UNIVERSITÄT ZU LÜBECK

From the Institute of Chemistry and Metabolomics

of the University of Lübeck

Director: Prof. Dr. rer. nat. Thomas Peters

An NMR study into virus-host interactions of murine Norovirus capsid and SARS-CoV-2 spike proteins

Eine NMR Studie zu Virus-Wirt-Wechselwirkungen von murinen Norovirus Kapsid- und SARS-CoV-2 Spike-Proteinen

Dissertation

for Fulfillment of

Requirements

for the Doctoral Degree

of the University of Lübeck

from the Department of Natural Sciences

Submitted by

Thorben Maaß

from Rendsburg

Lübeck, 2022

First referee: Prof. Dr. Thomas Peters
Second referee: Prof. Dr. Christian Hübner
Date of oral examination: 31.05.2023
Approved for printing, Lübeck, 12.12.2023

Thorben Maaß

Contents

Note:
An electronical Supplementary Information (SI) file was submitted along with this thesis. The SI contains figures, tables, and Notes which are referred to as “Fig. S x”, “Tab. S x”, and “Note S x”, where “x” is a placeholder for the respective figure number, table number, and note number. A data repository containing all data presented in this thesis is maintained by Prof. Dr. Thomas Peters and can be accessed after justified demand.

Acknowledgements and Contributions

Zuallererst möchte ich mich aufrichtig und herzlich bei **Prof. Dr. Thomas Peters** bedanken. Ich bin froh, auf den in dieser Arbeit vorgestellten Projekten gearbeitet zu haben. Ich sehe die Möglichkeit hierzu nicht als selbstverständlich an. Außerdem hat mich die außerordentlich gute Betreuung im positiven Sinne mit geformt und ich habe das Gefühl, wirklich viel gelernt zu haben.

Dann möchte ich mich bei **Dr. Alvaro Mallagaray** und **Dr. Robert Creutznacher** bedanken. Beide sind herausragende Wissenschaftler, von denen ich ebenfalls viel gelernt habe und die mir stets als wissenschaftliche Vorbilder dienten.

Besonders möchte ich hier auch **Patrick König** hervorheben, mit dem ich unzählige wissenschaftliche und nicht-wissenschaftliche Diskussionen geführt habe. Danke!

Außerdem möchte ich mich bedanken bei **Dr. Ashok Kumar Rout**, **Dr. Friedemann Flügge**, **Evelyn Wachholz**, **Thomas Neubert**, **Dr. Thorsten Biet, Petra Lipp, Dr. Barbara Richer, Wilfried Hellebrandt** und **Prof. Dr. Karsten Seeger.** Sie standen mir stets mit Rat und Tat zur Seite und haben sowohl die Labore als auch die Geräte zuverlässig am Laufen gehalten.

Zudem möchte ich mich bei allen anderen Mitarbeitenden des Instituts für Chemie und Metabolomics bedanken. Die Zeit am Institut hat mir viel Freude bereitet und ich werde gerne auf diesen Abschnitt zurückblicken.

I also enjoyed the numerous and fruitful collaborations. I explicitly want to mention **Dr. Guido Hansen**, who gave me unique and valuable insights into X-ray crystallography, **Dr. Matrin Frank**, who provided MD simulation data and analysis of corona virus spike proteins, **Prof. Dr. Thomas Krey** and **Dr. George Liam Ssebyatika,** with whom I collaborated successfully in the SARS-CoV-2 glycan project, **Dr. Martin Empting** and **Erik Kosche** from the field of medicinal chemistry, **Prof. Dr. Stefan Taube**, **Maryna Chaika**, **Jan Knickmann**, **Dr. Carmen Mirabelli**, and **Miranda Lane** from the field of norovirus research, **Dr. Chris Waudby**, who provided guidance and software code I used for NMR line shape analysis, **Prof. Dr. Charlotte Utrecht** and **Dr. Jasmin Dülfer** from the mass-spectrometry field, and **Dr. Hauke Paulsen** and **Leon Torben Westermann**, who performed MD simulations of the P-domain and shared their results with me.

Eine weitere wertvolle Erfahrung war die Zusammenarbeit mit **Leon Torben Westermann**, **Marlene Brückner** und **Lea Kim Breckwoldt**, die als Bachelor-, Blockpraktium-, und Master-Studierende meine Zeit sowohl mit wissenschaftlichen als auch mit nicht-wissenschaftlichen Beiträgen bereichert haben. Ich bin zutiefst überzeugt, dass die drei, wohin sie auch gehen mögen, die Zukunft signifikant mitgestalten werden. Ihr Mitwirken zu meiner Arbeit ist in Tabelle i zusammengefasst.

Für die uneingeschränkte Unterstützung und vielen schönen Momente außerhalb der Welt der Wissenschaft möchte ich mich bei meinen Eltern **Ingrid** und **Carsten**, bei meinem Bruder **Christoph,** bei dem restlichen engsten Familienkreis und bei meinen Freunden bedanken.

Der letzte und größte Dank gilt meiner langjährigen Partnerin **Marie-Joann Lafrenz**.

Tab. i: Supervised contributions of students to this thesis.

Person	**contributed to**	**data shown in**
Leon Torben Westermann (Bachelor thesis)	NMR assignment of MNV P-domain	Fig. 5.2.1, Fig. 5.2.2, Fig. S 2.1, Fig. S 2.2 Fig. S 2.8
Leon Torben Westermann (Master internship)	Determination of MNV P-domain methyl group order parameters	Fig. 5.4.14, Fig. 5.5.6, Fig. 5.5.7, Fig. S 4.11, Fig. S 4.12, Fig. S 5.2
Lea Kim Breckwoldt (Bachelor thesis)	SARS-CoV-2 RBD protein experiments into ligand binding	Fig. 5.6.6, Fig. 5.6.9, Fig. S 6.3
Marlene Brückner (Bachelor thesis)	SARS-CoV-2 spike protein experiments into ligand binding	Fig. 5.6.2, Fig. 5.6.3, Fig. 5.6.6, Fig. 5.6.7, Fig. 5.6.8, Fig. 5.6.10, Fig. S 6.1, Fig. S 6.2, Fig. S 6.4

1 Abstract

Human Noroviruses are a leading cause for gastroenteritis and are responsible for more than 200,000 deaths per year. Despite 40 years of research, neither drugs nor vaccines are available. Due to the lack of cost-effective cell culture systems for human noroviruses, murine norovirus (MNV) has become an important model system. The P-domain protein of MNV is part of the viral capsid and interacts with the cellular receptor and neutralizing antibodies. Interactions of the P-domain with host factors such as metal ions, GCDCA, and pH appear to affect immune escape and receptor binding. In addition, GCDCA has been shown to cause capsid contraction and dimerization of the P-domain. Here, NMR spectroscopy is used to study interactions with host factors. In a first step, side-chain resonance assignment of the MNV P-domain was performed using nuclear Overhauser effects (NOEs) from a 4D HMQC-NOESY-HMQC experiment in combination with site-directed mutagenesis and pseudo contact shifts (PCSs). The assignment process was supported by AMIGO, a new algorithm developed for automatic assignment of resonances from NOEs and PCSs. The assignment allowed the analysis of line shapes in $^{1}H,^{13}C$ HMQC spectra of the P-domain. This revealed a dissociation constant K_D of 7 µM and an off-rate constant k_{off} of 1.3 s^{-1} for dimerization of the P-domain, and a K_D of 11 µM and a k_{off} of approximately 26 s^{-1} for binding of GCDCA. Plotting chemical shift perturbations (CSPs) on a structural model of the P-domain suggests allosteric structural rearrangements in the receptor and antibody binding epitopes upon GCDCA binding. In addition, two metal ion binding sites with mM affinity for Mg^{2+} and Ca^{2+} were identified. One of the sites is located at the receptor binding site, the other one at the G'H' loop. Binding of metal ions to the G'H' loop causes dimerization of the P-domain and results in globally distributed CSPs like those observed upon binding of GCDCA. Metal ion binding is slow on the NMR time scale. In contrast, in the presence of GCDCA, the exchange is fast, and, at the same time, the metal ion affinity shifts from the mM to the µM range. The observations reflect a "global protein transition" induced by GCDCA or by metal ions. A complete thermodynamic and partial kinetic description of P-domain dimerization, metal ion binding, and GCDCA binding is presented. It is shown that the P-domain exists almost exclusively in its dimeric form at acidic pH values between 4 and 5, while it dissociates into monomers at neutral pH values. Simultaneously, protonation of the G'H' loop at acidic pH values impedes metal ion binding to this site. Analysis of side chain order parameters and CSPs indicate the presence of allosteric network-like connections of amino acid side chain methyl groups throughout the P-domain. It is likely that this network is linked to the underlying mechanism causing the global effects observed for metal ion and GCDCA binding.

As of late 2019, SARS-CoV-2 has triggered a global pandemic, and researchers around the world have shifted their focus to better understand this newly emerging pathogen. The viral spike protein is of particular interest as it binds to the cellular receptor and interacts with neutralizing antibodies, much like the P-domain of MNV. Some studies have suggested a role of glycans as attachment factors. Here, STD NMR experiments with the Wuhan spike protein are used to provide clarity on selectivity and affinity of glycan binding. The sialoglycans 3'SL, 6'SL, and Sialyl-$Lewis^X$ bind to the Wuhan spike protein with dissociation constants in the mM range. The absence of STD effects for the Omicron BA1 spike protein indicates a loss of binding due to evolutionary pressure. In addition, binding of human blood group antigens (HBGAs) to the Wuhan spike protein can be excluded. Furthermore, the sialoglycan mimetics Oseltamivir and Zanamivir, as well as Lifitegrast, a drug used for treatment of dry eye, were found to bind to the Wuhan spike protein and may act as lead compounds for drug development.

2 Zusammenfassung

Humane Noroviren sind eine der Hauptursachen für Gastroenteritis und verantwortlich für mehr als 200.000 Todesfälle pro Jahr. Trotz mehr als 40 Jahren Forschung gibt es weder Medikamente noch Impfstoffe. Aufgrund des Mangels an kostengünstigen Zellkultursystemen hat sich das murine Norovirus (MNV) zu einem wichtigen Modellsystem für die Untersuchung von humanen Noroviren entwickelt. Das P-Domänen-Protein (P-Domäne), das Teil des MNV Kapsids ist, interagiert mit dem zellulären Rezeptor und neutralisierenden Antikörpern. Interaktionen der P-Domäne mit Wirtsfaktoren wie Metallionen, GCDCA und dem pH-Wert scheinen die Immunabwehr und die Rezeptorbindung zu beeinflussen. Darüber hinaus wurde gezeigt, dass GCDCA die Kontraktion des Kapsids und die Dimerisierung der P-Domäne bewirkt. Hier wird NMR-Spektroskopie eingesetzt, um Wechselwirkungen mit Wirtsfaktoren zu untersuchen. In einem ersten Schritt wurden Seitenkettenresonanzen der P-Domäne mithilfe von Kern-Overhauser-Effekten (NOEs) aus einem 4D-HMQC-NOESY-HMQC-Experiment, ortsgerichteter Mutagenese und Pseudokontaktverschiebungen (PCSs) zugeordnet. Der Zuordnungsprozess wurde von AMIGO unterstützt, einem neuen Algorithmus, der für die automatische Zuordnung von Resonanzen aus NOEs und PCSs entwickelt wurde. Die Zuordnung ermöglichte die Analyse von Linienformen in $^{1}H,^{13}C$ HMQC-Spektren der P-Domäne. Dies ergab eine Dissoziationskonstante K_D von 7 µM und eine Dissoziationsratenkonstante k_{off} von 1.3 s^{-1} für die Dimerisierung der P-Domäne sowie eine K_D von 11 µM und einen k_{off} von ca. 26 s-1 für die Bindung von GCDCA. Die Lokalisierung von *Chemical Shift Perturbations* (CSPs) auf einem Strukturmodell der P-Domäne deutet auf allosterische strukturelle Umlagerungen in den Rezeptor- und Antikörperbindungsepitopen hin. Darüber hinaus wurden zwei Metallionen-Bindungsstellen mit mM-Affinität für Mg^{2+}- und Ca^{2+}-Ionen identifiziert. Eine davon befindet sich an der Rezeptorbindungsstelle, die andere an der G'H'-Schleife. Die Bindung von Metallionen an die G'H'-Schleife bewirkt die Dimerisierung der P-Domäne und führt zu globalen CSPs, ähnlich denen, die bei der GCDCA Bindung beobachtet wurden. Die Metallionenbindung erfolgt langsam auf der NMR-Zeitskala. Im Gegensatz dazu erfolgt die Bindung in Gegenwart von GCDCA schnell. Gleichzeitig verschiebt sich die Affinität für die Metallionen vom mM in den µM-Bereich. Die Ergebnisse spiegeln eine globale „Proteinveränderung“ wider, die durch GCDCA oder durch Metallionen induziert werden kann. Es wird eine vollständige thermodynamische und teilweise kinetische Beschreibung der Dimerisierung der P-Domäne, der Metallionenbindung und der GCDCA-Bindung vorgelegt. Es wird ferner gezeigt, dass die P-Domäne bei sauren pH-Werten zwischen 4 und 5 fast ausschließlich in ihrer dimeren Form vorliegt, während sie bei neutralen pH-Werten überwiegend monomer ist. Gleichzeitig behindert die Protonierung der G'H'-Schleife bei sauren pH-Werten die Bindung von Metallionen an diese Stelle. Die Analyse von Ordnungsparametern und CSPs deutet auf das Vorhandensein von Verbindungen zwischen den Aminosäureseitenketten hin. Diese erstrecken sich über die gesamte P-Domäne und sind vergleichbar mit einem allosterischen Netzwerk. Es ist wahrscheinlich, dass dieses Netzwerk mit dem zugrundeliegenden Mechanismus für die allosterischen Effekte der Metallionen- und GCDCA-Bindung verknüpft ist.

Ende 2019 hat SARS-CoV-2 eine globale Pandemie ausgelöst, und Forscher auf der ganzen Welt haben ihren Fokus darauf gerichtet, den neu aufgetretenen Erreger besser zu verstehen. Von besonderem Interesse ist das Spike-Protein. Ähnlich wie die P-Domäne von MNV bindet es an den zellulären Rezeptor und interagiert mit neutralisierenden Antikörpern. Massenspektrometrische Experimente deuten auf eine mögliche Rolle von Glykanen als Bindungsfaktoren hin. Hier wurden STD-NMR-Experimente mit dem Wuhan-Spike-Protein verwendet, um Klarheit über die Selektivität und Affinität

der Glykanbindung zu schaffen. Es wurde festgestellt, dass die Sialoglykane 3'SL, 6'SL und Sialyl-Lewis^X mit Dissoziationskonstanten im mM-Bereich binden. Das Fehlen von STD-Effekten für das Omikron BA1-Spike-Protein deutet auf einen Verlust der Bindung aufgrund von evolutionärem Druck hin. Darüber hinaus kann eine Bindung menschlicher Blutgruppenantigene (HBGAs) an das Wuhan-Spike-Protein ausgeschlossen werden. Die Sialoglykan-Mimetika Oseltamivir und Zanamivir sowie Lifitegrast, ein Medikament zur Behandlung des trockenen Auges, binden an das Wuhan-Spike-Protein und können als Leitverbindungen für die Arzneimittelentwicklung dienen.

3 Introduction

Outbreaks of new viruses pose a serious and recurring threat to modern society. To fundamentally understand virus infections, it is crucial to study the underlying molecular interactions with atomic resolution. NMR spectroscopy is one of the few techniques that can provide such insights. It can be applied to decipher interactions of specific viral proteins with receptors, attachment factors, and small-molecule host factors. Such knowledge in turn can provide new avenues for antiviral drug development. This chapter provides a biological background on murine norovirus (MNV) and SARS-CoV-2. In addition, selected NMR spectroscopy methods are presented and discussed.

3.1 Interaction of murine norovirus capsids with host-factors

The genus norovirus belongs to the *Caliciviridae*, a family of non-enveloped positive-stranded RNA viruses. The genus covers at least six genogroups, which are subdivided into ≥30 genotypes (de Graaf et al., 2016). Viruses of genogroups GI, GII, and GIV infect humans (huNoV) and are one of the most common causes of gastroenteritis (Robilotti et al., 2015). The annual costs associated with norovirus infections are estimated to exceed 60 billion US dollars (Bartsch et al., 2016). In addition to the economic burden, norovirus infections cause over 200,000 deaths in children under five years of age each year. Developing countries are particularly affected (Pires et al., 2015). More than 40 years after the discovery of noroviruses (Adler et al., 1969), there are neither vaccines nor drugs for acute treatment (Arias et al., 2013). The development of cell culture systems constitutes a significant advance in understanding the requirements of huNoV infections (Jones et al., 2015; Ettayebi et al., 2016). However, the available methods are costly and labor-intensive, impeding wider application.

Viruses of the closely related genogroup GV, the murine noroviruses (MNV), infect mice. They share the intestinal tropism with their human counterparts. Due to the availability of cell culture systems and animal models, they became the model system for huNoV research (Vashist et al., 2009). Like huNoV, the MNV genome encodes three open reading frames (ORFs) (Glass et al., 2009; Vashist et al., 2009). ORF1 encodes a polyprotein which is processed into a protease, a helicase, and an RNA polymerase, among others. ORF2 and ORF3 code for the proteins VP1 and VP2. The MNV genome additionally contains an ORF that encodes the virulence factor VF1 (McFadden et al., 2011).

Of particular interest for infection mechanisms of MNV is the VP1 protein, which builds up the viral capsid (Fig. 3.1.1). Initial structural models were obtained in 2008 and 2010 using cryo-electron microscopy (cryo-EM) (Katpally et al., 2008; Katpally et al., 2010). They show a T=3 icosahedral capsid consisting of 180 VP1 protein copies. The VP1 protein can be subdivided into two domains. The shell domain (S-domain) forms the shell of the capsid. The protruding domain (P-domain) protrudes from the capsid, hovers over the capsid shell (Katpally et al., 2010), and interacts with various host factors (Sherman et al., 2021).

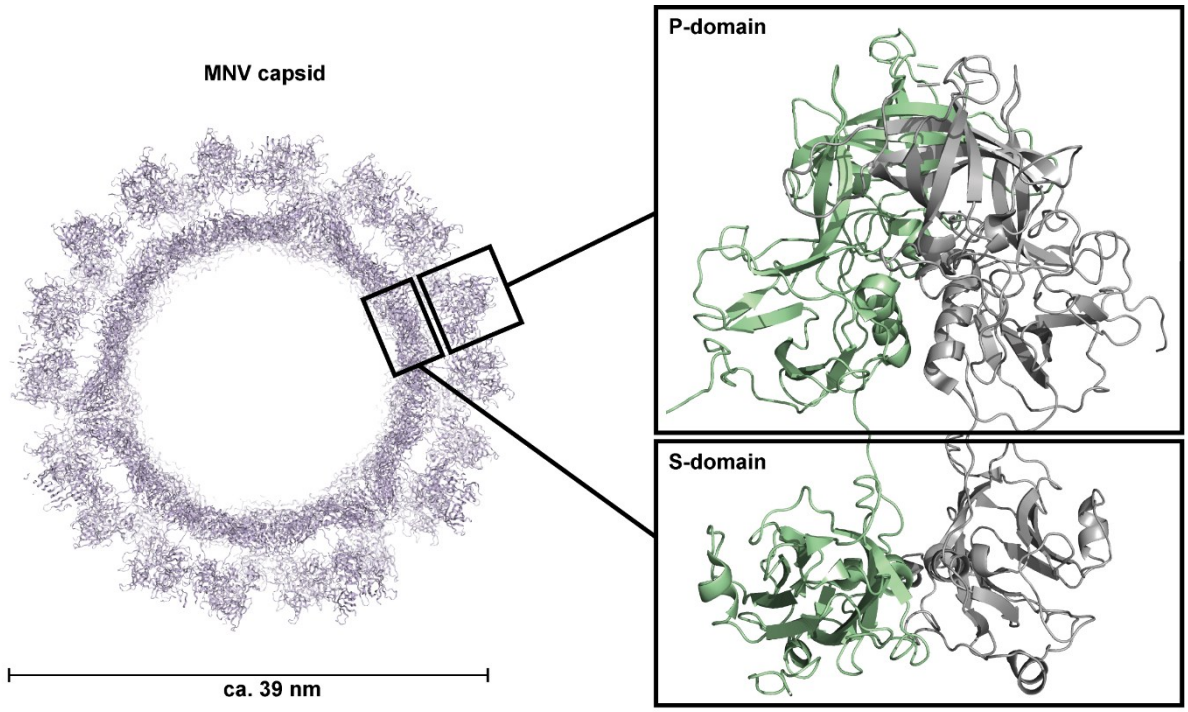


Fig. 3.1.1: Structural model of MNV capsid (pdb 6crj, Nelson et al., 2018). The T=3 icosahedral capsid is built up of 180 copies of the VP1 protein. The VP1 protein is shown on the right side and can be subdivided into the protruding domain (P-domain) and the shell domain (S-domain). The S-domain forms the capsid shell. The P-domain hovers above the shell.

A first crystal structure model of the apo P-domain was published in 2010 by Taube et al.. More recent studies yielded crystal structure models of the P-domain in complex with the CD300lf protein receptor and metal ions or in complex with the CD300lf receptor, metal ions, and the bile acid GCDCA (Nelson et al., 2018; Kilic et al., 2018). In contrast to the apo structure, these structures showed exposed loops at the tip of the P-domain in a different, *closed*, conformation (Fig. 3.1.2). At the time it was hypothesized that these loops would bind neutralizing antibodies (Kolawole et al., 2017). Indeed, a recent cryo-EM model of the capsid complexed with a neutralizing antibody confirmed this binding epitope (Williams et al., 2021a). Further cryo-EM studies of the MNV capsid revealed that GCDCA alone also brings the P-domain into the closed conformation (Sherman et al., 2019). The above-mentioned study by Nelson et al. (2018) additionally found that metal ions and GCDCA increase the affinity of the P-domain to the CD300lf receptor.

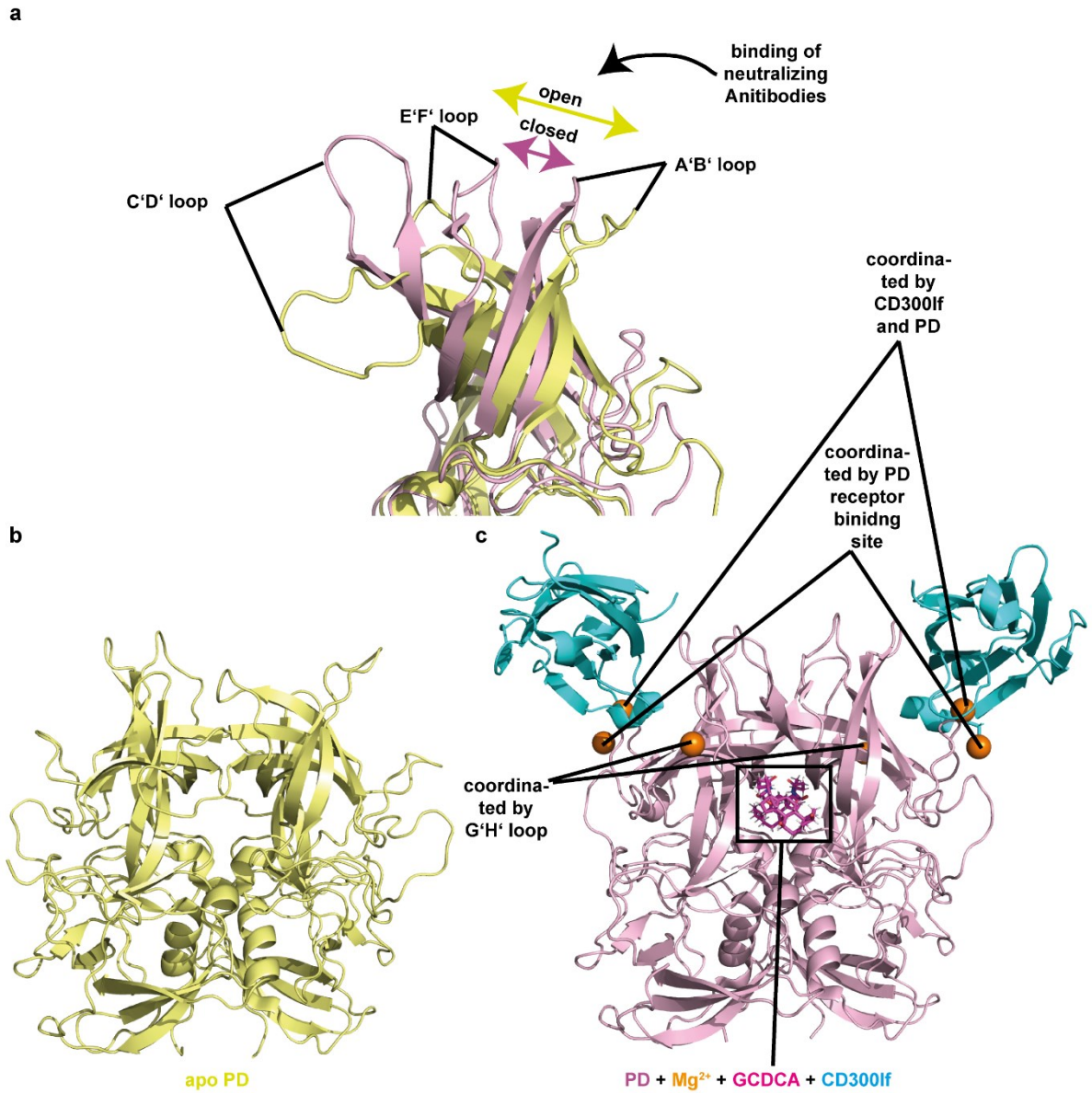


Fig. 3.1.2: MNV P-domain crystal structure models suggest a *closed* and an *open* conformation. (a) shows an overlay of P-domain (PD) crystal structure models in the apo state (b, pdb 3lq6, Taube et al., 2010) and in complex with CD300lf, Mg^{2+}, and GCDCA (c, pdb 6e47, Nelson et al., 2018). The *open* conformation binds neutralizing antibodies (Williams et al., 2021a; Creutznacher et al., 2022a). The *closed* conformation binds the CD300lf receptor (Nelson et al., 2018). The GCDCA binding pocket is formed by both amino acid chains of the dimeric P-domain. The P-domain dimer has four metal ion binding sites, two at the G'H' loops and two at the receptor-binding site. Both, P-domain and receptor, complex two further metal ions. It should be noted that the structural model of the P-domain in complex with metal ions, CD300lf, and GCDCA is nearly identical to models of the complex without GCDCA (pdb 6c6q, Nelson et al., 2018) or with only GCDCA but without the receptor and metal ions (Sherman et al., 2019).

In addition to inducing the closed conformation of the loops at the tip of the P-domain, GCDCA also triggers the capsid to adopt to a contracted conformation. Here, the P-domain is closely associated with the S-domain instead of floating above the capsid shell (from now on referred to as *contracted* vs. *expanded* conformations). This correlates with observations from NMR experiments and with results from size exclusion chromatography (Creutznacher, 2020), both showing that in solution the P-

domain is in an equilibrium of monomers and dimers with GCDCA shifting this equilibrium towards the dimeric state. Furthermore, functional assays revealed that GCDCA prevents the binding of monoclonal antibodies to MNV (Williams et al., 2021a; Creutznacher et al., 2022a). In another cryo-EM study, the MNV capsid was found to adopt a contracted conformation in the absence of GCDCA, similar to the contracted conformation previously observed in the presence of GCDCA (Snowden et al., 2020). The authors discussed the presence of metal ions as a possible reason but could not deduce an exact cause for their findings. In another cryo-EM study by Song et al., 2020, the capsid was found to be in the expanded conformation at pH values of 7 and 8 in the presence of EDTA. EDTA was added to chelate metal ions during preparation. At pH values of 6, 7, and 8 in the absence of EDTA, or at pH 6 in the presence of EDTA, the capsid was found to be contracted. The authors concluded that lower pH values cause the P-domains to settle down on the shell domain and that metal ions contribute to stabilize the contracted conformation. However, the contributions of the different factors were not further dissected. In addition, several kinds of metal ions were present during their study masking their individual contributions. The authors also speculate about another metal ion binding site between the P- and S-domains that might be responsible for the contraction. Yet another cryo-EM study compared capsid conformations at pH 5.0 and 7.4 (Williams et al., 2021b). Whereas the capsid was expanded at pH 7.4, it was contracted at pH 5.0. Furthermore, the P-domain adapted to the closed conformation at pH 5.0. This study also concluded that metal ions and a low pH were sufficient to stop binding of neutralizing antibodies to MNV, as previously described for GCDCA. The results on antibody binding, capsid contraction, and P-domain dimerization are summarized in Tab. 3.1.

Tab. 3.1: Summary of MNV P-domain and capsid properties.

Condition	P-domain Conformation (open vs. closed)	Capsid Conformation (contracted vs. expanded)	Antibody Binding (yes vs. no)	Receptor Affinity (normal vs. enhanced)	P-domain Dimerization (dimeric, monomeric, or mixture)
Neutral pH	open (Taube et al., 2010, Williams et al., 2021b)	expanded (Williams et al., 2021b)	yes (Williams et al., 2021b)	normal (Nelson et al., 2018)	?
GCDCA, Neutral pH	closed*** (Sherman et al., 2019)	contracted (Sherman et al., 2019)	no (Williams et al., 2021a; Creutznacher et al., 2022a)	?	dimeric (Creutznacher, 2020)
Metal Ions, Neutral pH	?**	contracted* (Song et al., 2020; Snowden et al., 2020)	no (Willams et al., 2021b)	enhanced (Nelson et al., 2018)	?
Metal Ions, GCDCA, and CD300lf or Metal Ions and CD300lf	closed (Nelson et all., 2018)	?	?	enhanced (Nelson et al., 2018)	?
Low pH	closed at pH 5.0 (Willams et al., 2021b)	contracted at pH 5.0 (Willams et al., 2021b)	no binding at pH 5.0 (Willams et al., 2021b)	?	mixture at pH 5.3 (Creutznacher, 2020)

*The authors do not link their observations exclusively to the presence of metal ions.

**The local resolutions of the cyro-EM structural models of Snowden et al., 2020, or Song et al., 2020, are significantly lower than 4 Å at the C'D', A'B', and E'F' loop. The authors do not provide information about the open or closed conformations.

*** C'D', A'B', and E'F' loops are "significantly distorted" and have "high B-values" (Sherman et al., 2019). Nevertheless, the authors conclude that the closed conformation is present.

3.2 SARS-CoV-2 spike protein interactions with glycans and small molecules

SARS-CoV-2 is an enveloped positive-strand RNA virus that was first described in early 2020 (Wu et al., 2020; Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). It is responsible for an ongoing global pandemic. First vaccines were approved by the U.S. Food and Drug Administration in December 2020 (Patel et al., 2022). Several monoclonal antibodies (mAbs) are available for acute treatment. However, mutations in variants of concern reduce vaccine-induced neutralizing antibody titers and can abolish neutralization with approved mAbs. The high manufacturing effort for large-scale mAb production is a further hurdle, preventing their broader application (Corti et al., 2021). This highlights the need for cost-effective small molecule inhibitors.

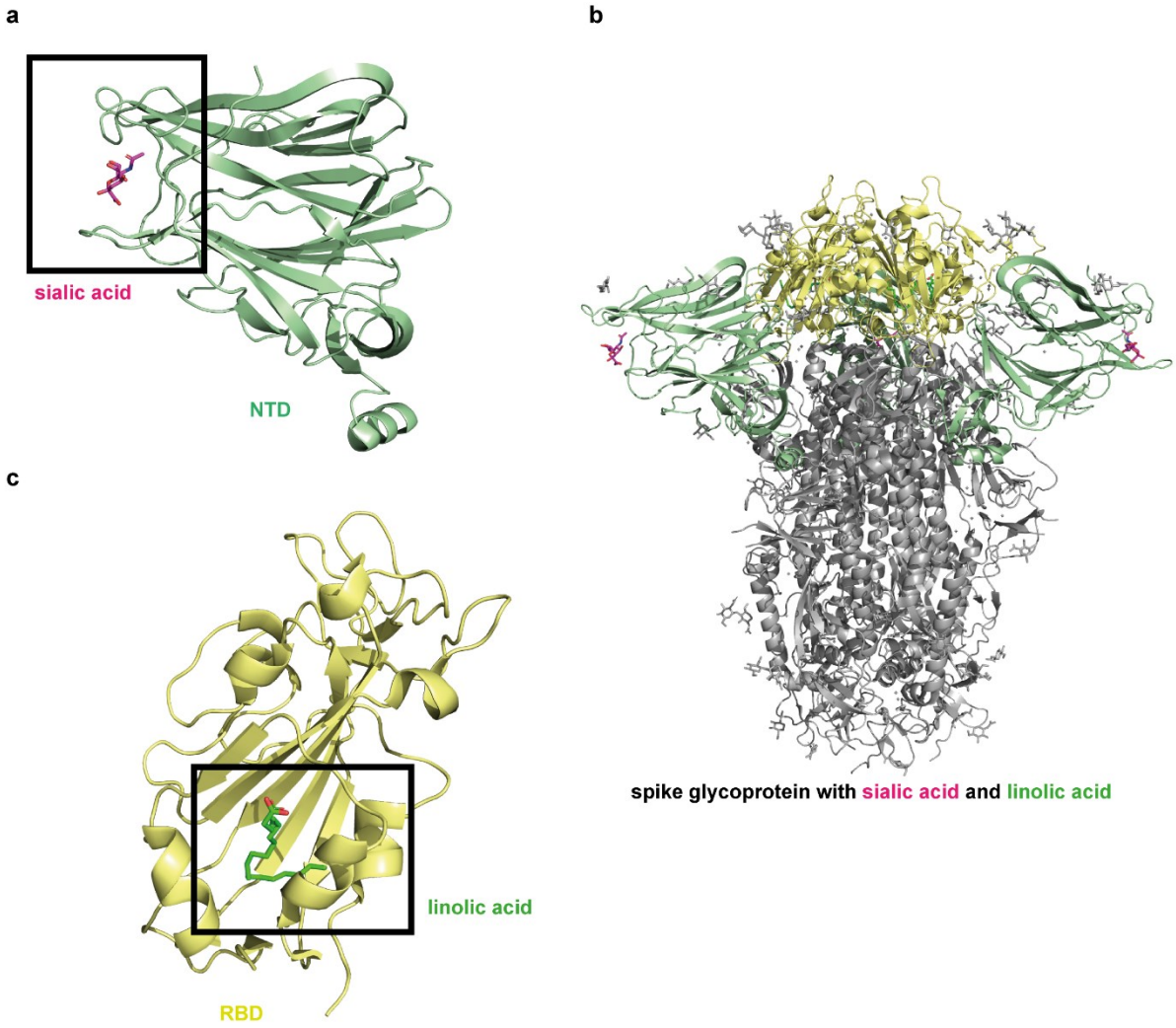


Fig. 3.2.1: Cryo-EM structural model of SARS-CoV-2 spike glycoprotein in complex with a sialic acid derivate and linolic acid. The NTD (a) of the spike protein (b, pdb 7qur, Buchanan et al., 2022) contains a sialic acid binding site. The RBD contains a fatty acid binding site (c).

The SARS-CoV-2 spike protein (Fig. 3.2.1b) is of particular interest as a drug target and for a deeper understanding of virus-host interactions. The receptor-binding domain (RBD) of the spike protein facilitates binding to the cellular receptor ACE2 (Shang et al., 2020; Jackson et al., 2022). It bears a well-conserved binding pocket for fatty acids (Fig. 3.2.1c), which may provide a promising target for small-molecule drugs (Toelzer et al., 2020). Recently, Lifitegrast, a drug used for the treatment of dry eye,

was found to bind to the fatty acid-binding pocket of the RBD (Creutznacher et al., 2022b). Interestingly, it also showed an antiviral effect in cell culture experiments (Day et al., 2021).

Next to fatty acid-binding, the binding of human blood group oligosaccharides (Wu et al., 2021; Nguyen et al., 2022) and sialoglycans (Li et al., 2021; Sun, 2021; Nguyen et al., 2022) to the RBD have been described in the literature. Mass spectrometry-derived dissociation constants lay in the micromolar range (Nguyen et al., 2022). However, another study employing saturation transfer difference (STD) NMR found that the N-terminal Domain (NTD) of the spike protein but not the RBD binds sialoglycans (Unione et al., 2022). This was further verified by a cryo-EM structure of the spike protein with a sialic acid derivate attached to the NTD (Fig. 3.2.1a; Buchanan et al., 2022). These findings call into question the mass spectrometry-derived results. A new approach for deriving dissociation constants from STD NMR experiments reported a dissociation constant in the low micromolar range for sialoglycan binding to the spike protein, suggesting an essential role in infection (Buchanan et al., 2022). An atomic force microscopy study (AFM) concluded a specificity for 9-O-acetylated sialic acids, claiming a dissociation constant in the low micromolar range (Petitjean et al., 2022). The authors conclude that non-9-O-acetylated sialic acids must have a rather low affinity. The discrepancy between the biophysical methods is apparent. Although there is qualitative evidence for an interaction between the NTD and sialogylcans, there is no consensus on its strength yet. Studies employing infection assays to investigate the potential role of sialoglycans came to different conclusions. One study found that sialic acids mildly impede the infection with SARS-CoV-2 (Chu et al., 2021). Two other studies reported slightly positive effects of sialoglycans on cell surfaces for infectivity (Saso et al., 2022; Nguyen et al., 2022). Interestingly, the NTD is targeted by several neutralizing antibodies (Chi et al., 2020; Wibmer et al., 2021), suggesting a functional role of the NTD (Jackson et al., 2022). In fact, the sialic acid-binding pocket identified in the cryo-EM study coincides with the so-called ‘NTD super site’, constituting the binding epitope for several neutralizing antibodies (Harvey et al., 2021; McCallum et al., 2021b; Buchanan et al., 2022). Especially variants of concern, including Omicron variants, bear mutations at this binding epitope and are less susceptible to neutralizing antibodies binding to this epitope (McCallum et al., 2021a; Willett et al., 2022; Cameroni et al., 2022).

3.3 Biomolecular solution-state nuclear magnetic resonance (NMR) spectroscopy

For general introductions into NMR spectroscopy, it is referred to textbooks by, e.g., Keeler, 2010, and Levitt, 2008. This chapter focuses on the aspects necessary for interpreting the results presented in this thesis.

Protein-based NMR experiments in solution are usually performed using isotopically labeled proteins. The experiments provide valuable information on structural features, protein-ligand interactions, protein-protein interactions, and protein dynamics (Williamson, 2013; Boswell and Latham, 2019). However, the development of protein labeling protocols is not always trivial (Mallagaray et al., 2019), and a resonance assignment is required. These factors are a bottleneck for the application of protein-based NMR techniques (Meyer and Peters, 2003). In contrast, ligand-based experiments can be performed without costly and labor-intensive labeling protocols but require large proteins and specific binding kinetics. They can provide information about protein-ligand interactions and ligand binding epitopes (Meyer and Peters, 2003).

3.3.1 Protein-based NMR

Different labeling strategies allow the use of different NMR techniques. Two techniques applied in this work are [U-^{2}H,^{15}N] or [U-^{15}N] labeling of the protein backbone and [^{1}H,^{13}C] labeling of protein methyl groups.

[U-^{2}H,^{15}N] or [U-^{15}N] labeled proteins carry ^{15}N instead of the naturally most abundant ^{14}N nuclei. [U-^{2}H,^{15}N] proteins are additionally perdeuterated to minimize the number of relaxation sources (Gardner and Kay, 1998). Amide deuterons are exchanged with protons by exposing the protein to H_2O-based buffers. However, this exchange is difficult to accomplish for larger proteins, and unfolding-refolding protocols may be required (Mallagaray et al., 2019).

[^{1}H,^{13}C] methyl group labeled proteins are usually perdeuterated and only methyl groups carry ^{1}H and ^{13}C nuclei (Gardner and Kay, 1998; Schütz and Sprangers, 2020). Methyl groups are attached to the ends of the corresponding amino acids giving them a high degree of flexibility (Sprangers et al., 2007; Boswell and Latham, 2019). In addition, methyl groups have three protons attached to one carbon atom that act as the origin of magnetization in the respective NMR experiments . In contrast, backbone amides carry only one proton, resulting in signals with a less favorable signal to noise ratio. All these properties contribute to obtaining sharp and strong signals for methyl groups, even in larger proteins (Sprangers et al., 2007; Boswell and Latham, 2019).

NMR experiments for heteronuclear chemical shift correlations in proteins

Backbone-labeled proteins or ^{13}C-methyl group-labeled proteins can be studied using 2D [^{1}H,^{15}N] TROSY HSQC experiments or 2D [^{1}H,^{13}C] HMQC experiments. Both experiments use a TROSY (transverse relaxation-optimized spectroscopy) effect, which allows studies of proteins with sizes well above 25 kDa (Pervushin et al., 1997; Tugarinov et al., 2003). In [^{1}H,^{15}N] TROSY HSQC spectra, [^{1}H,^{15}N] amides serve as *reporters* for their respective protein regions (Pervushin et al., 1997). In [^{1}H,^{13}C] HMQC experiments, [^{1}H,^{13}C] methyl groups are the reporters for the respective protein regions (Tugarinov et al., 2003). In the resulting spectra, the reporters appear as resonances with specific frequencies in each dimension, the *chemical shift*. The chemical shift depends on the chemical environment of the particular reporter (Williamson, 2013). For example, the chemical environment can be affected by changes in protein structure or binding of ligands, resulting in *chemical shift perturbations* (CSPs) (Williamson, 2013). More specifically, ^{15}N CSPs can, e.g., arise from rearrangements of the protein

backbone or the formation of hydrogen bonds to the amid or attached carbonyl group within the peptide bond (Williamson, 2013). ^{13}C methyl group CSPs of Ile, Val, Leu, and Met can among other things result from altered side-chain rotamer populations (Mulder, 2009; Butterfoss et al., 2010; Hansen et al., 2010; Hansen and Kay, 2011; Siemons et al., 2019). CSPs of Ala methyl groups directly correlate with the backbone conformation (Godoy-Ruiz et al., 2010; Pederson et al., 2017).

Identification and characterization of ligand binding using chemical shift perturbations

CSPs can be used to derive dissociation constants of protein-ligand interactions and to extract information about binding kinetics. The appearance of the CSPs depends on the *exchange regime*.

In the slow exchange regime, the difference in the chemical shifts Δν of two states, e.g., the apo protein and the ligand-bound protein, is much larger than the exchange rate k_{ex} (i.e., $k_{ex} << \Delta\nu$). The exchange rate is defined as shown in Eq. 1:

$$k_{ex} = k_{on} * [L] + k_{off}$$

(Eq. 1)

where k_{on} and k_{off} are the on-rate and off-rate of the binding reaction. [L] is the ligand concentration. In the slow exchange regime, a specific resonance can be observed for each state (Keeler, 2010). The intensities of the resonances can be used to quantify the population occupation of the two states. The populations can then be used to derive dissociation constants from the law of mass action (Williamson, 2013). One hurdle in the slow exchange regime is the transfer of the assignment from one state to the other. The assignment process is complex and time-consuming and is usually done for only one protein state. However, many authors use the nearest neighbor approach, in which a resonance of the unassigned state is assigned to the nearest resonances of the assigned state (Farmer et al., 1996; Lüttgen et al., 2002; Schoenle et al., 2021). This approach makes CSPs biased towards smaller values and underestimates ligand or conformational effects. It should also be noted that the resonance of the unassigned state can also be broadened beyond detection. However, this would likely also result from effects in close proximity to the reporter under scrutiny (e.g., backbone amide or methyl group, see above) and would therefore not lead to erroneous conclusions (Williamson, 2013).

In the fast exchange regime, k_{ex} is much larger than the difference in chemical shift between the two states (i.e., $k_{ex} >> \Delta\nu$). Here, the resonance in the spectrum appears at a population-weighted average between the chemical shifts of the apo state and bound state (Keeler, 2010). Tracking chemical shifts at different ligand concentrations thus allows a direct transfer of the assignments (Williamson, 2013). Furthermore, the occupancy of the states at specific ligand concentrations can be quantified if their chemical shifts are known. From this, dissociation constants can be derived from the law of mass action (Williamson, 2013). It should be noted that the absence of CSPs in the presence of a molecule of interest unambiguously proves that there is no binding at the given ligand concentration (Williamson, 2013).

Analyzing ligand binding kinetics using 2D line shape analysis with *TITAN*

Dissociation constants can also be determined from line shapes of resonances. Two-dimensional line shape analysis is superior to 1D line shape analysis as it overcomes the problem of signal overlap (Günther and Schaffhausen, 2002; Waudby et al., 2016). Line shape analysis is sensitive for exchange processes with k_{ex} ranging from the low microsecond to the high millisecond range (Waudby et al., 2016). It is especially effective in cases where apo-protein and bound-protein resonances are in

intermediate exchange, where CSP analysis cannot be applied (Williamson, 2013; Waudby et al., 2016). Furthermore, line shape analysis can reveal information about complex binding modes, such as multiple-step binding cooperativity (Arai et al., 2012).

In principle, the appearance of line shapes is, among other things, a function of the protein and ligand concentrations as well as the kinetics of the binding reaction (Günther and Schaffhausen, 2002). The entire evolution of magnetization and thus of line shapes can be described phenomenologically by the Bloch equations (Levitt, 2008). Combining the Bloch equations with chemical exchange terms yields the McConnell equations (McConnell, 1958). Reaction rates can be determined by fitting these equations to experimental NMR spectra. Waudby et al., 2016, developed *TITAN*, a software application that extends the analysis method. Assuming an underlying binding model, dissociation constants and off-rates can be obtained by fitting experimental NMR spectra to spectra simulated using a direct quantum mechanical approach. In these simulations, the complete evolution of magnetization is considered throughout the pulse sequence. Chemical exchange is accounted for by including a binding model-based exchange super operator and the McConnell equations. Detailed descriptions and a graphical user interface make *TITAN* a ready-to-use program.

Pseudo contact shifts

Paramagnetic *pseudo contact shifts* (PCSs) are changes of the chemical shift caused by hyperfine coupling of nuclear spins, e.g., of a protein, with spins of unpaired electrons of paramagnetic metals (Levitt, 2008). A detailed explanation can be found in reviews by Otting, 2010, and Clore and Iwahara, 2009. For the measurement of PCSs, the chemical shift of a protein resonance in the presence of a diamagnetic metal is subtracted from the chemical shift in the presence of a paramagnetic metal (Otting, 2010). This is important because next to PCSs, metal ion induced CSPs can also change the chemical shift (Otting, 2010).

The magnitude of the PCS depends on the length of the vector between the nucleus under scrutiny and the *paramagnetic center* as well as the angle Θ between this vector and the vector describing the external magnetic field. In isotropic solutions, the effect would be averaged out due to molecular tumbling. However, sufficient binding of certain paramagnetic metals to a protein leads to weak molecular alignment in the external magnetic field. This gives the solution an anisotropic character. The time-averaged alignment is described by the anisotropic part of the *magnetic susceptibility tensor* Δχ, which is directly related to the time-averaged Θ of all spins (Eq. 14-20 and 30 of Kramer et al., 2004; Orton et al., 2020). The location of the paramagnetic metal defines the paramagnetic center (Otting, 2010). In practise, Δχ and the paramagnetic centre can be determined with software packages such as *Paramagpy* (Orton et al., 2020) or *Numbat* (Schmitz et al., 2008) by using experimental PCSs and a structural model, e.g., that of a protein.

Consequently, PCSs directly encode structural information (Otting, 2010). In the simplest case, experimental PCSs can be compared to PCS theoretically obtained from a known Δχ, a known paramagnetic center, and different structural models. *Quality factors* (Q-factors, c.f. chapter 4.4.3) can then be used to test the agreement of a particular structural model with the structure in solution.

Assignment of methyl group labeled proteins

There are different approaches for the assignment of methyl group labeled proteins (Schütz and Sprangers, 2020). If no backbone assignment is available, the brute-force method is the systematic removal of individual methyl groups by mutagenesis (Amero et a., 2011). The corresponding spectra are compared to those of the wild-type protein. However, chemical shifts of neighboring methyl groups are often substantially perturbed, impeding interpretation (Sprangers and Kay, 2007). Historically, this approach has been advantageous for multi-domain complexes with a large molecular weight but a limited amount of spectrally different reporters (Sprangers et al., 2005). However, the method is costly, time-consuming, and labor-intensive.

A more robust approach is the *methyl-walk* strategy (Proudfoot et al., 2016). This method is based on nuclear Overhauser effects (NOEs) and a given high-resolution structural model. NOEs are usually obtained from 3D or 4D HMQC NOESY experiments and provide short-range distance restraints (Tugarinov et al., 2005; Wen et al., 2012). NOEs between methyl group resonances can be directly correlated with distances between methyl groups obtained from the structural model (Proudfoot et al., 2016). Once an initial assignment has been identified, it is possible to "walk" from methyl group to methyl group in the structural model and from methyl group resonance to methyl group resonance in the HMQC-NOESY spectrum. However, in NOE networks consisting of many methyl group resonances with distinct chemical shifts, the analysis becomes increasingly complex and time-consuming.

Paramagnetic NMR techniques such as the measurement of PCSs can be applied as a complementary method. In contrast to NOEs, PCSs provide long-range distance restraints. Spins more than 40 Å away from the paramagnetic center can exhibit PCSs (Otting, 2010). Comparing experimental PCSs with PCSs calculated from a known $\Delta\chi$, a known paramagnetic center, and a high-resolution structural model can validate existing assignments or extend the assignment (Flügge and Peters, 2018; Mühlberg et al., 2022).

The combination of PCSs with NOEs proved to be particularly advantageous for proteins with large methyl group networks, as shown in studies by Flügge and Peters, 2018, or Mühlberg et al., 2022. However, the manual analysis of NOEs and PCSs is time-consuming, and in the above studies, the data were analyzed using automated assignment protocols.

