)[21,44]

MRASCTPLKAPLRRPERLASSGRFAWVLLLAPLLLLPTSSDACDDPPRFVSMKPQGTLKP
SYSPGEQIVYECHLGFQPVTPGQVLALVCQDNNTWSSLQEGCKKRRCPTLADPTNGQVIL
VNGSTEFGSEVHYVCNNGYYLLGTNISYCEVSSGTGVNWSDNPPTCEKILCQPPPEIQNG
KYTNSHKDVFEYNEVVTYSCDPSNGPDEYSLVGESKLTCIGNGEWSSQPPQCKVVKCVYP
AIEHGTIVSGFGPKYYYKATVVLKCNEGFNLYGNSVVVCGENSTWEPELPKCIK
<VSIPPSTQSPIPSTQPPVPSVSVSTRSTQHPVPSVS > GHPPRP
TDASPPNGAEGLGAGYIVLVIVAVLIGVGLLLCLYCCFCRQRKKGIYVTGESHRQDILFSL

CCP1: CDDPPRFVSMKPQGTLKPSYSPGEQIVYECHLGFQPVTPGQVLALVCQDNNTWSSLQEGCKK
CCP2: RRCPTLADPTNGQVILVNGSTEFGSEVHYVCNNGYYLLGTNISYCEVSSGTGVNWSDNPPTCE
CCP3: KILCQPPPEIQNGKYTNSHKDVFEYNEVVTYSCDPSNGPDEYSLVGESKLTCIGNGEWSSQPPQCK
CCP4: VVKCVYPAIEHGTIVSGFGPKYYYKATVVLKCNEGFNLYGNSVVVCGENSTWEPELPKC
STP: IK <VSIPPSTQSPIPSTQPPVPSVSVSTRSTQHPVPSVS > GHPPRPTDASPPN
TM: GLGAGYIVLVIVAVLIGVGLL
CT: CCFCRQRKKGIYVTGESHRQDILFSL

Signal peptide: MRASCTPLKAPLRRPERLASSGRFAWVLLLAPLLLLPTSSDA (for Glycosylation)
Interaction with BVDV E2
TM domain
Cytoplasmic
N-Glycans (according to UniProt Q6VE48)/ **Putative *N*-Glycans / Putative *O*-Glycans**

Apparent MW: ~50-60 kDa and 90 kDa (splice variants)

WNV Protease

Amino acid sequence of WNV protease NS2B-NS3^{pro}

M W I E R T A D I T W E S D A E I T G S S E R V D V R L D D D G N F Q L M N D P G A P W A G G
G G S G G G G G G V L W D T P S P K E Y K K G D T T T G V Y R I M T R G L L G S Y Q A G A G V
M V E G V F H T L W H T T K G A A L M S G E G R L D P Y W G S V K E D R L C Y G G P W K L Q
H K W N G H D E V Q M I V V E P G K N V K N V Q T K P G V F K T P E G E I G A V T L D Y P T G T
S G S P I V D K N G D V I G L Y G N G V I M P N G S Y I S A I V Q G E R M E E P A P A G F E P E
M L L E H H H H H H

Number of amino acids: 246

Molecular weight: 26543.63 Da

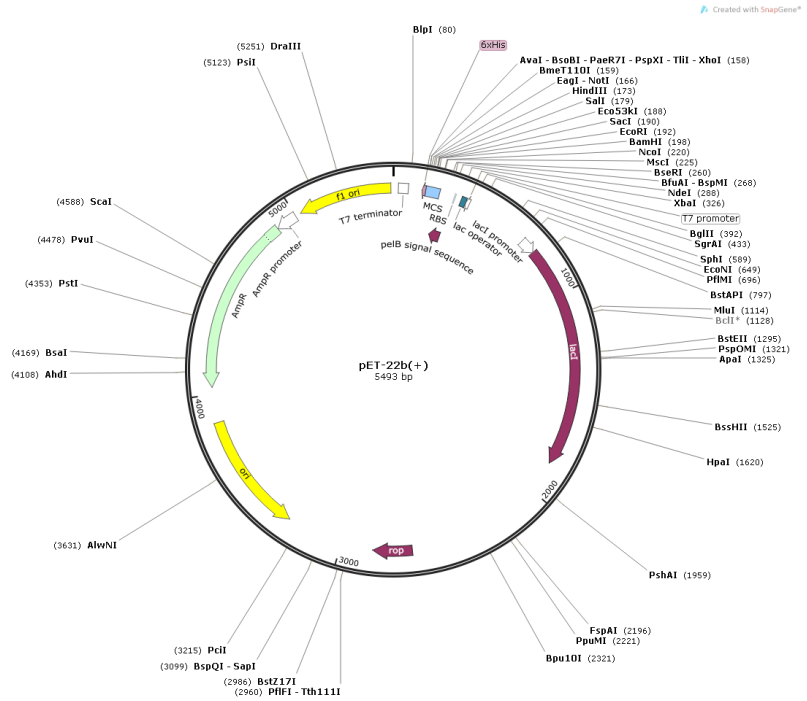
Ext. coefficient: 55920 M^{-1} cm^{-1}

Theoretical pI: 5.13

T D M W I E R T A D I T W E S D A E I T G S S E R V D V R L D D D G N F Q L M N D P G A P W A
G G G G S G G G G G G V L W D T P S P K E Y K K G D T T T G V Y R I M T R G L L G S Y Q A G A
G V M V E G V F H T L W H T T K G A A L M S G E G R L D P Y W G S V K E D R L C Y G G P W K
L Q H K W N G H D E V Q M I V V E P G K N V K N V Q T K P G V F K T P E G E I G A V T L D Y P T
G T S G S P I V D K N G D V I G L Y G N G V I M P N G S Y I S A I V Q G E R M E E P A P A G F E P
E M L R K K

Amino acid sequence of the WNV protease NS2B-NS3^{pro} used in this study. The first 238 residues are identical to the sequence of Su et al. (BMRB 11053)[132]. However, the first two amino acids (Thr and Asp) are missing. The last three amino acid residues Arg, Lys, and Lys (red) are replaced by Leu and Gln. In addition, the histidine-tag (grey) is missing in the latter construct. Backbone assignment is available for underlined amino acids.

Vector Map



MNV-1 P Dimer

Amino acid sequence of HIS-MOCR-TEV-MNV-1 domain

MHHHHHHSSGVDLAMSNMTYNNVFDHAYEMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIRSVMTSE
GIDLEFEDSGLMPDTKDDIRILQARIYEQLTIDLWEDAEDLLNEYLEEVEEYEEDEEGTENLYFQ///SNATIYRMVDLP
VIQPRLCTHARWPAPVYGLLVDPSLPSNPQWQNGRVHVDGTLLGTTPISGSWVSCFAAEAAYEFQSGTGEVATFT
LIEQDGSAYVPGDRAAPLGYPDFSGQLEIEVQTETTKTGDKLKVTTFEMILGPTTNADQAPYQGRVFASVTAAASLD
LVDGRVRAVPRSIYGFQDTIPEYNDGLLVPLAPPIGPFLPGEVLLRFRTYMRQIDTADAAAEAIDCALPQEFVSWFAS
NAFTVQSEALLLRYRNTLTGQLLFECKLYNEGYIALSYSGSGPLTFPTDGIFEVVSWVPRLYQLASVGSLATGRMLKQ

His-tag

MOCR

TEV cleavage site

MNV-1 P domain

HIS-MOCR-TEV-MNV-1 domain

Number of amino acids: 451

Molecular weight: 50302.12 Da

Theoretical pI: 4.34

MNV-1 domain (as monomer)

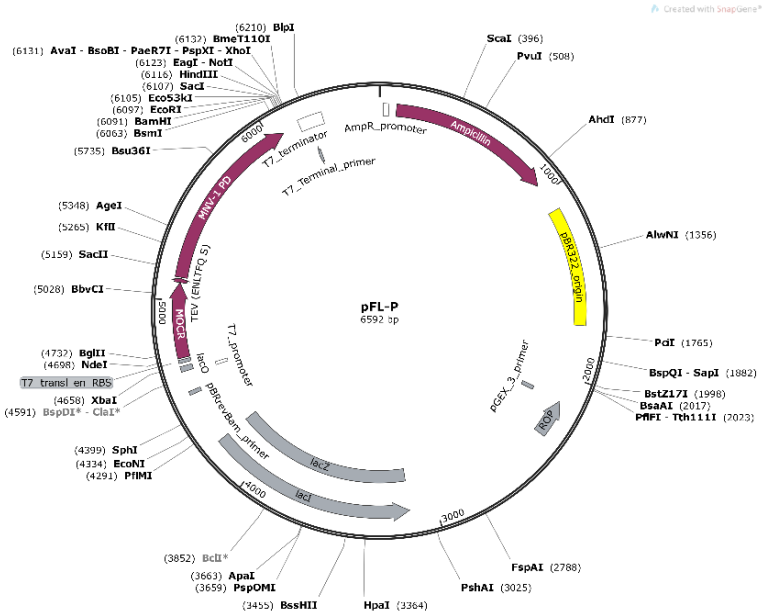
Number of amino acids: 320

Molecular weight: 34835.5 Da

Ext. coefficient: 48610 M^{-1} cm^{-1}

Theoretical pI: 4.68

Vector Map



Saga P Dimer

Amino acid sequence after cleavage with HRV3C protease

GPGSKPFTVPILTVEEMTNSRFPIPLEKLFTGPSGAFVVQPQNGRCTTDGVLLGTTQLSPVNICTFRGDVTHIAGSRN
YTMNLASLNWNNYDPTEEIPAPLGTPDFVGKIQGLLTQTTKGDGSTRGHKATVYTGSAPFTPKLGSVQFSTDTEND
FETHQNTKFTPVGVIQDGSTTHRNEPQQWVLPSYSGRNVHNVHLAPAVAPTFPGEQLLFFRSTMPGCSGYPNMD
LDCLLPQEWVQHFYQEAAPAQSDVALLRFVNPDTGRVLFECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSWVN
QFYTLAPM

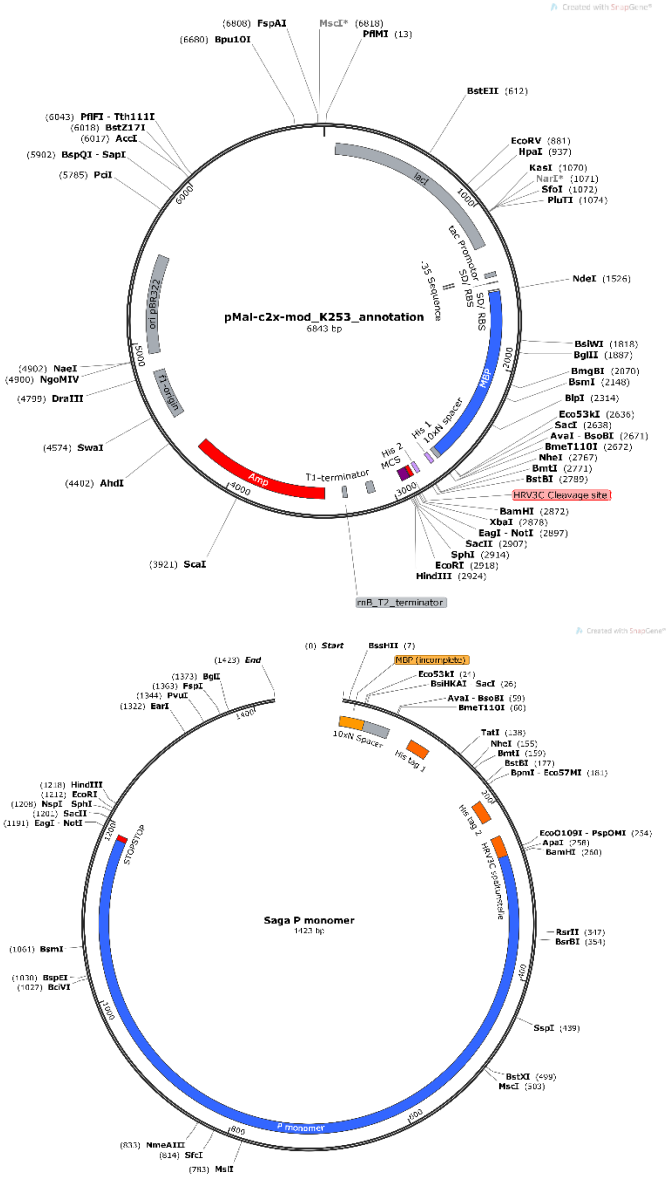
Number of amino acids: 310

Molecular weight: 34064.2 (as monomer)

Ext. coefficient: 70820 M^{-1} cm^{-1} (for P dimer)

Theoretical pI: 5.91

Vector Map



MI001 P Dimer

Amino acid sequence after cleavage with HRV3C protease

GPGSKPFSVPILTVEEMTNSRFPIPLEKLFTGPSSAFVVQPQNGRCTTDGVLLGTTQLSPVNICTFRGDVTHIAGTQE
YTMNLASQNWNNYDPTEEIPAPLGTPDFVGKIQGVLTQTTRRDGSTRGHKATVSTGSVHFTPKLGRIQFSTDTSND
FETGQNTRFTPVGVVQDGSTTHQNEPQQWVLPNYSGRDSHNVHLAPAVAPSFPGEQLLFFRSTMPGCSGYPNM
NLDCLLPQEWVQHFYQEAAPAQSDVALLRFVNPDTGRVLFECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSWV
NQFYTLAPM

Number of amino acids: 310

Molecular weight: 34149.2

Ext. coefficient: 34170 M^{-1} cm^{-1}

Theoretical pI: 5.91

Vector Map

Not available

Protein Sequence Alignment using the ExPAsy Web Tool

MNV-1 and Saga:

32.0% identity in 306 residues overlap; Score: 249.0; Gap frequency: 4.9%

```
MNV-1          11 LPVIQPRLCTHARWPAPVYGLLVDPSLPSNPQWQNGRVHVDGTLLGTTPISGSWVSCFAA
Saga            9 VPILTVEEMTNSRFPIPLEKLFTGPSGAFVVQPQNGRCTTDGVLLGTTQLSPVNICTFRG
                   *       *  * * *   *   **     * ****   ** *****  *      *

MNV-1          71 EAAYEFQSGTGEVATFTLIEQDGSAYVPGDRA-APLGYPDFSGQLEIEVQTETTK-----
Saga           69 DVTHIAGSRN---YTMNLASLNWNNYDPTEEIPAPLGTPDFVGKIQ-GLLTQTTKGDGST
                         *      *  *       * *     **** *** *       * ***

MNV-1         125 TGDKLKVTTFEMILGPTTNADQ-APYQGRVFASVTAAASLDLVDGRVRAVPRSIYGFQDT
Saga          125 RGHKATVYTGSAPFTPKLGSVQFSTDTENDFETHQNTKFTPVGVIQDGSTTHRNEPQQWV
                   * *  * *      *     *        *                          *

MNV-1         184 IPEYNDGLL-VPLAPPIGPFLPGEVLLRFRTYMRQIDTADAAAEAIDCALPQEFVSWFA
Saga          185 LPSYSGRNVHNVHLAPAVAPTFPGEQLLFFRSTMP-GCSGYPNMDLDCLLPQEWVQHFY
                   * *       * ***   *  *** ** **  *             ** **** *  *

MNV-1         242 SNAFTVQSEALLLRYRNTLTGQLLFECKLYNEGYIALSYSGSGPLTFPTDGIFEVVSWVP
Saga          243 QEAAPAQSDVALLRFVNPDTGRVLFECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSWVN
                    *   **   ***  *  **  ******   **      *   *  *  * *   ***

MNV-1         302 RLYQLA
Saga          303 QFYTLA
```

MI001 and Saga:

92.3% identity in 310 residues overlap; Score: 1568.0; Gap frequency: 0.0%

```
MI001           1 GPGSKPFSVPILTVEEMTNSRFPIPLEKLFTGPSSAFVVQPQNGRCTTDGVLLGTTQLSP
Saga            1 GPGSKPFTVPILTVEEMTNSRFPIPLEKLFTGPSGAFVVQPQNGRCTTDGVLLGTTQLSP
                  ******* ************************* *************************

MI001          61 VNICTFRGDVTHIAGTQEYTMNLASQNWNNYDPTEEIPAPLGTPDFVGKIQGVLTQTTRR
Saga           61 VNICTFRGDVTHIAGSRNYTMNLASLNWNNYDPTEEIPAPLGTPDFVGKIQGLLTQTTKG
                  ***************   ******* ************************** *****

MI001         121 DGSTRGHKATVSTGSVHFTPKLGRIQFSTDTSNDFETGQNTRFTPVGVVQDGSTTHQNEP
Saga          121 DGSTRGHKATVYTGSAPFTPKLGSVQFSTDTENDFETHQNTKFTPVGVIQDGSTTHRNEP
                  *********** ***  ******  ****** ***** *** ****** ******* ***

MI001         181 QQWVLPNYSGRDSHNVHLAPAVAPSFPGEQLLFFRSTMPGCSGYPNMNLDCLLPQEWVQH
Saga          181 QQWVLPSYSGRNVHNVHLAPAVAPTFPGEQLLFFRSTMPGCSGYPNMDLDCLLPQEWVQH
                  ****** ****  ********** ********************** ***********

MI001         241 FYQEAAPAQSDVALLRFVNPDTGRVLFECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSW
Saga          241 FYQEAAPAQSDVALLRFVNPDTGRVLFECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSW
                  ************************************************************

MI001         301 VNQFYTLAPM
Saga          301 VNQFYTLAPM
                  *********
```

MNV-1 and MI001:

33.3% identity in 306 residues overlap; Score: 276.0; Gap frequency: 4.9%

```
MNV-1          11 LPVIQPRLCTHARWPAPVYGLLVDPSLPSNPQWQNGRVHVDGTLLGTTPISGSWVSCFAA
MI001           9 VPILTVEEMTNSRFPIPLEKLFTGPSSAFVVQPQNGRCTTDGVLLGTTQLSPVNICTFRG
                   *       *  * * *   *   **     * ****   ** *****  *      *

MNV-1          71 EAAYEFQSGTGEVATFTLIEQDGSAYVPGDRA-APLGYPDFSGQLEIEVQTETTKT----
MI001          69 DVTH-IAGTQEY-TMNLASQNWNNYDPTEEIPAPLGTPDFVGKIQ-GVLTQTTRRDGST
                         ** *  *  *  *    * *     **** *** *     * * **

MNV-1         126 -GDKLKVTTFEMILGPTTNADQ-APYQGRVFASVTAAASLDLVDGRVRAVPRSIYGFQDT
MI001         125 RGHKATVSTGSVHFTPKLGRIQFSTDTSNDFETGQNTRFTPVGVVQDGSTTHQNEPQQWV
                   * *  * *      *     *        *                          *

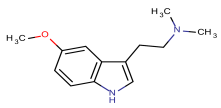
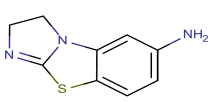
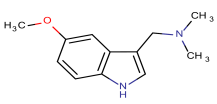
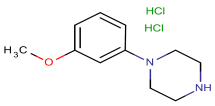
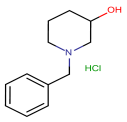
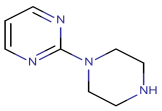
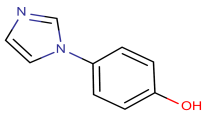
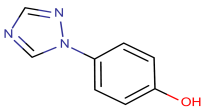
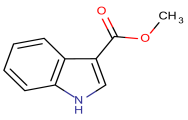
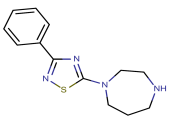
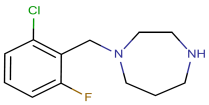
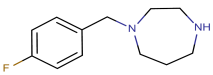
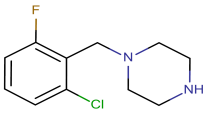
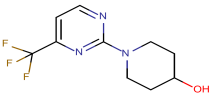
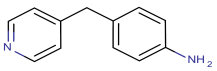
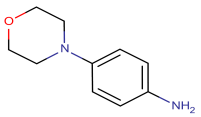
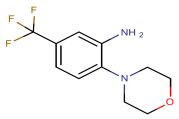
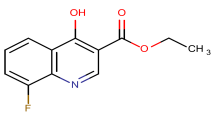
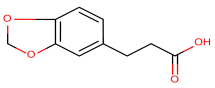
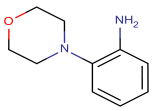
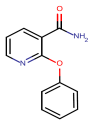
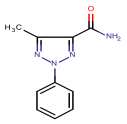
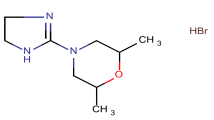
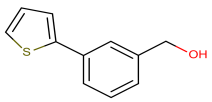
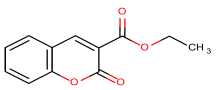
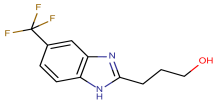
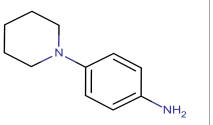
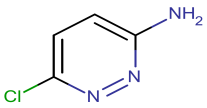
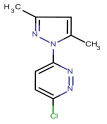
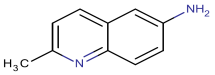
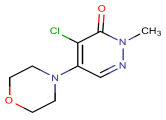
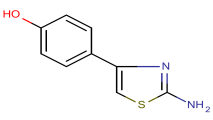
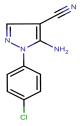
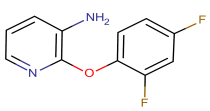
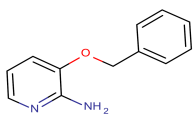
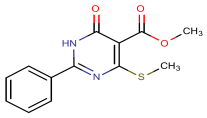
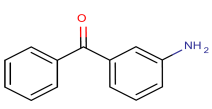
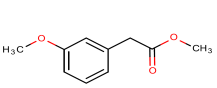
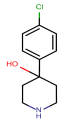
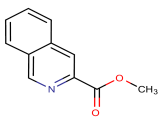
MNV-1         184 IPEYN-DGLLVPLAPPIGPFLPGEVLLRFRTYMRQIDTADAAAEAIDCALPQEFVSWFA
MI001         185 LPNYSGRDSHNVHLAPAVAPSFPGEQLLFFRSTMP-GCSGYPNMNLDCLLPQEWVQHFY
                   * *   *   * ***   *  *** ** **  *            ** **** *  *

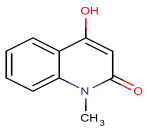
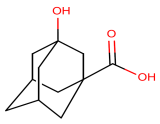
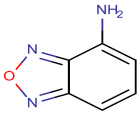
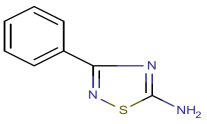
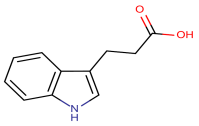
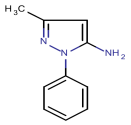
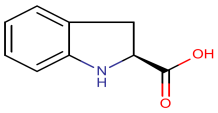
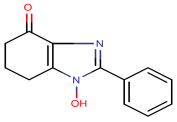
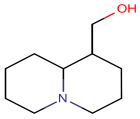
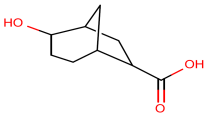
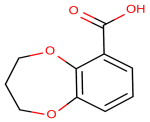
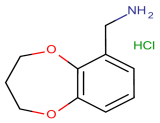
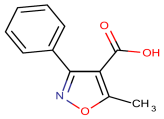
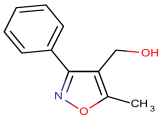
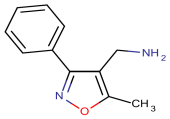
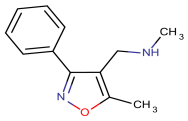
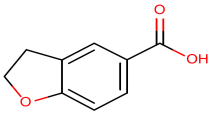
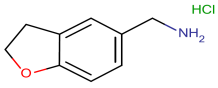
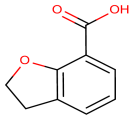
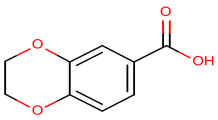
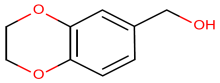
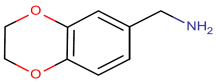
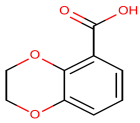
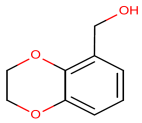
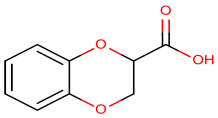
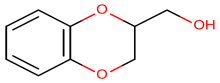
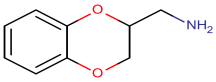
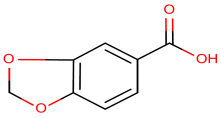
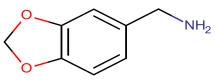
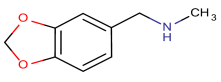
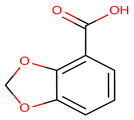
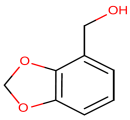
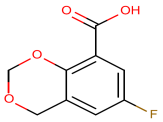
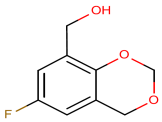
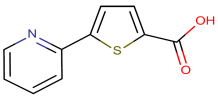
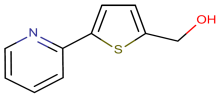
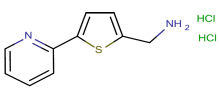
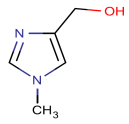
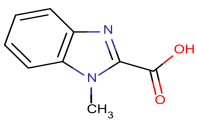
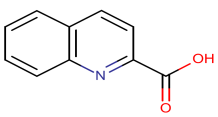
MNV-1         242 SNAFTVQSEALLLRYRNTLTGQLLFECKLYNEGYIALSYSGSGPLTFPTDGIFEVVSWVP
MI001         243 QEAAPAQSDVALLRFVNPDTGRVLFECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSWVN
                    *   **   ***  *  **  ******   **      *   *  *  * *   ***

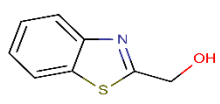
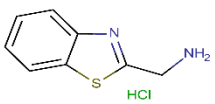
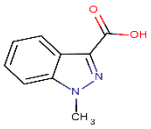
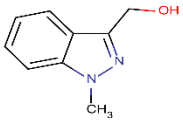
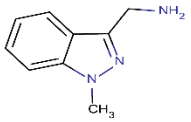
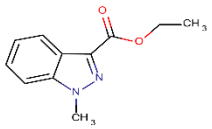
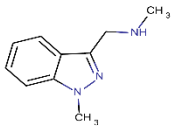
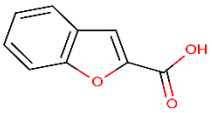
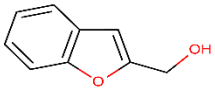
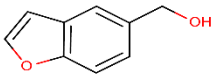
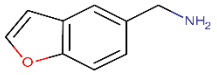
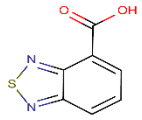
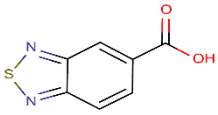
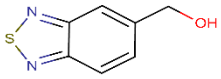
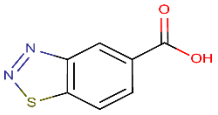
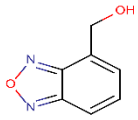
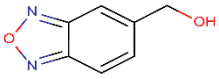
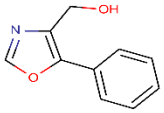
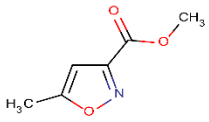
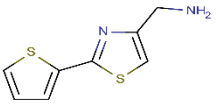
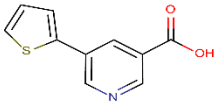
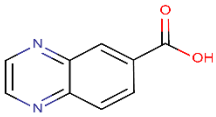
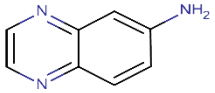
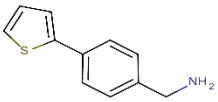
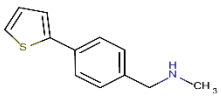
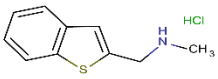
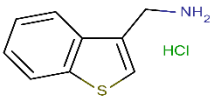
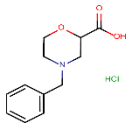
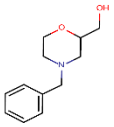
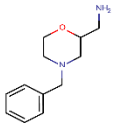
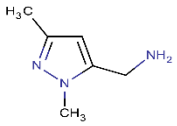
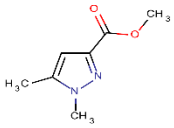
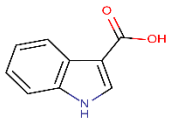
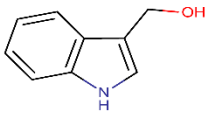
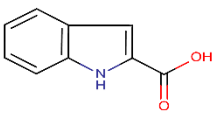
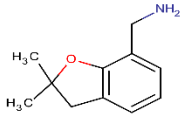
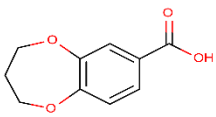
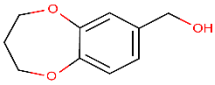
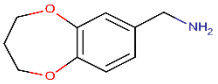
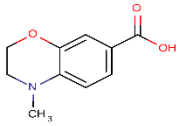
MNV-1         302 RLYQLA
MI001         303 QFYTLA
                    * **
```

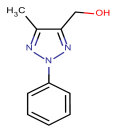
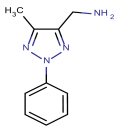
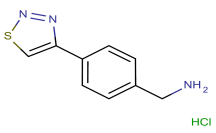
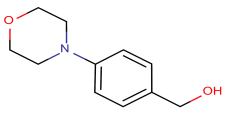
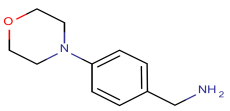
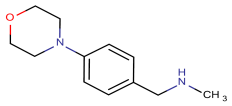
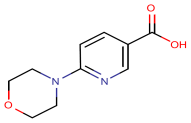
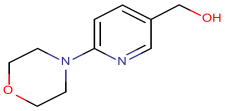
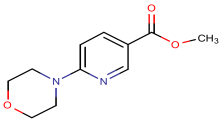
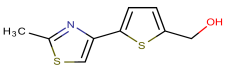
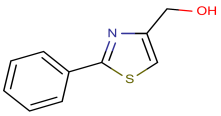
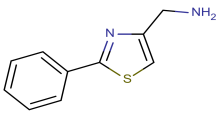
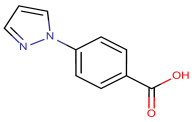
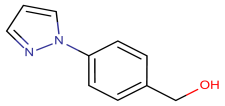
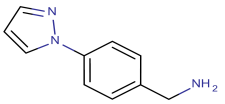
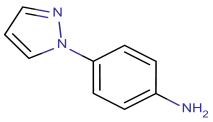
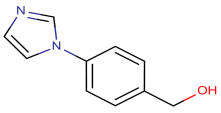
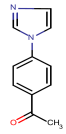
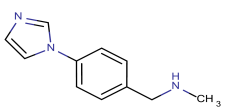
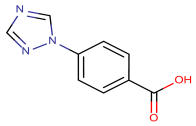
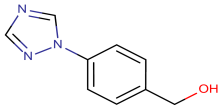
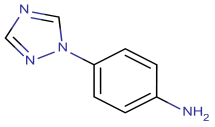
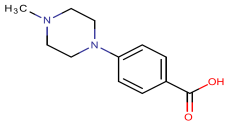
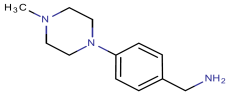
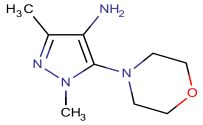
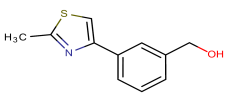
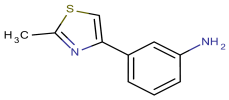
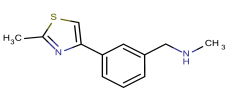
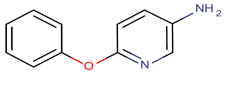
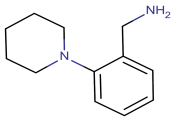
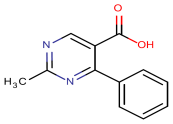
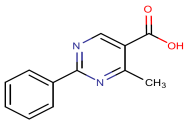
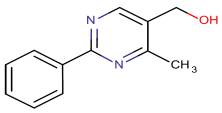
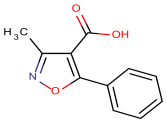
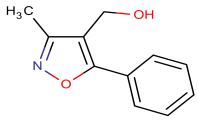
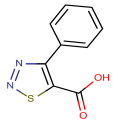
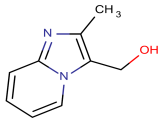
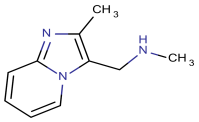
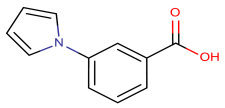
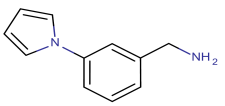
A2 Maybridge Library, Fragment Cluster and STD NMR/CSP hits

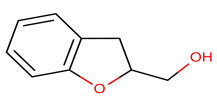
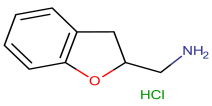
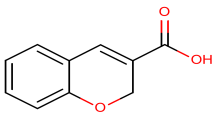
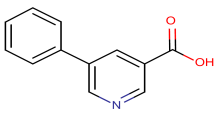
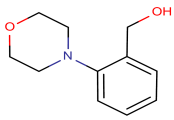
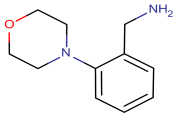
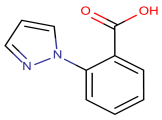
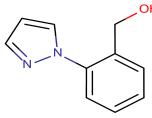
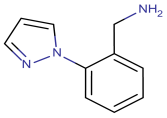
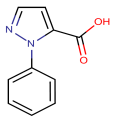
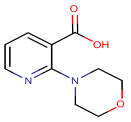
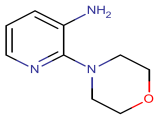
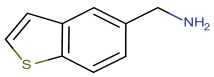
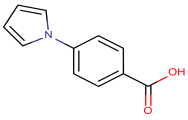
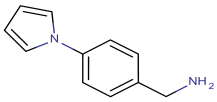
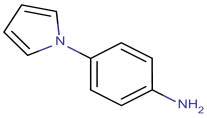
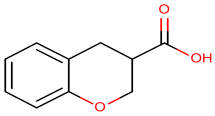
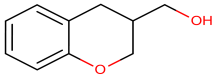
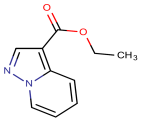
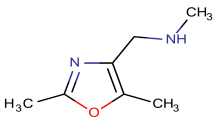
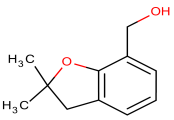
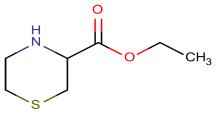
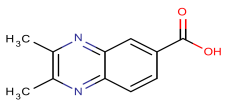
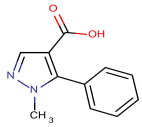
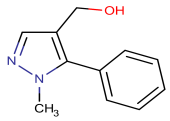
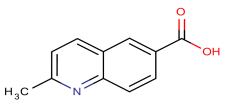
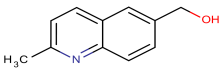
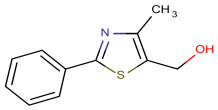
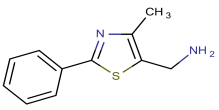
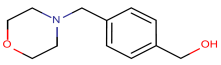
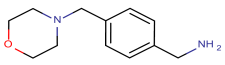
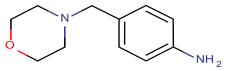
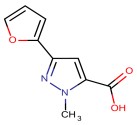
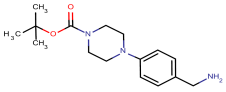
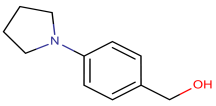
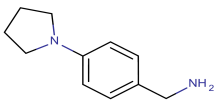
Maybridge CR2006 Library

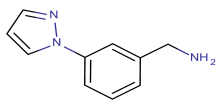
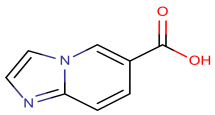
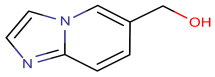
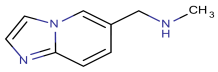
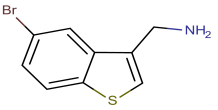
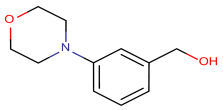
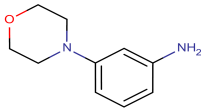
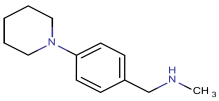
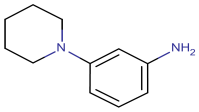
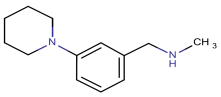
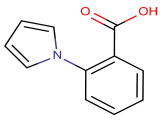
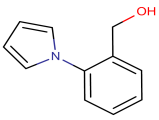
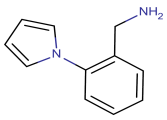
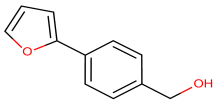
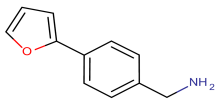
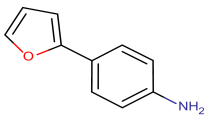
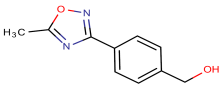
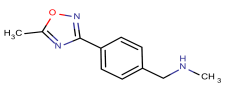
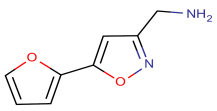
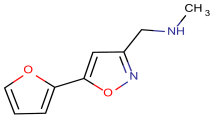
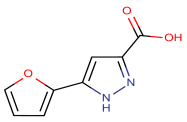
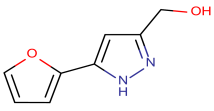
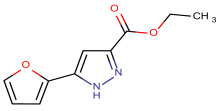
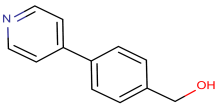
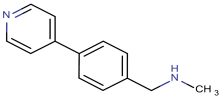
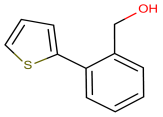
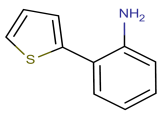
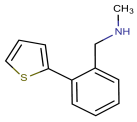
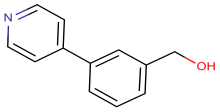
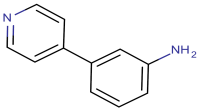
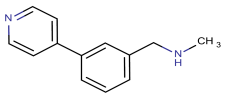
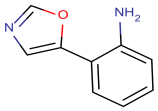
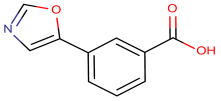
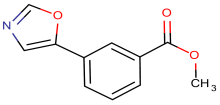
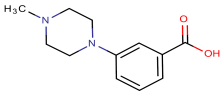
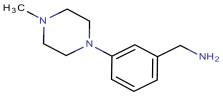
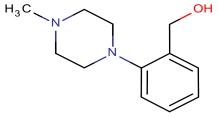
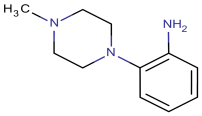
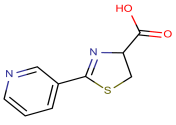
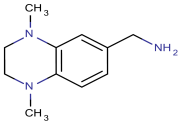
1 	2 	3 	4 	5 
6 	7 	8 	9 	10 
11 	12 	13 	14 	15 
16 	17 	18 	19 	20 
21 	22 	23 	24 	25 
26 	27 	28 	29 	30 
31 	32 	33 	34 	35 
36 	37 	38 	39 	40 
41	42	43	44	45
46	47	48	49	50

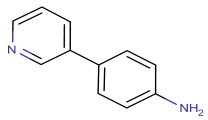
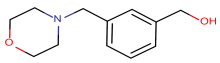
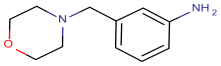
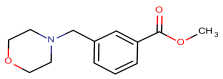
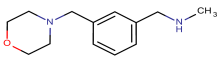
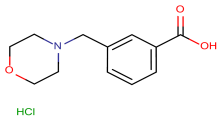
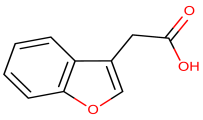
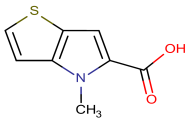
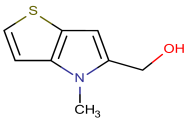
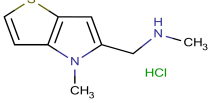
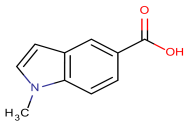
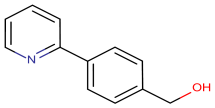
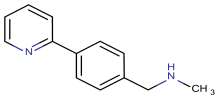
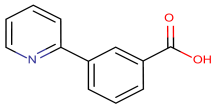
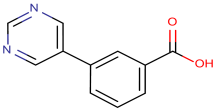
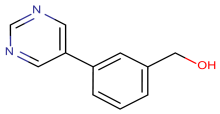
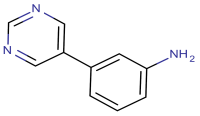
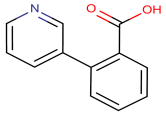
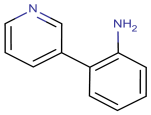
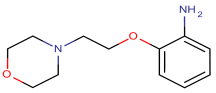
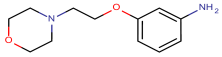
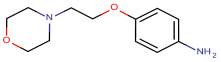
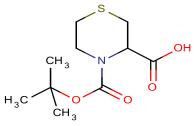
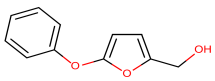
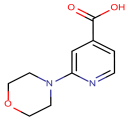
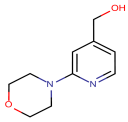
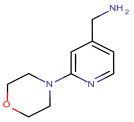
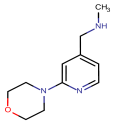
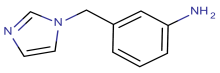
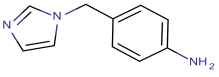
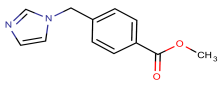
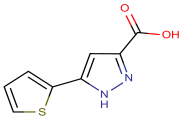
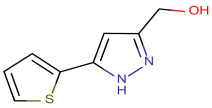
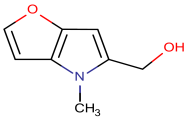
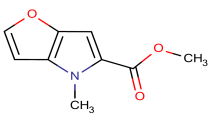
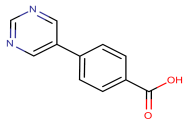
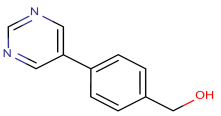
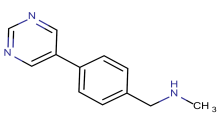
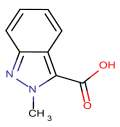
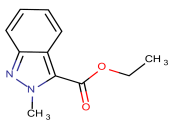
51 	52 	53 	54 	55 
56 	57 	58 	59 	60 
61 	62 	63 	64 	65 
66 	67 	68 	69 	70 
71 	72 	73 	74 	75 
76 	77 	78 	79 	80 
81 	82 	83 	84 	85 
86 	87 	88 	89 	90 
91	92	93	94	95
96	97	98	99	100

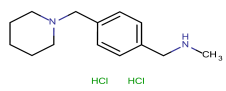
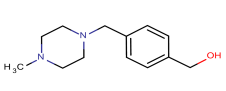
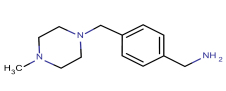
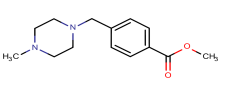
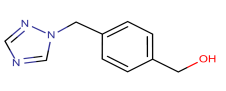
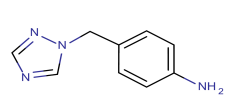
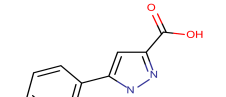
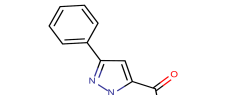
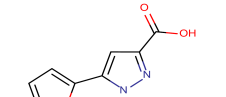
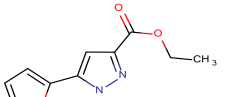
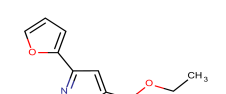
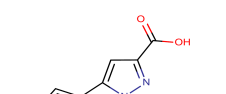
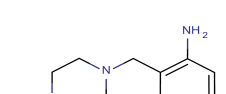
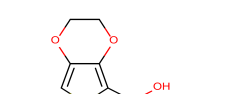
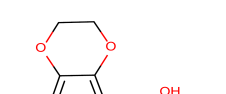
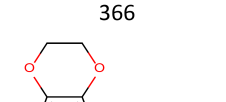
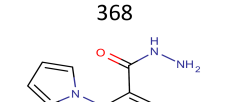
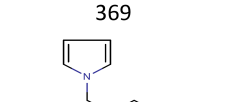
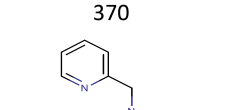
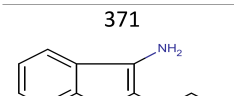
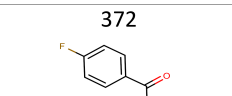
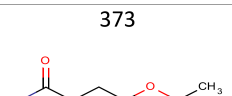
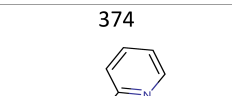
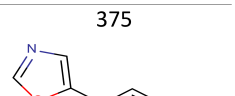
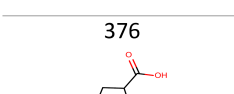
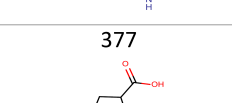
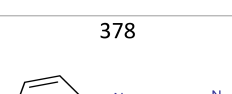
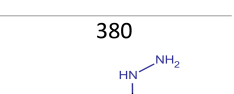
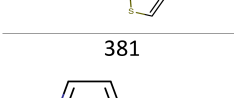
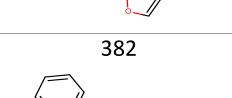
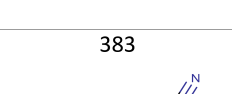
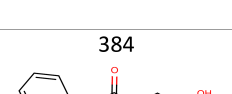
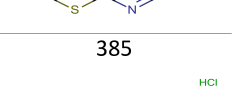
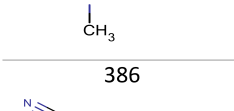
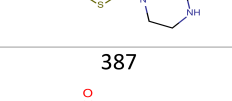
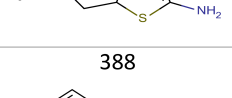
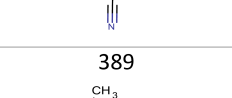
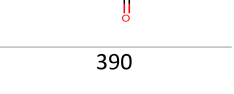
101 	102 	103 	104 	105 
106 	107 	108 	109 	110 
111 	112 	113 	114 	115 
116 	117 	118 	119 	120 
121 	122 	123 	124 	125 
126 	127 	128 	129 	130 
131 	132 	133 	134 	135 
136 	137 	138 	139 	140 
141	142	143	144	145
146	147	148	149	150

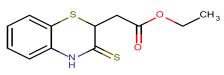
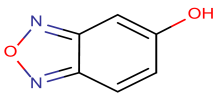
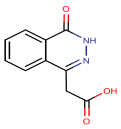
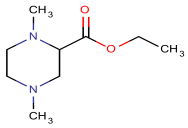
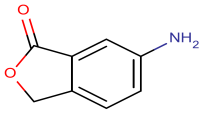
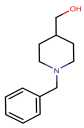
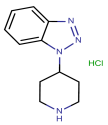
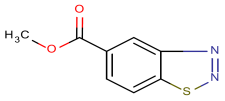
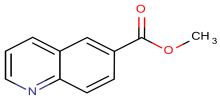
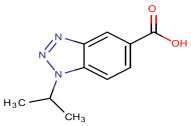
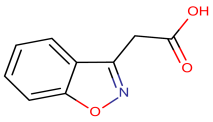
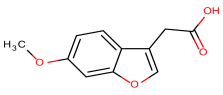
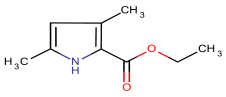
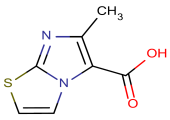
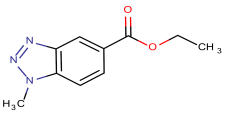
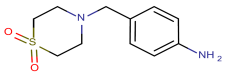
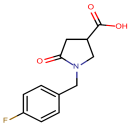
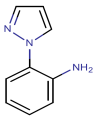
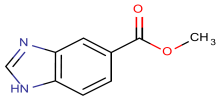
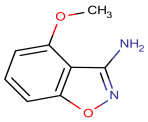
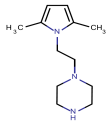
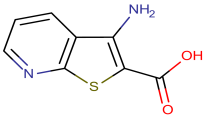
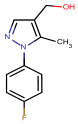
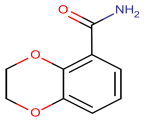
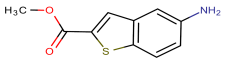
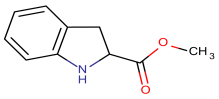
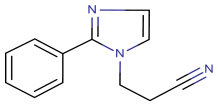
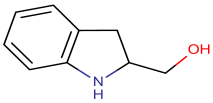
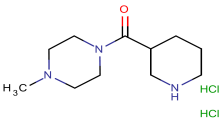
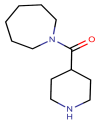
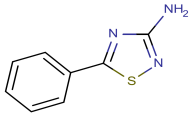
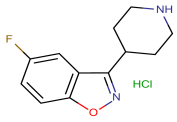
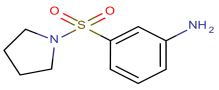
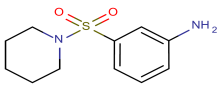
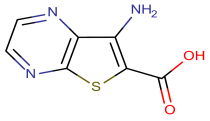
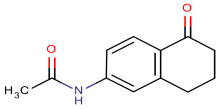
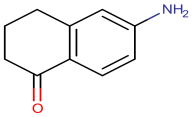
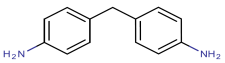
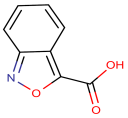
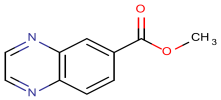
151 	152 	153 	154 	155 
156 	157 	158 	159 	160 
161 	162 	163 	164 	165 
166 	167 	168 	169 	170 
171 	172 	173 	174 	175 
176 	177 	178 	179 	180 
181 	182 	183 	184 	185 
186 	187 	188 	189 	190 
191	192	193	194	195
196	197	198	199	200

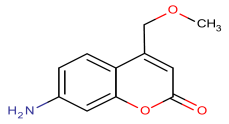
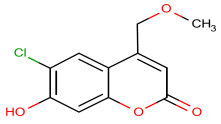
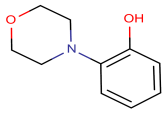
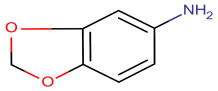
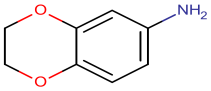
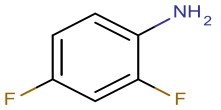
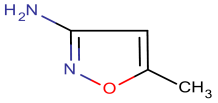
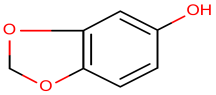
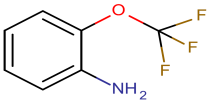
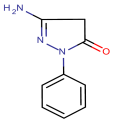
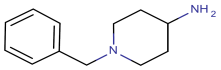
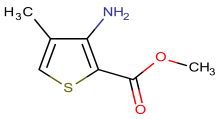
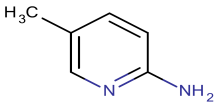
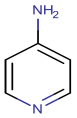
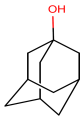
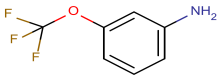
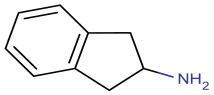
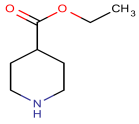
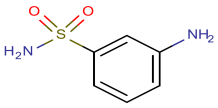
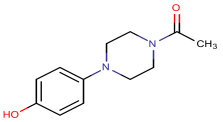
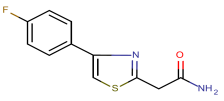
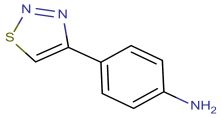
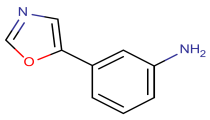
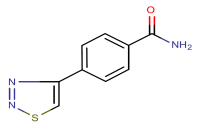
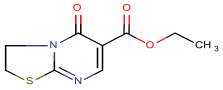
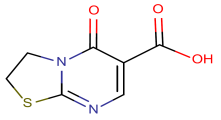
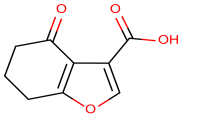
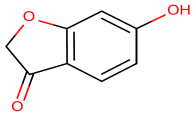
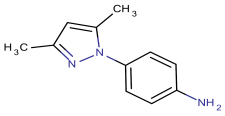
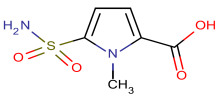
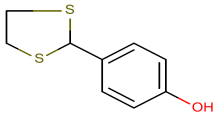
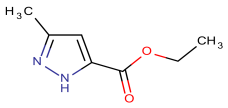
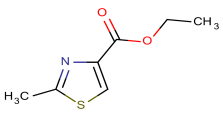
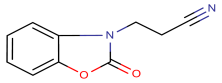
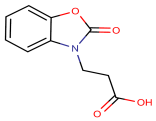
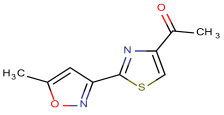
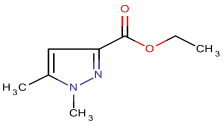
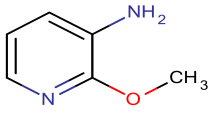
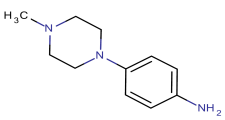
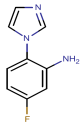
201 	202 	203 	204 	205 
206 	207 	208 	209 	210 
211 	212 	213 	214 	215 
216 	217 	218 	219 	220 
221 	222 	223 	224 	225 
226 	227 	228 	229 	230 
231 	232 	233 	234 	235 
236 	237 	238 	239 	240 
241	242	243	244	245
246	247	248	249	250

251 	252 	253 	254 	255 
256 	257 	258 	259 	260 
261 	262 	263 	264 	265 
266 	267 	268 	269 	270 
271 	272 	273 	274 	275 
276 	277 	278 	279 	280 
281 	282 	283 	284 	285 
286 	287 	288 	289 	290 
291	292	293	294	295
296	297	298	299	300

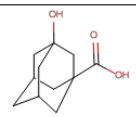
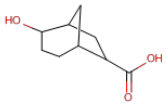
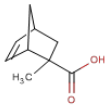
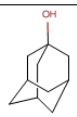
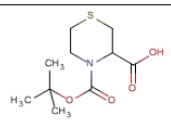
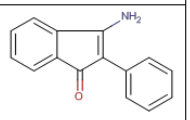
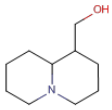
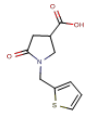
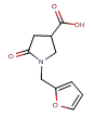
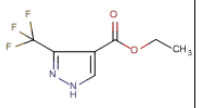
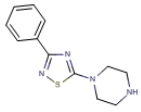
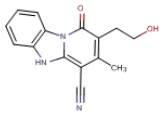
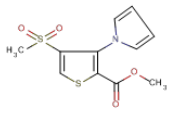
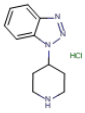
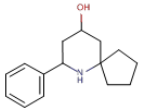
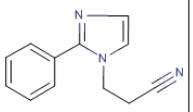
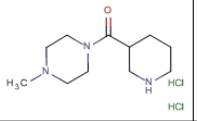
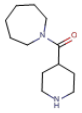
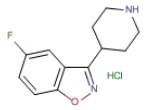
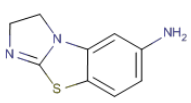
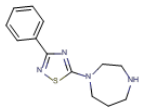
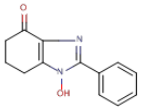
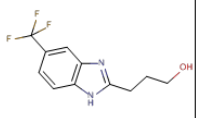
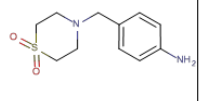
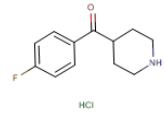
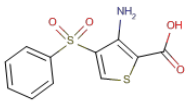
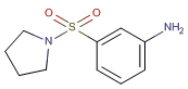
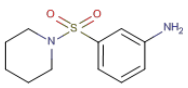
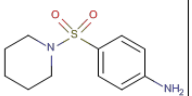
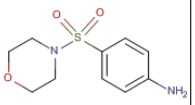
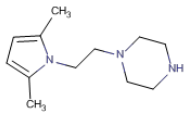
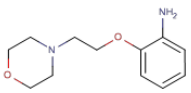
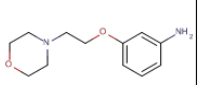
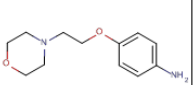
301 	302 	303 	304 	305 
306 	307 	308 	309 	310 
311 	312 	313 	314 	315 
316 	317 	318 	319 	320 
321 	322 	323 	324 	325 
326 	327 	328 	329 	330 
331 	332 	333 	334 	335 
336 	337 	338 	339 	340 
341	342	343	344	345
346	347	348	349	350

351 	352 	353 	354 	355 
356 	357 	358 	359 	360 
361 	362 	363 	364 	365 
366 	367 	368 	369 	370 
371 	372 	373 	374 	375 
376 	377 	378 	379 	380 
381 	382 	383 	384 	385 
386 	387 	388 	389 	390 
391	392	393	394	395
396	397	398	399	400

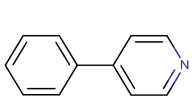
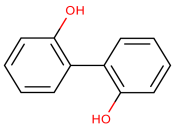
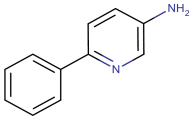
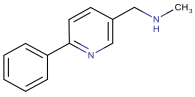
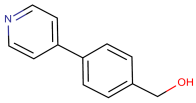
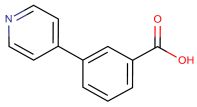
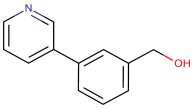
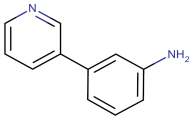
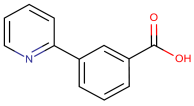
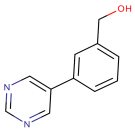
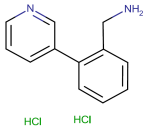
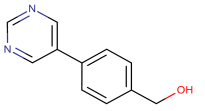
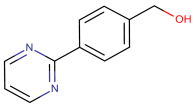
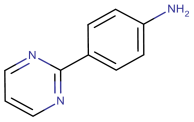
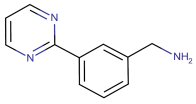
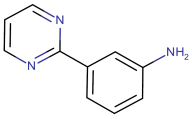
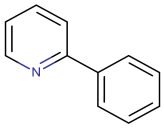
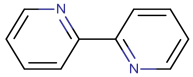
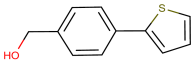
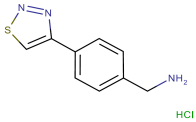
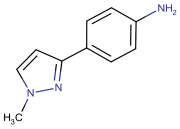
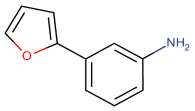
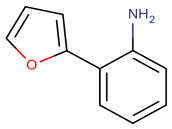
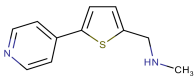
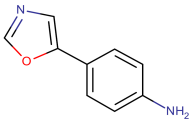
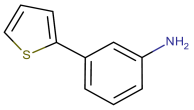
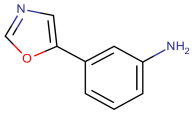
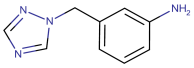
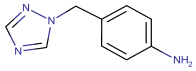
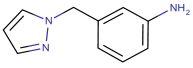
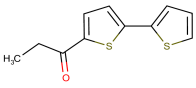
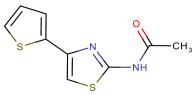
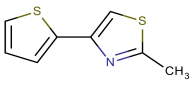
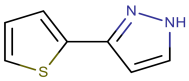
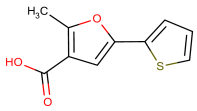
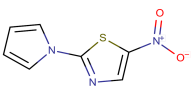
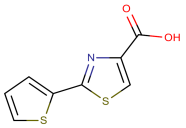
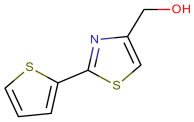
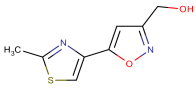
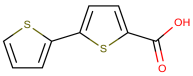
401 	402 	403 	404 	405 
406 	407 	408 	409 	410 
411 	412 	413 	414 	415 
416 	417 	418 	419 	420 
421 	422 	423 	424 	425 
426 	427 	428 	429 	430 
431 	432 	433 	434 	435 
436 	437 	438 	439 	440 
441	442	443	444	445
446	447	448	449	450

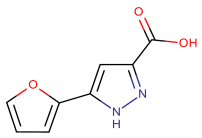
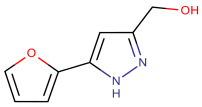
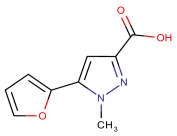
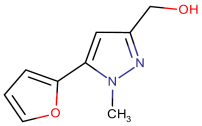
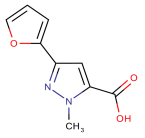
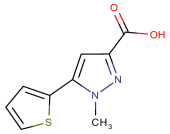
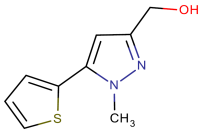
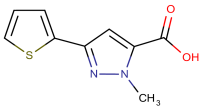
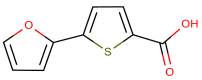
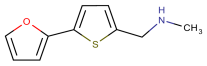
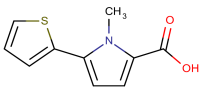
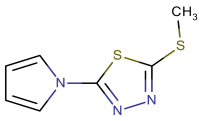
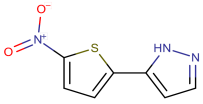
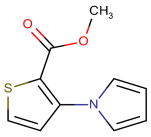
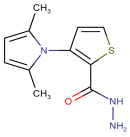
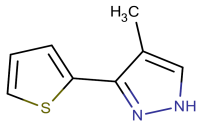
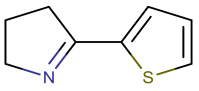
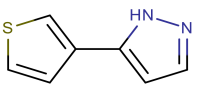
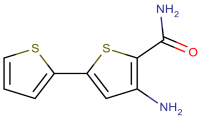
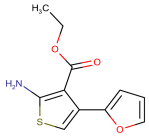
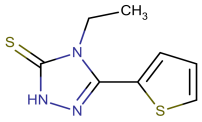
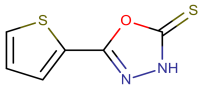
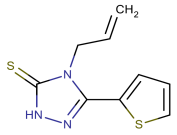
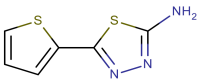
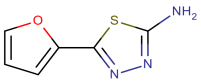
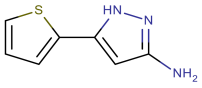
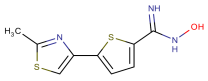
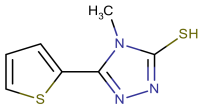
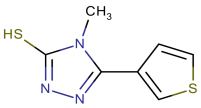
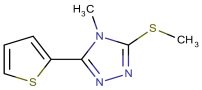
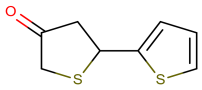
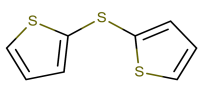
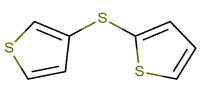
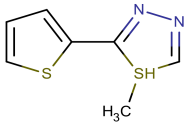
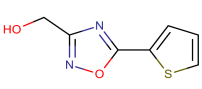
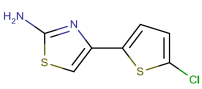
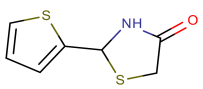
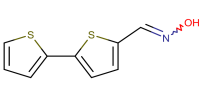
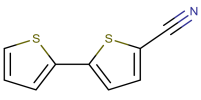
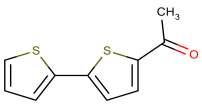
451 	452 	453 	454 	455 
456 	457 	458 	459 	460 
461 	462 	463 	464 	465 
466 	467 	468 	469 	470 
471 	472 	473 	474 	475 
476 	477 	478 	479 	480 
481 	482 	483 	484 	485 
486 	487 	488 	489 	490 
491	492	493	494	495
496	497	498	499	500

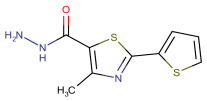
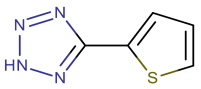
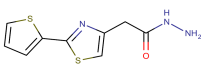
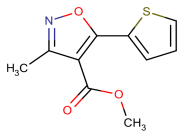
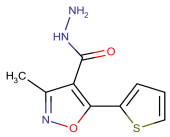
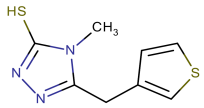
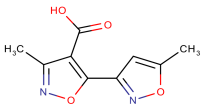
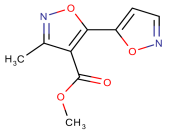
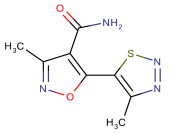
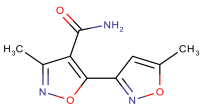
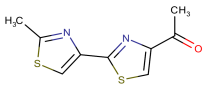
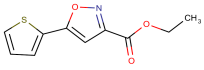
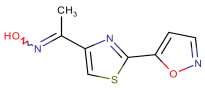
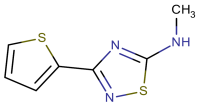
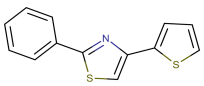
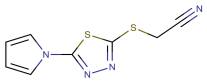
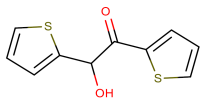
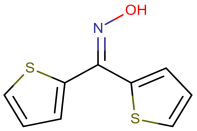
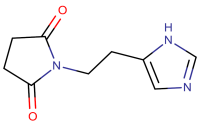
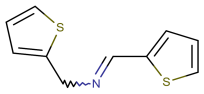
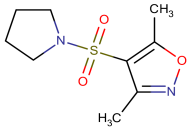
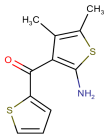
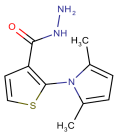
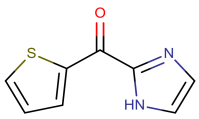
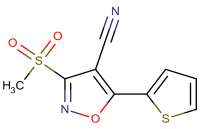
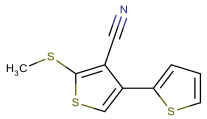
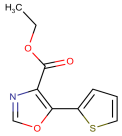
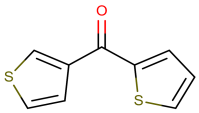
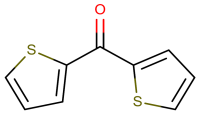
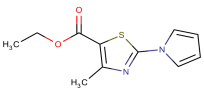
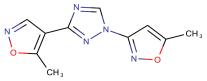
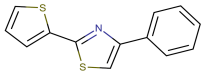
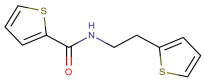
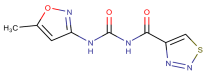
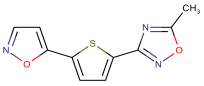
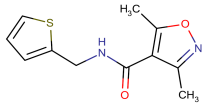
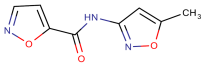
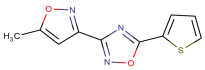
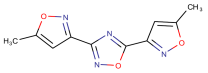
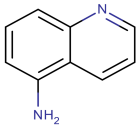
CR2006 Fragments: Scaffold Cluster S10 – Complex Structures

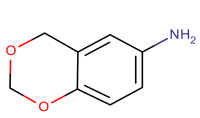
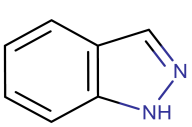
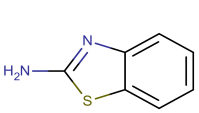
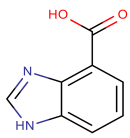
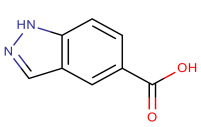
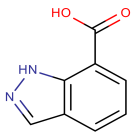
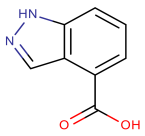
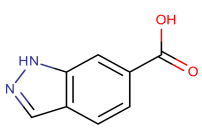
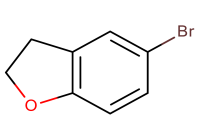
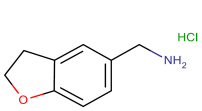
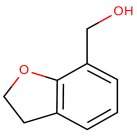
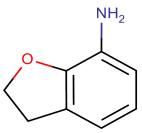
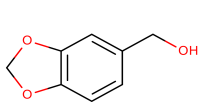
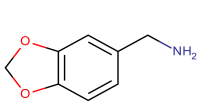
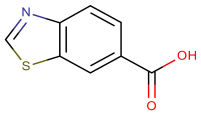
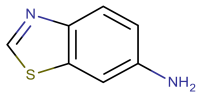
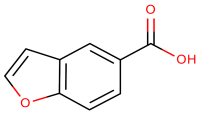
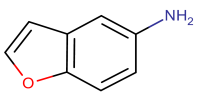
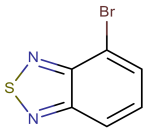
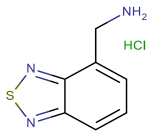
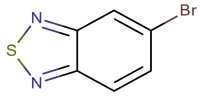
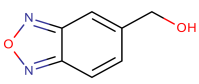
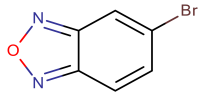
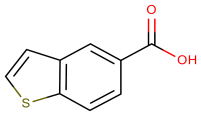
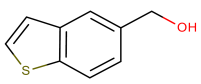
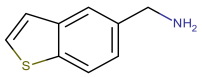
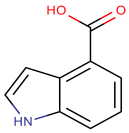
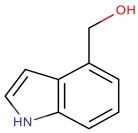
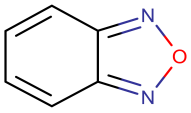
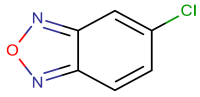
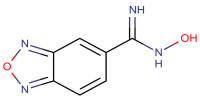
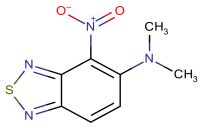
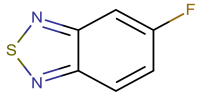
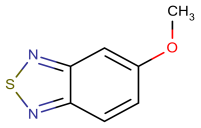
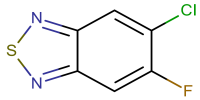
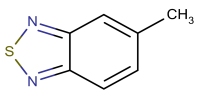
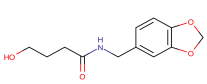
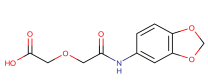
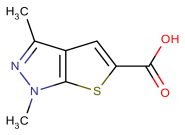
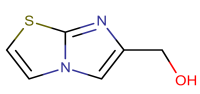
52	60	445	465	323	371
					
59	376	377	379	382	384
					
397	407	450	427	429	430
					
432	2	10	58	26	416
					
372	396	433	434	498	499
					
421	320	321	322		
					

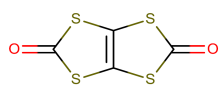
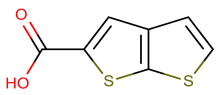
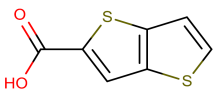
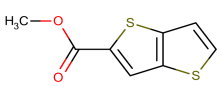
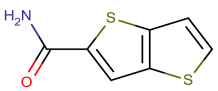
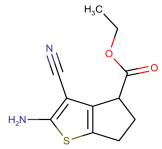
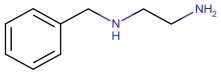
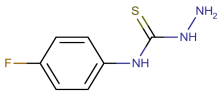
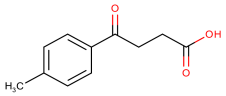
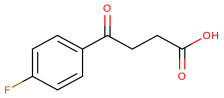
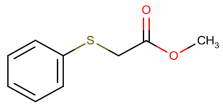
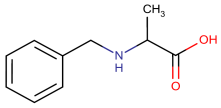
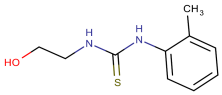
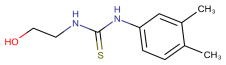
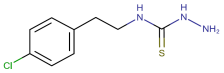
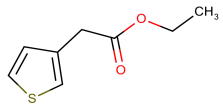
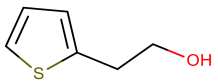
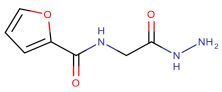
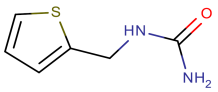
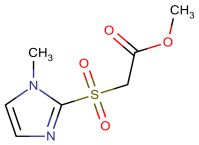
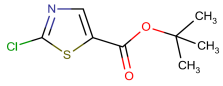
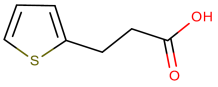
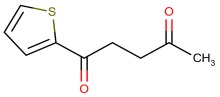
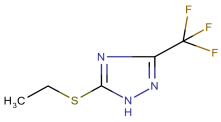
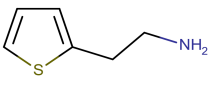
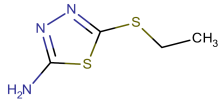
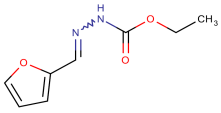
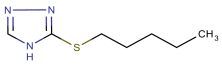
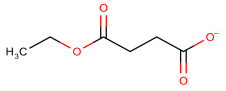
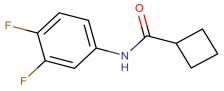
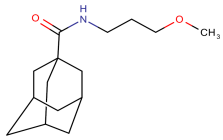
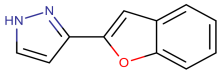
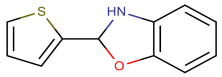
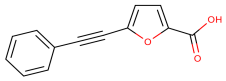
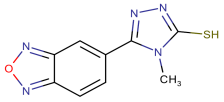
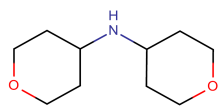
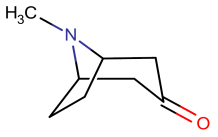
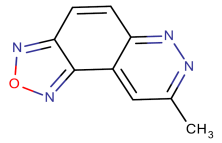
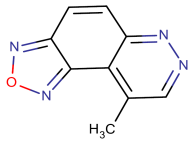
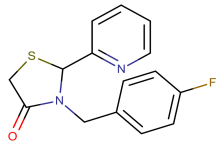
Maybridge TS2017 Library

TS001 	TS002 	TS003 	TS004 	TS005 
TS006 	TS007 	TS008 	TS009 	TS010 
TS011 	TS012 	TS013 	TS014 	TS015 
TS016 	TS017 	TS018 	TS019 	TS020 
TS021 	TS022 not available	TS023 	TS024 	TS025 
TS026 	TS027 	TS028 	TS029 	TS030 
TS031 	TS032 	TS033 	TS034 	TS035 
TS036 	TS037 	TS038 	TS039 	TS040 
TS041 	TS042	TS043	TS044	TS045
TS046	TS047	TS048	TS049	TS050

TS051 	TS052 	TS053 	TS054 	TS055 
TS056 	TS057 	TS058 	TS059 	TS060 
TS061 	TS062 	TS063 	TS064 	TS065 
TS066 	TS067 	TS068 	TS069 	TS070 
TS071 	TS072 	TS073 	TS074 	TS075 
TS076 	TS077 	TS078 	TS079 not available	TS080 
TS081 	TS082 	TS083 	TS084 	TS085 
TS086 	TS087 	TS088 	TS089 	TS090 
TS091 	TS092	TS093	TS094	TS095
TS096	TS097	TS098	TS099	TS100

TS101	TS102	TS103	TS104	TS105
				
TS106	TS107	TS108	TS109	TS110
				
TS111	TS112	TS113	TS114	TS115
				
TS116	TS117	TS118	TS119	TS120
				
TS121	TS122	TS123	TS124	TS125
				
TS126	TS127	TS128	TS129	TS130
				
TS131	TS132	TS133	TS134	TS135
				
TS136	TS137	TS138	TS139	TS140
				
TS141	TS142	TS143	TS144	TS145
				not available
TS146	TS147	TS148	TS149	TS150

TS151 	TS152 	TS153 	TS154 	TS155 
TS156 	TS157 	TS158 	TS159 	TS160 
TS161 	TS162 	TS163 	TS164 	TS165 
TS166 	TS167 	TS168 	TS169 	TS170 
TS171 	TS172 	TS173 	TS174 	TS175 
TS176 	TS177 	TS178 	TS179 not available	TS180 
TS181 	TS182 	TS183 	TS184 	TS185 
TS186 	TS187 	TS188 	TS189 	TS190 
TS191 	TS192	TS193	TS194	TS195
TS196	TS197	TS198	TS199	TS200

TS201	TS202	TS203	TS204	TS205
				
TS206	TS207	TS208	TS209	TS210
				
TS211	TS212	TS213	TS214	TS215
				
TS216	TS217	TS218	TS219	TS220
				
TS221	TS222	TS223	TS224	TS225
				
TS226	TS227	TS228	TS229	TS230
				
TS231	TS232	TS233	TS234	TS235
				
TS236	TS237	TS238	TS239	TS240
				
TS241	TS242			

Dendrogram of combined Maybridge CR2006/TS2017 Library

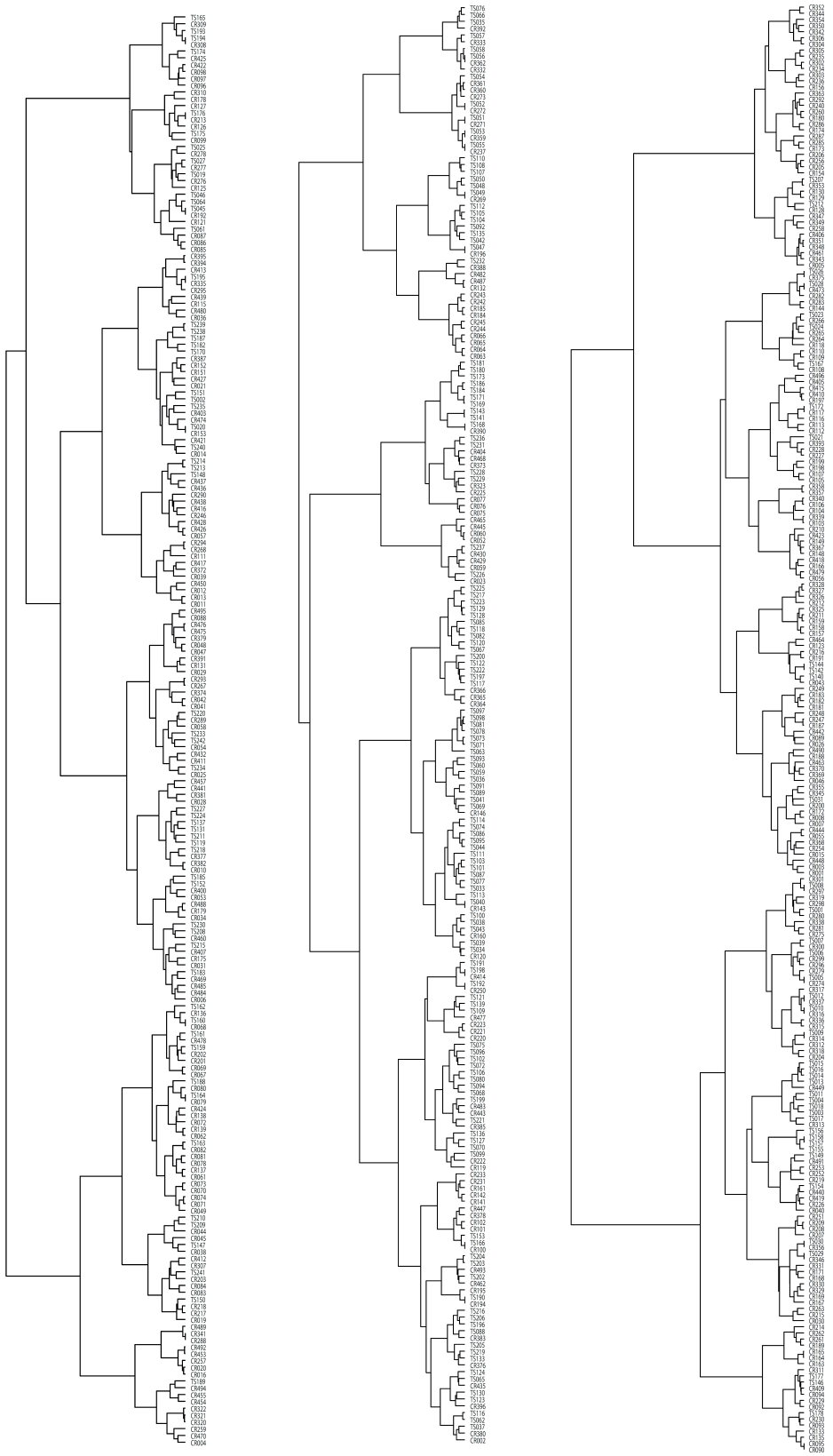


Figure A1 | Dendrogram of 657 water-soluble fragments

Mixtures for Fragment Screening

Table A1 | Maybridge library CR2006 – fragment mixtures

Number of compounds:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Mix 1	2	128	282	285	301	306	390	418	454	455	459	463	464	469						
Mix 2	19	41	55	57	61	69	81	92	96	115	121	144	157	163	181	411				
Mix 3	63	70	78	89	94	103	113	141	148	182	189	192	203	207	227	299				
Mix 4	73	95	108	112	133	184	204	214	237	252	308	315	325	332	364	412				
Mix 5	15	31	75	194	197	229	271	296	357	395	396	414	422	444	445	485				
Mix 6	21	43	53	100	135	191	212	242	259	318	336	358	362	410	435	476				
Mix 7	30	34	123	166	216	277	283	297	311	317	339	346	359	369	388	439				
Mix 8	25	38	40	49	98	106	119	159	168	222	250	273	304	335	394	405				
Mix 9	1	36	219	248	295	331	340	350	360	385	409	426	436	442	475	482				
Mix 10	288	321	354	361	415	419	437	483	487	491										
Mix 11	3	45	246	320	349	478	293	329	423	402										
Mix 12	54	172	179	249	266	280	303	322	368	370	374	391	392	400	448	460				
Mix 13	4	6	39	65	68	72	79	87	99	102	105	129	146	169	199	223				
Mix 14	66	77	107	111	120	126	130	136	152	153	156	209	215	258	269	352				
Mix 15	125	127	178	233	244	251	260	263	275	286	298	302	310	327	338	382				
Mix 16	7	14	47	64	76	86	213	245	265	278	281	287	305	328	347	387				
Mix 17	42	48	71	93	97	104	109	142	154	158	164	187	268	313	348	353				
Mix 18	74	82	101	110	117	118	143	149	151	160	167	183	188	294	366	406				
Mix 19	8	116	161	185	195	196	200	208	230	247	254	264	272	276	300	461				
Mix 20	171	198	228	231	243	253	262	274	309	316	326	333	351	378	441	453				
Mix 21	319	330	375	379	380	393	424	440	443	462	466	479	490							
Mix 22	267	279	312	337	365	367	438	465	470	473	492	494								
Mix 23	5	10	11	12	16	20	23	26	29	46	52	62	67	80	83	85	121	131	132	202
Mix 24	13	28	56	58	74	84	88	139	174	175	180	201	220	221	226	235	236	240	257	323
Mix 25	44	60	90	210	211	217	261	289	314	137	138	173	205	225	234	307	343	377	413	417
Mix 26	206	218	256	290	292	341	342	344	355	356	372	373	381	404	407	425	427	429	430	
Mix 27	93	309	319	363	368	383	402	428	447	449	450	457	468	474						
Mix 28	416	432	477	480	484	488	489	493	495	496	345	376	403							

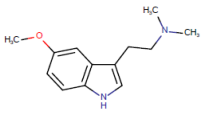
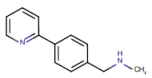
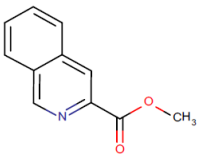
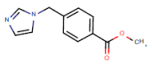
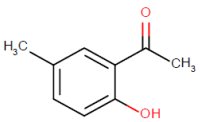
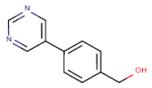
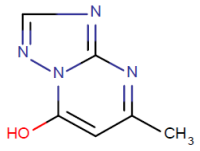
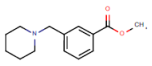
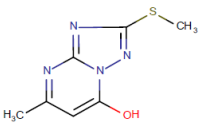
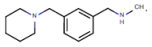
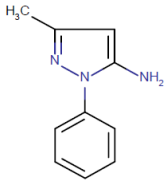
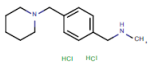
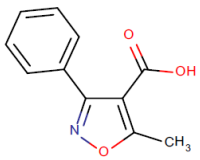
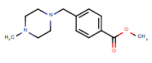
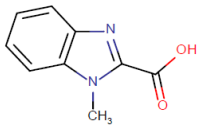
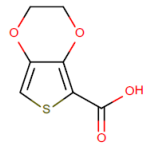
Table A2 | Maybridge library TS2017– fragment mixtures

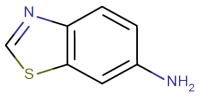
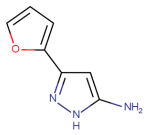
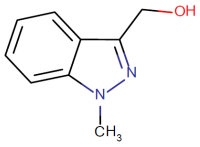
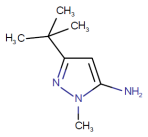
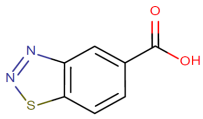
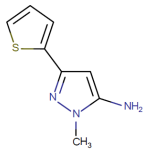
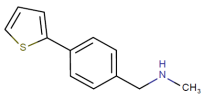
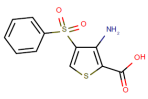
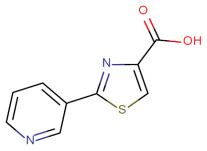
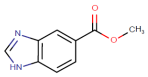
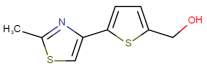
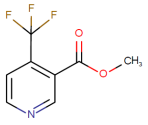
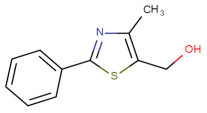
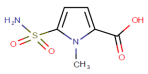
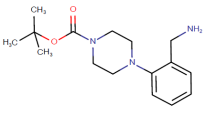
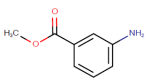
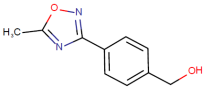
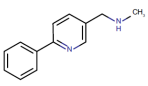
Number of fragments	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Mix 1	1	2	4	7	9	19	33	36	51	54	70	78	109	114	121	123	156	159
Mix 2	3	5	8	12	14	16	21	27	31	50	65	82	88	91	104	108	116	122
Mix 3	6	10	13	17	25	38	40	44	62	68	71	89	94	99	113	119	127	163
Mix 4	11	15	20	24	28	30	35	46	57	67	77	97	110	130	133	139	142	164
Mix 5	18	23	29	39	42	63	66	86	92	106	111	120	131	136	160	196	204	206
Mix 6	26	34	47	49	64	73	80	87	95	100	103	117	140	166	172	183	186	199
Mix 7	37	41	45	52	58	93	107	124	135	148	154	157	173	185	190	193	210	221
Mix 8	43	48	55	59	61	96	101	118	137	147	165	180	194	212	214	231	237	239
Mix 9	53	56	81	105	128	141	144	149	151	158	187	202	213	223	226	238		
Mix 10	60	72	76	98	129	161	181	188	215	217	220	224	236	208				
Mix 11	69	74	85	102	143	150	162	169	192	197	219	229	225					
Mix 12	75	112	152	168	175	207	211	218	228	235	242							
Mix 13	90	146	155	174	178	189	200	216	234	241	240							
Mix 14	134	153	170	182	191	227	209	230	233									
Mix 15	167	177	195	203	222	198	171	184	205									

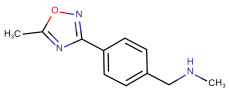
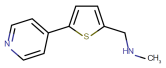
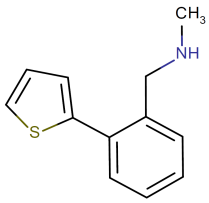
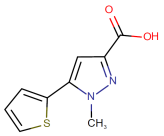
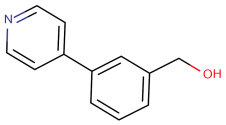
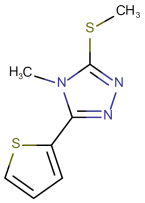
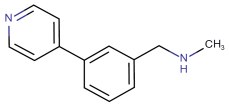
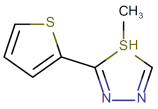
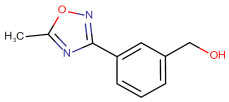
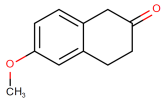
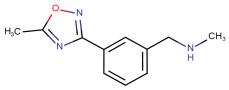
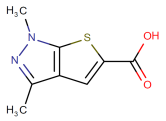
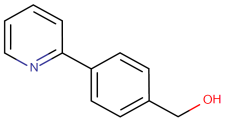
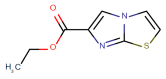
Red: manually allocated

Beige: Stock solution contains 5% perdeuterated DMSO

Compounds showing direct Irradiation

ID	Structure	log P	Solubility	ID	Structure	log P	Solubility
CR001		-1.36	+++	CR313		-0.9	+++
CR040		1.93	+++	CR331		1.81	+++
CR045		2.39	++	CR337		0.93	++
CR047		0.13	+++	CR342		-0.73	+++
CR048		1.44	+++	CR343		-4.42	+++
CR056		1.45	+++	CR351		-4.42	+++
CR063		-1.3	+++	CR354		-1.74	+++
CR089		-2.01	+++	CR364		-2.47	++

CR100		1.28	+++	CR388		0.63	+
CR104		0.73	+++	CR391		1.59	+++
CR115		-1.83	++	CR392		1.47	+++
CR125		-0.29	+++	CR396		-1.21	++
CR141		-2.05	+++	CR419		1.26	++
CR160		1.93	+++	CR442		1.64	+++
CR231		2.15	+++	CR480		-3.98	+++
CR246		-1.08	+++	CR491		1.15	+++
CR267		1.6	+++	TS004		-0.9	+++

CR268		-1.49	+++	TS025		-1.37	+++
CR278		-0.29	++	TS056		-1.74	+++
CR279		1.64	+++	TS081		2.23	+++
CR281		-1.28	+++	TS085		1.13	+++
CR293		1.6	+++	TS150		2.09	+++
CR294		-1.49	+++	TS190		-2.05	+++
CR312		2.02	+++	TS192		1.34	+++

Note: Compounds **CR173**, **CR287**, **CR344**, **CR348**, **CR349**, **CR406**, **CR429**, **CR480** should also be analyzed, as these fragments show broad unexpected signals in the 1D spectrum

List of STD NMR hits for WNV protease

Table A3 | List of positive hits from STD NMR screeing targeting the WNV protease

CR-#	Code	Lot No	H Acceptors	Hdonors	Rotatable Bonds	Clog P	Mw	Aq Sol Mm	Appearance	soluble	STD Signal Signal	STD Effect max	Protease act	S.E.	Reduction
1	AC 11641	320949	2	1	4	1,71	218,3	30,6	solid	WAHR	xx	0,0243	72%	4%	28%
8	AC 18635	301475	3	1	1	1,62	161,2	571,2	solid	WAHR	x	0,0107	83%	7%	17%
15	BTB 01214	277971	1	1	2	2,5	184,2	84,1	solid	WAHR	xx	0,0202	67%	9%	33%
25	BTB 05052	325631	4	0	3	1,7	218,2	6,8	solid	WAHR	xxx	0,0399	29%	4%	71%
34	BTB 08015	325630	2	1	2	2,36	222,2	7,8	solid	WAHR	x	0,0105	118%	3%	0%
38	BTB 09252	189209	3	0	4	1,46	180,2	4,3	liquid	WAHR	xxx	0,0602	105%	6%	0%
40	BTB 09392	315181	3	0	2	2,33	187,2	10,3	solid	WAHR	xxx	0,0384	67%	6%	33%
47	BTB 10447	28848	4	1	0	0,81	150,1	119,8	solid	WAHR	xx	0,025	39%	1%	61%
49	BTB 12187	325642	3	0	1	0,67	178,2	10,7	solid	WAHR	xxx	0,0467	97%	2%	3%
53	BTB 13048	320712	3	1	0	0,89	135,1	19,6	solid	WAHR	x	0,0056	47%	1%	53%
54	BTB 13395	279791	2	1	1	2,27	177,2	6,7	solid	WAHR	xxx	0,0399	67%	3%	33%
61	CC 00401	254327	4	1	1	1,11	194,2	4,1	solid	WAHR	x	0,0121	96%	5%	4%
65	CC 00513	299661	3	1	2	1,77	188,2	234,2	liquid	WAHR	x	0,0048	81%	5%	19%
72	CC 01313	291215	3	1	1	0,48	165,2	442,6	liquid	WAHR	x	0,0066	79%	7%	21%
73	CC 01401	283162	4	1	1	1,06	180,2	13,6	solid	WAHR	x	0,0122	96%	2%	4%
75	CC 01501	256765	4	1	1	1,1	180,2	48,3	solid	WAHR	x	0,019	84%	13%	16%
76	CC 01509	213803	3	1	1	0,99	166,2	176,3	solid	WAHR	x	0,0141	93%	9%	7%
79	CC 01613	300126	3	1	1	0,5	151,2	546,4	liquid	WAHR	x	0,0053	98%	2%	2%
81	CC 01701	314876	4	1	1	1,08	166,1	16,8	Solid	WAHR	x	0,0188	2%	1%	98%
89	CC 03801	300690	3	1	1	1,37	176,2	12,4	Solid	WAHR	xx	0,0229	91%	11%	9%
93	CC 04509	315526	2	1	1	2,07	159,2	111,0	solid	WAHR	x	0,0073	100%	21%	0%
98	CC 05623	312187	3	0	3	2,09	207,3	2,3	Solid	WAHR	xxx	0,059	73%	4%	27%
100	CC 05914	319577	1	1	0	0,66	150,2	43,0	solid	WAHR	xxx	0,044	96%	2%	4%
102	CC 06013	316536	2	1	1	1,07	200,7	578,1	solid	WAHR	x	0,0118	100%	14%	0%
104	CC 06309	313695	1	3	1	0,77	162,2	243,8	Liquid	WAHR	x	0,0034	87%	2%	13%
106	CC 06323	307859	3	0	3	1,86	204,2	3,2	Solid	WAHR	xxx	0,0359	85%	7%	15%
109	CC 06609	292354	2	1	1	1,23	148,2	69,6	liquid	WAHR	xx	0,0322	65%	6%	35%
112	CC 09101	312842	4	1	1	1,71	180,2	27,1	solid	WAHR	xxx	0,0359	12%	1%	88%
115	CC 09401	230716	4	1	1	1,31	180,2	27,1	solid	WAHR	x	0,0138	57%	5%	43%
118	CC 10209	303257	3	1	2	1,24	175,2	133,2	solid	WAHR	xx	0,0301	87%	3%	13%
119	CC 10422	208907	4	0	2	0,7	141,1	181,4	solid	WAHR	xx	0,0329	71%	12%	29%
120	CC 11513	275688	2	1	2	1,28	196,3	184,9	liquid	WAHR	xx	0,0222	54%	1%	46%
121	CC 11601	209808	3	1	2	1,87	205,2	2,3	solid	WAHR	xx	0,0273	71%	2%	29%
141	CC 13701	296953	4	1	2	1,72	206,2	11,4	solid	WAHR	xx	0,0225	71%	4%	29%
149	CC 15509	267197	2	1	2	2,01	188,2	49,6	solid	WAHR	x	0,0035	104%	3%	0%
151	CC 16209	226736	3	1	2	2,37	189,2	55,7	solid	WAHR	x	0,0045	86%	2%	14%
152	CC 16213	292488	3	1	2	2,02	188,2	363,5	liquid	WAHR	x	0,016	92%	3%	8%
166	CC 18414	267957	1	1	1	1,75	159,2	73,1	solid	WAHR	xx	0,025	32%	1%	68%
169	CC 18546	311565	2	1	3	1,9	187,2	188,5	Solid	WAHR	x	0,0034	80%	2%	20%
172	CC 18614	318114	2	1	1	1,12	160,2	360,6	solid	WAHR	x	0,0125	110%	2%	0%
178	CC 19446	307941	2	1	3	2,34	218,3	6,1	Liquid	WAHR	x	0,005	39%	3%	61%
179	CC 19614	297478	2	1	2	2,54	186,2	6,0	solid	WAHR	xx	0,0263	43%	5%	57%
181	CC 20101	272950	4	1	2	2,49	214,2	5,1	solid	WAHR	x	0,0033	66%	2%	34%
182	CC 20201	287624	4	1	2	2,79	214,2	5,1	solid	WAHR	x	0,0045	82%	8%	18%
185	CC 20309	286665	3	1	2	2,11	189,2	35,9	solid	WAHR	x	0,0096	112%	4%	0%
188	CC 21746	310269	2	1	2	1,1	175,2	178,8	Solid	WAHR	x	0,0056	87%	1%	13%
189	CC 21901	241886	2	1	2	2,91	187,2	1,9	solid	WAHR	x	0,0138	86%	5%	14%
191	CC 21914	321688	0	1	1	2,43	158,2	14,8	solid	WAHR	xxx	0,0362	59%	5%	41%
192	CC 22101	240727	2	1	2	1,8	193,2	2,4	solid	WAHR	x	0,0155	54%	2%	46%
196	CC 23109	290872	3	1	2	1,16	181,2	171,5	solid	WAHR	xx	0,0208	78%	7%	22%
204	CC 24201	239434	3	1	2	2,58	199,2	1,8	solid	WAHR	x	0,0084	57%	7%	43%
212	CC 24614	281541	2	1	1	0,56	179,2	165,1	solid	WAHR	xx	0,0309	68%	3%	32%
213	CC 24813	318974	1	1	1	1,46	163,2	71,1	liquid	WAHR	x	0,0158	70%	4%	30%
219	CC 26523	295787	3	0	3	1,1	190,2	4,3	Solid	WAHR	xxx	0,0407	94%	6%	6%
227	CC 28101	312046	3	1	2	1,77	202,2	2,5	solid	WAHR	x	0,0125	79%	6%	21%
243	CC 30509	310554	3	1	2	1,86	175,2	133,2	Solid	WAHR	x	0,0157	110%	10%	0%
244	CC 30513	315059	3	1	2	1,52	174,2	869,0	solid	WAHR	x	0,0164	60%	1%	40%
248	CC 31323	300259	3	0	3	1,63	190,2	4,3	Solid	WAHR	x	0,0183	85%	9%	15%
251	CC 32913	311075	2	1	2	1,65	173,2	596,1	liquid	WAHR	x	0,0098	89%	9%	11%

254	CC 33146	316016	2	1	2	0,83	161,2	663,5	Liquid	WAHR	x	0,0094	95%	9%	5%
262	CC 34309	302208	1	1	2	2,68	173,2	37,3	Solid	WAHR	x	0,0179	61%	3%	39%
273	CC 35023	302327	4	1	4	1,24	206,2	10,0	solid	WAHR	xx	0,0253	54%	4%	46%
274	CC 35409	300276	2	1	2	2,35	185,2	35,3	Solid	WAHR	x	0,0076	102%	2%	0%
279	CC 35609	300999	2	1	2	2,35	185,2	35,3	Liquid	WAHR	x	0,0139	81%	4%	19%
280	CC 35614	307170	1	1	1	2,1	170,2	14,0	Solid	WAHR	x	0,0157	95%	10%	5%
285	CC 35901	311456	3	1	2	1,34	220,3	6,2	Solid	WAHR	x	0,0166	81%	10%	19%
286	CC 35913	305424	2	1	2	0,76	205,3	397,3	liquid	WAHR	x	0,005	62%	10%	38%
287	CC 36009	300945	2	1	2	1,1	206,3	386,1	Solid	WAHR	x	0,0185	104%	2%	0%
288	CC 36014	308807	1	1	1	0,86	191,3	496,0	Solid	WAHR	xx	0,0307	72%	3%	28%
293	CC 39109	307346	4	1	2	1,73	190,2	92,9	Solid	WAHR	x	0,007	92%	14%	8%
295	CC 39222	303245	3	1	2	0,26	165,1	33,0	Solid	WAHR	xxx	0,071	104%	9%	0%
297	CC 39814	318944	1	1	1	2,1	170,2	14,0	Solid	WAHR	xx	0,0238	14%	3%	86%
300	CC 40009	300574	2	1	2	2,35	185,2	35,3	Solid	WAHR	xx	0,031	104%	9%	0%
304	CC 40122	311859	4	0	4	1,32	235,3	117,7	Solid	WAHR	x	0,0076	95%	7%	5%
305	CC 40146	313874	1	3	4	1,28	220,3	269,1	Liquid	WAHR	x	0,0054	86%	10%	14%
309	CC 40909	311052	1	1	1	0,59	167,2	63,6	Solid	WAHR	x	0,0056	96%	6%	4%
310	CC 40996	311975	1	1	2	0,65	216,7	109,2	Solid	WAHR	xx	0,0231	82%	11%	18%
313	CC 41546	315908	2	1	3	2,81	198,3	36,1	Liquid	WAHR	x	0,0139	76%	5%	24%
316	CC 41709	310322	3	1	2	1,7	186,2	189,9	Solid	WAHR	x	0,004	80%	1%	20%
317	CC 41714	311893	2	1	1	1,46	171,2	75,4	Solid	WAHR	xxx	0,0355	33%	1%	67%
328	CC 43146	315637	4	1	3	1	207,3	857,7	Liquid	WAHR	x	0,0061	94%	15%	6%
330	CC 43314	317360	1	1	2	1,46	173,2	22,0	Solid	WAHR	x	0,0184	106%	3%	0%
331	CC 43322	316793	3	0	4	1,97	216,2	2,4	Solid	WAHR	xxx	0,0359	35%	2%	65%
332	CC 43501	307385	3	2	2	1,21	194,2	12,0	solid	WAHR	x	0,0141	72%	6%	28%
333	CC 43509	308575	2	2	2	0,98	180,2	236,9	Solid	WAHR	x	0,0091	80%	3%	20%
337	CC 44309	310101	3	1	2	1,7	186,2	189,9	Solid	WAHR	x	0,0078	99%	2%	1%
346	CC 45414	313932	2	4	2	1,81	174,2	108,5	Solid	WAHR	x	0,0143	96%	4%	4%
348	CC 45513	314741	2	1	3	1,77	204,3	28,3	Liquid	WAHR	x	0,011	60%	6%	40%
351	CC 45596	312125	2	1	4	2,18	291,3	6,7	Solid	WAHR	x	0,0041	52%	2%	48%
352	CC 45609	311230	3	1	3	1,2	220,3	702,7	Solid	WAHR	x	0,0077	96%	11%	4%
354	CC 45622	310757	4	0	4	1,46	248,3	100,0	Liquid	WAHR	x	0,0193	76%	16%	24%
359	CC 46901	314652	4	1	2	1,12	192,2	10,3	Solid	WAHR	x	0,009	80%	5%	20%
360	CC 46923	315634	4	0	4	1,49	220,2	2,7	Solid	WAHR	xxx	0,0809	82%	13%	18%
361	CC 47023	315731	4	0	4	1,49	220,2	2,7	Solid	WAHR	x	0,0196	66%	6%	34%
362	CC 47101	315475	3	1	2	1,46	208,2	3,2	Solid	WAHR	x	0,013	99%	10%	1%
364	CC 47801	313428	1	4	1	0,24	186,2	8,9	Solid	WAHR	xxx	0,0644	45%	2%	55%
366	CC 47846	315962	3	1	2	0,4	185,2	180,2	Liquid	WAHR	xxx	0,094	64%	1%	36%
375	DFP 00120	319516	2	1	1	1,06	160,2	52,9	solid	WAHR	xx	0,0245	77%	2%	23%
379	GK 00722	325637	3	1	3	1,73	208,1	13,5	solid	WAHR	xx	0,0325	57%	1%	43%
382	HAN 00371	286930	3	1	2	2,29	246,3	8,5	solid	WAHR	x	0,0061	61%	3%	39%
390	CC 07114	318203	1	1	0	1,21	133,2	27,7	solid	WAHR	x	0,0188	108%	12%	0%
392	KM 00835	272951	1	1	1	1,16	179,2	25,3	solid	WAHR	xxx	0,0524	90%	4%	10%
394	KM 01494	125596	4	0	3	0,74	219,2	2,3	solid	WAHR	xx	0,021	67%	15%	33%
395	KM 01560	325645	4	1	1	0,37	191,2	17,2	solid	WAHR	x	0,0065	66%	8%	34%
396	KM 01757	180084	4	2	3	1,08	283,3	3,5	solid	WAHR	xxx	0,0642	78%	7%	22%
406	KM 09623	284387	2	1	3	1,91	205,3	35,5	liquid	WAHR	x	0,0137	92%	1%	8%
409	KM 10245	268692	3	0	2	2,33	187,2	5,6	solid	WAHR	xxx	0,0411	69%	1%	31%
412	KM 10844	272871	4	1	3	1,38	206,2	8,6	solid	WAHR	xx	0,021	80%	9%	20%
413	KM 10848	247356	2	1	3	1,37	167,2	4,1	solid	WAHR	x	0,0109	64%	5%	36%
436	NRB 00998	290657	2	1	1	0,8	203,2	7,2	solid	WAHR	x	0,0171	94%	8%	6%
438	NRB 04272	325651	0	2	2	2,56	198,3	9,9	solid	WAHR	x	0,0078	91%	3%	9%
440	RDP 00059	217675	4	0	2	0,95	188,2	41,1	solid	WAHR	xx	0,0236	77%	14%	23%
448	RJC 02157	280820	2	1	0	1,55	159,2	99,9	solid	WAHR	xx	0,0335	45%	2%	55%
455	SB 00353	275438	2	1	0	0,58	151,2	54,3	liquid	WAHR	x	0,0133	63%	4%	37%
461	SB 01177	263253	2	1	2	1,14	190,3	121,4	liquid	WAHR	x	0,0137	66%	6%	34%
462	SB 01297	312656	2	1	2	0,75	171,2	28,7	solid	WAHR	x	0,0063	65%	12%	35%
464	SB 01321	321589	1	1	0	0,42	94,1	774,1	solid	WAHR	xx	0,0275	86%	3%	14%
475	SEW 01371	41822	4	0	3	0,52	226,3	44,9	solid	WAHR	x	0,0154	106%	1%	0%
476	SEW 01382	200822	4	1	1	0,15	198,2	16,2	solid	WAHR	x	0,0094	93%	5%	7%
479	SEW 04169	257316	1	1	1	2,42	187,2	5,3	solid	WAHR	xx	0,0243	70%	5%	30%
482	SPB 01217	209034	3	1	3	0,86	154,2	75,4	solid	WAHR	xx	0,0227	101%	5%	0%
485	SPB 02696	195063	4	1	3	0,78	207,2	28,6	solid	WAHR	xx	0,025	72%	2%	28%

Twelve Most Potent Inhibitors of NS2B-NS3^{pro}

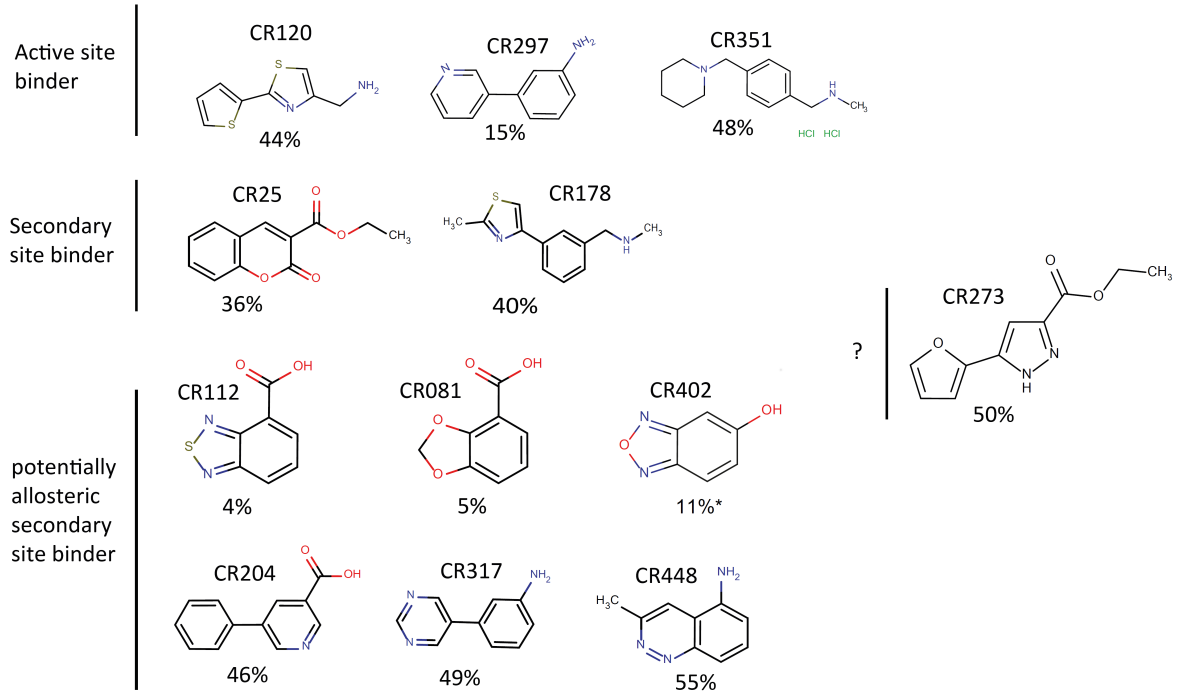


Figure A2 | Fragments exhibiting the highest inhibitory potency against the WNV protease
STD NMR competition assay revealed fragments, which compete for the active site of NS2B-NS3^{pro} (top) and putative second binding sites (below). Fragment **A** served as a competitor for the active site. Using a 20-fold excess of **A** over the concentration of fragments analyzed, the STD effect of fragments either decreased more than 25% (top), remained constant (middle) or increased more than 25% (bottom). The relative protease activity of 25 nM NS2B-NS3^{pro} supplemented with 900 µM fragment and 40 µM peptidic substrate is depicted below each chemical structure in percentage. The proteolytic activity of the WNV protease mixed with 40 µM substrate was set to 100%. The protease activity was measured twice. The average standard deviation of duplicate measurements was ~5-10%.