The algorithms developed for structure-based automatic assignment of protein methyl groups are summarized in the review by Pritišanac et al., 2020. They can be divided into NOE-based, paramagnetism-based, and hybrid approaches. The MAGMA (Pritišanac et al., 2017), MAGIC (Monneau et al., 2017), and MethylFLYA (Pritišanac et al., 2019) algorithms use distance restraints from NOEs. PRE-ASSIGN (Venditti et al., 2011), Possum (John et al., 2007), and PARAssign (Lescanne et al., 2017) are based on PCSs or PREs. However, each of the latter methods were only validated with the data set of a single protein. FLAMEnGO2.0 (Chao et al., 2012, 2014) and MAP-XSII (Xu and Matthews, 2013) can use both, NOE-based and paramagnetic restraints. Pritišanac et al., 2020, compare the performance of MethylFLYA, MAGMA, MAP-XSII, FLAMEnGO2.0, and MAGIC by testing them on five identical NOE data sets. All of them provided assignment results of sufficient quality. Nevertheless, none of them could provide 100 % assignment accuracy for all data sets. This underlines that the existing automated protocols must be applied with caution. The user should critically question the resulting assignments. However, the algorithms can represent a black box for the user. Then, due to the high complexity of the algorithms, it is often not easy to see through how the individual assignments were made.

Methyl group dynamics

Protein dynamics is another property that can be elucidated using the NMR toolbox. A range of NMR experiments can be utilized to cover different time windows (Boswell and Latham, 2019). One of the aims of this thesis is to assess dynamics on the pico- to nanoseconds time scale. In this time window side chain rotations and loop motions take place (Skjaerven et al., 2011; Boswell and Latham, 2019). The most commonly used parameter describing dynamics in this time window is the methyl group *order parameter* S^2. This parameter can be derived from relaxation rates of methyl groups (Boswell and Latham, 2019). It ranges between 0 and 1. A value of 0 corresponds to a methyl group symmetry axis moving within a large amplitude cone in the pico- to nanosecond time window (Fig. 3.3.1). A value of 1 corresponds to a methyl group with a symmetry axis moving within a narrow cone (Lee et al., 2000; Boswell and Latham, 2019). For [^{1}H,^{13}C] methyl group labeled proteins, S^2 order parameters can be obtained from "forbidden" double or triple quantum coherences (Tugarinov et al., 2007; Sun et al., 2011). Tugarinov and Kay, 2013, derived another approach based on the intensities of methyl group resonances in HMQC and HSQC spectra. In contrast to the previous methods, the acquisition of spectra required to derive all necessary parameters for the calculation of S^2 order parameters is comparatively fast. This makes the approach well applicable.

Interestingly, order parameters are directly linked to conformational entropy and can reveal, for example, entropy redistributions due to ligand binding (Frederick et al., 2007; Mittermaier and Kay, 2009). This relationship makes them a powerful tool for understanding complex molecular mechanisms such as entropy-driven allosteric networks (Capdevila et al., 2017).

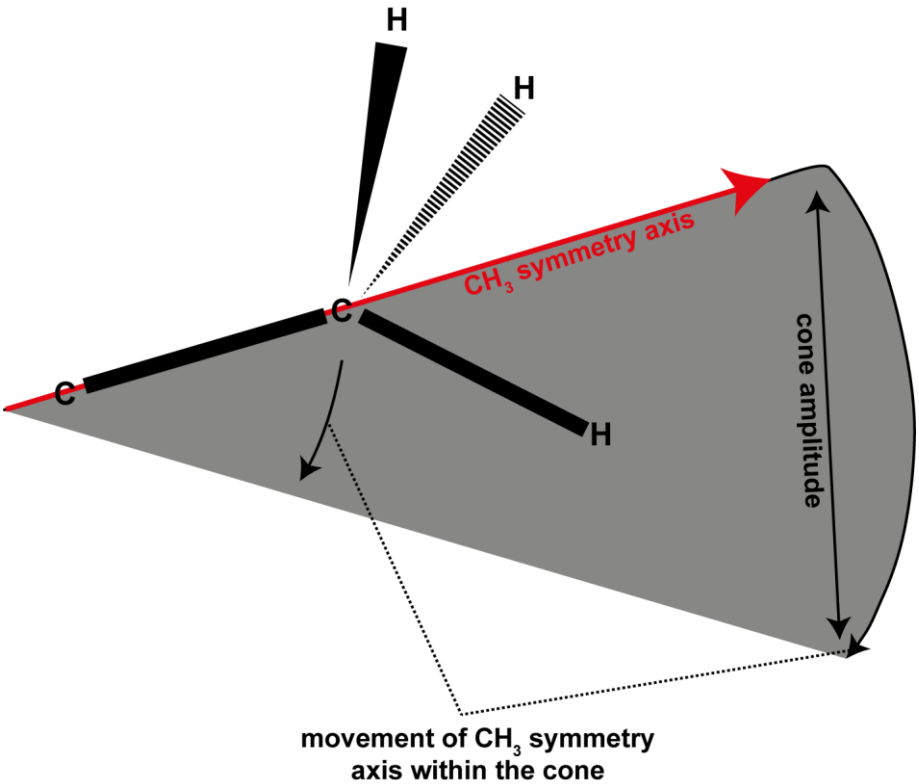


Fig. 3.3.1: The methyl group order parameter S2 describes the movement amplitude of the methyl group symmetry axis within a cone in the pico- to nanosecond time window. S^2 values are between 0, corresponding to large amplitude cones, and 1, corresponding to tight cones (Boswell and Latham, 2019).

3.3.2 Ligand-based NMR experiments using the saturation transfer difference (STD) experiment

STD NMR spectroscopy enables the characterization of ligand binding without isotope labeling of proteins. In the most straightforward approach, ligand binding epitopes and dissociations constants can be obtained from 1D proton STD spectra. Detailed descriptions of STD NMR experiments can be obtained from the literature (Mayer and Meyer, 1999; Meyer and Peters, 2003; Krishna and Jayalakshmi, 2008).

An STD spectrum results from the subtraction of an on-resonance spectrum from an off-resonance spectrum (Mayer and Meyer, 1999; Meyer and Peters, 2003). Protein ^{1}H resonances are more broadly distributed across the spectral window and also have substantially broader line widths than the ^{1}H resonances of small molecules. This property allows the selective irradiation of protein ^{1}H nuclei in samples containing both, proteins and small molecules. In the on-resonance experiment, protein nuclei are saturated by a selective on-resonance irradiation pulse. The saturation quickly spreads over the entire protein (spin diffusion). Spin diffusion originates from cross-relaxation and is therefore favored by the high molecular weights of proteins. If a small molecule binds to the protein, spin diffusion also spreads to ligand nuclei of ligands bound to the protein. Subsequently, remaining non-saturated spins are excited by a 90° pulse, and signal acquisition begins. In the resulting spectrum, only resonances belonging to non-saturated ^{1}H nuclei remain unaffected in their intensity.

In contrast, the off-resonance experiment involves irradiation at a frequency that does not saturate protein or ligand nuclei. All resonances are unaffected in their intensity in the resulting spectrum. Consequently, the difference spectrum, or STD spectrum, contains only resonance signals corresponding to nuclei of molecules acting as ligands.

The ratio of ligand signal intensities in the STD spectrum and the off-resonance spectrum can be used to quantify the STD effect (Meyer and Peters, 2003). It depends, among other factors, on distances to protein nuclei in the binding pocket. This results in an individual saturation profile for each ligand nucleus (Krishna and Jayalakshmi, 2008). Therefore, relative STD effects from a single spectrum can be used to characterize the ligand binding epitope (Mayer and Meyer, 2001). Other variables are ligand excess, the residence time of the ligand in the binding pocket, and the saturation time (Meyer and Peters, 2003; Angulo et al., 2010). The STD amplification factor can be calculated by multiplying the STD effect in percent by the ligand excess (Mayer and Meyer, 2001; Meyer and Peters, 2003). It provides an estimate of the fraction of ligand-bound protein. The experimental determination of STD amplification factors at different ligand concentrations can therefore be used to derive dissociation constants utilizing the law of mass action (Meyer and Peters, 2003; Neffe et al., 2007).

A limitation of the method is given by the residence time of the ligand in the binding pocket (Meyer and Peters, 2003). For very weakly binding molecules, the residence time is too short to allow sufficient saturation transfer. However, assuming a diffusion-controlled on-rate, dissociation constants in the mM range are still adequate to obtain STD effects. For strongly binding molecules, the residence time may become too long. This impedes signal amplification by multiple binding events to a single protein molecule within the saturation time. Therefore, it is challenging to detect ligand binding with dissociation constants below 100 pM. Another limitation is the molecular weight of the protein. It should be above 10 kDa since spin diffusion becomes more efficient at higher molecular weights (Meyer and Peters, 2003).

3.4 Objectives

For murine noroviruses (MNV), host factors such as metal ions, pH, and GCDCA appear to influence recognition of neutralizing antibodies and binding to the cellular receptor. GCDCA also triggers capsid contraction and dimerization of the P-domain, a protein domain protruding from the viral capsid. **The major objective of this study was to better understand the underlying molecular mechanisms of the interactions between the host factors and the P-domain.** A particular focus was placed on dissecting the effect of host factors on the dimerization of the P-domain. To this end, NMR spectroscopy experiments were identified as the main analytical tool to solve the scientific questions. As a first building block, a resonance assignment of the P-domain was necessary. Since all attempts to develop an unfolding-refolding protocol for the MNV P-domain had failed (Maaß, 2019), a backbone assignment was not considered. Instead, I focused on the assignment of [^{1}H,^{13}C] labeled side chain methyl groups. The following points had to be addressed:

(i) NMR assignment of the [^{1}H,^{13}C] methyl group labeled MNV P-domain using NOEs from a 4D HMQC-NOESY-HMQC experiment, PCSs, and site directed mutagenesis.

(ii) Development of an automated but comprehensible resonance assignment protocol for [^{1}H,^{13}C] methyl group labeled proteins and validation of the protocol with data sets from assigned proteins.

Based on the assignment, [^{1}H,^{13}C] HMQC or HSQC NMR spectra of the P-domain in the presence of different host factors were planned to deliver information on potential binding reactions and the dynamics of the P-domain. For the analysis of the spectra different methods were available:

(iii) 2D line shape analysis with *TITAN* (Waudby et al., 2016) to provide a thermodynamic and kinetic characterization of potential binding reactions.

(iv) Measurement of CSPs to localize host factor-induced effects on the P-domain and thermodynamically characterize the binding reactions.

(v) Determination of side order parameters to gain deeper insights on potential host factor binding effects on the conformational entropy of the P-domain.

Size exclusion chromatography had already been used to study the dimerization of the P-domain of MNV (Creutznacher, 2020) and was included into the portfolio of experimental methods. It was expected that the experiments would provide a mechanistic understanding of the effects of host factors on the MNV P-domain.

Glycans are discussed to act as host factors for SARS-CoV-2 infections. The spike glycoprotein of SARS-CoV-2 is located on the virus surface. The binding of glycans to the spike protein and, in particular, to the receptor binding domain (RBD) of the spike glycoprotein has been studied by several laboratories. However, the conclusions from these studies range from strong glycan binding to no glycan binding at all. These discrepancies call for further experiments into the binding of glycans to the spike glycoprotein and to the RBD. In addition, the spike protein represents an interesting antiviral target for the development of entry-inhibitors. I chose STD NMR experiments to shed light on these points. STD NMR experiments are ideally suited to study binding of small molecules to high molecular weight proteins such as the spike protein trimer. STD NMR experiments were planned with Wuhan and Omicron BA1 spike proteins to yield:

(i) Qualitative information on glycan or other small molecule binding to the respective spike proteins. This will also allow to draw conclusions on how the viral evolution affects glycan binding.

(ii) Ligand epitope mappings and thermodynamic characterization of binding glycans or small molecule ligands.

The RBD is the part of the spike protein that interacts with the cellular receptor. A recent study presented a protocol for backbone labeling of the RBD to yield an almost complete [^{1}H,^{15}N] backbone NMR assignment (Creutznacher, 2022b). This protocol lend itself well to test binding of glycans and other small molecules that could serve as lead compounds for drug development. Therefore, CSP fingerprint TROSY HSQC NMR experiments were included into the range of experiments to study small molecule binding to the RBD.

4 Material and Methods

4.1 AMIGO – an assignment tool for ^{13}C-methyl group labeled proteins

Input files for all data sets and the source code are stored on the server of the Institute of Chemistry and Metabolomics. They can be provided after justified demand. References to reference assignments for the validation and structural models used for running AMIGO are given in Tab S 1.1.

4.1.1 Testing data sets without additional structural restraints

To run AMIGO the user must select a range of cut-off distances for the different methyl group bearing amino acid types. For each data set, test runs were performed using only the 10 % "most unique" NOE graph building blocks as potential methyl walk starting points with cut-off distances ranging respectively from 3 Å to 7 Å to cut-off distances ranging from 3 Å to 11 Å in 0.2 Å steps. The weighting factor for NOEs (see Eq. 25) was set to 1. The weighting factors for additional restraints (see Eq. 26) were set to 0. For details on how to submit the respective parameters to AMIGO, see Fig. S 1.1. The files required to run AMIGO are shown in Fig. S 1.2. For the optimization of cut-off distances, the ratio of total assignments and assignments where the respective NOE graph building block exactly matched the structure graph building block was calculated for different rages of cut-off distances (Fig. S 1.3). The ratio is calculated by AMIGO and can be found in the "result.txt" output file. The range of cut-off distances that resulted in the highest ratio was used to perform the actual run. In this run all NOE graph building blocks were used as potential methyl walk starting points. For the validation, only methyl group clusters containing at least five methyl groups were considered. Assignments with no reference were considered unassigned (Fig. 5.1.11).

4.1.2 Testing data sets with additional structural restraints

See Figs. S 1.1 and S 1.2, and 4.1.1 for information on how to submit running parameters to AMIGO. The use of additional structural restraints such as PCSs makes assignments by AMIGO more robust to the choice of a range of cut-off distances (Fig. S 1.4). For the validation, a cut-off distance range from 3 Å to 9 Å in 0.2 Å steps was used (Fig. 5.1.11). All NOE graph building blocks were considered as potential methyl walk starting points. The weighting factor for NOEs was set to 1. The weighting factors for additional restraints (here PCSs) were set to 100. For the GTB data set the corresponding PCSs resulted from Ce^{3+}, Eu^{3+}, and Pr^{3+} (weighting factor 100, 100, 100, and 0 for *additional parameter 1*, *2*, *3*, and *4*; see Fig. S 1.1 iv). For MNV P-domain data set PCSs resulted from Ce^{3+}, Eu^{3+} and Sa^{3+} (weighting factor 100, 100, 100, and 0 for *additional parameter 1*, *2*, *3*, and *4*; see Fig. S 1.1 iv). For the LmUGP data set the PCSs resulted from Ce^{3+} and Eu^{3+} (weighting factor 0, 100, 100, and 0 for *additional parameter 1*, *2*, *3*, and *4*; see Fig. S 1.1 iv). Note that a weighting factor of 100 was used for PCSs so that the contributions of NOEs and PCSs to the *comparison score* (Eq. 28) are of equal magnitude. Clusters of less than five assigned methyl groups were only considered if the respective assignments were supported by additional restraints (here PCSs). Assignments of methyl groups with no reference were considered as unassigned.

4.1.3 Categorization of assignments for validation of AMIGO

To validate the assignments of AMIGO, the respective assignments were divided into four classes (c.f. 5.1.3). Correct assignments are assignments that are identical to reference assignments (c.f. Tab. S 1.1). Interchanged assignments correspond to methyl groups that have a measured NOE or an expected NOE (based on the final structure-based graph) to the actual methyl group. Thus, the methyl group is in close spatial vicinity to the real methyl group. Wrong assignments correspond to all other

assignments. Methyl groups for which no assignment was found and assignments of resonances without reference assignment were counted as unassigned. Only methyl groups that have NOEs were considered in the analysis.

4.2 Preparative laboratory methods

4.2.1 Media and chemicals

Tab. 4.1: Media

LB medium	10 g/L tryptone 5 g/L yeast extract 5 g/L NaCl H_2O (autoclaved)
TB medium	12 g/L tryptone 24 g/L yeast extract 54 g/L glycerol 0.55 g/L NaCl 3.3 g/L KH_2PO_4 16.63 g/L Na_2HPO_4 x 12 H_2O 1 g/L NH_4Cl (autoclaved)
M9 D_2O medium	5.2 g/L Na_2HPO_4 1.8 g/L KH_2PO_4 1 g/L NaCl 3 g/L NH_4Cl (or $^{15}NH_4Cl$ (Deutero), see chapter 4.2.4) 3 g/L D-glucose-d_7 (Deutero) 4 mM $MgSO_4$ 0.1 mM $CaCl_2$ 20 µg/ml vitamin B1 1 mM $MgCl_2$ 0.1 mg/L riboflavin 1 mg/L D-biotin 1 mg/L choline 1 mg/L folic acid D_2O (sterile filtrated)

M9 H_2O medium	5.2 g/L Na_2HPO_4
	1.8 g/L KH_2PO_4
	1 g/L NaCl
	3 g/L $^{15}NH_4Cl$
	3 g/L D-glucose
	4 mM $MgSO_4$
	0.1 mM $CaCl_2$
	20 µg/ml vitamin B1
	1 mM $MgCl_2$
	0.1 mg/L riboflavin
	1 mg/L D-biotin
	1 mg/L choline
	1 mg/L folic acid
	H_2O
	(sterile filtrated)

Tab. 4.2: Chemicals and Manufacturers.

Chemical	Manufacturers
Glycochenodeoxycholic acid (GCDCA)	Sigma-Aldrich
$CeCl_3$	Alfa Aesar
$EuCl_3$	Sigma-Aldrich
$LaCl_3$	Sigma-Aldrich
$SaCl_3$	Sigma-Aldrich
3'Sialyllactose trisaccharide with reducing end (3'SL)	Elicityl
6'Sialyllactose trisaccharide with reducing end (6'SL)	Elicityl
Sialyl-Lewis^X tetrasaccharide with reducing end (Sialyl-Lewis^X)	Elicityl
N-acetylneuraminic acid with reducing end (Neu5Ac)	Dextra Laboratories Ltd
9-O-acetyl-N-acetylneuraminic acid with reducing end (9-O-Ac-Neu5Ac)	Carbosynth Ltd
Oseltamivirphosphat	Sigma-Aldrich
Zanamivir	Sigma-Aldrich
Lifitegrast	AbMole BioScience
blood group A-tetrasaccharide Type I with reducing end (A type I)	Elicityl
blood group B-tetrasaccharide Type I with reducing end (B type I)	Elicityl
blood group H-disaccharide (H type)	in-house synthesis by Wilfried Hellebrandt

4.2.2 Bacterial Transformation and generation of cryo stocks

Generation of chemically competent *E. coli* BL21 (DE3) (Jeong et al., 2009) or DH5α cells was carried out by growing the respective cells overnight in LB-medium (Tab. 4.1) without ampicillin at 37 °C. 1 ml of the culture was used to inoculate 100 ml LB-medium without ampicillin. After reaching an optical density at 600 nm (OD_{600}) of 0.4, the new culture was incubated for 10 min on ice, and cells were harvested by centrifugation. The cell pellet was resuspended with ice-cold CM1 buffer (Tab. 4.3) and incubated for 20 min on ice. Cells were centrifuged, and the pellet was resuspended in 2 ml ice-cold CM2 buffer (Tab. 4.3). The bacteria suspension was aliquoted to 100 µl volumes and immediately stored at -80 °C.

LB agar plates were produced by adding 15 g agar to 1L LB medium and subsequent sterilization. The solution was cooled down to ca. 45 °C prior to addition of ampicillin to a final concentration of 100 µg/mL. Aliquots were distributed on Petri dishes and stored at 4° C.

Transformation was performed using aliquots of chemically competent *E. coli* BL21 (DE3) or DH5α cells. Cells were rapidly thawed and placed on ice. If not stated otherwise, 100 ng of plasmid DNA were added, and the cell suspension was incubated for 15 min on ice. After a heat-shock for 2 min at 42 °C, cells were incubated for 2 min on ice. 1 ml LB medium was added, and cells were grown for 1 h at 37 °C. 100 µL cell culture was spread on LB agar plates with ampicillin and cells were grown over night at 37 °C.

Bacterial cryo stocks were generated by inoculating LB medium containing 100 µg/ml ampicillin with the respective cells. The culture was grown until an OD_{600} of 1.5. 0.33 ml sterile 60 % v/v glycerol was added to 1 ml bacteria culture prior to storage at -80 °C.

Tab. 4.3: Buffers for generating chemically competent *E. coli* cells.

Buffer	Ingredients
CM1 buffer	10 mM sodium acetate
	50 mM $MnCl_2$
	5 mM NaCl
	(pH 5.6)
CM2 buffer	10 mM sodium acetate
	5 mM $MnCl_2$
	70 mM $CaCl_2$
	5 % v/v glycerol
	(pH 5.6)

4.2.3 Expression vectors and site directed mutagenesis

The pMAL-c2X expression plasmid (Walker et al., 2010) for MNV P-domain protein biosynthesis contained a gene for a fusion protein consisting of maltose-binding protein (MBP) and the GV CW1 murine norovirus VP1 P-domain (referred to as MNV P-domain, GenBank ID DQ285629, amino acids 228-530, N-terminal addition of GP peptide), separated by two His_8-tags and a HRV 3C protease cleavage site (LEVLFQGP) for protein purification.

The pMAL-c2X expression plasmid (Walker et al., 2010) for the RBD protein biosynthesis contained a gene for a fusion protein consisting of MBP and the SARS-CoV-2 spike protein receptor binding domain (referred to as RBD, GenBank ID YP_009724390.1, amino acids 319-527, N-terminal addition of GP peptide), separated by two His_8-tags and the HRV 3C cleavage site (see above).

pMAL-c2X expression plasmids additionally contained ampicillin resistance genes and isopropyl thiogalactoside-inducable (IPTG) *tac* promotors. GenBank (Benson et al., 2013) IDs were taken from www.ncbi.nlm.nih.gov on the 7th of November 2022.

Plasmids for expression and protein biosynthesis of MNV P-domain point mutants were generated by site-directed mutagenesis. The wild type plasmid was transformed into chemically competent *E. coli* DH5α cells. After transformation, cells from a single colony were grown over night at 37 °C in LB

medium (Tab. 4.1) with 100 µg/ml ampicillin. The wild type plasmid was then purified from *E. coli* DH5α cells to yield methylated template DNA. Plasmid purification was carried out using the NucleoSpin Plasmid (NoLid) kit (Macherey-Nagel) after instructions provided by the manufacturer. Ca. 50 ng template DNA was mixed on ice with 125 ng of the forward and of reverse primers designed for the respective point mutation (Tab. 4.4), 1 µl 10 mM dNTP mix (New England Biolabs), 5 µl 10x Pfu reaction buffer (Promega), 1 µl Pfu Polymerase (Promega) and sterile H_2O added to a final volume of 50 µl. Polymerase chain reaction (PCR) was carried out using an initial incubation time of 3 min at 95 °C. 40 s denaturation at 95 °C, 30 s annealing at 58 °C, and 18 min elongation at 72 °C were repeated 18 times. Final elongation was performed for 9 min at 72 °C. Afterwards, the mix was kept at 4 °C until digestion of methylated template DNA by addition of 1 µl 10 u/µl DpnI (Promega) for 1h at 37 °C. Afterwards, *E. coli* DH5α cells were transformed with 3 µl of the PCR products. The final plasmids were purified from single colonies as described above. The anticipated sequences were confirmed by commercial sequencing (LGC Genomics) using commercially available *malE* and *pTrcHis rev* sequencing primers (Eurofins).

Plasmids were transformed into *E. coli* BL21 (DE3) cells and stored as cryo stocks for protein expression.

Tab. 4.4: Primers for site-directed mutagenesis of MNV P-domain. All primers were commercially obtained (Eurofins). All mutants but the D410A and D440A mutant were used for the assignment of the MNV P-domain (c.f. chapter 5.2)

Mutant	Primer Sequence
A365G	Forward: 5'-GACGACCAATGGAGATCAAGCGCC-3'
	Reverse: 5'-GGCGCTTGATCTCCATTGGTCGTC-3'
A381G	Forward: 5'-GTGTGACAGCCGGAGCCTCTC-3'
	Reverse: 5'-GAGAGGCTCCGGCTGTCACAC-3'
A442G	Forward: 5'-CAGATTGATACTGGAGATGCAGCGGCG-3'
	Reverse: 5'-CGCCGCTGCATCTCCAGTATCAATCTG-3'
A444G	Forward: 5'-GATACTGCGGATGGAGCGGCG-3'
	Reverse: 5'-CGCCGCTCCATCCGCAGTATC-3'
A446G	Forward: 5'-GCGGATGCAGCGGGAGAAGCCATC-3'
	Reverse: 5'-GATGGCTTCTCCCGCTGCATCCGC-3'
I358L	Forward: 5'-CTTCGAGATGCTATTGGGCCCGACG-3'
	Reverse: 5'-CGTCGGGCCCAATAGCATCTCGAAG-3'
V352L	Forward: 5'-GATAAGCTGAAACTAACGACCTTCGAGATG-3'
	Reverse: 5'-CATCTCGAAGGTCGTTAGTTTCAGCTTATC-3'
I405L	Forward: 5'-GGTTTCAGGACACCCTACCCGAATACAAC-3'
	Reverse: 5'-GTTGTATTCGGGTAGGGTGTCCTGAAACC-3'
V304L	Forward: 5'-GAACGGGTGAACTAGCGACCTTTACCCTG-3'
	Reverse: 5'-CAGGGTAAAGGTCGCTAGTTCACCCGTTC-3'
V378L	Forward: 5'-GTGTTTGCCAGTCTAACAGCCGCAGCC-3'
	Reverse: 5'-GGCTGCGGCTGTTAGACTGGCAAACAC-3'
V387L	Forward: 5'-CTCTGGATCTGCTAGATGGCCG-3'
	Reverse: 5'-CGGCCATCTAGCAGATCCAGAG-3'
D410A	Forward: 5'-CGAATACAACGCTGGCCTGCTCG-3'
	Reverse: 5'-CGAGCAGGCCAGCGTTGTATTCG-3'

D440A	Forward: 5'-GCGCCAGATTGCTACTGCGGATG-3'
	Reverse: 5'-CATCCGCAGTAGCAATCTGGCGC-3'

4.2.4 Protein biosynthesis and purification

Unlabeled MNV P-domain protein

Cryo stocks of *E. coli* BL21 (DE3) cells carrying the pMAL-c2X plasmid encoding the MNV P-domain were used to inoculate 40 ml LB medium (Tab. 4.1) with 100 µg/ml ampicillin. Cells were grown over night at 37 °C. On the next day, TB medium (Tab. 4.1) with 100 µg/ml ampicillin was inoculated with the overnight culture resulting in an OD_{600} of 0.05. Cells were grown at 37 °C until an OD_{600} of 1.3. The temperature was decreased to 16 °C and after 1 h, IPTG was added to a final concentration of 0.1 mM to induce protein biosynthesis. On the next day, cells were harvested by centrifugation for 20 min with 5000 *g at 4 °C. Cell pellets were dried and stored at -80 °C until further processing.

[^{1}H,^{13}C] MILVA methyl group, [^{1}H,^{13}C] ILV methyl group, [^{1}H,^{13}C] MI*LVA methyl group or [U-^{2}H,^{15}N] and [^{1}H,^{13}C] MILVA methyl group labeled MNV P-domain proteins

For protein biosynthesis of [^{1}H,^{13}C] MILVA methyl group, [^{1}H,^{13}C] ILV methyl group or [^{1}H,^{13}C] MI*LVA methyl group labeled MNV P-domain proteins, cryo stocks of *E. coli* BL21 (DE3) cells carrying the pMAL-c2X plasmid encoding wild type or point mutant MNV P-domain proteins were used to inoculate 40 ml LB medium (Tab. 4.1) with 100 µg/ml ampicillin. Cells were grown over night at 37 °C. On the next day, 40 ml M9 D_2O medium with 100 µg/ml ampicillin was inoculated with a cell pellet obtained from the overnight culture resulting in an OD_{600} of 0.05. Cells were grown at 37 °C until an OD_{600} of 0.5. Cells were centrifuged at room temperature and resuspended in a volume of M9 D_2O medium with 100 µg/ml ampicillin corresponding to 4/5 of the final culture volume. The cells were grown at 37 °C until an OD_{600} of 0.75. A volume of M9 D_2O medium (Tab. 4.1) with 100 µg/ml ampicillin corresponding to 1/5 of the final culture volume was used to solve precursors listed in Tab. 4.5. The pH was adjusted to 7 and the solution was sterile filtrated. The temperature of the culture was decreased to 16 °C and the solution containing the precursors were added to the culture. Protein biosynthesis was induced after 1h by addition of 0.1 mM IPTG. Cells were grown until the stationary phase was reached, harvested by centrifugation for 20 min with 5000 *g at 4 °C and stored at -80 °C until further processing.

Protein biosynthesis of [U-^{2}H,^{15}N] and [^{1}H,^{13}C] MILVA methyl group labeled MNV P-domain proteins was carried out in a similar way. The only difference was the substitution of NH_4Cl with $^{15}NH_4Cl$ for M9 D_2O medium preparation.

Tab. 4.5: Final concentrations of precursors for [^{1}H,^{13}C] MILVA methyl group labelling.

Labelling scheme	Precursor	Concentration
L^{ProS} , V^{ProS} (L, V)	2-^{13}C-methyl-4-d_3-acetolactate (gift from Dr. Alvaro Mallagaray)	195 mg/L
I	2-ketobutyricacid-4-^{13}C-3, 3-d_2 (gift from Dr. Alvaro Mallagaray)	72 mg/L
I*	2-ketobutyricacid-$^{13}C_4$ (Sigma Aldrich)	72 mg/L
A	L-alanine-^{13}C-d_2 (Cambridge Isotope Laboratories)	0.6 g/L
	succinate-d_4 (Cambridge Isotope Laboratories)	3.75 g/L
M	L-methionine-(methyl-^{13}C) (Sigma Aldrich)	130 mg/L

Note that I* labelling leads to coupling in the ^{13}C dimension of [^{1}H,^{13}C] HMQC spectra. The labelling was used as the corresponding precursor was available in large quantities.

[U-^{15}N] labeled RBD protein

Cryo stocks of *E. coli* BL21 (DE3) cells carrying the pMAL-c2X plasmid encoding the RBD protein were used to inoculate 40 ml LB medium (Tab. 4.1) with 100 µg/ml ampicillin. Cells were grown over night at 37 °C. On the next day, M9 H_2O medium (Tab. 4.1) with 100 µg/ml ampicillin was inoculated with the overnight culture resulting in an OD_{600} of 0.1. Cells were grown at 37 °C. Protein biosynthesis was induced at an OD_{600} of 0.4 by addition of 1 mM IPTG. After cell growth reached the stationary phase, cells were harvested by centrifugation for 20 min with 5000 *g at 4 °C and stored at -80 °C until further processing.

Cell Lysis

All steps were performed on ice if not stated otherwise. Cell pellets for purification of MNV P-domain or RBD were resuspended in chilled sodium acetate-based lysis buffer or chilled Tris-based lysis buffer, respectively (Tab. 4.6). The suspension was incubated for 30 min at 4 °C with gentle shaking. Afterwards, cells were subjected to two lysis cycles using a high-pressure homogenizer (Thermo) applying a pressure of 1400 psi. Cell lysate for MNV P-domain purification were centrifuged for 90 min with 7000 *g at 4 °C. The supernatant was than subjected to further purification steps. Cell lysate for RBD purification was centrifuged for 60 min with 14000 *g at 4 °C. The supernatant was discarded and the RBD was purified from inclusion bodies.

Tab. 4.6: Cell lysis buffers.

Purification of	Buffer	Ingredients
MNV P-domain	acetate-based lysis buffer	20 mM sodium acetate 100 mM NaCl 10 mg/ L chicken egg white lysozyme (Merck Millipore) 160 U/L benzoase (Novagen) 0.48 mg/L leupeptin (Sigma) 0.48 mg/L aprotinin (Roth) (pH 5.3)
RBD	Tris-based lysis buffer	20 mM Tris 100 mM NaCl 10 mg/ L chicken egg white lysozyme (Merck Millipore) 160 U/L benzoase (Novagen) 48 mg/L leupeptin (Sigma) 48 mg/L aprotinin (Roth) (pH 8)

Solubilization and refolding of MBP RBD fusion protein

Information on buffer compositions can be obtained from Tab. 4.7. After cell lysis, pellets containing inclusion bodies were washed. Inclusion bodies containing pellets were resuspended in washing buffer. After 1h incubation under shaking at room temperature, the suspension was centrifuged for 1 h with 14000 *g at room temperature. The supernatant was discarded, and inclusion bodies were stored at -80 °C.

Pellets were thawed, solved in 25 ml solubilization buffer per 1 g of inclusion body, and incubated for 5 h under shaking at room temperature. The solution was flash-diluted 1:100 in cold refolding buffer and incubated over night at 4 °C prior to subjecting it to Ni-NTA chromatography.

Tab. 4.7: Buffers used for washing of inclusion bodies and solubilization and refolding of MBP RBD fusion protein.

Buffer	Ingredients
washing buffer	20 mM Tris 100 mM NaCl 1x BugBuster reagent from 10x concentrated stock (Merck Millipore) (pH 8)
solubilization buffer	20 mM Tris 100 mM NaCl 8 M urea 10 mM β-mercaptoethanol (pH 8.5)
refolding buffer	100 mM Tris 100 mM NaCl 5 mM reduced glutathione (Roth) 1 mM oxidized glutathione (Roth) (pH 8.5)

MNV P-domain and RBD purification

If not stated otherwise, information on buffer compositions can be obtained from Tab. 4.8. All chromatography steps were carried out using an ÄKTA pure system (GE Healthcare) at 4 °C. For MNV P-domain purification, the fusion protein was separated from the *E. coli* lysate employing two coupled 5 ml Ni-NTA columns (GE Healthcare). The columns were equilibrated with 2 column volumes acetate buffer. The lysate was loaded, and the columns were washed with 7 column volumes acetate buffer. A stepwise elution supplementing the acetate buffer with 10 % P-domain elution buffer for 10 column volumes was used to clear the columns from remaining *E. coli* proteins. Supplementation with 100 % P-domain elution buffer for 10 column volumes yielded the fusion protein.

Fractions containing the fusion proteins were pooled and filled up with acetate buffer to a final volume of 40 ml. 200 µl of 4 mg/ml His-tagged HRV 3C protease (gift from Dr. Robert Creutznacher) was added to cleave the fusion protein. The solution was dialyzed against 2 L acetate buffer with 5 mM β-mercaptoethanol over night at 4 °C with slight stirring. On the next day, a further purification step employing Ni-NTA chromatography was performed. The columns were equilibrated using two column volumes acetate buffer. The solution containing the His-tagged protease, the His-tagged MBP, and the MNV P-domain were loaded on the columns, with the P-domain not able to attach to the resin anymore. P-domain containing fractions were pooled and concentrated using 10 kDa centrifugal filters (Merck) to a final volume of 4 ml. Remaining proteins were eluted from the columns using stepwise gradients of 10% and 100% P-domain elution buffer as explained before.

For a last purification step, concentrated MNV P-domain proteins were subjected to size exclusion chromatography (SEC) employing a preparative Superdex 16/600 200 pg size exclusion column (GE Life Science). Acetate buffer was used for equilibration and as running buffer. The P-domain eluted from the column at an elution volume of ca. 80 to 90 ml.

For RBD purification, all steps were performed in a similar way with slight differences. Ni-NTA chromatography was carried out substituting acetate buffer and P-domain elution buffers with Tris imidazole buffer and RBD elution buffer, respectively. Dialysis was performed substituting acetate buffer containing 5 mM β-mercaptoethanol with RBD dialysis buffer. SEC was performed supplementing acetate buffer with RBD analysis buffer (Tab. 4.9).

Tab. 4.8: Buffers used for purification of MNV P-domain and RBD proteins.

Buffer	Ingredients
acetate buffer	20 mM sodium acetate 100 mM NaCl (pH 5.3)
P-domain elution buffer	20 mM Na_2HPO_4 100 mM NaCl 500 mM imidazole (pH 7.4)
Tris imidazole buffer	20 mM Tris 100 mM NaCl 20 mM imidazole (pH 8)
RBD elution buffer	20 mM Tris 100 mM NaCl 500 mM imidazole (pH 8)
RBD dialysis buffer	100 mM Tris 100 mM NaCl 1 mM reduced glutathione (Roth) 0.2 mM oxidized glutathione (Roth) (pH 8.5)

4.3 Analytical laboratory methods

4.3.1 Sample preparation

Information on buffer compositions, molecular weights, and molecular extinction coefficients are given in Tab. 4.9 if not stated otherwise. Sample pH values for H_2O-based buffers were determined using an Orion 8220BNWP pH-electrode (ThermoFisher) with an Orion Star A221 pH-meter (ThermoFisher). The electrode was calibrated with H_2O-based pH 4, 7 and 10 standard solutions (Roth) prior to measurements. For D_2O-based buffers, the obtained pH values (pH*) were converted into pH_{corr} values using Eq. 2 (Krezel and Bal, 2004):

$$pH_{corr} = 0.929 * pH^* + 0.41$$

(Eq. 2)

Note that in the literature, mostly pH* values are given for D_2O-based solutions. Using pH_{corr} values instead of pH* values serve the comparability to other studies carried out in H_2O. For solutions constituting out of a H_2O/D_2O mixture, the pH* is given.

Wuhan and Omicron BA1 SARS-CoV-2 spike proteins (sequences sown in Figs. S 6.5 and 6.6) were a gift from Dr. George Liam Ssebyatika, Institute of Biochemistry, University of Lübeck. [U-^{2}H,^{15}N] labeled RBD proteins (GenBank ID YP_009724390.1, amino acids 319-527, N-terminal addition of GP peptide) were a gift from Dr. Robert Creutznacher (spectra shown in Fig. 5.6.6c).

MNV P-domain, RBD, and spike proteins were concentrated using 10 kDa centrifugal filters (Merck Millipore). Buffer exchange into the proteins' respective analysis buffers (Tab. 4.9 if not stated otherwise) was performed using ZebraTM Spin Desalting columns (MWCO 40 kDa, Thermo Scientific) after the manufacturer's instructions or 10 kDa centrifugal filters (Merck, Millipore). If not stated otherwise, concentrations were determined using UV absorption at 280 nm measured with an UV spectrometer (Peqlab) and the respective molar extinction coefficients and molecular weights. Potential ligand molecules or metal ions were titrated to the protein samples from highly concentrated stocks in the same buffer with carefully adjusted pH, or pH_{corr}, or pH* values.

Tab. 4.9: Analysis buffers and molar extinction coefficients of MNV P-domain, RBD and spike proteins.

Protein (analysis method)	Buffer	Molar extinction coefficient and molecular weight
unlabeled P-domain (analytical SEC)	20 mM sodium acetate 100 mM NaCl (pH 5.3)	46.870 $M^{-1}cm^{-1}$ and 33.2 kDa
Isotope labeled* MNV P-domain protein (NMR)	20 mM sodium acetate-d_4 (Cambridge Isotope Laboratories) 100 mM NaCl (pH_{corr} 5.3)	46.870 $M^{-1}cm^{-1}$ and 33.2 kDa
Isotope labeled** RBD proteins (NMR)	75 mM Na_2PO_4 100 mM NaCl (pH* 6.2)	33.850 $M^{-1}cm^{-1}$ and 23.8 kDa
Spike proteins (NMR)	20 mM BisTris-d_{19} (Cambridge Isotope Laboratories) 100 mM NaCl (pH_{corr} 7.3)	140.960 $M^{-1}cm^{-1}$ and 139.6 kDa

*[^{1}H,^{13}C] MILVA methyl group, [^{1}H,^{13}C] MI*LVA methyl group or [U-^{2}H,^{15}N] and [^{1}H,^{13}C] MILVA methyl group labeled MNV P-domain proteins

**[U-^{2}H,^{15}N] or [U-^{15}N] labeled RBD proteins

4.3.2 Analytical size exclusion chromatography (SEC)

Analytical SEC was carried out on an ÄKTA pure system (GE Healthcare). Buffers are listed in Tab. 4.10. Protein solutions with P-domain concentrations listed in Tab. 4.10 were subjected to a Superdex 75 Increase 3.2/300 column (GE Life Sciences) pre-equilibrated in the respective buffers using a 10 µL sample loop by filling the entire loop with the protein solution. UV absorption was measured at 280 nm. The SEC run was performed at 4 °C employing a flow rate of 0.075 ml/min.

Tab. 4.10: Buffers used for analytical SEC.

Equilibration and running buffer	MNV P-domain concentration in µM	Chromatograms shown in
20 mM sodium acetate, 100 mM NaCl (pH 5.3)	5	Fig. 5.4.2b, green
20 mM sodium acetate, 100 mM NaCl, 5 mM $MgCl_2$ (pH 5.3)	5	Fig. 5.4.2b, blue
20 mM sodium acetate, 100 mM NaCl, 25 mM $MgCl_2$ (pH 5.3)	5	Fig. 5.4.2b, magenta
20 mM citric acid, 100 mM NaCl (pH 4)	21	Fig. 5.5.1b, red
20 mM citric acid, 100 mM NaCl (pH 6.2)	21	Fig. 5.5.1b, black

4.3.3 NMR spectroscopy

All pulse programs can be found in the data repository of this thesis. If not stated otherwise, all spectra were analysed and processed using *TopSpin* v3.6 (Bruker).

[^{1}H, ^{13}C] HMQC experiments (pulse program *metr-hmqcph_new.tb*) with [^{1}H,^{13}C] MILVA, [^{1}H,^{13}C] ILV or [^{1}H,^{13}C] MI*LVA methyl group labeled MNV P-domain proteins were carried out using 512 increments, 0.8 ppm as center of the spectral windows, and a sweep width of 3.5 ppm in the direct dimension. The respective parameters for the indirect dimension were 256 data points, 17 ppm and 18 ppm. The number of scans was either 4 or 8. The relaxation delay was set to 1.5 s. Further parameters are specified in the respective figure legends.

The 4D HMQC NOESY HMQC spectrum (pulse program *hmqcnoesyhmqc4d.tb*, Tugarinov et al., 2005) of [^{1}H,^{13}C] MILVA methyl group labeled MNV P-domain protein was acquired employing 30 % non-uniform-sampling (NUS) using a Poisson gap schedule (Hyberts et al., 2013, 2014) with 86 (^{13}C) x 84 (^{1}H) x 52 (^{13}C) increments in the indirect dimensions. 512 increments were used in the direct dimension. For optimization of the mixing time, 2D HMQC-NOESY spectra (pulse program *hmqcetgpno*) with 512 and 256 increments in the direct and indirect dimensions were acquired using mixing times of 50 ms, 100 ms, 200 ms, 300 ms, and 400 ms. Intensities of 13 randomly selected cross peaks were plotted against the mixing time (Fig. 3.6a in Westermann, 2020), yielding an optimized mixing time of 120 ms for the 4D HMQC-NOESY-HMQC experiment. The 4D spectrum was reconstructed using *TopSpin* v4.0.2's built-in recursive MDD (recursive-dimensional-decomposition) algorithm (Bruker).

STD spectra (pulse program *stddiffesgp.3*) of the unlabeled spike protein were acquired using an off-resonance frequency of 200 ppm and on-resonance frequency of -4 ppm. Saturation times were 2 s using a cascade of 50 ms Gaussian pulses separated by 1 ms. The relaxation delay was set to 5 s and excitation sculpting was used for suppression of the water signal. The spin lock filter included in the pulse program was set to 0 ms. Number of scans varied between 320 and 1920.

TRACT experiments (pulse programs *15n1h-tract-alpha* and *15n1h-tract-beta*, Lee et al., 2006) were carried out using [U-^{2}H,^{15}N] and [^{1}H,^{13}C] MILVA methyl group labeled MNV P-domain proteins. Buffers and protein concentrations are given in Tab. 4.11. The relaxation delay was set to 2s. 16 scans were measured for P-domain at a pH* of 5.3. 64 scans were measured for P-domain at s pH* of 4.6. The pulse sequence was repeated with 25 increasing delays up to 0.5 s (see respective figures).

Tab. 4.11: Sample conditions of TRACT experiments.

Sample	Shown in
140 µM P-domain, 20 mM soudium acetate, 100 mM NaCl, 30 mM $CaCl_2$, 10 % D_2O (pH* 5.3)	Fig. S 4.12
133 µM P-domain, 20 mM soudium acetate, 100 mM NaCl, 30 mM $CaCl_2$, 500 µM GCDCA, 10 % D_2O (pH* 5.3)	Fig. S 4.12
81 µM P-domain, 20 mM citric acid, 100 mM NaCl, 10% D_2O (pH*4.6)	Fig. 5.5.6
177 µM P-domain, 20 mM citric acid, 100 mM NaCl, 500 µM GCDCA, 10% D_2O (pH*4.6)	Fig. 5.5.6

Adapted versions of [^{1}H,^{13}C] HMQC (pulse program *hmqcph.amph4*) and HSQC (pulse program *hsqcph*) experiments and HMQC experiments with an additional delay to obtain R^S_{MQ} relaxation rates for determination of methyl group order parameters (Tugarinov and Kay, 2013, pulse program *2020.R.mq.am.tb.V2*) were carried out using [^{1}H,^{13}C] MILVA methyl group labeled MNV P-domain proteins. Sample compositions are given in Tab. 4.12. The relaxation delay was set to 1.5 s. The sweep width and spectral window centers were set to 3.75 ppm and 0.75 ppm for the direct dimension and to 18 ppm and to 17.5 ppm for the indirect dimension. For both dimensions, 512 increments were acquired. The number of scans was 8 except of the sample containing $CaCl_2$ and GCDCA, where 80 scans were acquired for HMQC and HSQC spectra and 32 for HMQC experiments with additional delays.

Tab. 4.12: Sample conditions for [^{1}H,^{13}C] HMQC and HSQC spectra used for order parameter determination.

Sample	Shown in
143 µM P-domain, 20 mM soudium acetate, 100 mM NaCl, 29 mM $CaCl_2$ (pH_{corr} 5.3)	Fig. S 4.11 and Fig. 5.4.14
136 µM P-domain, 20 mM soudium acetate, 100 mM NaCl, 29 mM $CaCl_2$, 500 µM GCDCA (pH_{corr} 5.3)	Fig. S 4.11 and Fig. 5.4.14
180 µM P-domain, 20 mM citric acid, 100 mM NaCl (pH_{corr} 4.6)	Fig. S 5.2 and Fig. 5.5.7
171 µM P-domain, 20 mM citric acid, 100 mM NaCl, 500 µM GCDCA (pH_{corr} 4.6)	Fig. S 5.2 and Fig. 5.5.7

[^{1}H,^{15}N] TROSY experiments (pule program *trosyetf3gpsi*) with [U-^{2}H,^{15}N] or [U-^{15}N] labeled RBD proteins were acquired with 2048 and 128 increments in the direct and indirect dimension for [U-^{15}N]-

labeled RBD and with 2048 and 256 for [U-^{2}H,^{15}N]-labeled RBD. Centers of the spectral window were set to 4.7 and 117 ppm and sweep widths were set to 16 ppm and 33 ppm, respectively. The relaxation delay was set to 1.5 s. Number of scans were 32 for [U-^{15}N]-labeled RBD and 24 for [U-^{2}H,^{15}N]-labeled RBD.

4.4 Data analysis

Corresponding *python* scripts for curve fitting and order parameter calculations are provided in the data repository of this thesis. Chemical shifts and peak intensities for chemical shift perturbation (CSP) mapping, fitting of order parameters, fitting of dissociation constants, fitting of apparent ionization constants, and determination of pseudo contact shifts (PCSs) were extracted using *CcpNmr Analysis* v2.4.2 software (Vranken et al., 2005). Intensities of STD NMR spectra were obtained using *TopSpin* v3.6 (Bruker).

4.4.1 STD NMR

STD amplification factors (STD-Af) were calculated according to Eq. 3:

$$STD - Af = \frac{I_{off} - I_{on}}{I_{off}} * \frac{L_t}{P_t}$$

(Eq. 3)

I_{off} and I_{on} are intensities in off-resonance and on-resonance spectra. L_t/P_t is the ligand excess.

For epitope mapping, all protons were normalized to the proton experiencing the highest STD-Af of the respective ligand.

4.4.2 Chemical shift perturbations (CSPs)

CSPs in Hz $\Delta\nu_{eucl}$ between resonances in [^{1}H, ^{13}C] HMQC spectra at different conditions are calculated using:

$$\Delta\nu_{eucl} = \sqrt{(\Delta\delta_H * \omega)^2 + (\Delta\delta_C * \omega * \gamma_C/\gamma_H)^2}$$

(Eq. 4)

with $\Delta\delta_H$ and $\Delta\delta_C$ being the chemical shift differences in the respective dimensions in ppm. γ_H and γ_C are the gyromagnetic ratios of the respective nuclei. ω is the spectrometer's frequency for ^{1}H nuclei (i.e. 600 MHz for all spectra shown in this thesis).

4.4.3 Pseudo contact shifts (PCSs)

^{1}H PCSs δ^{PCS} in ppm between certain resonances in [^{1}H,^{13}C] HMQC spectra of diamagnetic and paramagnetic samples were calculated using:

$$\delta^{PCS} = \delta(^1H)_{para} - \delta(^1H)_{dia}$$

(Eq. 5)

With $\delta(^1H)_{para}$ and $\delta(^1H)_{dia}$ being the ^{1}H chemical shifts under paramagnetic and diamagnetic conditions in ppm.

Magnetic susceptibility tensor parameters and paramagnetic centers of anisotropic samples were fitted using the *Paramagpy* software package (Orton et al., 2020) according to equations 6 to 8.

$$\delta^{PCS} = \frac{1}{4\pi r^5} * [x^2 - z^2, y^2 - z^2, 2xy, 2xz, 2yz] * \begin{bmatrix} \Delta\chi_{xx} \\ \Delta\chi_{yy} \\ \Delta\chi_{xy} \\ \Delta\chi_{xz} \\ \Delta\chi_{yz} \end{bmatrix}$$

(Eq. 6)

r is the distance between the paramagnetic center and the nucleus of interest. x, y, and z are the cartesian coordinates of the paramagnetic center in the protein frame. $\Delta\chi_{kk}$ with k ϵ [x,y,z] characterizes the anisotropic part of the magnetic susceptibility tensor Δχ.