List of STD NMR hits for Norovirus Screening Targets

Table A4 | List of positive hits from STD NMR screeing targeting NoV P dimers

CR#	Code	Average STD effects MNV-1 P dimers	Average STD effects Saga P dimers	Average STD effects MI001 P dimers
1	AC 11641	0.022	0.006	0.000
2	AC 12308	0.016	0.000	0.000
3	AC 12586	0.007	0.004	0.005
4	AC 12605	0.022	0.004	0.006
6	AC 13432	0.014	0.000	0.000
10	AW 00426	0.053	0.015	0.018
11	AW 00496	0.010	0.000	0.000
14	AW 00577	0.025	0.000	0.000
15	BTB 01214	0.018	0.014	0.000
25	BTB 05052	0.035	0.000	0.000
26	BTB 05684	0.033	0.000	0.017
30	BTB 06289	0.007	0.004	0.004
31	BTB 07325	0.000	0.004	0.000
36	BTB 08846	0.023	0.000	0.000
38	BTB 09252	0.024	0.000	0.000
39	BTB 09284	0.037	0.007	0.006
40	BTB 09392	0.033	0.002	0.000
42	BTB 09930	0.013	0.003	0.000
43	BTB 10019	0.040	0.000	0.000
46	BTB 10358	0.016	0.000	0.000
49	BTB 12187	0.031	0.000	0.000
53	BTB 13048	0.052	0.041	0.000
54	BTB 13395	0.039	0.000	0.025
57	BTB 14322	0.031	0.000	0.000
61	CC 00401	0.000	0.000	0.005
62	CC 00413	0.008	0.000	0.000
65	CC 00513	0.011	0.004	0.004
66	CC 00546	0.019	0.005	0.004
68	CC 00813	0.023	0.004	0.003
72	CC 01313	0.013	0.000	0.003
74	CC 01409	0.012	0.000	0.000
75	CC 01501	0.013	0.000	0.000
77	CC 01513	0.042	0.012	0.000
79	CC 01613	0.018	0.000	0.003
80	CC 01646	0.013	0.000	0.006
82	CC 01709	0.000	0.000	0.005
87	CC 02213	0.045	0.000	0.000
89	CC 03801	0.012	0.000	0.000
97	CC 05609	0.014	0.000	0.000
98	CC 05623	0.076	0.000	0.000
99	CC 05646	0.024	0.003	0.006
100	CC 05914	0.029	0.004	0.003
102	CC 06013	0.041	0.000	0.000
104	CC 06309	0.007	0.002	0.003
106	CC 06323	0.046	0.009	0.000
107	CC 06346	0.012	0.004	0.004
109	CC 06609	0.027	0.000	0.000
110	CC 07109	0.016	0.000	0.000
111	CC 07113	0.054	0.000	0.000
118	CC 10209	0.033	0.000	0.000
120	CC 11513	0.061	0.000	0.005
125	CC 12046	0.069	0.008	0.000
126	CC 12296	0.061	0.003	0.002
127	CC 12313	0.068	0.000	0.000
131	CC 12513	0.000	0.000	0.005
136	CC 27213	0.012	0.003	0.002
143	CC 14009	0.016	0.000	0.000
146	CC 14613	0.035	0.010	0.000
149	CC 15509	0.003	0.000	0.000
151	CC 16209	0.011	0.000	0.000
152	CC 16213	0.105	0.006	0.007
153	CC 16413	0.047	0.012	0.000
156	CC 17446	0.000	0.015	0.000
160	CC 18209	0.019	0.004	0.000
166	CC 18414	0.018	0.000	0.000

167	CC 18509	0.020	0.000	0.000
168	CC 18528	0.024	0.000	0.000
169	CC 18546	0.025	0.003	0.004
172	CC 18614	0.000	0.012	0.000
178	CC 19446	0.047	0.007	0.010
179	CC 19614	0.022	0.000	0.020
183	CC 20209	0.011	0.000	0.000
188	CC 21746	0.006	0.013	0.013
191	CC 21914	0.068	0.013	0.012
192	CC 22101	0.008	0.000	0.000
195	CC 22909	0.013	0.000	0.000
196	CC 23109	0.024	0.000	0.000
208	CC 24409	0.015	0.000	0.000
209	CC 24413	0.026	0.000	0.000
213	CC 24813	0.034	0.000	0.000
215	CC 25513	0.079	0.024	0.000
216	CC 25514	0.029	0.000	0.013
218	CC 25809	0.051	0.000	0.000
219	CC 26523	0.047	0.000	0.000
228	CC 28109	0.006	0.000	0.000
231	CC 29209	0.024	0.000	0.004
233	CC 29246	0.023	0.005	0.006
234	CC 29309	0.003	0.000	0.000
235	CC 29313	0.000	0.000	0.008
240	CC 30113	0.025	0.000	0.000
242	CC 30501	0.009	0.000	0.000
243	CC 30509	0.045	0.000	0.000
244	CC 30513	0.036	0.000	0.014
245	CC 30546	0.009	0.000	0.000
246	CC 30713	0.010	0.003	0.003
249	CC 31614	0.007	0.010	0.000
250	CC 32523	0.021	0.000	0.000
251	CC 32913	0.038	0.013	0.018
258	CC 33946	0.014	0.000	0.000
260	CC 34046	0.010	0.003	0.005
262	CC 34309	0.035	0.000	0.000
263	CC 34313	0.032	0.000	0.000
264	CC 34409	0.043	0.000	0.000
265	CC 34413	0.087	0.000	0.000
266	CC 34414	0.064	0.000	0.000
267	CC 34809	0.000	0.012	0.004
268	CC 34846	0.000	0.006	0.000
269	CC 34913	0.052	0.000	0.000
273	CC 35023	0.081	0.011	0.008
274	CC 35409	0.018	0.000	0.000
275	CC 35446	0.027	0.000	0.000
276	CC 35509	0.072	0.000	0.000
277	CC 35514	0.047	0.000	0.000
278	CC 35546	0.038	0.000	0.000
279	CC 35609	0.020	0.008	0.004
280	CC 35614	0.025	0.000	0.009
281	CC 35646	0.018	0.000	0.000
282	CC 35714	0.029	0.000	0.000
285	CC 35901	0.000	0.007	0.000
288	CC 36014	0.024	0.004	0.000
292	CC 38513	0.025	0.000	0.000
294	CC 39146	0.011	0.006	0.009
295	CC 39222	0.053	0.010	0.000
298	CC 39846	0.040	0.009	0.000
300	CC 40009	0.015	0.000	0.000
301	CC 40014	0.039	0.000	0.000
304	CC 40122	0.015	0.003	0.000
310	CC 40996	0.047	0.007	0.005
311	CC 41401	0.012	0.005	0.010
313	CC 41546	0.013	0.000	0.003
316	CC 41709	0.011	0.000	0.000
319	CC 41814	0.020	0.000	0.000
330	CC 43314	0.015	0.000	0.000
331	CC 43322	0.028	0.003	0.000
332	CC 43501	0.000	0.000	0.007
333	CC 43509	0.022	0.000	0.000
337	CC 44309	0.000	0.011	0.007
338	CC 44346	0.000	0.017	0.013
340	CC 44423	0.041	0.000	0.000
341	CC 44914	0.042	0.000	0.000

348	CC 45513	0.014	0.000	0.004
350	CC 45522	0.008	0.000	0.000
351	CC 45596	0.000	0.000	0.004
353	CC 45613	0.000	0.000	0.006
354	CC 45622	0.012	0.000	0.000
358	CC 46601	0.000	0.000	0.002
360	CC 46923	0.043	0.004	0.000
362	CC 47101	0.015	0.000	0.000
367	CC 48009	0.000	0.006	0.002
372	CD 09967	0.027	0.000	0.000
375	DFP 00120	0.017	0.000	0.000
378	GK 00309	0.034	0.000	0.000
379	GK 00722	0.008	0.000	0.000
382	HAN 00371	0.048	0.015	0.023
385	HTS 10971	0.017	0.005	0.000
390	CC 07114	0.035	0.024	0.000
391	KM 00494	0.000	0.002	0.000
392	KM 00835	0.000	0.008	0.003
394	KM 01494	0.035	0.000	0.000
396	KM 01757	0.000	0.016	0.000
400	KM 06374	0.033	0.000	0.000
405	KM 08380	0.026	0.000	0.000
409	KM 10245	0.043	0.000	0.000
415	MO 00117	0.047	0.000	0.000
418	MO 00205	0.022	0.000	0.000
419	MO 00298	0.036	0.004	0.000
423	MO 00765	0.005	0.000	0.000
426	MO 00997	0.000	0.000	0.007
427	MO 01011	0.017	0.004	0.005
432	MO 07029	0.087	0.000	0.000
436	NRB 00998	0.021	0.003	0.000
437	NRB 00999	0.026	0.000	0.000
438	NRB 04272	0.029	0.000	0.000
440	RDP 00059	0.008	0.006	0.000
447	RJC 00659	0.042	0.000	0.000
448	RJC 02157	0.000	0.006	0.004
450	RJC 03989	0.008	0.000	0.000
455	SB 00353	0.010	0.007	0.000
460	SB 01062	0.028	0.000	0.000
462	SB 01297	0.014	0.007	0.000
463	SB 01317	0.009	0.008	0.000
464	SB 01321	0.000	0.012	0.008
465	SB 01577	0.000	0.010	0.004
470	SB 02047	0.000	0.006	0.000
473	SEW 00990	0.040	0.050	0.000
475	SEW 01371	0.014	0.000	0.000
479	SEW 04169	0.017	0.004	0.000
482	SPB 01217	0.017	0.000	0.000
483	SPB 02288	0.016	0.000	0.000
487	SPB 05352	0.019	0.000	0.000
488	SPB 07839	0.014	0.000	0.000
491	TL 00154	0.026	0.003	0.000
495	TL 00659	0.019	0.000	0.000

Table A5 | List of positive hits from STD NMR screeing targeting Saga P dimers

ID TS2017 Library	Average STD effects Saga P dimers
TS001	0.040
TS002	0.038
TS003	0.023
TS004	0.022
TS008	0.036
TS010	0.015
TS012	0.009
TS013	0.023
TS014	0.082
TS015	0.022
TS016	0.013
TS019	0.061
TS020	0.037
TS021	0.013
TS025	0.026
TS026	0.241
TS028	0.023
TS035	0.031
TS036	0.019
TS039	0.031
TS047	0.020
TS065	0.014
TS066	0.025
TS068	0.051
TS074	0.156
TS078	0.021
TS081	0.028
TS085	0.163
TS093	0.080
TS098	0.015
TS104	0.050
TS108	0.013
TS110	0.025
TS113	0.016
TS121	0.010
TS124	0.022
TS150	0.158
TS152	0.027
TS153	0.033
TS162	0.078
TS168	0.024
TS170	0.019
TS175	0.026
TS183	0.013
TS192	0.125
TS205	0.061
TS207	0.021
TS214	0.013
TS225	0.128
TS231	0.007
TS238	0.025
TS239	0.009

Saga P Dimers - Comparison of STD NMR Competition Assays and CSP Experiments

Table A6 | Evaluation of STD NMR competition assays of Saga P dimers

Compound	CSP region			normalized STD effect	Std
	I	II	III		
CR001	**2a**	2b	(3)	81%	7%
CR039	**2a**	2b	(3)	119%	6%
CR055	**(1)**	-	-	n.d.	
CR065	**2a**	2b	(3)	111%	0%
CR066	**2a**	2b	(3)	86%	
CR100	**2a**	(3)		102%	2%
CR106	**2a**	3	-	75%	16%
CR120	**2a**	2b	(3)	n.d.	
CR125	**2a**	2b	-	100%	6%
CR136	**2b**	-	-	87%	18%
CR196	**3**	(2a)	-	77%	2%
CR251	**2a**	(2b)	(3)	101%	6%
CR254	**2a**	-	-	n.d.	
CR367	**2a**	(2b)	(3)	81%	4%
CR396	**3**	(2a)	(1)	120%	17%
CR440	**1**	(3)	-	86%	5%
TS039	**2b**	-	-	n.d.	
TS153	**3**	2a	(2b)	65%	8%
TS162	**3**	(2b)	-	62%	18%
TS189	**1**	2a	3	n.d.	
TS207	**2a**	2b	-	81%	12%

Dendrogram of Fragments that Bind to MNV-1 P Dimers

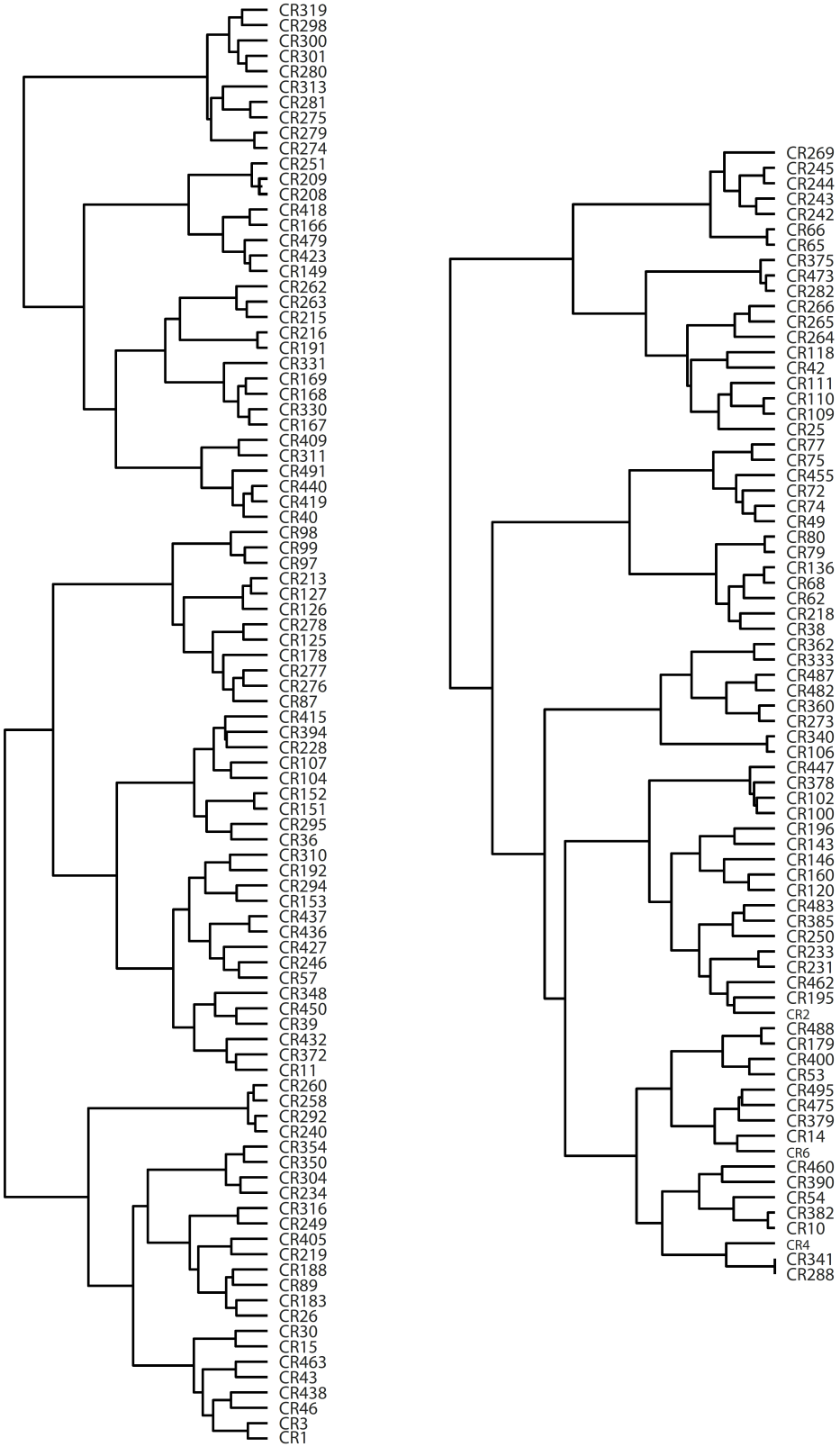


Figure A3 | STD NMR Hits of MNV-1 P dimer

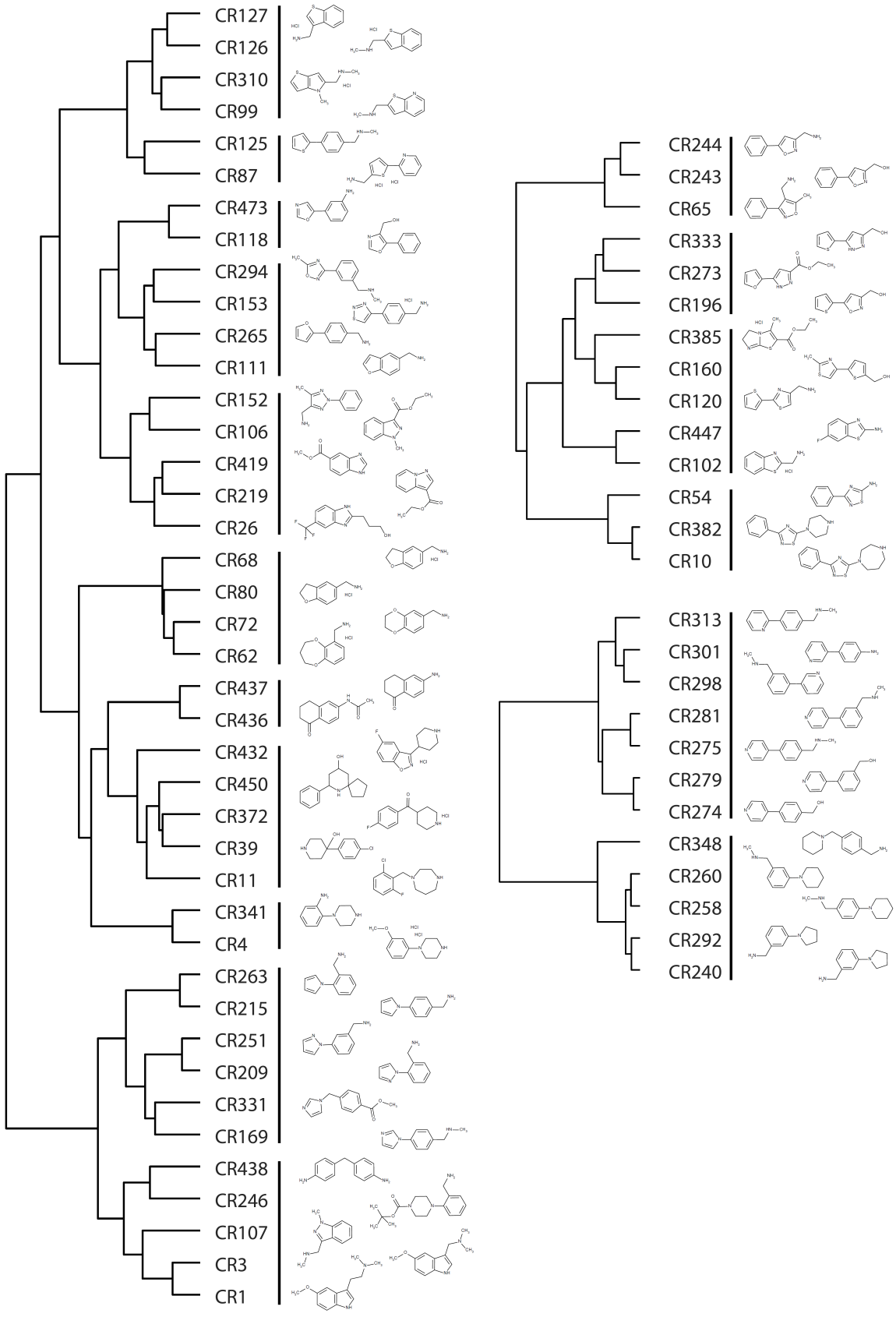


Figure A4 | Positive hits from SPR screening

Dendrogram of Fragments that Bind to Saga P Dimers

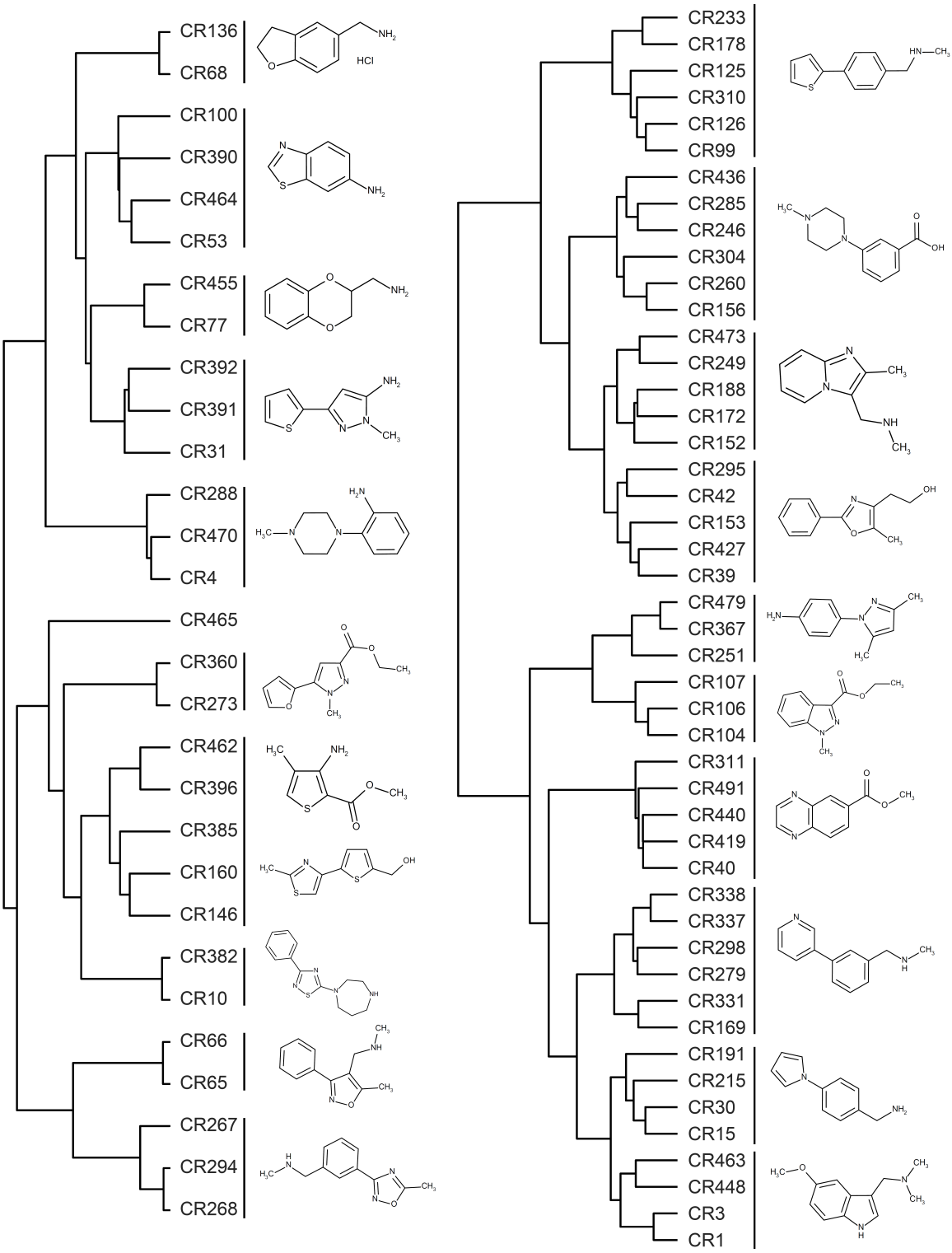


Figure A5 | Dendrogram of CR2006 fragments binding to Saga P dimers

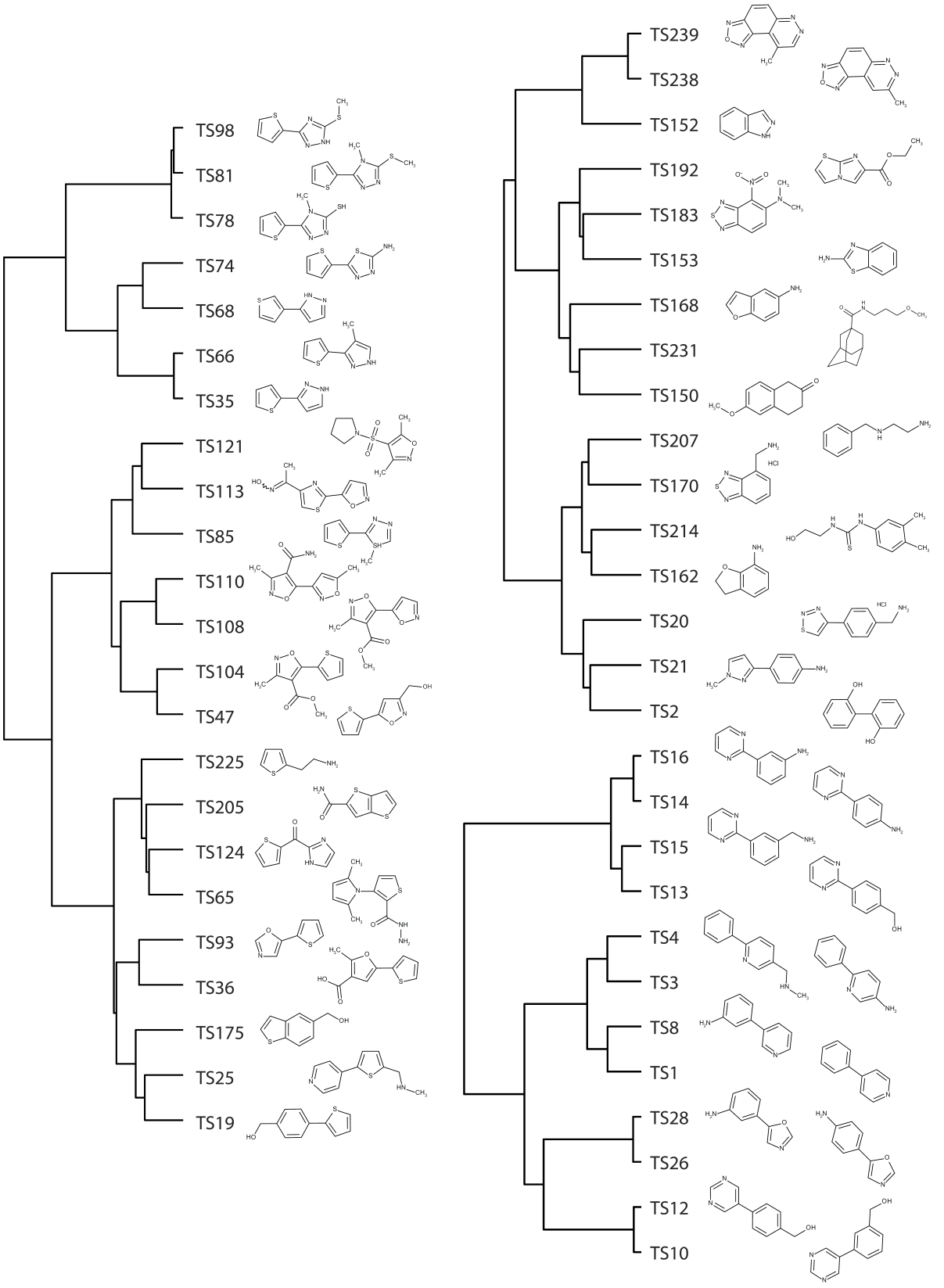


Figure A6 | Dendrogram of TS2017 fragments binding to Saga P dimers

Dendrogram of Fragments that Bind to MI001 P Dimers

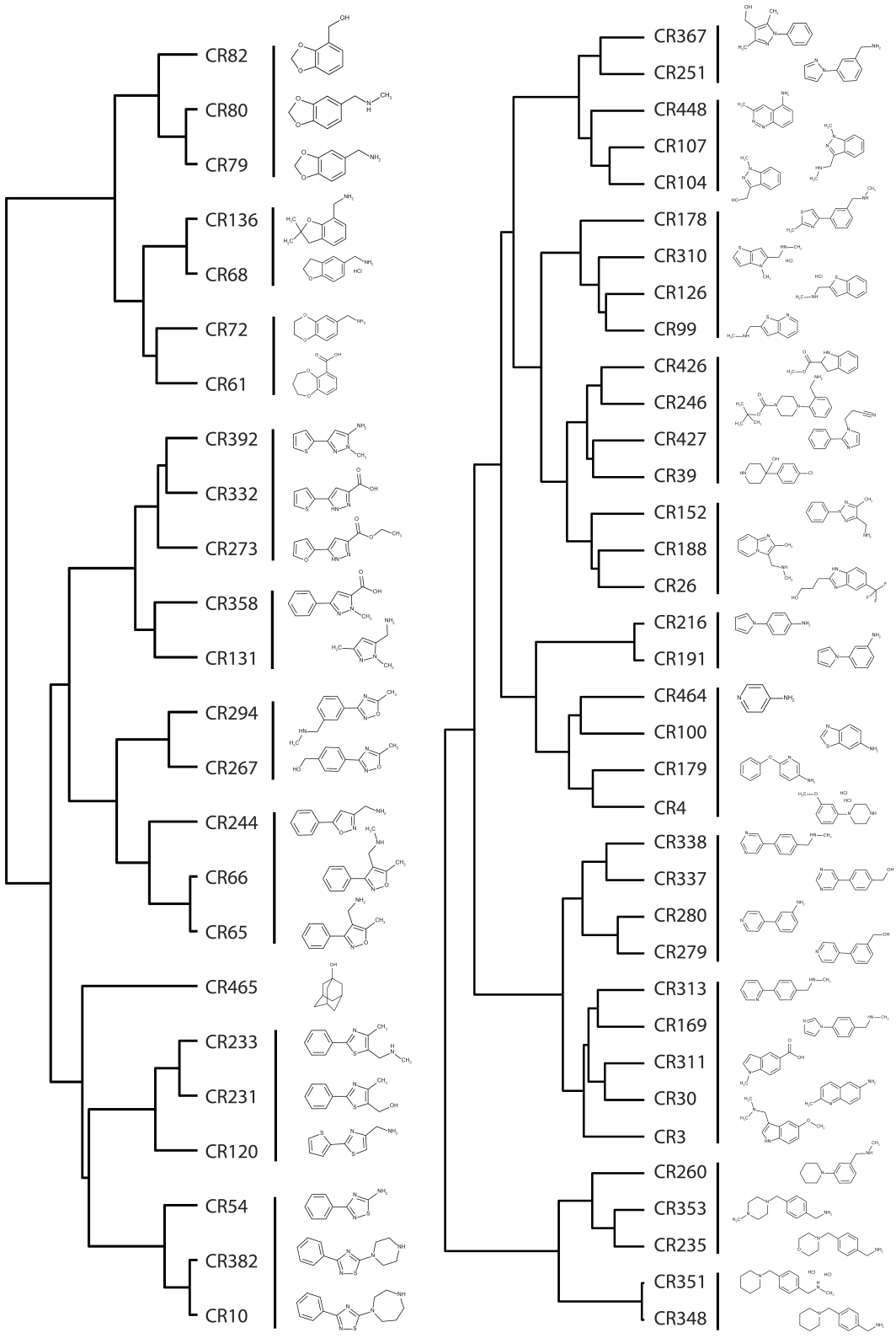


Figure A7 | Dendrogram of CR2006 fragments binding to MI001 P dimers

A3 2D Projections of 3D Plots

PCA – STD NMR Hits of Norovirus P Dimers

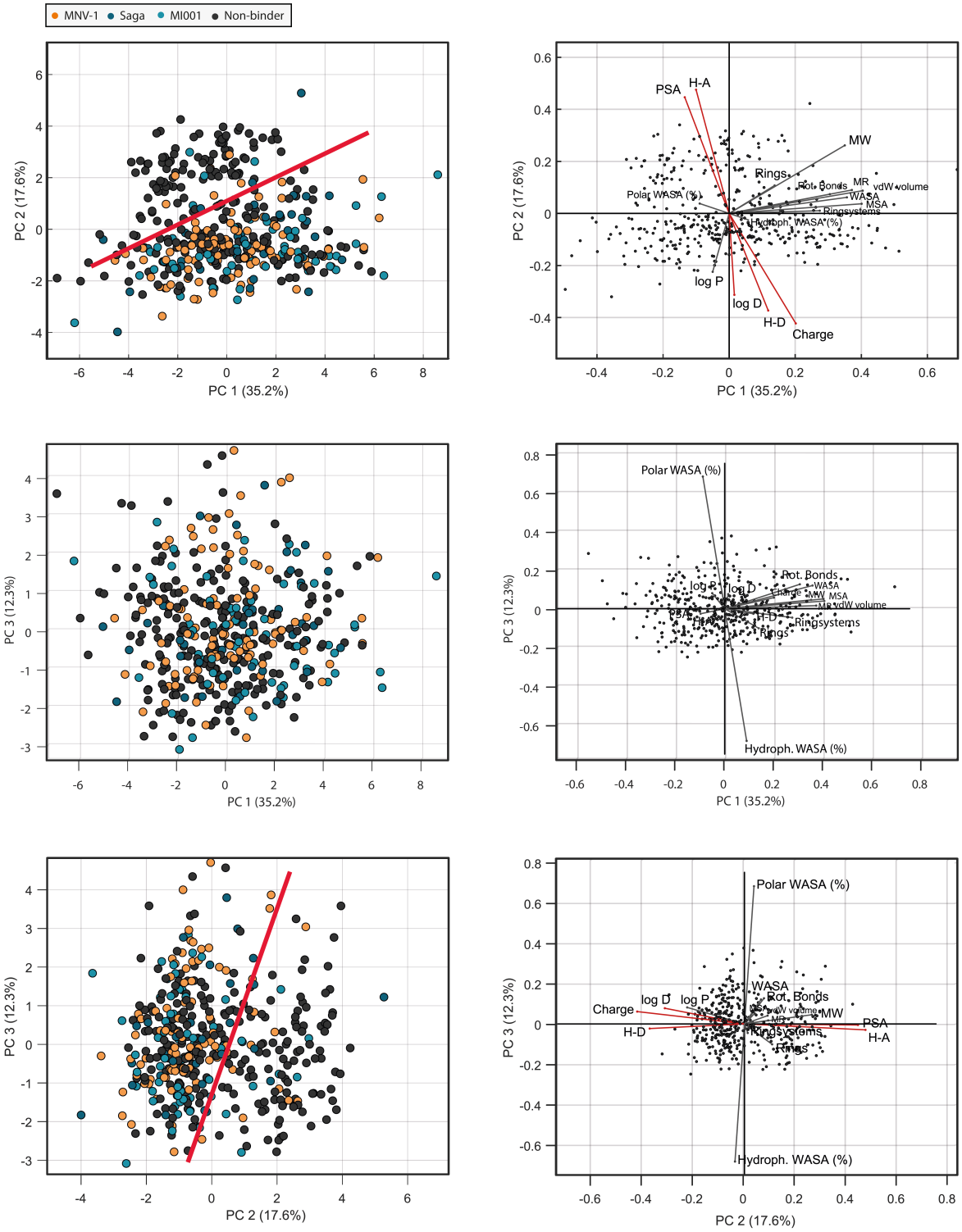


Figure A8 | Comparison of Hits based on the principal component analysis

PCA – Soluble and Poorly Soluble Fragments

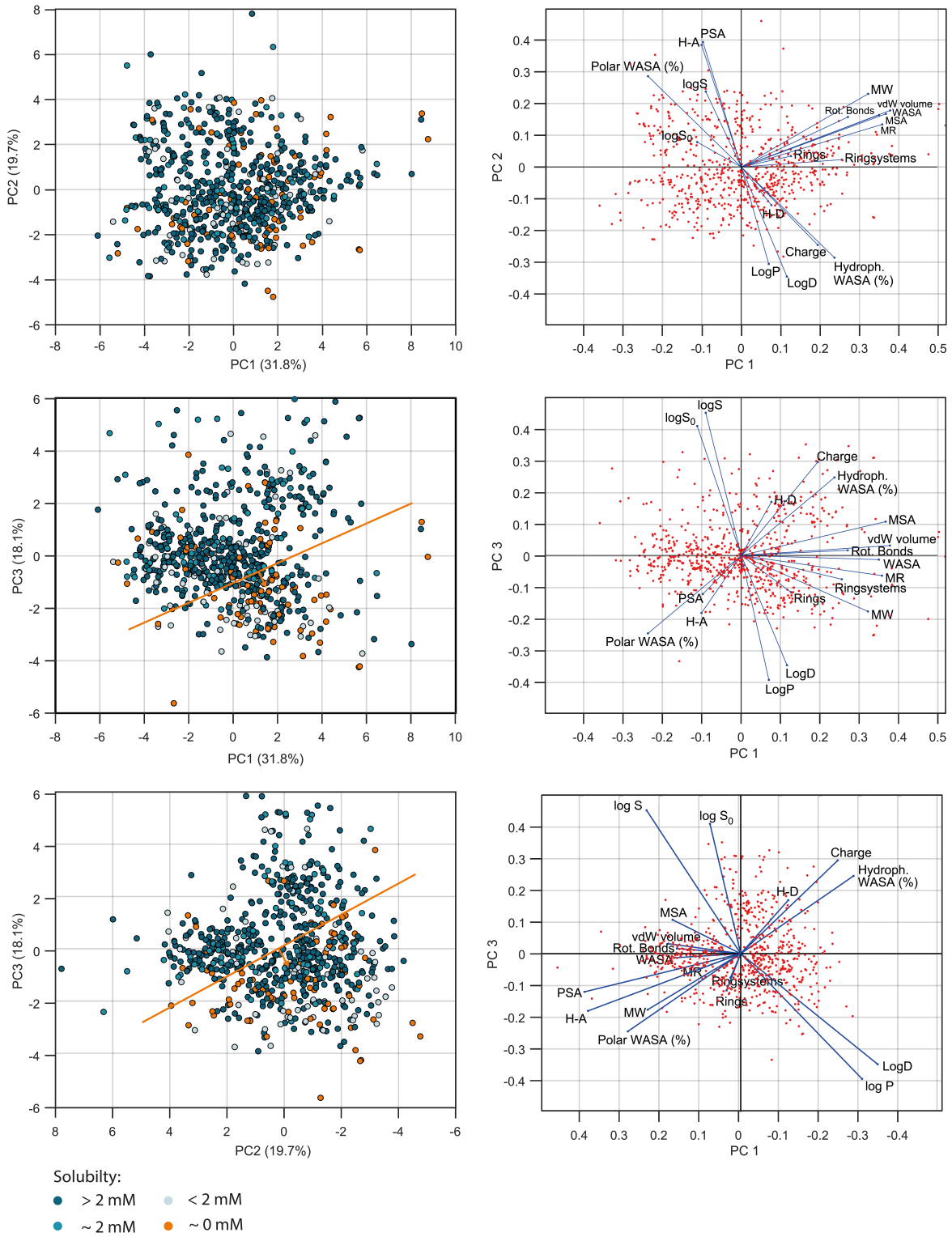


Figure A9 | PCA of the Solubility of Fragments

3D Histograms- STD NMR Hits of Norovirus P dimers

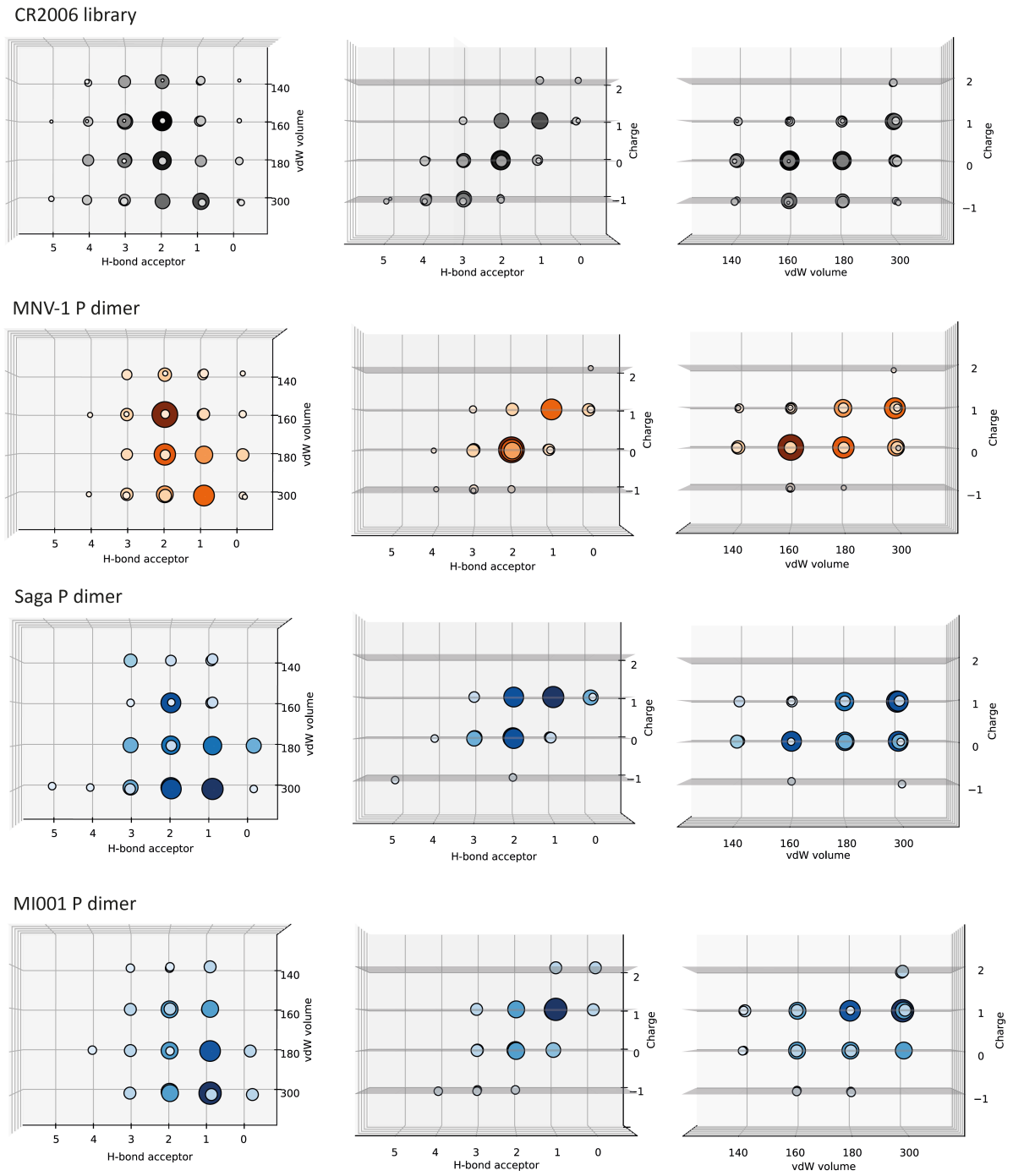


Figure A10 | Comparison of hits from the Norovirus screening based on structural properties

3D Histograms – Fragments of the Maybridge Libraries

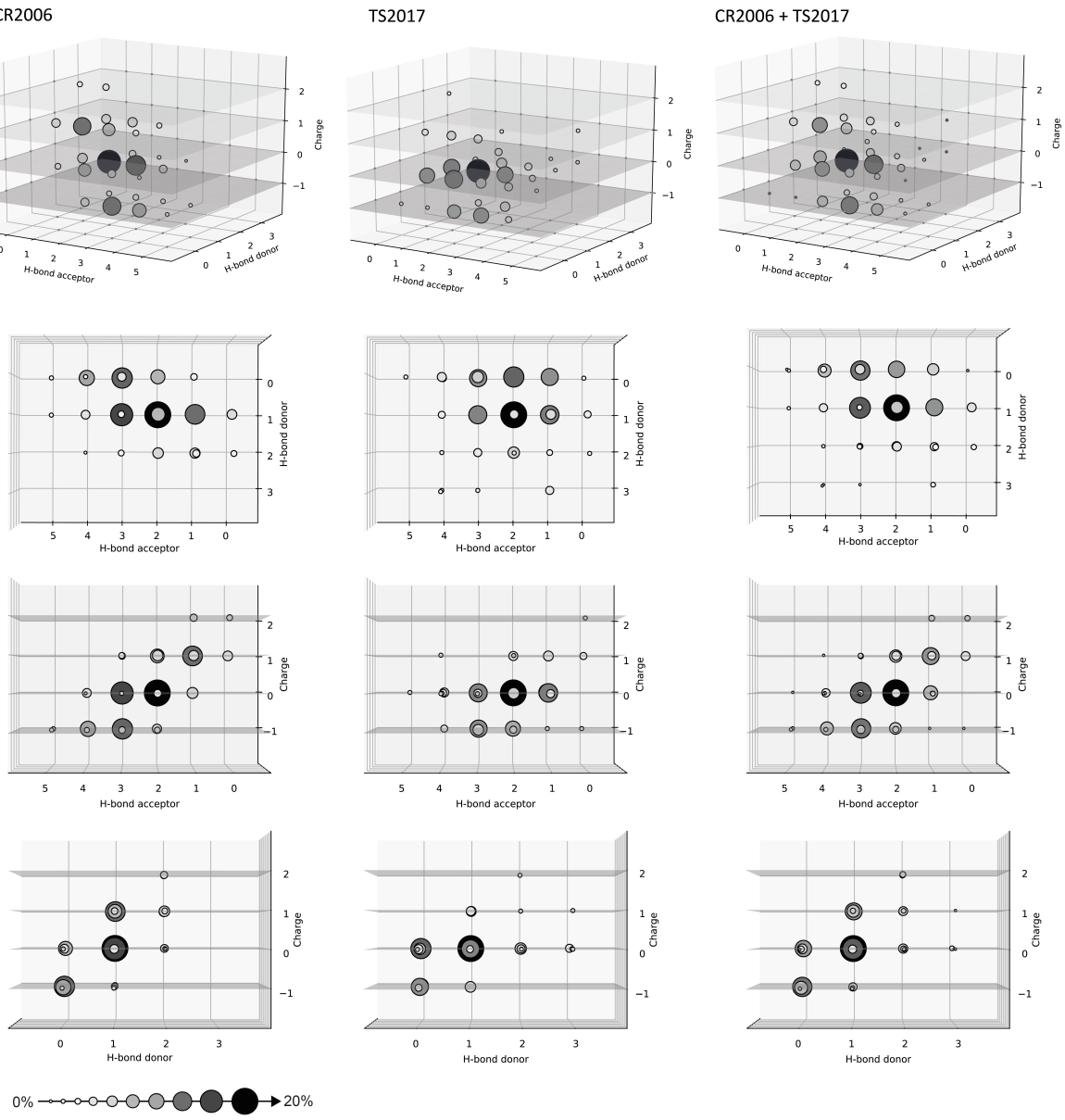


Figure A11 | Fragment selection for a library design

A4 Python, Matlab, and Maestro Scripts

calcChemParam_Library.py

```
### Tobias Schöne, 31.07.2017

#Calculates solubility, 223pri, MV and other Parmater of molecule/SMILE from Xcel file
provided by Maybidge
# using ChemAxon Suite and openbabel/ obprob in openbabel


import os
import sys
import os.path
from openpyxl import load_workbook


Library = 'M_TS_2017' #or M_TS_2017
Start = 1


if Library == 'M_CR_2006':
    Anzahl_Fragments = 500
    col_SMILES = 3
    Initials = 'CR'
if  Library == 'M_TS_2017':
    Anzahl Fragments = 238
    Initials = 'TS'
    col_SMILES = 8



#### replace comma with point
def replaceComma(logS_value):
    temp_array = []
    for j in range(0, len(logS_value)):
        if (logS_value[j] == ','):
            temp_array.append('.')
        else:
            temp array.append(logS value[j])
    S_value = ''.join(temp_array)  ### convert list to string
    S_value_float = float(S_value)
    return S_value_float


##### create matrix with Comp + SMILES

wb = load_workbook('C:/python/%s/%s.xlsx'%(Library,Library))
ws = wb.active


SMILES_array = [[]]

for comp in range(1,Anzahl_Fragments+1):
    SMILES array[comp-1].append(ws.cell(row=comp+  1, column=1).value)  ## comp ID
    if '.' In (ws.cell(row=comp + 1, column=col_SMILES).value):
        counter = 0
        for i in range(0, len((ws.cell(row=comp + 1, column=col_SMILES).value))):
            if ((ws.cell(row=comp + 1, column=col_SMILES).value)[i] == '.'):
                counter = counter +1
                if counter == 1:
                    a = i +1
                if counter == 2:
                    a = i +1
        SMILES_array[comp - 1].append((ws.cell(row=comp + 1, column=col_SMILES).value)[a:])
    else:
        SMILES array[comp-1].append(ws.cell(row=comp + 1, column=col SMILES).value) # SMILES
    SMILES array[comp - 1].append(ws.cell(row=comp + 1, column=2).value)  # Structure
    SMILES array.append([])
```

```
print(SMILES_array)
print('')
# create xcel sheet
os.system("copy EMPTY.xlsx %s_Param.xlsx" %Library)
print('')

wb = load_workbook('C:/python/%s_Param.xlsx' %Library)
ws = wb.active

ws.cell(row=1, column=1).value = 'Compound'
ws.cell(row=1, column=2).value = 'Structure'
ws.cell(row=1, column=3).value = 'SMILES'
ws.cell(row=1, column=4).value = 'canonical_SMILES'
ws.cell(row=1, column=5).value = 'MW'
ws.cell(row=1, column=6).value = 'logS'
ws.cell(row=1, column=7).value = 'logS_0'
ws.cell(row=1, column=8).value = 'Solubility (mM)'
ws.cell(row=1, column=9).value = 'Sol catarogy'
ws.cell(row=1, column=10).value = 'LogP chemaxon'
ws.cell(row=1, column=11).value = 'Charge at 7.0'
ws.cell(row=1, column=12).value = 'LogD chemaxon'
ws.cell(row=1, column=13).value = 'H-bond acceptor'
ws.cell(row=1, column=14).value = 'H-bond donor'
ws.cell(row=1, column=15).value = 'num_rings'
ws.cell(row=1, column=16).value = 'num_ringsystems'
ws.cell(row=1, column=17).value = 'num_rotable bonds'
ws.cell(row=1, column=18).value = 'molecular surface area (3D)'
ws.cell(row=1, column=19).value = 'van der Waals volume (3D)'
ws.cell(row=1, column=20).value = 'Topological polar surface area (2D)'
ws.cell(row=1, column=21).value = 'Water accessible surface area ASA'
ws.cell(row=1, column=22).value = 'ASA+'
ws.cell(row=1, column=23).value = 'ASA-'
ws.cell(row=1, column=24).value = 'ASA_hydrophobic'
ws.cell(row=1, column=25).value = 'ASA_polar'
ws.cell(row=1, column=26).value = 'Molar refractivity'
ws.cell(row=1, column=27).value = 'Name'


ws = wb.save('C:/python/%s_Param.xlsx' %Library)


### cxcalc usage: evaluate -e "logS('7.0', 'category')" "CC(O)=O" or cxcalc logs -U mol/l -i
true -H 7.0 'Smiles'
### frist output is easier to extract
### mview COC1=CC2=C(NC=C2CCNIC)C=C1 show structure in marvinview

wb = load_workbook('C:/python/%s_Param.xlsx' % Library)
ws = wb.active

for x in range(Start,Anzahl_Fragments+1):
    print("Fragment: %s" %(SMILES_array[x-1][0]))
    ###SMILES needs to be in quotes!!!
    SMILES = SMILES array[x-1][1]
    print("SMILE:", SMILES)

    wb = load_workbook('C:/python/%s_Param.xlsx' % Library)
    ws = wb.active


    #### Save logS in temp-file TS Sol
    os.system("cxcalc logs -i true -H 7.0 %s > temp.txt" %SMILES)
    with open('C:/python/temp.txt', 'r') as myfile:
        TS_Sol = myfile.readlines()

    os.system("cxcalc logs -U mol/l -i true -H 7.0 %s > temp_2.txt" %SMILES)
    with open('C:/python/temp_2.txt', 'r') as myfile:
        TS Sol mol L = myfile.readlines()

    logS = []
    logS2 = []
    logS = TS Sol[1]
    #print("Länge Sol:", len(logS))
    #print(logS)
    ### gep logS pH 7.0 value from temp.txt
    counter = 0
```

```
counter2 = 0
for i in range(2, len(logS) - 1):
    if (logS[i] == '\t'):
        counter = counter + 1
    if (counter == 2):
        counter2 = i
        #print(counter2)
        counter = counter + 1
logS2 = logS[counter2 +1: len(logS)-2]
#print(logS2)

### gep logS_0 value from temp.txt
logS_0 = []
counter = 0
counter3 = 0
for i in range(0, len(logS) - 7):
    if (logS[i] == '\t'):
        counter = counter + 1
    if (counter == 2):
        counter3 = i
        counter = counter + 1
    logS_0 = logS[2:counter3]
#print(logS_0)

### gep Sol in mol/L at pH 7.0  from temp 2.txt
logS mol L = []
logS2 mol L = []
logS mol L = TS Sol mol L[1]
#print("Länge_Sol:", len(logS))
#print(logS)
counter = 0
counter2 = 0
for i in range(2, len(logS_mol_L) - 1):
    if (logS_mol_L[i] == '\t'):
        counter = counter + 1
    if (counter == 2):
        counter2 = i
        #print(counter2)
        counter = counter + 1
logS2 mol L = logS mol L[counter2 +1: len(logS mol L)-2]
#print(logS2)

logS mM = ((replaceComma(logS2 mol L))*1000) ### should be in mM

if logS_mM > 50:
    Solubilty = 'Very high'
if logS_mM < 50 and logS_mM > 10:
    Solubilty = 'High'
if logS_mM < 10 and logS_mM > 1:
    Solubilty = 'Moderate'
if logS_mM < 1 and logS_mM > 0:
    Solubilty = 'Low'
if logS_mM == 0:
    Solubilty = 'Insoluble'


print("logS: ", replaceComma(logS2))
print("logS_0: ", replaceComma(logS_0))
print("logS_mM: ", logS_mM)
print('Solubility:', Solubilty)
print('')

### Molecular weight
os.system("cxcalc mass %s > temp_3.txt" %SMILES)
with open('C:/python/temp_3.txt', 'r') as myfile:
    MW = myfile.readlines()
mass = MW[1]
print("MW:  ", replaceComma(mass[2:]))

### logP
os.system("cxcalc logP -H 7.0 %s > temp_4.txt" %SMILES)
with open('C:/python/temp_4.txt', 'r') as myfile:
    logP temp= myfile.readlines()
logP = logP_temp[1]
print("logP:  ", replaceComma(logP[2:]))
```

```
### charge at pH 7.0
os.system("cxcalc formalcharge -H 7.0 %s > temp_5.txt" % SMILES)
with open('C:/python/temp_5.txt', 'r') as myfile:
    charge temp = myfile.readlines()
charge = charge_temp[1]
print("Charge at 7.0:", int(charge[2:]))

### logD
os.system("cxcalc logD -H 7.0 %s > temp_6.txt" % SMILES)
with open('C:/python/temp_6.txt', 'r') as myfile:
    logD_temp = myfile.readlines()
logD = logD_temp[1]
print("logD: ", replaceComma(logD[2:]))

### H-bond acceptor and donor
os.system("cxcalc acceptorcount -H 7.0 donorcount -H 7.0 %s > temp_7.txt" % SMILES)
with open('C:/python/temp_7.txt', 'r') as myfile:
    count_temp = myfile.readlines()
count = count temp[1]
print("Acceptor: ", count[2:3])
print("Donor: ", count[4:])

### ring count
os.system("cxcalc ringcount ringsystemcount %s > temp_8.txt" % SMILES)
with open('C:/python/temp_8.txt', 'r') as myfile:
    ringcount temp = myfile.readlines()
    ringcount = ringcount temp[1]
print("Rings: ", ringcount[2:3])
print("Ringsystems: ", ringcount[4:])

### rotatable bonds
os.system("cxcalc rotatablebondcount %s > temp_9.txt" % SMILES)
with open('C:/python/temp_9.txt', 'r') as myfile:
    rotatable_temp = myfile.readlines()
    rotBonds = rotatable_temp[1]
print("Rotatable bonds: ", rotBonds[2:])

### 226print226n226 surface area = van der Waals surface area
os.system("cxcalc msa %s > temp_10.txt" % SMILES)
with open('C:/python/temp_10.txt', 'r') as myfile:
    msa_temp = myfile.readlines()
    msa = msa temp[1]
print("MSA: ", replaceComma(msa[2:]))

### van der Waals volume 3D
os.system("cxcalc volume %s > temp_11.txt" % SMILES)
with open('C:/python/temp_11.txt', 'r') as myfile:
    volume temp = myfile.readlines()
    volume = volume temp[1]
print("Volume: ", replaceComma(volume[2:]))


### polar surface area
os.system("cxcalc psa %s > temp_12.txt" % SMILES)
with open('C:/python/temp_12.txt', 'r') as myfile:
    psa_temp = myfile.readlines()
    psa = psa_temp[1]
print("PSA: ", replaceComma(psa[2:]))


### wateraccesiblesurfacearea
os.system("cxcalc wateraccessiblesurfacearea %s > temp_13.txt" % SMILES)
with open('C:/python/temp_13.txt', 'r') as myfile:
    wateraccesiblesurfacearea temp = myfile.readlines()
    wateraccesiblesurfacearea = wateraccesiblesurfacearea temp[1]

wasa = []
ASA = []
wasa = wateraccesiblesurfacearea_temp[1]
counter = 0
counter2 = 0
for i in range(0, len(wasa) - 1):
    if (wasa[i] == '\t'):
        counter = counter + 1
```

```
        if (counter == 1):
            ASA Start = i+1
            counter = counter + 1
        if (counter == 2):
            ASA End = i+1
            ASA_plus_start = ASA_End + 1
        if (counter == 3):
            ASA plus end = i + 1
            ASA minus start = ASA plus end + 1
        if (counter == 4):
            ASA minus end = i + 1
            ASA_Hydrophobic_start = ASA_minus_end + 1
        if (counter == 5):
            ASA Hydrophobic end = i + 1
            ASA polar start = i + 2

    ASA = wasa[ASA Start: ASA End]
    ASA_plus = wasa[ASA_plus_start: ASA_plus_end]
    ASA_minus = wasa[ASA_minus_start: ASA_minus_end]
    ASA Hydrophobic = wasa[ASA Hydrophobic start: ASA Hydrophobic end]
    ASA polar = wasa[ASA polar start:]

    print(replaceComma(ASA))
    print(replaceComma(ASA_plus))
    print((replaceComma(ASA_minus)))
    print((replaceComma(ASA_Hydrophobic)))
    print((replaceComma(ASA_polar)))

    # REF 1: Ferrara P,.Apostolakis J., Caflisch A., Proteins 2002, 46, 24 - 33;
    # REF 2: Hasel W., Hendrickson T., Still W., Tetrahedron Computer Methodology, 1988, Vol.
1, Issue 2, 103 - 116


    ### Refractivity
    os.system("cxcalc refractivity %s > temp_14.txt" % SMILES)
    with open('C:/python/temp_14.txt', 'r') as myfile:
        refractivity temp = myfile.readlines()
        refractivity = refractivity temp[1]
    print("Refractivity: ", replaceComma(refractivity[2:]))


    ### Name
    os.system("cxcalc name %s > temp_15.txt" % SMILES)
    with open('C:/python/temp_15.txt', 'r') as myfile:
        name temp = myfile.readlines()
        name = name_temp[1]
    print("Name: ",name[2:])


    ### write Praam value into excel
    #wb = load_workbook('C:/python/SMILES/logS_TS2017.xlsx')
    k = x + 1
    ws.cell(row=k, column=1).value = SMILES array[x-1][0]    # comp ID
    ws.cell(row=k, column=2).value = SMILES array[x - 1][2] # Structure in ChemAxon
    ws.cell(row=k, column=3).value = SMILES array[x - 1][1]  # ChemAxon SMILES
    #ws.cell(row=k, column=4).value = 227print227n227 SMILES needs to be optimized
    ws.cell(row=k, column=5).value = replaceComma(mass[2:])  # 227print227n227 weight
    ws.cell(row=k, column=6).value = replaceComma(logS2)    ### write logS value with point
    ws.cell(row=k, column=7).value = replaceComma(logS 0)   ### write logS 0 value
    ws.cell(row=k, column=8).value = logS mM   ### write Solubiliy value (mM)
    ws.cell(row=k, column=9).value = Solubilty ### Categorie
    ws.cell(row=k, column=10).value = replaceComma(logP[2:])  ### logP at pH 7.0
    ws.cell(row=k, column=11).value = int(charge[2:]) ### Formal Charge
    ws.cell(row=k, column=12).value = replaceComma(logD[2:])  ### logD at pH 7.0
    ws.cell(row=k, column=13).value = int(count[2:3]) ### Acc count
    ws.cell(row=k, column=14).value = int(count[4:])  ### Donor count
    ws.cell(row=k, column=15).value = int(ringcount[2:3])  ### ring count
    ws.cell(row=k, column=16).value = int(ringcount[4:])  ### ringsystems
    ws.cell(row=k, column=17).value = int(rotBonds[2:]) ### rotatable bonds
    ws.cell(row=k, column=18).value = replaceComma(msa[2:])  ### 227print227n227 surface area
    ws.cell(row=k, column=19).value = replaceComma(volume[2:])  ### van der Waals volume
    ws.cell(row=k, column=20).value = replaceComma(psa[2:])  ### polar surface area
    ws.cell(row=k, column=21).value = replaceComma(ASA)  ### wateraccesiblesurfacearea ASA
    ws.cell(row=k, column=22).value = replaceComma(ASA_plus)  ### wateraccesiblesurfacearea
ASA
```

```
    ws.cell(row=k, column=23).value = replaceComma(ASA_minus)  ### wateraccesiblesurfacearea
ASA
    ws.cell(row=k, column=24).value = replaceComma(ASA Hydrophobic)  ###
wateraccesiblesurfacearea ASA
    ws.cell(row=k, column=25).value = replaceComma(ASA polar)  ### wateraccesiblesurfacearea
ASA
    ws.cell(row=k, column=26).value = replaceComma(refractivity[2:])  ### Refractivity
    ws.cell(row=k, column=27).value = name[2:]  ### Name of compound

    ws = wb.save('C:/python/%s_Param.xlsx' % Library)
    os.system("del temp*.txt")


ws = wb.save('C:/python/%s_Param.xlsx' %Library)
```

Matlab: Dendrogram

```
figure(1)
[num,txt,raw] = xlsread('CR2006_Library.xlsx');
X = linkage((squareform(pdist(num))),'ward','euclidean');
dendrogram(X,0, 'Orientation', 'left', 'Labels',txt)
```

Matlab: Principal component analysis

```
# Tobias Schöne, 18.10.2017

%Z-scoring
normMatrix_Zscore = Matrix;
for n = 1:size(Matrix,2)
    normMatrix_Zscore(:,n) = (Matrix(:,n) - mean(Matrix(:,n)))/std(Matrix(:,n));
end

%% Do PCA for normMatrix_Zscore
%do PCA on normalized data
[coeff,score,latent, tsquared, explained, mu] = pca(normMatrix_Zscore);

%plot results of PCA
class1 = soluble_cat(☺ == '0';
class2 = soluble_cat(☺ == '1';
class3 = soluble_cat(☺ == '2';
class4 = soluble_cat(☺ == '3';

blue00 = [198,220,226]/255;
blue0 = [0,145,168]/255;
blue = [0 97 122]/255;
orange = [236,116,4]/255;

figure(1)
plot(score(class2,1),score(class2,2),'Marker','o','MarkerFaceColor',blue00,'MarkerEdgeCo
lor', 'k', 'LineStyle','none' );
hold on
plot(score(class3,1),score(class3,2),'Marker','o','MarkerFaceColor',blue0,'MarkerEdgeCol
or', 'k', 'LineStyle','none' );
plot(score(class4,1),score(class4,2),'Marker','o','MarkerFaceColor',blue,'MarkerEdgeColo
r', 'k', 'LineStyle','none' );
plot(score(class1,1),score(class1,2),'Marker','o','MarkerFaceColor',orange,'MarkerEdgeCo
lor', 'k', 'LineStyle','none' );
xlabel('PC1')
ylabel('PC2')

%3D scatter plot
s=20; % size of marker
figure(2)
scatter3(score(class2,1),score(class2,2),score(class2,3),s,'Marker','o','MarkerFaceColor
',blue00,'MarkerEdgeColor', 'k' );
hold on
```

```
scatter3(score(class3,1),score(class3,2),score(class3,3),s,'Marker','o','MarkerFaceColor
',blue0,'MarkerEdgeColor', 'k' );
scatter3(score(class4,1),score(class4,2),score(class4,3),s,'Marker','o','MarkerFaceColor
',blue,'MarkerEdgeColor', 'k' );
scatter3(score(class1,1),score(class1,2),score(class1,3),s,'Marker','o','MarkerFaceColor
',orange,'MarkerEdgeColor', 'k' );
axis = 'equal';
PC1 = 229print("PC 1 (%.1f%%)",explained(1));
PC2 = 229print("PC 2 (%.1f%%)",explained(2));
PC3 = 229print("PC 3 (%.1f%%)",explained(3));
xlabel(PC1)
ylabel(PC2)
zlabel(PC3)

%biplot → show contribution of each parameter
figure(3)
biplot(coeff(:,1:3),'scores',score(:,1:3),'varlabels', Parameter);

%explained data
figure(4)
components = linspace(1,size(Matrix,2),size(Matrix,2));
bar(components,explained,'FaceColor',blue);
xlabel('Principal components')
ylabel('Percentage of variance explained by each PC (%)')
```

grepPeaksfrom_XML_to_EXCEL.py

```
# Tobias Schöne, 11.08.2017
# script copies chemical shifts from XML file (created by TopSpin using pp command)into excel
sheet
# need xcel file named EMPTY.xlsx to create new xcel file
# if Inensities of peaks are required set Intensity as True

import sys, os
import string
import xml.etree.ElementTree as ET
import ast  # change string to int or float
from openpyxl import Workbook
from openpyxl import load_workbook
import datetime

### read xml files with peaks ###

Initials = 'TS' ### write your intials
date = datetime.date.today().strftime("%Y%m%d")

target_sheet = 'PeakList_%s%s.xlsx' %(Initials, date)
os.system("cd C:/python/")
os.system("copy EMPTY.xlsx %s" %target_sheet)

Intensity = False

fragments = 243
counter = 1
counter2 = 2

wb = load_workbook('./%s' % target_sheet)
ws = wb.active

for i in range(1, fragments):
    wb = load_workbook('./%s' %target_sheet)
    ws = wb.active
    ws.cell(row=counter, column=1).value = i  ### write compound number into first column
    if os.path.isfile("./NMR_2017_242Fragments/%s/pdata/1/peaklist.xml" % i):
        file_peaks = open("./NMR_2017_242Fragments/%s/pdata/1/peaklist.xml" % i, 'r')

        tree = ET.parse(file_peaks)  ### need trees and roots to read in xml ###
        root = tree.getroot()

        intensity = []
```

```
        shifts = []

        low = 0.00
        high = 50.00

        for Peak1D in root[0].findall('Peak1D'):

            ### grep peak intensity
            intens = Peak1D.get('intensity')
            intens float = float(intens)
            intensity.append(intens float)

            ### grep chemical shifts
            shift = Peak1D.get('F1')
            shift float = float(shift)
            if ((intens_float >= low and intens_float < high and shift_float > 0.2) and (
                        shift_float > 5.3000 or shift_float < 4.3000)):
                shifts.append(shift_float)
            #print(shift_float)

        for j in range(0, len(shifts)):
            n = j + 2
            ws.cell(row=counter, column=n).value = shifts[j]  ### write shifts with 3 decimal
place
            if Intensity:
                ws.cell(row=counter2, column=n).value = intensity[j]  ### write intensity with
3 decimal place
        ws = wb.save(target sheet) #### save in between
    else:
        ws = wb.save(target_sheet)  #### save in between

    if Intensity:
        counter = counter + 2
        counter2 = counter2 + 2
    else:
        counter = counter + 1

ws = wb.save(target sheet)
ws = wb.close()
```

makeTanimotoMatrix.py

```
### Tobias Schöne, 11.08.2017
### store SMILES from xcel sheet >library< into .smi file
###  uses Obabel to create pairwise TANIMOTO coefficienten
### Execute in C:/python with makeTanimotoMatrix_TS.py

import sys
import os
import os.path
import string
from openpyxl import Workbook
from openpyxl import load_workbook

230print230 = 'M_CR_2006_TS2017_Library_only_soluble_fragments' # or M_CR_2006
Initials = 'CRTS'
target_sheet = '%s_TAN.xlsx' %(Initials)
os.system("cd C:/python/")
os.system("copy EMPTY.xlsx %s" %target_sheet)

#### create SMILES file from xcel file ########################################
## in windows back slash is still required
## SMILES files can be used later
os.system("mkdir C:\python\%s\SMILES" %(230print230))
wb = load_workbook('%s.xlsx' %(230print230))
ws = wb.active
fragment array = []

##### count number of fragments in xcel sheet and create fragment array
col = 11 # SMILES column
counter = 2
isOK = True
```

```
while isOK:
    if not (ws.cell(row=counter+1, column=col).value is None):
        fragment array.append(ws.cell(row=counter+1, column=1).value)
        counter = counter +1
    else:
        isOK = False
fragments = counter - 1
print("Number of Fragments: ", fragments-1)
#fragment array.remove(fragment array[len(fragment array)-1])
print(fragment_array)
print(len(fragment_array))

for i in range(0,len(fragment_array)):
    output = open('C:/python/%s.smi' % (fragment_array[i]), 'w')
    output.write(ws.cell(row=i + 3, column=col).value)

output.close()
wb.close()

#### Preapere excel matrix ##########################################################
wb TAN = load workbook(target sheet)
ws TAN = wb TAN.active
### fill first row with compound-ID
n = 1
for i in range(0,len(fragment_array)):
        k = n + 1 ###start with second row
        ws TAN.cell(row=1, column=k).value = (fragment array[i])
        n = n+ 1
ws_TAN = wb_TAN.save(target_sheet)
wb_TAN = load_workbook(target_sheet)
ws TAN = wb TAN.active

n = 1
k=0
### fill first column with compound-ID
for i in range(0,len(fragment_array)):
        k = n + 1
        ws TAN.cell(row=k, column=1).value = (fragment array[i])
        n = n + 1
ws TAN = wb TAN.save(target sheet)

#### Create Tanimoto matrix in excel ################################################
wb TAN = load workbook(target sheet)
ws TAN = wb TAN.active
counter = 0
for j in range(0,len(fragment_array)):
        wb_TAN = load_workbook(target_sheet)
        ws TAN = wb TAN.active
        counter2 = counter
        for i in range(counter2, len(fragment_array)):
            os.system("obabel %s.smi %s.smi -ofpt >temp.txt" % (fragment_array[counter],
fragment_array[counter2]))
        ### creates Tanimoto coefficient using openbabel
        ### stored in temp.txt
            with open('C:/python/temp.txt', 'r') as myfile:
                TS SMILES = myfile.readlines()
            extract_Tan = TS_SMILES[1]
            Tan = extract_Tan[30:38]
            print('counter', fragment_array[counter])
            print('counter2', fragment_array[counter2])
            Tan float = float(Tan)  ### brauche float/Zahl in Excel sonst wird es als Text
gespeichert
            print('TAN', Tan_float)
            x = counter2 + 2
            y = counter + 2
            ws TAN.cell(row = y, column = x).value = Tan float
            ws TAN.cell(row = x, column = y).value = Tan float
            os.system("del temp.txt") ###delete TAN-file
            counter2 = counter2 + 1
        ws_TAN = wb_TAN.save(target_sheet)
        counter = counter + 1
ws TAN = wb TAN.save(target sheet)
os.system("del TS*.smi")
```

poolCompounds_2014.py

```
# execute in python 2.6
# usage in linux shell: python poolcompounds.py -f <name.sdf>
# Author: Christoph Rademacher, 2008
# modified by Tobias Schöne, 2014

import sys, os
import string
from copy import copy
from scipy import rand, linspace

counter = 0
for element in sys.argv:
    counter = counter + 1
    if element == "-f":
        in filename = sys.argv[counter]

##  IMPORTANT PARAMETER  #####
bucket size = 18
min_Tscore = 0.55
peak_overlapp_tolerance = 0.08  # given in ppm

##  METHODS          #########

def returnMolNum(datafile):
    infile_input = open(datafile)
    return_array = []
    putName = 1
    for line in infile_input:
        if putName == 1:
            return_array.append(string.atoi(line.replace("\n", "")))
            putName = 0
        if line.startswith("$$$$"):
            putName = 1
    return return_array

def removeDoubleEntry(array):
    return array = []
    for element in array:
        if not (element in return_array):
            return array.append(element)
    return return_array

def saveCompoundsToFile(name, array):
    file = open("./Arrays/%s.txt" % name, "w")
    for compound in array:
        file.write("%d\n" % compound)
    file.close()

def readCompoundsFromFile(data_file):
    input = open(data_file, "r")
    return array = []
    for line in input:
        return_array.append(string.atoi(line.replace("\n", "")))
    return return_array

def returnPeaks(comp_num):
    file_peaks = open("peakList_20140801.txt", 'r')
    return_array = []
    for line in file_peaks:
        if string.atoi(line.split("\t")[0]) == comp_num:
            for element in (line.split("\t")[1:len(line.split("\t")) - 1]):
                return array.append(string.atof(element))
    return return_array

def noPeakOverlapp(comp_num, bucket_array):
    peaks in bucket = []
    for compound in bucket_array:
        for peak in returnPeaks(compound):
            peaks in bucket.append(peak)
    for peak in returnPeaks(comp_num):
        peak_is_ok = True
```

```
        for peakBucket in peaks_in_bucket:
            if (abs(peak - peakBucket) < peak_overlapp_tolerance):
                peak is ok = False
        if peak_is_ok:
            return peak_is_ok
    return False

def returnTanimotoScore(comp_num_1, comp_num_2):
    tanimotoMatrix = open("./definitions/tanimotoMatrix.txt", "r")
    row = 1
    for line in tanimotoMatrix:
        if row == comp_num_1:
            return string.atof(line.split("\t")[comp_num_2 - 1])
        row += 1

# dieser compound mindestens einen Tscore > 0.55 hat
def Tscore(comp_num, bucket_array):
    for compound in range(0, len(bucket_array)):
        if (returnTanimotoScore(comp_num, bucket_array[compound]) >= min_Tscore):
            return False
    return True

# Methode bekommt drei arrays. Erstes aus dem mit hoechster prioritaet genommen werden soll,
dann zwei als ersatz zum auffuellen
# Kriterien sind:
# 1. Nicht reaktiv
# 2. Keine Signalüberlappung für mindestens ein Signal (~Abstand 0.08 ppm)
# 3. Tscore > 0.55

def makeBucketsFrom(main_array, sub1_array, sub2_array, outfile):
    temp bucket = []
    temp main array = (main array)
    temp sub1 array = (sub1 array)
    temp sub2 array = (sub2 array)

    output = open(outfile, "w")
    output.write("Num\tTSum\t")  ##### Eventuell noch aendern <!>
    for i in range(1, bucket_size + 1):
        output.write("%d\t" % i)
    output.write("\n")
    bucket_counter = 1
    while len(temp_main_array) > 0:
        # fuellen der ersten position
        random position = int(rand(1)[0] * len(temp main array))
        temp bucket.append(temp main array[random position])
        temp_main_array.remove(temp_main_array[random_position])

        # fuellen der naechsten x positionen
        for position in range(1, bucket_size):
            if len(temp_main_array) > 0:
                accepted = False
                for compound in temp_main_array:
                    if not accepted:
                        if (noPeakOverlapp(compound, temp_bucket) and Tscore(compound,
temp bucket)):
                            temp bucket.append(compound)
                            temp_main_array.remove(compound)
                            accepted = True

        # fuelle nun alle offenen plaetze im array auf mit sub1 array
        for position in range(len(temp_bucket), bucket_size):
            if len(temp_sub1_array) > 0:
                accepted = False
                for compound in temp_sub1_array:
                    if not accepted:
                        if (noPeakOverlapp(compound, temp_bucket) and Tscore(compound,
temp bucket)):
                            temp bucket.append(compound)
                            temp_sub1_array.remove(compound)
                            accepted = True

        # fuelle nun alle offenen plaetze im array auf mit sub2 array
        for position in range(len(temp_bucket), bucket_size):
            if len(temp_sub2_array) > 0:
                accepted = False
```

```
                for compound in temp_sub2_array:
                    if not accepted:
                        if (noPeakOverlapp(compound, temp_bucket) and Tscore(compound,
temp bucket)):
                            temp bucket.append(compound)
                            temp_sub2_array.remove(compound)
                            accepted = True

        # wenn bucket aufgefuellt ist dann schreibe diesen in file und setze wieder auf leeren
bucket
        # gebe in der ausgabe datei die summe der paarweisen 234print234n koeff aus
        sum_Tscore = 0
        for counter in range(0, len(temp_bucket)):
            for counter_2 in range(0, len(temp_bucket)):
                if counter_2 > counter:
                    sum Tscore += returnTanimotoScore(temp bucket[counter],
temp bucket[counter 2])
        sum_Tscore = sum_Tscore / len(temp_bucket)  # gibt den durschnittlichen score aus
        # print outfile,"RotBonds:", rotatable_bonds_in_this_bucket, "length main_array:",
len(main array), "TanimotoScoreSum:", sum Tscore
        sum Tscore = 0
        output.write("%d\t%.4f\t" % (bucket_counter, sum_Tscore))
        bucket_counter += 1
        temp_bucket.sort()
        for element in temp_bucket:
            output.write("%d\t" % element)
        output.write("\n")
        temp bucket = []
    output.close()

############################
# 1. Determine Reactivity  ##
############################

### Acids = open("acids.txt","w")

do calc = True  # setze auf False, wenn die Acids/Bases schon berrechnet wurden

if do_calc:

    #### Acids ####
    acid_array = []
    acid_def = open("./definitions/acids.txt", "r")
    for acid in acid_def:
        if acid.startswith("EXC"):
            os.system(
                "obabel %s -o sdf -O tempacid.sdf -s '%s'" % (in_filename, ac -
id.split("\t")[1].replace("\n", "")))
            for element in returnMolNum("tempacid.sdf"):
                if element in acid_array:
                    acid_array.remove(element)
        else:
            os.system("obabel %s -o sdf -O tempacid.sdf -s '%s'" % (in_filename,
acid.replace("\n", "")))
            acid_array += returnMolNum("tempacid.sdf")
        os.system("rm tempacid.sdf")
    acid_array = removeDoubleEntry(acid_array)

    #### bases ####
    bases array = []
    bases_def = open("./definitions/bases.txt", "r")
    for bases in bases_def:
        if bases.startswith("EXC"):
            os.system(
                "obabel %s -o sdf -O tempbases.sdf -s '%s'" % (in_filename, ba -
ses.split("\t")[1].replace("\n", ""))
            for element in returnMolNum("tempbases.sdf"):
                if element in bases_array:
                    bases_array.remove(element)
        else:
            os.system("obabel %s -o sdf -O tempbases.sdf -s '%s'" % (in_filename,
bases.replace("\n", "")))
            bases_array += returnMolNum("tempbases.sdf")
        os.system("rm tempbases.sdf")
    bases_array = removeDoubleEntry(bases_array)
```

```
    #### electrophile ####
    electrophile array = []
    electrophile_def = open("./definitions/electrophiles.txt", "r")
    for electrophile in electrophile_def:
        if electrophile.startswith("EXC"):
            os.system("obabel %s -o sdf -O tempElectrophile.sdf -s '%s'" % (
            in_filename, electro - phile.split("\t")[1].replace("\n", "")))
            for element in returnMolNum("tempElectrophile.sdf"):
                if element in electrophile_array:
                    electrophile array.remove(element)
        else:
            os.system(
                "obabel %s -o sdf -O tempElectrophile.sdf -s '%s'" % (in_filename, electro -
phile.replace("\n", "")))
            electrophile_array += returnMolNum("tempElectrophile.sdf")
        os.system("rm tempElectrophile.sdf")
    electrophile array = removeDoubleEntry(electrophile array)

    #### nucleophile ####
    nucleophile array = []
    nucleophile_def = open("./definitions/nucleophiles.txt", "r")
    for nucleophile in nucleophile_def:
        if nucleophile.startswith("EXC"):
            os.system("obabel %s -o sdf -O tempNucleophile.sdf -s '%s'" % (
            in_filename, nucleo - phile.split("\t")[1].replace("\n", "")))
            # os.system("sleep 2")
            for element in returnMolNum("tempNucleophile.sdf"):
                if element in nucleophile_array:
                    nucleophile_array.remove(element)
        else:
            os.system(
                "obabel %s -o sdf -O tempNucleophile.sdf -s '%s'" % (in_filename, nucleo -
phile.replace("\n", "")))
            nucleophile_array += returnMolNum("tempNucleophile.sdf")
        os.system("rm tempNucleophile.sdf")
    nucleophile_array = removeDoubleEntry(nucleophile_array)

else:
    electrophile_array = readCompoundsFromFile("./Arrays/e.txt")
    nucleophile array = readCompoundsFromFile("./Arrays/n.txt")
    acid_array = readCompoundsFromFile("./Arrays/a.txt")
    bases_array = readCompoundsFromFile("./Arrays/b.txt")

#### Entferne alle Substanzen, die nicht zu Verfuegung stehen #######
notAvailable_data = open("notAvailable.txt", "r")
notAvailable_array = []
for line in notAvailable_data:
    notAvailable_array.append(string.atoi(line[0:line.find("\n")]))
notAvailable array 2 = copy(notAvailable array)
while len(notAvailable_array) > 0:
    counter = 0
    if notAvailable_array[0] in acid_array:
        acid array.remove(notAvailable array[counter])
    if notAvailable_array[0] in bases_array:
        bases array.remove(notAvailable array[counter])
    if notAvailable_array[0] in electrophile_array:
        electrophile_array.remove(notAvailable_array[counter])
    if notAvailable_array[0] in nucleophile_array:
        nucleophile array.remove(notAvailable array[counter])
    notAvailable array.remove(notAvailable array[counter])
#### Sortiere Arrays ########
acid array.sort()
bases_array.sort()
electrophile_array.sort()
nucleophile array.sort()
#### output ####
print "acid: %d" % len(acid_array)
print acid_array
print (" ")
print "bases: %d" % len(bases_array)
print bases_array
print (" ")
print "Electrophile: %d" % len(electrophile_array)
print electrophile_array
print (" ")
```

```
print "Nucleophile: %d" % len(nucleophile_array)
print nucleophile_array
print (" ")
a = len(acid array) + len(bases array) + len(electrophile array) + len(nucleophile array)
print a
saveCompoundsToFile("e", electrophile_array)
saveCompoundsToFile("n", nucleophile_array)
saveCompoundsToFile("a", acid_array)
saveCompoundsToFile("b", bases_array)

##################################################
##   check for intersections  ####################
##################################################
# e n array
e n array = []
for index in range(0, len(electrophile_array)):
    if electrophile_array[index] in nucleophile_array:
        e_n_array.append(electrophile_array[index])
for element in e_n_array:
    nucleophile array.remove(element)
    electrophile array.remove(element)
# a b array
a_b_array = []
for A in acid_array:
    if A in bases_array:
        a b array.append(A)
for A in a_b_array:
    acid array.remove(A)
    bases_array.remove(A)

compound array = []
for counter in range(1, 501):
    if counter not in notAvailable_array_2:
        compound array.append(counter)
# inert_array
inert_array = copy(compound_array)
for compound in compound_array:
    if compound in electrophile_array:
        inert array.remove(compound)
    else:
        if compound in nucleophile_array:
            inert_array.remove(compound)
        else:
            if compound in e_n_array:
                inert array.remove(compound)
# neutral_array
neutral_array = copy(compound_array)
for compound in compound_array:
    if compound in acid_array:
        neutral array.remove(compound)
    if compound in bases_array:
        neutral_array.remove(compound)
    if compound in a_b_array:
        neutral array.remove(compound)

# erstelle alle AE,AN,AEN,AI arrays
ae_array = []
an_array = []
aen array = []
ai array = []

for A in acid_array:
    if A in electrophile_array:
        ae_array.append(A)
    else:
        if A in nucleophile_array:
            an array.append(A)
        else:
            if A in e_n_array:
                aen_array.append(A)
            else:
                ai array.append(A)

# erstelle alle BE,BN,BEN,BI arrays
be_array = []
```

```
bn_array = []
ben array = []
bi array = []

for B in bases_array:
    if B in electrophile_array:
        be_array.append(B)
    else:
        if B in nucleophile_array:
            bn array.append(B)
        else:
            if B in e_n_array:
                ben_array.append(B)
            else:
                bi array.append(B)

# erstelle alle BE,BN,BEN,BI arrays
abe_array = []
abn_array = []
aben array = []
abi array = []

for AB in a_b_array:
    if AB in electrophile_array:
        abe array.append(AB)
    else:
        if AB in nucleophile_array:
            abn array.append(AB)
        else:
            if AB in e_n_array:
                aben array.append(AB)
            else:
                abi array.append(AB)

# erstelle alle neuE,neuN,neuEN,neuI arrays
neue_array = []
neun array = []
neuen array = []
neui array = []

for N in neutral_array:
    if N in electrophile_array:
        neue array.append(N)
    else:
        if N in nucleophile_array:
            neun_array.append(N)
        else:
            if N in e_n_array:
                neuen array.append(N)
            else:
                neui_array.append(N)

# speichern der array zur schnelleren berechnung
saveCompoundsToFile("ae", ae_array)
saveCompoundsToFile("an", an_array)
saveCompoundsToFile("aen", aen_array)
saveCompoundsToFile("ai", ai_array)
saveCompoundsToFile("be", be_array)
saveCompoundsToFile("bn", bn_array)
saveCompoundsToFile("ben", ben_array)
saveCompoundsToFile("bi", bi_array)
saveCompoundsToFile("abe", abe_array)
saveCompoundsToFile("abn", abn_array)
saveCompoundsToFile("aben", aben_array)
saveCompoundsToFile("abi", aben_array)
saveCompoundsToFile("inert", inert_array)
saveCompoundsToFile("neutral", neutral_array)
saveCompoundsToFile("neuE", neue_array)
saveCompoundsToFile("neuN", neun_array)
saveCompoundsToFile("neuEN", neuen_array)
saveCompoundsToFile("neuI", neui_array)

print "\nSTART################\n"
print_daten = True
if print_daten:
```

```
    # ausgabe der array_daten
    print "E: %d" % len(electrophile_array)
    print "N: %d" % len(nucleophile_array)
    print "EN: %d" % len(e_n_array)
    print "I: %d" % len(inert_array)
    print "A: %d" % len(acid_array)
    print "B: %d" % len(bases_array)
    print "AB: %d" % len(a_b_array)
    print "N: %d" % len(neutral_array)
    print "AE: %d" % len(ae_array), ae_array
    print "AN: %d" % len(an_array), an_array
    print "AEN: %d" % len(aen_array), aen_array
    print "AI: %d" % len(ai_array), ai_array
    print "BE: %d" % len(be_array), be_array
    print "BN: %d" % len(bn_array), bn_array
    print "BEN: %d" % len(ben_array), ben_array
    print "BI: %d" % len(bi_array), bi_array
    print "ABE: %d" % len(abe_array), abe_array
    print "ABN: %d" % len(abn_array), abn_array
    print "ABEN: %d" % len(aben_array), aben_array
    print "ABI: %d" % len(abi_array), abi_array
    print "neuE: %d" % len(neue_array), neue_array
    print "neuN: %d" % len(neun_array), neun_array
    print "neuEN: %d" % len(neuen_array), neuen_array
    print "neuI: %d" % len(neui_array), neui_array

################################################
###### Berechnung von Buckets   ################
################################################

makeBucketsFrom(ae_array, neue_array, ai_array, "./buckets/AE_buckets.txt")
makeBucketsFrom(be_array, neue_array, bi_array, "./buckets/BE_buckets.txt")
makeBucketsFrom(abe_array, neue_array, abi_array, "./buckets/ABE_buckets.txt")
makeBucketsFrom(an_array, neun_array, ai_array, "./buckets/AN_buckets.txt")
makeBucketsFrom(bn_array, neun_array, bi_array, "./buckets/BN_buckets.txt")
makeBucketsFrom(abn_array, neun_array, abi_array, "./buckets/ABN_buckets.txt")
makeBucketsFrom(aen_array, neuen_array, ai_array, "./buckets/AEN_buckets.txt")
makeBucketsFrom(ben_array, neuen_array, bi_array, "./buckets/BEN_buckets.txt")
makeBucketsFrom(aben_array, neuen_array, abi_array, "./buckets/ABEN_buckets.txt")
makeBucketsFrom(ai_array, neui_array, neui_array, "./buckets/AI_buckets.txt")
makeBucketsFrom(bi_array, neui_array, neui_array, "./buckets/BI_buckets.txt")
makeBucketsFrom(abi_array, neui_array, neui_array, "./buckets/ABI_buckets.txt")
makeBucketsFrom(neue_array, neui_array, neui_array, "./buckets/neuE_buckets.txt")
makeBucketsFrom(neun_array, abi_array, neui_array, "./buckets/neuN_buckets.txt")
print "\nEND###################\n"
# ausgabe der array_daten
print "E: %d" % len(electrophile_array)
print "N: %d" % len(nucleophile_array)
print "EN: %d" % len(e_n_array)
print "I: %d" % len(inert_array)

print "A: %d" % len(acid_array)
print "B: %d" % len(bases_array)
print "AB: %d" % len(a_b_array)
print "Neutral: %d" % len(neutral_array)

print "AE: %d" % len(ae_array), ae_array
print "AN: %d" % len(an_array), an_array
print "AEN: %d" % len(aen_array), aen_array
print "AI: %d" % len(ai_array), ai_array

print "BE: %d" % len(be_array), be_array
print "BN: %d" % len(bn_array), bn_array
print "BEN: %d" % len(ben_array), ben_array
print "BI: %d" % len(bi_array), bi_array

print "ABE: %d" % len(abe_array), abe_array
print "ABN: %d" % len(abn_array), abn_array
print "ABEN: %d" % len(aben_array), aben_array
print "ABI: %d" % len(abi_array), abi_array

print "neuE: %d" % len(neue_array), neue_array
print "neuN: %d" % len(neun_array), neun_array
print "neuEN: %d" % len(neuen_array), neuen_array
print "neuI: %d" % len(neui_array), neui_array
```

SMILES Definition according to Hann et al.

Acids:

[OD1][P,C,S](=O)
[NH1]([P,S]=O)([P,S]=O)
[nH]1cnoc1=O
[OD1]C1=NC=NO1
[ND1]1C=NOS1=O
[OD1]C1=NC(=O)CC1=O
[OD1]C1NC(=O)C(=O)C1
[nH1]1ncoc1=O
[nH1]1occc1=O
[OD1]c1oncc1=O
[OD1]C1=NN=CO1
[nH1]1[nH]cnc1=O
[OD1]C1=N[ND1]C=N1
[OD1]C1=NOC=C1
[nH1]1ccc(=O)o1
[nH1]nnn
[nH1](n)nn
[OD1]C1=NC(=O)NO1
[OD1]C1=NC(=O)ON1
[nH1]1cnnc1C(F)(F)F
[nH1]1cnc(n1)C(F)(F)F
[nH1]1C(=O)CC(=O)O1
[OD1]C1=CC(=O)NO1
[OD1]C1=CC(=O)ON1
[ND1]C1(=O)c2ccccc2S1(=O)=O
[OD1]C1=NS(=O)(=O)c2ccccc21
[OD1]C1=NC(=O)c2ccccc21
[OD1]C1=COC=CC1=O
[OD1]C1=NSN=C1
[OD1]NC(=O)
[NH]S(=O)(=O)C(F)(F)F
[NH]IS(=O)=O
[OD1]c1c[c,n]cc1

Bases:

[ND1][CX4]
[ND2]([CX4])[CX4]
[ND3]([CX4])([CX4])[CX4]
[C,c](=N)N
EXC [C,c](=N)N[C,S](=O)
EXC n(:c)(:c):a
EXC [N,n;+1]

Electrophiles:

[C;H1](=[O,S])[C,c]
[C,P;H1](=[O,S])[O,S]
C(=[O,S])[OD2,S]
[C](=O)([C,c,O,S])[C,c,O,S]
C(=[O,S])(N)Oc
EXC C(=O)[OD1]
EXC C(=O)[SD1]
C1(=O)NS(=O)(=O)[C,c]=,:[C,c]1
c(=O)(~c)~c
C(=O)-N-C=O

Nucleophiles:

[ND1][CX4]
[ND2]([CX4])[CX4]
[OD1][C,c,N]
C(#N)
EXC [OD1][P,C,S](=O)
EXC N=[O,C,N,S]
EXC [ND1]-[C,c,N]=[C,c,N,n,O,S]
EXC [OD1]C=C
EXC [OD1]NC=[O,S]

makeBinsforScreening_2017.py

```
### Tobias Schöne, 13.08.2017
### cmd in windows "python makeBinsForScreening.py"
### file C:/python/

### a new method allSignlsIsolated in method
    noPeakOverlapp guaratnees that all signals of fragment can be analyzed
### must be set manually

import os.path
from openpyxl import load_workbook

####    inputs:

### please define
Tanimoto = 0.6
shift tolerance = 0.025  # ppm
Bin Size = 30

Maybridge = 'M_TS_2017'  # or M_TS_2017
Initials = 'TS'  # CR or TS
### fragment sheet contains all fragment ID to be 240print240n + functional groups (e.g. NH2)
in column 6
Fragment_sheet = '%s_Library.xlsx' %Maybridge
num fragments = 242
Drug_Target = 'Saga'
Fragment_col = 1 ### see in xcel sheet
functionalGroup = 6   ### see column in excel sheet
Solublity = 11  ### see column in excel sheet

### load Excel
os.system("cd C:/python/")

# Open Xcel sheets, to save 240print240 time ######
# Xcel sheets need to be in file C:/python/ or same file as python script
# TAN-Matrix
wb_TAN = load_workbook('C:/python/%s/%s_TAN_%s_Fragments.xlsx' % (Maybridge, Initials,
num fragments))
ws TAN = wb TAN.active
#### 1H shifts
wb1 = load_workbook('./PeakList_TS2017.xlsx')
ws1 = wb1.active  ### sheet has to be active
#### 1H shifts
wb2 = load_workbook('./PeakList_TS2017_DMSO.xlsx')
ws2 = wb2.active  ### sheet has to be active
#### 240print240n240240 group
wb_group = load_workbook('C:/python/%s/%s' % (Maybridge, Fragment_sheet))
ws_group = wb_group.active
#### Fragments to screen
wb_STD = load_workbook('C:/python/%s/%s' % (Maybridge, Fragment_sheet))
ws STD = wb STD.active

#### Methods

#### grep TAN from excel sheet for two fragments
def grepTANfromXcel(comp1, comp2):
    x = int(comp1[2:]) + 5
    y = int(comp2[2:]) + 5
    ###search for first comp
    counter = 1
    counter2 = 1
    for i in range(2, x):
        if (ws_TAN.cell(row=1, column=i).value != comp1):
            counter = counter + 1
        if (ws_TAN.cell(row=1, column=i).value == comp1):
            b = counter + 1
    ###search for second comp
    for i in range(2, y):
        if (ws_TAN.cell(row=1, column=i).value != comp2):
            counter2 = counter2 + 1
        if (ws_TAN.cell(row=i, column=1).value == comp2):
            a = counter2 + 1
```

```
    return float((ws_TAN.cell(row=a, column=b).value))

def TanimotoTolerance(bucket_array, comp):
    for i in range(0, len(bucket_array)-1):
            if (grepTANfromXcel(bucket_array[i],comp) > Tanimoto):
                return False
    return True

## takes shifts from Xcel sheet for comp ##

def grepShiftsfromXcel(comp1):
    comp = int(comp1[2:])
    #os.system("cd C:/python/")
    #wb1 = load_workbook('./PeakList_TS2017.xlsx')
    #ws1 = wb1.active  ### sheet has to be active
    shifts = []
    for i in range(2, 25):
        a = ws1.cell(row=comp, column=i).value
        # print(a)
        if isinstance(a, float):  ### Prüft, ob Eingabe eine Zahl ist
            shifts.append(ws1.cell(row=comp, column=(i)).value)
        else:  ### Schaue in die DMSO PeakList nach
            #wb2 = load_workbook('./PeakList_TS2017_DMSO.xlsx')
            #ws2 = wb2.active  ### sheet has to be active
            for j in range(1, 242):
                if ws2.cell(row=j, column=1).value == comp:
                    for k in range(2, 25):
                        b = ws2.cell(row=j, column=k).value
                        # print('else', b)
                        if isinstance(b, float):  ### Prüft, ob Eingabe eine Zahl ist
                            shifts.append(ws2.cell(row=j, column=(k)).value)
            # print(shifts)
            return shifts
    #print(shifts)
    return shifts

##### gets functional group from excel sheet, must be prepared in advance ########
def grepGroup(comp1):
    ###search for comp
    b = 1
    counter = 1
    for i in range(0,len(STD_array)):
        if (ws_group.cell(row=i+2, column=1).value != comp1):
            counter = counter + 1
        if (ws_group.cell(row=i+2, column=1).value == comp1):
            b = counter + 1
            return (ws_group.cell(row=b, column=functionalGroup).value)
    return('grepGroup Error')


#### gets fragement-array of Bin XX and check if any peak 241print241
def oneSignlIsolated(bucket_array, comp_shift_array):
    for peak in comp_shift_array:
        peak is ok = True
        for peakBucket in bucket_array:
            if (abs(peak - peakBucket) < shift_tolerance):
                peak_is_ok = False

        if peak_is_ok:
            return peak_is_ok
    return False

#### gets fragement-array of Bin XX and check whether any peak 241print241
def allSignlsIsolated(bucket_shift_array, comp_shift_array):
    for peak in comp_shift_array:
        for peakBucket in bucket_shift_array:
            if (abs(peak - peakBucket) < shift_tolerance):
                return False
    return True

### set method oneSignlIsolated or allSignlsIsolated################
#### check overlap of signals in a global fashion
def noPeakOverlapp(list_matrix, bucket_array, comp):
    is_OK = True
    # produce shift_array of all fragments in bucket_array from Bin XX
```

```
    bucket_shift_temp_array2 = []
    for i in range(0,len(bucket_array)):
            bucket shift temp array = []

            num shifts = int((bucket array[i])[2:]) -1
            if list_matrix[num_shifts][1] > 0:    #### check whether 1H shifts are available
                for j in range(2,(list_matrix[num_shifts][1])+2):
                    bucket shift temp array.append(list matrix[num shifts][j])
                bucket shift temp array2 = bucket shift temp array2 + bucket shift temp array

    # Now create shift array of comp to analyze
    comp_shift_temp_array = []
    comp1 = int(comp[2:]) - 1
    if list_matrix[comp1][1] > 0:
        for i in range(2, (list_matrix[comp1][1]) + 2):
            comp shift temp array.append(list matrix[comp1][i])

    # check whether one isolated signal is available
    bucket_shift_temp_array3 = []

    ######  set method to allSignlsIsolated if desired
    if not (oneSignlIsolated(bucket_shift_temp_array2, comp_shift_temp_array)):
        is_OK = False
    else:
        bucket shift temp array3 = bucket shift temp array3 + bucket shift temp array2 +
comp shift temp array

    #### Now check whether the first X fragments in bucket array exhibit still at least one
isolated signal
    if is_OK:
        for i in range(0, len(bucket_array)):
            bucket shift temp array4 = []
            bucket shift temp array4 = bucket shift temp array4 + bucket shift temp array3 #
To save all fragments
            bucket_shift_temp_array_valid = []
            num_shifts2 =  int((bucket_array[i])[2:]) -1
            if list_matrix[num_shifts2][1] > 0:
                for j in range(2, (list_matrix[num_shifts2][1]) + 2):
                   bucket shift temp array valid.append(list matrix[num shifts2][j]) #
                   bucket shift temp array4.remove(list matrix[num shifts2][j])

                ######  set method to allSignlsIsolated if desire
                if not oneSignlIsolated(bucket_shift_temp_array4,
bucket shift temp array valid):
                    is OK = False
    return is_OK

def removeElementfromArray(array, comp):
    try:
        array.remove(comp)
        return array
    except:
        pass
    return array

def ElementOfArray(array, comp):
    for i in range(0,len(array)-1):
        if comp == array[i]:
            return True
    return False

####  Main program  ################################################
####    Create Fragment array    ###########################################
print('########################')
print('Create fragment array')
print('########################')
STD array = []
##### count number of hits in xcel sheet
counter = 0
isOK = True
while isOK:
    if not (ws_STD.cell(row=counter+1, column=Fragment_col).value is None):
        counter = counter +1
    else:
        isOK = False
```

```
number fragments = counter -1 # since first row is row name
#### depends one comp num!! Here it's TSXXX
for i in range(2, number_fragments + 2):
    if ((ws_STD.cell(row=i, column=Solublity).value) != 'No'):
        STD_array.append(ws_STD.cell(row=i, column=Fragment_col).value)
print("Fragments to screen for %s:" %Drug_Target)
print(STD_array)
print("")
print("Number of Fragments: ", len(STD_array))
print("")
print("")
print("")

####    Create base array and acid array
base array = []
acid array = []
inert_array_temp = []
inert_array = []
inert array temp = inert array temp + STD array
for i in range(0,(len(STD_array)-1)):
    if not (grepGroup(STD_array[i]) is None):
        if 'COO' in grepGroup(STD_array[i]):
            acid_array.append(STD_array[i])
            inert array temp.remove(STD array[i])

for i in range(0, (len(STD_array))):
    if not(grepGroup(STD_array[i]) is None):
        if 'NH' in grepGroup(STD_array[i]):
            base_array.append(STD_array[i])
            inert array temp.remove(STD array[i])

inert array = inert array + inert array temp
print('Base array: ')
print(base_array)
print('')
print('Acid array: ')
print(acid_array)
print('')
print(len(base array) + len(acid array) + len(inert array))

####    Create shift_matrix of all fragments with Hash-tag function
####    must be a #fragments:#fragments matrix (e.g. (242:242)
fragment array = []
for i in range(1,num_fragments+1):
    fragment_array.append(i)
#print(fragment_array)

list matrix = [[]]
for x in range(1,(num_fragments)+1):
    counter = 0
    list_matrix[x-1].append(x)  #### Write comb number in first place
    list matrix[x - 1].append(0)  ####  number of shifts is zero
    for y in range(2, 30): ### find up to 30 shifts
        a = ws1.cell(row=x,  column=y).value
        if isinstance(a, float):
            list_matrix[x-1].append(ws1.cell(row=x, column=y).value)
            counter = counter + 1
        else:  ### look in DMSO sheet
            for j in range(0,(num_fragments-1)):
                if ws2.cell(row=j+1, column=1).value == fragment_array[j]:
                    for k in range(2, 30):
                        b = ws2.cell(row=j+1, column=k).value
                        if isinstance(b, float):
                            counter = counter + 1
                            list matrix[x-1].append.ws2.cell(row=j+1, column=k).value
    list matrix[x-1][1] = counter  ### second position = number of shifts
    list matrix.append([])

print('Länge_Matrix:', len(list_matrix))

#########################  Create Bins
print('#########################')
print('Create bins for screening')
print('#########################')
```

```
os.system("cd C:/python/")
###### write excel sheet in file; copy EMPTY FILE
target_sheet = 'Bins_for_Screening_%s.xlsx' %Drug_Target
os.system("cd C:/python/")
os.system("copy EMPTY.xlsx %s" %target_sheet)
Max_Bins = round(number_fragments/Bin_Size)+6  #### 6 extra as buffer

for Bin_num in range(1,Max_Bins):
    comp array = []
    wb = load workbook(target sheet)
    ws = wb.active
    if (len(STD_array)>0):
        ws.cell(row=Bin_num, column=1).value = ('Bin%s' % Bin_num) #BinXX in first column
        print('#######  Bin:',Bin_num)
        ws.cell(row=Bin num, column=2).value = STD array[0]  # write first comp in bin
        comp array.append(STD array[0])
        ws = wb.save(target sheet)
        excel_counter2 = 3
        counter3 = 1
        counter4 = 1

        reactivity  = True
        group = 'X'  # any group except NH2 or COOH
        # if comp has got CCOH group don't accept comp with NH2
        if ElementOfArray(base_array, STD_array[0]):
            group = 'COO'
            reactivity = False
        # if comp has got NH2 group don't accept comp with COOH
        if ElementOfArray(acid_array, STD_array[0]):
            group = 'NH'
            reactivity = False

        while counter4 < len(STD_array):
            # check whether second element shows reacitivty
            if ElementOfArray(base_array,STD_array[counter4]) and reactivity:
                group = 'COO'
                reactivity = False
            if ElementOfArray(acid_array,STD_array[counter4]) and reactivity:
                group = 'NH'
                reactivity = False
            if ((counter3 < Bin_Size)
                    and (TanimotoTolerance(comp_array,STD_array[counter4]))
                    and (grepGroup(STD_array[counter4]) != group)
                    and (noPeakOverlapp(list_matrix, comp_array, STD_array[counter4]))):
                wb = load workbook(target sheet)
                ws = wb.active
                ws.cell(row=Bin_num, column=excel_counter2).value = STD_array[counter4]
                comp array.append(STD array[counter4])
                print('###############  Fragment: ', STD_array[counter4])
                excel counter2 = excel counter2 + 1
                ws = wb.save(target_sheet)
                counter3 = counter3 + 1
                removeElementfromArray(STD array, STD array[counter4])
                counter4 = counter4 -1 # because one element has been removed
            counter4 = counter4 + 1  # if no match check next compound
        removeElementfromArray(STD array, STD array[0])
        print('')
        print('Fragments to allocate:')
        print(STD_array)
        if(len(STD_array)>0):
            print('')
            print('############################')
            print('####   next Bin   ####')
            print('############################')
            print('')

print('################   Program accomplished     #######')
print('')
print('Tanimoto:', Tanimoto)
print('Signal tolerance:', shift_tolerance)
print('Manual allocation:', STD_array)
ws = wb.save(target sheet)
```

python2Maestro_plotCSP.py

```
# Tobias Schöne, 06.08.2017
# reads chemical shifts of assigned amino acid resiudes following ligand titration
# creates a Maestro script that has to be copied into the Maestro Command Script Editor


from openpyxl import load_workbook

#### open excel sheet
wb_CSP = load_workbook('C:/python/CSP_Saga_All_RC_Assignment.xlsx')
ws CSP = wb CSP.active

##### take aa and shifts from xcel sheet and write into array
isOK2 = True
col counter = 3
while isOK2:
    CSP AA = []
    CSP shift = []
    col_counter = col_counter + 1
    if (ws_CSP.cell(row=1, column=col_counter).value is None):
        isOK2 = False
    counter = 2
    isOK = True
    red = ws_CSP.cell(row=108, column=col_counter).value
    red2 = ws_CSP.cell(row=107, column=col_counter).value
    orange = ws CSP.cell(row=106, column=col counter).value
    while isOK:
        if not (ws_CSP.cell(row=counter, column=col_counter).value is None):
            CSP AA.append(ws CSP.cell(row=counter, column=1).value)
            CSP_shift.append(ws_CSP.cell(row=counter, column=col_counter).value)
            counter = counter + 1
        else:
            isOK = False
    if(counter == 4):
        print(CSP_AA)
    print(ws_CSP.cell(row=1, column=col_counter).value)
    print(CSP_shift)

    ### copy maestro 245print245n into txt.file
    output = open("MeastroCMD_%s_%.3f_%.3f_%.3f.txt" % (ws_CSP.cell(row=1,
column=col counter).value, red, red2, orange), "w")

    output.write("displayonlyatom backbone")
    output.write("\n")
    output.write("undisplayatom atom.ptype C")
    output.write("\n")
    output.write("undisplayatom atom.ptype O")
    output.write("\n")
    output.write("undisplayatom atom.ptype CO")
    output.write("\n")
    output.write("undisplayatom atom.ptype CA")
    output.write("\n")
    output.write("workspaceselectionreplace mol.num 1")
    output.write("\n")
    output.write("coloratomrgb red=128 green=128 blue=128 at.selected")
    output.write("\n")
    output.write("workspaceselectionclear")
    output.write("\n")
    output.write("workspaceselectionreplace mol.num 2")
    output.write("\n")
    output.write("coloratomrgb red=128 green=128 blue=128 at.selected")
    output.write("\n")
    output.write("workspaceselectionclear")
    output.write("\n")
    output.write("workspaceselectionreplace all")
    output.write("\n")
    output.write("undisplayatom ( withinbonds 1 ((at.selected) AND (( atom.disp ) ) ) ) AND ((
atom.ele H ) )")
    output.write("\n")

    ### copy maestro 245print245n into txt.file
    ### amino acids will be displayed as colored balls
    for i in range(0, len(CSP_AA)):
        output.write("find res.num %s" % CSP_AA[i])
        output.write("\n")
        if (CSP_shift[i] >= red):
```

```
            output.write("coloratomrgb red=255 green=46 blue=46 at.selected")
        if (CSP_shift[i] < red) and (CSP_shift[i] >= red2):
            output.write("coloratomrgb red=255 green=153 blue=153 at.selected")
        if (CSP_shift[i] < red2) and (CSP_shift[i] >= orange):
            output.write("coloratomrgb red=255 green=204 blue=102 at.selected")
        if (CSP_shift[i] < orange):
            output.write("coloratomrgb red=191 green=191 blue=191 at.selected")
        output.write("\n")
        output.write("repatom rep=cpk at.selected")
        output.write("\n")
        output.write("findnext")
        output.write("\n")
        if (CSP_shift[i] >= red):
            output.write("coloratomrgb red=255 green=46 blue=46 at.selected")
        if (CSP_shift[i] < red) and (CSP_shift[i] >= red2):
            output.write("coloratomrgb red=255 green=153 blue=153 at.selected")
        if (CSP_shift[i] < red2) and (CSP_shift[i] >= orange):
            output.write("coloratomrgb red=255 green=204 blue=102 at.selected")
        if (CSP_shift[i] < orange):
            output.write("coloratomrgb red=191 green=191 blue=191 at.selected")
        output.write("\n")
        output.write("repatom rep=cpk at.selected")
        output.write("\n")
    output.write("fit all")
    output.write("\n")
    output.write("workspaceselectionclear")
    output.write("\n")
    output.write("endundoblock")

output.close()
ws_CSP = wb_CSP.close()
```

Solubility_Boxplot_and_Histogram.py

```
# Tobias Schöne,  01.10.2017

"""
============================================================================================
Imports parameter from excel sheet and gives Box plots or histogram per parameter for soluble
and unsoluble fragments
Check excel in order to define column of desired parameter
============================================================================================
"""

import matplotlib.pyplot as plt
import numpy as np
import os.path
from openpyxl import load_workbook
from pylab import show, bar, 247print247, ylabel, xlabel, xticks, yticks, title, axes,
savefig,close
from matplotlib import pyplot


sheet = "M_CR_2006_TS2017_Library.xlsx"

wb = load workbook(sheet)
ws = wb.active

# Read data of excel
##### count number of rows in xcel sheet
counter = 0
isOK = True
while isOK:
    if not (ws.cell(row=counter+1, column=1).value is None):
        counter = counter +1
    else:
        isOK = False
number of rows = counter + 1
print("Number of Fragments:", number_of_rows - 3)  #738

Solubility col = 19   # look into excel sheet

# 1 = soluble , 0 = unsoluble or almost unsoluble
Parameter 1 = []
Parameter 0 = []
parmamter_col = int(input("Parameter_column: ")) ## 16
Name = str(input("Enter name of 247print247n247 (e.g. Charge): "))
plot = str(input("Enter y/n for Histogram: "))
for i in range(3, number_of_rows):
    ### MW
    if(ws.cell(row=i, column=Solubility_col).value == "++"
       or ws.cell(row=i, column=Solubility_col).value =="+++"
       or ws.cell(row=i, column=Solubility_col).value == "++++"):
        ### soluble fragments
        Parameter 1.append((ws.cell(row=i, column=parmamter col).value))

    else:
        ### not soluble or almost unsoluble
        Parameter_0.append((ws.cell(row=i, column=parmamter_col).value))


##############################################################################
### Prepare data for Histogram ###############################################
##############################################################################

# count Parmameter for discrete values (e.g. H-Bond, Rotatabel bonds
def xValues(array):
    values temp = []
    for i in range(np.amin(array), np.amax(array)+1):
        values_temp.append(i)
    #print(values_temp)
    return np.asarray(values_temp)

def getFrequency(array, values):
    values percent=[0]*len(values)
    values_temp2 = [0]*len(values)
```

```
    for j in range(0,len(values_temp2)):
        for i in range(0, len(array)):
            if(array[i]==values[j]):
                values temp2[j] = values temp2[j] + 1
    for j in range(0, len(values_temp2)):
        values_percent[j] = (values_temp2[j]/len(array)*100)
    return np.asarray(values_percent)

### method
if(plot == "y"):
    # plot Histogram
    w = 0.15
    fig = plt.figure(figsize=(9, 6))
    fig.add subplot(111)
    ax = plt.subplot(111)
    values = xValues(Parameter 1+Parameter 0)
    ax.bar(values - w, (getFrequency(Parameter 1,xValues(Parameter 1+Parameter 0))), width=w,
color='#1D70B7', edgecolor='black', linewidth=1)
    ax.bar(values, (getFrequency(Parameter_0,xValues(Parameter_1+Parameter_0))), width=w,
color='#CC3333', edgecolor='black', linewidth=1)
    plt.ylabel('Frequency (%)')
    plt.xlabel(Name)
    plt.xticks(values)
    if(int(np.amax((getFrequency(Parameter_1,xValues(Parameter_1+Parameter_0)))))< 45):
        steps = 5
    else:
        steps = 10
    plt.yticks(range(0,
int(np.amax((getFrequency(Parameter_1,xValues(Parameter_1+Parameter_0)))))+8, steps))
    for item in ([ax.title, ax.xaxis.label, ax.yaxis.label] +
                 ax.get xticklabels() + ax.get yticklabels()):
        item.set fontsize(28)
    savefig("Solubility_%s.pdf" % Name)
    plt.show()
else:
    # plot BoxPlot
    fontsize = 28
    data = [Parameter 1, Parameter 0]
    fig = plt.figure(figsize=(9, 6))
    ax = fig.add subplot(111)
    box = ax.boxplot(data, notch=False, patch_artist=True,
                     medianprops=dict(linestyle='-', linewidth=1.5, color='black'))

    colors = ['#1D70B7', '#CC3333']
    # plt.ylim((-8, -2))
    my_xticks = ['high', 'low']
    x = np.array([1, 2])
    plt.xticks(x, my xticks, fontsize=28)
    # plt.yticks(fontsize=fontsize)
    plt.yticks(fontsize=28)
    ylabel('%s' % Name, fontsize=fontsize)
    xlabel('Solubility', fontsize=fontsize)
    for patch, color in zip(box['boxes'], colors):
        patch.set facecolor(color)
    savefig("Solubility_%s.pdf" % Name)

    # print statistical moments
    print("Median soluble: ", np.median(Parameter_1))
    print("mean soluble: ", np.mean(Parameter_1))
    print("std soluble: ", np.std(Parameter_1))
    print("Median unsoluble: ", np.median(Parameter_0))
    print("mean soluble: ", np.mean(Parameter_0))
    print("std soluble: ", np.std(Parameter_0))

    # plot Distribution
    fig = plt.figure(figsize=(9, 6))
    fig.add subplot(111)
    ax = plt.subplot(111)
    ax.hist(Parameter_1, bins='auto', edgecolor='black', color='#1D70B7')
    plt.xticks(fontsize=fontsize)
    plt.yticks(fontsize=fontsize)
    ylabel('Counts', fontsize=fontsize)
    xlabel('logP', fontsize=fontsize)
    savefig("Solubility_%s_Distribution_Solubles.pdf"%Name)
```

```
    fig = plt.figure(figsize=(9, 6))
    fig.add subplot(111)
    ax = plt.subplot(111)
    ax.hist(Parameter_0, bins='auto', edgecolor='black', color='#CC3333')
    fontsize = 28
    plt.xticks(fontsize=fontsize)
    plt.yticks(fontsize=fontsize)
    ylabel('Counts', fontsize=fontsize)
    xlabel('logP', fontsize=fontsize)
    savefig("Solubility_%s_Distribution_Unsolubles.pdf" % Name)
    plt.show()
```

Boxplot_and_Histogram_STD_Hits.py

```
# Tobias Schöne, 01.10.2017

"""
============================================================================================
Imports parameter from excel sheet and gives Box plots or histogram per parameter
Check excel in order to define column of desired parameter
============================================================================================
"""

import matplotlib.pyplot as plt
import numpy as np
import os.path
from openpyxl import load_workbook
from pylab import show, bar, 249print249, ylabel, xlabel, xticks, yticks, title, axes,
savefig,close
from matplotlib import pyplot
from scipy import stats
from scipy.stats import chisquare
from scipy.stats import chi2_contingency
import numpy as np



## load excel sheets
library = "M_CR_2006_TS2017_Library.xlsx"
STD_Hits = "STD_Hits_MCR2006_Saga_MI001_MNV.xlsx"  ## stored in col 1-3
wb = load workbook(library)
ws = wb.active
wb STD = load workbook(STD Hits)
ws_STD = wb_STD.active

## inputs
parmamter_col = int(input("Parameter_column: ")) ## 16
Name = str(input("Enter name of 249print249n249 (e.g. Charge): "))
plot = str(input("Enter y/n for Histogram: "))

# create STD hits array
Saga Hits = 76
MI001 Hits = 60
MNV Hits = 168
Saga, MI001, MNV, Non_binder = [],[],[],[]

isOK = True
counter = 1
while isOK:
    if not (ws_STD.cell(row=counter + 1, column=1).value is None):
        MNV.append((ws_STD.cell(row=counter + 1, column=1).value))
        counter = counter + 1
    else:
        isOK = False

isOK = True
counter = 1
while isOK:
    if not (ws_STD.cell(row=counter+1, column=2).value is None):
        Saga.append((ws STD.cell(row=counter+1, column=2).value))
        counter = counter +1
    else:
        isOK = False
```

```
isOK = True
counter = 1
while isOK:
    if not (ws_STD.cell(row=counter + 1, column=3).value is None):
        MI001.append((ws_STD.cell(row=counter + 1, column=3).value))
        counter = counter + 1
    else:
        isOK = False
isOK = True
counter = 1
while isOK:
    if not (ws_STD.cell(row=counter + 1, column=4).value is None):
        Non_binder.append((ws_STD.cell(row=counter + 1, column=4).value))
        counter = counter + 1
    else:
        isOK = False

print(len(Non_binder))
print(Non_binder)
print(len(MNV))
print(MNV)
print(len(Saga))
print(Saga)
print(len(MI001))
print(MI001)


# Read data of Excel
##### count number of rows in xcel sheet
counter = 0
isOK = True
while isOK:
    if not (ws.cell(row=counter+1, column=1).value is None):
        counter = counter +1
    else:
        isOK = False
#print(counter)
number_of_rows = counter + 1
print("Number of Fragments:", number_of_rows - 3)

Fragment_ID = 1   # look into excel sheet
number_of_bins = 30

### Methods
def grepParameter(comp,col):
    end = int(comp[2:]) + 10
    ###search for first comp
    counter = 1
    b = 1
    # search in TS2017 Library
    for i in range(2, end):
        if (ws.cell(row=i, column=1).value != comp):
            counter = counter + 1
    # print(counter)
        if (ws.cell(row=i, column=1).value == comp):
            b = counter + 1
        #print("b: ", b)
    return ws.cell(row=b, column=col).value

# create Parameter array for each drug target + all fragments
Parameter Non binder, Parameter Saga, Parameter MI001, Parameter MNV = [],[],[],[]


for i in range(0, len(Non_binder)):
    Parameter Non binder.append(ws.cell(row=int((Non binder[i])[2:]) + 2,
column=parmamter col).value)
for i in range(0,len(Saga)):
    Parameter Saga.append(ws.cell(row=int((Saga[i])[2:])+2, column=parmamter col).value)
for i in range(0,len(MI001)):
    Parameter_MI001.append(ws.cell(row=int((MI001[i])[2:])+2, column=parmamter_col).value)
for i in range(0,len(MNV)):
    Parameter MNV.append(ws.cell(row=int((MNV[i])[2:])+2, column=parmamter col).value)

print('all')
print(len(Parameter_Non_binder))
```

```
print('MNV')
print(np.sort(Parameter_MNV))
print("Saga")
print(np.sort(Parameter_Saga))
print("MI001")
print(np.sort(Parameter_MI001))
##############################################################################
### methods for Histogram / Boxplot ##########################################
##############################################################################

# count Parmameter for discrete values (e.g. H-Bond, Rotatabel bonds
def xValues(array):
    values_temp = []
    for i in range(np.amin(array), np.amax(array)+1):
        values temp.append(i)
    #print(values temp)
    return np.asarray(values_temp)

def getFrequency(array, values):
    values percent=[0]*len(values)
    values temp2 = [0]*len(values)
    for j in range(0,len(values_temp2)):
        for i in range(0, len(array)):
            if(array[i]==values[j]):
                values temp2[j] = values temp2[j] + 1
    for j in range(0, len(values_temp2)):
        values percent[j] = (values temp2[j]/len(array)*100)
    return np.asarray(values_percent)

### method
if(plot == "y"):
    # plot Histogram
    w = 0.15
    fig = plt.figure(figsize=(9, 6))
    fig.add_subplot(111)
    ax = plt.subplot(111)
    values = xValues(Parameter Non binder)
    ax.bar(values - (w+w/2), (getFrequency(Parameter Non binder,values)), width=w,
color='#505050', edgecolor='black', linewidth=1)
    ax.bar(values - (w/2), (getFrequency(Parameter MNV,values)), width=w, color='#F49B49',
edgecolor='black', linewidth=1)
    ax.bar(values + (w/2), (getFrequency(Parameter_Saga,values)), width=w, color='#00617A',
edgecolor='black', linewidth=1)
    ax.bar(values + (w+w/2), (getFrequency(Parameter_MI001,values)), width=w, color='#6EA5CE',
edgecolor='black', linewidth=1)
    plt.ylabel('Frequency (%)')
    plt.xlabel(Name)
    plt.xticks(values)
    max = 65 ## define
    plt.yticks(range(0, max, 10))
    for item in ([ax.title, ax.xaxis.label, ax.yaxis.label] +
                 ax.get_xticklabels() + ax.get_yticklabels()):
        item.set fontsize(28)

    # plt.rc('legend', fontsize=20)  # legend fontsize
    savefig("%s_Histo_STD_HITS_TS20171019.pdf" % Name)
    plt.show()

### chi-test - contingency
obs1 = getFrequency(Parameter col1,values)
obs2 = getFrequency(Parameter col2, values)
obs = (obs1,obs2)
chi2, p, dof, ex = chi2 contingency(obs)
print('p-value =', p)



else:
    # plot BoxPlot
    fontsize = 28
    data = [Parameter_Non_binder, Parameter_MNV, Parameter_Saga, Parameter_MI001]
    fig = plt.figure(figsize=(9, 6))
    ax = fig.add subplot(111)
    box = ax.boxplot(data, notch=False, patch artist=True,
                     medianprops=dict(linestyle='-', linewidth=1.5, color='black'))
```

```
    # box = plt.boxplot(data, notch=False, patch_artist=True, medianprops=dict(linestyle='-',
linewidth=1.5, color='black'))
    colors = ['#505050', '#F49B49', '#00617A', '#6EA5CE']
    # plt.ylim((-8, -2))
    my_xticks = ['NonB', 'MNV-1', 'Saga', 'MI001']
    x = np.array([1,2,3,4])
    plt.xticks(x, my_xticks, fontsize=28)
    # plt.yticks(fontsize=fontsize)
    plt.yticks(fontsize=28)
    ylabel('%s' % Name, fontsize=fontsize)
    #xlabel('Drug target', fontsize=fontsize)
    for patch, color in zip(box['boxes'], colors):
        patch.set_facecolor(color)
    savefig("%s__BoxPlot_STD_HITS_TS20171019.pdf" % Name)

    # t-test
    t,p = stats.ttest ind(Parameter col1, Parameter col2, equal var=True)
    # If False, perform Welch's t-test, which does not assume equal population variance
    print('p-value =', p)

    # plot Distribution
    fig = plt.figure(figsize=(9, 6))
    fig.add subplot(111)
    ax1 = plt.subplot(111)
    ax1.hist(Parameter_Non_binder, bins=number_of_bins, edgecolor='black', color='#505050')
    ax2 = plt.subplot(111)
    ax2.hist(Parameter_MNV, bins=number_of_bins, edgecolor='black', color='#F49B49')
    ax3 = plt.subplot(111)
    ax3.hist(Parameter_Saga, bins=number_of_bins, edgecolor='black', color='#00617A')
    ax4 = plt.subplot(111)
    ax4.hist(Parameter_MI001, bins=number_of_bins, edgecolor='black', color='#6EA5CE')
    plt.xticks(fontsize=fontsize)
    plt.yticks(fontsize=fontsize)
    ylabel('Count', fontsize=fontsize)
    xlabel('%s' % Name, fontsize=fontsize)
    savefig("%s_Distribution_Histo_TS20171019.pdf"%Name)

    plt.show()
```

searchParameter.py

```
# Tobias Schöne, 13.10.2017
"""

"""
import sys
import os
import os.path
import string
from openpyxl import Workbook
from openpyxl import load_workbook

### Parameter
parmamter_col = int(input("Parameter_column: ")) ## 21
Name = str(input("Enter name of 252uppressi (e.g. Charge): "))

### load Excel24
library = "M_CR_2006_TS2017_Library.xlsx"
STD_Hits = "3D_all_WNVP_non-binders_for_3D_Plot_A_vdW_Q.xlsx" ## insert list of fragments
stored in first column
wb = load workbook(library)
ws= wb.active
wb STD = load workbook(STD Hits)
ws_STD = wb_STD.active

hits = []
counter = 1
isOK = True
while isOK:
    if not (ws_STD.cell(row=counter+1, column=1).value is None):
        hits.append((ws STD.cell(row=counter+1, column=1).value))
        counter = counter + 1
    else:
```

```
        isOK = False

##### count number of colums in xcel sheet
counter = 0
isOK = True
while isOK:
    if not (ws_STD.cell(row=1, column=counter+1).value is None):
        counter = counter +1
    else:
        isOK = False
#print(counter)
free_col = counter + 1
print('free_col: ', free_col)

##### count number of rows in xcel sheet
counter = 0
isOK = True
while isOK:
    if not (ws_STD.cell(row=counter+1, column=1).value is None):
        counter = counter +1
    else:
        isOK = False
#print(counter)
number_of_rows = counter
print('Number of Hits: ', number_of_rows-1)
print(len(hits))
print(hits)

### Methods
def grepParameter(comp,col):
###search for first comp
    counter = 1
    b = 1
    # search Libray
    for i in range(2, 750):
        if (ws.cell(row=i, column=1).value != comp):
            counter = counter + 1
            #print(counter)
        if (ws.cell(row=i, column=1).value == comp):
            b = counter + 1
            #print("b: ",b)
    return ws.cell(row=b, column=col).value

### prepare Excel sheet
ws STD.cell(row=1,column = free col).value = Name

for i in range(0,len(hits)):
    ws STD.cell(row=i+2, column=free col).value = grepParameter(hits[i],parmamter col)

ws STD= wb STD.save(STD Hits)
ws_STD = wb_STD.close()
ws= wb.close()
```