Coordinates of the respective nuclei for calculation of r were obtained from crystal structure models, as described in the results part. Corresponding self-written *python* scripts are provided in the data repositary of this thesis. x, y, z, and $\Delta\chi_{xx}$, $\Delta\chi_{yy}$, $\Delta\chi_{xy}$, $\Delta\chi_{xz}$, and $\Delta\chi_{yz}$ were fitted using experimentally derived ^{1}H PCSs according to Eq. 6. Uncertainties in x, y, and z were obtained using 1000 bootstrap iterations sampling 90 % of the data. In a second fitting round, x, y, and z were held constant on the values derived before and the fitting was repeated to derive values for $\Delta\chi_{xx}$, $\Delta\chi_{yy}$, $\Delta\chi_{xy}$, $\Delta\chi_{xz}$, and $\Delta\chi_{yz}$. The axial and rhombic components of the magnetic susceptibility tensor can be calculated from the obtained parameters using Eqs. 7 and 8.

$$\Delta\chi_{ax} = \Delta\chi_{zz} - \frac{1}{2}(\Delta\chi_{xx} - \Delta\chi_{yy})$$

(Eq. 7)

$$\Delta\chi_{rh} = \Delta\chi_{xx} - \Delta\chi_{yy}$$

(Eq. 8)

The orientation of Δχ's frame to the protein frame is characterized by the Euler angles α, β, and γ. Uncertainties in $\Delta\chi_{ax}$, $\Delta\chi_{rh,}$ α, β, and γ were determined using 1000 bootstrap iterations sampling 90 % of the data with x, y, and z hold constant as described above.

Quality factors (Q-factors) were derived employing calculated PCSs $\delta^{PCS,calc}$ based on the above mentioned best fit's parameters and experimental PCSs $\delta^{PCS,exp}$ using:

$$Q = \sqrt{\frac{\sum_i \left(\delta_i^{PCS,exp} - \delta_i^{PCS,cal}\right)^2}{\sum_i \left(\delta_i^{PCS,exp}\right)^2}}$$

(Eq. 9)

The index i is for summation over all spins.

Lines of best fit for correlation plots between experimental and calculated PCSs or experimental PCSs at different conditions were obtained using the linear regression tool implemented in *LibreOffice* v6.4.7.2.

4.4.4 2D line shape analysis

2D line shape analysis of [^{1}H,^{13}C] HMQC spectra was performed using the program *TITAN* (Waudby et al., 2016). Spectra were processed using *nmrPipe* (Delaglio et al., 1995). All *nmrPipe* shell scripts are provided in the data repository of this thesis. Parameter uncertainties were obtained by 100 bootstrap iterations using *TITAN*'s error estimation tools. Additional information can be obtained from the respective figures in chapter 5.

Analysis of MNV P-domain dimerization at a pH_{corr} of 5.3 as shown in chapter 5.3.2 or at pH_{corr} of 7.4 as shown in chapter 5.5.1 were carried out using *TITAN's* built-in protein dimerization model and spectra of 13 µM, 25 µM, 50 µM, 75 µM, 100 µM, and 226 µM MNV P-domain at pH_{corr} of 5.3 or spectra of 13 µM, 25 µM, 50 µM, 75 µM, 100 µM, 150 µM, and 230 µM MNV P-domain at pH_{corr} of 7.4. For the analysis of spectra at pH_{corr} of 5.3, chemical shifts and line widths of the monomeric P-domain (M) were fitted using spectra of 12.5 µM and 25 µM MNV P-domain samples. For the analysis of spectra at a pH_{corr} of 5.3, chemical shifts and line widths of dimeric P-domain (D) were fitted using spectra of 226 µM and 100 µM P-domain samples. In a last step, chemical shifts were hold constant on the before derived values. Line widths, the off-rate $k_{off;Dimerization}$, and the dissociation constant $K_{D;Dimerization}$ were fitted using spectra of all different P-domain concentrations with the line widths derived above as starting conditions. For the analysis of spectra at pH_{corr} of 7.4, chemical shifts and line widths of the monomeric P-domain (M) were fitted using the spectrum of the 13 µM P-domain sample. Chemical shifts and line widths of dimeric P-domain (D) were fitted using the spectrum of the 230 µM P-domain sample. The last fitting step was performed as for the spectra at a pH_{corr} of 5.3 with the only difference being that chemical shifts of the dimeric state were also fitted and not hold constant during the second step. Chemical shifts derived before served as starting conditions.

MNV P-domain GCDCA binding analysis as shown in 5.3.3 was performed using a *TITAN* extension implemented by Dr. Christopher A. Waudby, Institute of Structural and Molecular Biology, University College London, to account for the three or five state binding models used here (Creutznacher et al., 2022a). Spectra of P-domain in the presence of 0 µM, 25 µM, 50 µM, 75 µM, 125 µM, 250 µM, and 550 µM GCDCA were used for the analysis. Line widths and chemical shifts of P-domain monomers (M state) and dimers (D state) were fitted using the apo P-domain spectrum. Line widths and chemical shifts of the GCDCA bound P-domain (DL stare) were fitted using the spectrum of the P-domain sample containing the maximal GCDCA concentration. For cooperativity models, chemical shifts of additional states (D* and DL* states) were set to arbitrary positions. For a final fit, chemical shifts were hold constant on the values derived before. $K_{D;Dimerization}$ and $k_{off;Dimerization}$ were set to the values obtained before (see above and chapter 5.3.2) and hold constant. Line widths, dissociation constants, off-rates, and cooperativity parameters α and β (c.f. chapter 5.3.3) were fitted. The line widths values obtained before were used as start conditions for the fit. Analysis of GCDCA binding in citric acid buffer as explained in chapter 5.5.2 was carried out in a similar way applying the three-state model and using spectra of P-domain in presence of 0 µM, 25 µM, 50 µM, 75 µM, 125 µM, 180 µM, 300 µM, and 400 µM GCDCA.

P-domain metal ion binding analysis as described in 5.4.5 or 5.5.2 was performed employing the three-state binding model used for the GCDCA analysis (see above). Spectra of D410A MNV P-domain in presence of 0 mM, 2.5 mM, 5 mM, 10 mM, 15 mM, 20 mM, and 25 mM $MgCl_2$ or in presence of 0 mM, 5 mM, 10 mM, 15 mM, 20 mM, and 25 mM $CaCl_2$ or spectra of MNV P-domain at a pH_{corr} of 7.4 in presence of 0 mM, 0.5 mM, 3 mM, 5 mM, 7.5 mM, 12.5 mM, and 20 mM $CaCl_2$ were used. Chemical shifts and line widths of the monomer and dimer states were fitted using the spectra of the apo P-

domain. Chemical shifts and line widths of the metal ion bound state were fitted using spectra with the maximal metal ion concentration. In the final run, all spectra were used to fit chemical shifts of the bound states with the values derived before as staring values. Additionally, line widths of all states, off-rates, and dissociation constants were fitted. Chemical shifts of the monomeric and dimeric states were kept constant at the values derived before. $K_{D;Dimerization}$ and $k_{off;Dimerization}$ were kept constant to the values obtained before for the corresponding pH_{corr} values (see above, chapter 5.3.2, and chapter 5.5.1). All line widths and chemical shifts of the bound states obtained before were used as start conditions for the fit.

Analysis of metal ion bound P-domain GCDCA binding as shown in chapter 5.4.5 or GCDCA binding to the apo P-domain at a pH_{corr} of 4.4 as shown in chapter 5.5.2 were carried out using *TITAN's* built-in two step binding model. P-domain spectra in the presence of 25 mM $MgCl_2$ and 0 µM, 40 µM, 80 µM, 150 µM, and 300 µM GCDCA or P-domain spectra in presence of 25 mM $CaCl_2$ and 0 µM, 20 µM, 40 µM, 80 µM, 150 µM, and 300 µM GCDCA or spectra of P-domain at a pH_{corr} of 4.4 in presence of 0 µM, 30 µM, 60 µM, 90 µM, 150 µM, 210 µM, and 400 µM GCDCA were used for the analysis. For analysis of metal ion bound P-domain GCDCA binding or GCDCA binding at a pH_{corr} of 4.4, P-domain spectra in presence of 0 µM GCDCA were used to fit line widths and chemical shifts of the metal ion bound P-domain or the dimeric apo P-domain (D' or D state). Spectra of P-domain in the presence of the maximal GCDCA concentrations were used to fit line shapes and chemical shifts of the GCDCA or GCDCA and metal ion bound states (DL or DL' states). In the final run all spectra were used and chemical shifts of the metal ion bound P-domain or the apo P-domain at a pH_{corr} of 4.4 were kept constant at the values derived in the steps before. Chemical shifts of the GCDCA and metal ion bound states or the GCDCA bound state at pH_{corr} 4.4 were fitted using the values derived before as start conditions. Line widths of all states were fitted using the values derived before as start conditions. Additionally, off-rates and dissociation constants were fitted.

4.4.5 Deriving dissociation constants from binding isotherms

Dissociation constants K_D were derived by fitting experimentally obtained NMR observables Δ_{obs} of n signals at certain ligand and protein concentrations L_t and P_t using the following system of equations (Williamson et al., 2013):

$$\begin{bmatrix} \Delta_{obs,1} \\ \vdots \\ \Delta_{obs,n} \end{bmatrix} = \begin{bmatrix} \frac{K_{D,1}+L_t+P_t-\sqrt{\left((K_{D,1}+L_t+P_t)^2-4L_tP_t\right)}}{2P_t} * \Delta_{obs;max,1} \\ \vdots \\ \frac{K_{D,n}+L_t+P_t-\sqrt{\left((K_{D,n}+L_t+P_t)^2-4L_tP_t\right)}}{2P_t} * \Delta_{obs;max,n} \end{bmatrix}$$

(Eq. 10)

$\Delta_{obs;max}$ is the NMR observable at saturating ligand concentrations and is fitted for each signal individually. K_D values were globally fitted by setting $K_{D,1}$=…=$K_{D,n}$ if CSPs or intensities were used as NMR observables. For determination of K_D values from STD-Afs as NMR observable, the K_D parameter was not constrained, and all equations were fitted individually. The smallest obtained K_D was regarded as closest to the real K_D. Errors were estimated using the standard error based on the residuals of the fit. Fitting of equations was carried out using self-written *python* scripts employing the *lmfit, symfit,* and *numpy* libraries. The *matplotlib* library was used for plotting binding isotherms.

4.4.6 Determination of [^{1}H,^{13}C] methyl group order parameters S^2

Methyl group order parameters were determined as suggested by Tugarinov and Kay, 2013, using the equations given below with a more extensive error propagation analysis.

$$I_{HSQC}(t_1, t_2) =$$

$$\left[\frac{9}{4} * \exp\left(-2\tau R_{2,H}^{F}\right) + \frac{9}{4} * \exp\left(-2\tau R_{2,H}^{S}\right)\right]$$

$$* \exp\left(-2\tau R_{2,H}^{F}\right) * \exp\left(-t_1 R_{2,C}^{F}\right) * \exp\left(-t_2 R_{2,H}^{F}\right)$$

$$+[\frac{9}{4} * \exp\left(-2\tau R_{2,H}^{F}\right) - \frac{3}{4} * \exp\left(-2\tau R_{2,H}^{S}\right)]$$

$$* exp\left(-2\tau R_{2,H}^{F}\right) * \exp\left(-t_1 R_{2,C}^{S}\right) * \exp\left(-t_2 R_{2,H}^{F}\right)$$

$$+\left[\frac{9}{4} * \exp\left(-2\tau R_{2,H}^{F}\right) + \frac{9}{4} * \exp\left(-2\tau R_{2,H}^{S}\right)\right]$$

$$* \exp\left(-2\tau R_{2,H}^{S}\right) * \exp\left(-t_1 R_{2,C}^{F}\right) * \exp\left(-t_2 R_{2,H}^{S}\right)$$

$$+[-\frac{3}{4} * \exp\left(-2\tau R_{2,H}^{F}\right) + \frac{9}{4} * \exp\left(-2\tau R_{2,H}^{S}\right)]$$

$$* \exp\left(-2\tau R_{2,H}^{S}\right) * exp\left(-t_1 R_{2,C}^{S}\right) * exp\left(-t_2 R_{2,H}^{S}\right)$$

(Eq. 11)

$$I_{HMQC}(t_1, t_2) = 6 * \exp\left(-4\tau R_{2,H}^{F}\right) * \exp\left(-t_1 R_{MQ}^{F}\right) * \exp\left(-t_2 R_{2,H}^{F}\right)$$

$$+6 * \exp\left(-4\tau R_{2,H}^{S}\right) * \exp\left(-t_1 R_{MQ}^{S}\right) * \exp\left(-t_2 R_{2,H}^{S}\right)$$

(Eq. 12)

R denotes relaxation rates. The superscripts “F” and “S” correspond to fast and slow relaxation rates, respectively. The subscripts “2,H” and “2,C” characterize ^{1}H and ^{13}C single quantum coherence relaxation rates. The subscript “MQ” corresponds to the ^{1}H-^{13}C multiple quantum coherence relaxation rate. τ is a delay in the pulse program ($1/2*{}^{1}J_{CH}$= 3.6 ms). t_1 and t_2 are the acquisition times in the indirect and direct dimensions, respectively. It is assumed that the product of the ^{13}C resonance frequency ω_C and the global correlation time τ_C $\omega_C * \tau_C >> 1$, and that methyl rotation is infinitely fast. The relaxation rates in equations 11 and 12 are given with

$$R_{2,H}^{F} = \left(\frac{9}{20} * \frac{\gamma_H^4 \hbar^2 \tau_C}{r_{HH}^6} + \frac{1}{45} * \frac{\gamma_H^2 \gamma_C^2 \hbar^2 \tau_C}{r_{HC}^6}\right) * \left(\frac{\mu_0}{4\pi}\right)^2 * S^2 + R_{2,ext}$$

(Eq. 13)

$$R_{2,H}^{S} = \left(\frac{1}{45} * \left(\frac{\mu_0}{4\pi}\right)^2 * \frac{\gamma_H^2 \gamma_C^2 \hbar^2 \tau_C}{r_{HC}^6}\right) * S^2 + R_{2,ext}$$

(Eq. 14)

$$R_{2,C}^{F} = \left(\frac{1}{5} * \left(\frac{\mu_0}{4\pi}\right)^2 * \frac{\gamma_H^2 \gamma_C^2 \hbar^2 \tau_C}{r_{HC}^6}\right) * S^2 + R_{1,ext}$$

(Eq. 15)

$$R_{2,C}^{S} = \left(\frac{1}{45} * \left(\frac{\mu_0}{4\pi}\right)^2 * \frac{\gamma_H^2 \gamma_C^2 \hbar^2 \tau_C}{r_{HC}^6}\right) * S^2 + R_{1,ext}$$

(Eq. 16)

$$R_{MQ}^{F} = \left(\frac{4}{45} * \frac{\gamma_H^2 \gamma_C^2 \hbar^2 \tau_C}{r_{HC}^6} + \frac{9}{20} * \frac{\gamma_H^4 \hbar^2 \tau_C}{r_{HH}^6}\right) * \left(\frac{\mu_0}{4\pi}\right)^2 * S^2 + R_{2,ext}$$

(Eq. 17)

$$R_{MQ}^{S} = R_{2,ext} = \frac{8}{15} * \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^2 \gamma_C^2 \hbar^2 \tau_C}{r_{HDext}^6} + \frac{9}{20} * \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2 \tau_C}{r_{HHext}^6}$$

(Eq. 18)

$$R_{1,ext} = \frac{3}{20} * \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2 \tau_C}{r_{HHext}^6}$$

(Eq. 19)

where μ^0 is the vacuum permeability constant and $\hbar$ is the Planck constant. γ_H and γ_C are the gyromagnetic rations of ^1H and ^{13}C spins. r_{HH} and r_{HC} are the distances between ^{13}C and ^1H (1.135 Å) or ^1H and ^1H spins (1.813 Å) within a methyl group. $R_{1,ext}$ and $R_{2,ext}$ describe dipolar relaxation with external deuterons and protons. r_{HHext} and r_{HDext} are dependent of distances to external protons and deuterons (see below).

To account for the intensity in the final spectra, equations for I_{HMQC} and I_{HSQC} were integrated over the acquisitions times $t_{1,max}$ (0.0942302 s) and $t_{2,max}$ (0.1136640 s) of the indirect and direct dimensions. This is achieved by substituting the $\exp(-t_1R)$ and $\exp(-t_2R)$ terms in Eqs. 11 and 12 with $(1-\exp(-t_{1,max}R)/R$ and $(1-\exp(-t_{2,max}R)/R$ terms.

Gauss distributions of R^S_{MQ}

R^S_{MQ} was derived experimentally using the pulse program *2020.R.mq.am.tb.V2* (c.f. chapter 4.3.3) with relaxation delays Δt of 0 ms, 4 ms, 8 ms, 12 ms, 20 ms, 28 ms, 36 ms, 50 ms, 76 ms, 100 ms, and 200 ms and using resonance intensities and relaxation delays to fit R^S_{MQ} and A_0 to an exponential decay:

$$I(\Delta t) = A_0 * \exp\left(-R_{MQ}^{S} \Delta t\right)$$

(Eq. 20)

For error estimation, standard deviations and means of intensities of methyl group resonances were estimated from duplicate intensity measurements for each methyl group resonance and relaxation delay. Standard deviations and mean of the spectral noise were calculated from 100 noise peaks. Noise peaks were obtained from the spectrum with the longest relaxation delay. For each methyl group

resonance and relaxation delay, 1000 synthetical intensity values were generated by summing up randomly drawn values from synthetic Gauss distributions of the spectral noise and the respective resonance's intensity (estimated by the respective standard deviation and mean, see above). 1000 bootstrap iterations were carried out for the fitting process. In each iteration, one intensity value for each relaxation delay was drawn randomly from the corresponding synthetic intensity values. The obtained values and the corresponding relaxation delays were than used to fit R^S_{MQ}. This resulted in a Gauss distribution of R^S_{MQ} values defined by the standard deviation and mean of the bootstrap procedure. Spectra with different relaxation delays were acquired in a random order to avoid artifacts.

Gauss distributions of I_{HMQC} and I_{HSQC}

Gauss distributions of I_{HMQC} and I_{HSQC} values were estimated by the means and standard deviations of triplicate measurements.

Gauss distributions of the global correlation time τ_C

Determination of ^{15}N R_α and R_β relaxation rates was carried out by integrating the 8 ppm to 10 ppm region of TRACT spectra with 25 increasing delays Δt_{TRACT} and fitting the integral areas and delays to an exponential decay:

$$I(\Delta t) = A_0 * \exp\left(-R_i \ \Delta t_{TRACT}\right)$$

(Eq. 21)

with i ϵ [α,β]. Gauss distributions of R_α and R_β resulted from the obtained values and standard errors of the fit. A Gauss distribution of τ_C was generated by 1000 Monte Carlo iterations of calculating τ_C from randomly drawn values of R_α and R_β Gauss distributions by solving the following equation (Lee et al., 2006):

$$R_\beta - R_\alpha = 2\frac{\mu_0 \gamma_H \gamma_N h}{16\pi^2\sqrt{2} r_{HN}^3} * \frac{\gamma_N B_0 \Delta\delta_N}{3\sqrt{2}} * \left(1.6\tau_C + \frac{1.2\tau_c}{1+\tau_C^2\omega_N^2}\right) * (3 * \cos^2(\Theta) - 1)$$

(Eq. 22)

where γ_N is the gyromagnetic ration of ^{15}N spins, h is the Planck constant, r_{HN} is the distance between ^1H and ^{15}N spins within the N-H bond (1.02 Å), B_0 is the external magnetic field strength, $\Delta\delta_N$ is the difference of the two principal components of the axially symmetric ^{15}N chemical shift tensor (160 ppm), and ω_N is the ^{15}N resonance frequency. The angle Θ is assumed to be 17°.

Gauss distribution of r_{HHext}

Gauss distributions of r_{HHext} were obtained by calculating means and standard deviations resulting from 1000 Monte Carlo iterations of solving equation 18 with values for τ_C and R^S_{MQ} randomly drawn from their respective Gauss distributions.

Calculation of r_{HDext} from a crystal structure model

The crystal structural model of MNV P-domain in presence of GCDCA, metal ions, and CD300lf (pdb 6e47, Nelson et al., 2018) was used to obtain the mean distance r_{HDext} of the three protons of a certain methyl groups to the protein's deuterons using:

$$r_{HDext} = \frac{1}{3} \sum_{j=1-3} \sum_{i} \frac{1}{r_{H_j D_i}^6}$$

(Eq. 23)

where proton "j" corresponds to the methyl group of question and deuteron "i" can be any deuteron of the protein.

Calculating S^2 methyl group order parameters

For the calculation of methyl group order parameters, 1000 Monte Carlo iterations were performed employing the corresponding r_{HDext} and drawing random values from Gauss distributions of τ_C, R^S_{MQ}, r_{HHext}, I_{HMQC}, and I_{HSQC}. The values were used to fit S^2 globally according to equations 11 to 19. The resulting mean and standard deviation are given as the final S^2 value and its corresponding error, respectively. *Python* scripts employed for the analysis are given in the data repository of this thesis.

4.4.7 Deriving apparent pK_a values from CSPs

CSPs in Hz $\Delta\nu_{eucl}$ (see above) of methyl group resonances at different pH_{corr} values (c.f. chapter 5.5.3) were used to fit apparent ionization constants pK_a to an adopted version of the Henderson-Hasselbalch equation (Croke et al., 2011):

$$\Delta\nu_{eucl} = \Delta\nu_{eucl,low} - \frac{\Delta\nu_{eucl,low} - \Delta\nu_{eucl,high}}{1 + 10^{(n_{Hill} * (pk_a - pH))}}$$

(Eq. 24)

with $\nu\Delta_{eucl,low}$ and $\nu\Delta_{eucl,high}$ being the low pH and high pH CSP plateaus. n_{Hill} is the apparent Hill coefficient. pk_a values are regarded as apparent because methyl groups are not directly protonated. *Python* scripts employed for the analysis are given in the data repository of this thesis.

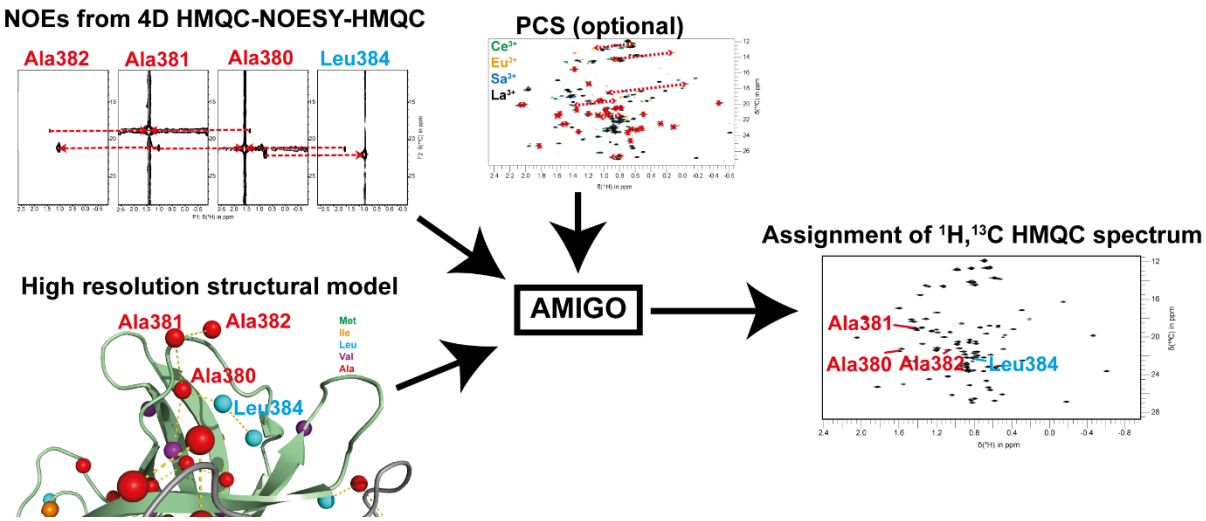
4.4.8 Illustration and RMSD analysis of structural models

Structural models were graphically illustrated using *pymol* (DeLano, 2002). Alignment of structural models and calculation of corresponding RMSD values was carried out using the "align" command. Three dimensional rotations were performed using the "turn" command. Distances were determined using the "distance" command.

5 Results

5.1 AMIGO – a tool for automated NMR assignment of methyl group labeled proteins

$^{1}H,^{13}C$ correlated NMR experiments allow the investigation of [$^{1}H,^{13}C$] methyl group labeled proteins with molecular weights of several hundred kDa (Sprangers et al., 2007). Thus, the amount of NMR-accessible protein systems increased significantly over the last 20 years. Once an [$^{1}H,^{13}C$] resonance assignment has been achieved, the approach can provide insights into, e.g., structural aspects, ligand binding, and dynamics. Structure-based assignment approaches use experimental distance restraints provided by NOEs. These restraints are compared to a high-resolution structural model with the aim to match experimental restraints and theoretical restraints as predicted from the model. The approach is referred to as the *methyl walk* method (Proudfoot et al., 2016). Additional restraints such as theoretical and experimental PCSs can be used to support this assignment strategy (Flügge and Peters, 2018; Mühlberg et al., 2022). In my thesis, I have developed a new automated assignment algorithm, AMIGO (**A**utomated **M**ethyl walk **i**dentification by reconstructing **G**raphs from *building bl**o**cks*, Scheme 5.1). Unlike other algorithms, AMIGO makes use of the very intuitive methyl walk method. This results in a comprehensive and understandable output for the user. Additional restraints such as PCSs can be employed to improve performance. To keep the computational time within reasonable limits, *building blocks* are used to reconstruct experimental and structural model-based NOE networks. The complete assignment is achieved by combining the individual building blocks in a LEGO manner. The performance of AMIGO was validated using 13 different data sets of assigned proteins.



Scheme 5.1: AMIGO uses experimental restraints such as NOEs and PCSs and a high-resolution structural model to generate methyl walks and assign resonances in the corresponding $^{1}H,^{13}C$ HMQC spectrum. The figures are taken from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.1.1 AMIGO workflow

A brief overview of AMIGO's general workflow, input provided by the user, and output generated is presented in Fig. 5.1.1.

The input comprises a high-resolution structural model from that the spatial coordinates of methyl groups are taken, a list of NOE-based distance restraints, and an assignment of resonance signals to amino acid types. As an option, lists containing information about additional restraints, e.g., PCSs or pre-assignments can be provided to support the assignment process. See Fig. S 1.2 for more information about providing input files to AMIGO.

AMIGO considers all methyl group resonances as potential starting points for methyl walks. From these starting points individual methyl walks are created. The best methyl walk is chosen based on the *total score*, which will be discussed in detail later on. The methyl groups are assigned to methyl group resonances according to this methyl walk. Resonances and methyl groups that are already assigned to each other are excluded from further analysis. The procedure is repeated with the remaining methyl group resonances as potential starting points until no further assignments can be identified.

As output (Fig. S 1.5), AMIGO provides a list of resonances assigned to methyl groups and the methyl walk applied for obtaining the assignments (Fig. S 1.5b). Additional restraints (e.g., PCSs) associated with the respective methyl groups and resonances are also listed and compared. This information assists understanding and comprehending the methyl walks generated by AMIGO and allows critical assessment of the assignment. In addition, experimental NOE networks and NOE networks based on the structural model are presented as graphs. They can be used to track the methyl walks proposed by AMIGO (Fig. S 1.5a and c). To further validate AMIGO's assignment, AMIGO also plots experimentally and theoretically determined additional restraints by applying the assignment (Fig. S 1.6).

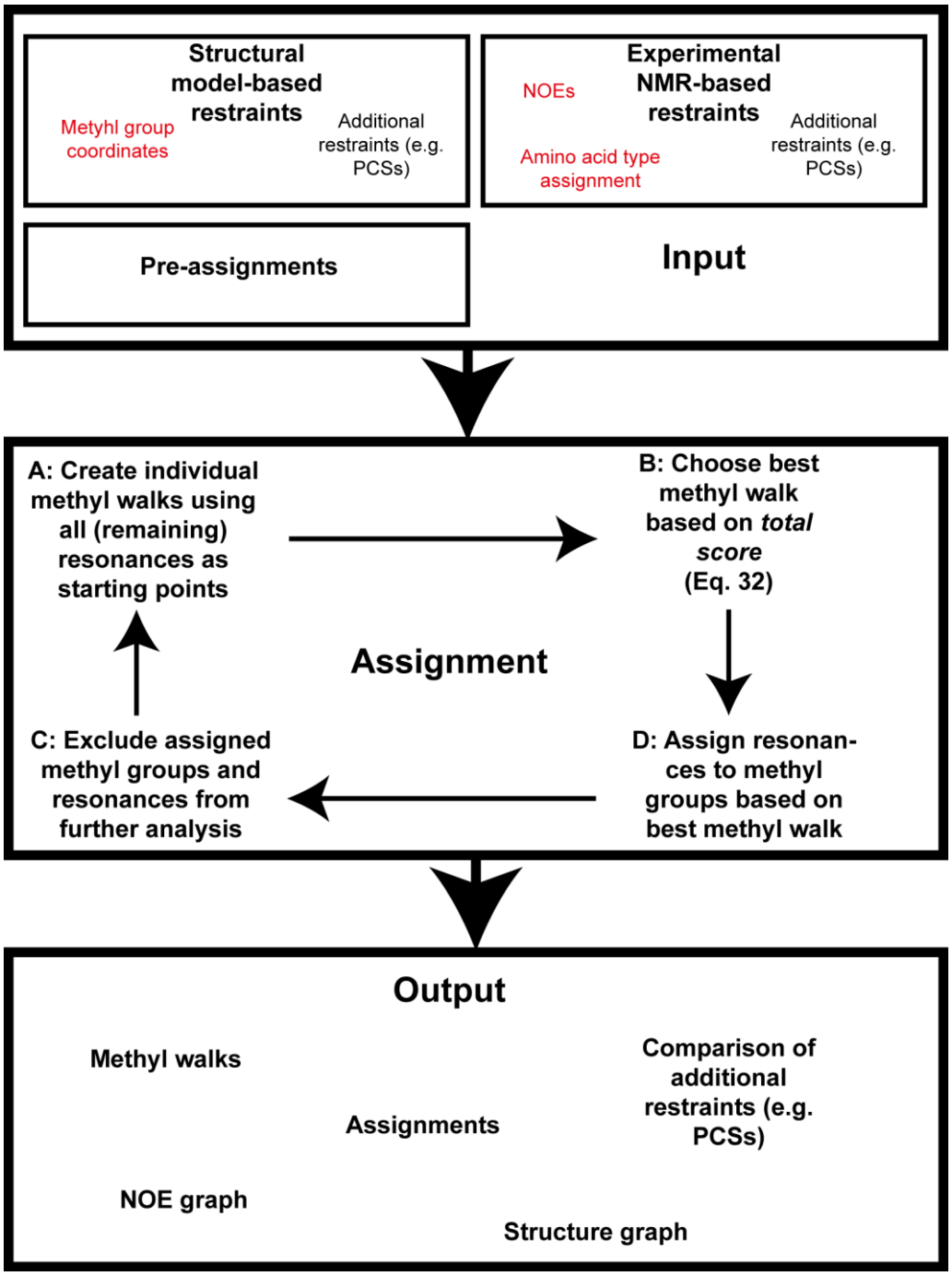


Fig. 5.1.1: AMIGO workflow. The user provides input data. Inputs highlighted in red are necessary; inputs written in black are optional. The algorithm starts methyl walks by trying out every methyl group resonance as a possible starting point and assigns resonances to methyl groups. AMIGO creates an output for the user making the assignments comprehensive. For further details about input files, see Fig S 1.2. For more information about output files, see Fig. S 1.5 and Fig. S 1.6.

5.1.2 AMIGO identifies methyl walks by disassembling and reconstructing experimental and theoretical NOE networks

AMIGO compares an experimental NOE network with theoretical NOE networks based on a high-resolution structural model. As with other algorithms, e.g., MAGMA or MAGIC, this is done by generating graphs (Pritišanac et al., 2017; Monneau et al., 2017). In contrast to the published algorithms, AMIGO samples a set of possible structure-based graphs instead of using a single graph. These graphs and a graph based on experimental NOEs are decomposed into *building blocks* and then reassembled to identify methyl walks.

The graph and the building block concept

The conversion of experimental NOE data into a graph results in an *NOE-based graph*. Resonances in [^{1}H,^{13}C] HMQC spectra are considered as nodes (Fig. 5.1.2a). Each node has attributes such as the corresponding amino acid type and optional additional restraints (e.g., experimental PCS; not shown in Fig. 5.1.2). Edges between nodes are created based on experimental NOE connectivities (Fig. 5.1.2a). For illustration purposes, in this chapter the amino acid types are reflected by the color of the nodes, i.e., orange corresponds to isoleucine, green to methionine, and blue to alanine.

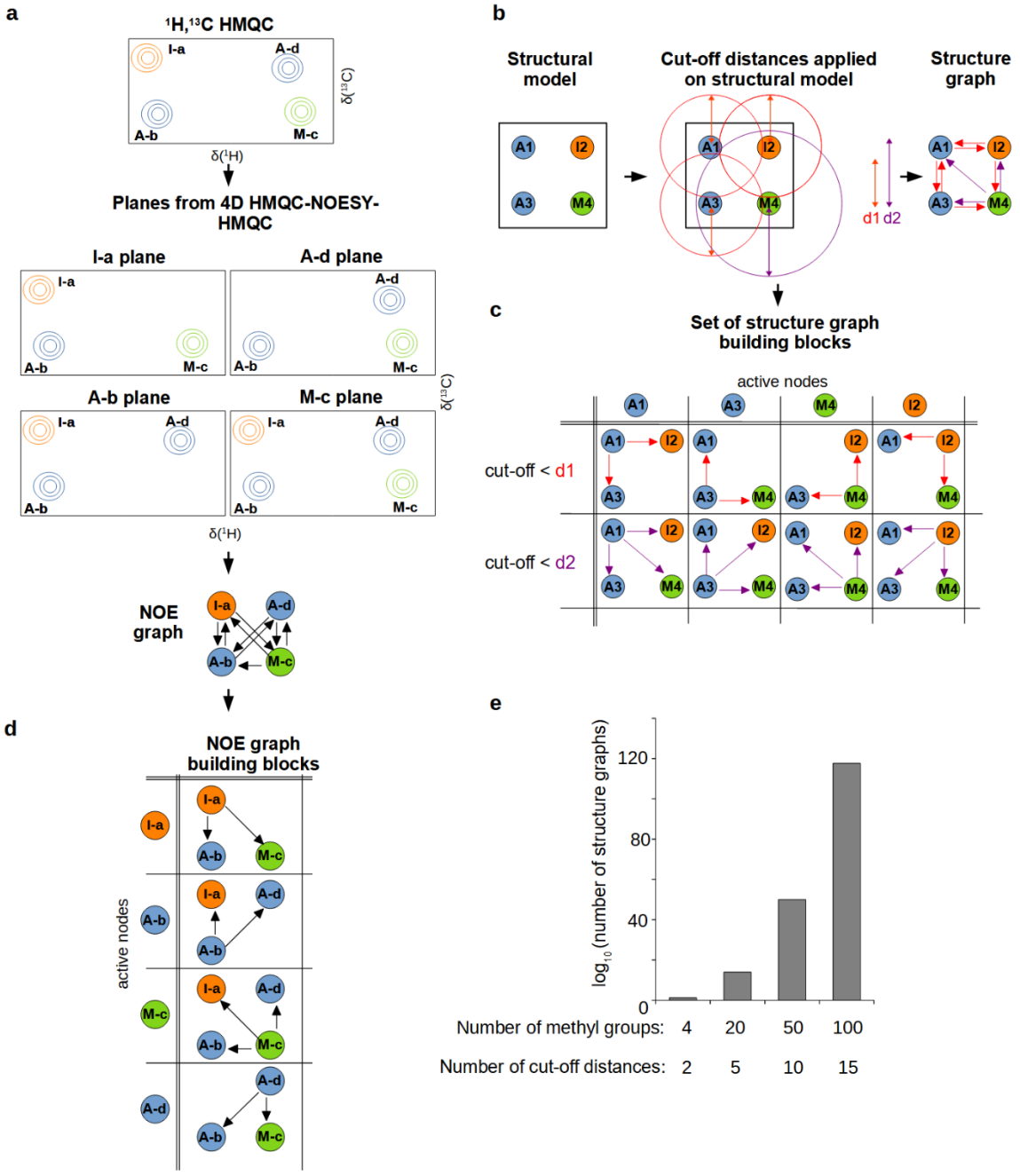


Fig. 5.1.2: Generation of structure-based and NOE-based graphs and dissection into *building blocks*. (a) Creation of NOE-based graphs. For each methyl group resonance in a 2D [^{1}H,^{13}C] HMQC spectrum, there is a node in the corresponding directed NOE graph. The nodes are connected by edges given by NOE connectivities obtained form, e.g., a 4D HMQC-NOESY-HMQC experiment. NOE connectivities can be symmetric, with an NOE cross peak observed starting from either of two methyl group resonances, or asymmetric, in which case an NOE cross peak is only observed starting from one of the two methyl group resonances, i.e., only in one dimension. This is symbolized by the direction of the arrows ("directed edges"). The resonances and corresponding nodes are labeled with a capital letter according to their amino acid type and a lowercase letter as an identifier. (b) Creation of a structure-based graph. For each methyl group in the structural model, there is a node in the structure-based graph. Nodes are connected by edges if they are close in space. The criterion for two methyl groups being close is regulated by a variable cut-off distance (here "d1" and "d2"). If only one cut-off distance is used, all connectivities are symmetric and are symbolized by bidirectional edges. In case different cut-off distances are used, connections may become asymmetric and are symbolized by an uni-directional edge. The

methyl groups and corresponding nodes are labeled with a capital letter according to their amino acid type and an integer as an identifier. (c) Example for a set of building blocks. Each building block is given by an *active node* and a cut-off distance that is applied to the corresponding methyl group in the structural model. Note that for the example shown in (b), 4 methyl groups and 2 cut-off distances result in 2^4 possible structure-based graphs (c.f. Fig. 5.1.3b) but only in 2*4 building blocks representing the set of possible structure-based graphs. From the set of building blocks, every structure-based graph in the set of graphs can be reassembled (c.f. Fig. 5.1.3). (d) shows the disassembling of the NOE graph into building blocks. Shown in (e) is how the number of possible structural graphs increases with the number of methyl groups and cut-off distances taken into account. For better visualisation, the amino acid types of the nodes is additionally encoded in the node's color, i.e., orange corresponds to isoleucine, green to methionine, and blue to alanine.

In contrast to the NOE graph, the creation of a suitable *structure graph* requires further considerations. Methyl groups occurring in the structural model are regarded as nodes (Fig. 5.1.2b). As with the NOE graph, every node has attributes such as the corresponding amino acid type and theoretically derived additional restraints. Edges are created based on cut-off distances that are compatible with the observation of corresponding NOEs (Fig. 5.1.2b). However, the observation of NOEs depends on many factors, with the distance between the two nuclei under scrutiny being only one of them (Neuhaus, 2011). Therefore, an individual cut-off distances is used for each individual methyl group, and multiple sets of cut-off distances are sampled to create a set of structural graphs which are in turn compared to the NOE-based graph. This variability introduces a problem: In the example shown in Fig. 5.1.2b, two possible cut-off distances (d1 and d2) are assumed. For a protein with only 4 methyl groups arranged in a rectangle, $m^n=2^4=16$ different structure graphs can be created (shown in Fig. 5.1.3b), where m is the number of different cut-off distances and n is the number of methyl groups. Examples of the number of possible structure graphs are shown in Fig. 5.1.2e. For a protein carrying 100 methyl groups and assuming 5 possible cut-off distances, extrapolation of the example would result in $7.9*10^{69}$ possible structure graphs.

To allow taking the entire set of possible structure-based graphs into account, a novel graph format was developed, where a set of graphs is disassembled into a set of *building blocks* describing the graphs (Fig. 5.1.2c). Each building block consists of an *active node* corresponding to a methyl group in the structure. Applying a certain cut-off distance on this methyl group results in building blocks that include the active node and a variable number of nodes corresponding to methyl groups within this cut-off distance based on the structural model. Combining individual building blocks allows assembly of each graph out of the set of possible structure-based graphs (Fig. 5.1.3). Consequently, the set of m^n structure graphs can be described by n*m building blocks (c.f. Fig. 5.1.3 a and b). The NOE-based graph can be similarly disassembled into building blocks (Fig. 5.1.2d). NOE graph building blocks and structure graph building blocks can be compared directly to identify methyl walks. Importantly, this simplification allows reasonable computing times.

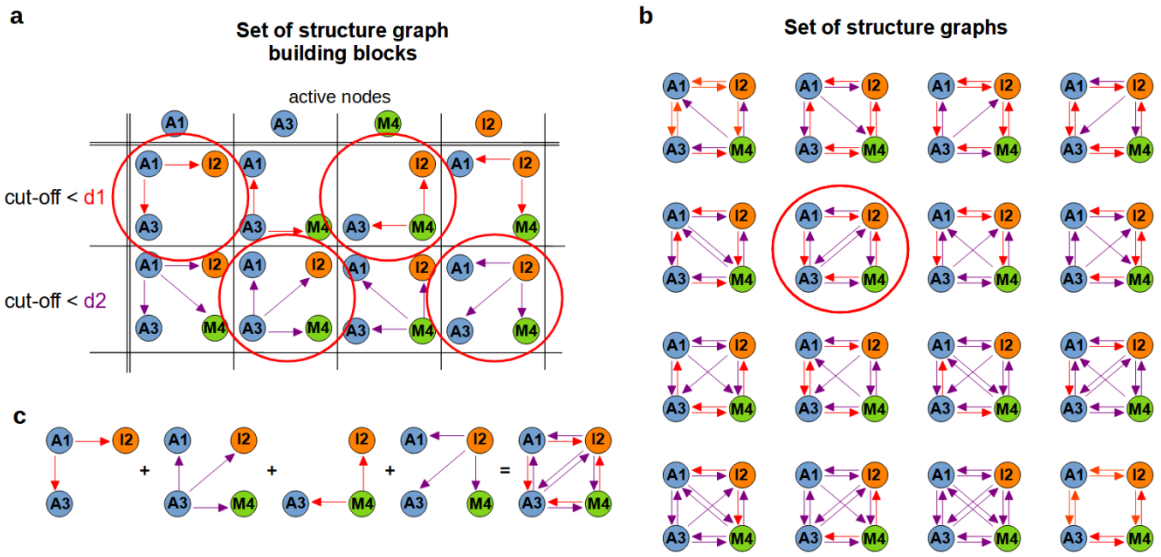


Fig. 5.1.3: Example for reconstructing a structure graph from structure graph building blocks. Recreation of a specific structure graph (c) being part of the set of possible structure graphs shown in (b) from the set of structure graph building blocks (a) describing this set of structure graphs.

Identification of methyl walks from graph building blocks

Identification of methyl walks using building blocks describing the set of structure graphs and the NOE graph is carried out in four steps. For a better understanding it is helpful to follow the example shown in Fig. 5.1.4 to Fig. 5.1.8.

Step 1 (Fig. 5.1.4**)**

A given NOE graph building block (NGBB) is compared with all structure graph building blocks (SGBBs) whose active nodes have the same amino acid type as the active node of the NGBB. The comparison is evaluated by calculating the *comparison score* (Box 1, Eq. 28). Briefly, the comparison score is based on similarities of edge patterns, adjacent amino acid types, and experimental and theoretical additional restraints (e.g., PCSs). It is divided into an NOE-based term (Eq. 25) and an additional restraints-based term (Eq. 26 or Eq. 27). This way the best matching SGBB is identified.

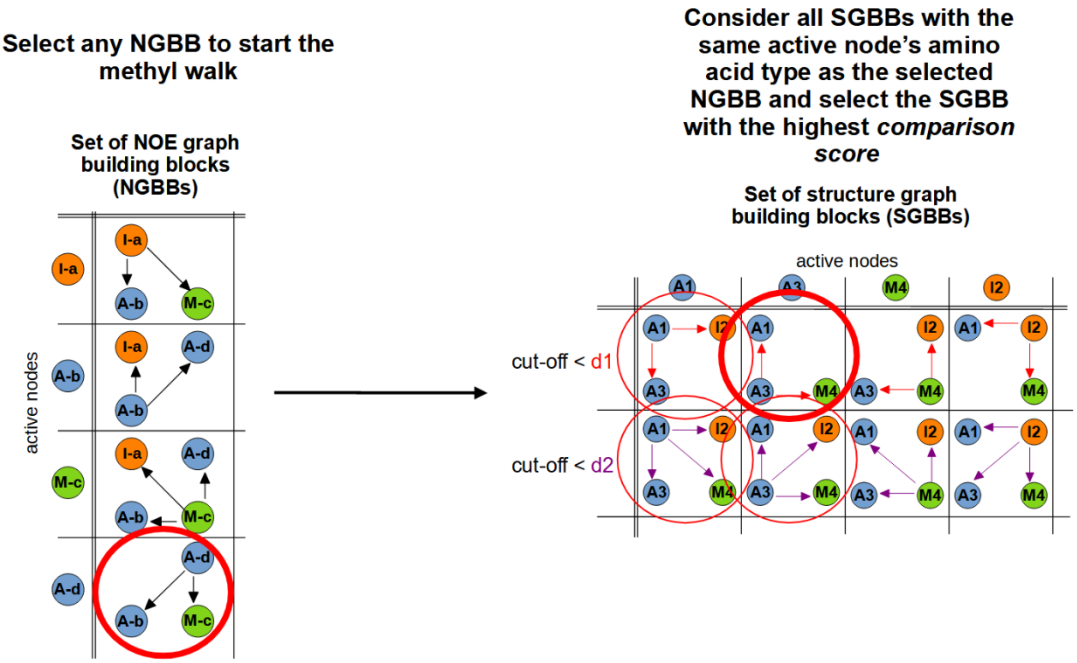


Fig. 5.1.4: Step 1 of the identification of methyl walks from graph building blocks. See main text to step 1 for more explanations. Red circles indicate building blocks that are considered, red circles with thick lines highlight the building blocks that are selected.

Step 2a (Fig. 5.1.5)

The active node of the SGBB that fits best is assigned to the active node of the NGBB. In addition, the two building blocks are used to start the reconstruction of the NOE graph and a structure graph matching the NOE graph. The two active nodes that have been assigned to each other also form the starting point for the methyl walk.

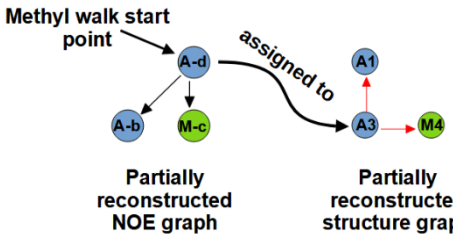


Fig. 5.1.5: Step 2a of the identification of methyl walks from graph building blocks. See main text to step 2a for more explanations.

Step 3 (Fig. 5.1.6**)**

The third and fourth steps continue the methyl walk. In step three, the partially reconstructed NOE graph is searched for unassigned nodes. The *uniqueness score* (Box 2, Eq. 29) of the corresponding NGBBs are calculated, and the NGBB with the lowest uniqueness score is identified. The uniqueness score reflects the relative abundance of the NGBB's active node's amino acid. A rarely occurring amino acid type and edges within that NGBB to nodes with other rare amino acid types yields a low

uniqueness score. The lower the uniqueness score, the “more unique” is the NGBB, and the lower are chances of erroneous assignments when continuing the methyl walk starting with this NGBB.

By continuing the methyl walk with an NGBB being “as unique as possible”, the chance of erroneous assignments is reduced. By only considering NGBBs whose active nodes correspond to nodes in the partially reconstructed NOE graph, there is always an already assigned node (here “A-d” which was assigned in the previous step) within the corresponding NGBB, which is necessary to create methyl walks.

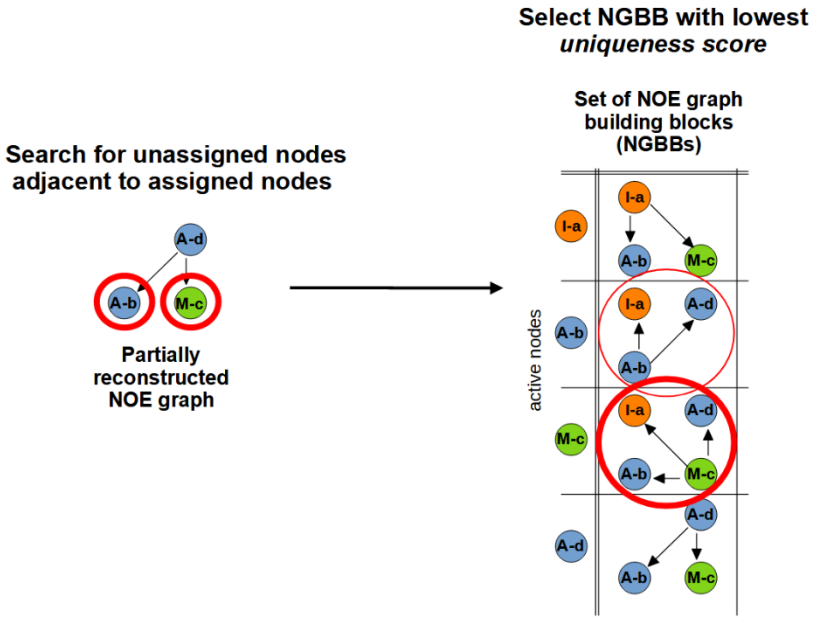


Fig. 5.1.6: Step 3 of the identification of methyl walks from graph building blocks. See main text to step 3 for more explanations. Red circles indicate building blocks (right) or active nodes (left) that are considered, red circles with thick lines highlight the ones being selected.