3D_histo _for_A_D_Q.py

```
# Tobias Schöne, 14.10.2017

"""
================================================================================
plots scatter plots of three different parameters
the size and color of bubbles indicate the frequency within a given data set
================================================================================
"""

import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
import numpy
import os.path
from openpyxl import load_workbook
from pylab import show, bar, 253print253, ylabel, xlabel, xticks, yticks, title, axes,
savefig, close
import matplotlib
```

```
# load excel sheet and count number of Fragments
FragmentSheet = '3D_Plotting_M_CR_2006_Library_only_soluble_Fragments_A_D_Q_nach_vdW_volume'
cmap = 'binary'
wb = load_workbook('%s.xlsx'%(FragmentSheet))
ws = wb.active
counter = 1
isOK = True
while isOK:
    if not (ws.cell(row=counter+1, column=1).value is None):
        counter = counter + 1
    else:
        isOK = False
number of compounds = counter -1
print(number_of_compounds)
matrix fragments = numpy.zeros((number of compounds,3))
for i in range(0,number_of_compounds):
    matrix_fragments[:][i] = [ws.cell(row=i+2, column=2).value,ws.cell(row=i+2,
column=3).value,ws.cell(row=i+2, column=4).value]
# prepare matrix
x A = [0,1,2,3,4,5]
y D = [0,1,2,3]
z_Q = [-1,0,1,2]
_x,_y,_z = [],[],[]
matrix values = numpy.zeros(((len(x A)*len(y D)*len(z Q)),3))
counter = 0
for i in range(0,len(x_A)):
    for j in range(0,len(y_D)):
        for k in range(0,len(z_Q)):
            _x.append(x_A[i])
             y.append(y D[j])
             z.append(z Q[k])
            matrix values[:][counter] = [x A[i], y D[j],z Q[k]]
            counter = counter + 1

def index2D(array, matrix):
    counter = 0
    index = 0
    for i in range(0,len(matrix)):
        if not ((array == matrix[:][i]).all()):
            counter = counter + 1
        else:
            index = counter
    return index

# calculate ratio for each point
number_of_values = (len(x_A)*len(y_D)*len(z_Q))
List ratio = [0]*number of values
List colors = [0]*number of values
for i in range(0,number_of_compounds):
    List_ratio[(index2D((matrix_fragments[:][i]),matrix_values))] =
List_ratio[(index2D((matrix_fragments[:][i]),matrix_values))] + 1
    List colors[(index2D((matrix fragments[:][i]), matrix values))] = List colors[(
    index2D((matrix fragments[:][i]), matrix values))] + 1
print(len(matrix_values))
print(matrix_values)
print(List_ratio)
print(sum(List_ratio))
### scale counts
scale factor = 4000
sum = sum(List ratio)
for i in range(0,number_of_values):
    List_ratio[i] = (List_ratio[i]/sum) * scale_factor
    List_colors[i] = (List_colors[i]/sum)

print(np.sort(List_colors))

#plot figure
alpha_bubles = 1
#cmap = 'binary'
fig= plt.figure(1)
ax = Axes3D(fig)
s=  List ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
```

```
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x axis surface = np.arange(-1,7,1)
y_axis_surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x_axis_surface, y_axis_surface)
alpha value = 0.3
ax.plot_surface(xx, yy, z_Q[0], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z_Q[1], color='white',alpha=0.3)
ax.plot_surface(xx, yy, z_Q[2], color='white',alpha=0.15)
ax.plot_surface(xx, yy, z_Q[3], color='white',alpha=0.1)

#define axes
ax.set xticks(x A)
ax.set yticks(y D)
ax.set zticks(z Q)
ax.set_xlabel('H-bond acceptor')
ax.set_ylabel('H-bond donor')
ax.set_zlabel('Charge')
ax.set xlim(-1,6,1)
ax.set ylim(-1, 4,1)
ax.set_zlim(-2,3,1)

#plot 2D projection
x,y,z = [],[],[]
for i in range(0,number_of_compounds):
    x.append(matrix fragments[i][0])
    y.append(matrix_fragments[i][1])
    z.append(matrix_fragments[i][2])
size = 3
ax.plot(x, z, color ='gray', marker='.', markersize=size, linewidth=0, zdir='y', zs=4)
ax.plot(y, z, color ='gray', marker='.', markersize=size, linewidth=0, zdir='x', zs=-1)
ax.plot(x, y, color ='gray', marker='.', markersize=size, linewidth=0, zdir='z', zs=-2)
ax.view_init(16,-58)

#plot figure
fig= plt.figure(2)
ax = Axes3D(fig)
s=  List ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x_axis_surface = np.arange(-1,7,1)
y axis surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x axis surface, y axis surface)
#define axes
ax.set_xticks(x_A)
ax.set_yticks(y_D)
ax.set zticks(z Q)
ax.set_xlabel('H-bond acceptor')
ax.set_ylabel('H-bond donor')
#ax.set zlabel('Charge')
ax.set_xlim(-1,6,1)
ax.set_ylim(-1, 4,1)
ax.set zlim(-2,3,1)
ax.view init(89.99,90.01)

#plot figure
fig= plt.figure(3)
ax = Axes3D(fig)
s=  List ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x axis surface = np.arange(-1,7,1)
y axis surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x_axis_surface, y_axis_surface)
ax.plot_surface(xx, yy, z_Q[0], color='white',alpha=0.4)
```

```
ax.plot_surface(xx, yy, z_Q[1], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z_Q[2], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z_Q[3], color='white',alpha=0.4)
#define axes
ax.set xticks(x A)
ax.set_yticks(y_D)
ax.set_zticks(z_Q)
ax.set_xlabel('H-bond acceptor')
#ax.set ylabel('H-bond donor')
ax.set_zlabel('Charge')
ax.set xlim(-1,6,1)
ax.set_ylim(-1, 4,1)
ax.set_zlim(-2,3,1)
ax.view init(0.01,90.01)

#plot figure
fig= plt.figure(4)
ax = Axes3D(fig)
s=  List_ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x axis surface = np.arange(-1,7,1)
y axis surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x axis surface, y axis surface)
ax.plot_surface(xx, yy, z_Q[0], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z_Q[1], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z_Q[2], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z_Q[3], color='white',alpha=0.4)
#define axes
ax.set xticks(x A)
ax.set_yticks(y_D)
ax.set_zticks(z_Q)
#ax.set xlabel('H-bond acceptor')
ax.set_ylabel('H-bond donor')
ax.set_zlabel('Charge')
ax.set xlim(-1,6,1)
ax.set_ylim(-1, 4,1)
ax.set_zlim(-2,3,1)
ax.view init(0,-0.01)

plt.show()
```

3D_histo _plot_for_A_vdW_Q.py

```
# Tobias Schöne, 14.10.2017

"""
================================================================================
plots scatter plots of three different parameters
the size and color of bubbles indicate the frequency within a given data set
================================================================================

"""

import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
import numpy
import os.path
from openpyxl import load_workbook
from pylab import show, bar, 256print256, ylabel, xlabel, xticks, yticks, title, axes,
savefig, close
import matplotlib

# load excel sheet and count number of Fragments
##############################################

# binning of data >>> MW
```

```
def binning(value):
    a = 0
    if(value<140):
        a = 0
    if (140<=value<160):
        a = 1
    if (160<=value<180):
        a = 2
    if (180<=value):
        a = 3
    return a

FragmentSheet = '3D_STD_Hits_MCR2006_MNV_for_3D_Plot_A_vdW_Q'
cmap = 'Blues'  # https://matplotlib.org/examples/color/colormaps_reference.html
wb = load_workbook('%s.xlsx'%(FragmentSheet))
ws = wb.active
counter = 1
isOK = True
while isOK:
    if not (ws.cell(row=counter+1, column=1).value is None):
        counter = counter + 1
    else:
        isOK = False
number_of_compounds = counter -1
print(number_of_compounds)
matrix fragments = numpy.zeros((number of compounds,3))
for i in range(0,number_of_compounds):
    matrix fragments[:][i] = [ws.cell(row=i+2, column=2).value,binning(ws.cell(row=i+2,
column=3).value),(ws.cell(row=i+2, column=4).value)]
print(matrix_fragments)

# prepare matrix
x   = [0,1,2,3,4,5]
y   = [0,1,2,3]
z__ = [-1,0,1,2]
_x,_y,_z = [],[],[]
matrix values = numpy.zeros(((len(x  )*len(y  )*len(z  )),3))
counter = 0
for i in range(0,len(x__)):
    for j in range(0,len(y  )):
        for k in range(0,len(z__)):
            _x.append(x__[i])
             y.append(y  [j])
             z.append(z  [k])
            matrix values[:][counter] = [x  [i], y  [j],z  [k]]
            counter = counter + 1

def index2D(array, matrix):
    counter = 0
    index = 0
    for i in range(0,len(matrix)):
        if not ((array == matrix[:][i]).all()):
            counter = counter + 1
        else:
            index = counter
    return index

# calculate ratio for each point
number of values = (len(x  )*len(y  )*len(z  ))
List ratio = [0]*number of values
List colors = [0]*number of values
for i in range(0,number_of_compounds):
    List_ratio[(index2D((matrix_fragments[:][i]),matrix_values))] =
List_ratio[(index2D((matrix_fragments[:][i]),matrix_values))] + 1
    List colors[(index2D((matrix fragments[:][i]), matrix values))] = List colors[(
    index2D((matrix fragments[:][i]), matrix values))] + 1
print(len(matrix_values))
print(matrix_values)
print(List_ratio)
print(sum(List_ratio))
### scale counts
scale factor = 4000
sum = sum(List ratio)
for i in range(0,number_of_values):
    List_ratio[i] = (List_ratio[i]/sum) * scale_factor
```

```
    List_colors[i] = (List_colors[i]/sum)

#plot figure
alpha bubles = 1
#cmap = 'binary'
fig= plt.figure(1)
ax = Axes3D(fig)
s=  List ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x  xis surface = np.arange(-1,7,1)
y axis surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x  xis surface, y axis surface)
alpha_value = 0.3
ax.plot_surface(xx, yy, z__[0], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[1], color='white',alpha=0.3)
ax.plot_surface(xx, yy, z__[2], color='white',alpha=0.15)
ax.plot_surface(xx, yy, z__[3], color='white',alpha=0.1)

#define axes
my_yticks = ['140', '160', '180', '300']
plt.yticks(y  , my yticks, fontsize=10)
ax.set xticks(x  )
#ax.set yticks(y  )
ax.set_zticks(z__)
ax.set_xlabel('H-bond acceptor')
ax.set_ylabel('vdW volume')
ax.set_zlabel('Charge')
ax.set xlim(-1,6,1)
ax.set ylim(-1, 4,1)
ax.set_zlim(-2,3,1)

#plot 2D projection
x,y,z = [],[],[]
for i in range(0,number_of_compounds):
    x.append(matrix fragments[i][0])
    y.append(matrix_fragments[i][1])
    z.append(matrix_fragments[i][2])
size = 3
ax.plot(x, z, color ='gray', marker='.', markersize=size, linewidth=0, zdir='y', zs=4)
ax.plot(y, z, color ='gray', marker='.', markersize=size, linewidth=0, zdir='x', zs=-1)
ax.plot(x, y, color ='gray', marker='.', markersize=size, linewidth=0, zdir='z', zs=-2)
ax.view_init(10,-40)

#plot figure
fig= plt.figure(2)
ax = Axes3D(fig)
s=  List_ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x  xis surface = np.arange(-1,7,1)
y axis surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x  xis surface, y axis surface)
#define axes
my_yticks = ['140', '160', '180', '300']
plt.yticks(y__, my_yticks, fontsize=10)
ax.set xticks(x  )
#ax.set yticks(y  )
ax.set zticks(z  )
ax.set_xlabel('H-bond acceptor')
ax.set_ylabel('vdW volume')
ax.set_zlabel('Charge')
ax.set xlim(-1,6,1)
ax.set ylim(-1, 4,1)
ax.set zlim(-2,3,1)
ax.view_init(89.99,90.01)
```

```
#plot figure
fig= plt.figure(3)
ax = Axes3D(fig)
s=  List ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x  xis surface = np.arange(-1,7,1)
y_axis_surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x__xis_surface, y_axis_surface)
ax.plot_surface(xx, yy, z__[0], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[1], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[2], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[3], color='white',alpha=0.4)
#define axes
my_yticks = ['140', '160', '180', '300']
plt.yticks(y  , my yticks, fontsize=10)
ax.set xticks(x  )
#ax.set yticks(y  )
ax.set_zticks(z__)
ax.set_xlabel('H-bond acceptor')
ax.set_ylabel('vdW volume')
ax.set_zlabel('Charge')
ax.set xlim(-1,6,1)
ax.set ylim(-1, 4,1)
ax.set_zlim(-2,3,1)
ax.view_init(0.01,90.01)

#plot figure
fig= plt.figure(4)
ax = Axes3D(fig)
s=  List_ratio
ax.scatter(_x,_y,_z,s=s,c=List_colors,cmap=cmap,alpha=alpha_bubles, edgecolor='black',
linewidths=1)
# create plane surface
point  = np.array([1, 2, 3])
normal = np.array([1, 1, -1])
d = -point.dot(normal)
x__xis_surface = np.arange(-1,7,1)
y axis surface = np.arange(-1,5,1)
xx, yy = np.meshgrid(x  xis surface, y axis surface)
ax.plot_surface(xx, yy, z__[0], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[1], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[2], color='white',alpha=0.4)
ax.plot_surface(xx, yy, z__[3], color='white',alpha=0.4)
#define axes
my_yticks = ['140', '160', '180', '300']
plt.yticks(y__, my_yticks, fontsize=10)
ax.set_xticks(x__)
#ax.set yticks(y  )
ax.set zticks(z  )
ax.set_xlabel('H-bond acceptor')
ax.set_ylabel('vdW volume')
ax.set_zlabel('Charge')
ax.set_xlim(-1,6,1)
ax.set ylim(-1, 4,1)
ax.set zlim(-2,3,1)
ax.view init(0,-0.01)

plt.show()
```

Correlation_Analysis_Heatmap.py

```
### Tobias Schöne, 23.10.2017

import matplotlib.pyplot as plt
import numpy as np
import os.path
from openpyxl import load_workbook
from pylab import show, bar, 260print260, ylabel, xlabel, xticks, yticks, title, axes,
savefig,close
from matplotlib import pyplot


library = "CSP_Fragments_Correlation_sortiert_nach_corrcoef.xlsx"
Name = str(input("Please instert name of xcel sheet: "))
wb = load workbook(library)
ws = wb.active
os.system("cd C:/python/")
os.system("copy EMPTY.xlsx %s.xlsx" %Name)

wb2 = load_workbook("%s.xlsx" %Name)
ws2 = wb2.active

### prepare matrix and grep fragment IDs
parameter = []
counter = 0
isOK = True
while isOK:
    if not (ws.cell(row=1, column=counter+1).value is None):
        ws2.cell(row=counter+2, column=11).value = (ws.cell(row=1, column=counter+1).value)
        ws2.cell(row=1, column=counter+2).value = (ws.cell(row=1, column=counter + 1).value)
        parameter.append(ws.cell(row=1, column=counter+1).value)
        counter = counter +1
    else:
        isOK = False
print(parameter)
dimension = counter
print(dimension)

### create correlation coefficient matrix
corr_matrix = np.zeros((dimension,dimension))
counter = 0
for x in range(1,dimension+1):
    counter = counter + 1
    for j in range(counter,dimension+1):
        array1 = []
        array2 = []
        isOK = True
        counter2 = 1
        while isOK:
            if not (ws.cell(row=counter2+1, column=1).value is None):
                array1.append((ws.cell(row=counter2+1, column=x).value))
                array2.append((ws.cell(row=counter2+1, column=j).value))
                counter2 = counter2 + 1
            else:
                isOK = False
        corrcoef = (np.corrcoef(array1, array2)[0][1])
        corr matrix[x-1][j-1] = corrcoef
        corr matrix[j-1][x - 1] = corrcoef
        ws2.cell(row=j+1, column=x+1).value = corrcoef
        ws2.cell(row=x + 1, column=j+1).value = corrcoef
ws2 = wb2.save("%s.xlsx" %Name)


### plot heatmap
plt.imshow(corr_matrix, cmap='RdYlBu')
my xticks = my yticks = parameter
x = np.arange(0,dimension)
y = np.arange(0,dimension)
plt.xticks(x, my_xticks, fontsize=9, rotation='vertical')
plt.yticks(y, my xticks, fontsize=9)
#plt.imshow(corr matrix, cmap='coolwarm', interpolation ='nearest')
plt.colorbar()
plt.show()
```

A5 NMR Pulse Programs and AU Programs

Zgesgp

```
;zgesgp
;avance-version (10/02/12)
;1D sequence
;water suppression using excitation sculpting with gradients
;T.-L. Hwang & A.J. Shaka, J. Magn. Reson.,
;   Series A 112 275-279 (1995)
;
;$CLASS=HighRes
;$DIM=1D
;$TYPE=
;$SUBTYPE=
;$COMMENT=
prosol relations=<triple>
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

“p2=p1*2”
„d12=20u“
„TAU=de+p1*2/3.1416+50u“
„acqt0=0“
baseopt_echo
1 ze
2 30m
 d12 pl1:f1 BLKGRAD
 d1
 p1 ph1

 50u UNBLKGRAD
 p16:gp1
 d16 pl0:f1
 (p12:sp1 ph2:r):f1
 4u
 d12 pl1:f1
 p2 ph3
 4u
 p16:gp1
 d16
 TAU
 p16:gp2
```

```
 d16 pl0:f1
 (p12:sp1 ph4:r):f1
 4u
 d12 pl1:f1

 p2 ph5
 4u
 p16:gp2
 d16

 go=2 ph31
 30m mc #0 to 2 F0(zd)
 4u BLKGRAD
exit

ph1=0
ph2=0 1
ph3=2 3
ph4=0 0 1 1
ph5=2 2 3 3
ph31=0 2 2 0

;pl0 : 0W
;pl1 : f1 channel – power level for pulse (default)
;sp1 : f1 channel – shaped pulse 180 degree
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel – 180 degree high power pulse
;p12: f1 channel – 180 degree shaped pulse (Squa100.1000)   [2 msec]
;p16: homospoil/gradient pulse
;d1 : relaxation delay; 1-5 * T1
;d12: delay for power switching                    [20 usec]
;d16: delay for homospoil/gradient recovery
;NS: 8 * n, total number of scans: NS * TD0
;DS: 4
;use gradient ratio:    gp 1 : gp 2
;                    31 :   11
;for z-only gradients:
;gpz1: 31%
;gpz2: 11%
;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;$Id: zgesgp,v 1.8 2010/02/12 15:03:56 ber Exp $
```

Stddiffesgp.3

```
;avance-version (11/01/24)
;pseudo 2D sequence
;   for saturation transfer difference
;with shaped pulse train for saturation on f2 channel
;alternating between on and off resonance
;   to be defined by fq2list
;with spoil sequence to destoy unwanted magnetization
;water suppression using excitation sculpting with gradients
;with spinlock to suppress protein signals
;(use parameterset STDDIFFESGP.3)
;M. Mayer & B. Meyer, Angew. Chem. Int. Ed. 38, 1784-1788 (1999)
;M. Mayer & B. Meyer, Angew. Chem. 111, 1902-1906 (1999).
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>
define list<frequency> stdlist=<$FQ2LIST>

"p2=p1*2"
"d12=20u"
"d11=30m"
"p29=d29"
"l5=d20/p42"
"d31=p42*l5"
"DELTA1=d1-d31"
"TAU=de+p1*2/3.1416+50u"

"nbl=td1"

"acqt0=0"
baseopt_echo

1 ze
  3m stdlist:f2 st0
2 6m
3 6m
4 d11
  6m

5 50u UNBLKGRAD
  4u pl10:f1
  (p17 ph2)
  (p17*2 ph3)
  4u
  p30:gp1
  10m pl1:f1
  4u BLKGRAD


  DELTA1


6 (p42:sp9 ph4):f2
  4u
  lo to 6 times l5

  p1 ph1
  4u pl29:f1
  (p29 ph5)

  50u UNBLKGRAD
  p16:gp2
  d16 pl0:f1
  (p40:sp10 ph6:r):f1
  4u
  d12 pl1:f1

  p2 ph7

  4u
  p16:gp2
  d16
  TAU
  p16:gp3
  d16 pl0:f1
  (p40:sp10 ph8:r):f1
  4u
  d12 pl1:f1

  p2 ph9

  p16:gp3
  d16
  4u BLKGRAD


goscnp ph31

3m stdlist.inc
  3m stdlist:f2 st
  lo to 3 times nbl
  3m ipp1 ipp5 ipp6 ipp7 ipp8 ipp9
ipp31
  3m stdlist.res
  lo to 4 times ns
  d11 wr #0
  3m rppall
  3m zd
  lo to 5 times l4
exit


ph1=0 2
ph2=0
ph3=1
ph4=0
ph5=1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3
ph6=0 0 1 1
ph7=2 2 3 3
ph8=0 0 0 0 1 1 1 1
ph9=2 2 2 2 3 3 3 3
ph31=0 2 2 0 2 0 0 2
```

Comments:

```
;pl0 : 0W
;pl1 : f1 channel – power level for pulse (default)
;pl2 : f2 channel – power level for pulse (default)      [120 dB]
;pl10: f1 channel – power level for TOCSY-spinlock
;pl29: f1 channel – power level for trim pulse
;sp9 : f2 channel – shaped pulse  for saturation         [40 – 60 dB]
;sp10: f1 channel – shaped pulse 180 degree
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel – 180 degree high power pulse
;p16: homospoil/gradient pulse
;p17: f1 channel – trim pulse                    [2.5 msec]
;p29: f1 channel – trim pulse
;p30: gradient pulse                             [3 msec]
;p40: f1 channel – 180 degree shaped pulse (Squa100.1000) [2 msec]
;p42: f2 channel -  shaped pulse for saturation          [50 msec]
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                         [30 msec]
;d12: delay for power switching                  [20 usec]
;d16: delay for homospoil/gradient recovery
;d20: saturation time
;d29: spinlock time                              [10 – 50 msec]
;d31: saturation time as executed
;l4: l4 = number of averages = (total number of scans) / NS
;l5: loop for saturation: p42 * l5 = saturation time
;NS: 8 * n
;DS: 4
;td1: number of experiments
;NBL: NBL = number of irradiation frequencies
;define FQ2LIST (irradiation frequencies)
  (list has to be stored in "/u/exp/stan/nmr/lists/f1")

;use gradient ratio:    gp 1 : gp 2 : gp 3
;                       40 :   31 :   11

;for z-only gradients:
;gpz1: 40%
;gpz2: 31%
;gpz3: 11%
;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;this pulse program produces a ser-file (PARMOD = 2D)
;The STD experiment is protected by international patents owned by:
;Alepharma Licensing, Raamfeld 67, 22397 Hamburg, Germany.
;For commercial use (direct or indirect) please contact the company for
;licensing information at:
;E-mail: info@alepharma-licensing.com,
;Fax: +49 4060847812,
;Tel: +49 1701685158 or +49 1712788867.

;$Id: stddiffesgp.3,v 1.9.2.1 2011/01/24 15:49:24 ber Exp $
```

AU program: Stdproc_2

```
/*     stdsplit                                    */
/*     Short Description :                         */
/*     Splitting STD pseudo 2D datasets            */
/*     Keywords :                                  */
/*     STD                                         */
/*     Description/Usage :                         */
/*     Splits STD pseudo 2D data into reference spectrum and   */
/*     difference(s)                               */
/*     Author(s) :                                 */
/*     Name          : Wolfgang Bermel             */
/*     Organisation    : Bruker Analytik           */
/*     Email         : wolfgang.bermel@bruker.de   */
/*                                                                                 */
/*     Name          : Thorsten Biet               */
/*     Organisation    : Universitaet Lübeck       */
/*     Email         : 265print265n.biet@chemie.uni-luebeck.de   */
/*     Name          Date    Modification:         */
/*     ber           050712  created               */
/*                Biet          090318  zusaetzliche Abfrage         */
/*
$Id: stdsplit,v 1.1.2.1 2006/04/18 13:52:56 rke Exp $
*/
char path[PATH_MAX];
int td1, td2;
int counter;
int nexpno, refexpno, byteorder, parmode;
int refrow;
int method;
double dcin;
/***** get dataset and parameters *****/
GETCURDATA
FETCHPARS("BYTORDA",&byteorder)
FETCHPARS("TD",&td2)
td2 = ( (td2 + 255) / 256 ) * 256;

FETCHPAR1S("TD",&td1)
if ( td1 < 2 )
  {
  STOPMSG("not enough increments");
  }
FETCHPAR("PARMODE",&parmode);

if ( parmode != 1 )
  {
  STOPMSG("Dataset is not a 2D");
  }
/***** get input *****/
refexpno = expno * 100;
GETINT("First dataset to be stored in expno : ",refexpno);
nexpno = refexpno + 1;
refrow = 1;
GETINT("Reference experiment (off-resonance) stored in row : ",refrow);

/***** read rows *****/
for ( counter=1; counter<=td1 ; counter++ )
  {
  if ( counter == refrow )
    {
    RSER(counter, refexpno, 1);
    }
  else
    {
    RSER(counter, nexpno, 1);
    nexpno += 1;
    }
```

```
  }

/***** calculate difference *****/
nexpno = refexpno + 1;
method = -1;
GETINT("Method: Add(standard)=1 or Substract=-1",method);
if (method==1)
        {
        dcin=1.0;
        }
        else if (method==-1)
        {
        dcin=-1.0;
        }
        else
        {
        STOPMSG("The method must be either 1 (addition) or -1 (subtraction)")
        }

for ( counter=1; counter<td1 ; counter++ )
  {
  DATASET (name,nexpno,procno,disk,user);
  DATASET2 (name,refexpno,procno,disk,user)
  DATASET3 (name,nexpno,procno,disk,user);

  STOREPAR("DC",dcin);
  ADDFID;
  nexpno += 1;
  }

/***** print message *****/
nexpno = refexpno + 1;
if ( td1 == 2)
  {
  (void)266print(path,"reference experiment in expno %d\n difference experiment in expno %d",refexpno, nexpno);
  }
else
  {
  (void)266print(path,"reference experiment in expno %d\n difference experiments in expno %d to %d",refexpno, nexpno, nexpno+td1-2);
  }

QUITMSG(path)
```

Trosyetf3gpsi.2

```
;trosyetf3gpsi.2
;avance-version (10/02/12)
;2D H-1/X correlation via TROSY
;   using sensitivity improvement
;phase sensitive using Echo/Antiecho gradient selection
;using f3 – channel
;(use parameterset )
;
;T. Schulte-Herbrueggen & O.W. Sorensen, J. Magn. Reson. 144,
;   123 – 128 (2000)
;(M. Czisch & R. Boelens, J. Magn. Reson. 134, 158-160 (1998) )
;(K. Pervushin, G. Wider & K. Wuethrich, J. Biomol. NMR 12,
;   345-348 (1998) )
;(A. Meissner, T. Schulte-Herbrueggen, J. Briand & O.W. Sorensen, Mol. Phys. 96,
;   1137-1142 (1998) )
;(J. Weigelt, J Am Chem Soc 120, 10778-10779 (1998) )
;(M. Rance, J.P. Loria & A.G. Palmer III, J. Magn. Reson. 136, 91-101 (1999) )
;(G. Zhu, X.M. Kong & K.H. Sze, J. Biomol. NMR 13, 77-81 (1999) )
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=

prosol relations=<triple>
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>
define list<gradient> EA2 = { 0.8750 1.0000}
define list<gradient> EA4 = { 1.0000 0.6667}
define list<gradient> EA6 = { 0.6595 1.0000}

“p2=p1*2”
“p22=p21*2”
“d11=30m”
“d26=1s/(cnst4*4)”
„in0=inf1/2“
„d0=6u“
```

```
„DELTA1=d26-p16-d16“
“DELTA2=d25-p16-d16”
“DELTA3=d26-p11-p16-d16-8u”
#   ifdef LABEL_CN
“DELTA=d0*2+p8-p21*4/3.1416+8u”
#   else
„DELTA=d0*2-p21*4/3.1416+6u“
#   endif /*LABEL_CN*/
1 ze
2 d11
3 d1 pl1:f1
  50u UNBLKGRAD
  (p1 ph3)
  p16:gp1
  d16
  DELTA1
  (center (p2 ph2) (p22 ph1):f3 )
  DELTA1
  p16:gp1
  d16
  (p1 ph2)
  p16:gp2*EA2
  d16
  (p21 ph5):f3
  DELTA
  (p22 ph1):f3

#   ifdef LABEL_CN
  d0 gron0
  2u groff
  (p8:sp13 ph1):f2
  d0 gron0*-1
  2u groff
#   else
  d0 gron0
  d0 gron0*-1
  2u groff
#   endif /*LABEL_CN*/

  4u
  p16:gp2*-1*EA2
```

```
 d16
 (p1 ph6)
 p16:gp3
 d16
 DELTA2
 (center (p2 ph2) (p22 ph2):f3 )
 DELTA2
 p16:gp3
 d16
 (p1 ph1)
 p16:gp4*EA4
 d16

 (p21 ph7):f3
 p16:gp5
 d16
 DELTA3 pl0:f1
 (p11:sp1 ph4:r):f1
 4u
 4u pl1:f1
 (center (p2 ph2) (p22 ph8):f3 )
 4u pl0:f1
 (p11:sp1 ph4:r):f1
 4u
 DELTA3
 p16:gp5
 d16 pl1:f1
 (p21 ph9:r):f3
 p16:gp6*EA6
 d16
 4u BLKGRAD
 go=2 ph31
 d11 mc #0 to 2
   F1EA(calgrad(EA2) & calgrad(EA4) & calgrad(EA6) & calph(ph6, +180) & calph(ph7, +180), caldel(d0,
+in0) & calph(ph5, +180) & calph(ph31, +180))
exit

ph1=0
ph2=1
ph3=2
ph4=3
```

```
ph5=0 2
ph6=3
ph7=0 0 2 2
ph8=1 1 3 3
ph9=3 3 1 1
ph31=0 2 2 0


;pl0 : 0W
;pl1 : f1 channel – power level for pulse (default)
;pl3 : f3 channel – power level for pulse (default)
;sp1: f1 channel – shaped pulse  90 degree  (H2O on resonance)
;sp13: f2 channel – shaped pulse 180 degree (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel – 180 degree high power pulse
;p8 : f2 channel – 180 degree shaped pulse for inversion (adiabatic)
;p11: f1 channel -  90 degree shaped pulse           [1 msec]
;p16: homospoil/gradient pulse                      [1 msec]
;p21: f3 channel -  90 degree high power pulse
;p22: f3 channel – 180 degree high power pulse
;d0 : incremented delay (2D)                        [6 usec]
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O                            [30 msec]
;d16: delay for homospoil/gradient recovery
;d25 : 1/(4J')NH,
;       compensation delay for suppression of other Trosy peaks
;d26 : 1/(4J)NH
;cnst4: = J(NH)
;inf1: 1/SW(N) = 2 * DW(N)
;in0: 1/(2 * SW(N)) = DW(N)
;nd0: 2
;NS: 4 * n
;DS: 16
;td1: number of experiments
;FnMODE: echo-antiecho

;use gradient ratio: gp 0 : gp 1 : gp 2 : gp 3 : gp 4 : gp 5 : gp 6
;                    3 :    3 :   80 :    5 :   30 :    7 :30.13
;for z-only gradients:
;gpz0: 3%
;gpz1: 3%
```

;gpz2: 80%
;gpz3: 5%
;gpz4: 30%
;gpz5: 7%
;gpz6: 30.13%
;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100
;gpnam5: SMSQ10.100
;gpnam6: SMSQ10.100
 ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
; option –DLABEL_CN (eda: ZGOPTNS)
;preprocessor-flags-end

;Processing
;PHC0(F1): 22.5
;$Id: trosyetf3gpsi.2,v 1.3 2010/02/12 15:03:56 ber Exp $

Metr-hmqcph.tb

```
;metr-hmqcph.tb
;avance-version (09/04/17)
;Methyl-TROSY HMQC with optional Water 272uppression
;2D H-1/X correlation via heteronuclear zero and double quantum
;   coherence
;phase sensitive
;with decoupling during acquisition
;
;V. Tugarinov, P.W. Hwang, J.E. Ollerenshaw & L.E. Kay,
;                J. Magn. Reson. 125, 10420-10428 (2003)
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>
“p2=p1*2”
“d2=3.6m”
“d11=30m”
“d12=20u”
“d13=4u”
„d0=3u“
„in0=inf1/2“
„DELTA1=d2-p19-d16-d13“
1 ze
2 d11 do:f2
3 4u
# ifdef WATER
  d11 fq=cnst20 (bf ppm) :f1
  d11 pl9:f1
  d1 cw:f1 ph29
  d13 do:f1
  d13 pl2:f2
# else
  d1 pl2:f2
# endif
```

```
  p3:f2 ph0
  d13 UNBLKGRAD
# ifdef WATER
  4u
  p16:gp1
  d16 fq=0 :f1
# else
  4u
  p16:gp1
  d16
# endif
  d12 pl1:f1
  p1 ph1
  d13
  p19:gp2
  d16
  DELTA1
  p3:f2 ph3
  d0
  p2 ph2
  d0
  p3:f2 ph4
  DELTA1
  p19:gp2
  d16 pl12:f2
  d13 BLKGRAD
  go=2 ph31 cpd2:f2
  d11 do:f2 mc #0 to 2 F1PH(calph(ph3, +90) & calph(ph29, +90), caldel(d0, +in0))
exit
ph0=0
ph1=0
ph2=0
ph3=0 2
ph4=0 0 2 2
ph29=0
ph31=0 2 2 0


;pl1 : f1 channel – power level for pulse (default)
;pl2 : f2 channel – power level for pulse (default)
```

```
;pl9 : f1 channel – power level for presaturation
;pl12: f2 channel – power level for CPD/BB decoupling
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel – 180 degree high power pulse
;p3 : f2 channel -  90 degree high power pulse
;p16: homospoil/gradient pulse                    [1 msec]
;p19: gradient pulse 2                           [500 usec]
;d0 : incremented delay (2D)                     [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d2 : 1/(2J)XH                                   [3.6 ms]
;d11: delay for disk I/O                         [30 msec]
;d12: delay for power switching                  [20 usec]
;d13: short delay                                [4 usec]
;d16: delay for homospoil/gradient recovery
;cnst20: chemical shift of residual Water peak (in ppm)   [4.7 ppm]
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 4 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEQ
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel – 90 degree pulse for decoupling sequence

;for z-only gradients:
;gpz1: 30%
;gpz2: 40%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100

                              ;preprocessor-flags-start
;WATER: for optional Watersupression start experiment with
;          option –DWATER (eda: ZGOPTNS)
                              ;preprocessor-flags-end

;$Id: hmqcphpr,v 1.5 2009/07/02 16:40:44 ber Exp $
```

A6 SPR Experiments

Adsorption Experiments

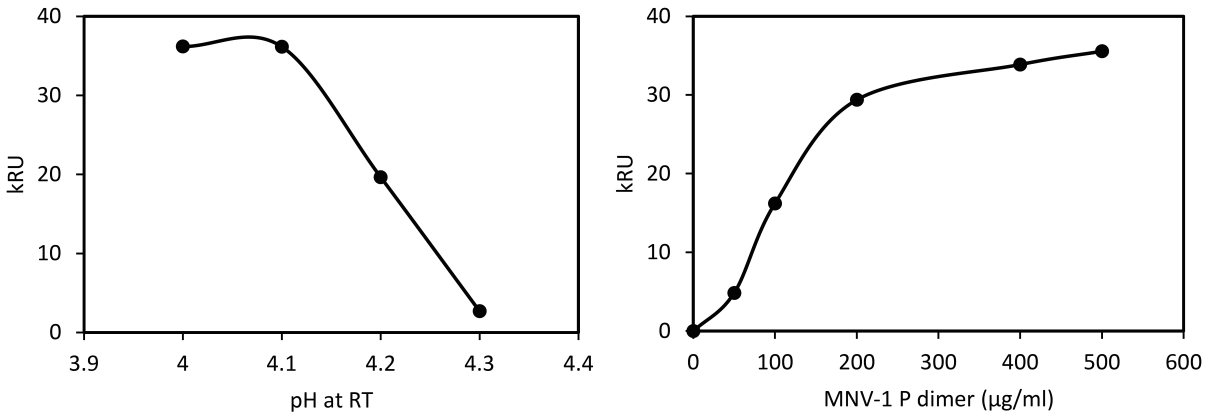


Figure A12 | Adsorption of MNV-1 P dimers to the chip matrix depending on pH and protein concentration

Sensograms

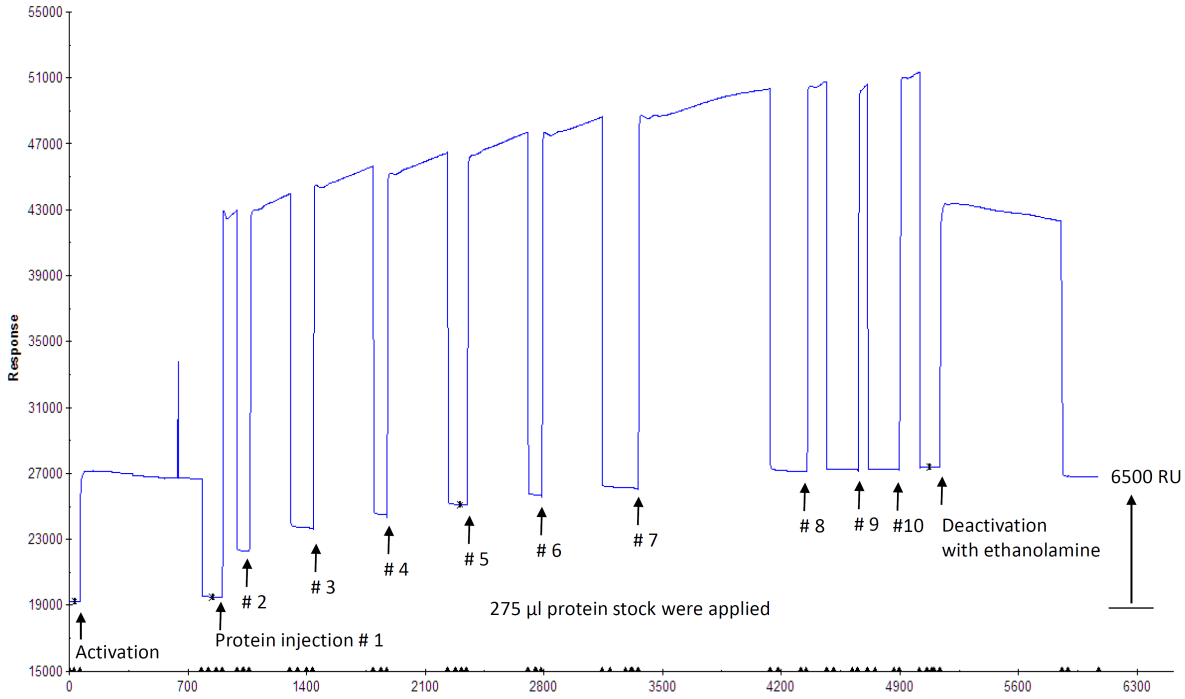


Figure A13 | SPR sensogram of the immobilization of MNV-1 using the amine coupling methods
The carboxy-methylated surface was activated by injection of 120 µl of freshly prepared mixture containing 0.05 M N-Hydroxysuccinimide (NHS) and 0.2 M 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) at flow rate of 10 µl/ min. Freshly purified MNV-1 P dimers (~7 mg/ml) were diluted with 10 mM sodium acetate to reach a final concentration of ~400 µg/ml and a final pH of 4.1. Immobilization was conducted at a flow rate of 5 µl/min with several injections of 5 – 20 µl until intended response level was reached (~6500 RU). Remaining activated groups were deactivated by injection of 120 µl of 1 M ethanolamine (pH 8.5) at flow rate of 10 µl/min.

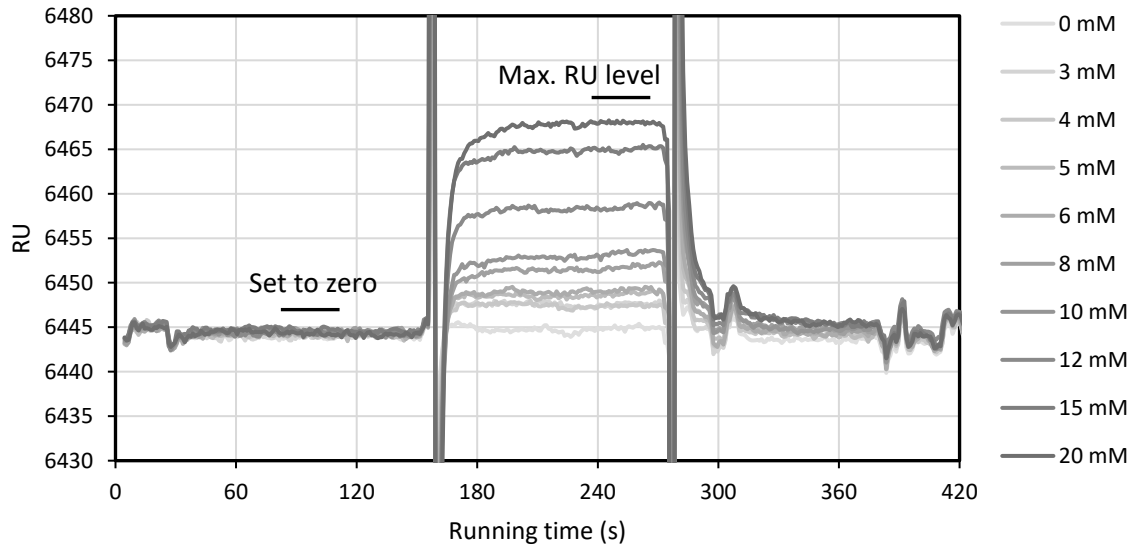


Figure A14 | SPR Sensogram of 3'-sialyllactose titration

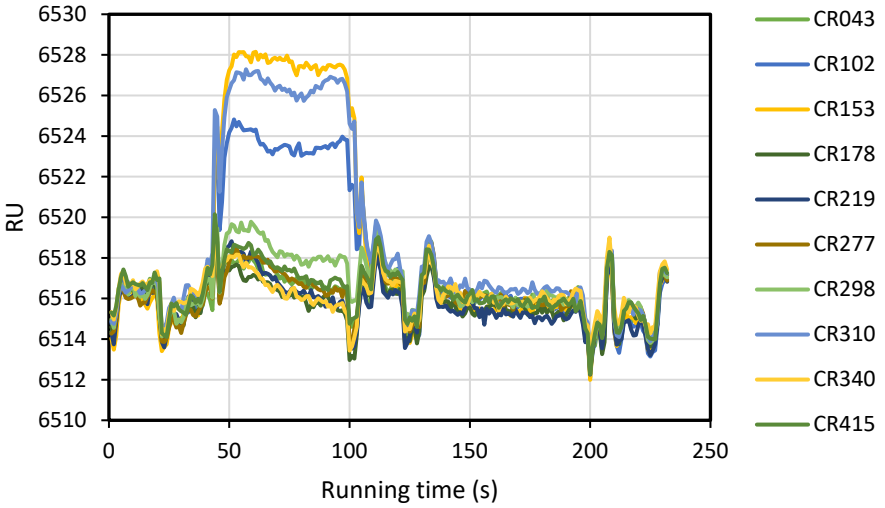


Figure A15 | Exemplary SPR sensograms of different fragments

Biacore Methods

```
DEFINE LOOP cycle_probe
   LPARAM  %pos
         R1A1
         R1A2
         R1A3
         R1A4
         R1A5
         R1A6
         R1A7
         R1A8
         R1A9
         R1A10
         R2D2  !432 cntrl
END

DEFINE APROG Screening
   PARAM %pos
   KEYWORD Postition %pos
   FLOW 10
   1:20 RPOINT Baseline –b
   1:30 QUICKINJECT %pos 10 !Volumen 10uL
   2:25 RPOINT Signal –w0:00:30
   3:30 RPOINT Ende –w0:00:30
   4:00 WASH i             !IFC washen
   5:00 RPOINT cyc_ende
END
MAIN
  RACK 1 THERMO_A
  RACK 2 THERMO_A
  DETECTION 2-1,3-1,4-1
    LOOP cycle_probe STEP
       APROG Screening %pos

    ENDLOOP

END
```

A7 Chemical Shift Perturbation Experiments

Assignment of $^{1}H,^{15}N$-TROSY-HSQC Spectrum of NS2B-NS3^{pro} based on BMRB Entry 11053

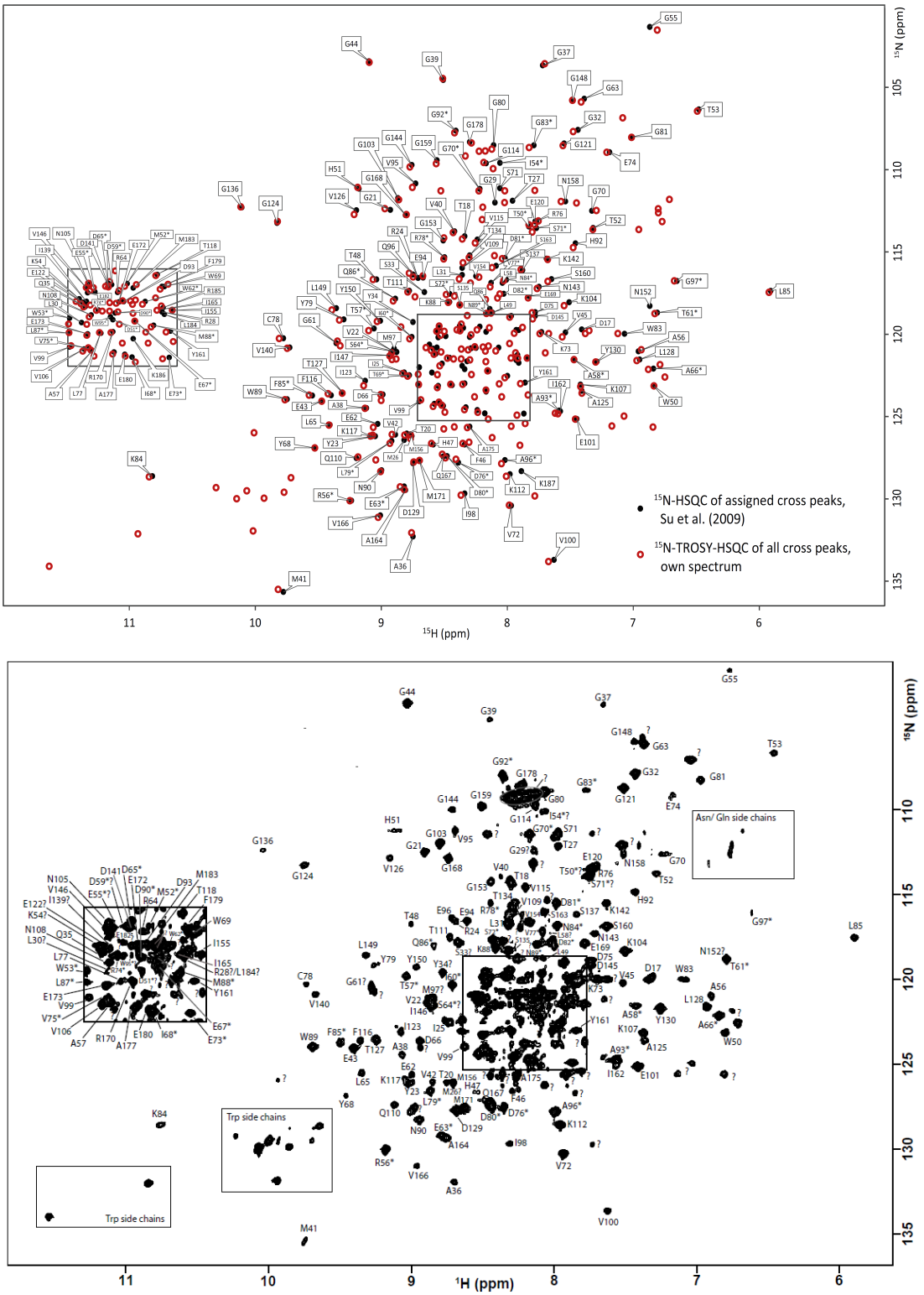


Figure A16 | Assignment of $^{1}H,^{15}N$-TROSY-HSQC spectrum of NS2B-NS3^{pro}
$^{1}H,^{15}N$-TROSY-HSQC spectrum of 130 µM NS2B-NS3^{pro} in the presence of 1.6 mM fragment **A** at 25 °C, in 25 mM sodium phosphate buffer (10% D_2O), pH 7.0. Backbone assignment is based on BMRB entry 11053.

$^{1}H,^{15}N$-TROSY-HSQC spectra of NS2B-NS3^{pro} in Complex with Different Fragments

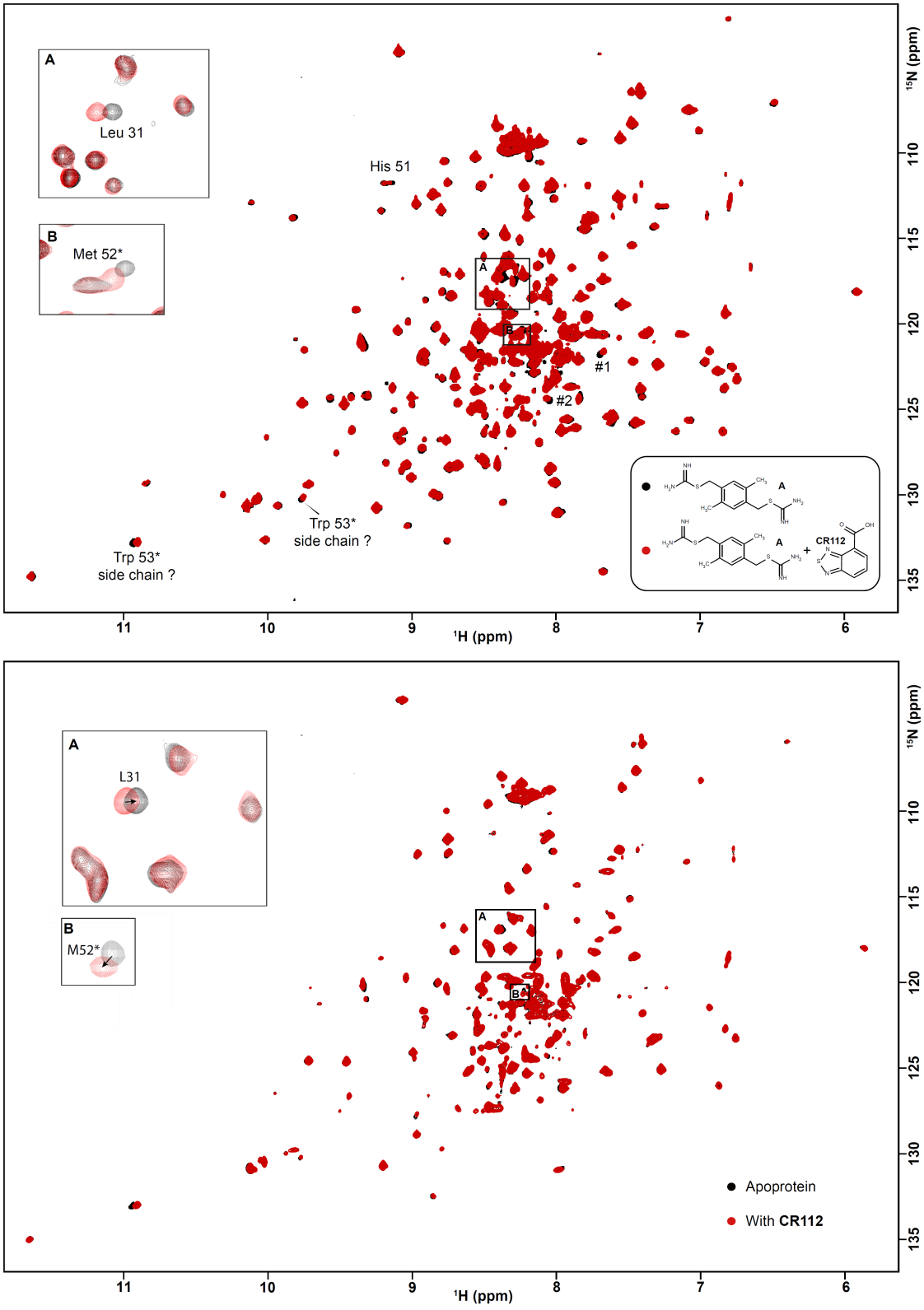


Figure A17 | Chemical shift perturbations of uniformly ^{15}N-labeled NS2B-NS3^{pro} upon binding of CR112
Top) Superposition of $^{1}H,^{15}N$-TROSY-HSQC spectra of 130 µM WNV NS2B-NS3^{pro} saturated with 1.6 mM fragment **A** alone (black) and in complex with both, 1.6 mM fragment **A** and ~10 mM fragment **CR112** (red). Side chain nitrogen atoms are not assigned. However, some of the unassigned NH resonances show significant chemical shift perturbations in the presence of **CR112** (e.g. signal #1 and #2). Bottom) Superimposed $^{1}H,^{15}N$-TROSY-HSQC spectrum of 130 µM NS2B-NS3^{pro} alone (black) and in complex with ~10 mM fragment **CR112** (red).

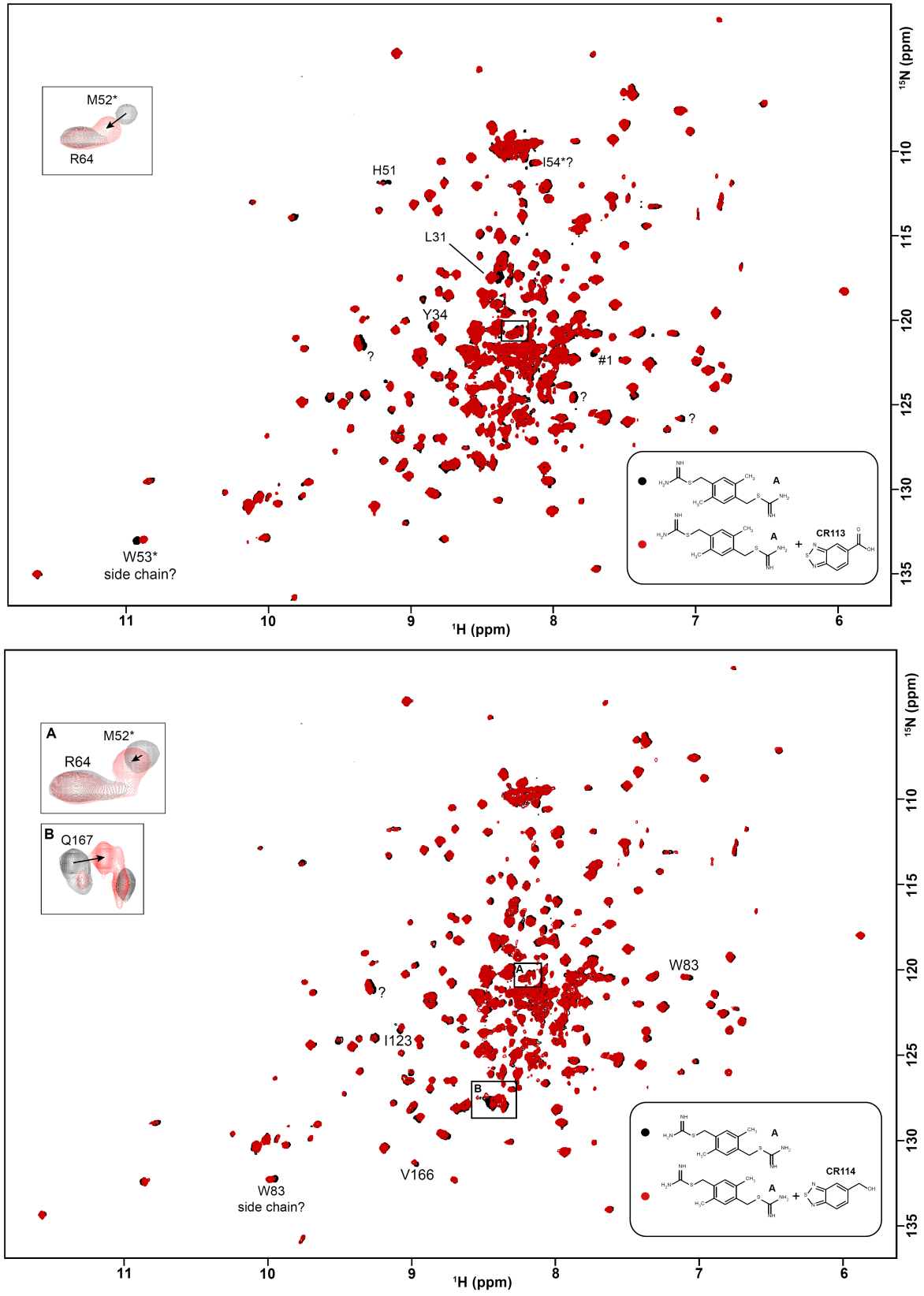


Figure A18 | Chemical shift perturbations of uniformly ^{15}N-labeled NS2B-NS3^{pro} upon binding of CR113 or CR114
Superimposed ^{1}H,^{15}N-TROSY-HSQC spectra of 130 µM WNV NS2B-NS3^{pro} saturated with 1.6 mM fragment **A** alone (black) and in complex with both, 1.6 mM **A** and a specific fragment as depicted (red). For details, see experimental section **2.3.7**.

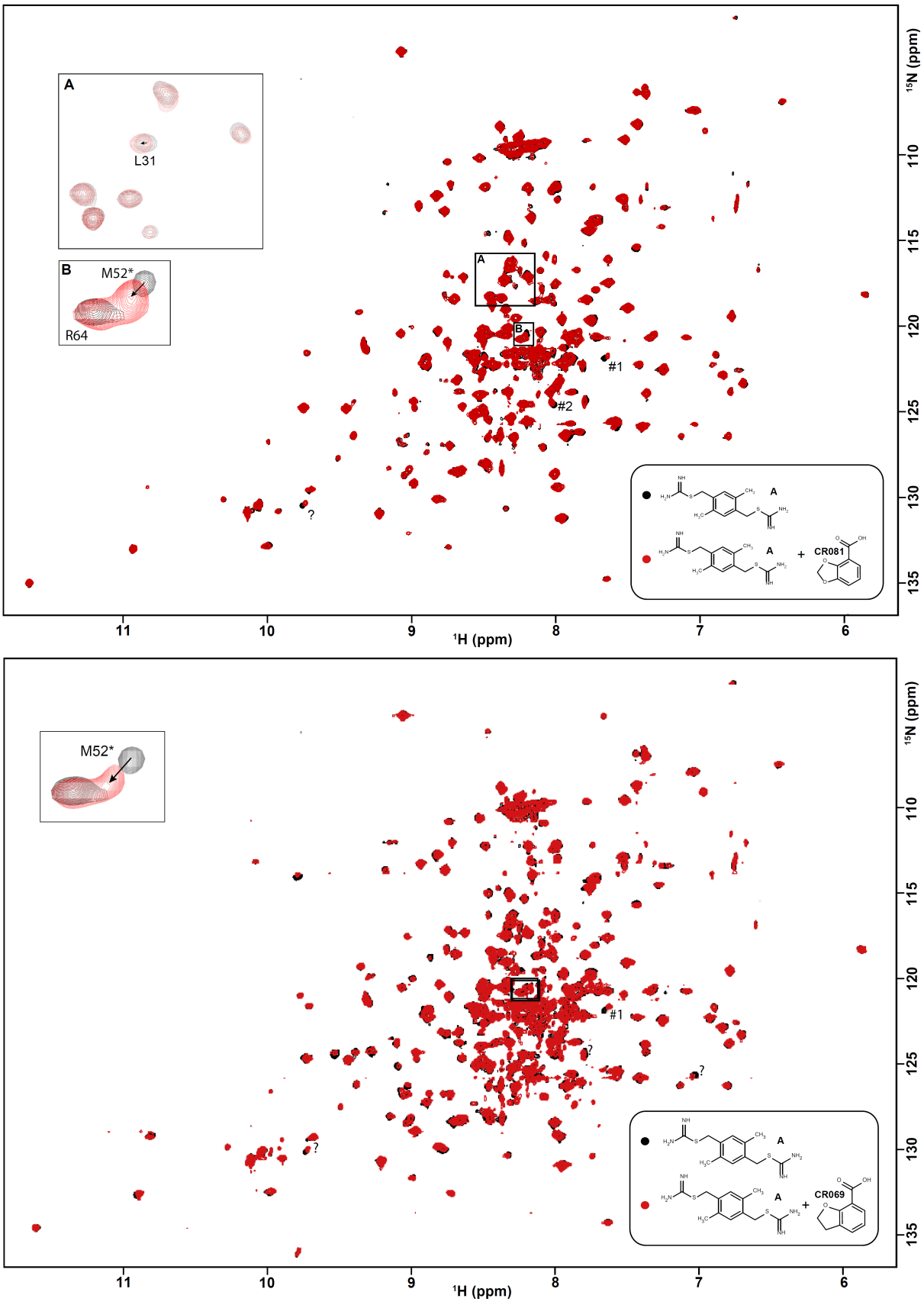


Figure A19 | CR081 and CR069 (Figure A18 continued)

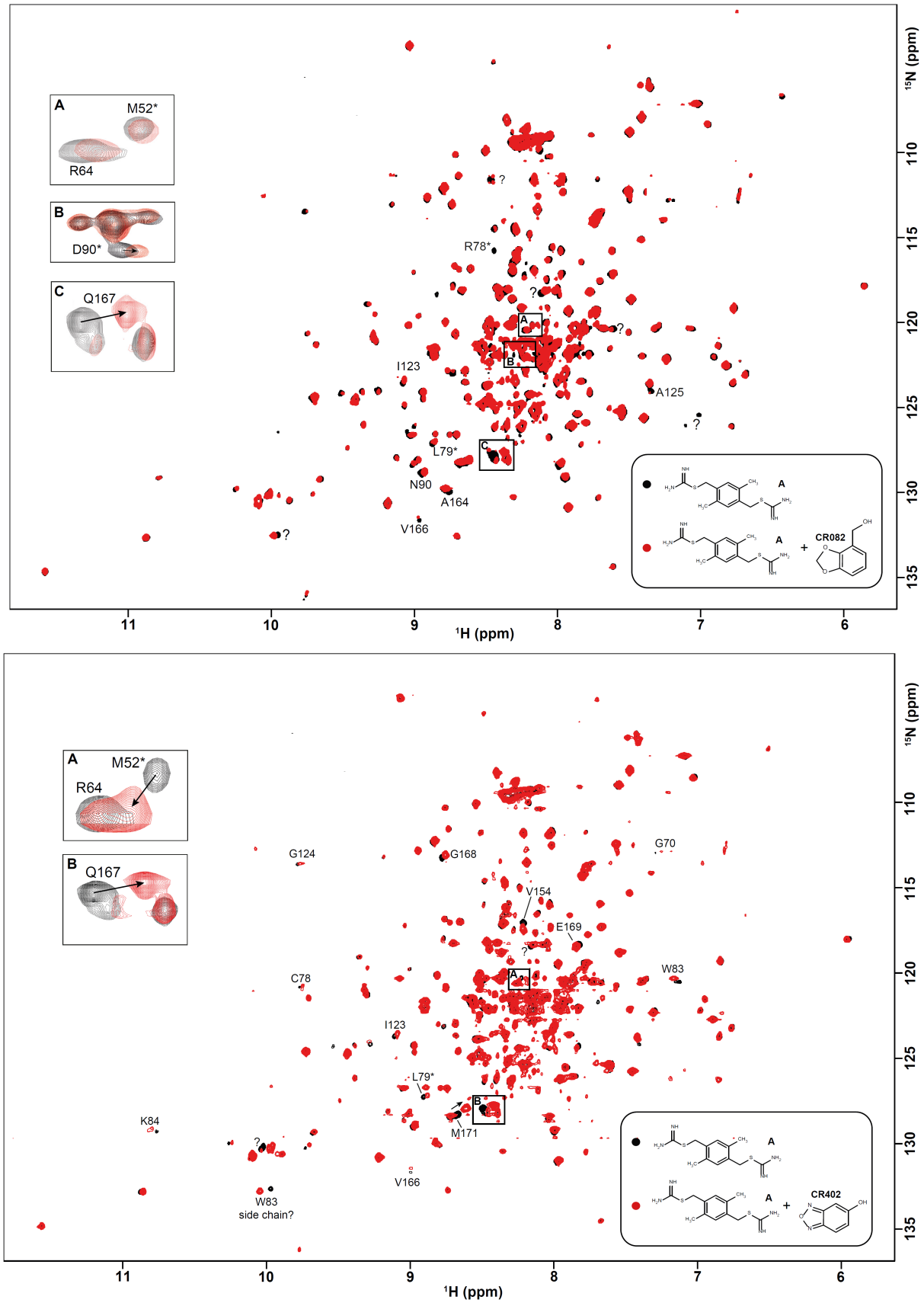


Figure A20 | CR082 and CR402 (Figure A18 continued)

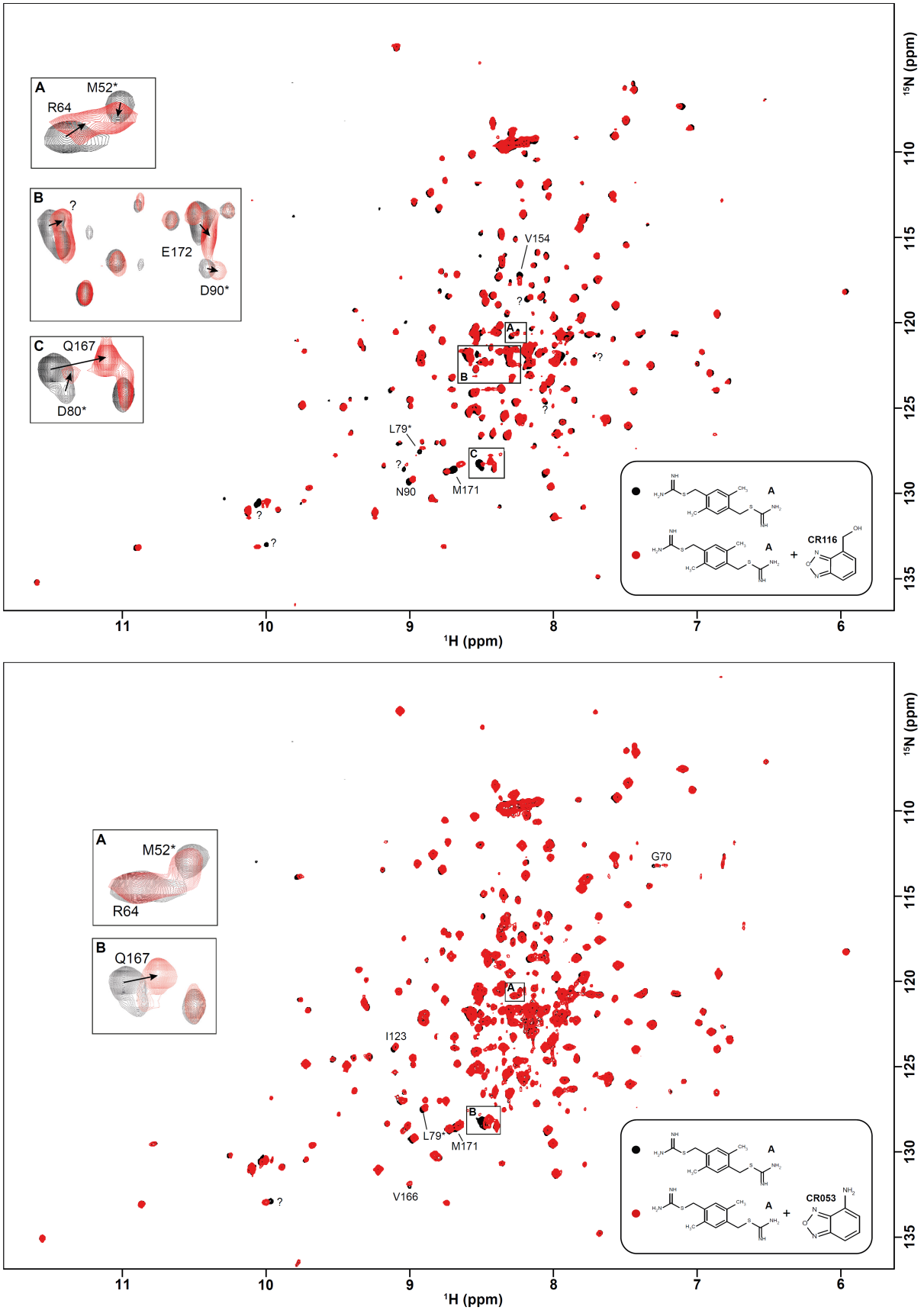


Figure A21 | CR116 and CR053 (Figure A18 continued)

Assignment of U-[^{2}H,^{15}N]-labeled Saga P Dimer and TROSY-HSQC Spectra

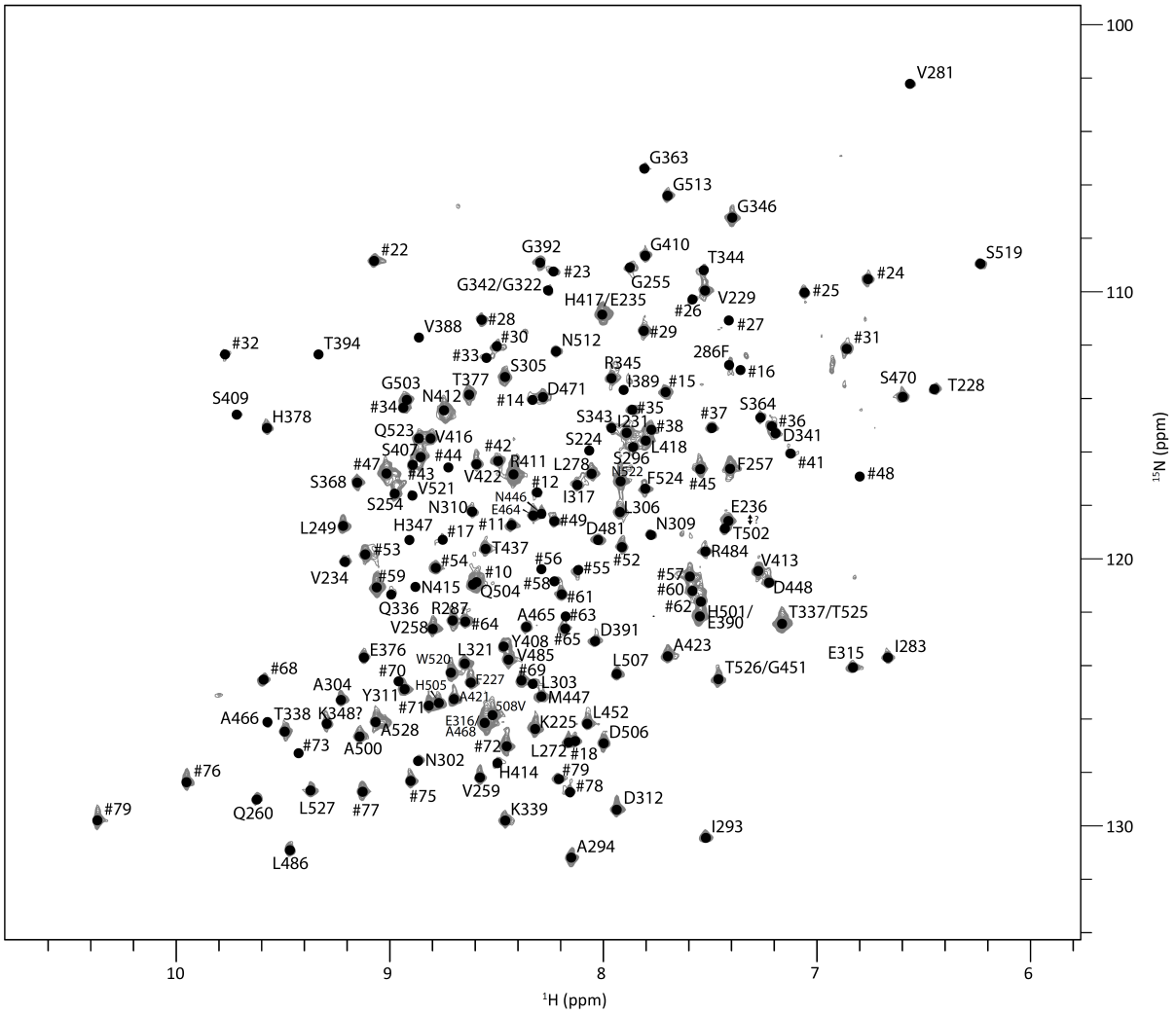


Figure A22 | Assignment of uniformly [^{2}H,^{15}N]-labeled Saga P dimer

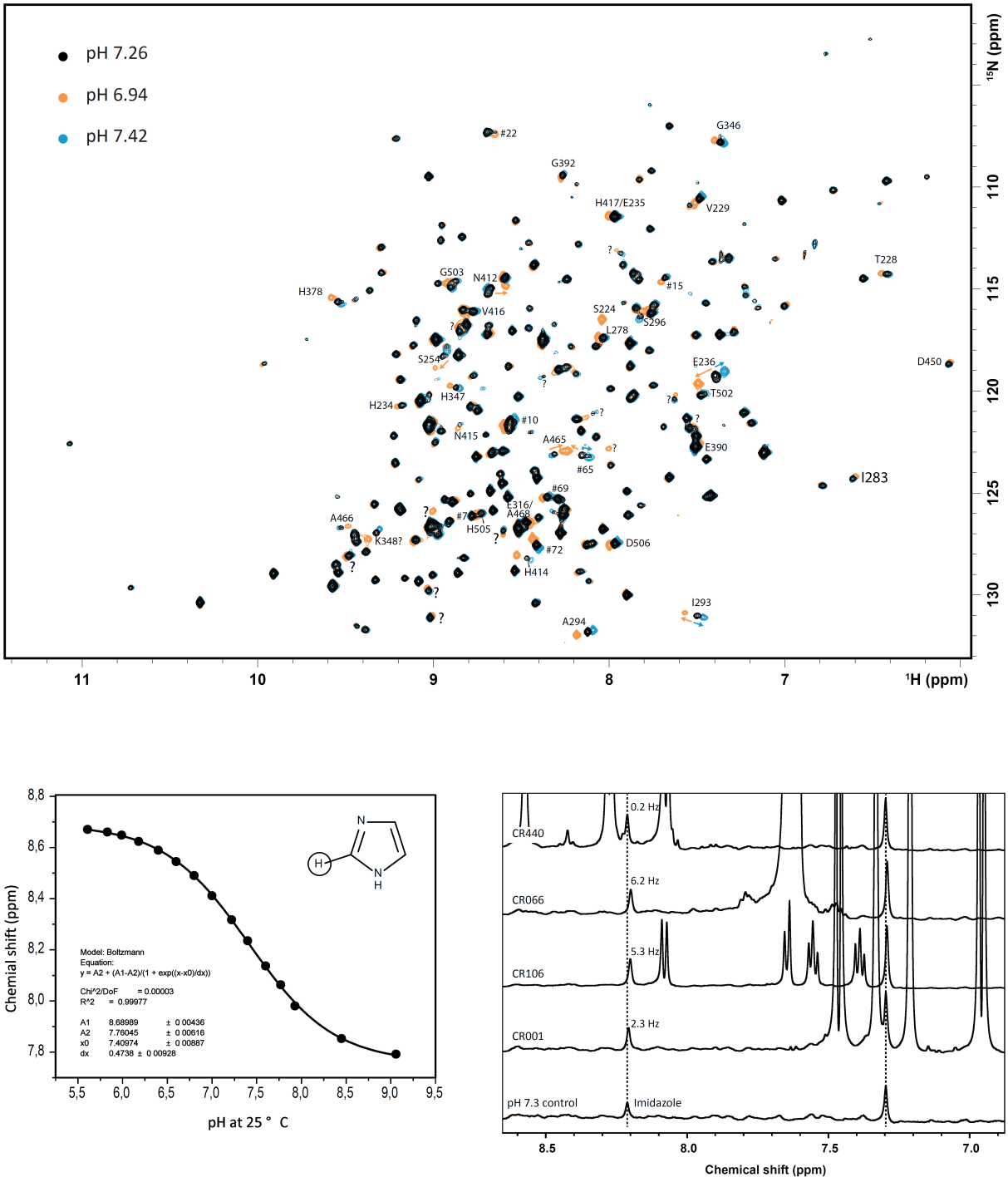


Figure A23 | Chemical shift perturbations upon pH differences
The pH titration of isotopically labeled Saga P dimers was conducted by Robert Creutznacher (2017, unpublished data) in 100 mM sodium phosphate buffer supplemented with and 8% D_2O. Data of the pH titration of imidazole (lower panel) were kindly provided by Christph Müller-Hermes. Exemplary 1D NMR spectra of pH control via imidazole are given for the fragments **CR001**, **CR106**, **CR066**, and **CR440**.

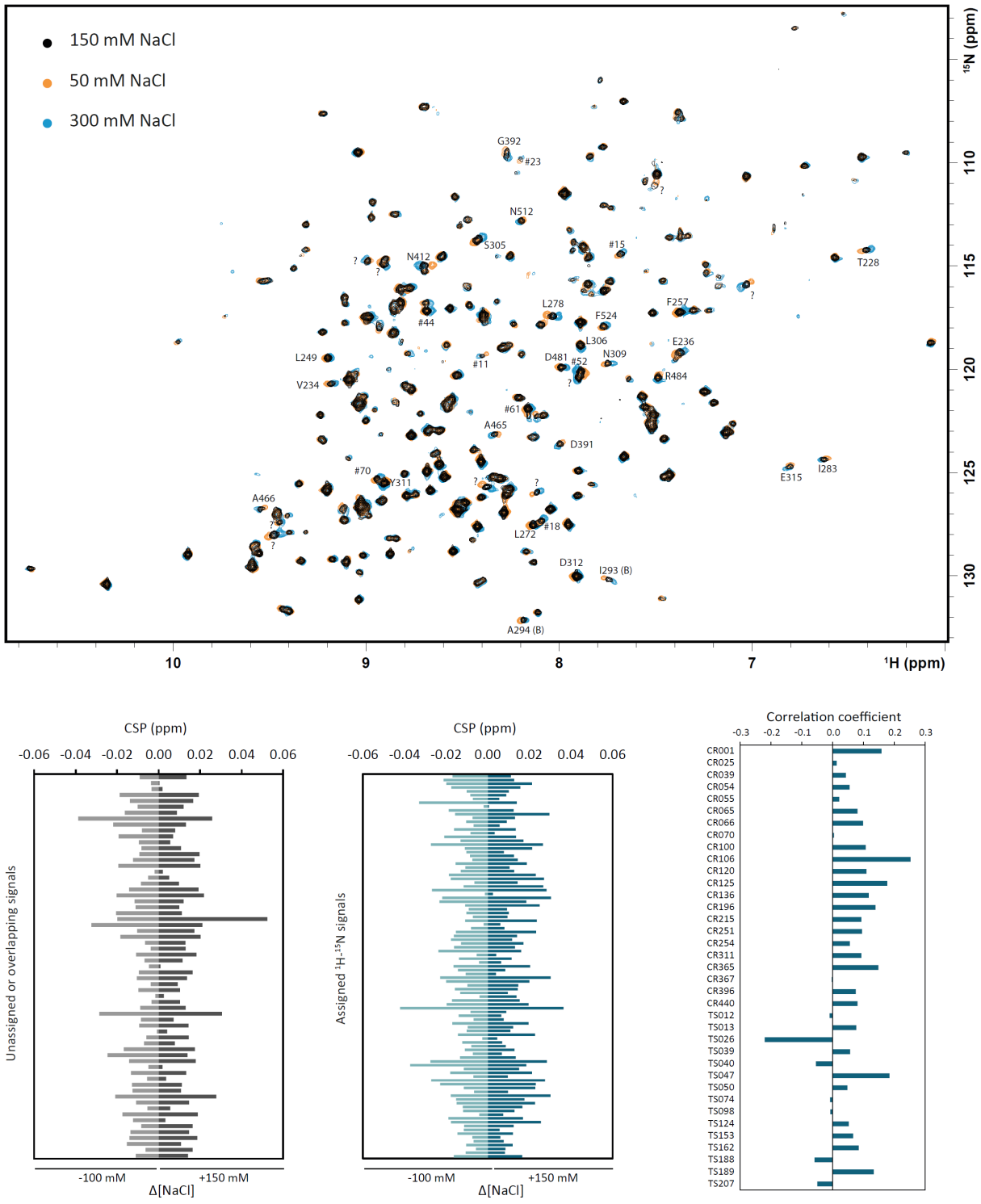


Figure A24 | Chemical shift perturbation upon different sodium chloride concentration
Row data of NaCl titration of isotopically labeled Saga P dimers were kindly provided by Robert Creutznacher (2017, unpublished data). $^{1}H,^{15}N$-TROSY-HSQC spectra of ~100 µM Saga P dimers were recorded in 20 mM sodium phosphate at increasing amount of NaCl. On the right side, Pearson's correlation coffecients are given, displaying the correlation of CSP profiles of fragments and the CSP profile derived from increasing the salt concentration from 150 to 300 mM NaCl. Due to the small coefficients (< |0.3|) ionic-strength-dependent CSP artefacts are highly unlikely.

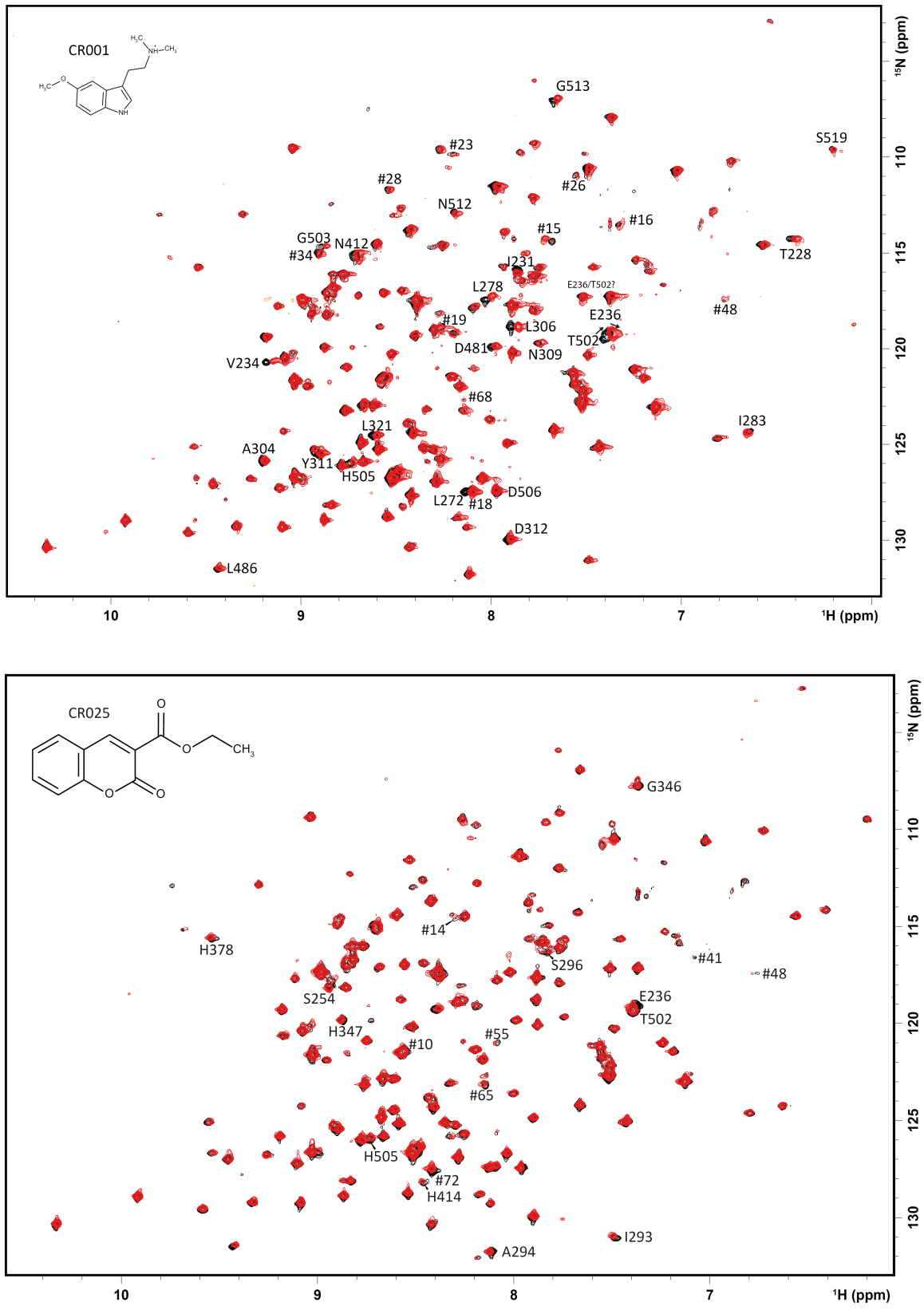


Figure A25 | Chemical shift perturbations of uniformly ^{15}N-labeled Saga P dimers upon binding of CR001 and CR025
Superimposed ^{1}H,^{15}N-TROSY-HSQC spectra of 50 – 60 µM Saga P dimers saturated with a specific fragment as depicted (red). For details, see experimental section **3.3.1.**

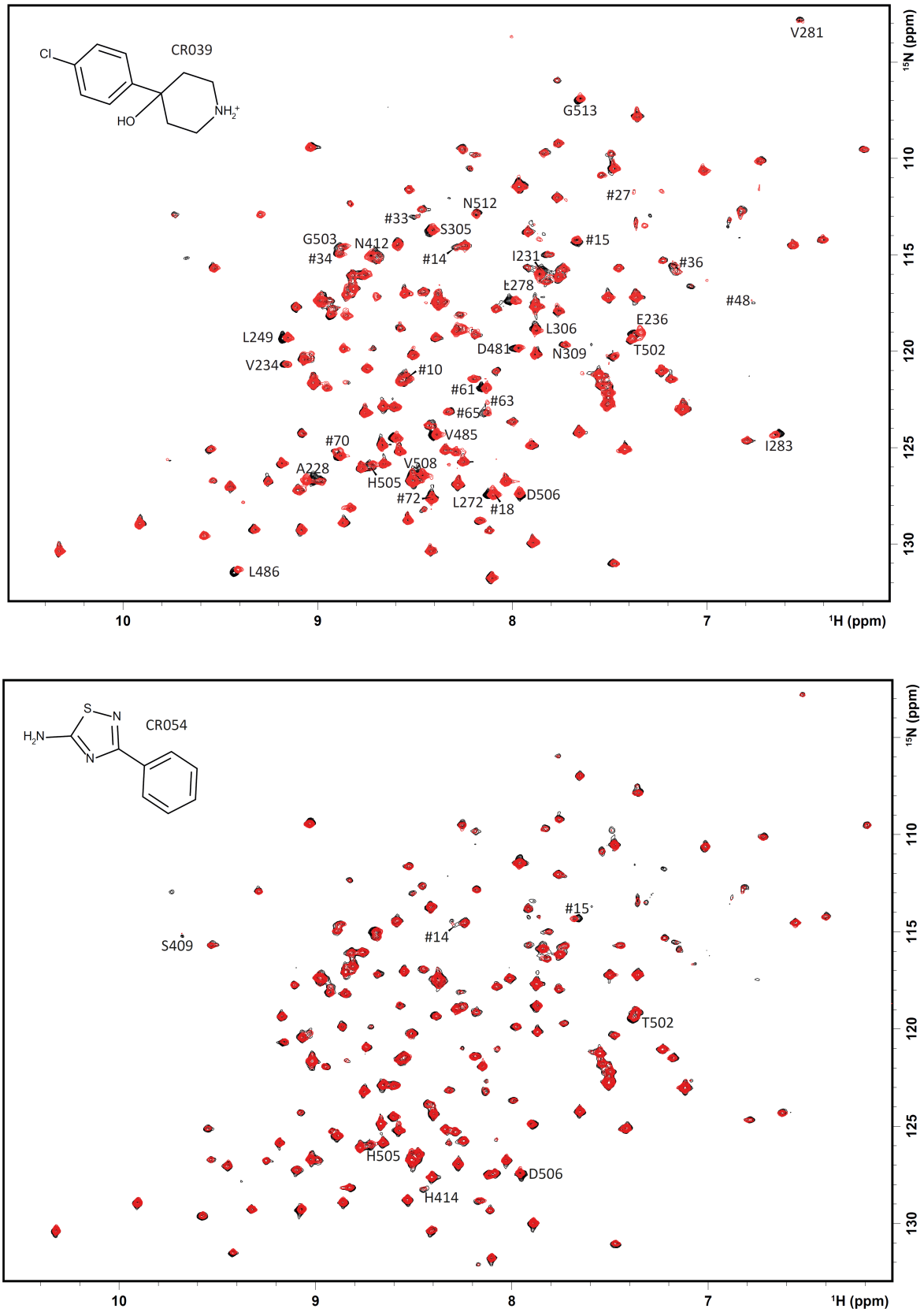


Figure A26 | CR039 and CR054 (Figure A25 continued)

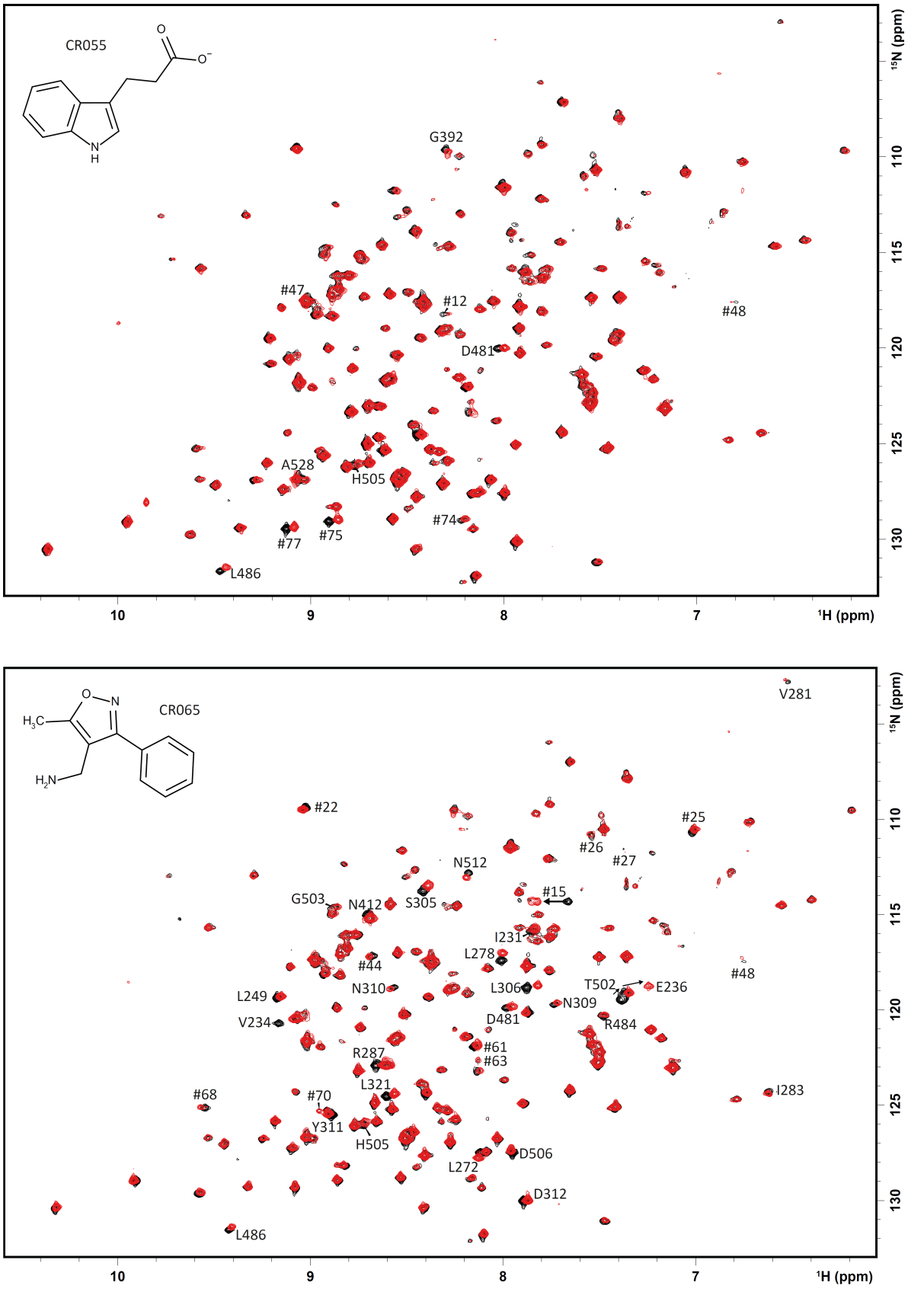
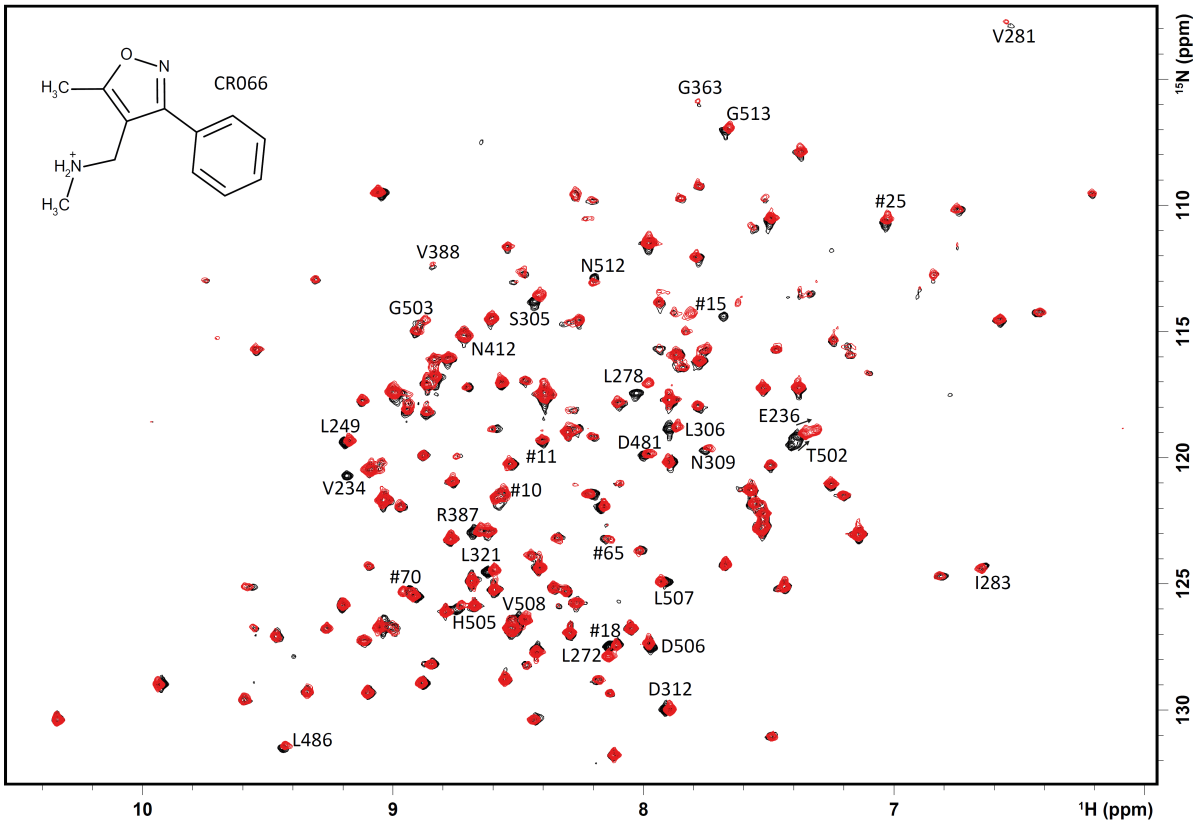


Figure A27 | CR055 and CR065 (Figure A25 continued)



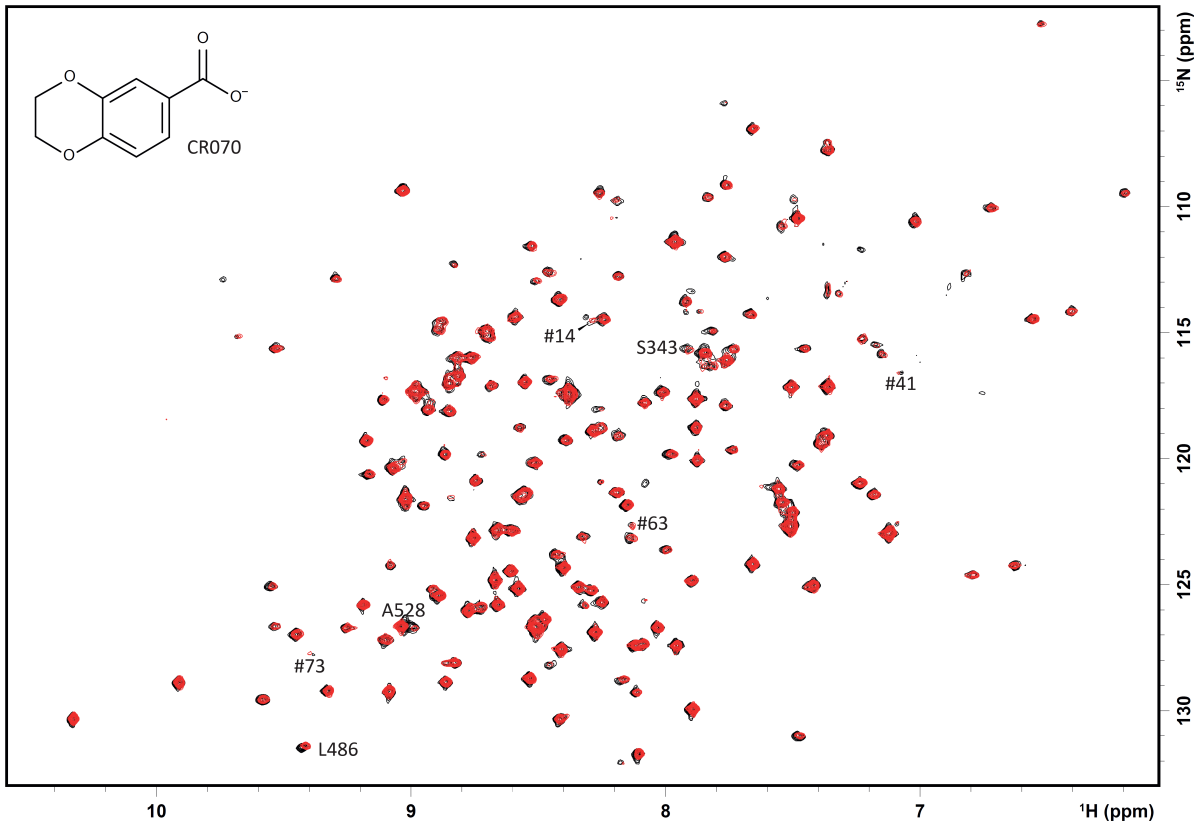
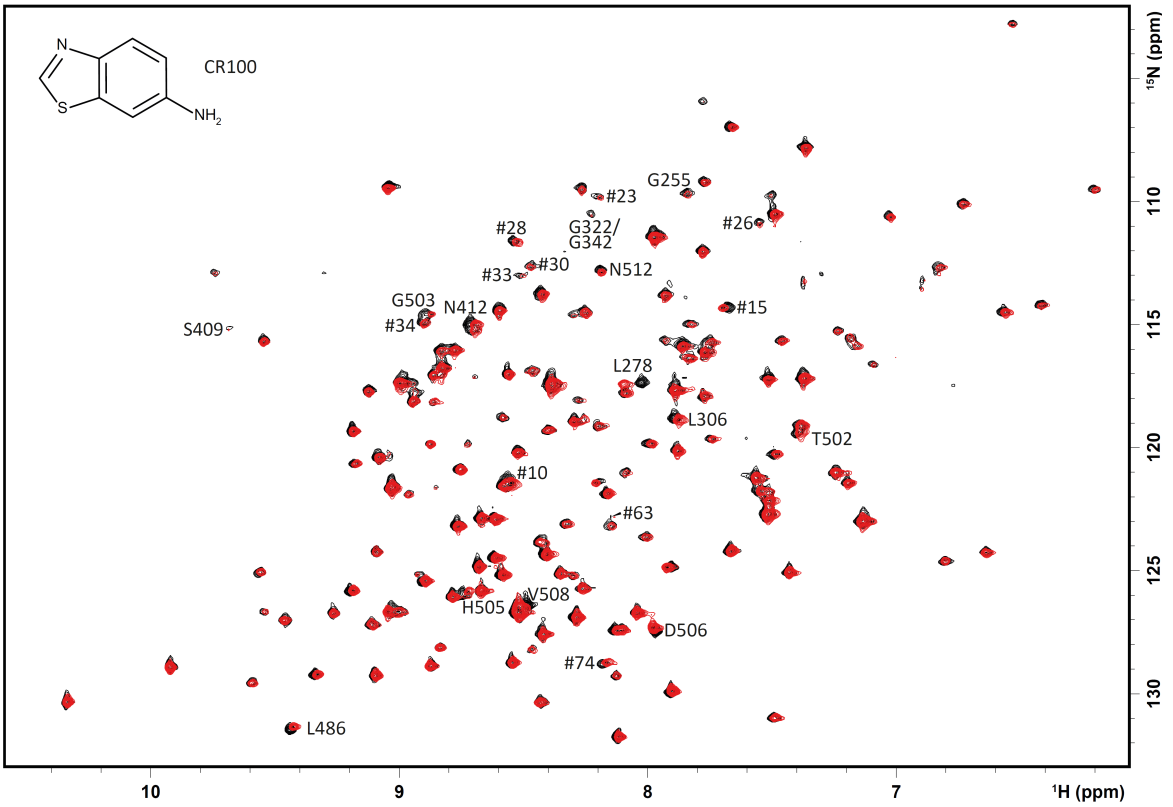


Figure A28 | CR066 and CR070 (Figure A25 continued)



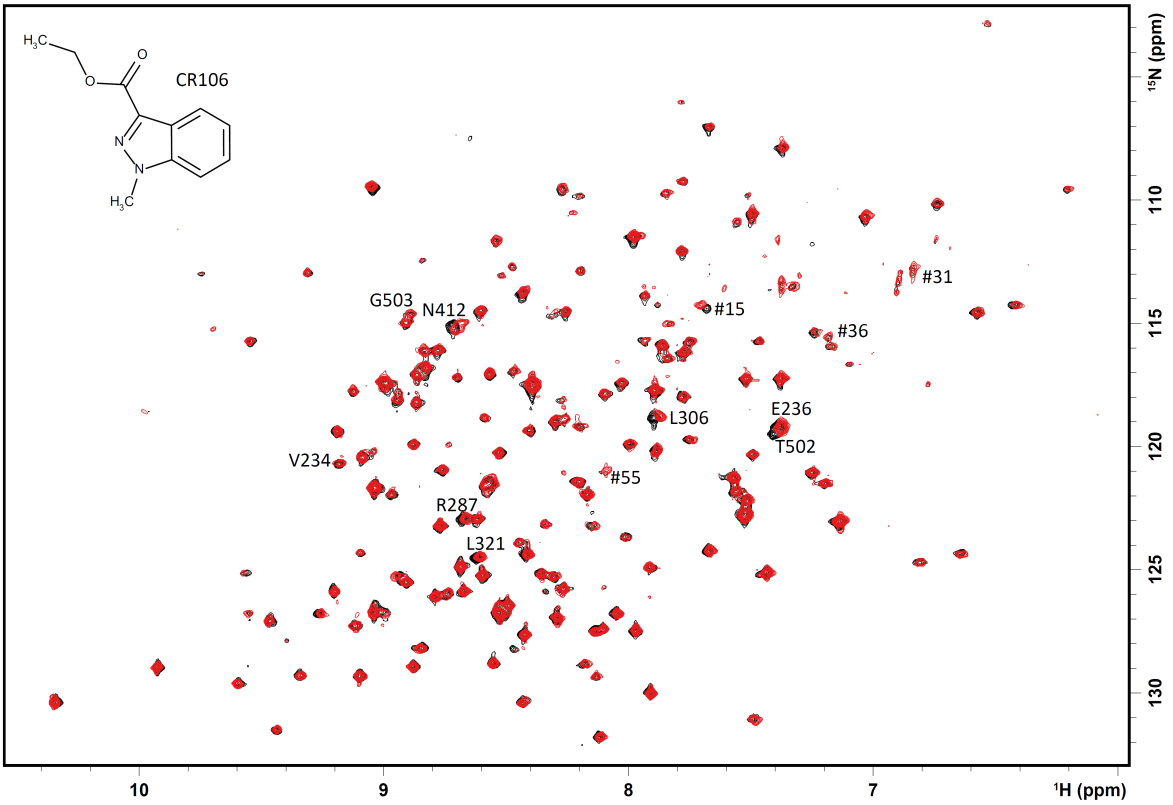


Figure A29 | CR100 and CR106 (Figure A25 continued)

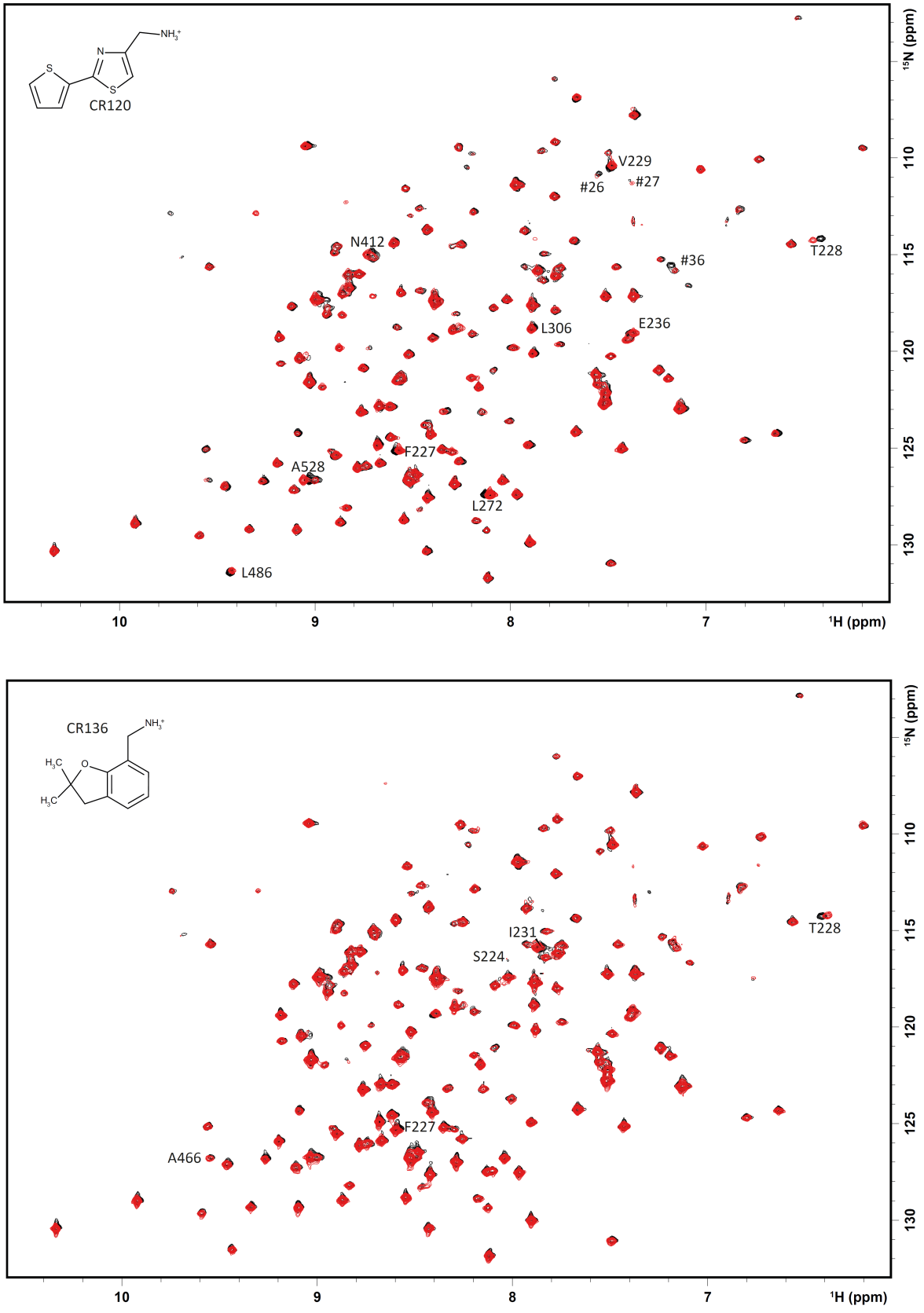
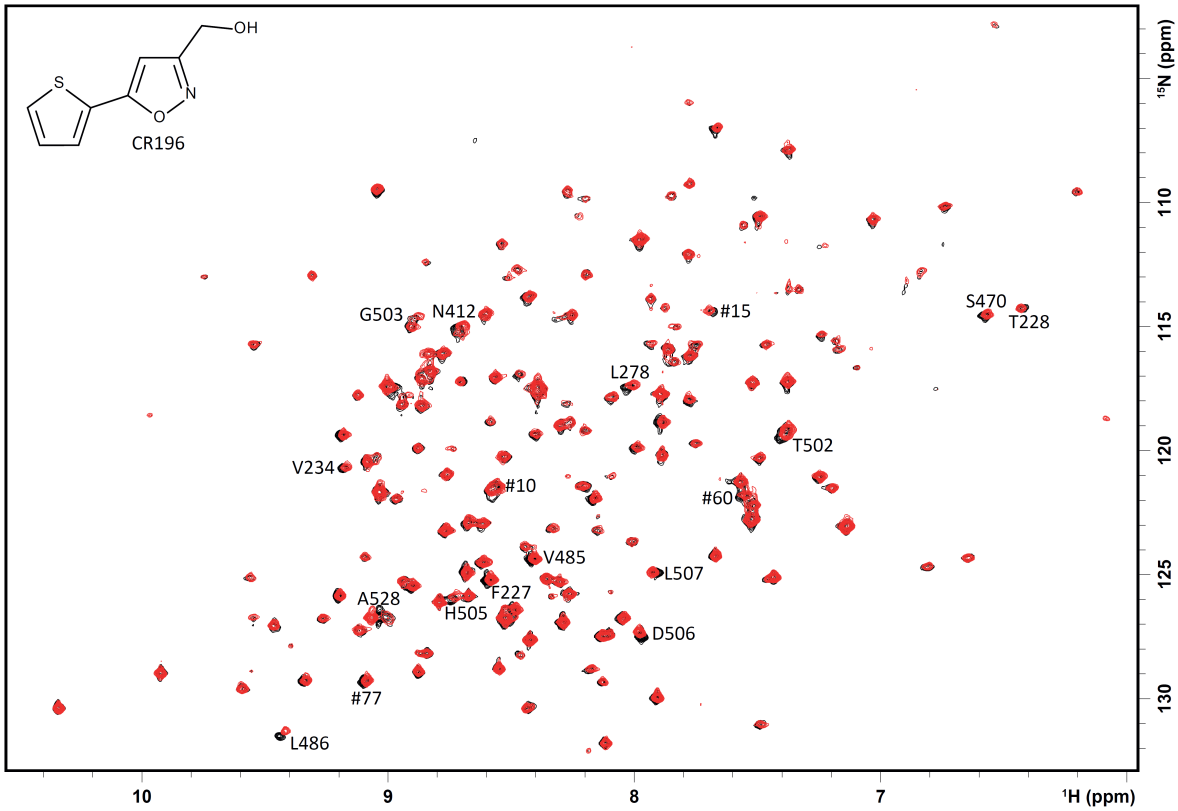


Figure A30 | CR120 and CR136 (Figure A25 continued)



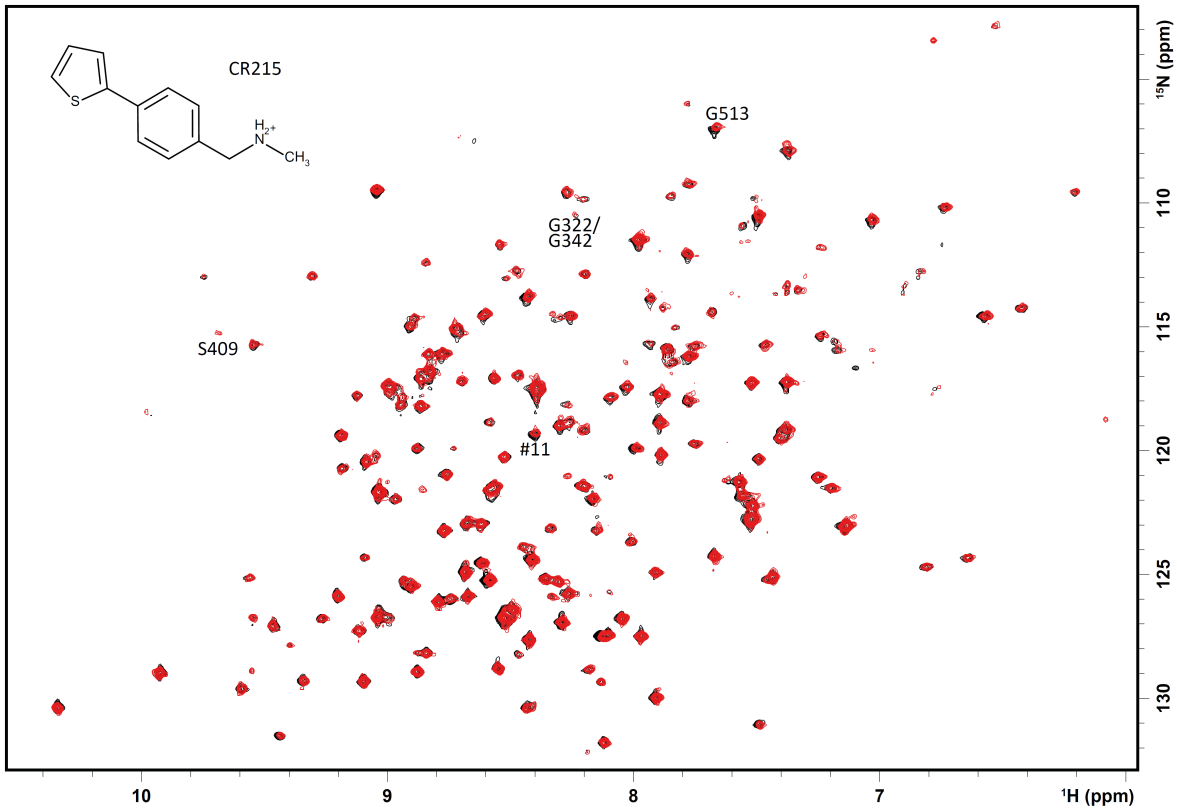


Figure A31 | CR196 and CR215 (Figure A25 continued)

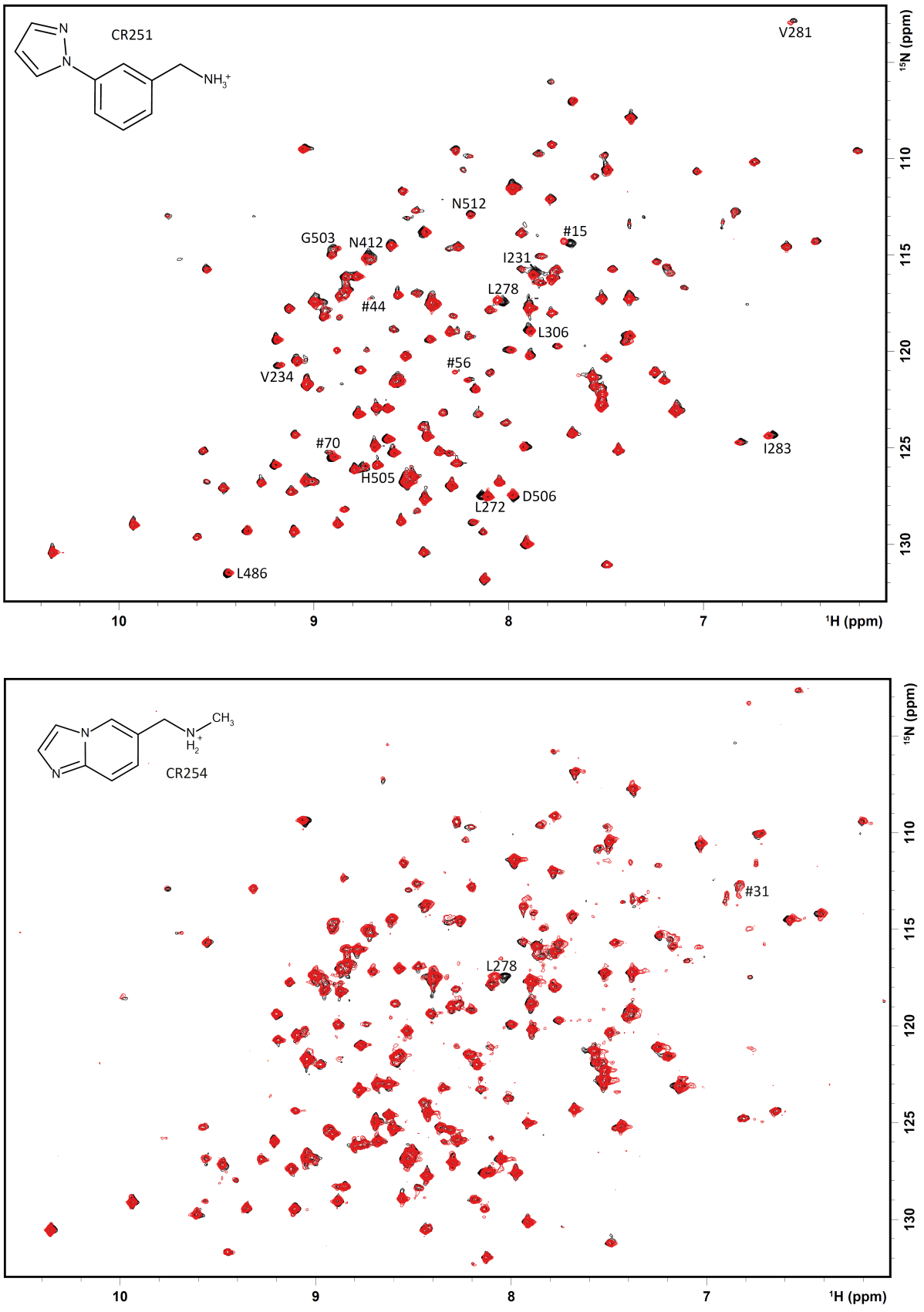


Figure A32 | CR251 and CR254 (Figure A25 continued)