Step 4 (Fig. 5.1.7**)**

In the fourth step, the identified NGBB from step three is compared to SGBBs using the comparison score explained in step one. However, the search space is greatly reduced by considering only a subset of SGBBs. First, the amino acid type of the SGBBs’ active node must match the amino acid type of the NGBB’s active node. Second, only SGBBs whose active nodes are unassigned and correspond to a node in the partially reconstructed structure graph are considered. That node in the partially reconstructed graph therefore automatically fulfils another requirement. It is adjacent to a node which is assigned to a node being part of the identified NGBB (“A3”, which is assigned to “A-d” in Fig. 5.1.5). The comparison score (Eq. 28) is then used to identify the best matching SGBB. It should be noted that this way chains of assigned nodes and thus methyl walks are created, and calculation time is significantly reduced.

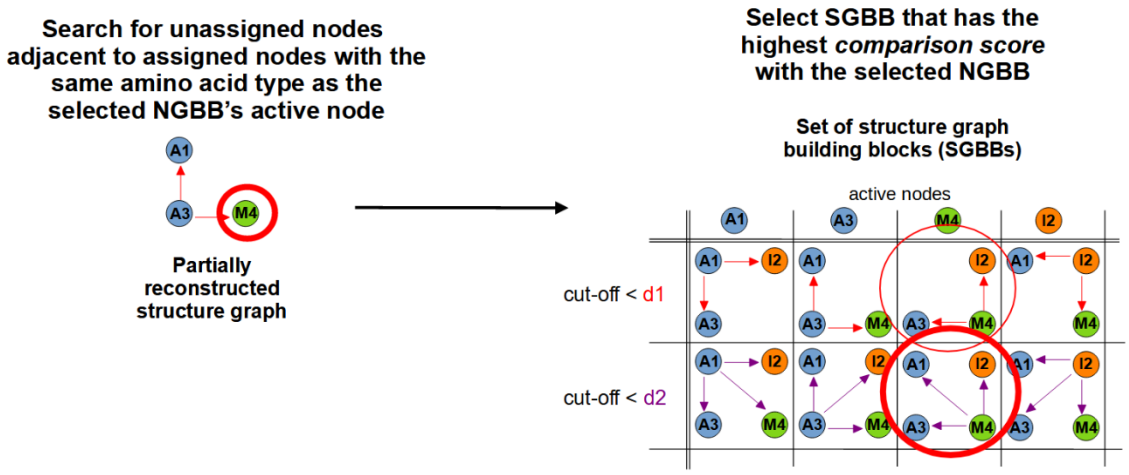


Fig. 5.1.7: Step 4 of the identification of methyl walks from graph building blocks. See main text to step 4 for more explanations. Red circles indicate building blocks (right) or active nodes (left) that are considered, red circles with thick lines highlight the ones being selected.

Step 2b (Fig. 5.1.8**)**

The algorithm proceeds with step two. The active nodes of the NGBB and the best matching SGBB are assigned to each other, and the respective building blocks are used to continue the assembly of the NOE graph and a matching structure graph.

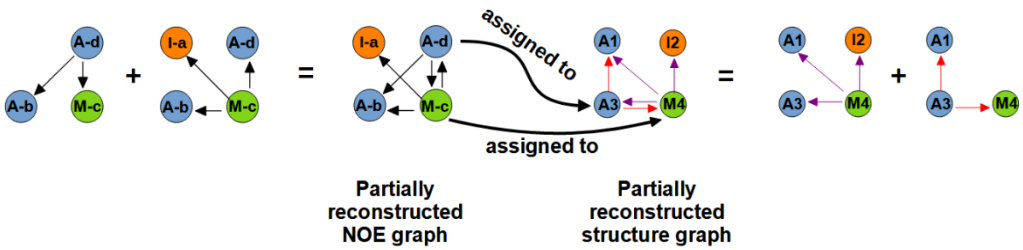


Fig. 5.1.8: Step 2b of the identification of methyl walks from graph building blocks. See main text to step 2b for more explanations.

Steps 2b, step 3, and step 4 are repeated until no more assignments can be identified. AMIGO considers every methyl group as potential starting point. Each starting point results in an individual methyl walk. The respective methyl walks and resulting assignments are validated against each other by calculating the *total score* (Box 3, Eq. 32). Only that methyl walk with the highest total score and therefore best solution is used for the final assignment (see also Fig. 5.1.1).

In contrast to other algorithms, the method presented here is an intelligent implementation of the methyl walk strategy, making assignments highly understandable and allowing the user to supervise the results. Furthermore, the consideration of different cut-off distances for each individual methyl group is expected to decrease differences in structure graphs and NOE graphs, thus enabling an advance in graph-based automated assignment theory.

Box 1: The *comparison score*.

See Fig. 5.1.9 for an exemplary calculation. The NMR-based term is defined as

$$comparison\ score_{NOE}(NGBB, \mathrm{SGBB}) = W_{NOE}{*}2{*}c - |c - E_{structure}| - |c - E_{NOE}|$$

Eq. 25

with c being the number of common edges of the two building blocks, $E_{structure}$ being the SGBB's total number of edges, E_{NOE} being the NGBB's total number of edges. W_{NOE} is the weight factor for NOE-based restraints, which is typically set to 1 (see chapter 4.1).

The additional restraints-based term is only used if additional restraints are provided and is defined as:

$$comparison\ score_{additional\ restraints}(NGBB, \mathrm{SGBB}) = \sum_{i=1}^{4} W_i{*}|a_{i,exp} - a_{i,theo}|$$

Eq. 26

W_1, W_2, W_3 and W_4 being weight factors for the four optional sets of additional restraints. $a_{i,\mathrm{exp}}$ and $a_{i,theo}$ are experimentally and theoretically determined additional restraints of the NGBB's and the SGBB's active nodes. If a particular $a_{i,\mathrm{exp}}$ or $a_{i,theo}$ is not provided, the corresponding additional restrain-based term is calculated by using

$$comparison\ score_{additional\ restraints}(NGBB, \mathrm{SGBB}) = \sum_{i=1}^{4} 10 * W_i{*}\left|\mu_{i,\mathrm{exp}} - \mu_{i,theo}\right|$$

Eq. 27

with $\mu_{i,\mathrm{exp}}$ and $\mu_{i,theo}$ being the mean of the absolute values of the other i-ths experimentally or theoretically determined set of additional restraints. Thereby, AMIGO is not biased towards assignments without additional restraints. This procedure has empirically proven to be suitable.

The comparison score is then given by

$$comparison\ score(\mathrm{NGBB}, \mathrm{SGBB}) = comarison\ score_{NOE} - comparison\ score_{additional\ restraints}$$

Eq. 28.

Fig. 5.1.9: Exemplary calculation of the *comparison score*. NOE graph building block (left) and a structure graph building block (right) for exemplarily computing the *comparison score*$_{NOE}$ (Eq. 25) The number of common edges *c* is 2 as there is an edge from an isoleucine node to a methionine node and an edge from an isoleucine node to an alanine node occurring in both building blocks. The number of total edges $\boldsymbol{E_{structure}}$ and $\boldsymbol{E_{NOE}}$ is 3 and 2. As a result, the *comparison score*$_{NOE}$ of the pair of building blocks equals $\boldsymbol{3{*}W_{NOE}}$.

Box 2: The *uniqueness score*

An exemplary calculation is illustrated in Fig. 5.1.10. The uniqueness score is defined as

$$uniqueness\ score(NGBB, G_{NMR}) = {p_{active}}^{4*} \prod_{i=1}^{k} p_{i,neighbor}$$

Eq. 29

where k is the number of neighbors of the active node in the NGBB for that the uniqueness score is calculated. This NGBB corresponds to a set of NGBBs termed G_{NMR}. G_{NMR} typically includes all building blocks having unassigned active nodes that are part of a partially reconstructed NOE graph. This is important for the continuation of the methyl walk.

$$p_{active} = n_{NGBB}/n_{total}$$

Eq. 30

with n_{NGBB}being the amount of NGBBs in G_{NMR} having the same active node amino acid type as B_{NMR}, and n_{total} being the total amount of NGBBs in G_{NMR}. Thus, Eq. 30 calculates the probability of an active node having a certain amnio acid type.

$$p_{i,neighbor} = n_{neighbor,i}/n_{total,neighbor}$$

Eq. 31

Where $n_{total,neighbor}$ is the total number of nodes neighbored to an active node in all NGBBs in G_{NMR}. $n_{neighbor,i}$ is the number of nodes neighbored to active nodes in all NGBBs in G_{NMR} that have the same amino acid type as the i-ths neighbor of the active node in B_{NMR}. Thus Eq. 31 calculates the probability that the i-ths neighbor of the active node has a certain amino acid type.

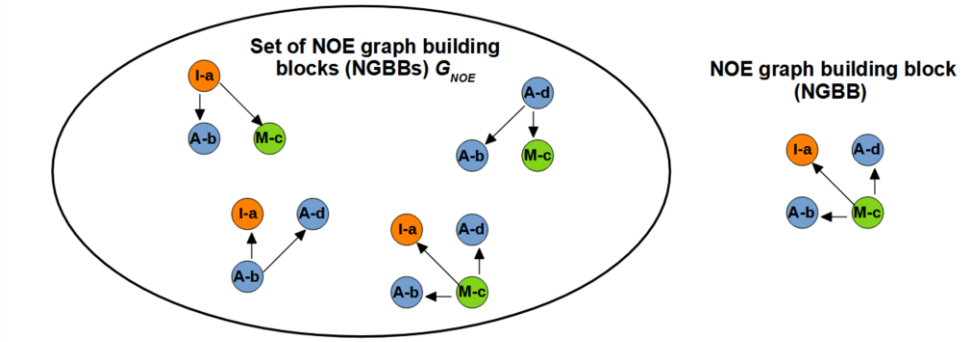


Fig. 5.1.10: Example for calculating the *uniqueness score*. The uniqueness score is calculated for a particular NOE graph building block (right) being part of a set of NGBBs termed G_{NOE} (left) applying Eq. 29. *k* equals 3, n_{NGBB} equals 1, and n_{total} equals 4. Consequently, p_{active} equals ¼. The first, second and third neighbor of the active node in the particular NGBB (right) are A-d, I-a and A-b and have the amino acid types alanine, isoleucine and alanine. $p_{neighbor,1}$ corresponding to the first neighbor, i.e., A-d, is then calculated as follows. As there are 5 edges from an active node to an alanine ($n_{neighbor,1}$) and 9 edges in total ($n_{total,neighbor}$) in the set of NGBBs G_{NMR}, $p_{neighbor,1}$ equals 5/9. $p_{neighbor,2}$ and $p_{neighbor,3}$ are calculated in the same way. This results in a uniqueness score of 2.7*10^{-4} for the particular NGBB (right) making this building block the one with the lowest uniqueness score in G_{NMR}.

Box 3: The *total score*.
The total score is calculated using

$$total\ score(A_{methyl\ walk}) = \sum_{i=1}^{N} comparisonscore(NGBB_i, SGBB_{,i})$$

Eq. 32

with $A_{methyl\ walk}$ being a set of N building block pairs which active nodes have been assigned to each other $(NGBB_i, SGBB_i)$ based on the methyl walk for that the total score is calculated. The methyl walk and corresponding assignments providing the highest total score are chosen for the final assignment.

5.1.3 Testing AMIGO using benchmark data sets

AMIGO was validated using 13 benchmark data sets. The resulting assignments were classified into four categories: not assigned methyl groups, correctly assigned methyl groups, incorrectly assigned groups, and methyl groups assigned to a methyl group close to the actual methyl group (referred to as "interchanged" in Fig. 5.1.11a). An example for the last category is shown in Fig. 5.1.11b (see also chapter 4.1.3). Here, for the huNoV P-dimer, the resonance corresponding to Leu432 was incorrectly assigned to Leu507. The two methyl groups are around 5 Å apart and have the same amino acid type. This is a typical error that AMIGO is prone to. However, this type of error would not particularly limit the conclusions about certain protein regions, e.g., the localization of ligand epitopes.

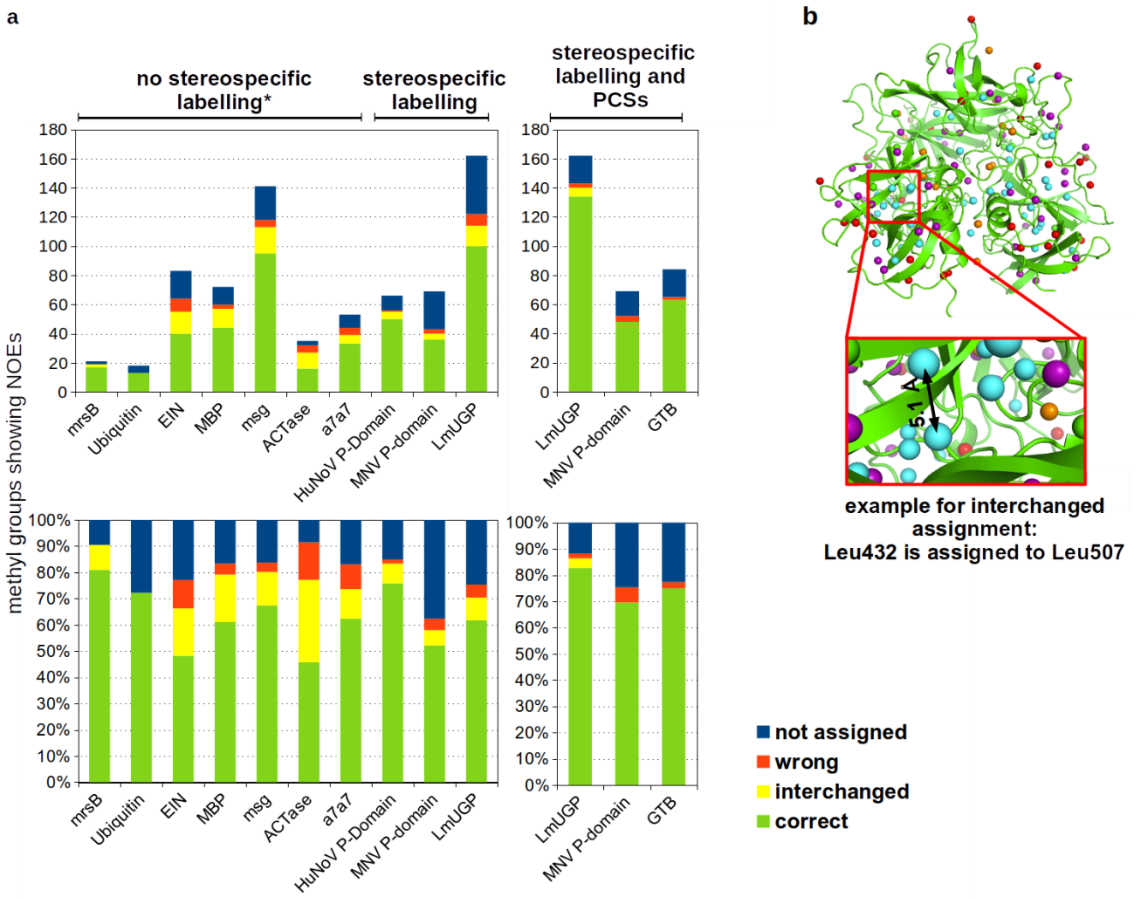


Fig. 5.1.11: Validation of AMIGO. (a) shows AMIGOS performance on different data sets. Assignments are classified by four categories. The upper panels show the total number of assignments; the lower panels show the number of assignments in %. (b) shows an example for the "interchanged" category. For the huNov P-dimer (pdb 4x06, Singh et al., 2015; methyl groups of different amino acid types highlighted as colored spheres), the resonance of Leu432 was assigned to Leu507. Both methyl groups have the same amino acid type and are about 5 Å apart. Reference assignments for the respective data sets can be found in Tab. S 1.1. (*) See main text for further explanations on benchmark data sets without stereospecific labelling.

Stereospecific pro-S leucine and valine labeling strategies only have been developed in recent years (Gans et al., 2010). They provide advantageous properties such as less spectral crowding and are often more straightforward to assign (Flügge and Peters, 2018; Müller-Hermes et al., 2020). AMIGO was therefore designed to assign stereospecifically pro-S labeled proteins. Seven of the 13 benchmark data sets were taken from the MAGMA benchmark set (Pritišanac et al., 2017). The corresponding proteins were not stereospecifically labeled. Using the NOE lists provided with MAGMA, it is not possible to separate NOEs into those originating from pro-S and pro-R methyl groups. NOEs originating from a pair of pro-S and pro-R methyl groups of a single amino acid were therefore combined and attributed to the pro-S methyl group. Thus, AMIGO can be tested with these data sets. However, this processing leads to severe artifacts in the corresponding data sets, as now often significantly more NOEs originate from a given pro-S-methyl group (i.e., NOEs to other methyl groups that have actually only originated from the pro-R methyl group are now additionally and artificially attributed to the pro-S methyl group). Nevertheless, AMIGO identifies methyl walks in these data sets. Accuracy of assignments is protein-dependent, ranging from 50 % for ATCase to 100 % for Ubiquitin (Fig. 5.1.11a). For data sets of stereospecifically pro-S-labeled proteins, accuracy is generally improved and ranges from 83% for

LmUGP to 94% for MNV P-domain. Best results were obtained by including PCSs as additional restraints. Here, the accuracy ranges between 97 % for GTB, 95 % for MNV P-domain, and 91% for LmUGP. AMIGO can be used to identify methyl walk-based assignments, which is especially helpful for proteins bearing a large number of labeled methyl groups. In contrast to other algorithms, AMIGO's output is highly transparent. I recommend a semi-automated strategy, where assignments and methyl walks of AMIGO are supervised by the user and serve as a guidance during the assignment process.

AMIGO's calculation time depends on protein size and the number of labeled methyl groups, but was finished within less than 24 h for all benchmark data sets on an Intel® Core™ i9-9900K Processor (3.6 GHz).

5.2 MNV P-domain [^{1}H,^{13}C] MILVA methyl group NMR assignment

A major prerequisite and bottleneck for studying proteins using NMR spectroscopy is the assignment of resonance signals to specific nuclei. NMR assignment of the MNV P-domain was particularly difficult for several reasons. First, the MNV P-domain exists in an equilibrium of monomers and dimers. Both species produce distinct signals in NMR spectra leading to spectral crowding. To solve this problem, all assignment experiments were carried out in the presence of GCDCA. GCDCA is a small molecule that has been shown to trigger MNV P-domain dimerization (Creutznacher, 2020). Second, ^{15}N backbone labeled proteins are usually employed for NMR studies. Such proteins can be assigned using triple resonance experiments (Cavanagh et al., 1995). However, for proteins as large as the MNV P-domain dimer (ca. 66 kDa), unfolding-refolding protocols are necessary to place ^{1}H nuclei at the backbone amide groups and ^{2}H nuclei elsewhere (Mallagaray et al., 2019). All attempts to establish such a protocol for the MNV P-domain have failed (Maaß, 2019). Hence, the project was pursued by following ^{13}C methyl group labeling strategies. The assignment of Met, Ile, Leu^{ProS}, Val^{ProS}, and Ala (MILVA) [^{1}H,^{13}C] methyl labeled MNV P-dimers was achieved using a 4D HMQC-NOESY-HMQC experiment and yielded the assignment of 55 of 103 methyl group resonances. The generation of point mutants and the measurement of pseudo contact shifts (PCSs) extended the assignment. Finally, 77 of 103 resonances were assigned (Fig. 5.2.1). Of note, the MNV P-domain was found in two different forms caused by Pro361 cis-trans configurational isomerism.

5.2.1 Assignments based on the 4D HMQC-NOESY-HMQC experiment

A 4D HMQC-NOESY-HMQC spectrum contains information about distances between [^{1}H,^{13}C] methyl groups. The comparison of this information with a high-resolution structural model enables a structure-based assignment. The approach is supported by knowledge about the corresponding amino acid types. It is called the *methyl walk* approach (Proudfoot et al., 2016) and was employed to assign the MNV P-domain.

A high-resolution crystal structure of CW3 MNW P-domain (Fig. 5.2.1a) complexed with the CD300lf receptor, metal ions, and GCDCA (pdb 6e47, Nelson et al., 2018) was used as a structural model. Except for a single amino acid substitution (K296E), the CW3 MNV P-domain amino acid sequence is identical to the CW1 MNV P-domain used here (Thackray et al., 2007). Due to the high degree of dimer symmetry, it was expected that 100 resonances would occur reporting for the 200 MILVA methyl groups of the dimer. Surprisingly, 103 methyl group resonances were observed in [^{1}H,^{13}C] HMQC spectra of MILVA labeled P-domain (Fig. 5.2.1b).

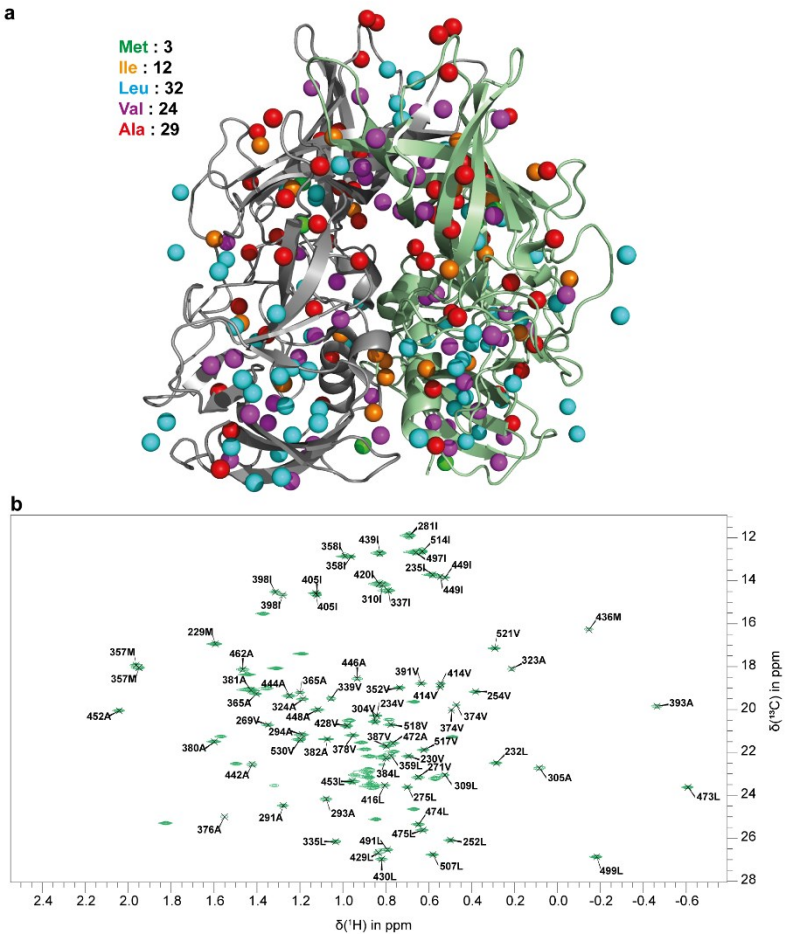


Fig. 5.2.1: Assignment of [^{1}H,^{13}C] HMQC resonances of MILVA labeled MNV P-domain. (a) shows the high-resolution crystal structure (pdb 6e47, Nelson et al., 2018) of the MNV P-domain with MILVA methyl groups color-coded. (b) Partial assignment of MILVA methyl groups of MNV P-domain. The sample contained 500 µM MILVA P-domain and saturating amounts of GCDCA. The spectrum was recorded on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K. The figure and figure legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

[^{1}H, ^{13}C] HMQC spectra of selectively A, LV, V, MI (Maaß, 2019), ILV, and MILVA labeled MNV P-domain samples were compared to assign resonances to the respective amino acid types (Fig. S 2.1). For alanine, all 29 expected methyl group resonances were identified. For leucine, only 28 of 32 expected resonances were observed. For methionine, valine, and isoleucine, 4 of 3, 26 of 24, and 16 of 12 expected resonances were identified. Lack of resonances can be explained by unfavorable relaxation properties at certain sides of the protein (Proudfoot et al., 2016). However, exceeding the number of expected resonance signals may indicate different protein forms or errors in the coding gene. Sequencing of the gene construct encoding the MNV P-domain excluded the latter explanation.

For assignment of resonances to specific methyl groups of the protein, a 4D HMQC-NOESY-HMQC spectrum was acquired for 7 days, providing NOEs as structural restraints. Together with the high-resolution crystal structure model (pdb 6e47, Nelson et al., 2018) and information on the amino acid types, methyl walks were identified. For this purpose, F1(^{1}H)/F2(^{13}C) planes (Fig. 5.2.2a) are extracted from the 4D spectrum with F3(^{1}H)/F4(^{13}C) frequencies of a particular methyl group resonance in the

2D [^{1}H,^{13}C] HMQC spectrum (Fig. 5.2.2c). The peaks appearing at F1=F4 and F2=F3 are auto peaks. The remaining peaks are NOE cross peaks and provide information about distances to the surrounding methyl groups. Setting the F3(^{1}H)/F4(^{13}C) frequencies to the frequency of the cross peak (red dashed arrow in Fig. 5.2.2a) leads to another F1(^{1}H)/F2(^{13}C) plane corresponding to a spatially neighboring methyl group. The identification of similar patterns in the structural model (Fig. 5.2.2b) allows to walk from methyl group to methyl group in the structure and from resonance to resonance in the 4D spectrum. With the support of AMIGO (c.f. chapter 5.1), methyl walks were identified, and 55 resonances were unambiguously assigned (Fig. 5.2.2d).

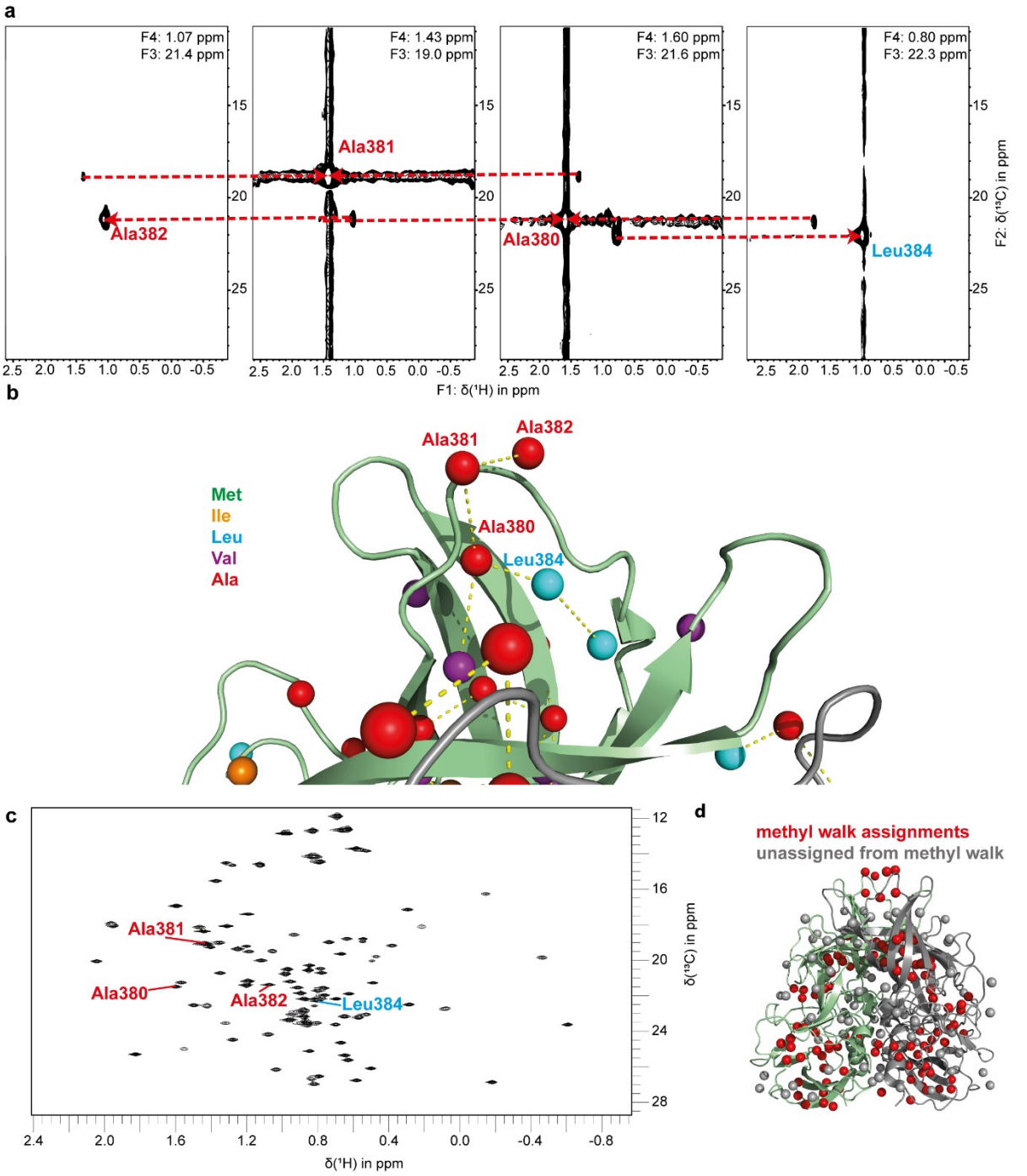


Fig. 5.2.2: Structure-based assignment of MILVA-labeled MNV P-domain methyl group resonances in [^{1}H,^{13}C] HMQC spectrum using a 4D HMQC-NOESY-HMQC spectrum and the methyl walk approach. By comparing structural restraints given by the 4D HMQC-NOESY-HMQC spectrum (a) to methyl-methyl distances in a high-resolution crystal structure (b) (pdb 6e47, Nelson et al., 2018), methyl group resonances in the [^{1}H,^{13}C] HMQC spectrum (c) were assigned to MNV P-domain methyl groups using the methyl walk strategy. (a) and (b) show an exemplary methyl walk from Ala382 to Leu384. The approach yielded unambiguous assignment of 55 resonances to P-domain methyl groups (d). Spectra were acquired with 500 µM MILVA-labeled P-domain at 298 K on a Bruker Avance III HD 600 MHz spectrometer with cryo probe in the presence of saturating amounts of GCDCA. The figure and the legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.2.2 Assignments by site-directed mutagenesis and metal ion titration reveals Pro361 cis/trans configuration

Due to the unfavorable spatial distribution of the methyl groups, certain loop regions could not be assigned using the methyl walk approach (Fig. 5.2.2d). However, these loop regions are of considerable biological interest as they provide binding sites for metal ions and the 300CDlf receptor (Nelson et al., 2018). Therefore, 11 isotope-labeled P-domain point mutant proteins (Tab. 4.4 and Fig. S 2.3) were generated and purified. The spectra were compared to the WT (wildtype) protein to obtain further assignments.

For five mutants, V304L, A442G, V387L, V378L, and A444G, a single methyl group resonance disappeared, making the assignment straightforward (Fig. S 2.3b-f). The A381G point mutant served as a control for the assignment of Ala381 by the methyl walk approach (Fig. S 2.3g). Notably, two resonances disappeared for V352L, I358L, A365G, I405L, and A446G (Fig. S 2.3 h-k and Fig. 5.2.3d). One of the two disappearing resonances for V352L was unambiguously assigned to V391 using the methyl walk strategy (Fig. S 2.2). The second resonance that disappears for A446G was assigned to Ala448 by PCSs (see below). In these cases, the disappearance of another peak is likely due to secondary effects of the mutation. Such effects have also been described for other proteins (Sprangers and Kay, 2007). Interestingly, for I405L, both resonances were assigned by methyl-methyl NOEs. Here, the I405L mutant served as a control and verified that both resonances belong to the same methyl group. Ala365 and Ile358 are located in the same region of the protein, but no independent assignment was available. As already suggested in chapter 5.2.1, this strongly indicates the presence of a second protein form.

To further test for a second protein form, metal ions were titrated to the protein, and chemical shift perturbations (CSPs) were followed in [^{1}H,^{13}C] HMQC spectra. As expected from the crystal structure model (Fig. 3.1.1), two binding sites per monomer were identified (Fig. S 2.4a). These binding sites were found to differ in affinity. A first set of resonances, including Ala442 and Ala444, reaches saturation at 550 µM $CaCl_2$ (Fig. S 2.4b and c; for assignment of Ala442, see below). This is in good agreement with a dissociation constant K_D of 138 µM reported in the literature (Creutznacher et al., 2021). A second set, including, e.g., Ala365 and Ile358, approached saturation at 20 mM $CaCl_2$ (Fig. S 2.4 d and e). Tracking CSPs induced by lanthanide ion titration revealed a similar pattern distinguishing the two binding sites. In contrast to $CaCl_2$, $LaCl_3$ induced fast exchange CSPs for one resonance and slow exchange CSPs for the other resonance presumably belonging to the same methyl group at the low-affinity binding site. This is exemplified in Fig S 2.5c for Ile358 and Ile405. This behavior was also observed for six other resonance pairs, including the pair corresponding to Ala365. For the remaining pairs, one of the two resonances was assigned by methyl-methyl NOEs (i.e., Met357, Val374, Ile389, Val414, and Ile449), and the other was spectrally directly adjacent to the assigned resonance (Fig. S 2.5b). These observations suggested the presence of a second form of the protein.

Different hypotheses were tested to identify the orign of the two forms. Based on resonance intensities of different samples of different ages, the population of both forms is approximately equal and constant over several months. Therefore, spontaneous asparagine deamidation as observed for human norovirus P-domain proteins (Mallagaray et al., 2019) was ruled out. Chemical exchange between the two forms was also excluded as the populations of the two forms were constant over a temperature range of 298 K to 323 K (spectra can be found in the data repository of this thesis). Surprisingly, examination of different crystal structure models revealed that Pro361 is present in the trans configuration in some structures (e.g., 3lq6, Taube et al., 2010), and in the cis configuration in other structures (e.g., 6e47, Nelson et al., 2018). Furthermore, in one structure (i.e. pdb 6xw5,

Koromyslova et al., 2020), the configuration is cis for one amino acid chain and trans for the other amino acid chain of the dimer (Fig. 5.2.3 a and b). Consequently, it can be concluded that the appearance of two resonances for a single methyl group represents the presence of Pro361 cis/trans isomers. The difference in the chemical shift is most pronounced for Ala365 (Fig. 5.2.3d). As can be seen by aligning the two amino acid chains in *pymol*, the methyl group is dislocated by about 10 Å (Fig. 5.2.3c). Mapping all methyl groups that are split into two peaks on the structural model clearly shows that they are located at the same site of the protein around Pro361 (Fig. S 2.5a).

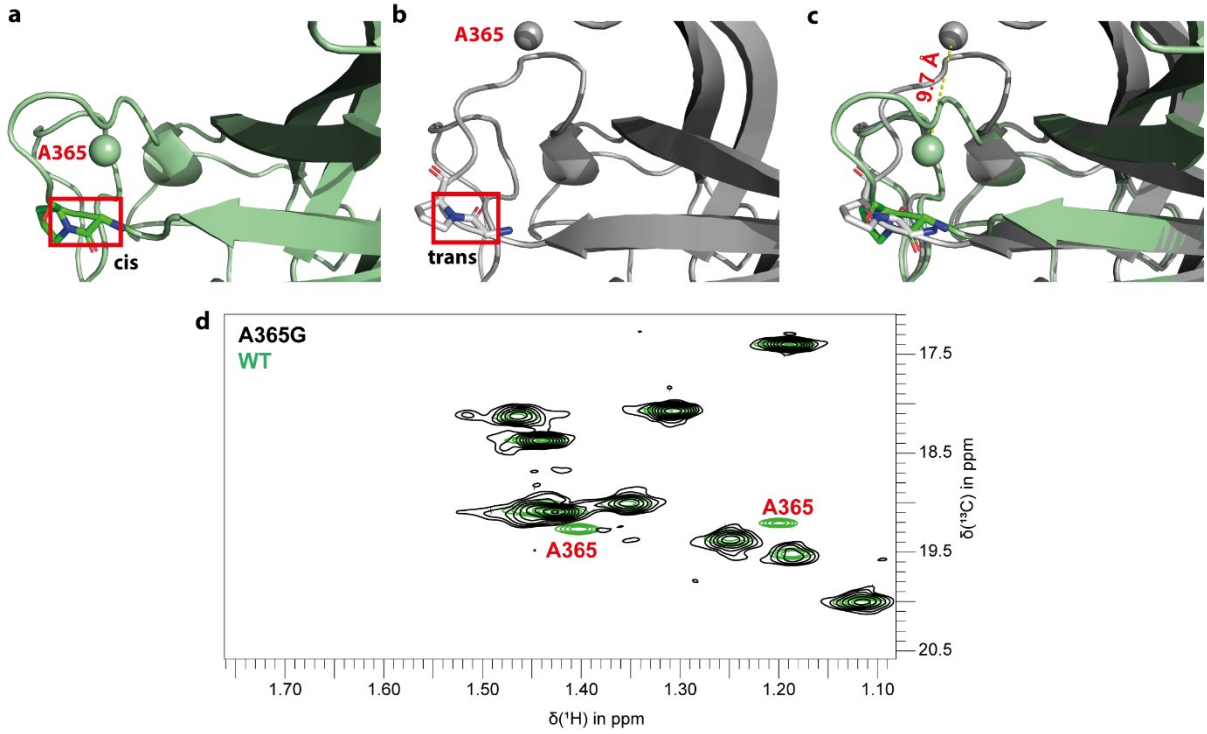


Fig. 5.2.3: MNV P-domain proteins are a mixture of Pro361 cis/trans isomers. (a) and (b) show the A and B chain of the MNV P-domain in the crystal structure model by Koromyslova et al., 2020 (pdb 6xw5). The red boxes indicate the Gly360-Pro361 peptide bond in the cis or trans configuration. In (c), an overlay of the two chains (RMSD 0.319 nm) is shown. The Ala365 methyl group is dislocated by ca. 10 Å due to the different Pro361 configurations. (d) Overlay of [^{1}H,^{13}C] HMQC spectra of WT (wildtype, green), and A365G mutant (black) P-domain of MNV shows disappearance of two peaks upon mutation. The observation of two peaks for one methyl group reflects the presence of a mixture of trans-Pro361 and cis-Pro361 isomers (see also main text and Fig. S 2.5). The spectrum of A365G was acquired with 45 µM A-labeled A365G P-domain and saturating amounts of GCDCA. For details of the spectrum of WT P-dimers, see Fig. 5.2.1. Both spectra were acquired at 298 K on a Bruker Avance III HD 600 MHz spectrometer equipped with a cryogenic probe. The figure and the legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.2.3 Assignments using pseudo contact shifts

Next to site-directed mutagenesis and the methyl walk approach, measurement of pseudo contact shifts (PCSs) provide another strategy to obtain methyl group assignments. In short, coupling with unpaired electrons of paramagnetic metals binding to the protein of interest leads to a shift in the methyl group chemical shift. The magnitude of the shift encodes structural information such as the length and orientation of the vector connecting the paramagnetic center (i.e. the position of the paramagnetic metal ion) with the methyl group of interest (Otting, 2010). If the anisotropic part of the magnetic susceptibility tensor Δχ and coordinates of the paramagnetic center are known, theoretical PCSs can be back-calculated from a structural model. The comparison of theoretically derived PCSs with measured PCSs can be used to identify and validate assignments (Flügge and Peters, 2018).

Analysis of PCSs of the MNV P-domain resonances required special attention to the concentration of paramagnetic ions and pre-processing of the structure. First, the P-domain dimer has four metal ion binding sites (see Fig S 2.4). Two binding sites have µM affinity for Ca^{2+} and two have mM affinity for Ca^{2+}. Assuming that paramagnetic metal ions have similar affinities, the paramagnetic metal ion concentrations should be around 400 µM to saturate only the high-affinity binding site. At this concentration, the low-affinity binding site is most likely not significantly occupied. Second, the aim was to use a single paramagnetic center in the fitting process. This allows the use of established Δχ and paramagnetic center parameter fitting protocols (Orton et al., 2020). Therefore, the structural model (pdb 6e47, Nelson et al., 2018) was processed using the high degree of dimer symmetry (Fig. 5.2.4a). Distances between each atom in the one monomer and the high-affinity metal ion binding to the same monomer (distance d_1, metal ion B) were compared to distances between this metal ion and the corresponding atom in the other monomer (distance d_2). Only those atoms with the smaller distance d_1 or d_2 were included in the processed structure (red mesh in the right panel of Fig. 5.2.4a).

^{1}H PCSs were measured in the presence of Ce^{3+}, Eu^{3+}, and Sa^{3+}. La^{3+} was used as a diamagnetic reference. The paramagnetic metals were chosen based on the axial component ΔX_{ax} of Δχ reported in the literature. The signs of ΔX_{ax} values for Ce^{3+} and Sa^{3+} are opposite to those for Eu^{3+}. Moreover, the absolute values are ca. 20-fold larger for Eu^{3+} and Ce^{3+} compared to Sa^{3+} (Otting, 2010). Therefore, Sa^{3+} was expected to shift resonances weakly in one direction, making it easy to trace the resonance back to the diamagnetic control. PCSs induced by Ce^{3+} should have the same sign but are much larger. For Eu^{3+} the amplitude is similar to Ce^{3+} but with the opposite sign. These properties were instrumental in identifying the vectors on which resonances shifted upon addition of the different paramagnetic metals (Fig. 5.2.4b).

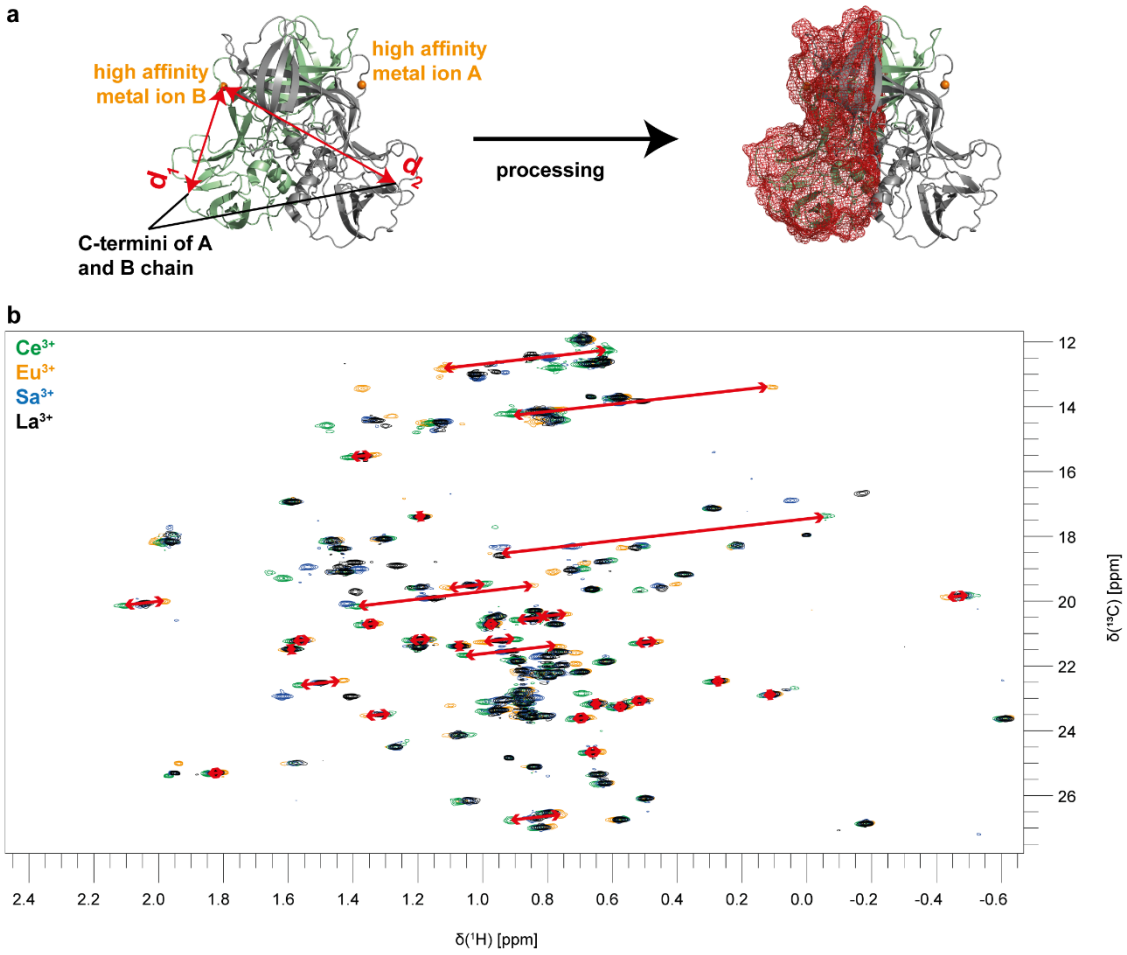


Fig. 5.2.4: Measuring PCSs and preparing a crystal structure model for tensor fitting. (a) crystal structure model of MNV P-dimers (pdb 6e47, Nelson et al., 2018) with bivalent metal ions shown as orange spheres. The A and B chain are shown in grey and in pale green. Only metal ions binding to the high-affinity binding sites with a dissociation constant for Ca^{2+} ions of 138 µM (Creutznacher et al. 2021) are shown. It is assumed that only these two binding sites are significantly occupied by lanthanide ion concentrations of 400 µM $LaCl_3$, 400 µM $SaCl_3$, 450 µM $EuCl_3$, and 400 µM $CeCl_3$. At these concentrations, occupation of the low-affinity metal ion binding sites can be neglected (see Fig. S 2.4). The structural model is processed for Δχ and paramagnetic center fitting using the program *Paramagpy* (Orton et al. 2020). For the fitting process, the distances d_1 and d_2 (red arrows) between the metal ion B to symmetrically positioned atoms in chains A (gray) and B (green) are compared. This is exemplified for the C-termini. Only the atom with the shorter distance is included in the fitting (illustrated by red mesh around the processed structure). (b) [^{1}H,^{13}C] HMQC spectra of MNV P-domain in the presence of saturating amounts of GCDCA and lanthanide ions. PCSs vectors are exemplarily highlighted by red arrows. The identification of vectors was carried out as explained in the text. The samples contained 31 µM MILVA-labeled P-domain, except for the sample containing Eu^{3+} where the concentration was 38 µM. Spectra were acquired on a Bruker 600 MHz Avance III HD spectrometer equipped with a cryogenic probe. The figure and the legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

Based on the experimental Ce^{3+} induced ^{1}H PSCs measured for 55 of the 63 assigned methyl groups, paramagnetic center and Δχ-parameters were fitted (Tab. 5.1 and Fig. 5.2.5a). A Q-factor of 0.12 confirmed the assignment and indicated that the structural model corresponds well with the structure in solution. The paramagnetic center was fitted close to the high-affinity metal ion found in the crystal

structure (Tab. 5.1). Although the γ-angle of the principal axis system was not well defined (168.6 ± 18.8 °), it was possible to assign Ile439 unambiguously (Fig. S 2.6). The corresponding methyl group is the only unassigned isoleucine methyl group within a 20 Å sphere around the high-affinity metal ion (Fig. S 2.6c). The only unassigned isoleucine methyl group resonance sufficiently affected by addition of paramagnetic metals (Fig. S 2.6a) and by titration with bivalent metal ions (Fig. S 2.6b) was therefore assigned to Ile439. Inclusion of the new assignment in a second round of fitting resulted in a similar Δχ and paramagnetic center, but a much better-defined γ-angle (Tab. 5.1 and Fig. 5.2.5b).

Theoretical PCSs derived from the second round of fitting provided new information to obtain further assignments. For the A446G P-domain mutant, two alanine resonances disappeared (see above and Fig. S 2.7a). However, these resonances did not respond to lanthanide titration in agreement with the other resonances, which are split into two signals due to the Pro361 cis/trans isomerism (see above). This indicates that secondary effects of the mutation cause the disappearance of the other resonance in this case. Consequently, it is likely that the other resonance disappearing corresponds to a methyl group in close spatial proximity. This is confirmed by CSPs upon bivalent metal ion titration. Both resonances most likely belong to methyl groups at the high-affinity metal ion binding site (Fig. S 2.7c), restricting the assignment possibilities of the two resonances to the remaining unassigned alanine methyl groups of Ala445, Ala446, and Ala448 located in the metal ion binding loop (Fig. S 2.7b). The comparison of theoretical and measured PCSs induced by Ce^{3+} allows the assignment of the two resonances to Ala446 and Ala448 (Fig. S 2.7d).

Calculated PCSs for the remaining two unassigned isoleucine methyl groups are +0.034 ppm for Ile310 and -0.012 ppm for Ile514 (Fig. S 2.8a). This is in good agreement with measured PCSs of the two remaining unassigned isoleucine resonances (+0.029 ppm and -0.009 ppm, Fig. S 2.8b), allowing them to be assigned. Re-analysis of the Ile514 resonance using the 4D-HMQC-NOESY-HMQC spectrum resulted in the assignment of Val234 (Fig. S 2.8 c and d). Using the new assignments for a third round of Δχ and paramagnetic center parameter fitting further improved the Q-factor, yielded well-defined tensor parameters, and locates the paramagnetic center very close to the metal ion in the crystal structure (Tab. 5.1 and Fig. 5.2.5c).

Tab. 5.1: Parameters of Δχ fitting for Ce3+.

		First round	Second round	Third round
ΔX_{ax} in 10^{-32} m^3		-1.60±0.05	-1.60±0.05	-1.60±0.05
ΔX_{rh} in 10^{-32} m^3		-0.66±0.07	-0.69±0.07	-0.74±0.04
Coordinates of origin in Å	X	-16.4±0.9	-16.6±0.8	-16.3±0.5
	Y	2.8±1.4	2.9±0.7	2.9±0.4
	Z	-44.9±1.4	-45.4±1.1	-45.3±0.7
Orientation of principal axis of tensor in °	α	109.8±0.7	107.5±1.0	109.6±0.3
	β	66.0±3.2	64.9±3.1	66.8±2.4
	γ	168.6±18.8	155.6±4.1	155.9±2.0
Amount of methyl groups included		55	56	61
Q-factor		0.12	0.12	0.10
Distance to metal ion position in the crystal structure in Å*		2.2	2.1	1.9

*For errors of the position of the paramagnetic center, see row for coordinates of origin. The resolution of the crystal structure is 2 Å (pdb 6e47, Nelson et al., 2018).