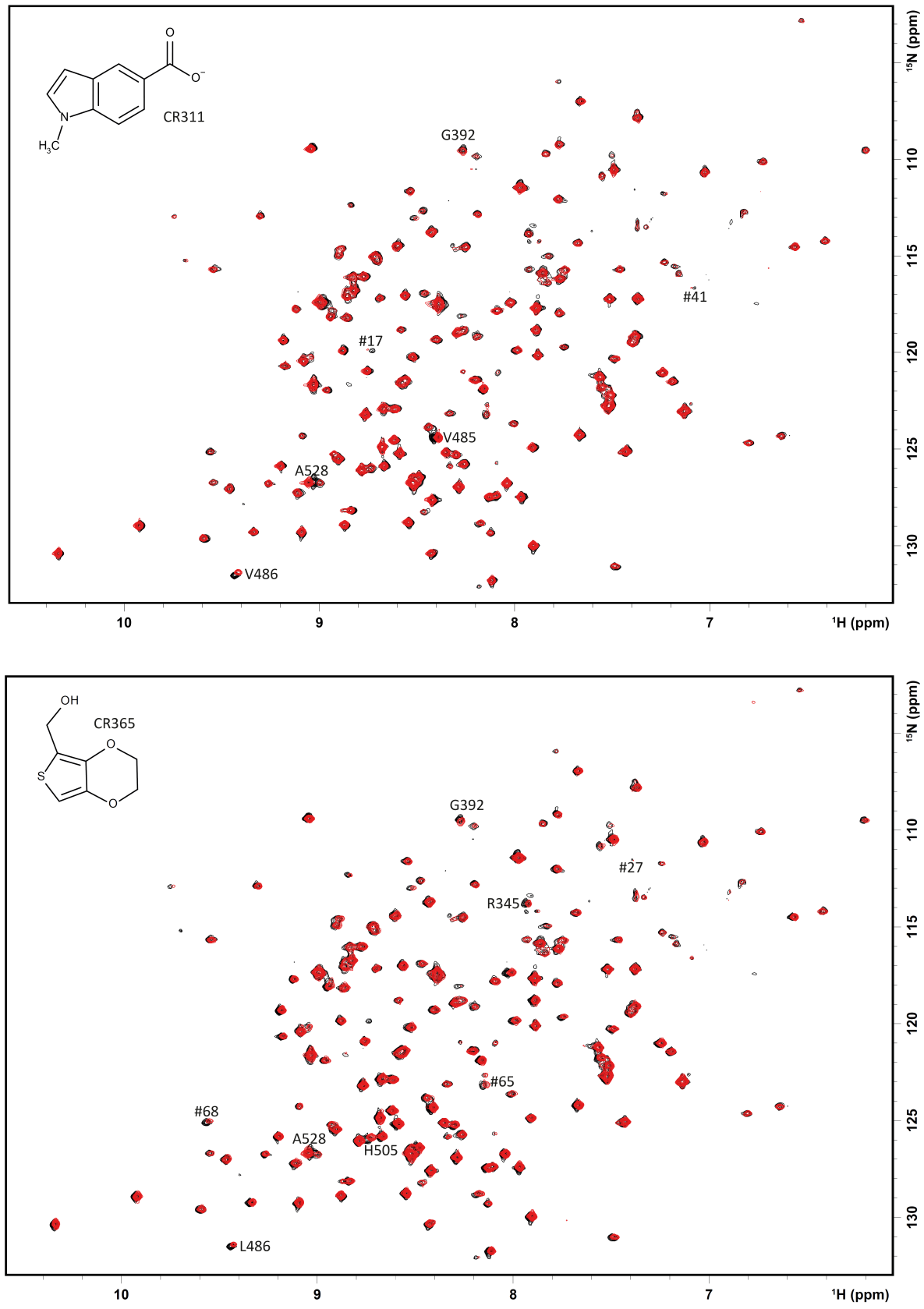


Figure A33 | CR311 and CR365 (Figure A25 continued)

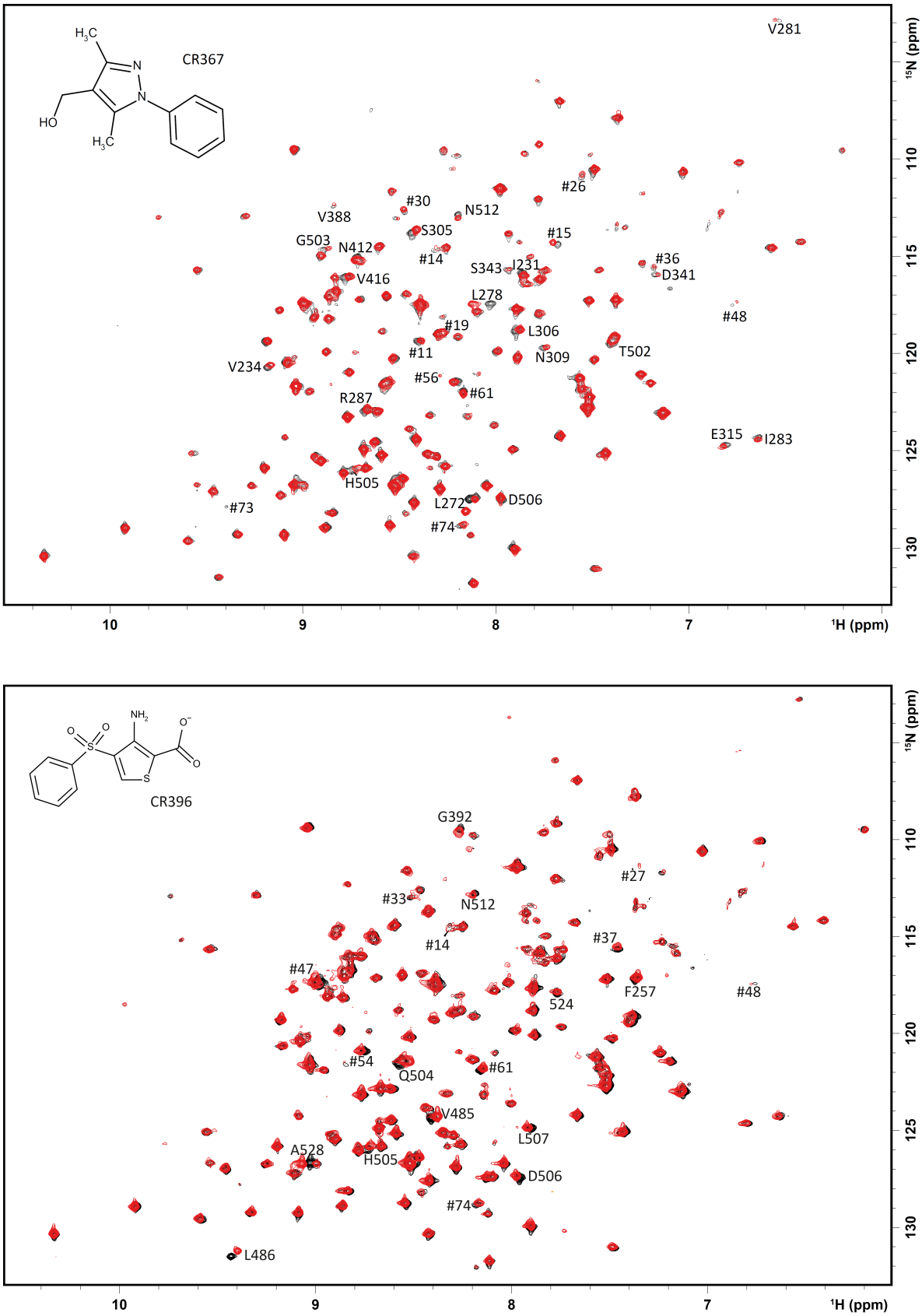


Figure A34 | CR367 and CR396 (Figure A25 continued)

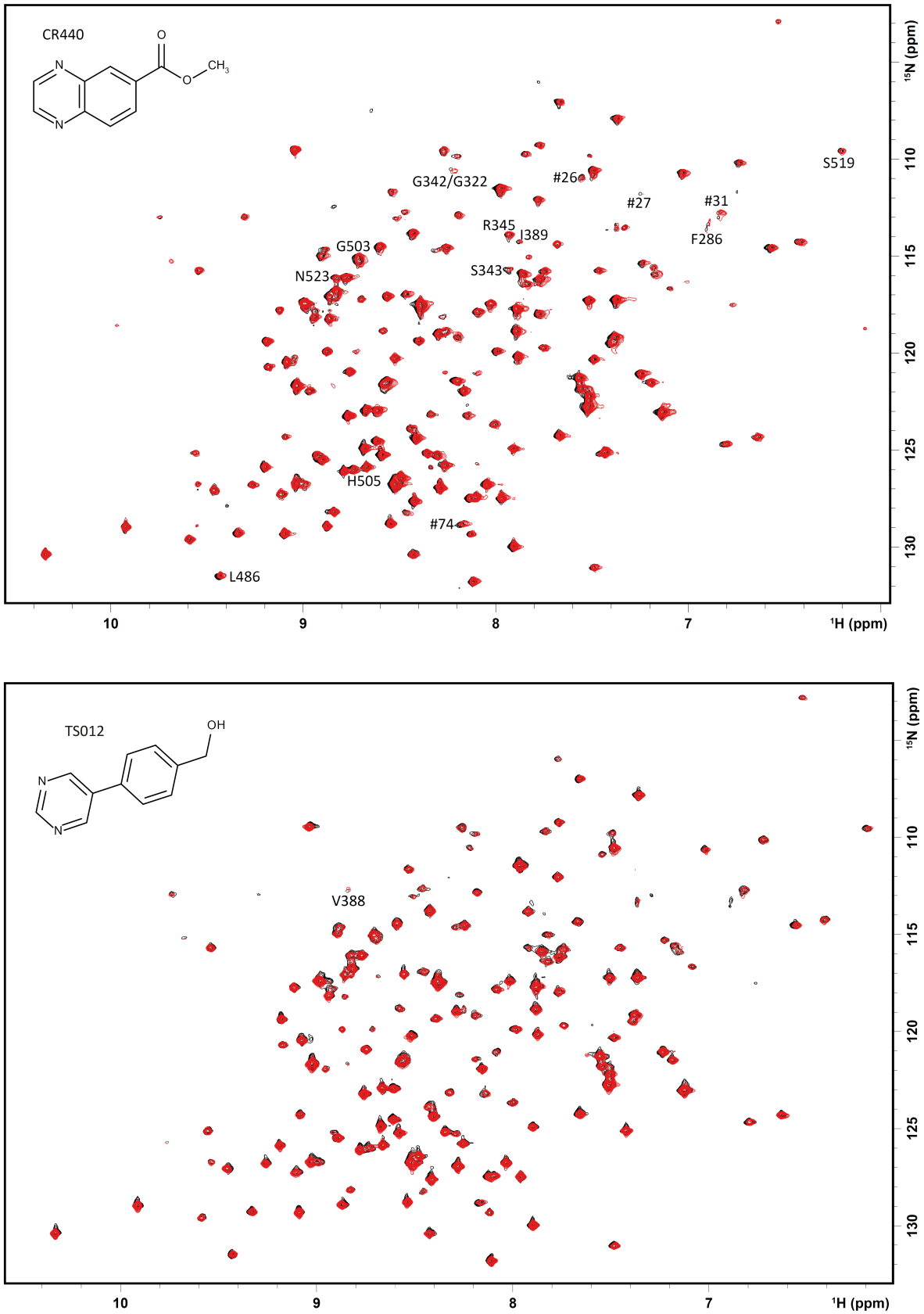


Figure A35 | CR440 and TS012 (Figure A25 continued)

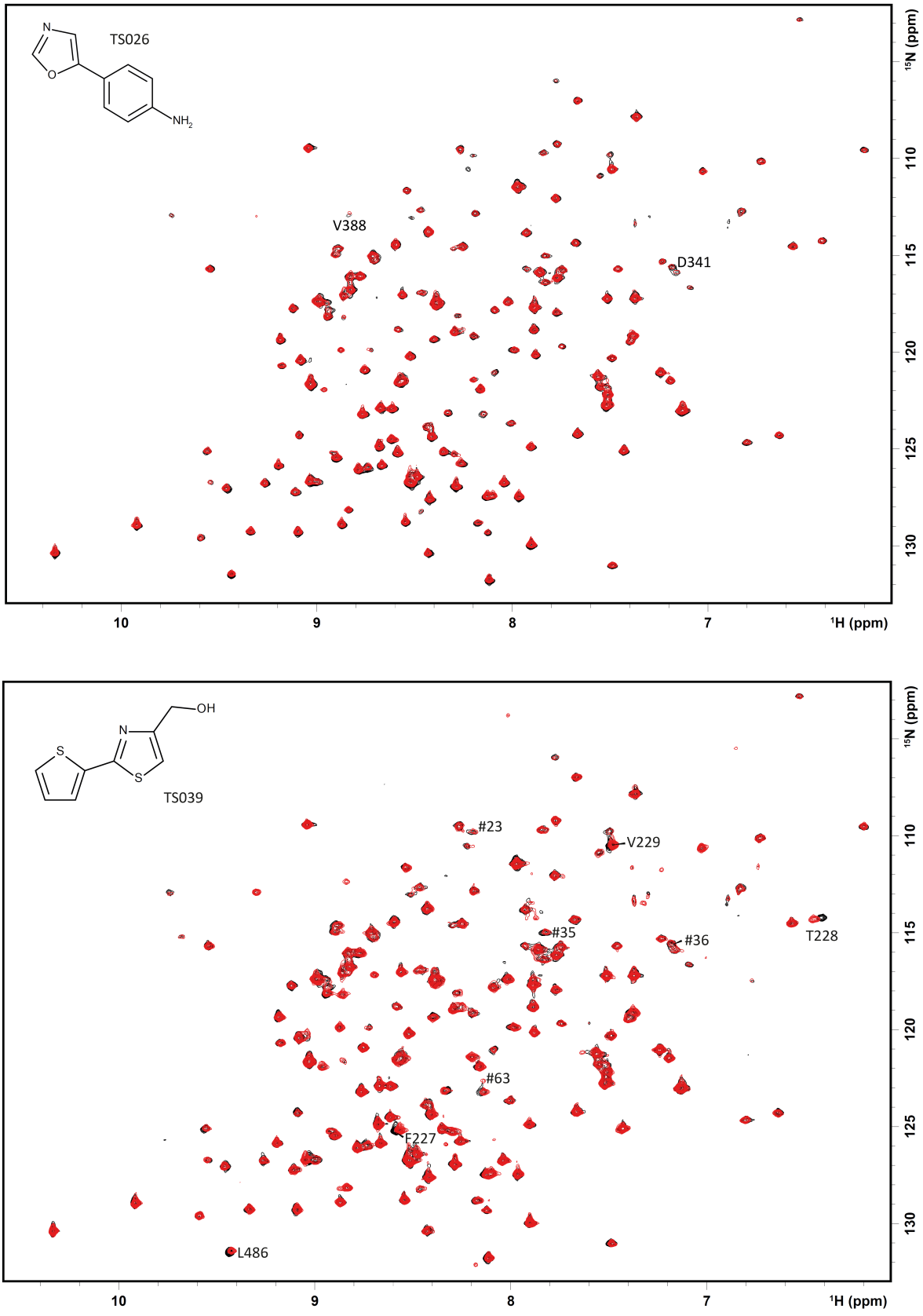
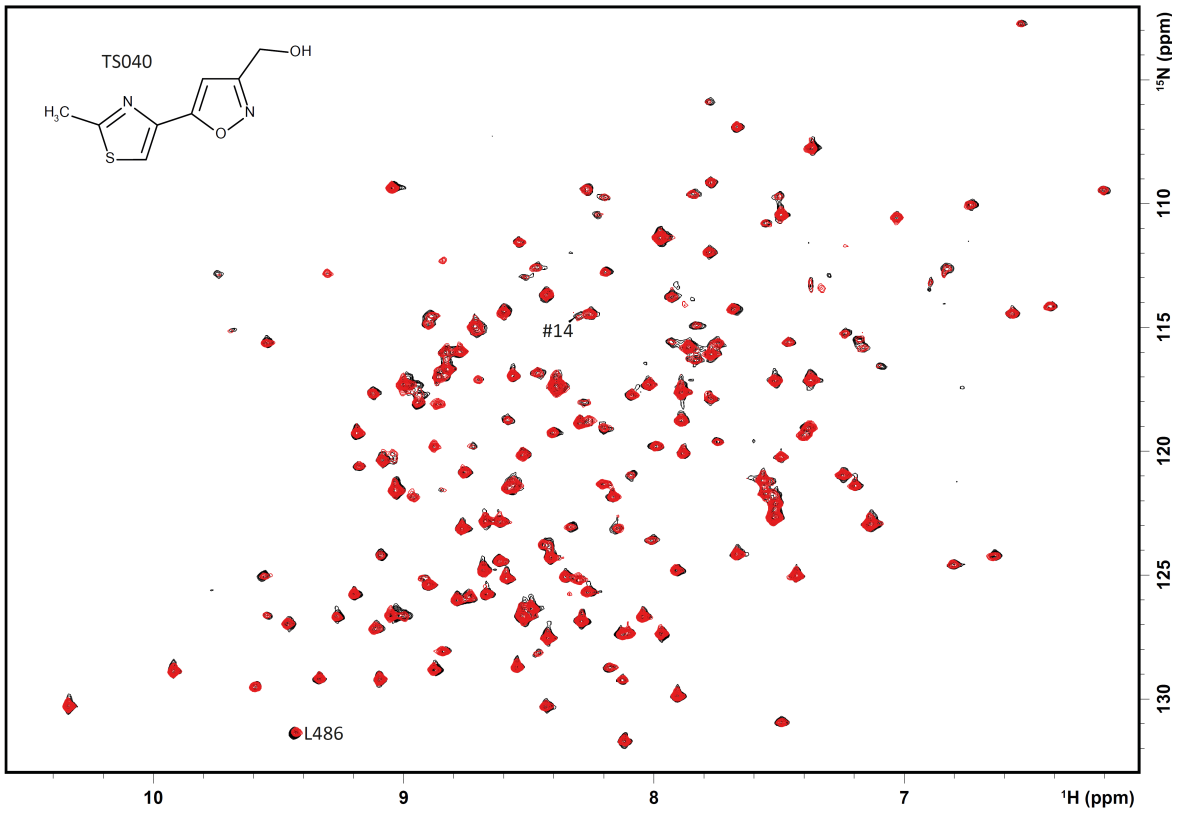


Figure A36 | TS026 and TS039 (Figure A25 continued)



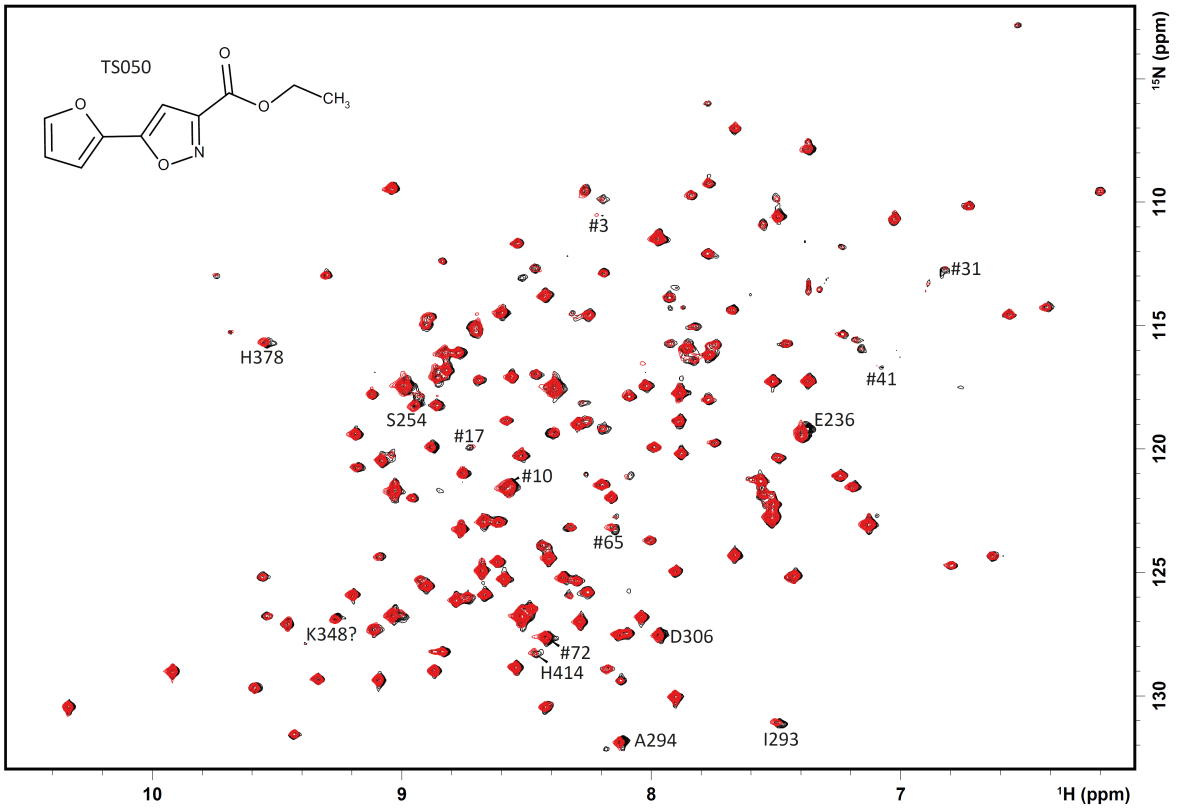


Figure A37 | TS040 and TS050 (Figure A25 continued)

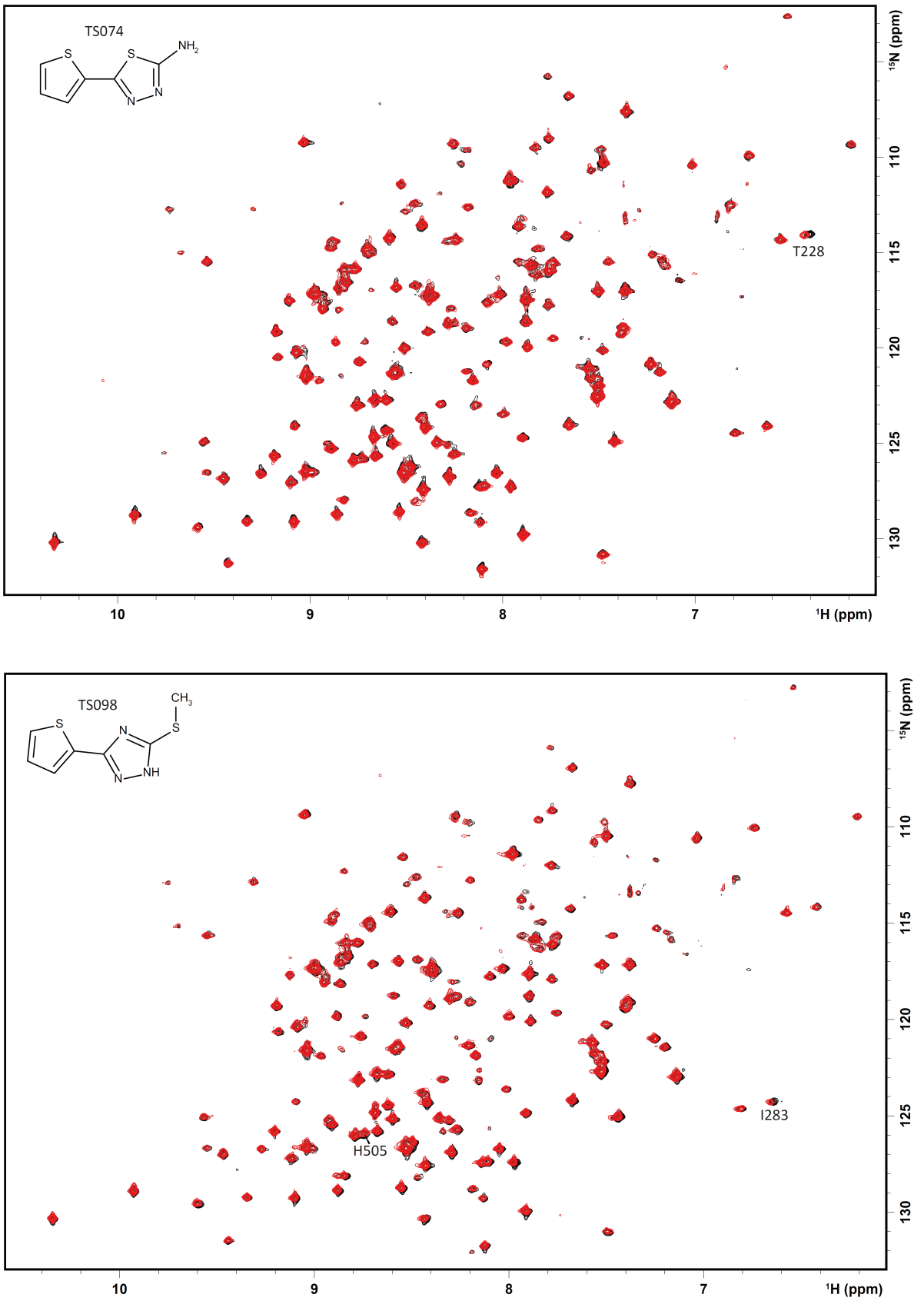
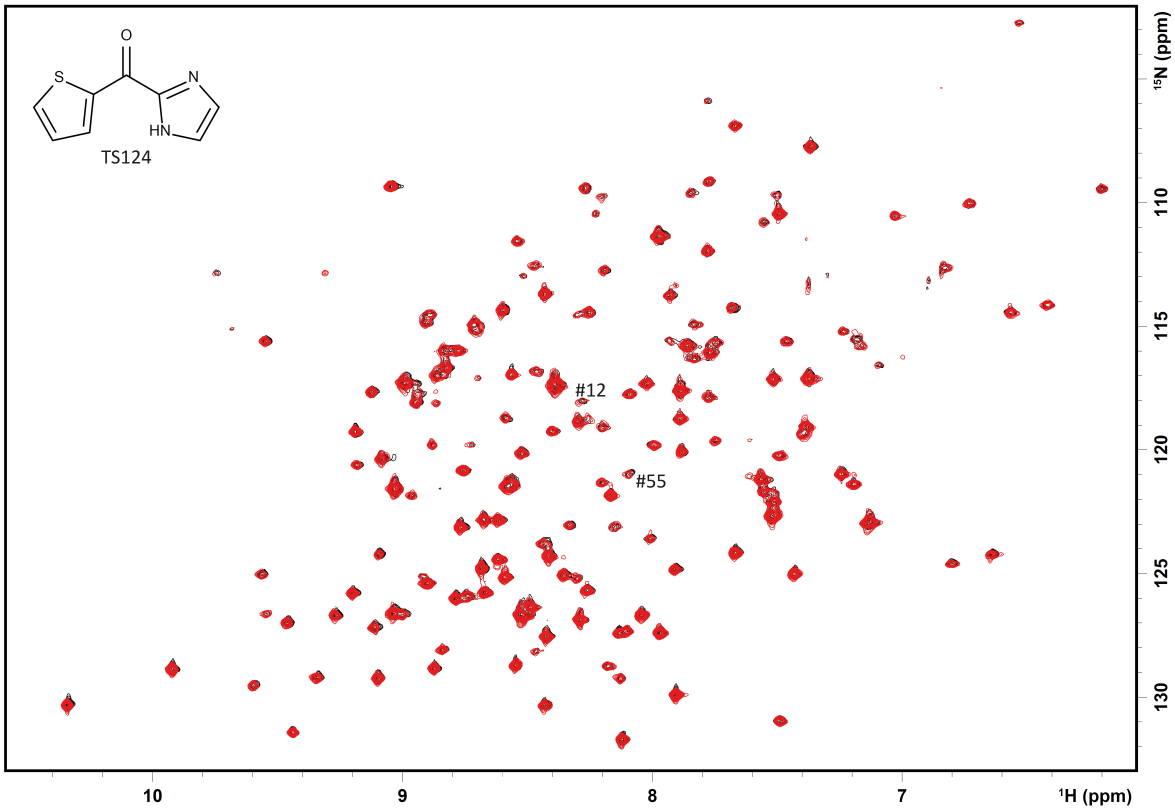


Figure A38 | TS074 and TS098 (Figure A25 continued)



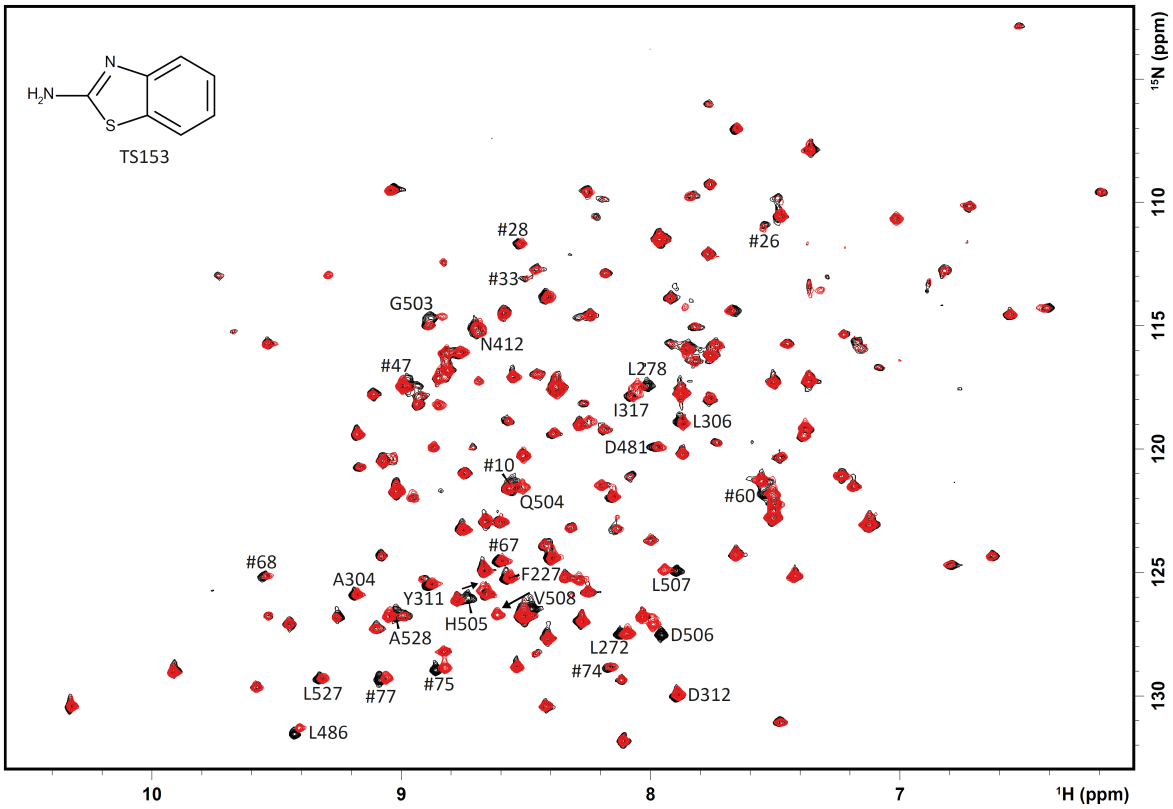


Figure A39 | TS124 and TS153 (Figure A25 continued)

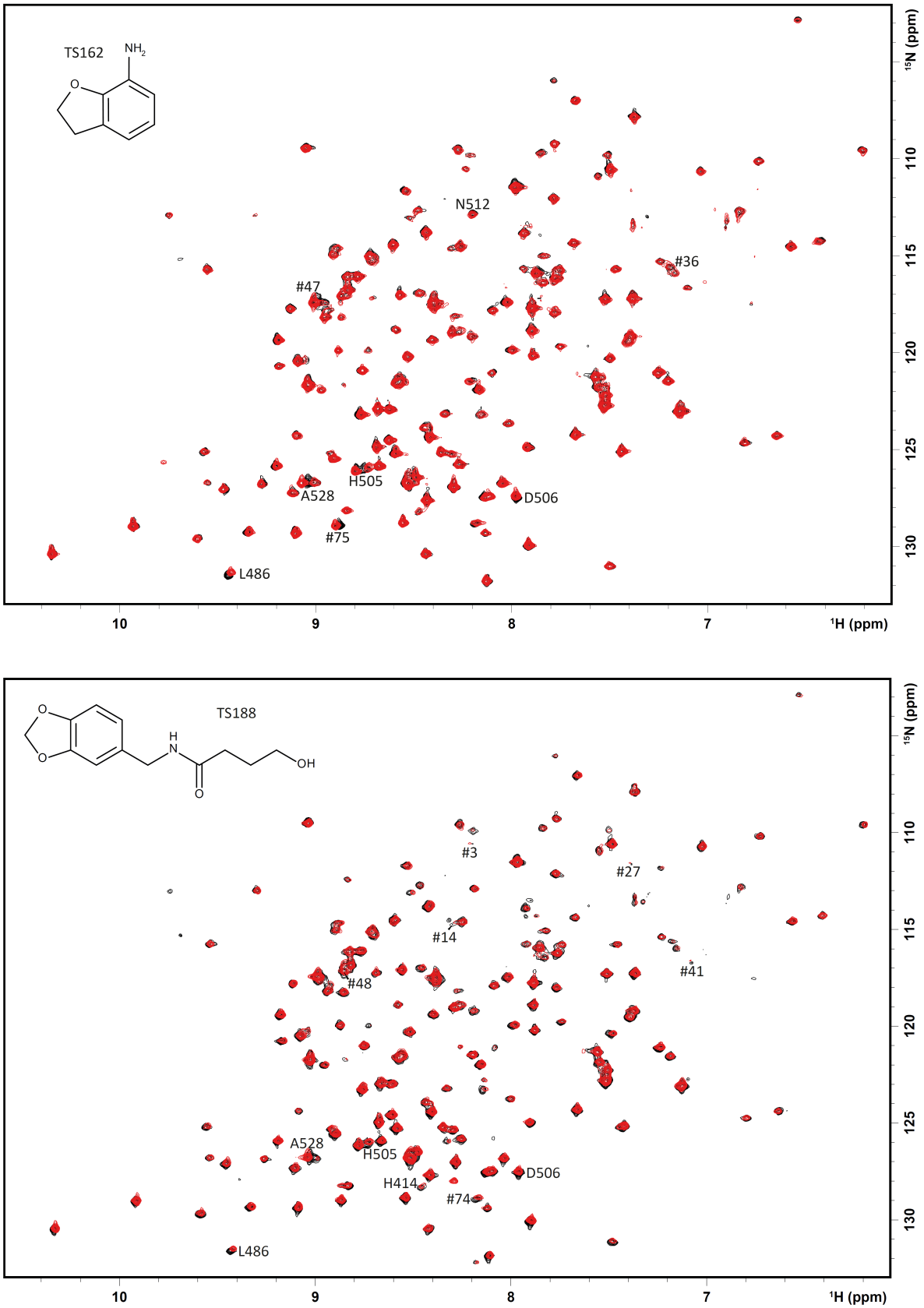


Figure A40 | TS162 and TS188 (Figure A25 continued)

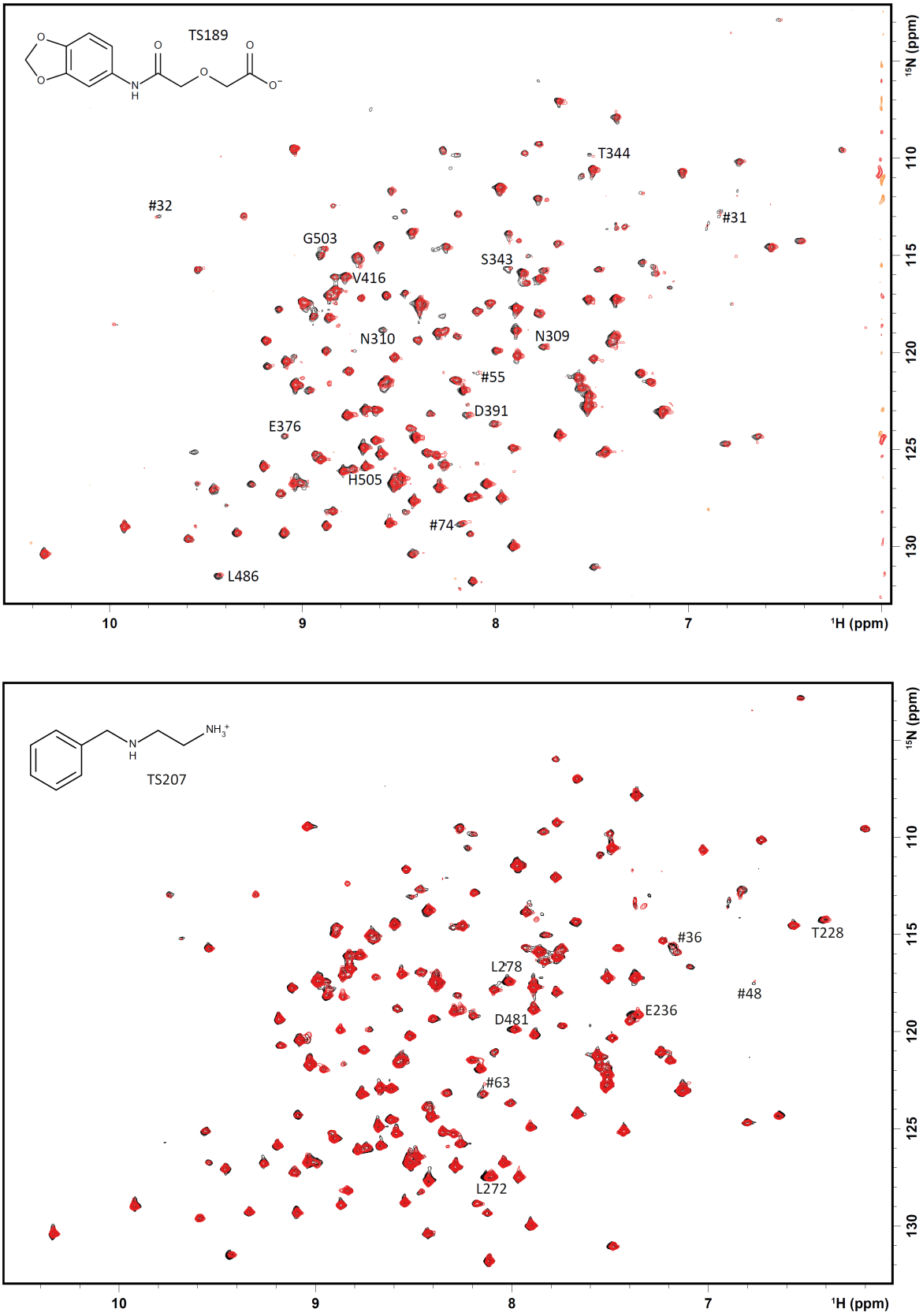


Figure A41 | TS189 and TS207 (Figure A25 continued)

A8 Docking and Virtual Screening

WNV Protease Screening

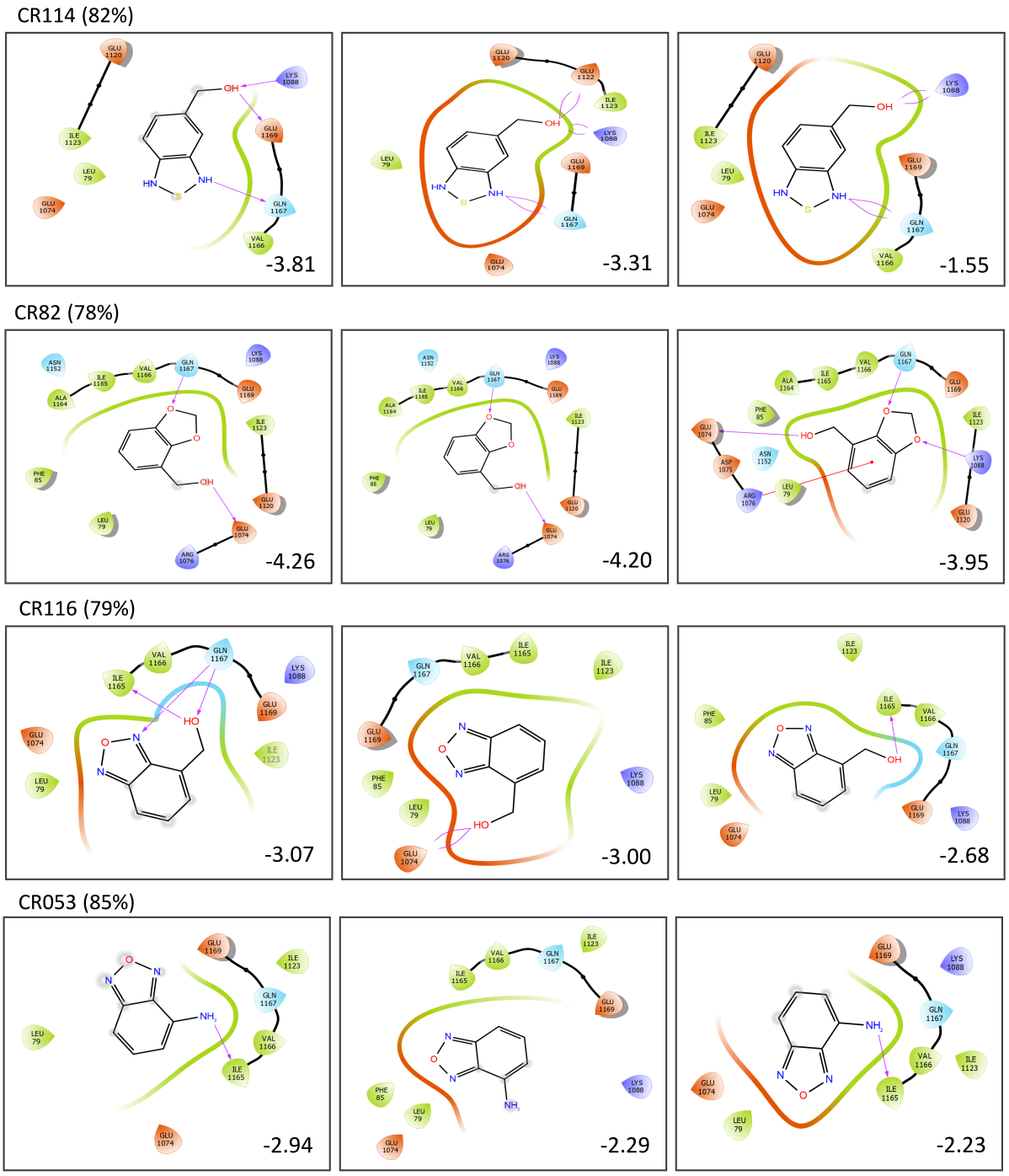


Figure A42 | Docking of SAR fragments into binding pocket Q167
These fragments seem to bind in the binding pocket (Q167) as seen for fragment **CR402**. However, neither the docking scores (right bottom corner) nor the docking poses (best three) could explain why **CR402** exhibits a higher inhibitory potency (% values).

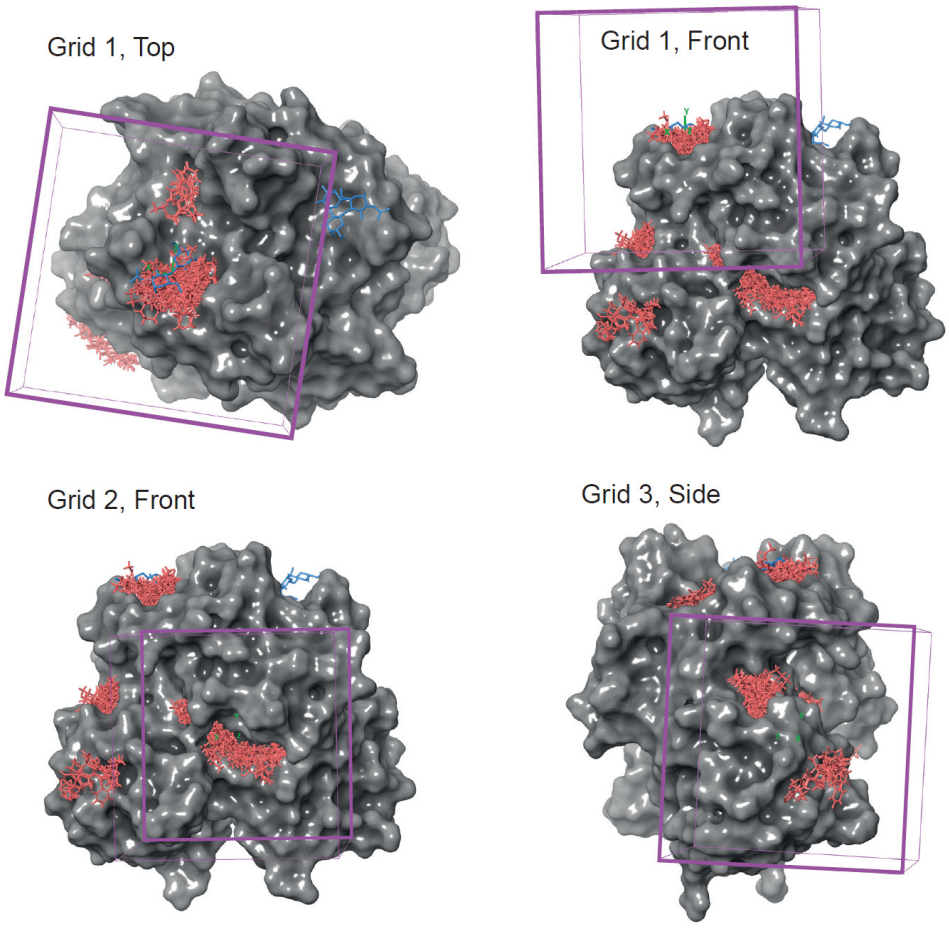


Figure A43 | Receptor grids of Saga P dimer used in the virtual screening

Table A7 | Position and dimensions of receptor grids of the Saga P dimer

Receptor Grid	Grid Center	Size (outer Box, Å)
1 – Fucose binding site	119.8, 4.8, 109.8	35x35x35
2a- Leucine 306	103.9, -25.3, 102.8	27x27x27
2b – Isoleucine 231	103.9, -40.0, 102.8	27x27x27
3 – Glycine 503	128.5, -25.4, 101.5	27x27x27

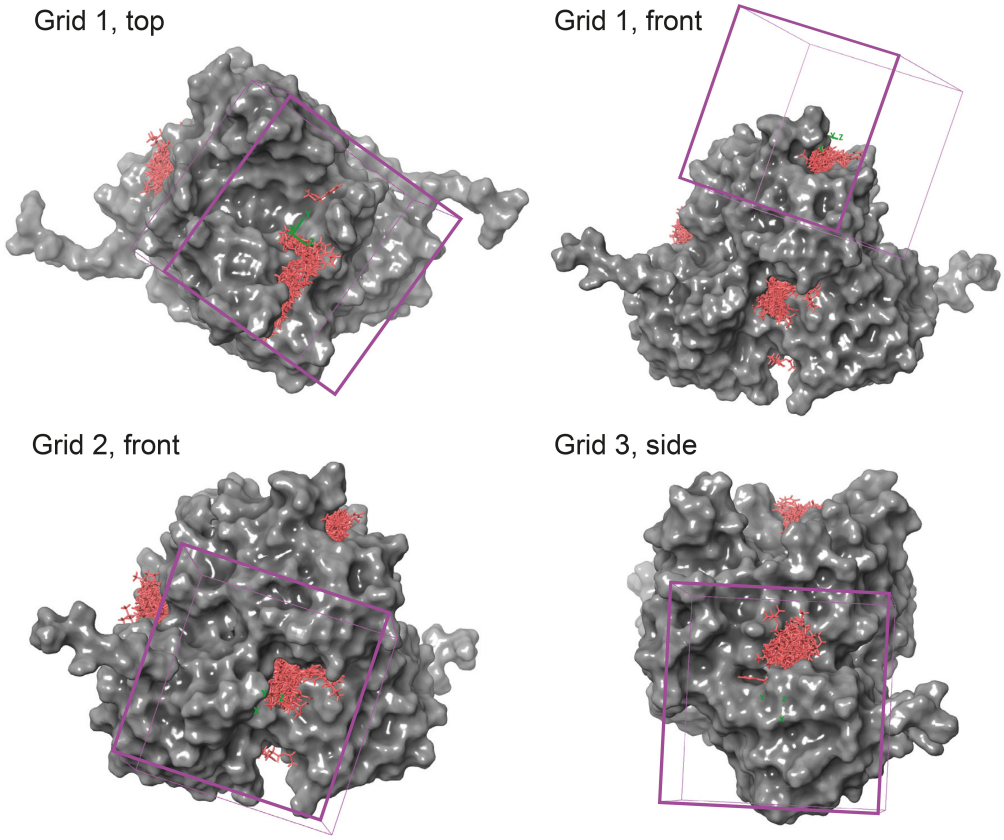


Figure A44 | Receptor grids of MNV-1 P dimer used in the virtual screening

Table A8 | Position and dimensions of receptor grids of the MNV-1 P dimer

Receptor Grid	Grid Center	Size (outer Box, Å)
1 – GM3 binding site	-18, 0, 14	42x42x42
2- Leucine 306	25, -16, 20	42x42x42
3 – Glycine 503	24, -2, -3	42x42x42

Table A9 | Dihedral angles of 3'-sialyllactose docked in binding sites of MNV-1 P dimers

Grid	NeuNAc(2→3)Gal ϕ/ψ (deg)	Gal(1→4)Glc ϕ/ψ (deg)	Docking score/ Emodel
1	-56.0/ -1.9	30.6/ -33.4	-7.1/ -44.3
2	62.2/ -19.1	54.8/ -16.3	-9.7/ -51.8
3	-10.4/ -48.3	59.5/ -5.2	-9.3 / -64.5
"free 1"[a]	-156.9/ -24.5	163.0/ 5.3	-
„free 2"[b]	-68/ -3	n.d.	-

[a]global minimum based on Maestro, Ligand Preparation tool

[b]globel minimum according to REF[298]

n.d. = not determined

Definition of dihedral angles:

ϕ = H1-C1-O1-Cx (for NeuNAc C1-C2-O2-Cx)

ψ = C1-O1-Cx-Hx (for NeuNAc C2-O2-Cx-Hx)

A9 Lists of Chemicals, Materials, and Equipments

Chemicals	Vendor
^{15}N-Amoniumchlorid	Cambridge Isotopes
2-Keto-3-2D-4-^{13}C-Butyrat	Institute of Chemistry and Metabolsim
2-Keto-3-2H-4-^{13}C-Butyrat	CortecNet
2-β-mercaptoethanol	Sigma Aldrich
3'-sialyllactose	Carbosynth
4% Paraformaldehyde	Affymetrix
7-amino-4-methylcoumarin (AMC)	Institute of Biochemistry
Alkyne sugars	Institute of Chemistry and Metabolsim
Alpha-ketoisovalerate (Na^+) (3-Methyl-^{13}C; 3,4,4,4-D4)	Cambridge Isotopes
AMC-Substrate nonapeptide	Institute of Biochemistry
Ammonium Persulfate (APS)	Sigma-Aldrich
Ampicillin	Gerbu
Ascorbate	Roth
Azdio sugars	Institute of Chemistry and Metabolsim
Biotin-alkyne	Institute of Chemistry and Metabolsim
Biotin-azide	Institute of Chemistry and Metabolsim
Bis-Acrylamide/Acrylamide 30:1	Roth
Bromphenol blue	Merck
BTTAA	Institute of Chemistry and Metabolsim
Calcium chloride	Roth
Calium chloride	Merck
Casamin acid	Institute of Chemistry and Metabolsim
Chloric acid (HCl	Roth
Choline chloride (B4)	Sigma Aldrich
Cobalamin (B12)	Sigma Aldrich
Copper sulfate pentahydrat	Merck
DAPI (40,6-Diamidino-2-phenylindole dihydrochloride)	Sigma Aldrich
D-Biotin	Sigma Aldrich
Deut. Glucose ^{12}C	Cambridge Isotopes
Deuteriumoxid 99,85%D / 99,9%D (D_2O)	Euriso-Top/ CortecNet
D-Glucose	Roth
Disodium hydrogen phoshate dihydrate	Merck
DMSO (Dimetyl sulfoxide)	Sigma Aldrich
D-Pantothenic acid	Sigma Aldrich
DSS-d_6 4,4-dimethyl-4-silapentane-1-sulfonic acid	Sigma Aldrich
DTT (Dithiothreitol)	Sigma Aldrich
EDC (1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide HCl)	Carbosynth
Ethylenediaminetetraacetic acid (EDTA)	Merck
Folic acid	Sigma Aldrich
Fragment A (S01870SC)	Maybridge, Thermo Fisher Scientific
GelCode® Blue Stain Reagent	BioRad
Gelcode® Blue Stain Reagent	Thermo scientific
Glycerol >99%	Sigma Aldrich
Glycin	Merck
HEPES (2-(4-(2-Hydroxyethyl)-1-piperazinylethansulfon acid)	Gerbu
Imidazole	Fluka

Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Roth
Kalium chloride	Merck
Kalium dihydrogen phosphate (Anhydrat)	Roth
Magnesium chloride	Merck
Magnesium sulfate	Merck
Mangan(II)-Chloride (H_20)	Fluka
Maybridge Library CR2006/ TS2017	Maybridge, Thermo Fisher Scientific
Methanol	Sigma-Aldrich
Methyl α-L-fucopyranoside	Carbosynth
Mineral stock	Institute of Chemistry and Metabolsim
Myo-inositol	Sigma Aldrich
Natrium chloride	Roth
NHS (N-hydroxysuccinimide)	GE Healthcare
Nicotinamide	Sigma Aldrich
N-Octyl-D-glucopyranoside	Riedel
Penicillin/Streptomycin (NaCl 10000 U/ml)	GE Healthcare
Polysorbat 20 (Tween 20)	Sigma
Protease inhibitor mix (cOmplete)	Roche
Pyridoxal hydrochloride	Sigma Aldrich
Riboavin	Sigma Aldrich
SDS (sodium dodecyl sulfate polyacrylamide)	Serva
Sodium Desoxycholat	Sigma Aldrich
Sodium dihydrogen phosphate Dihydrate	Merck
Sodium sulfate	Merck
Tetramethylethylenediamine (TEMED)	Sigma Aldrich
Tetrametyhlendiamin	Sigma Aldrich
THPTA	Thermo Scientific Fisher
tris(hydroxymethyl)aminomethane (Tris) /HCl	Roth
TSP-d_4 (Na^+) 3-(Trimethylsilyl)-2,2,3,3-deuteropropionate	Sigma Aldrich
Urea	Gerbu
Vitamin B1	Sigma Aldrich
Vitamine mix (100x)	Sigma-Aldrich/ Institute of Chemistry and Metabolsim
Western Lightning® Plus-ECL (Chemiluminescence)	Perkin Elmer

Cells, Media, Biomolecules, Columns	Vendor
BL21(DE3) E.coli cells (transformed)	Institute of Chemistry and Metabolsim
HEK293T cells	Institute of Virology and Cell Biology
Huh7 cells	Institute of Virology and Cell Biology
MDBK cells	Institute of Virology and Cell Biology
Agar Agar	Roth
Dulbecco's Modified Eagle Medium [+]4,5g/l D-Glucose, [+] L-Glutamine, [-] Pyruvate	Thermo Fisher Scientific
Eco E.coli-OD2 DN	Silantes
Horse Serum	Thermo Fisher Scientific
Fetal Calf Serum	Thermo Fisher Scientific
Typtone	Roth
Yeast extract	Roth
Pierce® High Sensitiv Streptavidin-HRP	Thermo Fisher Scientific
α-CD46 (CA17/1/6) (mouse)	Institute of Virology and Cell Biology
α-E_2 (mouse)	Institute of Virology and Cell Biology
α-mouse- Alexa Fluor®488	Thermo Scientific Fisher
α-mouse-HRP (Goat)	Institute of Virology and Cell Biology
α-mouse-Indocarbocyanin/Cy3 (Goat)	Dianova
Protein marker (Precision Plus Protein™ Dual Xtra Standards)	BioRad
PNGase F	New England BioLab
Bovine Serum Albumin	Serva
TEV protease	Institute of Virology and Cell Biology (Taube Lab)
HRV 3C protease	Institute of Virology and Cell Biology (Taube Lab)
MI001 P dimers	Institute of Virology and Cell Biology (Taube Lab)
HiLoad® 26/600 Superdex® 75 prep grade	GE Healthcare Life Science
Ni-NTA Superflow	Qiagen
HisTrap FF column (Ni)	GE Healthcare Life Science
IgG Sepharose™ 6 Fast Flow	GE Healthcare Life Science
Column (empty)	BioRad

Materials (Standard)	Vendor
1,5 mL tubes	Eppendorf
10 mL syringes	Roth
10/15 cm cell culture dish	Greiner bio-one
15 mL tubes	Sarstedt
20 mL syringes	Roth
5 mL syringes	Roth
50 mL tubes	Sarstedt
2L/ 1L/ 0.5 L glass bottles	Schott Duran
Blunt needles	Becton Dickinson S.A.
Centrifuge lter Amicon Ultra15, 10kDa	Merck Millipore
Centrifuge lter Amicon Ultra4 10 kDa	Merck Millipore
Cuvettes	Sarstedt
Dialyze cassette (Slide-A-Lyzer™ dialysis cassette	Thermo Fisher Scientific
Glas plates (SDS PAGE)	BioRad
Microtiter plates (96-well)	Greiner bio-one
Multi-well plates	Greiner bio-one
Nalgene® cryogenic vials (1.2 ml)	Sigma Aldrich
Nitrocellulose membrane	BioRad
NMR tubes, 3mm	Hilgenberg
Rotilabo-syringe lter 0,22 µm	Roth
Rotilabo-syringe lter 0,8 µm	Roth
SnakeSkin Dialysis Tubing, 10.000 MWCO	Thermo Scientic
TC-Flask T175	Sarstedt
Neubauer count chamber	OpticLabor

Device	Type	Vendor
Autoclave	Laboklav 55 MSV	SHP Steriltechnik AG
Bench	MSC Advantage	Thermo Scientific
Balances	PCB 1000-2	Kern & Sohn GmbH
	MC 210 S	Sartorius
Centrifuges	Minispin	Eppendorf
	Centrifuge 5810	Eppendorf
	Centrifuge 5415 D	Eppendorf
	Sorvall RC-5B Plus	Thermo Scientic
	Universal 320 R	Hettich Zentrifugen
	Ultrazentrifuge	Beckmann
Flourescence microscope	AxioCam MRm,	Zeiss
French Press Cell Disrupter		Thermo Electron
Gel caster	Protein Electrophoresis Systems	Bio-Rad
Heating block	Thermostat 5320	Eppendorf
	Thermomixer 5436	
Ice machine		Scotsman
Incubator	Refrigerated Incubator	New Brunswick Scientic
	Shaker Innova 4330	
	37°C, 5% CO_2, 65% RH	Labotec

Freezer/ refrigerator	4 °C -60 °C	Liebherr Zapf
Luminescent Image Analyzer	LAS-4000 mini ImageQuantTMLAS 4000 (Software)	Fujifilm GE Healthcare
Lyophilisator	Alpha 1-2 LD plus	Martin Christ Gefriertrocknungsanlagen GmbH
Magnet heating plate	C-MAG HS7 IKAMAG RET	IKA Labortechnik
Millipore unit	SG-Reinstwassersystem	SG
NMR spectrometer	Avance DRX 500 with TCI cryo probe NMR sampler: SampleXpress Avance 250	Bruker
pH-Meter	pH 320 Orion Star A221	WTW Thermo Scientic
Photometer	V-550 NanoDrop1000	Jasco peqlab
Pump system	ÄKTApurifier Fraction Collector Frac-950 ÄKTA™ pure (Paula) Econo Gradient Pump	GE Healthcare Life Sciences GE Healthcare Life Sciences GE Healthcare Life Sciences BioRad
Shaker	KS 125 Basic	IKA Labortechnik
SpectroSize™ 300 device	Dynamic Light Scattering	XtalConcepts
Sonicator	Sonorex TK 52 Sonoplus HD200 UW200 MS 73/D sonotrode	Bandelin Bandelin Bandelin Sonoplus
SPR spectrometer	Biacore 3000 CM5 sensor chips	Biacore/ GE Healthcare
Water bath	Cell Culture	GFL

Bibliography

1. Mukhopadhyay, S., Kuhn, R. J. & Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat Rev Microbiol* 3, 13–22 (2005).
2. Suthar, M. S., Diamond, M. S. & Gale, M. West Nile virus infection and immunity. *Nat Rev Microbiol* 11, 115–128 (2013).
3. Grove, J. & Marsh, M. The cell biology of receptor-mediated virus entry. *J Cell Biol* 195, 1071–1082 (2011).
4. de Haan, C. A. M. *et al.* Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. *J Virol* 79, 14451–6 (2005).
5. Plochmann, K. *et al.* Heparan Sulfate Is an Attachment Factor for Foamy Virus Entry. *J Virol* 86, 10028–10035 (2012).
6. Compton, T., Nowlin, D. M. & Cooper, N. R. Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology* 193, 834–841 (1993).
7. Harrington, P. R., Lindesmith, L., Yount, B., Moe, C. L. & Baric, R. S. Binding of Norwalk virus-like particles to ABH histo-blood group antigens is blocked by antisera from infected human volunteers or experimentally vaccinated mice. *J Virol* 76, 12335–43 (2002).
8. Tan, M. & Jiang, X. Norovirus and its histo-blood group antigen receptors: An answer to a historical puzzle. *Trends in Microbiology* 13, 285–293 (2005).
9. Hansman, G. S. *et al.* Structural Basis for Norovirus Inhibition and Fucose Mimicry by Citrate. *J Virol* 86, 284–292 (2012).
10. Mallagaray, A. *et al.* Attachment of norovirus to histo blood group antigens: a cooperative multistep process. *Angew Chem Int Ed Engl* 54, 12014–12019 (2015).
11. Fiege, B. *et al.* Molecular details of the recognition of blood group antigens by a human norovirus as determined by STD NMR spectroscopy. *Angew Chem Int Ed* 51, 928–932 (2012).
12. Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. Infectious entry pathway of influenza virus in a canine kidney cell line. *J Cell Biol* 91, 601–613 (1981).
13. Yolken, R. H., Willoughby, R., Wee, S. B., Miskuff, R. & Vonderfecht, S. Sialic acid glycoproteins inhibit in vitro and in vivo replication of rotaviruses. *J Clin Invest* 79, 148–154 (1987).
14. Willoughby, R. E. Rotaviruses preferentially bind O-linked sialylglycoconjugates and sialomucins. *Glycobiology* 3, 437–445 (1993).
15. Taube, S. *et al.* Ganglioside-linked terminal sialic acid moieties on murine macrophages function as attachment receptors for murine noroviruses. *J Virol* 83, 4092–4101 (2009).
16. Han, L. *et al.* Gangliosides are ligands for human noroviruses. *J Am Chem Soc* 136, 12631–12637 (2014).
17. Burckhardt, C. J. *et al.* Drifting motions of the adenovirus receptor CAR and immobile integrins initiate virus uncoating and membrane lytic protein exposure. *Cell Host Microbe* 10, 105–117 (2011).
18. López, S. & Arias, C. F. Multistep entry of rotavirus into cells: A Versaillesque dance. *Trends in Microbiology* 12, 271–278 (2004).
19. Coyne, C. B. & Bergelson, J. M. Virus-induced Abl and Fyn kinase signals permit coxsackievirus entry through epithelial tight junctions. *Cell* 124, 119–131 (2006).
20. Heldwein, E. E. & Krummenacher, C. Entry of herpesviruses into mammalian cells. *Cellular and Molecular Life Sciences* 65, 1653–1668 (2008).
21. Maurer, K., Krey, T., Moennig, V., Thiel, H. J. & Rumenapf, T. CD46 is a cellular receptor for bovine viral diarrhea virus. *J Virol* 78, 1792–1799 (2004).
22. Mei, M. *et al.* Identification of novel viral receptors with cell line expressing viral receptor-binding protein. *Sic Rep* 5, (2015).