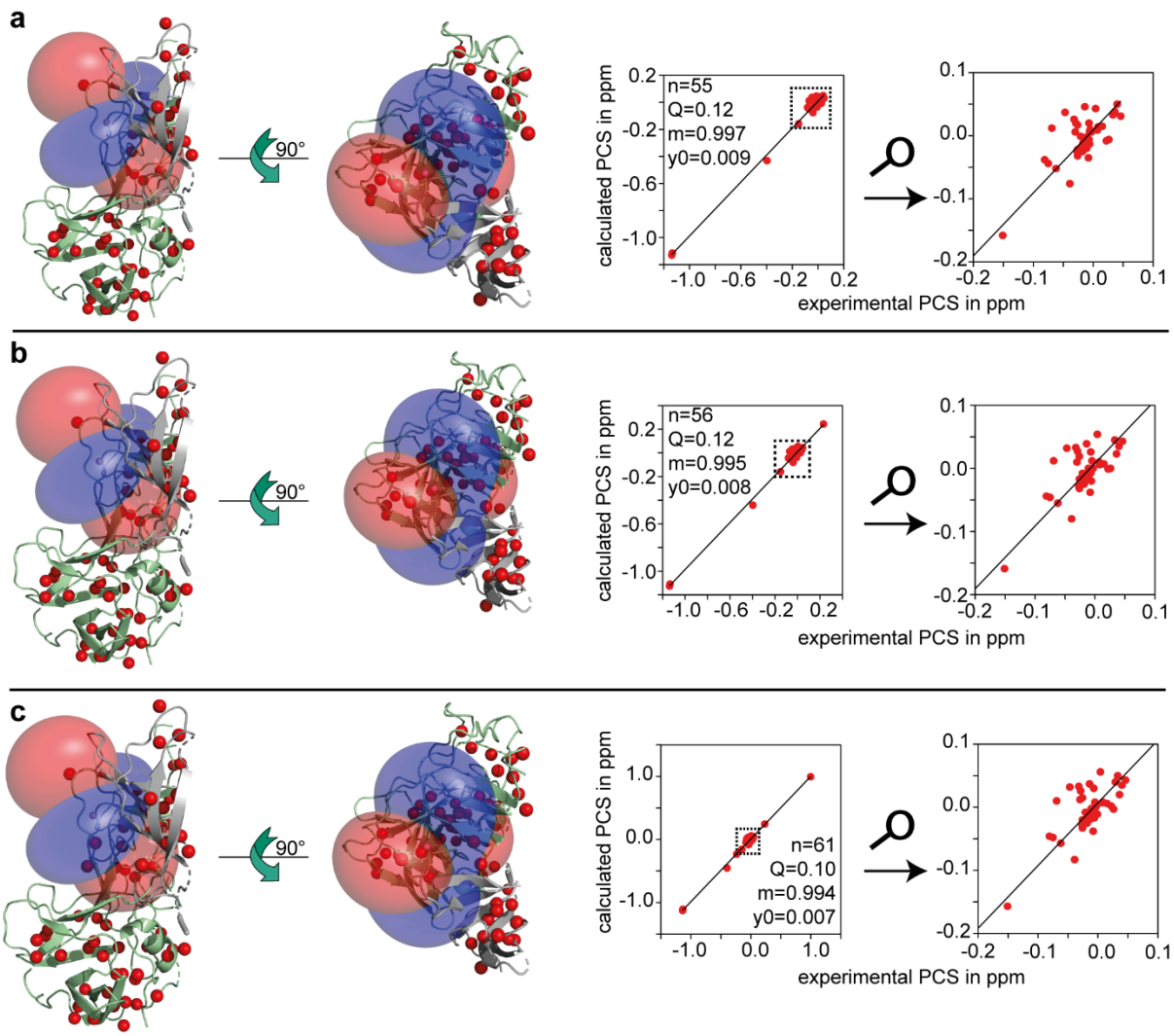


Fig. 5.2.5: Assignment of [^{1}H,^{13}C] methyl group resonances of MNV P-domain using Ce^{3+} PCSs. (a) Using experimental ^{1}H PCSs of 55 methyl groups assigned via methyl-methyl NOEs (Fig. 5.2.2) and point mutations (Fig. S 2.3), Δχ and the paramagnetic center parameters for Ce^{3+} were determined (Tab. 5.1, "first round"). As a result, Ile439 assignment using PCSs and metal ion titration was possible (Fig. S 2.6). (b) In a second round of fitting, the PCS of Ile439 was included (Tab. 5.1, "second round") resulting in a better-defined γ-angle and enabling further assignments explained in the text and in Figs. S 2.7 and S 2.8. (c) A third round of fitting included PCSs of the newly assigned methyl groups (i.e., Ala446, Ala448, Ile514, Ile310, and Val234), resulting in a well-defined γ-angle and a Q-factor of 0.10 (Tab. 5.1). For details of the assignment process, see the main text. The figure and the legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

Theoretical and experimental PCSs induced by Sa^{3+} and Eu^{3+} were consulted to validate the assignment further. To scrutinize the impact of the second high-affinity metal ion (metal ion A in Fig. 5.2.4a or Fig. 5.2.6a), atoms were gradually excluded based on different criteria: In the first round, atoms with the smaller distance of d_1 or d_2 were included (see above). Here, the restriction criterion $|d_1-d_2| > 0$ Å does not lead to the exclusion of further atoms (Fig. 5.2.6 b.i). In the second round, only atoms were included that additionally fulfilled the condition $|d_1-d_2| > 2.5$ Å (Fig. 5.2.6 c.i). In the third round, the criterion was set to $|d_1-d_2| > 5$ Å (Fig. 5.2.6 d.i). As PCSs show an r^{-5} dependence, where r is the distance

between the paramagnetic center and the atom in question, it was expected that the remaining atoms would be more and more exclusively affected by metal ion B. This and the correctness of the assignment are very well reflected in the resulting Q-factors (Fig. 5.2.6 b-d, ii-iv).

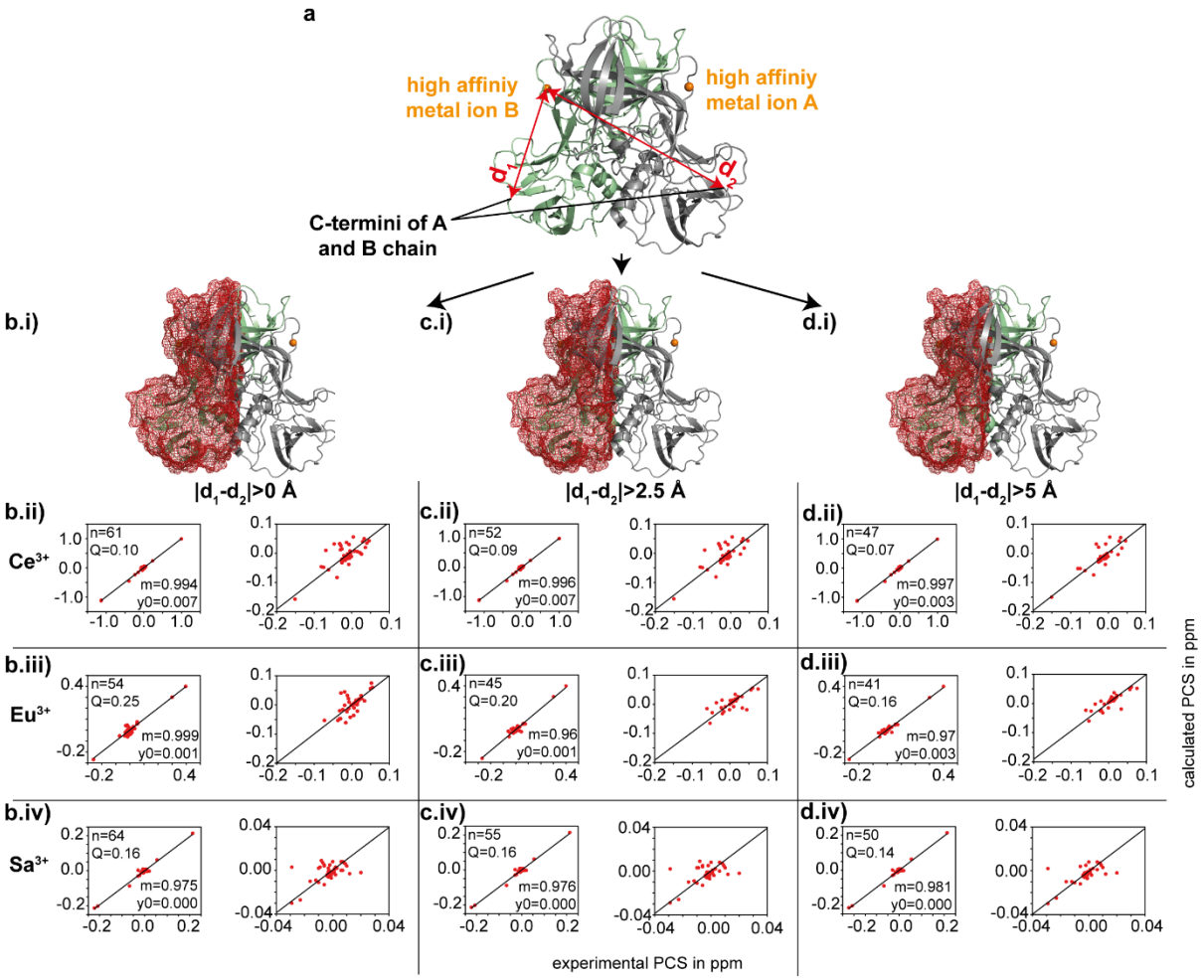


Fig. 5.2.6: Validation of assignment using PCSs of Ce^{3+}, Eu^{3+}, and Sa^{3+}. The structural model with metal ions shown as orange spheres((a), pdb 6e47, Nelson et al., 2018) was processed to fit Δχs with *Paramagpy* to a single paramagnetic center as described before with a further criterion. For the fitting of parameters, only atoms were considered with the smaller of the two distances d_1 and d_2 and fulfilling the condition $|d_1-d_2|>0$ Å (covered by the red mesh in (b.i), same processing as in Fig. 5.2.4 a). The criterion for structure processing was altered to $|d_1-d_2|>2.5$ Å (atoms covered by red mesh shown in (c.i)) and to $|d_1-d_2|>5$ Å (atoms covered by red mesh shown in (d.i)). As PCSs depend on r^{-5} (i.e. the distance between a nucleus of interest and the paramagnetic metal ion), the influence of the metal ion binding to the A chain (metal ion A in (a)) on the calculation of the Δχ and the coordinates of the paramagnetic center is decreased due to structural processing conditions from (b.i) to (d.i). This is reflected in the Q-factors of the Δχ tensors of Ce^{3+} (b.ii-d.ii), Eu^{3+} (b.iii-d.iii) and Sa^{3+} (b.iv-d.iv). n is the amount of assigned methyl group resonances for that PCSs were measured and therefore used to determine the respective Δχ and paramagnetic center. Tensor and paramagnetic center parameters are given in Tab. S 2.1 The figure and the legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.3 Thermodynamics, kinetics, and CSP analysis of MNV P-domain dimerization and GCDCA binding

Recent experiments have shown that MNV P-domains in solution exist in an equilibrium of monomers and dimers. Remarkably, the equilibrium is completely shifted to the dimeric state by the binding of a small molecule, GCDCA (Creutznacher, 2020). GCDCA binding and P-domain dimerization in solution correlates with immune evasion and capsid contraction in the virus context (Sherman et al., 2019; Williams et al., 2021a; Creutznacher et al., 2022a). At the time, no MNV P-domain NMR assignment was available. This prevented a deeper understanding of the underlaying mechanisms. With the assignment in hand (c.f. chapter 5.2), the methyl group labeled P-domain provides a possibility to close this gap. In this chapter, the nearest neighbour approach (Williamson, 2013) was used to assign apo P-domain resonances. Detailed analysis of 2D line shapes and chemical shift perturbations (CSPs) allowed characterization of thermodynamics and kinetics of dimerization and GCDCA binding, as well as insights into allosteric effects caused by GCDCA binding.

5.3.1 Apo P-domain assignment

As explained in chapter 5.2, direct assignment of the apo P-domain is impeded by the existence of monomers and dimers in solution. Therefore, all assignment experiments were carried out in the presence of saturating amounts of GCDCA. However, observing chemical shifts of methyl groups using different GCDCA levels allowed the assignment to be transferred to 58 apo P-domain resonances (Tab. S 3.1). Many resonances are in slow exchange on the NMR time scale (Fig. 5.3.1a). Here, Euclidian distances between the apo and the GCDCA bound P-domain differ in the range of 50 to 150 Hz. Backtracking to the apo form using the nearest neighbor approach (Williamson, 2013) was only possible if the resonances occur sufficiently isolated spectral regions. Other resonances are in intermediate (Fig. 5.3.1 bi and bii) or fast-to-intermediate (Fig. 5.3.1 b.iii) exchange on the NMR time scale. They were assigned by following chemical shifts at different GCDCA concentrations (Williamson, 2013). Assignments of resonances that do not undergo chemical shift perturbations were directly transferred from the GCDCA bound form (Fig. 5.3.1c). It is noteworthy that the obtained chemical exchange regimes limit the exchange rate of GCDCA binding to 10 Hz < $k_{exchange}$ < 100 Hz (Fig. 5.3.1).

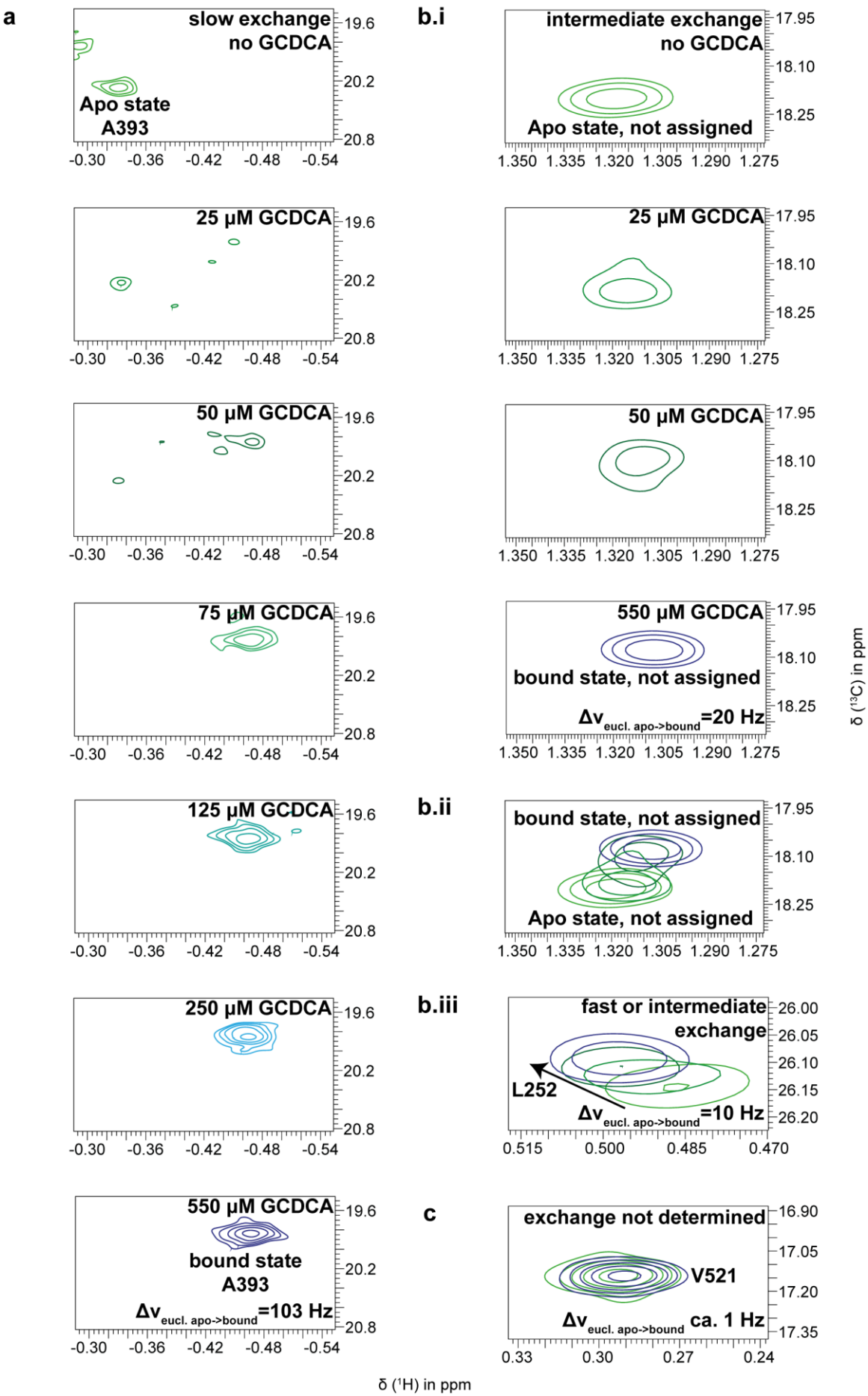

a
slow exchange
no GCDCA
Apo state
A393
25 µM GCDCA
50 µM GCDCA
75 µM GCDCA
125 µM GCDCA
250 µM GCDCA
550 µM GCDCA
bound state
A393
Δν eucl. apo->bound =103 Hz
b.i
intermediate exchange
no GCDCA
Apo state, not assigned
25 µM GCDCA
50 µM GCDCA
550 µM GCDCA
bound state, not assigned
Δν eucl. apo->bound =20 Hz
b.ii
bound state, not assigned
Apo state, not assigned
b.iii
fast or intermediate
exchange
L252
Δν eucl. apo->bound =10 Hz
c
exchange not determined
V521
Δν eucl. apo->bound ca. 1 Hz
δ (13C) in ppm
δ (1H) in ppm

Fig. 5.3.1: [^{1}H,^{13}C] HMQC resonances reflect slow to fast-to-intermediate exchange upon titration of P-domain with GCDCA. (a) Slow exchange: The Euclidian distance $\Delta\nu_{Eucl}$ (103 Hz) between free and bound state is much larger than the GCDCA exchange rate $k_{exchange}$, leading to separate signals for the bound and the free state of the P-domain. (b.i-ii) Slow-to-intermediate exchange: The Euclidian distance $\Delta\nu_{Eucl}$ (20 Hz) between free and bound state is slightly larger than the exchange rate, causing signals of low intensity being visible during the titration. Note that b.i and b.ii show the same resonance. (b.iii) Fast or fast-to-intermediate exchange: Signals shift during titration with slightly varying intensities, indicating that Euclidian distance $\Delta\nu_{Eucl}$ (< 10 Hz) is smaller than the exchange rate. (c) The Euclidian distance $\Delta\nu_{Eucl}$ is low. Therefore, this peak would be classified as showing no CSP. It can be concluded that the exchange rate should be larger than 10 Hz and below 103 Hz, which agrees with on- and off-rates given in chapter 5.3.3. For details about the acquisition and sample compositions, see Fig. 5.3.4. Figure and legend are adapted from Creutznacher et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.3.2 2D line shape analysis of P-domain dimerization in solution

Chemical exchange between P-domain monomers and dimers is slow on the NMR time scale, so that assignment transfer to the monomeric form is difficult. However, apo dimer resonances of 9 assigned and 4 unassigned methyl groups (Fig. 5.3.2a) were sufficiently isolated. For these resonances, a corresponding monomer resonance in close spectral proximity was unambiguously identified by applying the nearest neighbour approach (Williamson, 2013). The spectral regions of these monomer-dimer resonance pairs were quantum-mechanically simulated with the program *TITAN* (Waudby et al., 2016) and fitted to measured spectra at different P-domain concentrations. A simple protein dimerization model was used for this analysis (Fig. 5.3.2c). The 3D representations of measured and fitted spectral regions corresponding to Val269 are shown in Fig. 5.3.2b (see Fig. S 3.1 for all spectral regions). While the relative intensity of the monomer resonance increases, the intensity of the dimer resonance decreases with falling protein concentrations. The analysis yielded an off-rate $k_{off;Dimerization}$ of 1.25 ± 0.21 $s^{-}1$ and a dissociation constant $K_{D;Dimerization}$ of 6.9 ± 0.6 µM.

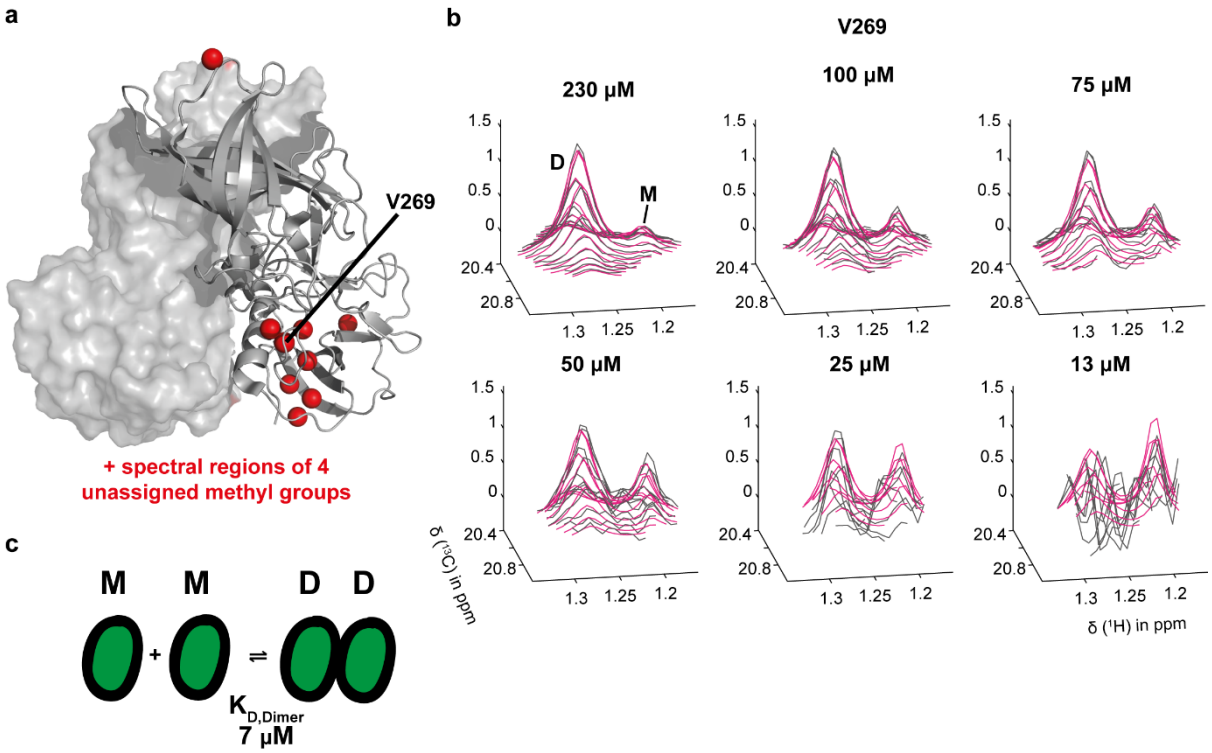


Fig. 5.3.2: *TITAN* 2D line shape analysis of P-domain dimerization yields thermodynamic and kinetic parameters. (a) shows methyl groups for that the spectral regions were quantum-mechanically simulated and fitted to measured spectra using *TITAN* (Waudby et al., 2016). (b) shows an overlay of the fitted (pink) and measured (grey) spectra of different P-domain concentrations for the spectral region corresponding to Val269. Assuming a two-state dimerization model (c) yields an off-rate $k_{off;Dimerization}$ of 1.25 ± 0.21 s^{-1} and a dissociation constant $K_{D;Dimerization}$ of 6.9 ± 0.6 µM. Resonances in the spectral regions corresponding to the monomeric and dimeric forms are labeled, respectively. Note that the 3D visualization was chosen here as the signal-to-noise ratio of spectra with low protein concentrations resulted in very noisy 2D top-view visualizations used in other figures of this thesis. Spectral regions of the other 12 methyl groups used in the analysis are shown in Fig. S 3.1. Spectra were recorded on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K. The figure is adapted from Creutznacher et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.3.3 CSP and 2D line shape analysis of P-domain GCDCA binding

Aiming at characterizing the P-domain GCDCA interaction, chemical shift perturbations (CSPs) between apo P-domain and GCDCA bound P-domain resonances were determined (Tab. S 3.1). Three of the 58 assigned apo P-domain resonances were not sufficiently resolved, e.g., Leu232 (Fig. 5.3.3a), and CSPs were not determined. For 12 of the 77 assigned GCDCA bound P-domain resonances, the exchange is slow and no apo peak was found in the corresponding spectral region (Fig. 5.3.3b). The CSP was assumed to be larger than 20 Hz. An alternative explanation might be peak broadening beyond detection as a result of altered dynamics. However, this would also report for changes in the respective protein region and therefore not affect the validity of conclusions (Williamson, 2013). Seven of the 77 assigned peaks were in crowded spectral regions. Assignment transfer was not possible, so that no CSP determination was possible either.

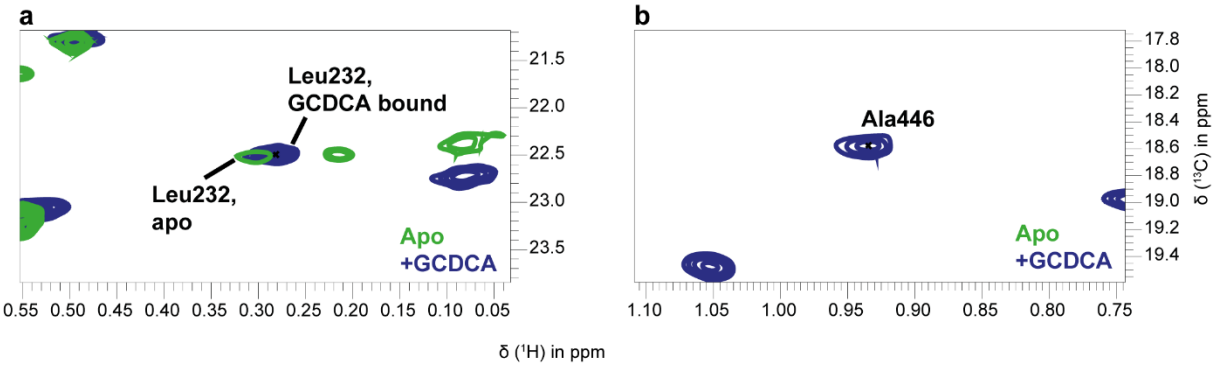


Fig. 5.3.3: Overlay of sections of [^{1}H,^{13}C] spectra of P-domain with GCDCA bound and in the apo form. (a) The Leu232 resonance of the apo-form is not sufficiently resolved. Therefore, CSPs cannot be determined. (b) The resonance of Ala446 is an example of well-isolated resonances where no corresponding apo signal can be identified in close spectral vicinity (<20 Hz). It is concluded that the apo signal shows a CSP with a Euclidian distance $\Delta\nu_{Eucl}$ larger than 20 Hz. Alternatively, the apo-state signal might be broadened beyond detection. This would also report for changes at the respective protein region and would not distort conclusions (Williamson, 2013). For details about the acquisition and sample compositions, see Fig. 5.3.4. Legend and figure are adapted from Creutznacher et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

CSPs obtained from [^{1}H,^{13}C] HMQC spectra (Fig. 5.3.4a) were classified according to the magnitude of the Euclidean distances $\Delta\nu_{Eucl}$ and plotted on the structural model (Fig. 5.3.4b). As expected, most resonances with larger CSPs belong to methyl groups around the GCDCA binding pocket, e.g., Met436 or Ala393 (Tab. S 3.1). However, some of these methyl groups are located more than 20 Å away from the GCDCA, and CSPs clearly reflect long-range effects. These long-range effects can be observed for, e.g., Ala365 and Ala380 (Fig. 5.3.4c). According to published structural models, the corresponding protein regions constitute binding epitopes for the receptor and monoclonal antibodies (Nelson et al., 2018; Williams et al., 2021a). Other methyl groups that clearly exhibit CSPs are located at the lower part of the P-domain. Interestingly, these regions hardly differ in available structural models with and without GCDCA (Fig. 3.1.2b and c). Indirect effects of GCDCA instead of effects caused by GCDCA binding to P-domain are unlikely and can be excluded, since the spectral effects are reversible (Fig. S 3.2).

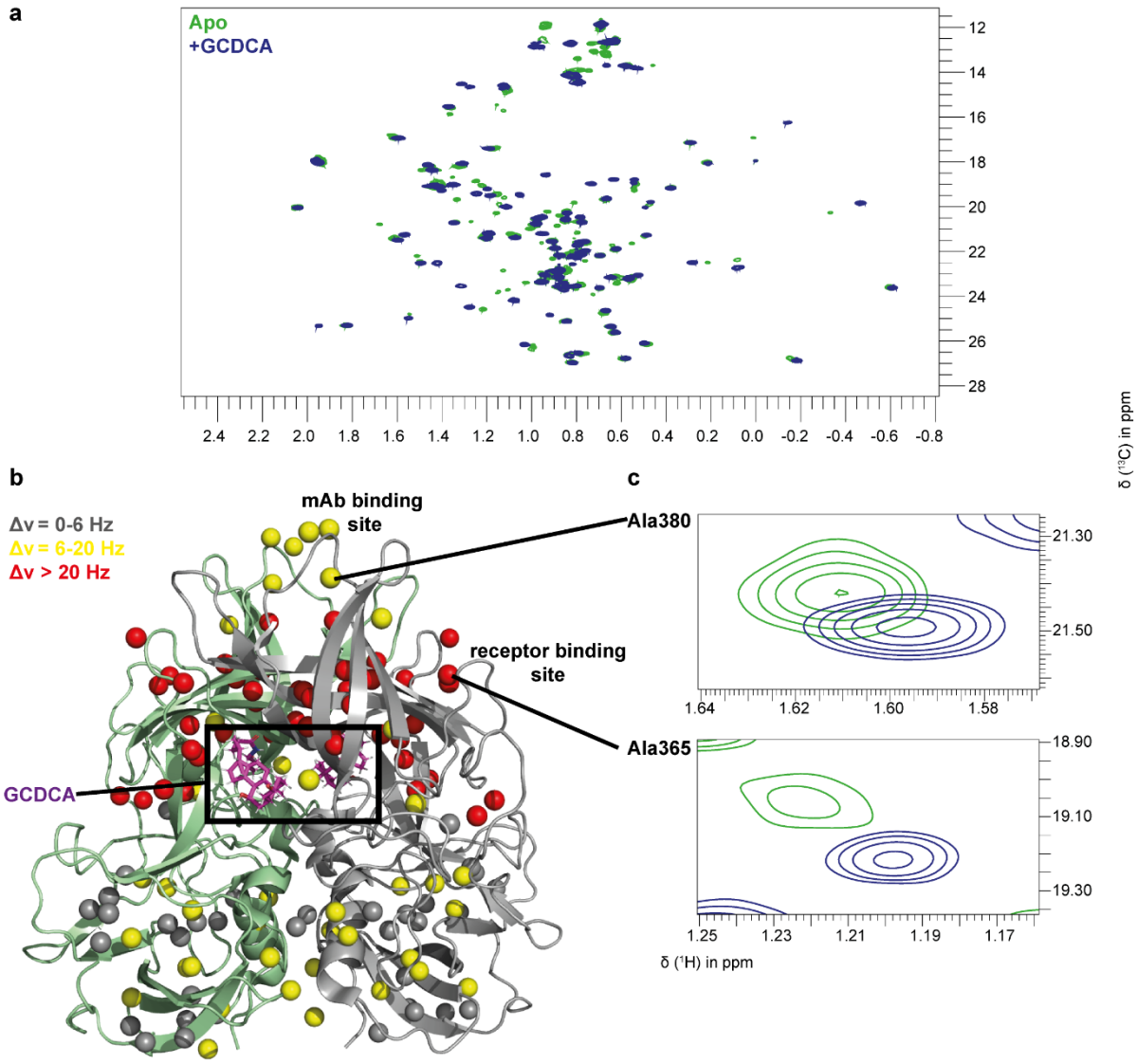


Fig. 5.3.4: Chemical shift perturbations (CSPs) in MNV CW1 P-domains upon GCDCA binding. [^{1}H,^{13}C] HMQC spectra of 50 µM methyl group-labeled P-domains titrated with 550 µM GCDCA (a, green and blue spectra, respectively) reveal major changes upon GCDCA binding reflected by CSPs. CSPs were quantified as low, (grey), medium (orange), or strong (red) CSPs Δν and mapped on the crystal structure model of the MNV P-domain (b, pdb 6e47, Nelson et al., 2018). Significant CSPs can be observed in regions up to 20 Å apart from the GCDCA binding pocket at the binding site for monoclonal antibodies (mAb; Williams et al., 2021a) or at the receptor binding site (Nelson et al., 2018) (b). Spectra were recorded on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K. Legend and figure are adapted from Creutznacher et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

To not only localize GCDCA effects on P-domain but to obtain insights into thermodynamics and kinetics of the binding reaction, 2D line shape analysis with *TITAN* (Waudby et al., 2016) was performed. Recent studies in the laboratory revealed that only the dimeric and not the monomeric P-domain binds GCDCA. Thereby, the monomeric species in solution is depleted by GCDCA binding (Creutznacher, 2020). That only the dimeric P-domain binds GCDCA is supported by the fact that the binding pocket is formed by both amino acid chains of the dimeric P-domain (Fig. 5.3.4b; Nelson et

al.,2018). The study by Creutznacher, 2020, also identified three P-domain states in unassigned [^{1}H,^{15}N] TROSY HSQC spectra. The states reflect the monomeric, the dimeric, and the GCDCA bound P-domain. However, further analysis was impeded by crowded spectra and the lack of any resonance assignments. For a subset of 11 methyl groups (Fig. 5.3.5), the three states are distinguishable in less crowded [^{1}H,^{13}C] HMQC spectra of the P-domain. The first state belongs to the monomeric protein ("M" in Fig. 5.3.5), the second state corresponds to the dimeric form ("D" in Fig. 5.3.5), and the third one reports for the GCDCA bound dimeric P-domain ("DL" in Fig. 5.3.5). The simplest binding model would therefore have three states (Fig. 5.3.5a, three-state model). Assuming no inter-subunit communication, P-domain dimers with only one GCDCA molecule bound would not result in extra peaks reporting for additional states. In contrast, strong cooperativity could explain the absence of such extra peaks. Taking this consideration into account in a second model (Fig. 5.3.5b, five-state model), additional states characterized by P-domain dimers with a single bound GCDCA molecule (DL* and D* in Fig. 5.3.5) would be only weakly populated and thus hidden in the spectral noise.

For the equilibrium between monomers and dimers an off-rate $k_{off;Dimerization}$ of 1.25 ± 0.21 s^{-1} and dissociation constant $K_{D;Dimerization}$ of 6.9 ± 0.6 μM were determined (c.f. Fig. 5.3.2). The three-state and five-state models and the parameters derived for the monomer-dimer equilibrium were used for 2D line shape analysis with *TITAN* (Waudby et al. 2016). Spectral regions of the 11 methyl groups were quantum-mechanically simulated and fitted to measured [^{1}H,^{13}C] HMQC spectra of the P-domain at different GCDCA concentrations.

The three-state model yielded a dissociation constant $K_{D,GCDCA}$ of 10.5 ± 0.5 μM and an off- rate $k_{off,GCDCA}$ of 25.8 ± 3.5 s^{-1}. The $K_{D,GCDCA}$ of 11 μM and the off-rate of $k_{off,GCDCA}$ of 26 s^{-1} translate into an on-rate of ca. 2*10^{6} M/s. This agrees well with the exchange rate estimated from chemical exchange regimes (10 Hz < $k_{exchange}$ < 105 Hz, Fig. 5.3.1).

For the analysis using the five-state model, chemical shifts of singly bound P-domain dimer states (D* and DL*) were set to arbitrary positions in the spectral noise. The off-rate and the dissociation constant of the second binding event are defined by Eq. 33:

$$K_{D2,GCDCA} = K_{D1,GCDCA} \cdot 10^{\alpha} \text{ and } k_{off2,GCDCA} = k_{off1,GCDCA} \cdot 10^{\beta}$$

Eq. 33

with $K_{D2,GCDCA}$ and $k_{off2,GCDCA}$ being the dissociation constant and the off-rate of the second binding event and $K_{D1,GCDCA}$ and $k_{off1,GCDCA}$ being the respective parameters of the first binding event. A first parameter fitting approach resulted in a β of zero, indicating similar off-rates for both binding events. Since the uncertainties for most parameters were rather high, β was held constant at zero at a second approach to reduce the number of parameters. This approach yielded values of $K_{D1,GCDCA}$ = 3 ± 2 mM, k_{off} = 14 ± 32 s^{-1}, and α = - 4.7 ± 0.6. An α of - 4.7 ± 0.6 reflects the approach-intrinsic strong cooperativity and results in a $K_{D2,GCDCA}$ of ca. 60 nM (Fig. 5.3.5).

For both models, simulated spectra are very similar to the measured ones showing that the two models adequately describe the data (Fig. 5.3.5c). However, parameters deduced by the five-state model are not as sharply defined as for the three-state model, indicating that the fitting problem for the five-state model is underdetermined. Aiming at both, sufficiency and simplicity, the three-state model is the favored model.

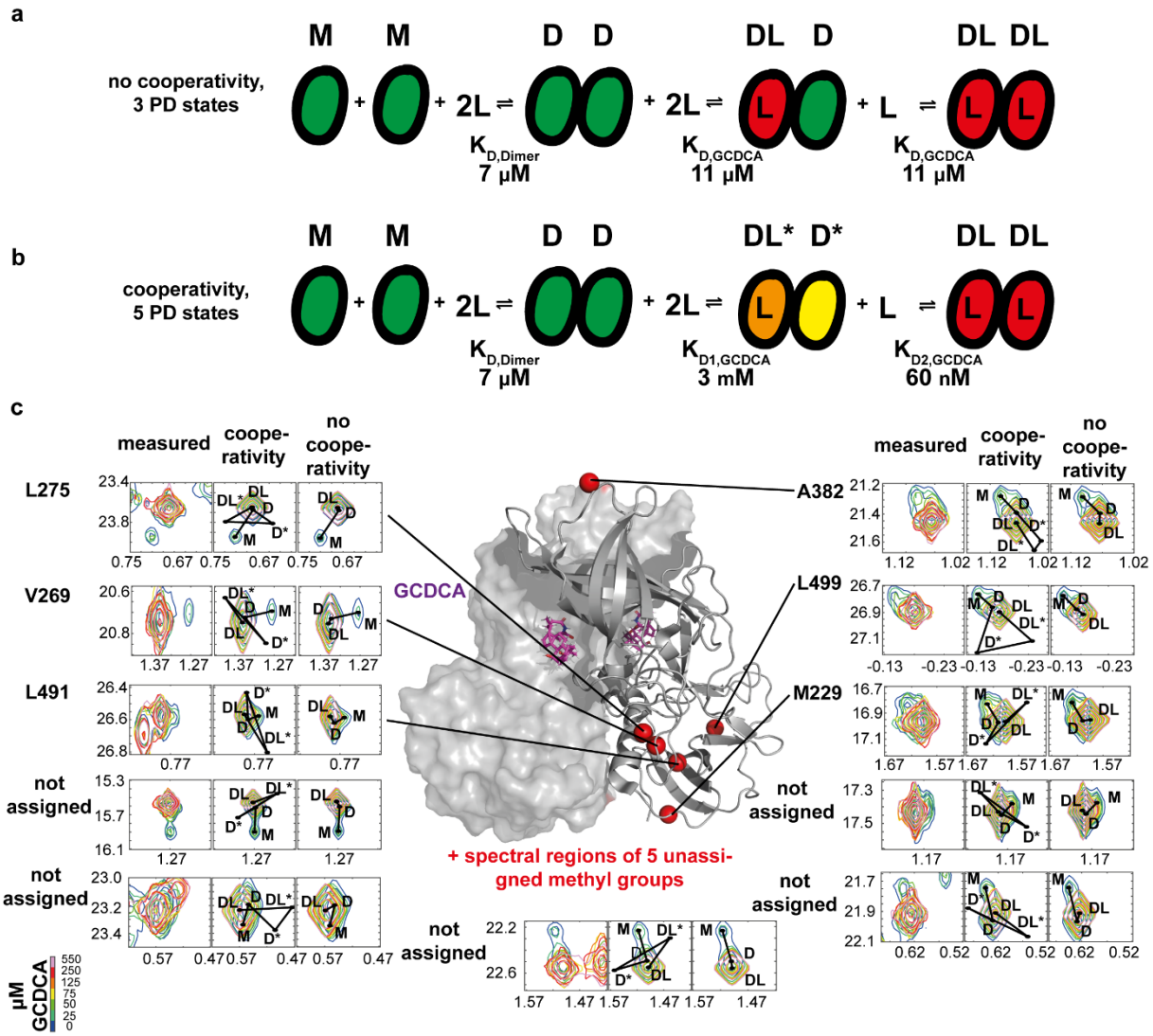


Fig. 5.3.5: Thermodynamic and kinetic characterization of possible GCDCA binding models using 2D line shape analysis with *TITAN*. (a) [^{1}H,^{13}C] HMQC-based GCDCA titration data (Fig. 5.3.4) match two binding models: a three-state binding model in which dimerization is followed by two consecutive, independent binding events with a dissociation constant K_D of 10.5 µM and a model with strong cooperativity in which binding of the first GCDCA molecule increases the affinity for the second GCDCA molecule. In both models, the monomeric protein (M) is binding-incompetent and is depleted at higher ligand concentrations by removal of unbound dimer (D) through subsequent binding of GCDCA. (b) Spectral regions of 11 methyl groups in which monomeric (M), dimeric (D), and ligand-bound states (DL) were at least partially resolved were fitted to both models using *TITAN* (Waudby et al., 2016; Creutznacher et al., 2022a). Note, that strong cooperativity inherently implies the existence of two unobservable P-domain states (D* and DL*) of the singly-bound dimer. Signal positions corresponding to these states were positioned randomly during the fit procedure. The dissociation constant $K_{D,Dimer}$ was determined independently by *TITAN* analysis of a P-domain concentration series (Fig. S 3.1 and Fig. 5.3.2). For details about the acquisition and sample compositions, see Fig. 5.3.4. Figure and legend are adapted from Creutznacher et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

5.4 P-domain dimerization and GCDCA binding is controlled by bivalent metal ion binding

The results from the previous chapter and results presented in Creutznacher, 2020, should also be considered in the context of cryo-electron microscopy and X-ray crystallography studies (Taube et al., 2010; Nelson et al., 2018; Sherman et al., 2019). Taken together, these studies provide a picture in which GCDCA simultaneously causes structural rearrangements, tight P-domain dimer interactions, and capsid contraction. Furthermore, these changes are associated with immune evasion (Williams et al., 2021a; Creutznacher et al., 2022a). At the time, other cryo-electron microscopy studies (Song et al., 2020; Snowden et al., 2020) were performed in the presence of bivalent metal ions. As in the presence of GCDCA, the resulting structural models showed the MNV capsid in the contracted conformation. However, metal ions were not unambiguously identified as the cause of the contraction. The four metal ion binding sites per dimer seen in a crystal structure model have already been validated in NMR experiments for MNV P-domain assignment in presence of GCDCA (c.f. Fig. S 2.4; Fig. 3.1.2, Nelson et al., 2018). This chapter aims at a better understanding of interactions between the P-domain and metal ions. The resonance assignment of the metal ion bound P-domain was transferred from spectra of P-domains with GCDCA bound and with GCDCA and metal ions bound. NMR spectroscopy and size exclusion chromatography revealed the formation of dimers upon increasing metal ion concentrations. Long-range chemical shift perturbations (CSPs) were observed at the antibody and receptor binding sites. Point mutants were created to selectively switch off the individual metal ion binding sites of the P-domain. This allowed the localization of the binding site responsible for the observed long range effects and dimerization at the G'H' loop. Using intensities, line shapes, and CSPs in [^{1}H,^{13}C] HMQC spectra of the P-domain at different metal ion and GCDCA concentrations showed thermodynamic and kinetic cross-talks between bivalent metal ion and GCDCA binding. PCSs and methyl group order parameters were determined to characterize GCDCA effects on the metal ion-bound P-domain. Analysis of CSPs of P-domain bound to different metal ions and GCDCA suggests the existence of an allosteric network-like connection of methyl groups throughout the protein.

5.4.1 Assignment of [^{1}H,^{13}C] methyl resonance signals of Mg^{2+} and Ca^{2+} bound P-domain states

For CSP and line shape analysis, the assignment of metal ion bound P-domain resonances is required. Four P-domain states with different exchange regimes between the specific states were considered, i.e., the apo P-domain dimer (D), the metal ion bound P-domain (D'), the GCDCA bound P-domain (DL) and the metal ion and GCDCA bound P-domain (DL') (Fig. 5.4.1). The exchange between the assigned GCDCA bound state and the GCDCA and Mg^{2+} or Ca^{2+} bound states is fast on the NMR time scale (c.f. Fig. 5.4.7c). Therefore, assignment was straightforward for 75 of the 77 assigned resonances. Only resonance assignments of Leu359 and Leu384 were not transferred due to spectral overlap (Fig. S 4.1a). The exchange between the metal ion bound and GCDCA and metal ion bound states is slow (c.f. Fig. 5.4.10d). However, for both states the P-domain is exclusively present as a dimer (see below) and most resonances are sufficiently resolved and isolated (c.f. Fig. 5.4.10a). The resonances of Mg^{2+} or Ca^{2+} bound P-domains were compared with the resonances of the GCDCA bound P-domain, the GCDCA and Mg^{2+} bound P-domain, and the GCDCA and Ca^{2+} bound P-domain. This allowed the assignment transfer for 74 resonances to the Mg^{2+} bound P-domain state and the Ca^{2+} bound P-domain state. Again, no transfer was possible for Leu359 and Leu384 due to spectral overlap (Fig. S 4.1a). Additionally, transfer was not possible for one of the two Ile405 resonances, as this resonance split into two resonances with lower intensity in the Mg^{2+} bound state (Fig. S 4.1b). This observation remains

unexplained. The splitting was not observed for the Ca^{2+} bound state (Fig S 4.1c). However, for the Ca^{2+} bound state, one of the two Val374 resonances was broadened beyond detection (Fig. S 4.1d). Transfer from the metal ion-bound states to the apo state using the nearest neighbor approach led to an extension of the apo state assignment obtained before (c.f. chapter 5.3.1). This allowed the unambiguous assignment of the resonances of Ile337, Ala444, Ala309, and Val391 of the apo P-domain (see Fig. S 4.2). A total of 62 apo P-domain resonances were assigned.

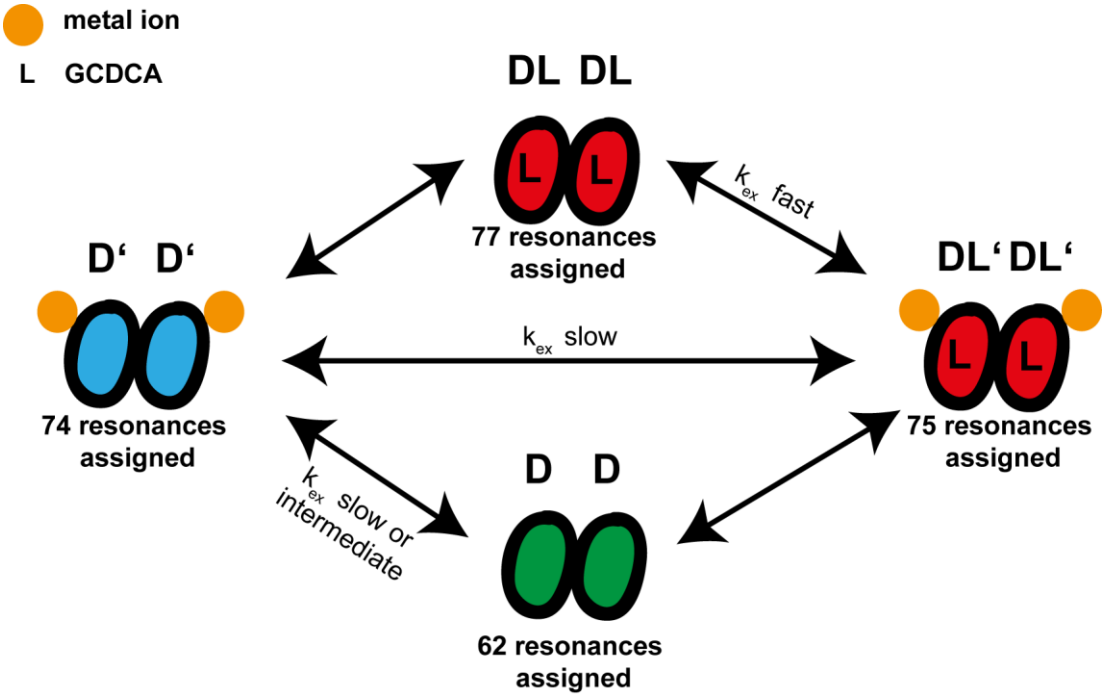


Fig. 5.4.1: Different P-domain states used for assignment transfer to the metal ion bound P-domain states. D corresponds to the apo dimer, D' to the metal ion bound dimer, DL to the GCDCA bound dimer and DL' to the metal ion and GCDCA bound dimer. Chemical exchange k_{ex} between the DL and DL' states is fast on the NMR time scale. Chemical exchange between the D' and DL' states is slow on the NMR time scale, but most resonances are sufficiently resolved and isolated, enabling transfer of assignments. Chemical exchange between the D state and the D' state is intermediate or slow. Note that there is no direct exchange between the D' and the DL states and the D and the DL' states. However, the nearest neighbor approach can be used to transfer assignments between these states.

5.4.2 Bivalent metal ions trigger P-domain dimerization

At the level of P-domain in solution, GCDCA leads to the dimerization of the protein (Creutznacher, 2020). The question therefore arises whether metal ions would also trigger P-domain dimerization. Monitoring monomer and dimer resonances in [^{1}H,^{13}C] HMQC spectra of different $MgCl_2$ concentrations clearly shows a gradual decrease of monomer resonance intensities (Fig. 5.4.2a). The same behavior is observed when $MgCl_2$ is replaced with $CaCl_2$ (Fig. S 4.3). These findings indicate P-domain dimerization in the presence of bivalent metal ions. A shift of the molecular weight in size exclusion chromatography supports this conclusion (Fig. 5.4.2b).

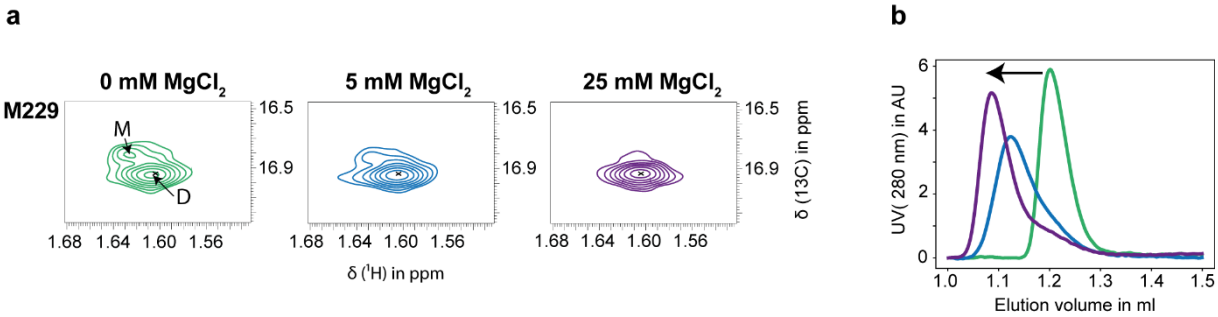


Fig. 5.4.2: $MgCl_2$ causes dimerization of the P-domain. Upon addition of $MgCl_2$ to the P-domain, monomer resonances (M) disappear, whereas dimer resonances (D) remain (a). Dimer formation was confirmed using size exclusion chromatography (SEC) (b). The P-domain was found to elute from the column at higher molecular weights upon addition of 5 mM and 25 mM $MgCl_2$ (same color code as in (a)). Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 75 µM MILVA labeled MNV P-domain.

5.4.3 CSP analysis of the interaction between P-domains and bivalent metal ions reveal long-range effects throughout the protein

Analysis of CSPs due to metal ion titration reveals a pattern similar to that by GCDCA. To determine CSPs, the 62 assigned apo resonances were compared with the corresponding Mg^{2+} or Ca^{2+} bound state resonances (Fig. 5.4.3a, Fig. S 4.4b, and Tabs. S 4.1 and S 4.2). The analysis was performed in the same way as described in chapter 5.3.3. In brief, CSPs were not determined if one of the two resonances was not sufficiently resolved. If a resonance in the bound form was well isolated and no corresponding apo resonance was found in that spectral region, the CSP was assumed to be larger than 20 Hz. The respective resonances are highlighted in Tabs. S 4.1 and S 4.2. As explained above, CSPs were classified according to the Euclidean distances $\Delta\nu_{Eucl}$ and plotted on a structural model. Comparison of how methyl groups are affected by the respective metal ions shows almost identical patterns for Mg^{2+} and Ca^{2+} (Fig. 5.4.3b and S 4.4a). As can be seen from the figures, CSPs are observed throughout the protein. Not only the metal ion binding areas are affected, but also regions remote from these binding sites. The details of the analysis will be discussed below for the Mg^{2+} bound state. CSPs are observed at the receptor-binding site. This correlates well with one of the metal ion binding sites being close to the receptor/P-domain interface (Fig. 5.4.3c and d). Interestingly, as in the case of GCDCA, CSPs were also found in the E'F' loop, which is the binding epitope for neutralizing antibodies and at lower parts of the P-domain (Fig. 5.4.3b, e and f). It should be noted that the distribution of affected methyl groups throughout the protein may reflect a "global transition". The term is in quotation marks because CSPs do not necessarily result from structural rearrangements but from changes in the chemical environment. Therefore, "global transition" should be understood as an abstract form of transition, but one that affects the entire protein. Such a transition would not necessarily involve large-scale structural rearrangements. The possibility of a "global transition" is discussed below in more detail.

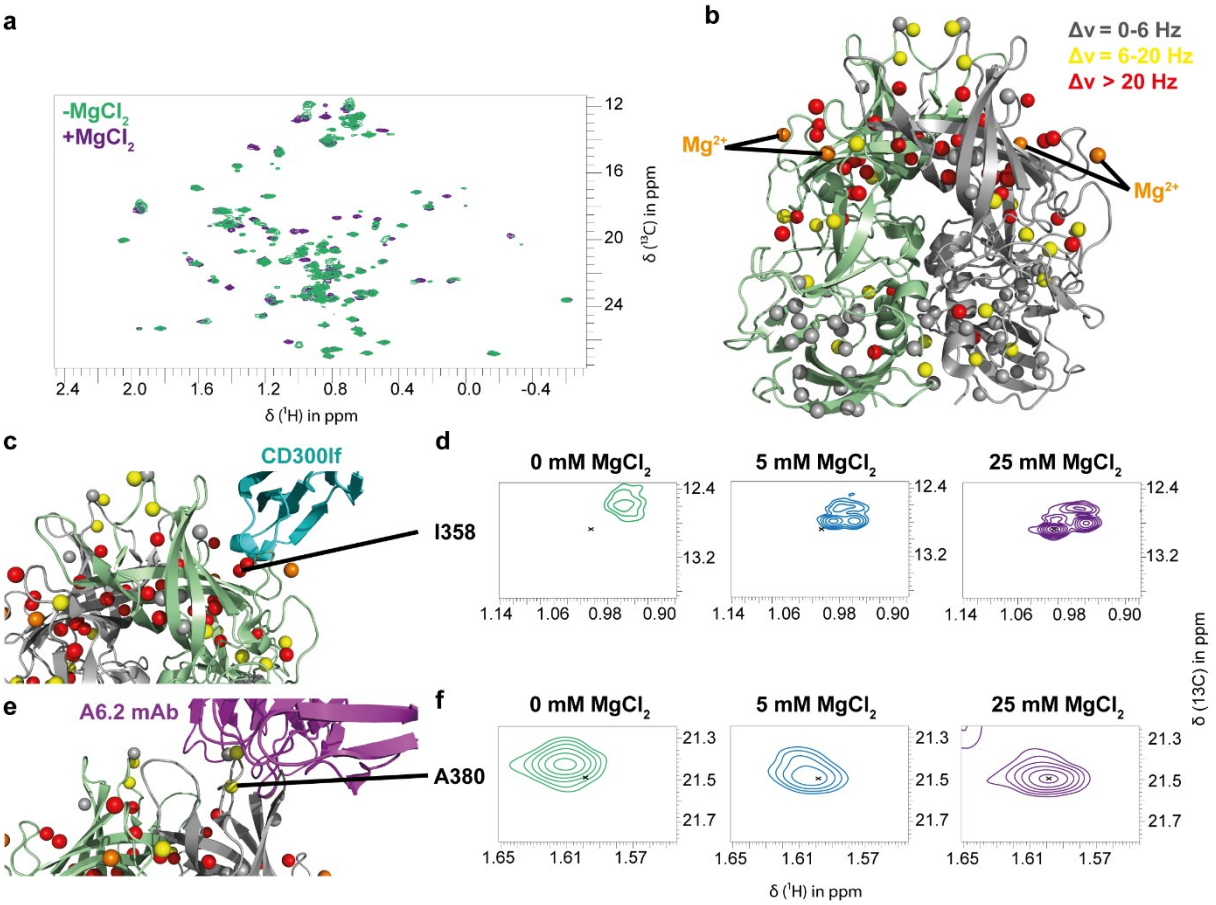


Fig. 5.4.3: Global effects of Mg^{2+} binding to MNV P-domain. (a) [^{1}H,^{13}C] HMQC spectra of P-domain in the presence and absence of $MgCl_2$. (b) CSPs of methyl group resonances caused by $MgCl_2$ plotted on structural model (pdb 6e47, Nelson et al., 2018). $MgCl_2$ affects almost all parts of the P-domain. (c) Methyl groups located at the receptor binding site are affected, e.g., Ile358. Note that the peak splitting of Ile358 is due to Pro361 cis/trans isomerism (c.f. Fig. 5.2.3). The CD300lf receptor is shown in cyan. CSPs of the corresponding resonance are shown in (d). (e) Additionally, methyl groups located at the antibody binding site are affected, e.g., Ala382. The Fab fragment of a monoclonal antibody (mAb A6.2; pdb 7l5j, Williams et al., 2021a) was modelled in the structure and is shown in magenta. The resonance corresponding to Ala380 is shown in (f). Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 75 µM MILVA labeled MNV CW1 P-domain. The structural model of the P-domain antibody complex shown in (e) was generated by aligning the P-domains in structural models of the P-domain in complex with metal ions, GCDCA, and CD300lf (pdb 6e47, Nelson et al., 2018) and the MNV capsid in complex with a monoclonal antibody (pdb 7l5j, Williams et al., 2021a) using *pymol* (DeLano, 2002). The resulting P-domain coordinates from pdb 6e47 and the antibody coordinates from pdb 7l5j were used for the illustration.

5.4.4 Metal ion binding to the G'H' loop causes dimerization and long-range effects

From CSP and SEC experiments, it is not clear which of the two metal ion binding sites is responsible for promoting dimerization and long-range CSPs. Therefore, D440A and D410A P-domain point mutants were synthesized to selectively switch off metal ion binding.

According to the crystal structure model by Nelson et al., 2018, the metal ion at the G'H' loop is coordinated by Gln438 and Asp440 (Fig. 5.4.6d). The G'H' loop is also part of the GCDCA binding site and attaches the other P-domain chain within the dimer. Therefore, it is tempting to assume that metal ion binding to this loop causes dimerization and the described long-range effects, e.g., in the E'F' loop. Substituting Asp440 with an alanine lead in poor P-domain biosynthesis yields and resulted in NMR

spectra that are challenging to interpret, as many peaks are broadened beyond detection (Fig. 5.4.4a). In addition, no monomer peaks were observed in the apo state. Spectra in the presence and absence of $MgCl_2$ showed no major changes. The only assigned resonance clearly showing CSPs is Val414 (Fig. 5.4.4b and c) near Asp410, which coordinates the remaining metal ion at the receptor epitope (Fig. 5.4.6c). The binding site at the G'H' loop is evidently switched off. Asp440 is highly conserved in P-domains of MNV strains (Fig. S 4.13) and might constitute a critical residue for protein stability.

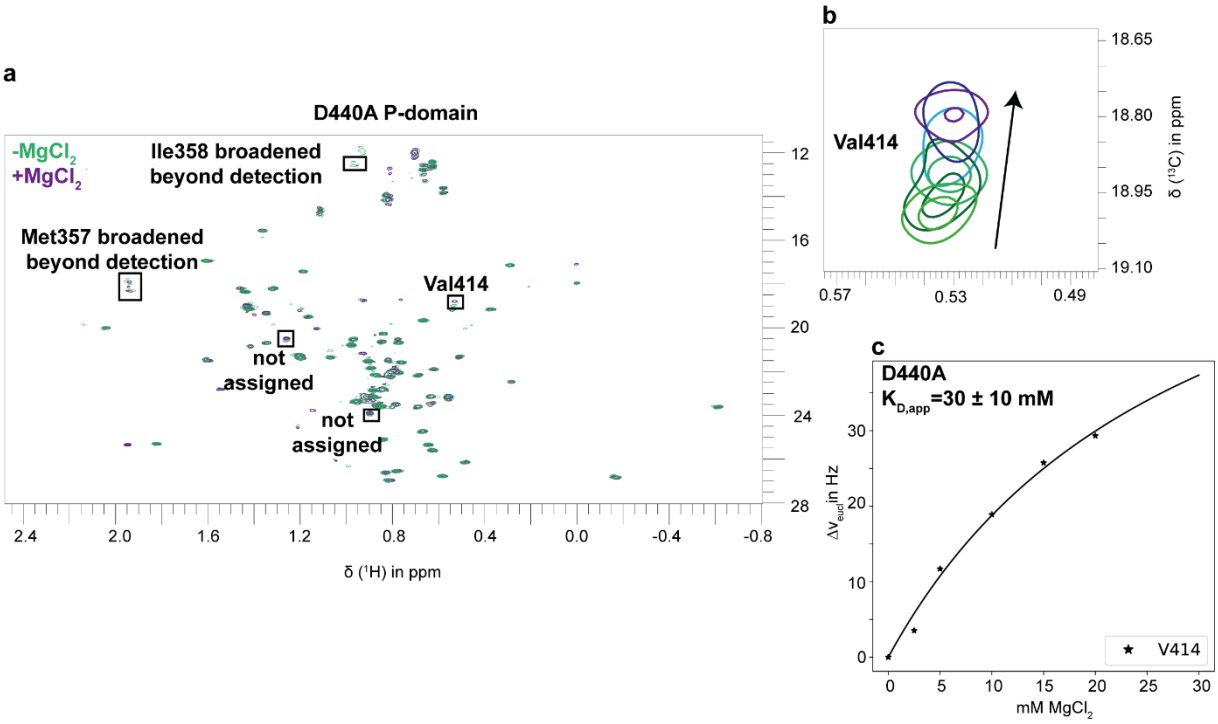


Fig. 5.4.4: Mg^{2+} binding to D440A P-domain is restricted to the binding site at the receptor epitope. (a) [^{1}H,^{13}C] HMQC spectra of D440A P-domain in the presence and absence of $MgCl_2$ indicate Mg^{2+} binding at the receptor binding epitope. The only assigned resonance showing CSPs is Val414 (b) located at this site. Many peaks are broadened beyond detection, as indicated for Ile358 and Met357. No major changes as observed for the WT (wildtype) protein can be concluded (c.f. Fig. 5.4.3a). (c) An apparent dissociation constant $K_{D,app}$ of 30 ± 10 mM was derived assuming a two-state binding model and using CSPs measured for Val414. Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 27 µM MI*LVA labeled MNV CW1 P-domain (for details on MI*LVA labelling, see chapter 4.2.4).

In contrast to the D440A P-domain, for D410A P-domain all resonances were well-resolved allowing a detailed analysis of NMR spectra in the presence and absence of metal ions (c.f. Fig. 5.4.6b). Titration of the D410A P-domain mutant in the presence of GCDCA with up to 25 mM $MgCl_2$ revealed no significant CSPs for amino acids close to corresponding metal ion binding site (Fig. 5.4.5a), showing that this site is inactive. This can be easily rationalized as the Asp410 carboxy group coordinating the metal ion (Fig. 5.4.6c) is removed from the P-domain molecule by the D410A mutation.

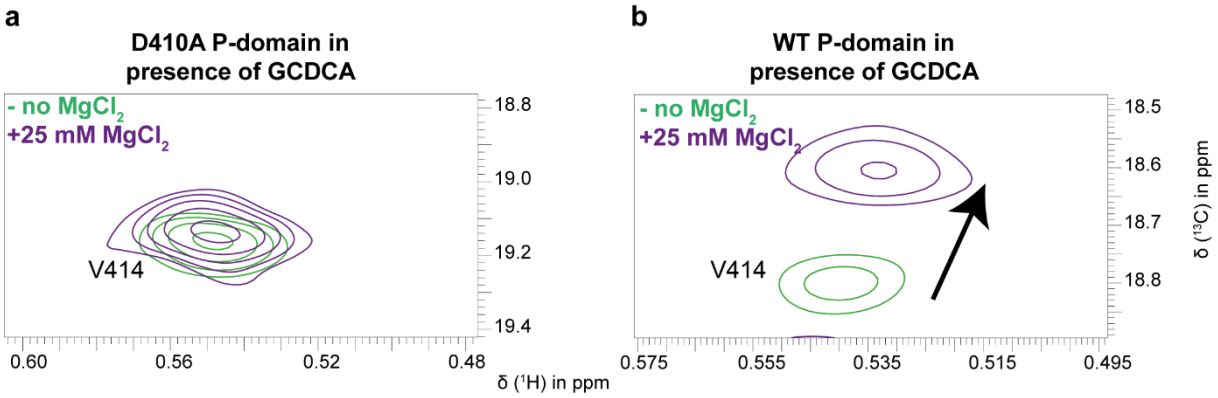


Fig. 5.4.5: The D410A mutant shows no binding of Mg^{2+} to the binding site at the receptor epitope. (a) shows the resonance corresponding to Val414 of the D410A P-domain in the presence and absence of $MgCl_2$. In comparison to the WT (wildtype) P-domain (b), the resonance of Val414 does not move significantly due to $MgCl_2$ addition indicating no significant binding to D410A P-domain at this binding site. Note that Val414 is close to Asp410, coordinating the metal ion (Fig. 5.4.6c). Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 36 µM MI*LVA labeled MNV CW1 D410A P-domain or 75 µM MILVA labeled MNV CW1 WT P-domain in presence of saturating amounts of GCDCA (for details on MI*LVA labelling, see chapter 4.2.4).

Consequently, CSPs observed for this mutant most likely originate exclusively from metal ion binding to the remaining binding site at the G'H' loop (Fig. 5.4.6d). CSPs were measured between the apo form and the Mg^{2+} bound form of D410A P-domain (Fig. 5.4.6a and b, Tab. S 4.3). The observed effects are very similar to those observed for the WT (wildtype) P-domain. Monomer resonances are depleted (Fig. 5.4.6e), and methyl groups show CSPs throughout the protein. This is particularly interesting for, e.g., Ile358 located at the receptor-binding site and Ala380 at the antibody binding site (Fig. 5.4.6a), as it underlines the long-range character of G'H' loop metal ion binding effects. Similar observations were obtained for Ca^{2+} binding to D410A P-domain (spectra can be found in the data repository of this thesis). The observation of almost no metal ion binding effects in spectra of D440A P-domain and very similar spectral effects for WT and D410A P-domain clearly leads to the conclusion that the G'H' loop metal ion binding site is responsible for dimerization and long-range effects. Reconsidering the idea of a "global transition", this transition would be triggered by metal ion binding to the G'H' loop. The binding would cause an enhancement of dimer interactions and long-range effects at the receptor-binding and antibody-binding site.

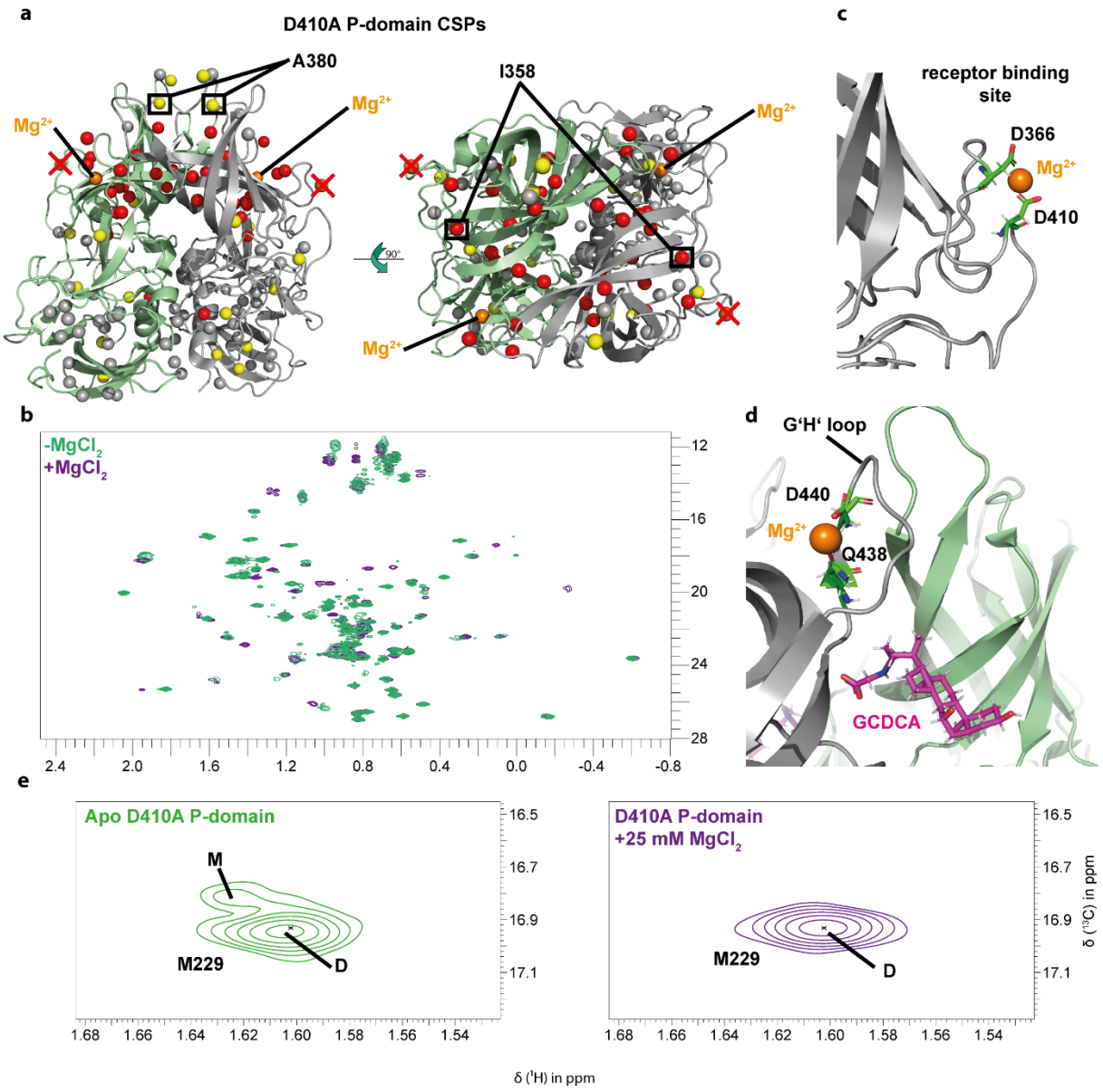


Fig. 5.4.6: Metal ion binding site at the G'H' loop causes long-range effects. D410A P-domain was generated to selectively switch off metal ion binding at the receptor epitope by replacing one of the two aspartates coordinating the metal ion at this site with an alanine (c, see also Fig. 5.4.5). [^{1}H,^{13}C] HMQC spectra of D410A P-domain in the presence and absence of $MgCl_2$ (b) and plotting CSPs on the structural model (a) (pdb 6e47; Nelson et al., 2018) show similar patterns as observed for the WT (wildtype) P-domain. This is especially of interest for methyl groups at the receptor binding site (e.g., Ile358) and the antibody binding site (e.g., Ala380). (c) shows the metal ion binding side that is switched off by the D410A mutation. (d) shows the remaining metal ion binding site of D410A P-domain. Interestingly, the loop (G'H' loop) coordinating the metal ion is also involved in GCDCA binding and attaches to the other P-domain chain in the dimer. Mg^{2+} binding to the G'H' loop causes the protein to dimerize (e). Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 62 µM MI*LVA labeled MNV CW1 P-domain (for details on MI*LVA labelling, see chapter 4.2.4)

5.4.5 Thermodynamic and kinetic cross-talk between metal ion binding and GCDCA binding to the P-domain

For the determination of dissociation constants characterizing metal ion binding to the G'H' loop, WT (wildtype) and D410A P-domain proteins were employed. Resonances of the metal ion bound P-domain and of the apo P-domain were in intermediate or slow exchange. Using intensities of suitable metal ion-bound P-domain resonances in slow exchange (Fig. 5.4.7a) at different $MgCl_2$ concentrations provided apparent dissociation constants $K_{D;app;Mg}$ of 11 ± 6 mM for the D410A P-domain (Fig. 5.4.7b) and of 5.7 ± 1.7 mM for the WT P-domain (Fig. 5.4.8, right panel). The values were determined using a two-step binding model. The same analysis was performed for $CaCl_2$ and yielded an apparent dissociation constant $K_{D,app;Ca}$ of 2.7 ± 2.4 mM and of 7.8 ± 3 mM for the D410A and the WT P-domain, respectively (Fig. 5.4.9). It should be noted that a two-step binding model is theoretically not suitable to describe metal ion binding to apo P-domain molecules. As explained in chapter 5.4.2, metal ion binding and dimerization are clearly linked, so neglecting the presence of monomers in the binding model is incorrect. Therefore, the derived dissociation constants are considered to be apparent. However, the protein concentrations are significantly higher than the dissociation constant describing P-domain dimerization (c.f. chapter 5.3.2). The presence of monomers can thus be ignored. Furthermore, the apparent dissociation constants for metal ions are much higher than the protein concentrations. Thus, the exact protein concentration has no significant influence on the derived dissociation constants, even if it slightly changes during the titration due to the formation of new dimers. Inclusion of the monomer species by applying a three-state binding model (c.f. chapter 5.3.3) to 2D line shape analysis with *TITAN* (Waudby et al., 2016) gave a dissociation constant $K_{D,Mg2+}$ of 3.8 ± 0.2 mM for D410A P-domain Mg^{2+} binding (Fig. S 4.5). A dissociation constant $K_{D,Ca2+}$ of 1.7 ± 0.2 mM was derived for the binding of D410A P-domain to Ca^{2+} (Fig. S 4.6). However, for both metal ions the three states (i.e., monomer, apo dimer, and metal ion bound dimer) could only be observed for a subset of resonances. For these resonances, the dimeric apo state and the dimeric bound state were not well resolved, so that the off-rates could only be determined with relatively high uncertainties. Comparing the dissociation constants derived with the different approaches and using WT and mutant P-domain molecules, the values ranged from 3.8 to 11 mM for Mg^{2+} and from 1.7 to 7.8 mM for Ca^{2+}.

As pointed out before, the G'H' loop is involved in both metal ion binding and GCDCA binding. To test for a possible cross-talk between metal ion binding and GCDCA binding, metal ion titration was repeated in the presence of saturating amounts of GCDCA. The observations are in strong contrast to the experiments without GCDCA. First, the exchange is fast on the NMR time scale (Fig. 5.4.7c). This is particularly interesting as some resonances, e.g., Met436, were in slow or slow-to-intermediate exchange without GCDCA and are now in fast or fast-to intermediate exchange (Fig. 5.4.7f and g). Second, no long-range CSPs were observed (Fig. 5.4.7d, c.f. Fig. 5.4.6a). Third, the resulting dissociation constant $K_{D;app;Mg}$ of WT and D410A P-domains is significantly reduced, reflecting much stronger metal ion P-domain interactions (Fig. 5.4.7e and Fig. 5.4.8; c.f. Fig. 5.4.7b). All dissociation constants of Mg^{2+} ions binding to the G'H' loop measured in the presence or absence of GCDCA are summarized in Tab. 5.2.

Tab. 5.2: Dissociation constants of Mg^{2+} binding to the G'H' loop of the P-domain.

P-domain Protein	GCDCA present?	K_D / mM	Method / Data shown in
WT	+	0.76 ± 0.09	CSPs / Fig. 5.4.8
	-	5.7 ± 1.7	Intensities / Fig. 5.4.8
D410A	+	0.62 ± 0.04	CSPs / Fig. 5.4.7
	-	11 ± 6	Intensities / Fig. 5.4.7
	-	3.8 ± 0.2	Line Shapes / Fig. S 4.5

Of special interest is the combination of these three findings, which provides evidence for the aforementioned possibility of a "global transition" of the P-domain. Apparently, this global transition can be triggered either by GCDCA (c.f. chapter 5.3.3) or by bivalent metal ions and is reflected by methyl group CSPs throughout the protein and by dimer formation. The transition is slow and must be the rate-limiting step for metal ion binding. This becomes clear by considering that the exchange is rather slow in absence of GCDCA (k_{ex} << 88 Hz, Fig. 5.4.7g) and rather fast in the presence of GCDCA (k_{ex} >> 96 Hz, Fig. 5.4.7f). The shift from slow to fast would on first glance seem to contradict the much higher affinity in the presence of GCDCA, but it makes sense because GCDCA already triggers the transition. This way, the metal ion can directly bind to the protein and the exchange is therefore fast. Considering that in the absence of GCDCA $k_{exchange}$ is << 88 Hz (Fig. 5.4.7g), and $k_{exchange}$ is given with $k_{exchange} = k_{off} + k_{on} * [L]$ (c.f. chapter 3.3.1), it follows that $k_{off;Mg}$ must be smaller than 88 s^{-1}. Using a $K_{D;app;Mg}$ of 11 mM and the relationship $K_D = k_{off}/k_{on}$ limits the on-rate $k_{on,Mg}$ to $8*10^3$ $M^{-1}s^{-1}$. This is much slower than a diffusion-controlled on-rate (Meyer and Peters, 2003) and can be explained by a rate-limiting global transition associated with metal ion binding. In the presence of GCDCA, the global transition has already occurred. Therefore, the on-rate is faster and CSPs are only visible for methyl groups near the metal ion binding site at the G'H' loop. Considering the relationship $K_D = k_{off}/k_{on}$, a faster on-rate would lead to a lower K_D and thus the affinity is higher, which is exactly what is observed: Mg^{2+} dissociation constants $K_{D;app;Mg}$ in the presence of GCDCA are 0.62 ± 0.04 mM (Fig. 5.4.7e) and 0.76 ± 0.09 mM (Fig. 5.4.8, left panel) for the D410A P-domain and the WT P-domain, respectively. The values are regarded to be apparent because the commercially available GCDCA is contaminated with small amounts of Ca^{2+} ions (Creutznacher et al., 2021). Therefore, it can be assumed that the "real" dissociation constants are even somewhat lower than the values reported here (see Note S 4.1 for the underlaying calculations).

A dissociation constant $K_{D;Ca}$ of 0.138 mM for Ca^{2+} in the presence of GCDCA has already been derived by Creutznacher et al., 2021. By comparing this constant to the values presented here in the absence of GCDCA ($K_{D,app;Ca}$ between 1.7 and 7.8 mM), it can be concluded that GCDCA enhances Mg^{2+} and Ca^{2+} binding in a similar way. All dissociation constants of Ca^{2+} ions binding to the G'H' loop measured in the presence or absence of GCDCA are summarized in Tab. 5.3.

Tab. 5.3: Dissociation constants of Ca^{2+} binding to the G'H' loop of the P-domain.

P-domain Protein	GCDCA present?	K_D / mM	Method / Data shown in
WT	+	0.138*	CSPs / (Creutznacher et al., 2021)
	-	7.8 ± 3	Intensities / Fig. 5.4.9
D410A	-	2.7 ± 2.4	Intensities / Fig. 5.4.9
	-	1.7 ± 0.2	Line Shapes / Fig. S 4.6

*This value is taken from Creutznacher et al., 2021.

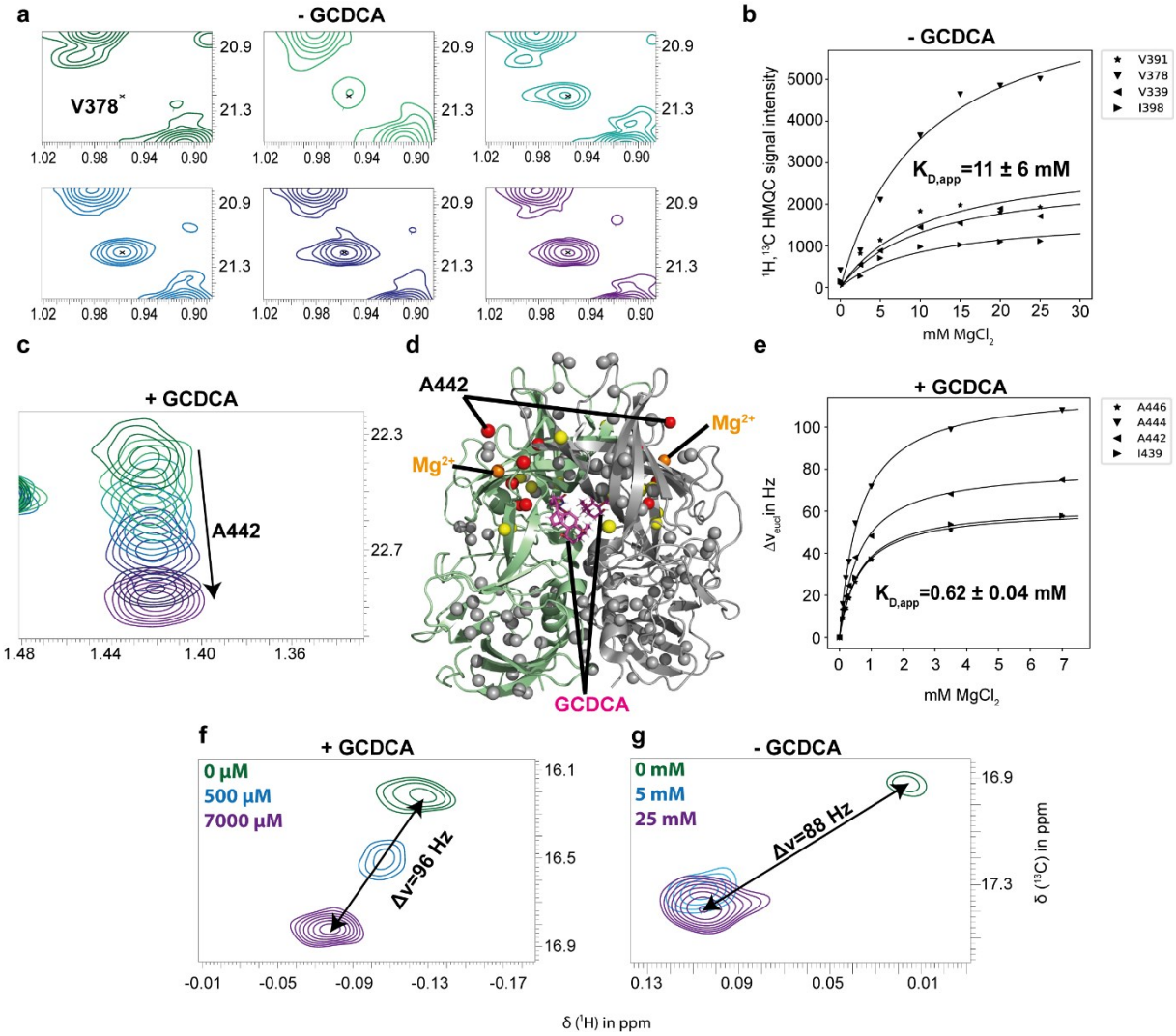


Fig. 5.4.7: GCDCA controls kinetics and thermodynamics of Mg^{2+} binding to the G'H' loop of D410A P-domain. The D410A P-domain lacks the receptor epitope metal ion binding site. Thus, all effects are caused by metal ion binding to the G'H' loop (c.f. Fig. 5.4.5). Titrating $MgCl_2$ to the apo protein causes dimerization and global structural rearrangements of the D410A P-domain (c.f. Fig. 5.4.6). Signals are in intermediate to slow exchange. Only intensities of signals in slow exchange (exemplarily shown in a) were used for determining the dissociation constant of Mg^{2+} binding, yielding an apparent dissociation constant $K_{D;app;Mg}$ of about 11 mM by fitting the law of mass action to resonance intensities (b). Analysis of CSPs due to $MgCl_2$ titration to the GCDCA bound D410A P-domain was straight forward as all signals are in fast exchange (c). Plotting CSPs on the structural model (pdb 6e47; Nelson et al., 2018) shows rather local than global effects of Mg^{2+} binding at the G'H' loop. This is in strong contrast to the changes observed in absence of GCDCA (c.f. Fig. 5.4.6). Using fast exchange CSPs for fitting of the law of mass action yielded an apparent dissociation constant $K_{D,app}$ of 0.62 mM in the presence of GCDCA (e). It is obvious by the difference of the dissociation constants determined in the presence (e) and absence (b) of GCDCA that GCDCA enhances the Mg^{2+} affinity. (f) and (g) show the resonance of Met436 upon $MgCl_2$ titration in the presence and absence of GCDCA, respectively. Whereas in the presence of GCDCA the exchange is fast, in absence of GCDCA the exchange is slow to intermediate. Strikingly, Δν is still in the same order of magnitude. The exchange rate k_{ex} is expected to be >> 96 Hz in the presence of GCDCA and <<88 Hz in the absence of GCDCA. Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 62 µM MI*LVA labeled D410A MNV CW1 P-domain in the absence of GCDCA and 36 µM MI*LVA labeled D410A MNV CW1 P-domain in presence of 250 µM GCDCA (for details on MI*LVA labelling, see chapter 4.2.4). The sample of the spectrum with 3.5 mM $MgCl_2$ used in (c) and (e) contained 18 instead of 36 µM MI*LVA labeled D410A MNV CW1 P-domain. In (c), the spectrum with 1 mM $MgCl_2$ is not shown to avoid crowding.

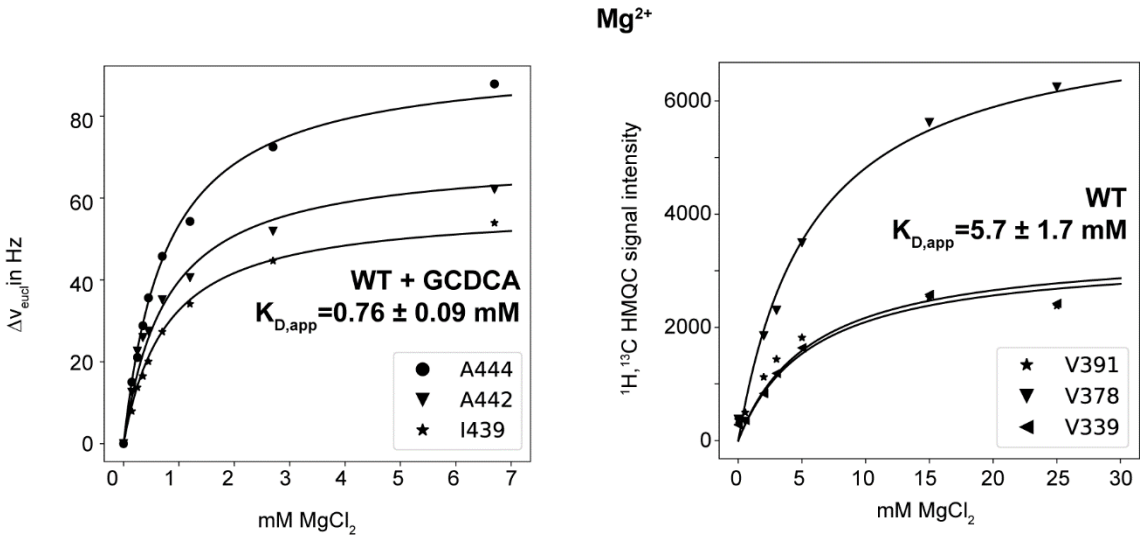


Fig. 5.4.8: Dissociation constants of Mg^{2+} binding to the G'H' loop of the WT (wildtype) P-domain protein in the presence and absence of GCDCA. The corresponding [^{1}H,^{13}C] HMQC spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 75 µM MILVA labeled MNV CW1 P-domain in absence of GCDCA and 25 µM MILVA labeled MNV CW1 P-domain in presence of GCDCA.

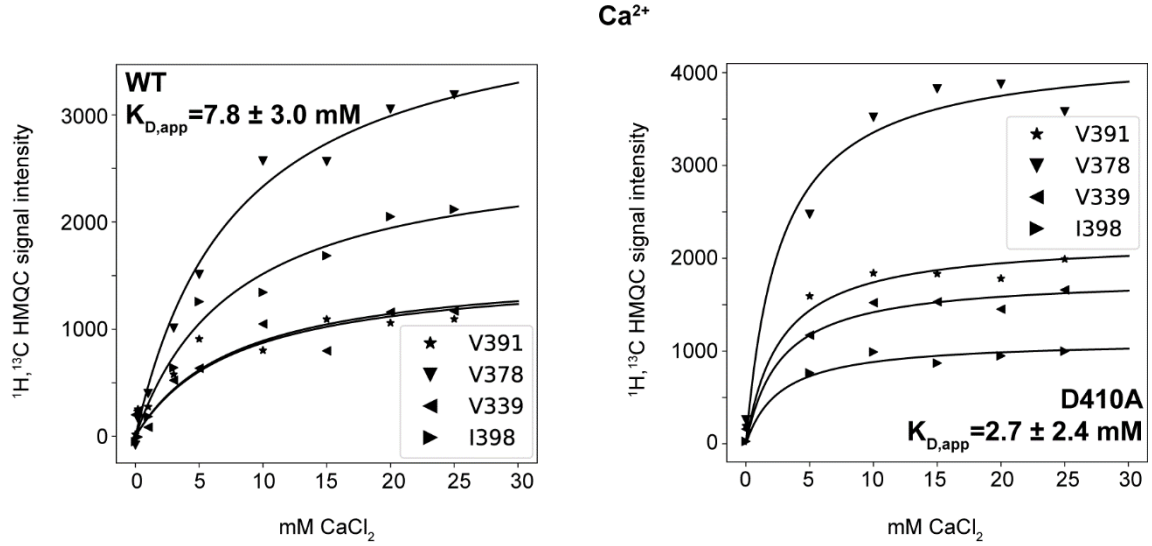


Fig. 5.4.9: Dissociation constants of Ca^{2+} binding to the G'H' loop of WT (wildtype) and D410A P-domain proteins in the absence of GCDCA. The corresponding [^{1}H,^{13}C] HMQC spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 39 µM MILVA labeled MNV CW1 P-domain or 18 µM MI*LVA labeled D410A MNV CW1 P-domain (for details on MI*LVA labelling, see chapter 4.2.4).

Following basic laws of thermodynamics and considering that GCDCA affects the G'H' loop metal ion affinity, one would conclude that metal ions must also influence the GCDCA affinity. To test this, GCDCA was titrated to the WT P-domain in the presence of saturating amounts of $MgCl_2$ or $CaCl_2$. Measurement of CSPs induced by GCDCA and mapping them on a structural model shows a similar pattern to that observed when GCDCA was titrated to the apo protein (Fig. 5.4.10a and b, Tab. S 4.4; c.f. Fig. 5.3.4). However, methyl group resonances of the E'F' loop responsible for antibody binding are not affected. The same picture emerges from titrating GCDCA to the Ca^{2+} bound P-domain (Fig. S 4.9, Tab. S 4.5). This is consistent with the idea that metal ions alone may trigger the above described "global transition". This transition causes the E'F' loop to already exist in the "final" state. Other parts

of the P-domain may require the binding of GCDCA to fully adopt to the final state. Possible reasons of GCDCA induced CSPs of resonances of the metal ion bound P-domain will be further discussed in chapters 5.4.7 and 5.4.8.

GCDCA binding was further investigated by employing 2D line shape analysis with *TITAN* (Waudby et al. 2016). Spectral regions of 9 suitable resonances were quantum-mechanically simulated and fitted to the acquired spectra using a two-state binding model (Fig. 5.4.10c and d and Figs. S 4.7 and S 4.8). The approach resulted in a dissociation constant $K_{D,GCDCA}$ of 4.4 ± 0.6 µM and an off-rate $k_{off,GCDCA}$ of 17.2 ± 1.4 s^{-1} for GCDCA binding in the presence of $MgCl_2$ and a dissociation constant $K_{D,GCDCA}$ of 0.15 ± 0.05 µM and an off-rate $k_{off,GCDCA}$ of 3.9 ± 0.9 s^{-1} for GCDCA binding in the presence of $CaCl_2$. The values differ from the dissociation constant $K_{D,GCDCA}$ of 10.5 ± 0.5 µM and the off-rate $k_{off,GCDCA}$ of 25.8 ± 3.5 s^{-1} determined in the absence of metal ions (c.f. chapter 5.3.3). All dissociation constants and off-rates of GCDCA in the presence and absence of metal ions are summarized in Tab. 5.4.

Tab. 5.4: Dissociation constants and off rates of GCDCA binding to the P-domain in the presence and absence of metal ions.

P-domain Protein	Metal ion present	K_D / µM	k_{off} / s^{-1}	Method / Data shown in
WT	-	10.5 ± 0.5	25.8 ± 3.5	Line Shapes / Fig. 5.3.5
	Mg^{2+}	4.4 ± 0.6	17.2 ± 1.4	Line Shapes / Fig. 5.4.10
	Ca^{2+}	0.15 ± 0.05	3.9 ± 0.9	Line Shapes / Fig. 5.4.10

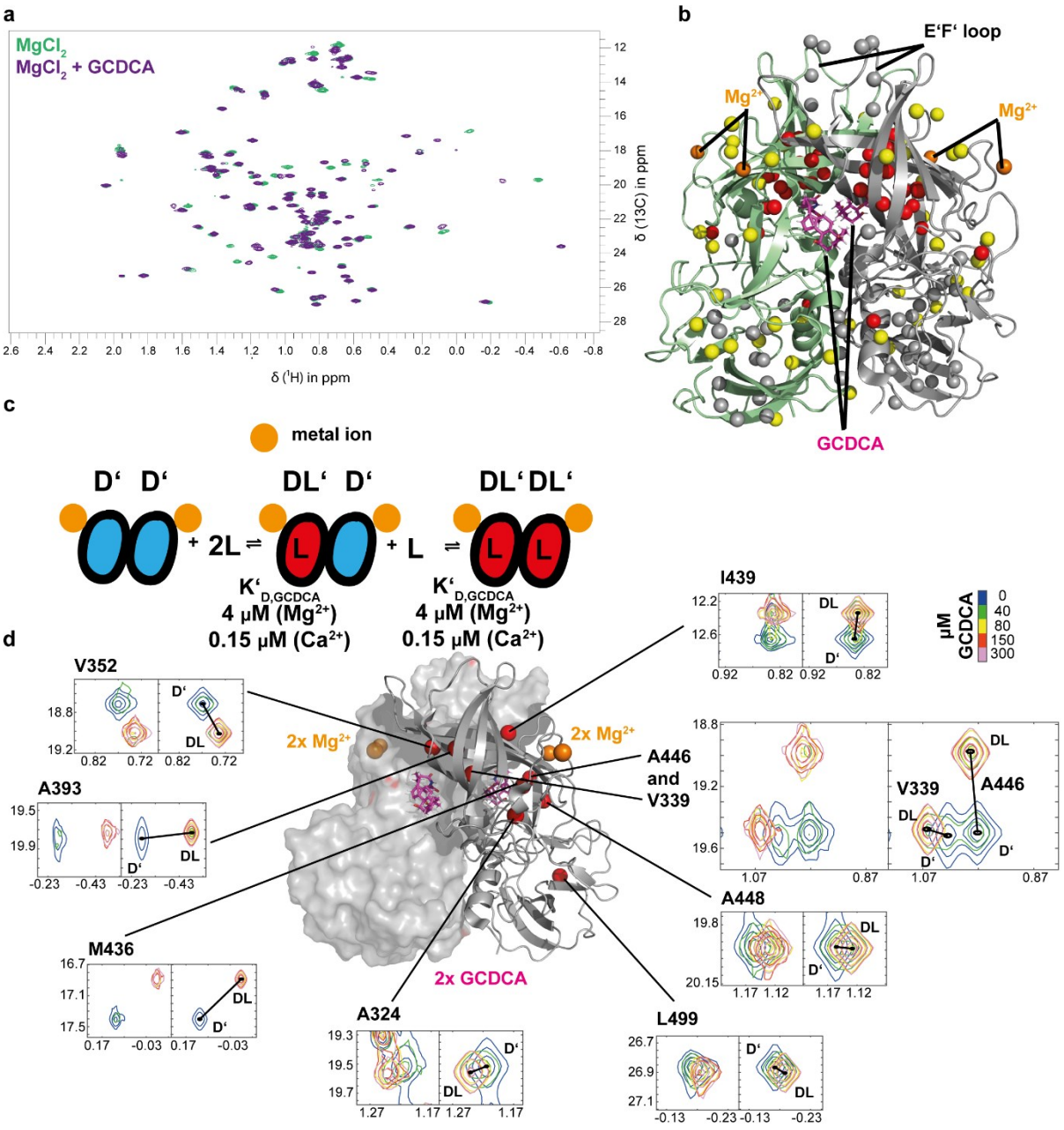


Fig. 5.4.10: Effects of metal ions on P-domain GCDCA interactions. (a) shows [^{1}H,^{13}C] HMQC spectra of Mg^{2+} bound P-domain in the presence and absence of GCDCA. The CSPs can be plotted on a structural model (6e47; Nelson et al., 2018) and show that almost the entire P-domain, but not the E'F' loop, which is responsible for antibody binding, is affected (b). As Mg^{2+} already causes P-domain dimerization, a simple two step binding model can be used to perform 2D line shape analysis with *TITAN* (Waudby et al., 2016), where the metal ion bound P-domain (D') binds GCDCA (DL') (c and d). A $K_{D,GCDCA}$ of 4.4 ± 0.6 µM and a $k_{off,GCDCA}$ of 17.2 ± 1.4 s^{-1} were derived for GCDCA binding to the P-domain in the presence of Mg^{2+}. This K_D value is smaller than the one observed in absence of $MgCl_2$ (c.f. chapter 5.3.3). The experiments and the analysis were also carried out substituting $MgCl_2$ with $CaCl_2$. A $K_{D,GCDCA}$ of 0.15 ± 0.05 and a $k_{off,GCDCA}$ of 3.9 ± 0.9 s^{-1} were determined, differing in an order of magnitude from the values obtained without metal ions (see also Figs. S 4.7 and S 4.8). The CSP pattern in the presence of Ca^{2+} was found to be similar to the CSP pattern in the presence of Mg^{2+} (Fig. S 4.9). Spectra shown here were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 75 µM MILVA labeled MNV CW1 P-domain and 25 mM $MgCl_2$. The left panels in (d) show the experimental spectra. The right panels show the spectra simulated and fitted with *TITAN*.