23. Trauger, S. a, Wu, E., Bark, S. J., Nemerow, G. R. & Siuzdak, G. The identification of an adenovirus receptor by using affinity capture and mass spectrometry. *Chembiochem* 5, 1095–9 (2004).
24. Sarangi, A., Bupp, K. & Roth, M. J. Identification of a retroviral receptor used by an envelope protein derived by peptide library screening. *Proc Natl Acad Sci U S A* 104, 11032–11037 (2007).
25. Haga, K. *et al.* Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells. *Proc Natl Acad Sci* 113, E6248–E6255 (2016).
26. Frei, A. P. *et al.* Direct identification of ligand-receptor interactions on living cells and tissues. *Nat Biotechnol* 30, 997–1001 (2012).
27. MacKinnon, A. L. & Taunton, J. in *Current Protocols in Chemical Biology* (2009).
28. Kayser, H. *et al.* Biosynthesis of a nonphysiological sialic acid in different rat organs, using N-propanoyl-D-hexosamines as precursors. *J Biol Chem* 267, 16934–16938 (1992).
29. Keppler, O. T. *et al.* Biosynthetic modulation of sialic acid-dependent virus-receptor interactions of two primate polyoma viruses. *J Biol Chem* 270, 1308–1314 (1995).
30. Herrmann, M., von der Lieth, C. W., Stehling, P., Reutter, W. & Pawlita, M. Consequences of a subtle sialic acid modification on the murine polyomavirus receptor. *J Virol* 71, 5922–5931 (1997).
31. Keppler, O. T. *et al.* Elongation of the N-acyl side chain of sialic acids in MDCK II cells inhibits influenza A virus infection. *Biochem Biophys Res Commun* 253, 437–442 (1998).
32. Keppler, O. T., Horstkorte, R., Pawlita, M., Schmidt, C. & Reutter, W. Biochemical engineering of the N-acyl side chain of sialic acid: biological implications. *Glycobiology* 11, 11R–18R (2001).
33. Hall, M. K., Reutter, W., Lindhorst, T. & Schwalbe, R. A. Biochemical engineering of the N-acyl side chain of sialic acids alters the kinetics of a glycosylated potassium channel Kv3.1. *FEBS Lett.* 585, 3322–3327 (2011).
34. Prescher, J. A., Dube, D. H. & Bertozzi, C. R. Chemical remodelling of cell surfaces in living animals. *Nature* 430, 873–877 (2004).
35. Dube, D. H., Prescher, J. A., Quang, C. N. & Bertozzi, C. R. Probing mucin-type O-linked glycosylation in living animals. *Proc Natl Acad Sci* 103, 4819–4824 (2006).
36. Laughlin, S. T. *et al.* Metabolic Labeling of Glycans with Azido Sugars for Visualization and Glycoproteomics. *Methods in Enzymology* 415, 230–250 (2006).
37. Prescher, J. A. & Bertozzi, C. R. Chemical Technologies for Probing Glycans. *Cell* 126, 851–854 (2006).
38. Lauehlin, S. T. & Bertozzi, C. R. Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via staudinger ligation. *Nat Protoc* 2, 2930–2944 (2007).
39. Hsu, T.-L. *et al.* Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proc Natl Acad Sci* 104, 2614–2619 (2007).
40. Hossler, P., Khattak, S. F. & Li, Z. J. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19, 936–949 (2009).
41. Luchansky, S. J., Goon, S. & Bertozzi, C. R. Expanding the diversity of unnatural cell-surface sialic acids. *ChemBioChem* 5, 371–374 (2004).
42. Zhao, X. *et al.* Labeling of enveloped virus via metabolic incorporation of azido sugars. *Bioconjug Chem* 26, 1868–1872 (2015).
43. Wang, F. I., Deng, M. C., Huang, Y. L. & Chang, C. Y. Structures and Functions of Pestivirus Glycoproteins: Not Simply Surface Matters. *Viruses* 7, 3506–3529 (2015).
44. Krey, T. *et al.* Function of bovine CD46 as a cellular receptor for bovine viral diarrhea virus is determined by complement control protein 1. *J Virol* 80, 3912–3922 (2006).
45. Pande, A. *et al.* The glycosylation pattern of baculovirus expressed envelope protein E2 affects its ability to prevent infection with bovine viral diarrhoea virus. *Virus Res* 114, 54–62 (2005).
46. Kielian, M., Maramorosch, K. & Mettenleiter, T. C. *Advances in Virus Research*. (2015).
47. Agnello, V., Abel, G., Elfahal, M., Knight, G. B. & Zhang, Q. X. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci U. S. A.* 96, 12766–12771 (1999).

48. Li, Y., Wang, J., Kanai, R. & Modis, Y. Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. *Proc Natl Acad Sci U S A* 110, 6805–6810 (2013).
49. Mathapati, B. S. *et al.* Entry of bovine viral diarrhea virus into ovine cells occurs through clathrin-dependent endocytosis and low pH-dependent fusion. *Vitr. Cell. Dev. Biol. - Anim.* 46, 403–407 (2010).
50. Tscherne, D. M., Evans, M. J., Macdonald, M. R. & Rice, C. M. Transdominant Inhibition of Bovine Viral Diarrhea Virus Entry. *J Virol* 82, 2427–2436 (2008).
51. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385 (2003).
52. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* 571–607 (2005). doi:10.1385/1592598900
53. Laemmli. Laemmli Buffer Recipe. *Nature* 5–6 (1970).
54. Marschall, R. Auf dem Weg zum " Rezeptor-Fischen " des Bovinen Virusdiarrhoe Virus mittels metabolischer Markierung und Klick-Chemie On the way to " receptor-fishing " of bovine viral diarrhea virus via metabolic labeling and click chemistry. *Master's thesis* (2016).
55. Besanceney-Webler, C. *et al.* Increasing the efficacy of bioorthogonal click reactions for bioconjugation: A comparative study. *Angew Chem Int Ed* 50, 8051–8056 (2011).
56. Soriano Del Amo, D. *et al.* Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J Am Chem Soc* 132, 16893–16899 (2010).
57. Hong, V., Steinmetz, N. F., Manchester, M. & Finn, M. G. Labeling live cells by copper-catalyzed alkyne--azide click chemistry. *Bioconjug Chem* 21, 1912–1916 (2010).
58. Kimple, M. E., Brill, A. L. & Pasker, R. L. Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* (2013).
59. Han, S. S. *et al.* Physiological effects of Ac4ManNAz and optimization of metabolic labeling for cell tracking. *Theranostics* 7, 1164–1176 (2017).
60. Kaewsapsak, P., Esonu, O. & Dube, D. H. Recruiting the Host's Immune System to Target Helicobacter pylori's Surface Glycans. *ChemBioChem* 14, 721–726 (2013).
61. Lalonde, M. E. & Durocher, Y. Therapeutic glycoprotein production in mammalian cells. *Journal of Biotechnology* 251, 128–140 (2017).
62. Croset, A. *et al.* Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J. Biotechnol.* 161, 336–348 (2012).
63. Steentoft, C. *et al.* Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32, 1478–1488 (2013).
64. Breidenbach, M. A. *et al.* Targeted metabolic labeling of yeast N-glycans with unnatural sugars. *Proc Natl Acad Sci* 107, 3988–3993 (2010).
65. Péanne, R. *et al.* Assessing ER and golgi N-glycosylation process using metabolic labeling in mammalian cultured cells. *Methods Cell Biol* 118, 157–176 (2013).
66. Chang, P. V. *et al.* Metabolie labeling of sialic acids in living animals with alkynyl sugars. *Angew Chem Int Ed* 48, 4030–4033 (2009).
67. Lin, S. H. & Guidotti, G. Chapter 35 Purification of Membrane Proteins. *Methods Enzymol.* 463, 619–629 (2009).
68. Hjelmeland, L. M. Solubilization of native membrane proteins. *Methods Enzymol.* 182, 253–264 (1990).
69. Bass, D. M. & Greenberg, H. B. Strategies for the identification of icosahedral virus receptors. *Journal of Clinical Investigation* 89, 3–9 (1992).
70. Skeene, K. *et al.* One Filter, One Sample, and the N- and O-Glyco(proteo)me: Toward a System to Study Disorders of Protein Glycosylation. *Anal Chem* 89, 5840–5849 (2017).
71. Rich, R. L., Miles, A. R., Gale, B. K. & Myszka, D. G. Detergent screening of a G-protein-coupled receptor using serial and array biosensor technologies. *Anal Biochem* 386, 98–104 (2009).
72. Cho, K. H. *et al.* Maltose neopentyl glycol-3 (MNG-3) analogues for membrane protein study. *Analyst* 140, 3157–3163 (2015).

73. Chang, V. T. *et al.* Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure* 15, 267–273 (2007).
74. Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nat Rev Drug Discov* 3, 660–672 (2004).
75. Hann, M. M., Leach, A. R. & Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J Chem Inf Comput Sci* 41, 856–864 (2001).
76. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: The impact of fragments on drug discovery. *Nat Rev Drug Discov* 15, 605–619 (2016).
77. Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* 467, 596–599 (2010).
78. Navratilova, I. & Hopkins, A. L. Fragment screening by surface plasmon resonance. *ACS Med Chem Lett.* 1, 44–48 (2010).
79. Neumann, T., Junker, H.-D., Schmidt, K. & Sekul, R. SPR-based Fragment Screening: Advantages and Applications. *Curr Top Med Chem* 7, 1630–1642 (2007).
80. Kuhn, P., Wilson, K., Patch, M. G. & Stevens, R. C. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Current Opinion in Chemical Biology* 6, 704–710 (2002).
81. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* 42, 864–890 (2003).
82. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science (80-.).* 274, 1531–1534 (1996).
83. Roegener, J. *et al.* Ultrasensitive detection of unstained proteins in acrylamide gels by native UV fluorescence. *Anal Chem* 75, 157–159 (2003).
84. Skinner, A. L. & Laurence, J. S. High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. *Journal of Pharmaceutical Sciences* 97, 4670–4695 (2008).
85. Dias, D. M. & Ciulli, A. NMR approaches in structure-based lead discovery: Recent developments and new frontiers for targeting multi-protein complexes. *Prog Biophys Mol Biol* 6107, 12 (2014).
86. Gossert, A. D. & Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy* 97, 82–125 (2016).
87. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int Ed* 38, 1784–1788 (1999).
88. Dalvit, C., Fogliatto, G. P., Stewart, A., Veronesi, M. & Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* 21, 349–359 (2001).
89. Rademacher, C. Investigations into Viral Entry Mechanisms and Carbohydrate-Protein Interactions using NMR Experiments. *Dissertation* (2008).
90. Rademacher, C. *et al.* Targeting norovirus infection-multivalent entry inhibitor design based on NMR experiments. *Chemistry (Easton).* 17, 7442–7453 (2011).
91. Rademacher, C. & Peters, T. Molecular recognition of ligands by native viruses and virus-like particles as studied by NMR experiments. *Top Curr Chem* 273, 183–202 (2008).
92. Tugarinov, V. & Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* 125, 13868–13878 (2003).
93. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1, 749–754 (2006).
94. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci* 94, 12366–12371 (1997).
95. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73, 1–16 (2013).

96. Liedberg, B., Nylander, C. & Lunström, I. Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators* 4, 299–304 (1983).
97. Goddard, J. P. & Reymond, J. L. Enzyme assays for high-throughput screening. *Current Opinion in Biotechnology* 15, 314–322 (2004).
98. Michelini, E., Cevenini, L., Mezzanotte, L., Coppa, A. & Roda, A. Cell-based assays: Fuelling drug discovery. *Analytical and Bioanalytical Chemistry* 398, 227–238 (2010).
99. Kim, R. & Skolnick, J. Assessment of programs for ligand binding affinity prediction. *J. Comput. Chem.* 29, 1316–1331 (2008).
100. Tang, G. W. & Altman, R. B. Knowledge-based fragment binding prediction. *PLoS Comput Biol* 10, (2014).
101. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie - International Edition* 41, 2644–2676 (2002).
102. Durrant, J. D. & McCammon, J. A. NNScore: A neural-network-based scoring function for the characterization of protein-ligand complexes. *J Chem Inf Model* 50, 1865–1871 (2010).
103. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 49, 6177–6196 (2006).
104. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 49, 534–553 (2006).
105. Lim, S. P. & Shi, P. Y. West Nile virus drug discovery. *Viruses* 5, 2977–3006 (2013).
106. Brinton, M. A. The molecular biology of West Nile Virus: a new invader of the western hemisphere. *Annu Rev Microbiol* 56, 371–402 (2002).
107. Hayes, E. B. & Gubler, D. J. West Nile virus: epidemiology and clinical features of an emerging epidemic in the United States. *Annu Rev Med* 57, 181–194 (2006).
108. Shiryaev, S. & Strongin, A. Structural and funtional parameters of the fllaviviral protease: a promising antiviral drug target. *Futur Virol* 5, 593–606 (2010).
109. Bakonyi, T., Ivanics, E., Erdélyi, K. & Nowotny, N. Lineage 1 and 2 Strains of Encephalitic West Nile Virus, Central Europe. *Emerg. Infect. Dis.* 12, 618–623 (2006).
110. Diamond, M. S., Shrestha, B., Mehlhop, E., Sitati, E. & Engle, M. Innate and adaptive immune responses determine protection against disseminated infection by West Nile encephalitis virus. *Viral Immunol* 16, 259–278 (2003).
111. Mostashari, F. *et al.* Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. *Lancet* 358, 261–264 (2001).
112. Lanciotti, R. S. *et al.* Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science (80-.).* 286, 2333–2337 (1999).
113. Sejvar, J. J. The long-term outcomes of human West Nile virus infection. *Clin Infect Dis* 44, 1617–1624 (2007).
114. Murray, K. *et al.* Persistent infection with West Nile virus years after initial infection. *J Infect Dis* 201, 2–4 (2010).
115. Lim, P. Y., Behr, M. J., Chadwick, C. M., Shi, P. Y. & Bernard, K. A. Keratinocytes are cell targets of West Nile virus in vivo. *J Virol* 85, 5197–5201 (2011).
116. Gamino, V. & Hofle, U. Pathology and tissue tropism of natural West Nile virus infection in birds: a review. *Vet Res* 44, 1297–9716 (2013).
117. Suthar, M. S. *et al.* A systems biology approach reveals that tissue tropism to West Nile virus is regulated by antiviral genes and innate immune cellular processes. *PLoS Pathog* 9, 7 (2013).
118. Mackenzie, J. Wrapping things up about virus RNA replication. *Traffic* 6, 967–977 (2005).
119. Nall, T. A. *et al.* Enzymatic characterization and homology model of a catalytically active recombinant West Nile virus NS3 protease. *J Biol Chem* 279, 48535–48542 (2004).
120. Bollati, M. *et al.* Structure and functionality in flavivirus NS-proteins: perspectives for drug design. *Antivir Res* 87, 125–148 (2010).

121. Gorbalenya, A. E., Donchenko, A. P., Koonin, E. V & Blinov, V. M. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Res* 17, 3889–3897 (1989).
122. Liu, W. J., Sedlak, P. L., Kondratieva, N. & Khromykh, A. A. Complementation analysis of the flavivirus Kunjin NS3 and NS5 proteins defines the minimal regions essential for formation of a replication complex and shows a requirement of NS3 in cis for virus assembly. *J Virol* 76, 10766–10775 (2002).
123. Assenberg, R. *et al.* Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. *J Virol* 83, 12895–12906 (2009).
124. Wengler, G., Czaya, G., Farber, P. M. & Hegemann, J. H. In vitro synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. *J Gen Virol* 72, 851–858 (1991).
125. Chappell, K. J., Stoermer, M. J., Fairlie, D. P. & Young, P. R. Generation and characterization of proteolytically active and highly stable truncated and full-length recombinant West Nile virus NS3. *Protein Expr Purif* 53, 87–96 (2007).
126. Erbel, P. *et al.* Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat Struct Mol Biol* 13, 372–373 (2006).
127. Aleshin, A. E., Shiryaev, S. A., Strongin, A. Y. & Liddington, R. C. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein Sci* 16, 795–806 (2007).
128. Hammamy, M. Z., Haase, C., Hammami, M., Hilgenfeld, R. & Steinmetzer, T. Development and characterization of new peptidomimetic inhibitors of the West Nile virus NS2B-NS3 protease. *ChemMedChem* 8, 231–241 (2013).
129. Robin, G. *et al.* Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. *J Mol Biol* 385, 1568–1577 (2009).
130. Garces, A. P. & Watowich, S. J. Intrinsic flexibility of West Nile virus protease in solution characterized using small-angle X-ray scattering. *Biochemistry* 52, 6856–6865 (2013).
131. Su, X. C. *et al.* NMR analysis of the dynamic exchange of the NS2B cofactor between open and closed conformations of the West Nile virus NS2B-NS3 protease. *PLoS Negl Trop Dis* 3, 561 (2009).
132. Su, X. C. *et al.* NMR study of complexes between low molecular mass inhibitors and the West Nile virus NS2B-NS3 protease. *Febs J* 276, 4244–4255 (2009).
133. Schöne, T. Identifikation allosterischer Inhibitoren gegen die WNV-Protease auf Basis STD NMR experimenten. (2014).
134. Chappell, K. J., Stoermer, M. J., Fairlie, D. P. & Young, P. R. Mutagenesis of the West Nile virus NS2B cofactor domain reveals two regions essential for protease activity. *J Gen Virol* 89, 1010–1014 (2008).
135. Radichev, I. *et al.* Structure-based mutagenesis identifies important novel determinants of the NS2B cofactor of the West Nile virus two-component NS2B-NS3 proteinase. *J Gen Virol* 89, 636–641 (2008).
136. Chappell, K. J., Stoermer, M. J., Fairlie, D. P. & Young, P. R. Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies. *J Biol Chem* 281, 38448–38458 (2006).
137. Kang, C. *et al.* Exploring the binding of peptidic West Nile virus NS2B-NS3 protease inhibitors by NMR. *Antivir Res* 97, 137–144 (2013).
138. Bastos Lima, A. *et al.* Dual inhibitors of the dengue and West Nile virus NS2B-NS3 proteases: Synthesis, biological evaluation and docking studies of novel peptide-hybrids. *Bioorg Med Chem* 23, 5748–5755 (2015).
139. Behnam, M. A., Graf, D., Bartenschlager, R., Zlotos, D. P. & Klein, C. D. Discovery of Nanomolar Dengue and West Nile Virus Protease Inhibitors Containing a 4-Benzyloxyphenylglycine Residue. *J Med Chem* 58, 9354–9370 (2015).
140. Kouretova, J. *et al.* Effects of NS2B-NS3 protease and furin inhibition on West Nile and Dengue virus replication. *J. Enzyme Inhib. Med. Chem.* 32, 712–721 (2017).

141. Nitsche, C. *et al.* Peptide-Boronic Acid Inhibitors of Flaviviral Proteases: Medicinal Chemistry and Structural Biology. *J Med Chem* 60, 511–516 (2017).
142. Cregar-Hernandez, L. *et al.* Small molecule pan-dengue and West Nile virus NS3 protease inhibitors. *Antivir Chem Chemother* 21, 209–217 (2012).
143. Dou, D. *et al.* Design, synthesis, and in vitro evaluation of potential West Nile virus protease inhibitors based on the 1-oxo-1,2,3,4-tetrahydroisoquinoline and 1-oxo-1,2-dihydroisoquinoline scaffolds. *J Comb Chem* 12, 836–843 (2010).
144. Ekonomiuk, D. *et al.* Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. *J Med Chem* 52, 4860–4868 (2009).
145. Ekonomiuk, D. *et al.* Discovery of a non-peptidic inhibitor of west nile virus NS3 protease by high-throughput docking. *PLoS Negl Trop Dis* 3, 13 (2009).
146. Ezgimen, M. *et al.* Characterization of the 8-hydroxyquinoline scaffold for inhibitors of West Nile virus serine protease. *Antivir Res* 94, 18–24 (2012).
147. Mueller, N. H. *et al.* Identification and biochemical characterization of small-molecule inhibitors of west nile virus serine protease by a high-throughput screen. *Antimicrob Agents Chemother* 52, 3385–3393 (2008).
148. Nitsche, C., Steuer, C. & Klein, C. D. Arylcyanoacrylamides as inhibitors of the Dengue and West Nile virus proteases. *Bioorg Med Chem* 19, 7318–7337 (2011).
149. Johnston, P. A. *et al.* HTS identifies novel and specific uncompetitive inhibitors of the two-component NS2B-NS3 proteinase of West Nile virus. *Assay Drug Dev Technol* 5, 737–750 (2007).
150. Koh-Stenta, X. *et al.* Identification of covalent active site inhibitors of dengue virus protease. *Drug Des. Devel. Ther.* 9, 6389–6399 (2015).
151. Gao, Y., Samanta, S., Cui, T. & Lam, Y. Synthesis and in vitro evaluation of West Nile virus protease inhibitors based on the 1,3,4,5-tetrasubstituted 1H-pyrrol-2(5H)-one scaffold. *ChemMedChem* 8, 1554–1560 (2013).
152. Samanta, S., Cui, T. & Lam, Y. Discovery, synthesis, and in vitro evaluation of West Nile virus protease inhibitors based on the 9,10-dihydro-3H,4aH-1,3,9,10a-tetraazaphenanthren-4-one scaffold. *ChemMedChem* 7, 1210–1216 (2012).
153. Samanta, S., Lim, T. L. & Lam, Y. Synthesis and in vitro evaluation of West Nile virus protease inhibitors based on the 2-{6-[2-(5-phenyl-4H-{1,2,4]triazol-3-ylsulfanyl)acetylamino]benzothiazol-2-ylsul fanyl}acetamide scaffold. *ChemMedChem* 8, 994–1001 (2013).
154. Sidique, S. *et al.* Structure-activity relationship and improved hydrolytic stability of pyrazole derivatives that are allosteric inhibitors of West Nile Virus NS2B-NS3 proteinase. *Bioorg Med Chem Lett* 19, 5773–5777 (2009).
155. Shiryaev, S. A., Cheltsov, A. V., Gawlik, K., Ratnikov, B. I. & Strongin, A. Y. Virtual ligand screening of the National Cancer Institute (NCI) compound library leads to the allosteric inhibitory scaffolds of the West Nile Virus NS3 proteinase. *Assay Drug Dev Technol* 9, 69–78 (2011).
156. Davis, B. J. & Roughley, S. D. Fragment-Based Lead Discovery. *Annual Reports in Medicinal Chemistry* (2017). doi:10.1016/bs.armc.2017.07.002
157. Arroyo, X., Goldflam, M., Feliz, M., Belda, I. & Giralt, E. Computer-aided design of fragment mixtures for NMR-based screening. *PLoS One* 8, 13 (2013).
158. Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* 6, 211–219 (2007).
159. Harner, M. J., Frank, A. O. & Fesik, S. W. Fragment-based drug discovery using NMR spectroscopy. *J Biomol NMR* 56, 65–75 (2013).
160. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat Rev Drug Discov* 1, 727–730 (2002).
161. Cala, O., Guilliere, F. & Krimm, I. NMR-based analysis of protein-ligand interactions. *Analytical and Bioanalytical Chemistry* 406, 943–956 (2014).
162. Cala, O. & Krimm, I. Ligand-orientation based fragment selection in STD NMR screening. *J Med Chem* 58, 8739–8742 (2015).

163. Krimm, I., Lancelin, J. M. & Praly, J. P. Binding evaluation of fragment-based scaffolds for probing allosteric enzymes. *J Med Chem* 55, 1287–1295 (2012).
164. Landström, J. *et al.* Small molecules containing hetero-bicyclic ring systems compete with UDP-Glc for binding to WaaG glycosyltransferase. *Glycoconj. J.* 29, 491–502 (2012).
165. Rademacher, C. *et al.* NMR-based exploration of the acceptor binding site of human blood group B galactosyltransferase with molecular fragments. *Glycoconj. J.* 27, 349–358 (2010).
166. Venkitakrishnan, R. P., Benard, O., Max, M., Markley, J. L. & Assadi-Porter, F. M. Use of NMR saturation transfer difference spectroscopy to study ligand binding to membrane proteins. *Methods Mol. Biol.* 914, 47–63 (2012).
167. Vogtherr, M. & Peters, T. Application of NMR based binding assays to identify key hydroxy groups for intermolecular recognition. *J Am Chem Soc* 122, 6093–6099 (2000).
168. Schöne, T. *et al.* STD-NMR experiments identify a structural motif with novel second-site activity against West Nile virus NS2B-NS3 protease. *Antiviral Res.* 146, 174–183 (2017).
169. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41, 207–234 (2005).
170. Schwartz, M. A. & Hayton, W. L. Relative stability of hetacillin and ampicillin in solution. *J Pharm Sci* 61, 906–909 (1972).
171. Bajusz, D., Racz, A., Heberger, K., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* 7, 15–69 (2015).
172. Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today* 11, 1046–1053 (2006).
173. Cherkasov, A. *et al.* QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry* 57, 4977–5010 (2014).
174. Hansch, C., Maloney, P. P., Fujita, T. & Muir, R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* 194, 178–180 (1962).
175. Todeschini, R. & Consonni, V. *Molecular Descriptors for Chemoinformatics*. *Molecular Descriptors for Chemoinformatics* 2, (2010).
176. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46, 3–26 (2001).
177. Ferrara, P., Apostolakis, J. & Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* 46, 24–33 (2002).
178. Hasel, W., Hendrickson, T. F. & Still, W. C. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.* 1, 103–116 (1988).
179. Viswanadhan, V. N., Ghose, A. K., Revankar, G. R. & Robins, R. K. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. *J Chem Inf Comput Sci* 29, 163–172 (1989).
180. Klopman, G. *et al.* Computer Automated log P Calculations Based on an Extended Group Contribution Approach. *J Chem Inf Comput Sci* 34, 752–781 (1994).
181. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* 8, 876–877 (2003).
182. Hann, M. *et al.* Strategic pooling of compounds for high-throughput screening. *J Chem Inf Comput Sci* 39, 897–902 (1999).
183. Grimm, L. L. Ligandenbindung und Proteindynamik bei humaner Blutgruppe B Galactosyltransferase und West-Nil Virus Protease NS2B-NS3. *PhD thesis* (2015).
184. Czisch, M. & Boelens, R. Sensitivity enhancement in the TROSY experiment. *J Magn Reson* 134, 158–160 (1998).
185. Pervushin, K. V, Wider, G. & Wuthrich, K. Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J Biomol NMR* 12, 345–348 (1998).

186. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* 47, 1739–1749 (2004).
187. Kim, M. O., Nichols, S. E., Wang, Y. & McCammon, J. A. Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC. *J. Comput. Aided. Mol. Des.* 27, 235–246 (2013).
188. Olsson, M. H. M., SØndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions. *J. Chem. Theory Comput.* 7, 525–537 (2011).
189. Dixon, M. The determination of enzyme inhibitor constants. *Biochem. J.* 55, 170–171 (1953).
190. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res* 37, 24 (2009).
191. Rasia, R. M., Brutscher, B. & Plevin, M. J. Selective Isotopic Unlabeling of Proteins Using Metabolic Precursors: Application to NMR Assignment of Intrinsically Disordered Proteins. *ChemBioChem* 13, 732–739 (2012).
192. Shiryaev, S. A. Expression and purification of a two-component flaviviral proteinase resistant to autocleavage at the NS2B-NS3 junction region. *Protein Expr Purif.* 52(2), 334–339 (2007).
193. Kang, C., Gayen, S., Chen, A. S. & Huang, Q. West Nile Virus (WNV) protease and membrane interactions revealed by NMR spectroscopy. *Biochem Biophys Res Commun* 423, 799–804 (2012).
194. Mittal, S., Chowhan, R. K. & Singh, L. R. Macromolecular crowding: Macromolecules friend or foe. *Biochim. Biophys. Acta - Gen. Subj.* 1850, 1822–1831 (2015).
195. Kim, Y. M. *et al.* NMR analysis of a novel enzymatically active unlinked dengue NS2B-NS3 protease complex. *J Biol Chem* 288, 12891–12900 (2013).
196. Jayalakshmi, V. & Krishna, N. R. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J Magn Reson* 155, 106–118 (2002).
197. Baurin, N. *et al.* Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. *J Chem Inf Comput Sci* 44, 2157–2166 (2004).
198. Ezgimen, M. D., Mueller, N. H., Teramoto, T. & Padmanabhan, R. Effects of detergents on the West Nile virus protease activity. *Bioorg. Med. Chem.* 17, 3278–3282 (2009).
199. Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* 13, 105–121 (2014).
200. Higueruelo, A. P., Schreyer, A., Bickerton, G. R., Blundell, T. L. & Pitt, W. R. What can we learn from the evolution of protein-ligand interactions to aid the design of new therapeutics? *PLoS One* 7, 11 (2012).
201. Hajduk, P. J. Fragment-based drug design: how big is too big? *J Med Chem* 49, 6972–6976 (2006).
202. Hopkins, A. L., Groom, C. R. & Alex, A. *Ligand efficiency: a useful metric for lead selection*. (Drug Discov Today. 2004 May 15;9(10):430-1., 2004).
203. Proudfoot, J. R. The evolution of synthetic oral drug properties. *Bioorg Med Chem Lett* 15, 1087–1090 (2005).
204. Leeson, P. D. & Davis, A. M. Time-related differences in the physical property profiles of oral drugs. *J Med Chem* 47, 6338–6348 (2004).
205. Zhang, W. *et al.* Identification of the binding site of an allosteric ligand using STD-NMR, docking, and CORCEMA-ST calculations. *ChemMedChem* 8, 1629–1633 (2013).
206. Jayalakshmi, V. & Krishna, N. R. CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J. Magn. Reson.* 168, 36–45 (2004).
207. Jayalakshmi, V., Biet, T., Peters, T. & Krishna, N. R. Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization. *J Am Chem Soc* 126, 8610–8611 (2004).

208. Yuan, Y., Wen, X., Sanders, D. A. R. & Pinto, B. M. Exploring the mechanism of binding of UDP-galactopyranose to UDP-galactopyranose mutase by STD-NMR spectroscopy and molecular modeling. *Biochemistry* 44, 14080–14089 (2005).
209. Yuan, Y. *et al.* Investigation of binding of UDP-Gal f and UDP-[3-F]Gal f to UDP- galactopyranose mutase by STD-NMR spectroscopy, molecular dynamics, and CORCEMA-ST calculations. *J Am Chem Soc* 130, 3157–3168 (2008).
210. Sheng, C. & Zhang, W. Fragment informatics and computational fragment-based drug design: an overview and update. *Med Res Rev* 33, 554–598 (2013).
211. Ludwig, C. *et al.* SALMON: Solvent accessibility, ligand binding, and mapping of ligand orientation by NMR spectroscopy. *J Med Chem* 51, 1–3 (2008).
212. Pardridge, W. M. Drug transport across the blood-brain barrier. *J. Cereb. Blood Flow Metab.* 32, 1959–72 (2012).
213. Zheng, D. P. *et al.* Norovirus classification and proposed strain nomenclature. *Virology* 346, 312–323 (2006).
214. Wobus, C. E. *et al.* Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS Biol.* 2, (2004).
215. Karst, S. M. & Wobus, C. E. Viruses in Rodent Colonies: Lessons Learned from Murine Noroviruses. *Annu. Rev. Virol.* 2, 525–548 (2015).
216. Martella, V. *et al.* Detection and molecular characterization of a canine norovirus. *Emerg. Infect. Dis.* 14, 1306–1308 (2008).
217. Ryan, M. L. *et al.* Symptomatic and asymptomatic rotavirus and norovirus infections during infancy in a chilean birth cohort. *Pediatr. Infect. Dis. J.* 28, 879–884 (2009).
218. Fiege, B. Insights into Binding of Attachment Factors and Inhibitors to a Human Norovirus from NMR and SPR Experiments. (2012).
219. Robilotti, E., Deresinski, S. & Pinsky, B. A. Norovirus. *Clin. Microbiol. Rev.* 28, 134–164 (2015).
220. Cardemil, C. V., Parashar, U. D. & Hall, A. J. Norovirus Infection in Older Adults: Epidemiology, Risk Factors, and Opportunities for Prevention and Control. *Infectious Disease Clinics of North America* 31, 839–870 (2017).
221. Goodfellow, I. & Taube, S. in *Viral Gastroenteritis* 355–378 (Elsevier, 2016).
222. Vongpunsawad, S., Venkataram Prasad, B. V. & Estes, M. K. Norwalk Virus Minor Capsid Protein VP2 Associates within the VP1 Shell Domain. *J Virol* 87, 4818–4825 (2013).
223. McFadden, N. *et al.* Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4. *PLoS Pathog.* 7, (2011).
224. Singh, B. K., Leuthold, M. M. & Hansman, G. S. Human Noroviruses' Fondness for Histo-Blood Group Antigens. *J Virol* 89, 2024–2040 (2015).
225. Prasad, B. V. *et al.* Antiviral targets of human noroviruses. *Curr. Opin. Virol.* 18, 117–125 (2016).
226. Berger, A. & Mainou, B. Interactions between Enteric Bacteria and Eukaryotic Viruses Impact the Outcome of Infection. *Viruses* 10, 19 (2018).
227. Murakami, K. *et al.* Norovirus Binding to Intestinal Epithelial Cells Is Independent of Histo-Blood Group Antigens. *PLoS One* 8, (2013).
228. Almand, E. A., Moore, M. D. & Jaykus, L.-A. Norovirus Binding to Ligands Beyond Histo-Blood Group Antigens. *Front. Microbiol.* 8, 1–10 (2017).
229. Müller, B., Klemm, U., Mas Marques, A. & Schreier, E. Genetic diversity and recombination of murine noroviruses in immunocompromised mice. *Arch. Virol.* 152, 1709–1719 (2007).
230. Taube, S. *et al.* Murine noroviruses bind glycolipid and glycoprotein attachment receptors in a strain-dependent manner. *J Virol* 86, 5584–5593 (2012).
231. Gerondopoulos, A., Jackson, T., Monaghan, P., Doyle, N. & Roberts, L. O. Murine norovirus-1 cell entry is mediated through a non-clathrin-, non-caveolae-, dynamin- And cholesterol-dependent pathway. *J. Gen. Virol.* 91, 1428–1438 (2010).

232. Weerasekara, S., Prior, A. M. & Hua, D. H. Current tools for norovirus drug discovery. *Expert Opinion on Drug Discovery* 11, 529–541 (2016).
233. Feng, X. & Jiang, X. Library screen for inhibitors targeting norovirus binding to histo-blood group antigen receptors. *Antimicrob Agents Chemother* 51, 324–331 (2007).
234. Zhang, X. F. *et al.* Tannic acid inhibited norovirus binding to HBGA receptors, a study of 50 Chinese medicinal herbs. *Bioorganic Med. Chem.* 20, 1616–1623 (2012).
235. Weichert, S. *et al.* Structural Basis for Norovirus Inhibition by Human Milk Oligosaccharides. *J Virol* 90, 4843–4848 (2016).
236. Zhang, X. F. *et al.* Inhibition of histo-blood group antigen binding as a novel strategy to block norovirus infections. *PLoS One* 8, 2013 (2013).
237. Lindesmith, L. C. *et al.* Broad Blockade Antibody Responses in Human Volunteers after Immunization with a Multivalent Norovirus VLP Candidate Vaccine: Immunological Analyses from a Phase I Clinical Trial. *PLoS Med.* 12, (2015).
238. Blazevic, V. *et al.* Development and maturation of norovirus antibodies in childhood. *Microbes Infect.* 18, 263–269 (2016).
239. Lindesmith, L. C. *et al.* Monoclonal Antibody-Based Antigenic Mapping of Norovirus GII.4-2002. *J Virol* 86, 873–883 (2012).
240. Garaicoechea, L. *et al.* Llama nanoantibodies with therapeutic potential against human norovirus diarrhea. *PLoS One* 10, (2015).
241. Koromyslova, A. D. & Hansman, G. S. Nanobodies targeting norovirus capsid reveal functional epitopes and potential mechanisms of neutralization. *PLoS Pathog.* 13, e1006636 (2017).
242. Lindesmith, L. C. *et al.* Particle Conformation Regulates Antibody Access to a Conserved GII.4 Norovirus Blockade Epitope. *J Virol* 88, 8826–8842 (2014).
243. Wegener, H. *et al.* Human norovirus GII.4 (MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology* 4, 1–11 (2017).
244. Hansman, G. S. *et al.* Crystal Structures of GII.10 and GII.12 Norovirus Protruding Domains in Complex with Histo-Blood Group Antigens Reveal Details for a Potential Site of Vulnerability. *J Virol* 85, 6687–6701 (2011).
245. Tan, M., Hegde, R. S. & Jiang, X. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *J Virol* 78, 6233–6242 (2004).
246. Tan, M. & Jiang, X. The p domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors. *J Virol* 79, 14017–14030 (2005).
247. Tan, M. *et al.* Terminal modifications of norovirus P domain resulted in a new type of subviral particles, the small P particles. *Virology* 410, 345–352 (2010).
248. Bereszczak, J. Z. *et al.* Structure, stability and dynamics of norovirus P domain derived protein complexes studied by native mass spectrometry. *J Struct Biol* 177, 273–282 (2012).
249. Li, Y. *et al.* Establishment of a novel method without sequence modification for developing NoV P particle-based chimeric vaccines. *Protein Expr Purif* 121, 73–80 (2016).
250. Tan, M., Meller, J. & Jiang, X. C-terminal arginine cluster is essential for receptor binding of norovirus capsid protein. *J Virol* 80, 7322–7331 (2006).
251. Sklenář, V. & Bax, A. Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* 74, 469–479 (1987).
252. Anglister, J., Grzesiek, S., Ren, H., Klee, C. B. & Bax, A. Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. *J. Biomol. NMR* 3, 121–126 (1993).
253. Blasi, E., Barluzzi, R., Bocchini, V., Mazzolla, R. & Bistoni, F. Immortalization of murine microglial cells by a v-raf/v-myc carrying retrovirus. *J. Neuroimmunol.* 27, 229–237 (1990).
254. Reed, L. J. & Muench, H. A simple method of estimating fifty per cent endpoints. *Am. Jounal Hyg.* 27, 493–497 (1938).

255. Ramakrishnan, M. A. Determination of 50% endpoint titer using a simple formula. *World J Virol* 5, 85 (2016).
256. Harder, E. *et al.* OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* 12, 281–296 (2016).
257. Howard, D. L. & Kjaergaard, H. G. Hydrogen bonding to divalent sulfur. *Phys Chem Chem Phys* 10, 4113–4118 (2008).
258. Skora, L., Mestan, J., Fabbro, D., Jahnke, W. & Grzesiek, S. NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. *Proc Natl Acad Sci U S A* 110, E4437-45 (2013).
259. Gioia, D., Bertazzo, M., Recanatini, M., Masetti, M. & Cavalli, A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules* 22, 2029 (2017).
260. Huang, N. & Jacobson, M. P. Binding-site assessment by virtual fragment screening. *PLoS One* 5, (2010).
261. Barelier, S. *et al.* Increasing chemical space coverage by combining empirical and computational fragment screens. *ACS Chem Biol* 9, 1528–1535 (2014).
262. Kumar, A., Voet, A. & Zhang, K. Y. J. Fragment Based Drug Design: From Experimental to Computational Approaches. *Curr. Med. Chem.* 19, 5128–5147 (2012).
263. Yuriev, E. & Ramsland Paul A. *Structural Glycobiology*. (2013).
264. Kadirvelraj, R., Foley, B. L., Dyekjær, J. D. & Woods, R. J. Involvement of water in carbohydrate-protein binding: Concanavalin A revisited. *J Am Chem Soc* 130, 16933–16942 (2008).
265. Clarke, C. *et al.* Involvement of Water in Carbohydrate - Protein Binding. *J Am Chem Soc* 12238–12247 (2001). doi:10.1021/ja004315q
266. Kim, M. & Cho, A. E. The role of water molecules in stereoselectivity of glucose/galactose-binding protein. *Sic Rep* 6, (2016).
267. Keseru, G. M. *et al.* Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *Journal of Medicinal Chemistry* 59, 8189–8206 (2016).
268. Hall, R. J., Mortenson, P. N. & Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* 116, 82–91 (2014).
269. Siegal, G., AB, E. & Schultz, J. Integration of fragment screening and library design. *Drug Discovery Today* 12, 1032–1039 (2007).
270. Davis, B. J. & Erlanson, D. A. Learning from our mistakes: The 'unknown knowns' in fragment screening. *Bioorganic and Medicinal Chemistry Letters* 23, 2844–2852 (2013).
271. Boyd, S. M. & De Kloe, G. E. Fragment library design: Efficiently hunting drugs in chemical space. *Drug Discov. Today Technol.* 7, (2010).
272. Morley, A. D. *et al.* Fragment-based hit identification: Thinking in 3D. *Drug Discovery Today* 18, 1221–1227 (2013).
273. Fuller, N. *et al.* An improved model for fragment-based lead generation at AstraZeneca. *Drug Discovery Today* 21, 1272–1283 (2016).
274. Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: Increasing saturation as an approach to improving clinical success. *J Med Chem* 52, 6752–6756 (2009).
275. Firth, N. C., Brown, N. & Blagg, J. Plane of best fit: A novel method to characterize the three-dimensionality of molecules. *J. Chem. Inf. Model.* 52, 2516–2525 (2012).
276. Sauer, W. H. B. & Schwarz, M. K. Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity. *J Chem Inf Comput Sci* 43, 987–1003 (2003).
277. Rishton, G. M. Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today* 2, 382–384 (1997).
278. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53, 2719–2740 (2010).

279. Baell, J. & Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* 513, 481–483 (2014).
280. Uchiyama, S., Santa, T., Okiyama, N., Fukushima, T. & Imai, K. Fluorogenic and fluorescent labeling reagents with a benzofurazan skeleton. *Biomedical Chromatography* 15, 295–318 (2001).
281. Metz, J. T., Huth, J. R. & Hajduk, P. J. Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. *J. Comput. Aided. Mol. Des.* 21, 139–144 (2007).
282. Xia, X., Maliski, E. G., Gallant, P. & Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J Med Chem* 47, 4463–4470 (2004).
283. Huth, J. R. *et al.* ALARM NMR: A rapid and robust experimental method to detect reactive false positives in biochemical screens. *J Am Chem Soc* 127, 217–224 (2005).
284. Feng, B. Y., Shelat, A., Doman, T. N., Guy, R. K. & Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* 1, 146–148 (2005).
285. McGovern, S. L., Helfand, B. T., Feng, B. & Shoichet, B. K. A specific mechanism of nonspecific inhibition. *J Med Chem* 46, 4265–4272 (2003).
286. Dalvit, C., Caronni, D., Mongelli, N., Veronesi, M. & Vulpetti, A. NMR-Based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening. *Curr Drug Discov Technol* 3, 115–124 (2006).
287. Hou, T. J., Xia, K., Zhang, W. & Xu, X. J. ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. *J Chem Inf Comput Sci* 44, 266–275 (2004).
288. Shoghi, E., Fuguet, E., Bosch, E. & Ràfols, C. Solubility-pH profiles of some acidic, basic and amphoteric drugs. *Eur. J Pharm Sci* 48, 291–300 (2013).
289. Domnick, A. Studies of murine norovirus host cell glycan interactions. *Master's thesis* (2016).
290. Liu, J. & Song, J. Insights into protein aggregation by NMR characterization of insoluble SH3 mutants solubilized in salt-free water. *PLoS One* 4, (2009).
291. Kastelic, M., Kalyuzhnyi, Y. V., Hribar-Lee, B., Dill, K. A. & Vlachy, V. Protein aggregation in salt solutions. *Proc Natl Acad Sci* 112, 6766–6770 (2015).
292. Charbonneau, M. E. *et al.* Small molecule deubiquitinase inhibitors promote macrophage anti-infective capacity. *PLoS One* 9, (2014).
293. Interlandi, G. & Thomas, W. E. Mechanism of allosteric propagation across a β-sheet structure investigated by molecular dynamics simulations. *Proteins Struct Funct Bioinforma* 84, 990–1008 (2016).
294. Tolonen, E. *et al.* Allosteric transition and binding of small molecule effectors causes curvature change in central ??-sheets of selected enzymes. *J. Mol. Model.* 17, 899–911 (2011).
295. Durrant, J. D., Friedman, A. J., Rogers, K. E. & McCammon, J. A. Comparing neural-network scoring functions and the state of the art: Applications to common library screening. *J. Chem. Inf. Model.* 53, 1726–1735 (2013).
296. Verdonk, M. L. *et al.* Docking performance of fragments and druglike compounds. *J Med Chem* 54, 5422–5431 (2011).
297. Halgren, T. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem* 47, 1750–1759 (2004).
298. Haselhorst, T., Weimar, T. & Peters, T. Molecular recognition of sialyl Lewisx and related saccharides by two lectins. *J Am Chem Soc* 123, 10705–10714 (2001).
299. Jones, M. K. *et al.* Human norovirus culture in B cells. *Nat Protoc* 10, 1939–1947 (2015).

Curriculum Vitae

Personal Details

Name	Tobias Schöne, Dr. rer. nat.
Address	Travelmannstr. 38/40, DE-23564 Lübeck
Contact	tobias.schoene@gmail.com
Date of birth	14.06.1988 (Schwerin)
Nationality	German



Education and Qualifications

12/2014 – 11/2017	**PhD thesis**, "Virus Receptor Identification *via* Click Chemistry and Fragment-based Drug Discovery using NMR Spectroscopy" Supervisor: Prof Dr. rer. nat. Thomas Peters, Institute of Chemistry, University of Lübeck, Germany (final grade: 1.1, magna cum laude)
10/2012 – 11/2014	**Master of Science** in Molecular Life Sciences, University of Lübeck, Germany (final grade: 1.4) Thesis: "Targeting West Nile Virus Infection – Screening for Allosteric Inhibitors Based on STD NMR Experiments " (Institute of Chemistry, Prof Dr. rer. nat. Thomas Peters)
10/2009 – 09/2012	**Bachelor of Science** in Molecular Life Science, University of Lübeck, Germany (final grade: 1.5) Thesis: "Analyses of activated sugar nucleotides in neuronal cells under hyperglycemic conditions *via* NMR spectroscopy and synthesis of per-*O* acetylated *N*-azidoacetylglucosamine" (Institute of Chemistry, Prof Dr. rer. nat. Thomas Peters)
07/2007 – 05/2009	**Military services**, Joint Support Service in Beelitz and administrator for civil-military-cooperation in Schwerin, Germany
1998 – 2007	**General higher education entrance qualification**, David-Franck Gymnasium, Sternberg, Germany (final grade: 1.8)

Personal and Scientific Skills

Languages	German (mother tongue), English (fluent)
IT	MS office, Adobe Illustrator, Maestro (Schrödinger), Instant JChem, Origin, ImageJ, Python programming, Matlab programming (moderate)
Methodologies	**Structural Biology:** Nuclear magnetic and surface plasmon resonance spectroscopy, crystallography (moderate), protein biosynthesis, cheminformatics, virtual docking/ screening **Virology and Cell Biology:** Animal and human cell culture, viral receptor identification**,** virus purification, DNA/ RNA cloning

Work and Teaching Experience

05/2016 – 02/2017	Supervision of Master thesis[1]
04/2016 – 09/2016	Supervision of Bachelor thesis[1]
2015 – 2017	Lecture assistant[1] for the practical course in a) Biophysical Chemistry, **Topic:** *In silico* molecular modelling b) Organic Chemistry, **Topic:** Protein determination
04/2012 & 07/2014	**Public Relations**, ScienceStation Lübeck, *Wissenschaft im Dialog*
01/2014 – 03/2014	Purification of Dengue virus serotype 2 (BSL3-conditions), practical course (master's module)[4]
09/2013 – 11/2013	Production and analysis of Bovine Virus Diarrhoe Virus (BVDV) clones (BSL2-conditions), practical course (master's module)[2]
04/2013 – 05/2013	Student assistant for the practical course in Biological Chemistry[1] **Topic:** Chemical synthesis
10/2011 – 06/2012	Tutor of a student exercise course for physics[3]
11/2011 – 12/2011	Student assistant for the practical course in Biophysical Chemistry, **Topic:** surface plasmon resonance and fluorescence spectroscopy
11/2010 – 08/2011	Student assistant for providing lecture notes and establishing of a student practical course in Biological Chemistry[1]
11/2010 – 07/2011	Student assistant for the practical course in Inorganic and Organic Chemistry[1]
10/2009	Two-week internship in the diagnostic laboratory MVZ West-Mecklenburg, Schwerin, Germany

[1] Institute of Chemistry, [2] Institute of Virology and Cell Biology,
[3] Institute of Medical Engineering, University of Lübeck
[4] School of Biological Science Nanyang Technological University, Singapore

Postgraduate Training

2015 – 2017	Courses at the center for internal further education, Lübeck a) Funding program EU 2020 b) PhD finished, what's next? c) Didactical training d) Rhetoric course, presentation and voice training
11/2015	Annual Bruker user meeting, Ettlingen, Germany
10/2014	GNMR summer school – Helmholtz-Center, Munich, Germany
04/2011 & 08/2011	**Exist-priME Cup** (master-cup and professional-cup), Management training and entrepreneur simulation

Scholarship and Student Engagement

07/2017	**Winner of poster presentation –** *Uni im Dialog,* University of Lübeck
09/2016 – 02/2017	**Member of the organizing committee** of the VIROCARB PhD symposium (3-day conference, 30 participants), Heinrich-Pette Institute, Hamburg, Germany
06/2012 – 11/2013	**Chairman of the organizing committee** of the 8th biomedical students' symposium (4-day conference, 250 participants), **Topic:** "From Infection to Therapy – Trends in Virology" University of Lübeck, Germany
10/2012	**Award for particular collegiate dedication**, University of Lübeck
12/2011 – 11/2014	**Scholarship** of the 'Studienförderwerk Klaus Murmann', Stiftung der Deutschen Wirtschaft, Berlin, Germany
09/2010 – 02/2014	**Member of the student council** for Molecular Life Science

International and Intercultural Experience

03/2017	Member of the delegation for the student exchange of East European medical students, Kishinev, Moldova and Czernowitz, Ukraine (Head: Prof. Dr. med. Karl-Friedrich Klotz, University of Lübeck)
08/2016	6-day International Conference on Magnetic Resonance in Biological Systems, Kyoto, Japan
01/2014 – 03/2014	Internship (master's module), School of Biological Science, Nanyang Technological University, Singapore
06/2009 – 08/2009	Environmental and social project organized by the wilderness committee in Vancouver, British Columbia, Canada

List of Publications

Poster:

Tobias Schöne, Alvaro Mallagaray, Ricarda Marschall, Fabienne Irrgang, Olaf Isken, Nobert Tautz, Thomas Peters (2017). *Rezeptorfischen mittels Klickchemie*. *Uni im Dialog*, University of Lübeck

Tobias Schöne, Lena Lisbeth Grimm, Thorsten Biet, Naoki Sakai, Linlin Zhang, Rolf Hilgenfeld, Thomas Peters (2016). STD NMR experiments identify a structural motif with novel second site activity against West Nile Virus NS2B-NS3^{pro} protease. 17th International Conference on Magnetic Resonance in Biological Systems.

Tobias Schöne, Lena Lisbeth Grimm, Thorsten Biet, Naoki Sakai, Linlin Zhang, Rolf Hilgenfeld, Thomas Peters (2016) STD NMR experiments identify fragments binding to a second site of West Nile Virus NS2B-NS3^{pro} protease. Discussion Meeting of the GDCh, Düsseldorf, Germany.

Tobias Schöne, Lena Lisbeth Grimm, Thorsten Biet, Naoki Sakai, Linlin Zhang, Rolf Hilgenfeld, Thomas Peters (2014). Targeting West Nile Virus Infection – Screening for Allosteric Inhibitors Based on STD NMR Experiments. Discussion Meeting of the GDCh, Berlin Germany.

Oral Presentations:

Tobias Schöne, Henrik Wegener, Stefan Taube and Thomas Peters (2017). STD NMR experiments identify fragments binding to murine and human norovirus P dimer. VIROCARB PhD symposium, Heinrich-Pette Institute, Hamburg, Germany

Tobias Schöne, Lena Lisbeth Grimm, Thorsten Biet, Naoki Sakai, Linlin Zhang, Rolf Hilgenfeld, Thomas Peters (2015). Targeting West Nile Virus Infection – Screening for Allosteric Inhibitors Based on STD NMR Experiments. SILVER progress reporting meeting, Leiden, Netherlands
Tobias Schöne, Lena Lisbeth Grimm, Thorsten Biet, Naoki Sakai, Linlin Zhang, Rolf Hilgenfeld, Thomas Peters (2015). Targeting West Nile Virus Infection – Screening for Allosteric Inhibitors Based on STD NMR Experiments. DZIF-Workshop, Next Generation Flaviviral Protease Inhibitors, Frankfurt am Main, Germany

Journals:

Tobias Schöne, Lena Lisbeth Grimm, Thorsten Biet, Naoki Sakai, Linlin Zhang, Rolf Hilgenfeld, Thomas Peters. STD NMR experiments identify a structural motif with novel second site activity against West Nile Virus NS2B-NS3^{pro} protease. Antiviral Research, 2017, 146:174-183

Henrik Wegener, Alvaro Mallagaray, **Tobias Schöne**, Julia Lockhauserbäumer, Hao Yan, Charlotte Uetrecht, Grant Hansman, Stefan Taube. Human Norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates. Glycobiology, 2017, 1-11