5.4.6 Dissociation constants of metal ion binding to the P-domain receptor binding site

Metal ion binding to the G'H' loop is switched off for the D440A P-domain (Fig. 5.4.4). The only assigned resonance with $MgCl_2$ induced CSPs belongs to Val414, which is located at the receptor-binding site. Fitting CSPs at different $MgCl_2$ concentrations using a two-state binding model results in a $K_{D,app;Mg}$ of 29 ± 10 mM for the receptor epitope metal ion binding site (Fig. 5.4.4c). The value is regarded as apparent as only one assigned methyl group showed CSPs and the highest $MgCl_2$ concentration is below the resulting dissociation constant. The addition of GCDCA significantly improves the quality of the spectra and results in well-resolved resonances (Fig. 5.4.11a). Fitting CSPs of suitable resonances of methyl groups located at the receptor-binding site to a two-state binding model yielded a $K_{D;app;Mg}$ of 20 ± 10 mM (Fig. 5.4.11c). Repeating the experiments and fits using the same resonances with the WT protein resulted in a $K_{D,app;Mg}$ of 20 ± 5 mM. The derived values indicate that there is no cross-talk between GCDCA binding and the binding of metal ions to the receptor-binding site. Substitution of $MgCl_2$ with $CaCl_2$ resulted in a $K_{D,app;Ca}$ of 9.5 ± 2.4 mM for the WT protein (Fig. 5.4.11d). The dissociation constants of metal ions binding to the P-domain receptor binding site are summarized in Tab. 5.5.

Tab. 5.5: Dissociation constants of metal ions binding to the P-domain receptor binding site.

P-domain Protein	GCDCA present?	Metal Ion	K_D / mM	Method / Data shown in
WT	+	Mg^{2+}	20 ± 5	CSPs / Fig. 5.4.11
	+	Ca^{2+}	9.5 ± 2.4	CSPs / Fig. 5.4.11
D440A	+	Mg^{2+}	20 ± 10	CSPs / Fig. 5.4.11
	-	Mg^{2+}	29 ± 10	CSPs / Fig. 5.4.4

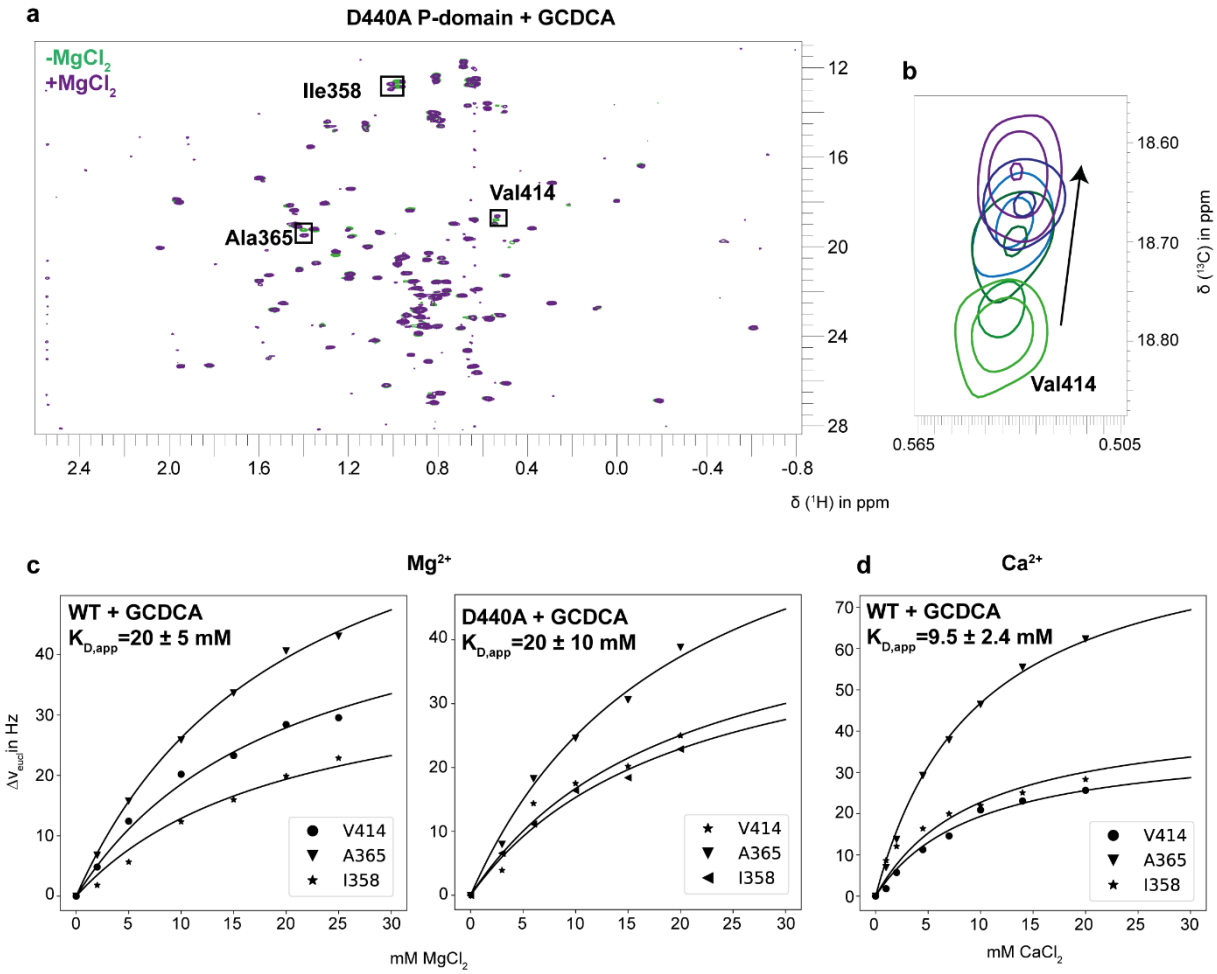


Fig. 5.4.11: Bivalent metal ion binding to the WT (wildtype) and D440A P-domain receptor binding site. (a) Adding GCDCA to the D440A P-domain results in well resolved spectra. This contrasts with the spectra of the apo form (c.f. Fig. 5.4.4). Using chemical shift perturbations of suitable resonances (b) corresponding to methyl groups around the receptor binding site yield dissociation constants for Mg^{2+} around 20 mM in the presence of GCDCA (c) and a dissociation constant of ca. 10 mM for Ca^{2+}. Spectra shown here were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 13 µM MI*LVA labeled D440A P-domain, 54 µM MILVA labeled MNV CW1 P-domain for titration with $MgCl_2$ and 39 µM MILVA labeled MNV CW1 P-domain for titration with $CaCl_2$ (for details on MI*LVA labelling, see chapter 4.2.4).

5.4.7 PCSs suggest no structural differences between P-domains with metal ions bound and P-domains with GCDCA and metal ions bound

The previous chapters have illustrated the long-range character of CSPs between the P-domain, to which only metal ions are bound, and the P-domain, to which metal ions and GCDCA are bound (c.f. Fig. 5.4.10). The following experiments and calculations aim at narrowing down on the origin of these global long-range CSPs.

Direct interpretation of CSPs is complex, and they might “only” reflect side-chain rearrangements not visible in the available structural models. Further NMR experiments were performed to gain more insights into differences between metal ion-bound and GCDCA and metal ion-bound P-domain states. First, the possibility that CSPs reflect more extensive structural rearrangements was tested by

measuring pseudo contact shifts (PCSs). Substitution of $MgCl_2$ with diamagnetic $LaCl_3$ resulted in similar P-domain spectra, indicating that lanthanides affect the P-domain similarly to bivalent metals (Fig. S 4.10a). In the next step, the paramagnetic metal ions Sa^{3+} and Ce^{3+} were employed to induce PCSs in methyl group resonances (Fig. S 4.10b). PCSs are directly dependent on the orientation and distance of the individual methyl groups to the paramagnetic metal ion. Large-scale structural rearrangements would alter the two factors, and PCSs would change for methyl groups in the affected protein regions. Sa^{3+} induced PCSs served to define PCSs vectors, as already explained in chapter 5.2.3. Ce^{3+} PCSs are generally more pronounced and were used for the structural analysis. PCSs in the presence and in the absence of GCDCA show a strong correlation, indicating no major structural rearrangements (Fig. 5.4.12a). Outliers exclusively correspond to methyl groups near the metal ion binding site at the G'H' loop. Due to the close distance to the metal ion, these methyl groups are also susceptible to PCSs through minor structural rearrangements. Small rearrangements at this site are not surprising, as the GCDCA could easily reorientate individual amino acids of the G'H' loop (Fig. 5.4.12b). Given the strong correlation of PCSs, CSPs measured between metal ion-bound and metal ion and GCDCA bound P-domain states most likely reflect small-scale structural rearrangements such as side chain reorientations. This characterization of the two states is consistent with structural models for metal ion and CD300lf bound P-domains in the presence and absence of GCDCA (Nelson et al., 2018). Calculating RMSD values of the two models results in an RMSD of 0.248 nm^2 (pdb 6e47 and 6c6q, Nelson et al., 2018). The PCSs in the presence and absence of GCDCA thus indicate almost identical structures with and without GCDCA, also in the absence of the CD300lf receptor.

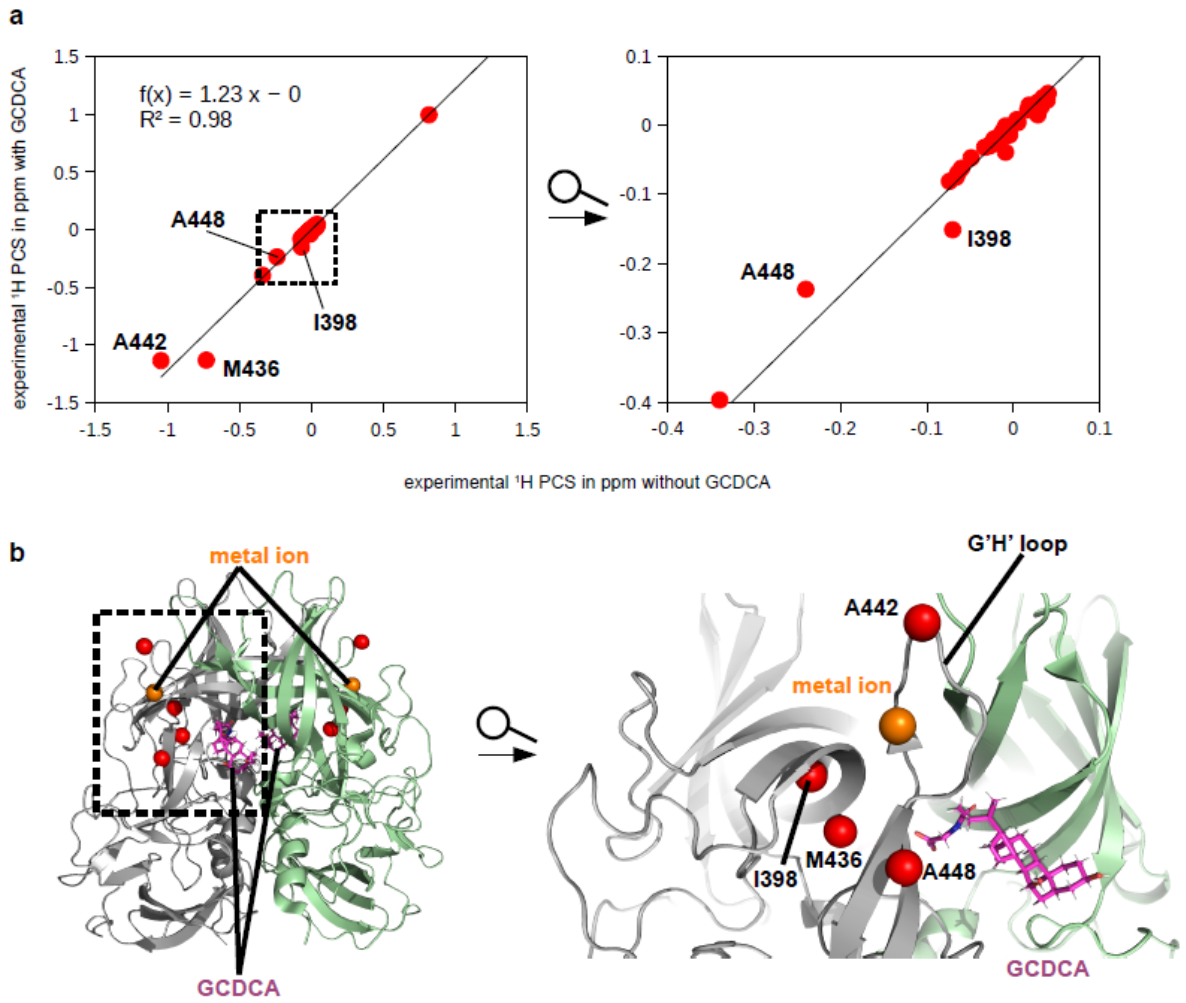


Fig. 5.4.12: Ce^{3+} induced pseudo contact shifts (PCSs) in the absence and in the presence of GCDCA indicate no major conformational changes upon the addition of GCDCA. (a) shows a correlation plot of PCSs in the presence and absence of GCDCA. PCSs are in general larger in the presence of GCDCA as can be seen by the slope of the line. This can be explained by stronger alignment in the presence of GCDCA as GCDCA enhances P-domain metal ion interactions. The correlation is good (R^2= 0.98), and outliers are located exclusively close to the metal ion binding site at the G'H' loop (b). As PCSs strongly depend on the distance to the metal ion, this region is susceptible for larger PCS differences. Slight rearrangements at the G'H' loop might be caused by GCDCA binding to it. No large-scale or global structural rearrangements by GCDCA binding to the metal ion-bound P-domain can be concluded. The structures with and without GCDCA must be highly similar. PCSs in the absence of GCDCA were obtained from [^{1}H,^{13}C] HMQC spectra shown in Fig. S 4.10. PCSs in the presence of GCDCA are shown in Fig. 5.2.4. The structural model of Nelson et al., 2018 (pdb 6e47) was used to for the illustration.

PCSs were than used to compare the solution structure of the P-domain in the presence of metal ions with three available structural models from X-ray crystallography and cryo-electron microscopy (Fig. 5.4.13 and Tab. 5.6). Using the respective models, Ce^{3+} induced PCSs, and the approach explained in chapter 5.2.3, Δχ and paramagnetic center parameters were fitted with *Paramagpy* (Orton et al., 2020). The derived parameters were then used to calculate theoretical PCSs based on the corresponding structural models. Agreement with the solution structure of the P-domain in the presence of metal ions was quantified using the Q-factor. The Q-factor depends on differences between experimental PCSs and theoretical PCSs. A Q-factor of 0 means perfect agreement and a Q-factor of 1 means no

agreement. Consequently, well-defined parameters and low Q-factors can only be obtained if the structural model matches the structure in solution.

The first model is a crystal structure model of the apo P-domain (Taube et al., 2010). The analysis yielded a reasonable Q-factor of 0.17 (Fig. 5.4.13a). The second model was obtained from a cryo-electron microscopy model of the MNV capsid in the presence of a monoclonal antibody (Williams et al., 2021a). The P-domain part of this model is almost identical to the P-domain of the first model. As expected, the resulting Q-factor is close to that derived for the first model (Fig. 5.4.13b). For both models, several tensor parameters were rather poorly defined (Tab. 5.6). The third model is the crystal structure model of P-domain in the presence of metal ions, GCDCA, and the CD300lf receptor (see above, pdb 6e47, Nelson et al., 2018). The analysis resulted in a better Q-factor of 0.11 and well-defined tensor parameters (Fig. 5.4.13c, Tab. 5.6).

Tab. 5.6: Parameters from Δχ fitting for Ce^{3+}.

		Apo P-domain	P-domain with mAb	P-domain with CD300lf, Mg^{2+} and GCDCA*
pdb code		3lq6 (Taube et al., 2010)	7l5j (Williams et al. 2021a)	6e47 (Nelson et al. 2018)
ΔX_{ax} in 10^{-32} m^3		-2.81±0.36	-2.60±0.45	-1.325±0.018
ΔX_{rh} in 10^{-32} m^3		-1.01±0.20	-0.57±0.28	-0.679±0.025
Coordinates of origin in Å	X	3.94±2.2	94.9±1.0	-16.3±0.5
	Y	-6.45±0.43	90.8±0.9	2.9±0.4
	Z	-5.27±1.3	99.4±1.5	-45.3±0.7
Orientation of principal axis of tensor in °	α	20.3±5.7	40±15	109.88±0.19
	β	94.5±3.0	24.1±8.3	74.1±1.3
	γ	163.3±9.3	24±27	170.4±7.6
Q-factor		0.17	0.18	0.11

*identical to structure with P-domain with CD300lf and Mg^{2+} (RMSD = 0.248 nm^2)

As explained above, crystal structure models of metal ion and CD300lf bound P-domains are almost identical with and without GCDCA. A similarly low Q-factor was observed when using the same structural model and PCSs measured in the presence of GCDCA (c.f. chapter 5.2.3). The relatively low Q-factors and the strong correlation of PCSs with and without GCDCA lead to the conclusion that metal ion bound P-domain structures in solution are highly similar in the presence and absence of GCDCA. Moreover, they must be very similar to the crystal structure models of the P-domain in complex with metal ions and CD300lf.

One of the major differences between the first two models and the third model is that the cleft between the A'B' loop and E'F' loop is closed (Fig. 5.4.13d and e). The closed conformation is thought to be unrecognized by monoclonal antibodies (Sherman et al., 2019; Williams et al., 2021a). One methyl group found in this loop is the Ala382 methyl group. Due to the conformational change the Ala382 methyl group relocates by more than 15 Å. Interestingly, the calculated and experimental PCSs for this methyl group agree only for the third model (Fig. 5.4.13a, b, and c). This supports the presence of the closed P-domain conformation in solution in the presence of bivalent metal ions. The preference of the closed conformation in the presence of bivalent metal ions is consistent with results from ELISA assays showing no antibody binding in the presence of bivalent metal ions (Williams et al., 2021b).

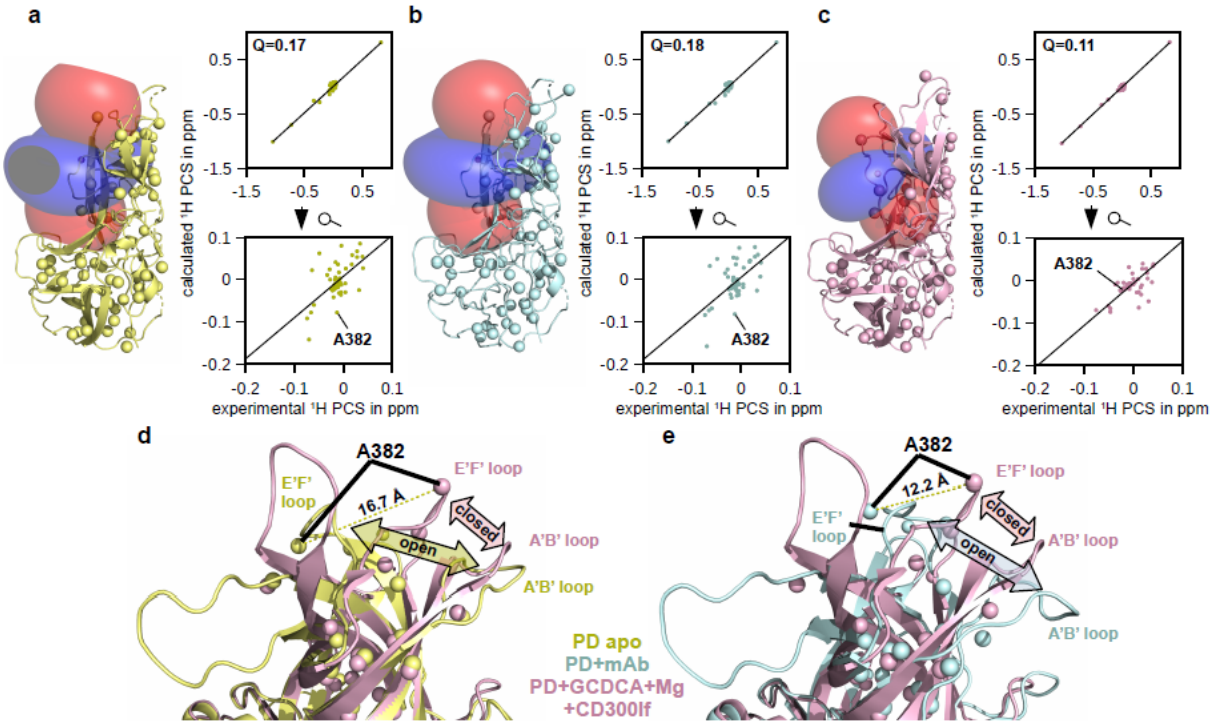


Fig. 5.4.13: P-domains in the presence of metal ions exists in a conformation that does not allow binding of neutralizing antibodies. PCSs were measured in the presence of $CeCl_3$ (Fig. S 4.10b) and Δχs and the position of paramagnetic centers were determined using structural models of the apo P-domain (a, pdb 3lq6, yellow), P-domain in the presence of a monoclonal antibody (mAb) (b, pdb 7l5j, cyan) and in the presence of GCDCA, metal ions and the CD300lf receptor (c, pdb 6e47, purple). Note that structural models of P-domain in complex with CD300lf and metal ions in the presence and absence of GCDCA are almost identical (pdb 6c6q, RSMD=0.248 nm^2). The best Q-factor was observed for the P-domain in the presence of GCDCA, metal ions, and CD300lf. For, e.g., Ala382, theoretical and measured PCSs do not match using structural models of the apo P-domain or of the P-domain in the presence of the mAb (a and b). Whereas the E'F' loop, where A382 is located, is in the open conformation for these two structural models, it is in a closed conformation in the presence of GCDCA, metal ions, and CD300lf (d, e). Note that the color coding in (d) and (e) is the same as in (a), (b), and (c). For Δχ tensor and paramagnetic centre fitting, the structural model of P-domain in the presence of GCDCA, metal ions and the CD300lf receptor (c, pdb 6e47) was processed as explained in chapter 5.2.3. Identical atoms were included for the calculations with structural models of apo P-domain (a) and P-domain in complex with the mAb (b). See Tab. 5.6 for tensor and paramagnetic center parameters.

5.4.8 Methyl group order parameters of metal ion bound P-domain reveal differences in pico- to nanosecond dynamics by GCDCA binding

Differences in chemical shifts of metal ion bound P-domain in the presence and absence of GCDCA are not caused by large-scale structural rearrangements. It is known that methyl group CSPs also report for small-scale side-chain rearrangements (Mulder, 2009; Butterfoss et al., 2010; Hansen et al., 2010; Hansen and Kay, 2011; Siemons et al., 2019), and such changes likely cause the CSPs observed here (c.f. chapter 5.4.7). Even if these changes have no significant impact on the structure, they might be associated with a change in the protein side-chain dynamics. This would alter the "entropic profile" of the P-domain (Capdevila et al., 2017). Methyl group order parameters S^2 report on dynamics in the pico- to nanosecond window. Side-chain rotations lie within this window (Boswell and Latham, 2019). Order parameters were determined as described by Tugarinov and Kay, 2013, with adjustments in the error estimation. In brief, Gauss distributions of all parameters required for the calculation of order parameters were estimated directly from multiple intensity measurements (I_{HMQC} and I_{HSQC}) or indirectly by fitting parameters to experimental intensity values according to Eq. 18, 20, 21 and 22, and using the fitted value and its standard error to estimate the mean and the standard derivation of the corresponding Gauss distribution (for parameters r_{HHext}, R^S_{MQ}, and τ_C). r_{HDext} values, which depend inversely on distances between a methyl group's protons and deuterium nuclei of the protein were calculated according to Eq. 23 based on a crystal structure model (pdb 6e47; Nelson et al., 2018). More details on the calculations can be found in chapter 4.4.6. Order parameters of each methyl group were than fitted according to Eq. 11-19 using 1000 bootstrap iterations by randomly drawing values from the Gaussian distributions of the above parameters and the r_{HDext} value derived from the structural model.

For each methyl group, order parameter differences between the metal ion-bound and metal ion and GCDCA bound state were calculated. The standard deviations of the order parameters of both states were summed up to give the final errors (Fig. 5.4.14c). The differences were then categorized into five groups. The first three groups are resolved with a significance level of 0.158. It should be noted that the size of the significance level was chosen for simplicity. It reflects the idea of non-overlapping error bars of the compared order parameters, which are given with one standard deviation, respectively. Due to the relatively large significance level, conclusions for single methyl groups should not be over-interpreted if the error bar of the difference is close to the threshold of the respective category. However, this does not affect the overall picture when comparing the order parameters of many methyl groups of the two P-domain states. The first group comprises differences in order parameters larger than 0.2 ($\Delta S>0.2$; pale red bars and spheres in Fig. 5.4.14a and c). In the second and third groups, order parameters were increased or decreased ($\Delta S>0$, pale yellow; $\Delta S<0$, pale blue). For the fourths group, there was no change in order parameters with a significance level of 0.044 ($0.02>\Delta S>-0.02$; grey spheres and asterisk). No conclusions were drawn for the fifths group because the errors were too large compared to the differences (not shown in Fig. 5.4.14a, white bars in c).

Significant changes in the dynamics of methyl groups were found throughout the protein. It is striking that most methyl groups show increased flexibility upon addition of GCDCA. Assuming the obtained CSPs report for side-chain rearrangements (c.f. chapter 5.4.7), a simultaneous change in side-chain dynamics of the same or neighboring methyl groups seems obvious (Fig. 5.4.14). According to the literature, differences in methyl group order parameters are directly linked to differences in conformational entropy (Capdevila et al., 2017). Following this line of argument, GCDCA binding to the metal ion bound P-domain would not cause significant structural rearrangements, but rather entropy

redistribution within the protein. The entropic profiles of the available reporter methyl groups thus indicate a state of higher conformational entropy in the presence of GCDCA. The ability of GCDCA to cause this redistribution even at sites remote from the binding pocket calls for further investigations into the presence of an allosteric network-like linkage of methyl groups within the P-domain.

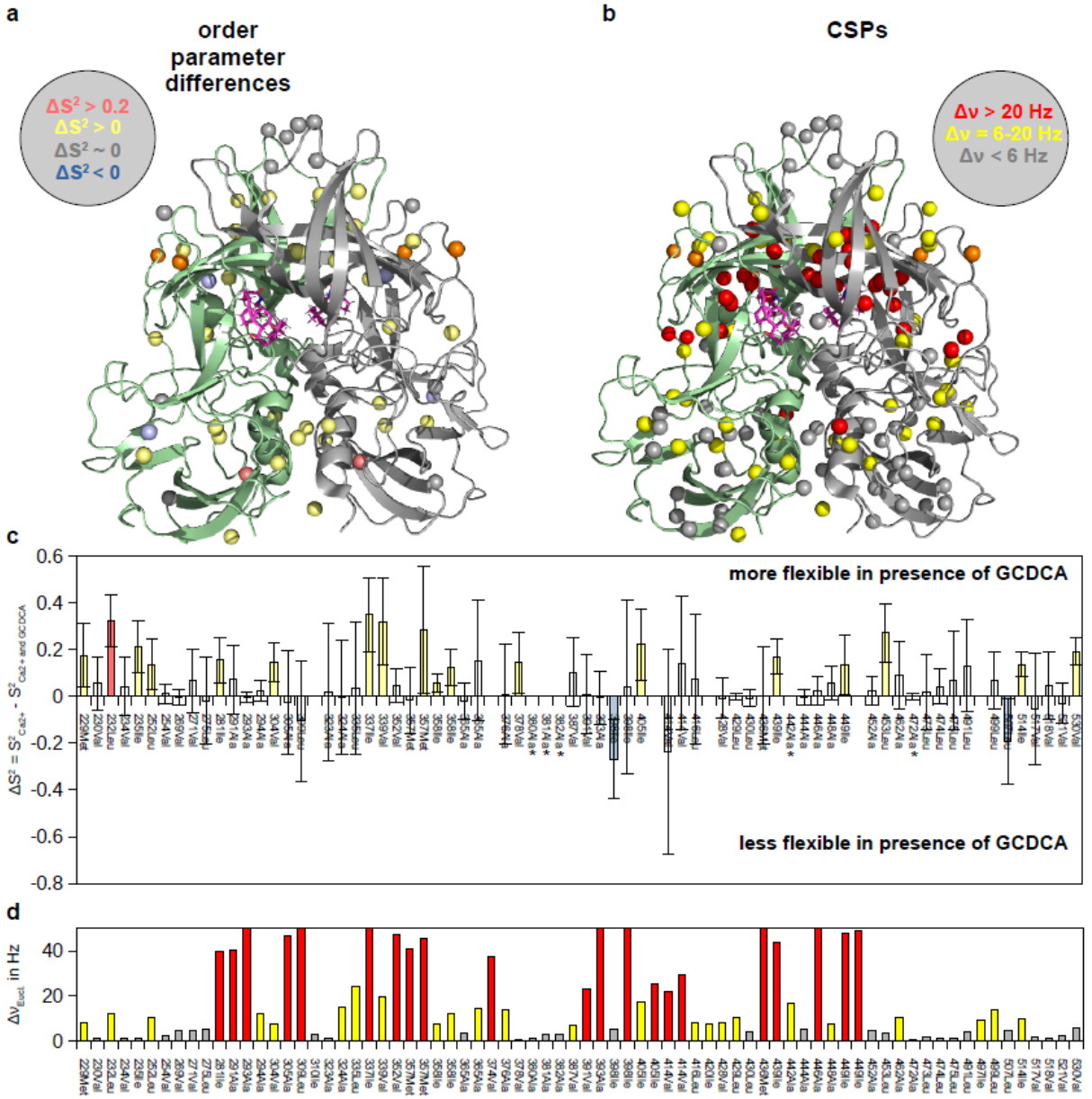


Fig. 5.4.14: Methyl group order parameters reflect GCDCA altering the dynamic profile of Ca^{2+} bound P-domain. (a) shows methyl group S^2 order parameter differences of the Ca^{2+} bound P-domain with and without GCDCA. (b) Methyl group CSPs of the Ca^{2+} bound P-domain caused by GCDCA. (c) Bar plot of order parameter differences in the presence and absence of GCDCA. Notably, most methyl groups increase in flexibility due to GCDCA binding. The color code is as shown in (a). (d) shows CSPs between the two states. The CSP color coding is shown in (b) and matches the color code of CSP plots throughout this thesis. CSPs were determined from HMQC spectra acquired for S^2 determination (c.f. Tab. 4.12 and chapter 4.3.3). The molecular correlation time τ_C required for S^2 calculations of the two states is 22 ns (Fig. S 4.12). S^2 order parameters of the individual states are shown in Fig. S 4.11. The corresponding spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe. Information about sample conditions is given in Tab. 4.12. A structural model of Nelson et al., 2018, (pdb 6e47) was used for the illustration.

5.4.9 Methyl group CSPs suggest allosteric network-like communication throughout the P-domain

Allosteric networks can transmit ligand-induced effects throughout a protein and induce CSPs or altered protein dynamics at remote sites (Capdevila et al., 2017). As explained in detail earlier, long-range CSPs were observed for GCDCA, Mg^{2+}, Ca^{2+}, and La^{3+}. It should be noted that unspecific bulk-effects of metal ions can be excluded. This can be concluded from experiments with the D440A P-domain, where no globally distributed CSPs were observed in the presence of metal ions (c.f. Fig. 5.4.4). Differences in order parameters between the Ca^{2+}-bound and the Ca^{2+} and GCDCA-bound P-domain show a shift towards a state of higher conformational entropy in the presence of GCDCA. Remarkably, many methyl groups distant from the GCDCA binding site were affected, which could be explained by an allosteric network-like connection between these remote sites and the GCDCA binding pocket. In this chapter, further clues to the existence of an allosteric network-like connection of methyl groups within the P-domain will be presented.

A typical feature of allosteric networks are linear movements of chemical shifts caused by similar ligands (e.g., Boswell et al., 2018). P-domain proteins in the presence of different metals and GCDCA clearly show such characteristic signs (Fig. 5.4.15). On the basis of ten selected methyl group resonances shown in Fig. 5.4.15, the presence of such a network is very well possible. The selected methyl groups are located at the G'H' loop (Ala442 and Ala444), at the C'D' loop adjacent to the G'H' loop (Val352), at the receptor-binding site (Ala365), at the dimer interface (Ile514, Ile281), or at other sites of the P-domain (Val339, Leu499, and Leu232). The local distribution of affected methyl groups emphasizes that the underlying network extends across the entire P-domain.

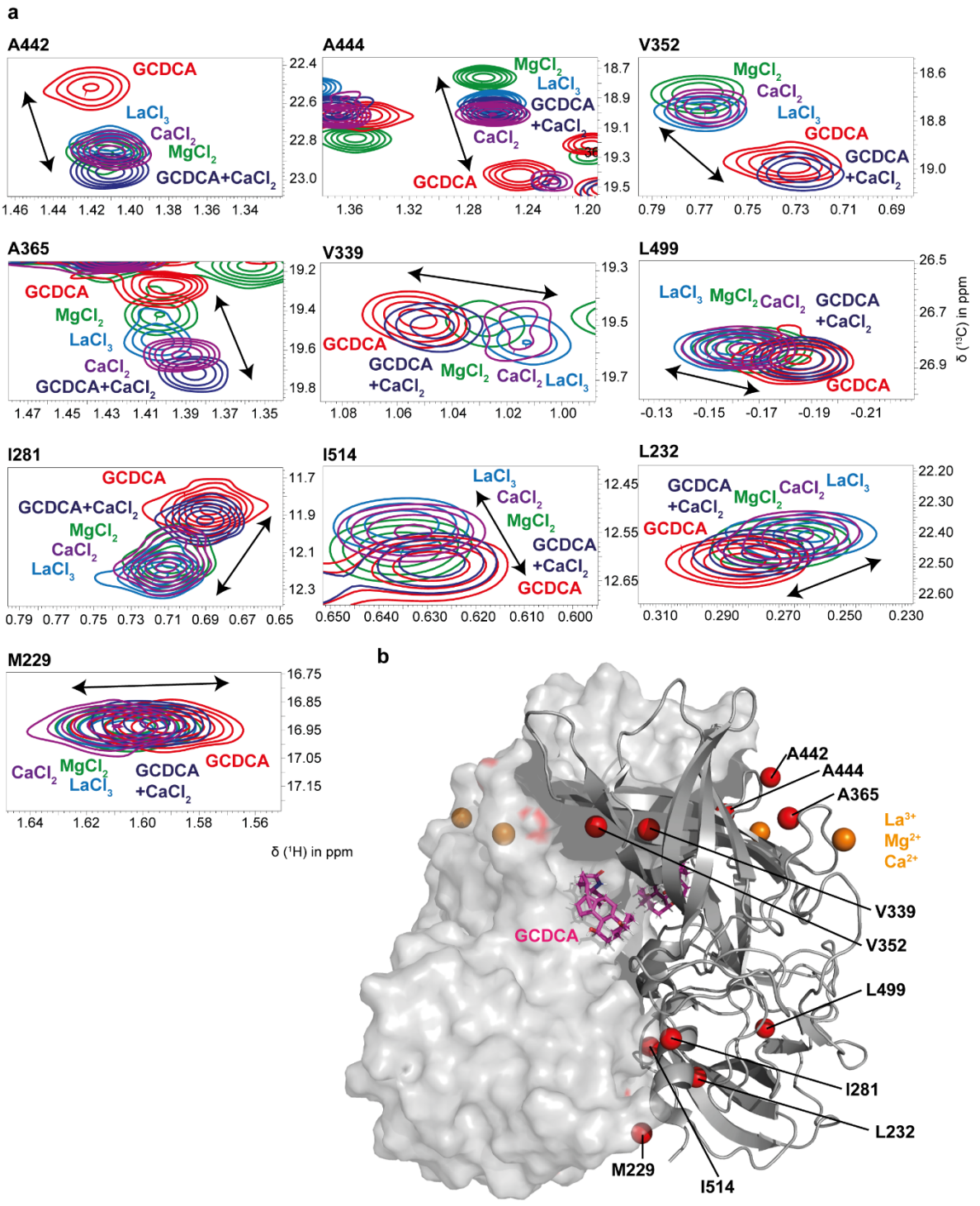


Fig. 5.4.15: Allosteric network-like connection of methyl groups reflected by linear CSPs of methyl groups throughout the P-domain. (a) shows resonances of methyl groups distributed all over the protein (b, pdb 6e47, Nelson et al., 2018). Chemical shifts move linearly by addition of different ligands triggering the "global protein transition" (c.f. chapter 5.4.5), indicating the presence of an allosteric network-like connection of methyl groups. For details on the acquisition of spectra see Fig. S 4.10, Fig. S 4.9, Fig. 5.4.3, and Fig. 5.3.4.

5.5 pH-dependant control of P-domain dimer formation and of binding of metal ions to the G'H' loop

The previous chapters showed how GCDCA and metal ion binding affect P-domain dimerization in solution. The formation of dimers at the level of the P-domain in solution parallels the contraction of the MNV capsid in cryo-electron microscopy models (Sherman et al., 2019; Song et al., 2020; Snowden et al., 2020). A recent cryo-electron microscopy study (Williams et al., 2021b) shows that acidic pH values also trigger the MNV capsid to transition to the contracted form. Thus, the question arises whether dimer formation can be induced by acidifying the pH, too. The dimerization of P-domains was studied at different pH values using 2D line shape analysis of P-domain resonances, resonance intensities, and size exclusion chromatography (SEC). Moreover, line shape analysis of metal ion binding and GCDCA binding to the P-domain at different pH values was performed to scrutinize a possible effect of the pH on ligand binding. Using chemical shift perturbations (CSPs) induced by altering the pH value allowed to derive apparent ionization constants for the G'H' loop of the P-domain. The observation of pH-dependent P-domain dimerization also enabled further analysis of dynamics, showing that GCDCA redistributes the entropic profile of the apo P-domain.

5.5.1 2D line shape analysis and SEC reveal pH-dependent P-domain dimerization

To support the hypothesis that P-domain dimerization depends on the pH, NMR spectra of the P-domain at different pH values were acquired. NMR experiments with large proteins such as the P-domain require experiments to be performed in D_2O buffers instead of H_2O buffers (Schütz and Sprangers, 2020). To make the experiments presented in this chapter comparable to cryo-electron microscopy studies performed in H_2O buffers, pH_{corr} values are given in this chapter (Krezel and Bal, 2004; see chapter 4.3.1). Tracking resonance intensity ratios of dimer and monomer resonances (I_D/I_M) at different pH_{corr} values shows a clear pH_{corr} dependence (Fig. 5.5.1a). When the pH_{corr} is shifted from 3.0 to 4.0, the intensity ratio increases. This indicates the formation of dimers. At a pH_{corr} value of 5.3, the dimers start to dissociate again, as reflected by a decreased intensity ratio. Interestingly, a pH_{corr} of 5.3 was used for the NMR experiments with the P-domain presented in the previous chapters. This value apparently represents neither optimal conditions for the formation, nor for the dissociation of dimers. Observation of intensity ratios at more neutral pH_{corr} values suggests further dissociation into monomers. The intensity ratio is lowest at a pH_{corr} of 7.4. Since NMR peak intensities might also depend on relaxation properties or slightly altered P-domain concentrations, size exclusion chromatography (SEC) was employed to verify the pH dependant dimerization (Fig. 5.5.1b). The elution volume of the P-domain is shifted by adjusting the pH from 6.2 to 4.0, as it is expected because of the higher molecular weight of the dimer.

As mentioned above, P-domain NMR experiments presented in previous chapters were performed at a pH_{corr} of 5.3. Analysing the intensity ratios in Fig. 5.5.1a, dissociation into monomers and formation of dimers is balanced at a pH_{corr} of 5.3. The lowest intensity ratio was obtained at a pH_{corr} of 7.4, indicating a significantly higher population of the monomers. At a pH_{corr} of 5.3 a dissociation constant $K_{D,Dimerization}$ of 7 µM and an off-rate $k_{off,Dimerization}$ of 1 s^{-1} were determined using 2D line shape analysis with *TITAN* (c.f. chapter 5.3.2). Repeating the analysis at a pH_{corr} of 7.4 yielded a $K_{D,Dimerization}$ of 45 ± 1.6 µM and an off-rate $k_{off,Dimerization}$ of 4.3 ± 0.3 s^{-1} (Fig. 5.5.1c and d). Comparison of the values obtained at the two conditions clearly shows a strong dependence on the pH_{corr}. However, a pH_{corr} of 5.3 does not constitute optimal conditions for dimer formation. Repeating the line shape analysis approach at pH_{corr} values of 4.4 or 4.6 would probably lead to even lower dissociation constants. Since this would

require very low P-domain concentrations and therefore long measurement times to detect the monomeric species, NMR spectroscopy reaches its limits here.

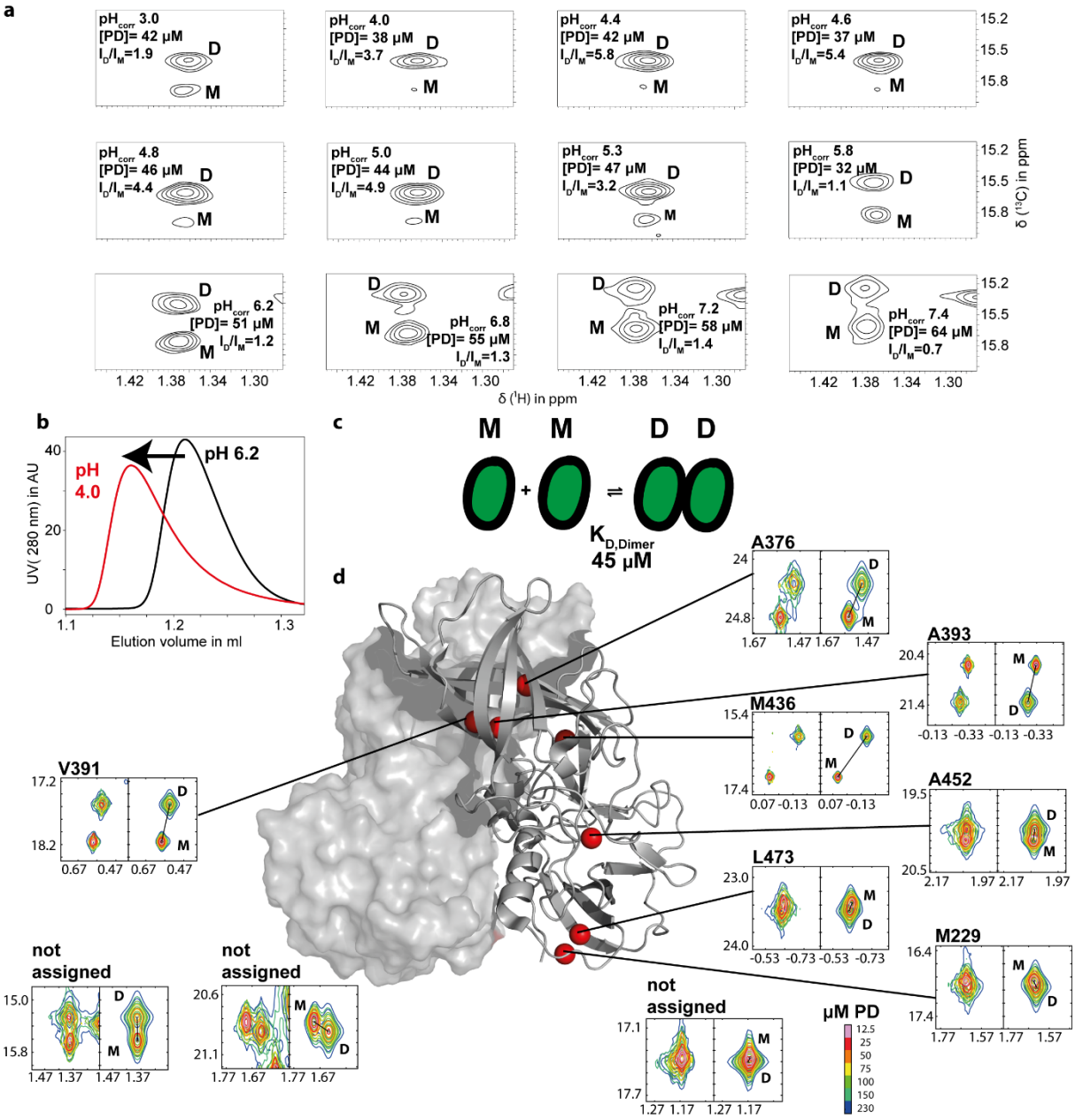


Fig. 5.5.1: Shifting from acidic to neutral pH triggers P-domain dimer dissociation. Relative intensities in [^{1}H,^{13}C] HMQC spectra of monomer and dimer resonances are affected by the pH. This reflects dimer dissociation at very acidic (pH_{corr} 3) or at rather neutral pH_{corr} values (pH_{corr} 5.8 -7.4) (a). [PD] is the P-domain concentration and I_D/I_M is the intensity ratio of monomer (M) and dimer (D) resonances. pH dependent dimerization is supported by size exclusion chromatograms of P-domain carried out at different pH values (b). (c,d) 2D line shape analysis of P-domain [^{1}H,^{13}C] HMQC spectra at different protein concentrations and pH_{corr} 7.4 yielded a dissociation constant K_D of 45 ± 1.6 µM and a k_{off} of 4.7 ± 0.3 s^{-1}. The dissociation constant is almost one order of magnitude higher than the one obtained in acidic conditions (pH_{corr} 5.3, K_D=7 µM, k_{off}=1.7 s^{-1}). The left panels show the measured spectra, and the right panels show the simulated spectra. Spectra were recorded in 20 mM citric acid-d_4 (Cambridge Isotope Laboratories), 100 mM NaCl for pH_{corr} values ranging from 3 to 5.8 and in 20 mM Bis-Tris-d_{19} (Cambridge Isotope Laboratories), 100 mM NaCl for pH_{corr} values ranging from 6.2 to 7.4. The spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe. A structural model of the P-domain by Nelson et al., 2018, (pdb 6e47) was used for the illustration in (d).

5.5.2 pH influences on thermodynamics and kinetics of P-domain metal ion and GCDCA binding

Dimerization, metal ion binding to the G'H' loop, and GCDCA binding are closely linked (c.f. chapters 5.3.3 and 5.4.4). The influence of pH on the formation of P-domain dimers leads to the question of whether the pH also affects the binding of GCDCA and metal ions. In the following, binding of GCDCA and metal ions to the P-domain will be scrutinized within the physiologically relevant pH ranges. Whereas bile acids occur, e.g., in the intestine at pH 4.4 - 5.2 (McConnell et al., 2008; Di Ciaula et al., 2017) bivalent metal ions are also present in the cytoplasm of cells (Romani and Scarpa, 1992) and in blood (Wang et al., 2002), where the pH is around 7.4 (Wolfe, 1961; Yoshida et al., 1993), and in the stomach (Moe, 2016), where the pH of mice is between 3 and 4 (McConnell et al., 2008).

As the metal ion binding to the G'H' loop (c.f. Fig. 5.4.6d) is coordinated by the side chain carboxy group of Asp440, it is reasonable to assume that protons can compete with metal ions binding to this loop. The comparison of [^{1}H,^{13}C] HMQC spectra of the P-domain at a pH_{corr} of 4 (i.e. at a high proton concentration) in the presence and absence of $CaCl_2$ shows almost no CSPs. Only small CSPs for Ala365 located at the receptor epitope metal ion binding site were observed (Fig. 5.5.2a). Therefore, significant metal ion binding to the G'H' loop can be excluded at a pH_{corr} of 4. On the contrary, almost every resonance experienced Ca^{2+} induced CSPs at a pH_{corr} of 7.4 (Fig. 5.5.2b). The effects are very similar to those observed for GCDCA binding or metal ion binding at a pH_{corr} of 5.3 (c.f. chapters 5.3.3 and 5.4.3). Since the monomeric species is significantly populated at a pH_{corr} of 7.4 (c.f. chapter 5.5.1), monomer resonances are visible for most methyl groups. In many cases, it is possible to observe the three states of the monomeric P-domain, the dimeric P-domain, and the dimeric metal ion-bound P-domain. This enables the quantification of thermodynamics and kinetics using 2D line shape analysis. Utilizing a three-state model (Fig. 5.5.2c), spectral regions of nine suitable resonances were quantum-mechanical simulated and fitted to the measured spectra. The analysis yielded a dissociation constant $K_{D,Ca2+}$ of 1.09 ± 0.02 mM and a $k_{off,Ca2+}$ of 27.5 ± 0.4 s^{-1} (Fig. 5.5.2d).

In summary, pH clearly influences metal ion binding to the G'H' loop. No binding of metal ions was detected at a pH_{corr} of 4 (see above). At a pH_{corr} of 5.3, a $K_{D,app;Ca}$ of 7 ± 3 mM was determined from intensities and a $K_{D;Ca}$ of 1.7 ± 0.2 mM was determined from line shapes (c.f. chapter 5.4.5). Line shape analysis at a pH_{corr} of 7.4 yielded a K_D of 1 mM (see above). The dissociation constants as well as on and off-rates are summarized in Tab. 5.7.

Notably, the off-rate of 28 s^{-1} together with the dissociation constant $K_{D,Ca2+}$ of 1 mM determined at a pH_{corr} of 7.4 translate in an on-rate $k_{on,Ca2+}$ of $2.8*10^4$ $M^{-1}s^{-1}$. Estimations for the on-rate at a pH_{corr} of 5.3 for Mg^{2+} set the value below $8*10^3$ $M^{-1}s^{-1}$ (c.f. chapter 5.4.4), which is consistent with what is observed at a pH_{corr} of 7.4 for Ca^{2+}. Such an on-rate is far away from diffusion-controlled (Meyer and Peters, 2003), validating the conclusion of a rate-limiting "global transition" taking place before metal ions can bind to the P-domain (c.f. chapter 5.4).

Tab. 5.7: Thermodynamics and kinetics of metal ions binding to the P-domain at different pH_{corr} values.

pH_{corr}	$K_{D,Ca2+}$	$k_{off,Ca2+}$	$k_{on,Ca2+}$	Data shown in (method)
4	No binding	/	/	Fig. 5.5.2 (CSPs)
5.3	1.7 ± 0.2 mM	3052 ± 2998 s^{-1}	/	Fig. S 4.6 (line shapes)
	8 ± 3 mM	$k_{off,Mg2+}$<< 88 s^{-1} (value for Mg^{2+})	$k_{on,Mg2+}$ < 8*10^3 $M^{-1}s^{-1}$ (value for Mg^{2+})	Fig. 5.4.7 and main text of chapter 5.4.4 (CSPs)
7.4	1.09 ± 0.02 mM	27.5 ± 0.4 s^{-1}	Ca. 2.8*10^4 $M^{-1}s^{-1}$	Fig. 5.5.2 (line shapes)

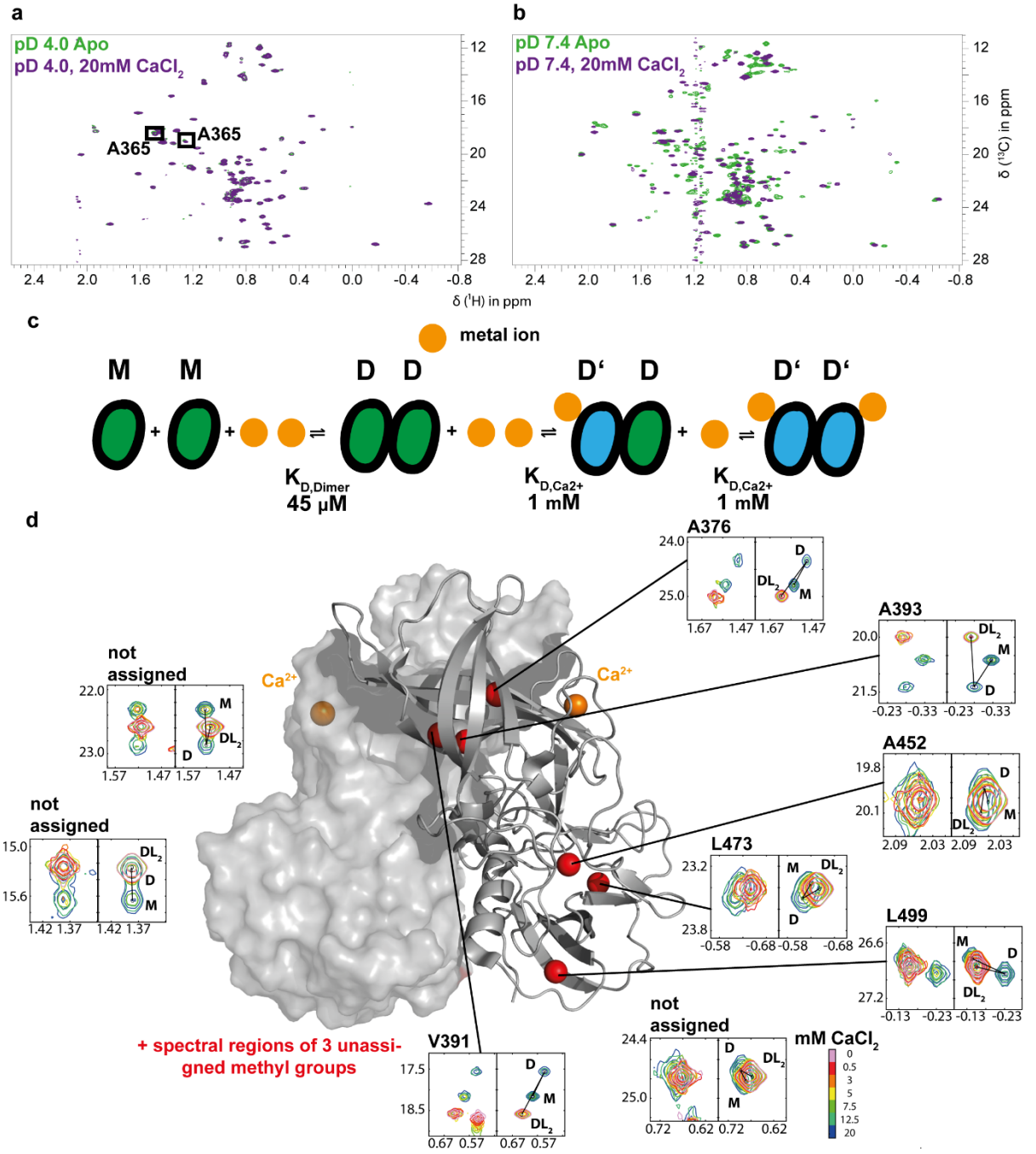


Fig. 5.5.2: Binding of metal ions is strongly regulated by the pH value. [^{1}H,^{13}C] HMQC spectra of P-domain at a pH_{corr} of 4.0 show no significant global changes due to addition of 20 mM $CaCl_2$ (a). CSPs observed for Ala365 only report for binding to the receptor epitope metal ion binding site (see chapter 5.4). However, at a pH_{corr} of 7.4, addition of metal ions causes drastic changes in [^{1}H,^{13}C] HMQC spectra (b). This is consistent with what was observed at a pH_{corr} of 5.3 (c.f. Fig. 5.4.3). Acquisition of P-domain spectra of the sample shown in (b) in the presence of different $CaCl_2$ concentrations and 2D line shape analysis allowed to determine a $K_{D,Ca2+}$ of 1.09 ± 0.02 mM and a $k_{off,Ca2+}$ of 27.5 ± 0.4 s^{-1} (c, d). The resulting on-rate $k_{on,Ca2+}$ for the metal ion is approximately $2.8*10^4$ $M^{-1}s^{-1}$. The left panels in (d) show the measured spectra, and the right panels show the simulated spectra. Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 52 µM or 58 µM MILVA labeled MNV CW1 P-domain at pH_{corr} values of 4.0 (20 mM acetate-d_4, 100 mM NaCl) or of 7.4 (20 mM BisTris-d_{19}, 100 mM NaCl). A structural model of the P-domain by Nelson et al., 2018, (pdb 6e47) was used for the illustration.

In a next step, the influence of pH_{corr} on the binding of GCDCA was investigated. GCDCA was titrated to P-domain in citric acid buffers at a pH_{corr} of 5.3 and at a pH_{corr} of 4.4. The two pH_{corr} values reflect the upper and lower limits of pH values obtained in gastrointestinal tracts of mice (McConnell et al., 2008). For the analysis, it is necessary to consider the following: the previous analysis of GCDCA binding to P-domain at a pH_{corr} of 5.3 was performed in acetate buffer (c.f. chapter 5.3.3). However, commercially available GCDCA is contaminated with small amounts of Ca^{2+} ions (Creutznacher et al., 2021). It is estimated that ca. 150 µM Ca^{2+} ions were added to the sample at the end of the GCDCA titration. Because of the thermodynamic and kinetic cross-talk between metal ion and GCDCA binding (c.f. chapter 5.4.5), the derived values should be critically questioned. To account for this problem, citric acid buffer instead of acetate buffer was used here to investigate the effect of the pH on GCDCA binding. Citric acid is known to effectively chelate Ca^{2+} ions (Welling et al., 2014). Comparing P-domain spectra in the presence of GCDCA using either citric acid buffer or acetate buffer shows that chemical shifts of particular resonances are affected that are also affected by $CaCl_2$ titration (Fig. 5.5.3a). In Citric acid buffer, these resonances shift towards the unbound state, emphasizing the chelation of Ca^{2+} (Fig. 5.5.3b). Titration of GCDCA to the P-domain in citric acid buffer should therefore compensate for the contamination of GCDCA by Ca^{2+} ions and allow a "clean" analysis.

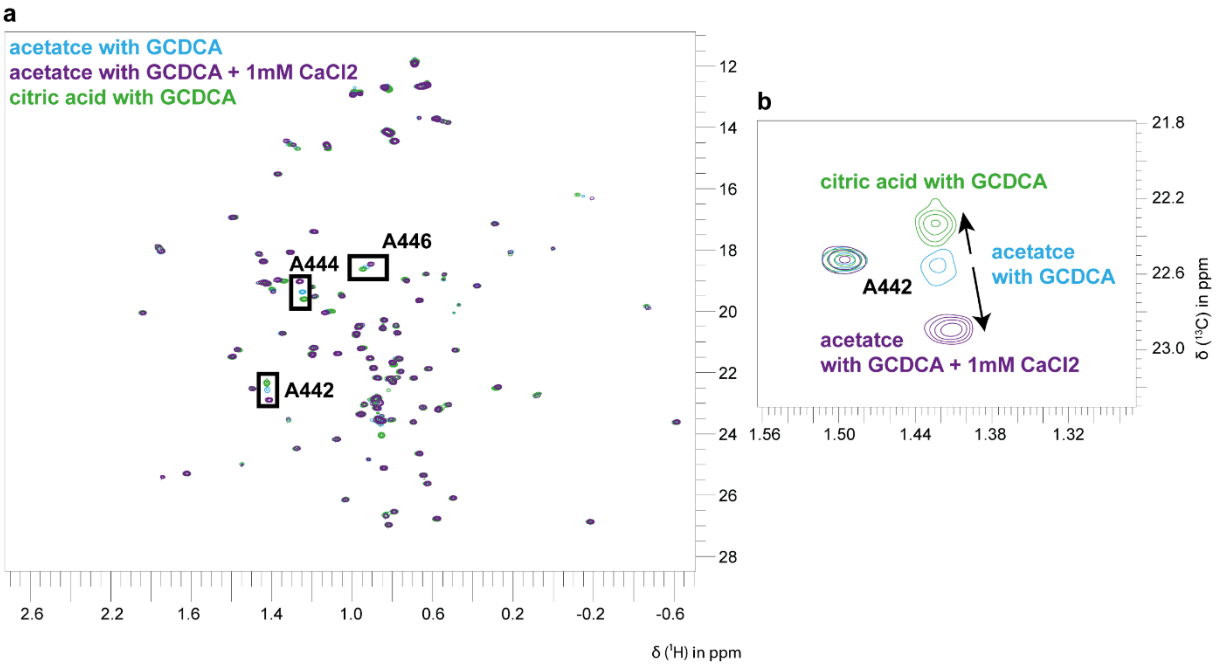


Fig. 5.5.3: Ca^{2+} ions in commercially purchased GDCA are chelated by citric acid buffer. (a) shows [^{1}H,^{13}C] HMQC spectra of P-domain in the presence of GCDCA. As pointed out in Creutznacher et al., 2021, commercially purchased GCDCA is contaminated with Ca^{2+} ions. Comparing the spectra recorded of samples in 20 mM acetate-d_4 buffer, 100 mM NaCl at a pH_{corr} of 5.3 with and without 1 mM $CaCl_2$ with a spectrum of a sample in 20 mM deuterated citric acid buffer, 100 mM NaCl at a pH_{corr} of 5.3 reveals that the Ca^{2+} ions are chelated by the citric acid buffer, as indicated by CSPs. (b) shows a zoom of (a) into the spectral region corresponding to the Ala442 resonance. Spectra of samples in citric acid were acquired with 47 µM P-domain and 300 µM GCDCA. Spectra of samples in acetate buffer were acquired with 34 µM P-domain and 250 µM GCDCA. All spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe.

For the analysis at a pH_{corr} of 5.3 in citric acid buffer, 11 suitable resonances were selected for that the three states of the monomeric P-domain, the dimeric P-domain, and the GCDCA bound P-domain were

at least partially resolved. The corresponding spectral regions were quantum-mechanically simulated and fitted to measured spectra using *TITAN* (Waudby et al., 2016). The analysis yielded a dissociation constant $K_{D,GCDCA}$ of 15 ± 1 µM and an off-rate $k_{off,GCDCA}$ of 17 ± 4 s^{-1} (Fig. S 5.1), not significantly differing from the analysis at a pH_{corr} of 5.3 in acetate buffer (c.f. chapter 5.3.3). The obtained dissociation constant and off-rate result in an on-rate $k_{on,GCDCA}$ of $1.1*10^6$ $M^{-1}s^{-1}$.

At a pH_{corr} of 4.4 in citric acid buffer, almost no monomer resonances are detectable, and the presence of monomeric P-domain can be neglected. Therefore, a two-step binding model instead of a three step binding model was used (Fig. 5.5.4a), yielding a dissociation constant $K_{D,GCDCA}$ of 16 ± 2 µM and an off-rate $k_{off,GCDCA}$ of 0.4 ± 0.8 s^{-1} (Fig. 5.5.4b). Due to the high uncertainty obtained for the off-rate, only an upper limit for the on-rate can be calculated. Assuming an off-rate of 1.2 s^{-1} results in an on-rate $k_{on,GCDCA} < 8*10^4$ $M^{-1}s^{-1}$. Comparing the values to the analysis at a pH_{corr} of 5.3 shows that dissociation constants are not affected, whereas there is a significant difference between the respective off and the on-rates. A shift of the off-rate towards smaller values was also observed for GCDCA binding in the presence of metal ions (c.f. Fig. 5.4.10). It is tempting to hypothesise that the compensation of the negative charges at the G'H' loop are responsible for the altered binding kinetics. However, since the protonation state of GCDCA also changes between pH 4.4 and 5.3 (Hofmann, 1963), it is difficult to draw mechanistic conclusions.

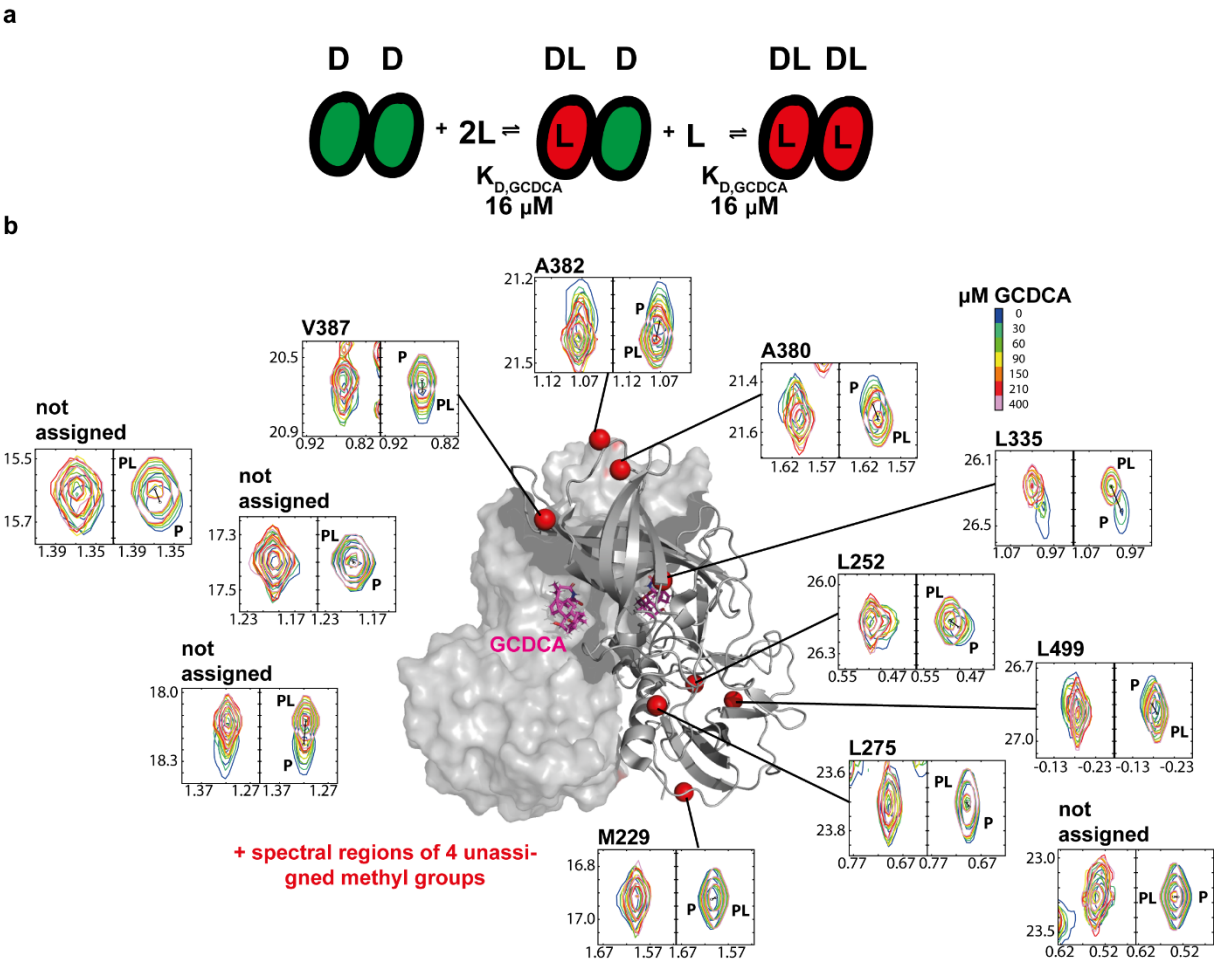


Fig. 5.5.4: P-domain GCDCA binding kinetics but not thermodynamics is affected by pH. Line shape analysis of P-domain spectra at pH_{corr} of 4.4 using different GCDCA concentrations resulted in a K_D of 15.5 ± 2.3 µM and a k_{off} of 0.4 ± 0.8 s^{-1}. Whereas the K_D is comparable to what was observed at a pH_{corr} of 5.3 (11 µM or 15 µM ;c.f. Fig. 5.3.5, Fig. S 5.1), the off-rate of the binding is significantly slower compared to the values determined for pH_{corr} 5.3 (26±4 s^{-1} or 17±4 s^{-1}). The left panels show the measured spectra and the right panels show the simulated spectra. Notably, this off-rate is comparable to what was observed for GCDCA binding in the presence of Ca^{2+} ions (c.f. Fig. 5.4.10). Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 64 µM MILVA labeled MNV CW1 P-domain in 20 mM deuterated citric acid, 100 mM NaCl, pH_{corr} 4.4. A structural model of the P-domain by Nelson et al., 2018, (pdb 6e47) was used for the illustration.

5.5.3 G'H' loop protonation parallels with P-domain dimerization

The G'H' loop is involved in GCDCA binding and metal ion binding, as illustrated in Fig. 5.5.5a. Both events result in dimerization of the P-domain (Creutznacher, 2020; chapters 5.3.2 and 5.4.2). It is also noteworthy that metal ion binding to the G'H' loop is abolished at a pH_{corr} value of 4 (c.f. chapter 5.5.2) and that pH_{corr} values between 4 and 5 also induce P-domain dimerization (c.f. 5.5.1). Protons and metal ions could both compensate for the negative charges of the G'H' loop residues Asp440, Asp443, and Glu447 (c.f. Fig. 5.5.5a). This could be the mechanism responsible for the formation of P-domain dimers and abolishment of metal ion binding at acidic pH values. In the literature, ionization constants pk_a of 3.9 and 4.3 are given for the side chain carboxy groups of aspartic acid and glutamic acid residues (PubChem IDs 5960 and 33032). However, such values can be very different in the context of proteins and must therefore be verified experimentally (Sereda et al., 1993).

For the determination of ionization constants of the acidic residues at the G'H' loop, [^{1}H,^{13}C] HMQC spectra of the P-domain at different pH_{corr} values were acquired. The methyl groups of Ala442, Ala444, and Ala446 are located at the G'H' loop. The corresponding resonances are only well resolved in the presence of GCDCA. Tracking methyl group CSPs in the presence of GCDCA shows pH_{corr} dependant CSPs (Fig. 5.5.5c). For pH_{corr} values between 3 and 6.2, citric acid buffers were used. Bis-Tris buffers were used for pH_{corr} values ranging between 6.2 and 7.4. Differences in chemical shifts between the two buffers were corrected by comparing spectra of the P-domain in both buffers at the same pH_{corr} (pH_{corr} of 6.2). Apparent pk_s values can be fitted using CSPs at different pH_{corr} values and an adopted version of the Henderson-Hasselbalch equation (Fig. 5.5.5b). The derived values are as follows: a pk_a of 5 for Ala442, and pk_a values of 4.5 and 4.6 for Ala444 and Ala448. It should be noted that the alanine methyl groups are not directly protonated. Furthermore, CSPs do not follow straight vectors but curves. This is probably caused by simultaneous protonation events at different sites, as expected by the three acidic G'H' loop residues (Fig. 5.5.5a). Therefore, the derived pk_s values must be considered as apparent values. Nevertheless, it can be clearly stated that the protonation state of the G'H' loop changes between a pH_{corr} of 6 and a pH_{corr} of 4. This is the same pH range in which P-domain dimerization and abolishment of metal ion binding is observed and therefore might explain pH dependent P-domain dimer formation and abolishment of metal ion binding at a pH_{corr} of 4.

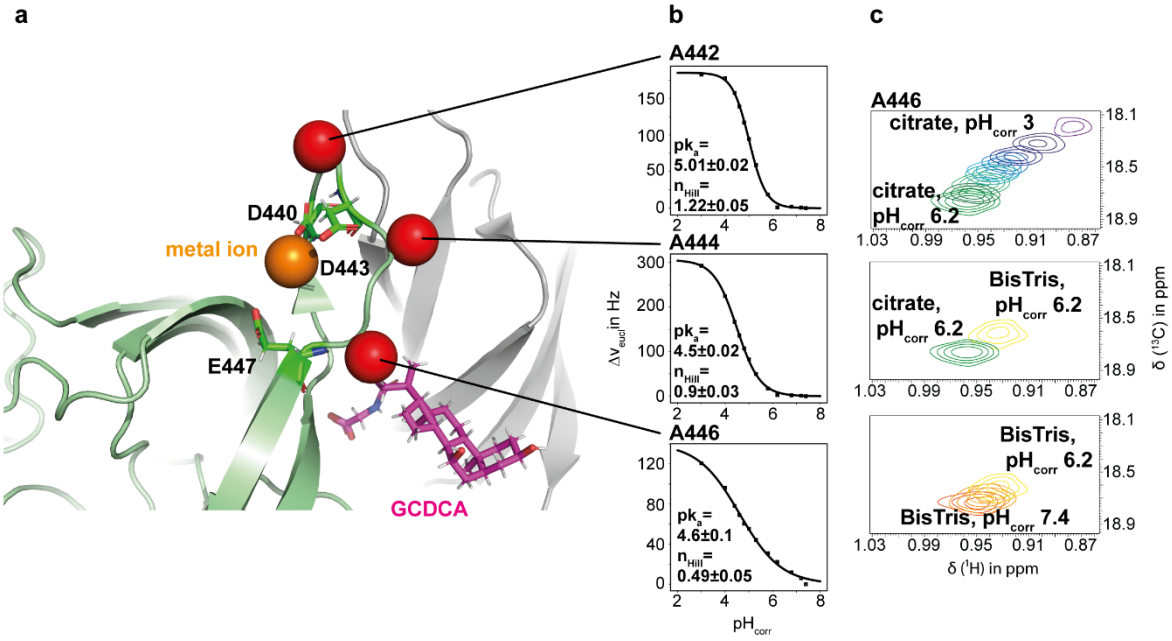


Fig. 5.5.5: Protonation of G'H' loop parallels with dimer formation. (a) shows the G'H' loop which is involved in metal ion and GCDCA binding (pdb 6e47, Nelson et al., 2018). Both events lead to dimerization of the P-domain. The loop contains three acidic residues (Asp440, Asp443 and Glu447). Alanine methyl groups within the loop are highlighted as red spheres. The corresponding resonances show CSPs due to the alteration of the pH_{corr} value. CSPs can be used to fit the Henderson-Hasselbalch equation (Croke et al., 2011, see chapter 4.4.7) to derive ionization constants pk_a. Even though the CSPs probably do not report for protonation of a single side, it can be concluded that the G'H' loop changes its protonation state between pH_{corr} 4 and 6. (c) shows CSPs due to changes of pH_{corr} exemplarily for A446. As origin, the position of the resonance at a pH_{corr} of 3 was used and the pH_{corr} was increased to 6.2 (upper panel). BisTris buffer was used to follow CSPs in the neutral range (pH_{corr} 6.2-7.4, lower panel). To account for the impact of the different buffers, spectra were acquired at a pH_{corr} of 6.2 in both, citrate buffer and BisTris buffer (panel in the middle). For details on the acquisition of the respective NMR spectra, see Fig. 5.5.1. In addition to the information given there, saturating amounts of GCDCA were added to the samples.

5.5.4 GCDCA redistributes pico- to nanosecond dynamics of P-domain

The pH-dependent formation of P-domain dimers allowed further investigations into the effects of GCDCA on the pico- to nanosecond dynamics of P-domain methyl groups. Methyl group order parameters of the apo P-domain and the GCDCA-bound P-domain were determined as explained previously (c.f. chapter 5.4.8) and compared. For the calculation of S^2 order parameters, the determination of the molecular correlation time τ_C is necessary. Since this is generally done by analyzing 1D TRACT spectra of [U-^{2}H,^{15}N]-labeled proteins (Lee et al., 2006), it is not possible to resolve τ_C for different co-existing protein species such as monomers and dimers in the case of the P-domain. Instead, τ_C would be averaged over all protein states in the sample and reliable order parameters cannot be calculated from these values. However, monomer resonance intensities at a pH_{corr} of 4.6 or pH_{corr} 4.4 suggest, that almost all P-domain proteins exist in the dimeric state at these pH_{corr} values. This is also confirmed by molecular correlation times τ_C of the P-domain at a pH_{corr} of 4.6 in the presence and absence of GCDCA (Fig. 5.5.6), which indicate an almost identical molecular weight.

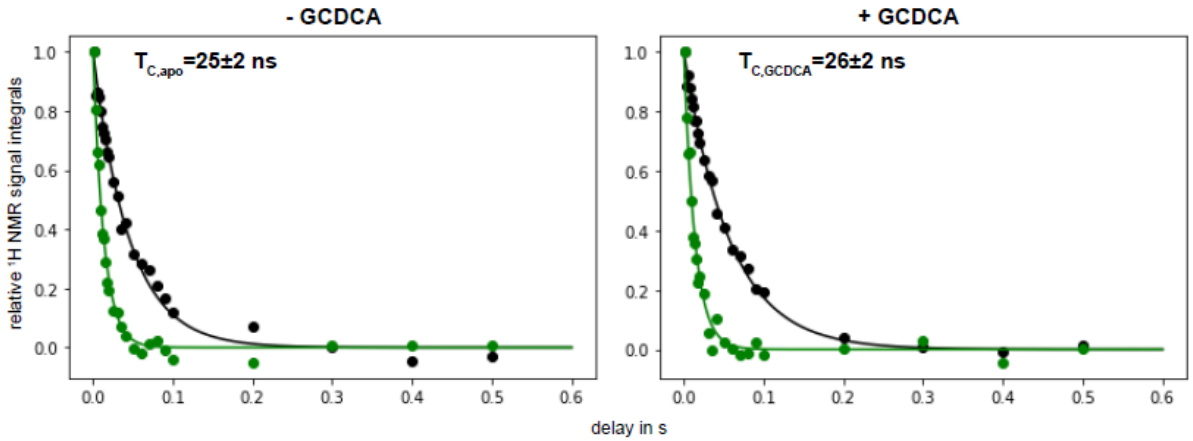


Fig. 5.5.6: TRACT experiments show identical molecular correlation times τ_C in the presence and absence of GCDCA at a pH of 4.6. Sample conditions, experimental details, and details on the analysis are given in chapters 4.3.3 and 4.4.6.

Comparing methyl group CSPs due to the addition of GCDCA indicates that the P-domain at a pH_{corr} of 4.6 is similarly affected by the GCDCA as it is described at a pH_{corr} of 5.3 (c.f. chapter 5.3.3 and Fig. 5.5.7b and d). Using the derived molecular correlation times, S^2 order parameters were calculated for the apo P-domain state and the GCDCA-bound P-domain state. As can be seen by plotting order parameter differences on a crystal structure model, methyl groups throughout the protein are affected (Fig. 5.5.7a and c). Strikingly, almost all methyl groups experience an increase in flexibility and thus make a positive contribution to the conformational entropy of the protein. This supports the results presented in chapter 5.4.8, where order parameters of Ca^{2+} bound and Ca^{2+} and GCDCA bound P-domains were compared. It is evident that changes in both observables, CSPs and order parameters, are present throughout the protein. This emphasizes the existence of allosteric network-like connections between the methyl groups of the P-domain.

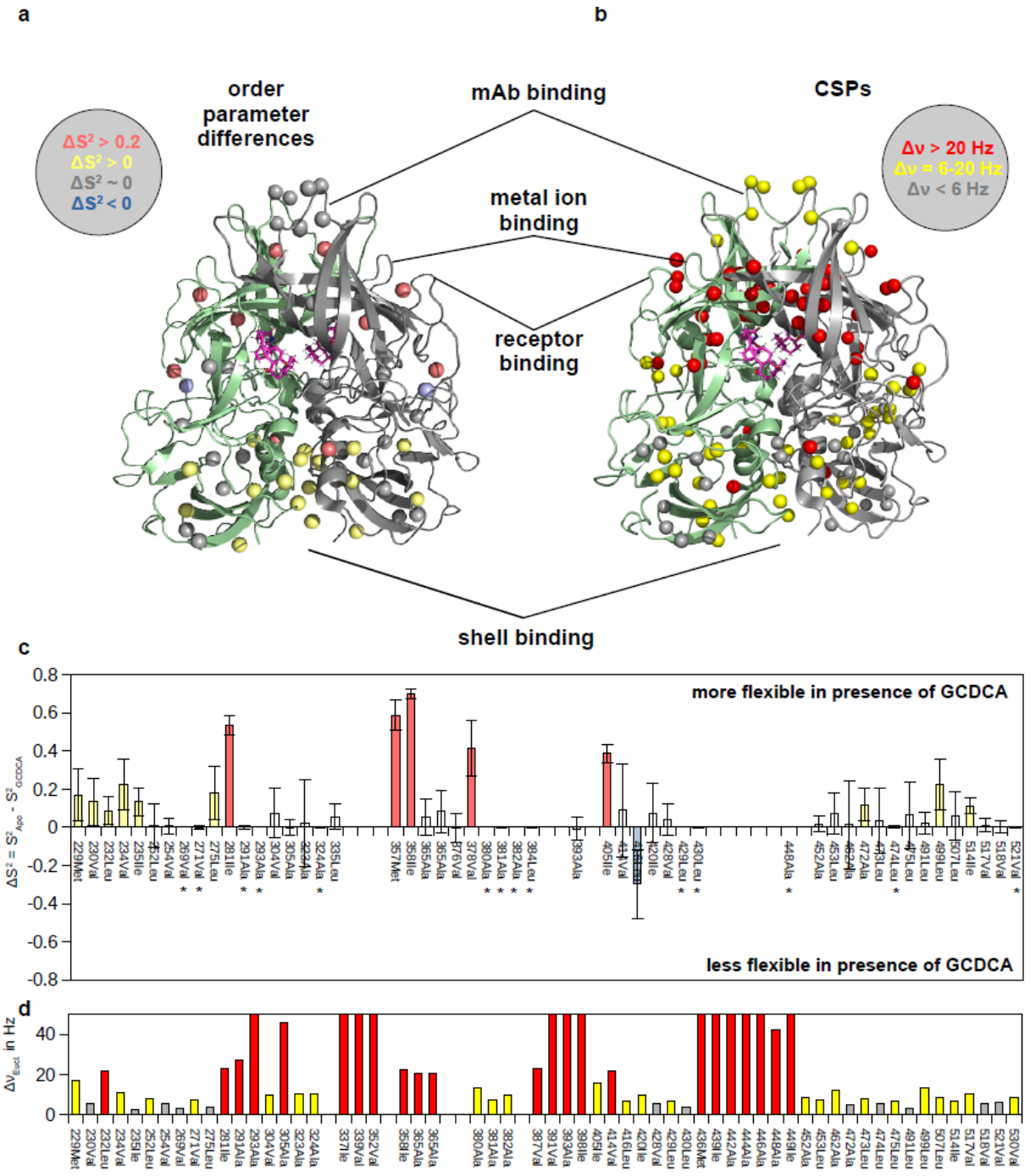


Fig. 5.5.7: Methyl group order parameters and CSPs of P-domain in the presence and absence of GCDCA. (a) shows methyl group S^2 order parameter differences of P-domain with and without GCDCA (pdb 6e47, Nelson et al., 2018). (b) Methyl group CSPs of P-domain caused by GCDCA (pdb 6e47, Nelson et al., 2018). (c) Bar plot of order parameter differences in the presence and absence of GCDCA. Notably, most methyl groups increase in flexibility due to GCDCA binding. The color code is as shown in (a). (d) shows CSPs between the two states. The CSP color coding is as follows: grey (0-6Hz), yellow (6-20 Hz), and red (more than 20 Hz) and matches the color code in CSP plots throughout this thesis. CSPs were determined from HMQC spectra acquired for S^2 determination (c.f. Tab. 4.12 and chapter 4.3.3). S^2 order parameters of the individual states are shown in Fig. S 5.2. The corresponding spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe. Information about sample conditions is given in Tab. 4.12.

5.6 SARS-CoV-2 spike glycoprotein and RBD interactions with glycans and lead compounds for drug development

In late 2019, a new virus of the *Coronaviridea* family, SARS-CoV-2, emerged and has since caused millions of deaths and heavy economic damage worldwide. Given the severe consequences of the outbreak, researchers from almost every country are shifting their focus to gain better understanding of SARS-CoV-2. One important viral protein is the spike glycoprotein. It is embedded in the viral membrane and interacts with host factors such as the ACE2 protein receptor and neutralizing antibodies. Thus, it takes a role that is analogue to that of the P-domain of MNV. These properties make it a promising drug target. Among the variety of small molecules that might have the potential to act as a lead compound for drug development is Lifitegrast. Chemical shift perturbation (CSP) experiments with the isotope-labeled receptor binding domain (RBD) of the spike protein produced in *E. coli* were performed in the Peters laboratory by Dr. Robert Creutznacher (Creutznacher et al., 2022b). The experiments show that Lifitegrast binds to the highly conserved fatty acid binding pocket of the RBD. In this thesis it is verified by STD NMR that Lifitegrast does not only bind to the non-glycosylated RBD, but also to the glycosylated spike protein, where some RBD sites are buried (Fig. 3.2.1b). Another finding is that sialoglycans also interact directly with the spike protein (original data not shown because of redundancy with Fig. 5.6.2; published in Creutznacher et al., 2022b). This is of great interest as it is suspected that sialoglycans act as co-factors for SARS-CoV-2 (Sun, 2021). The potential role of sialoglycans acting as co-factors called for a more profound biophysical characterization of the binding reaction. STD NMR was employed to derive dissociation constants of sialoglycans with the spike protein of the Wuhan variant. Remarkably, STD NMR experiments with the Omicron BA1 spike protein revealed a loss of sialoglycan binding. In addition, the sialoglycan mimetics Oseltamivir and Relenza were found to interact with the spike protein but not with the RBD. Finally, interactions of human blood group antigens (HBGAs) with the RBD or the spike protein can be ruled out by CSP and STD NMR experiments.

5.6.1 Wuhan SARS-CoV-2 spike glycoprotein interacts with Lifitegrast

As explained above, a small molecule, Lifitegrast (structure shown in Fig. 5.6.1b), which binds to the RBD, was identified in the Peters laboratory by Dr. Robert Creutznacher (Creutznacher et al., 2022b). It could serve as a lead compound for drug development. The RBD construct was produced in *E. coli.* It has been shown to bind to the ACE2 receptor with similar thermodynamics and kinetics to a glycosylated RBD construct produced in insect cells, indicating proper folding and integrity of the protein. However, some RBD sites are buried in the spike context (Fig. 3.2.1b) and the RBD produced in *E. coli* is not glycosylated. Therefore, it was necessary to test Lifitegrast binding in the employing the fully glycosylated spike protein.

STD NMR spectra show binding of Lifitegrast and prove that the binding pocket is accessible in the spike context. Comparing line widths in off-resonance spectra of Lifitegrast with and without spike protein shows strong line broadening, which impedes ligand epitope mapping (Fig. 5.6.1a). However, saturation transfer is clearly visible for the methyl group and aromatic protons of Lifitegrast. Due to the molecular weight of more than 420 kDa, STD NMR constitutes a well-suited method to study spike protein interactions with low molecular weight ligands. Strikingly, STD spectra of sufficient quality can be acquired within less than 1.5 hours, allowing screening or titration of ligands in a relatively short time.

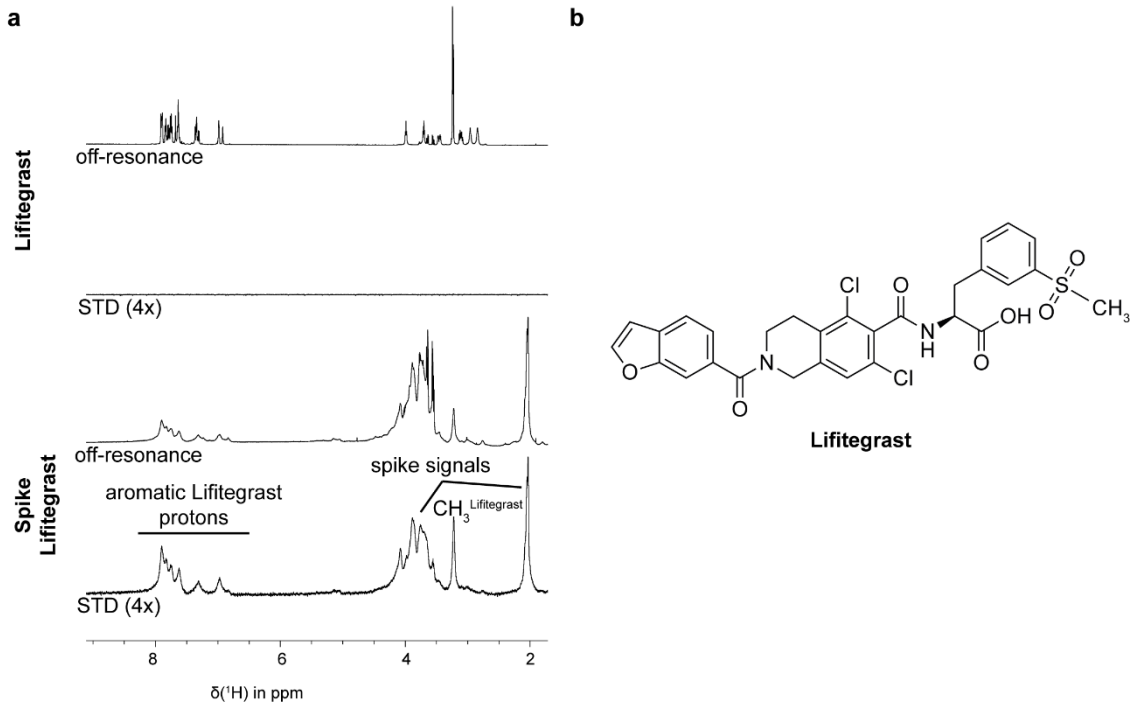


Fig. 5.6.1: STD NMR verifies the accessibility of the Lifitegrast binding pocket in the context of the Wuhan SARS-CoV-2 spike protein. (a) shows STD and off-resonance spectra of Lifitegrast in the presence and absence of the spike protein. Line broadening effects in the presence of the protein impedes ligand epitope mapping. However, saturation transfer is clearly visible for the aromatic protons and the methyl group of Lifitegrast. (b) Chemical structure of Lifitegrast. Spectra were acquired using 1 mM Lifitegrast, 45 µM spike protein on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K. On-resonance frequencies were set to -4 ppm. Assignments are based on Creutznacher et al., 2022b. The spike protein concentration was estimated by comparing intensities of spike signals to intensities of a sample with a known concentration acquired under identical conditions. Figure and legend are adapted from Creutznacher et al., 2022b.

5.6.2 STD NMR-derived dissociation constants and ligand epitope mapping of sialoglycans, mimetics, and human blood group antigens

The binding of sialoglycans and ABO(H) human blood group antigens (HBGAs) to the RBD and the spike protein is controversial in the literature. Mass spectrometry data suggest binding of sialoglycans and HBGAs to the RBD with dissociation constants in the mid µM range (Nguyen et al., 2022). Atomic force microscopy (AFM) provides a dissociation constant of 5 µM for 9-O-acetylated 5N-acetyl-neuraminic acid (9-O-Ac-Neu5Ac) to the spike protein, but only a low affinity for non-9-O-acetylated Neu5Ac (Petitjean et al., 2022). An STD NMR study from the Jiménez-Barbero group showed that sialoglycans bind to the N-terminal domain (NTD) of the spike protein and not to the RBD (Unione et al., 2022). A combined cryo-electron microscopy and "uSTD" NMR study confirmed the location of the binding site at the NTD. In this study, a dissociation constant of non-9-O-acetylated sialoglycans in the low µM range was derived (Buchanan et al., 2022). These results are contradictory and require a solid investigation of SARS-CoV-2 spike protein interactions with sialoglycans and HBGAs.

In this thesis, a classical and well-established approach to determine dissociation constants form STD NMR spectra was employed (Neffe et al., 2007). STD experiments were performed with the Wuhan

SARS-CoV-2 spike protein. Using STD Amplification factors (STD-Afs) of sialoglycans at different concentrations resulted in mM dissociation constants for 6' sialyl lactose (6'SL, Fig. 5.6.2a), 3' sialyl lactose (3'SL, Fig. 5.6.2b), and Sialyl-Lewis^X (Fig. 5.6.2c).

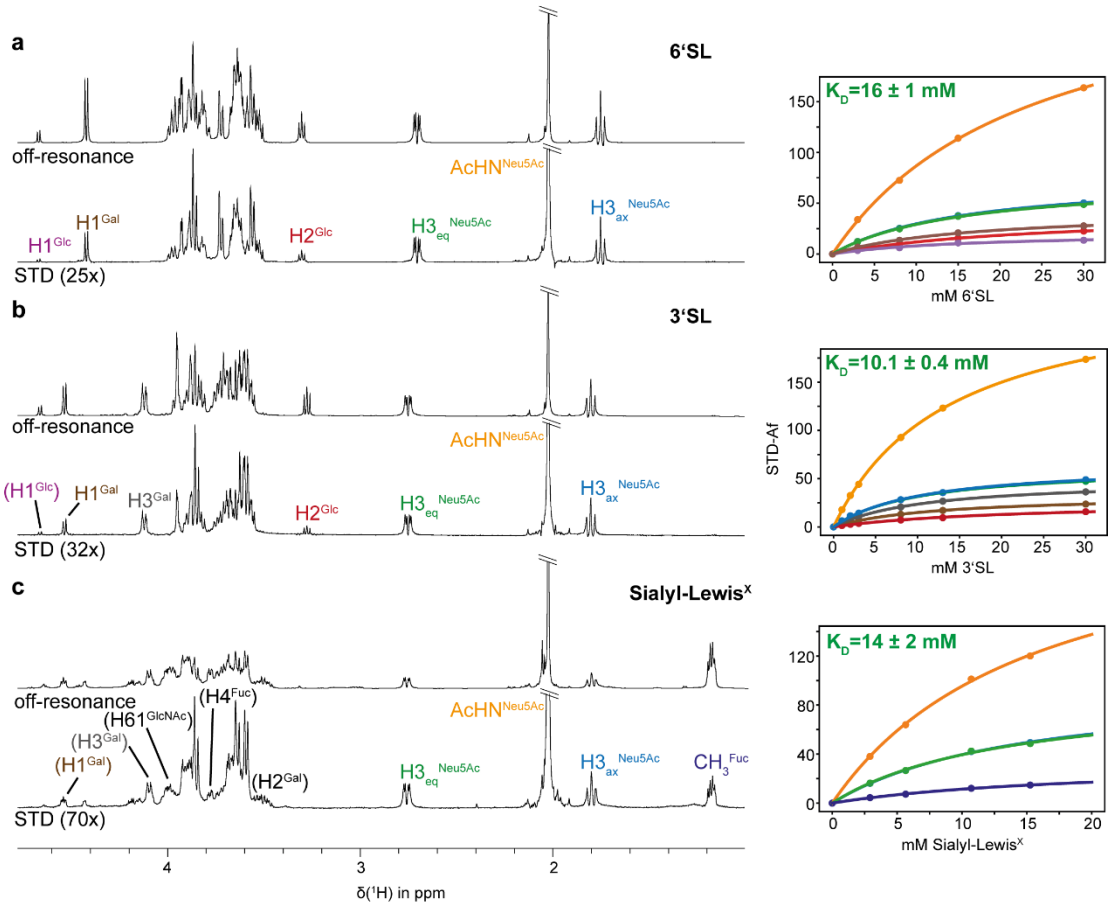


Fig. 5.6.2: STD-NMR derived binding affinities of sialoglycans to the Wuhan SARS-CoV-2 spike protein. Off-resonance, STD spectrum and binding isotherm of 6'SL (a), 3'SL (b), and Sialyl-Lewis^X tetrasaccharide (c). Binding isotherms and K_Ds are color-coded according to the respective labels in the STD spectra. Individual fits of the respective protons are summarized in Tab. S 6.1. Brackets indicate protons only used for epitope mappings shown in Fig. 5.6.3 but not for fitting binding isotherms. Negative controls are shown in Fig. S 6.1 and S 6.2. Spectra were acquired using 21 µM, 19 µM, and 12 µM spike glycoprotein and 30 mM of 3'SL or 6' SL and 15.2 mM Sialyl-Lewis^X tetrasaccharide on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K. On-resonance frequencies were set to -4 ppm. Assignments are based on literature data (Nakamura et al., 2000; Haselhorst et al., 2001; Fiege et al., 2012; Unione et al., 2022). It should be noted that the Sialyl-Lewis^X compound's purity is given with only >80 % by the manufacturer. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).The Sialyl-Lewis^X data was added to the figure.

The resulting binding epitopes (Fig. 5.6.3) agree well with the binding epitopes of partially ^{13}C-labeled sialoglycans obtained from 2D STD spectra (Unione et al., 2022). In addition, they prove the importance of the acetyl group for the binding reaction, as STD effects were most pronounced here for all sialoglycans tested.

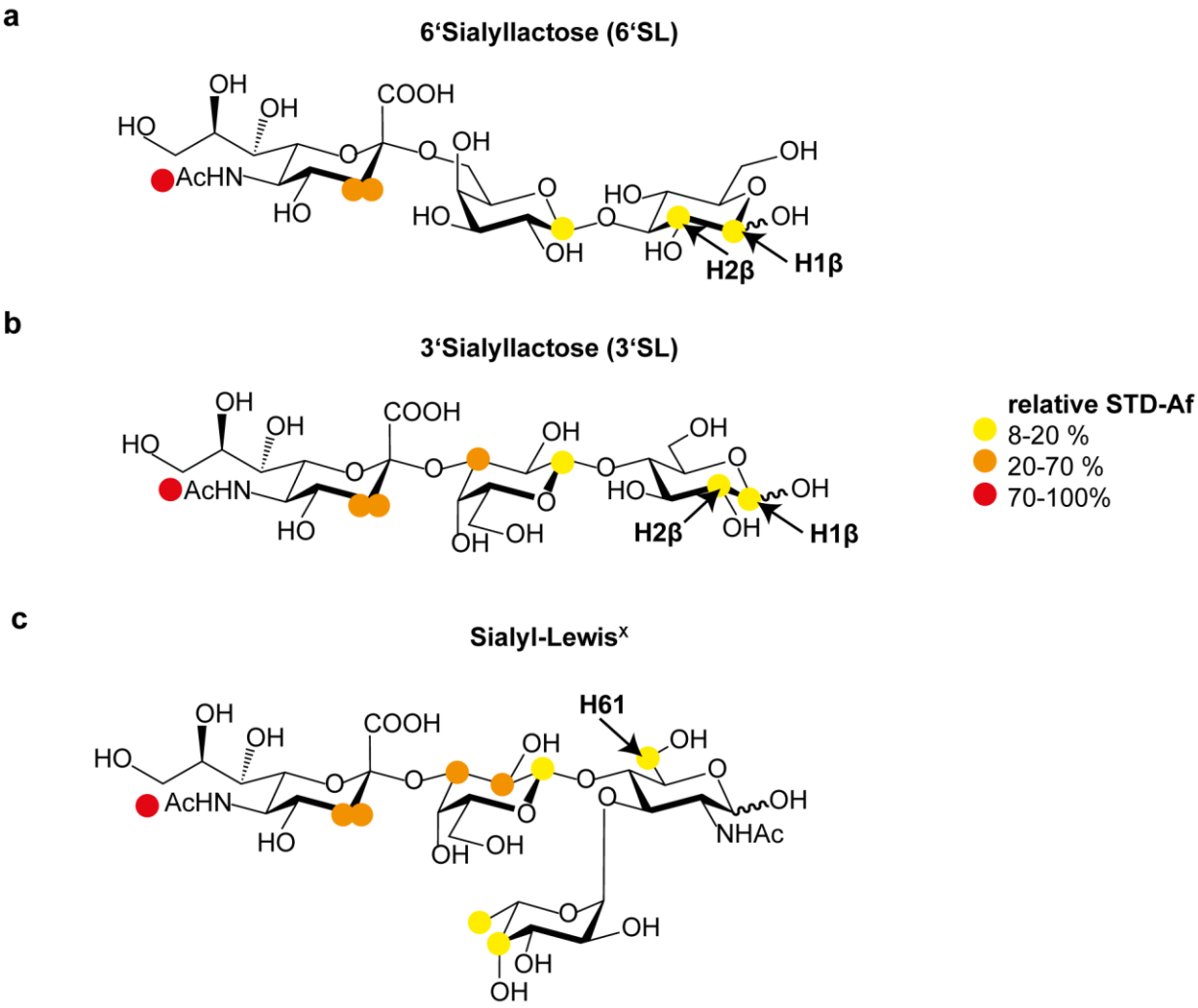


Fig. 5.6.3: STD NMR-derived epitope mappings of sialoglycan ligands binding to the Wuhan SARS-CoV-2 spike protein. STD spectra of 6'SL, 3'SL, and Sialyl-Lewis^x used for epitope mapping are shown in Fig. 5.6.2. All signals used for epitope mappings are highlighted in in Fig. 5.6.2. STD effects of other protons were not considered in the epitope mappings since the corresponding signals were not sufficiently resolved. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/). The Sialyl-Lewis^x data was added to the figure.

Sialoglycans are a prominent co-factor of influenza viruses. The sialoglycan mimetics Oseltamivir and Zanamivir target the neuraminidase of influenza viruses (Jackson et al., 2011). STD NMR experiments show that both molecules also bind to the SARS-CoV-2 spike protein (Fig. 5.6.4a and Fig. 5.6.5a). The binding epitopes (Fig. 5.6.4b and Fig. 5.6.5c) differ from those obtained for the sialoglycans, where the acetyl group experienced the greatest saturation transfer. In contrast, Oseltamivir and Zanamivir binding epitopes show the largest saturation transfer for the H2 and H5 protons, respectively (Fig. 5.6.4b and Fig. 5.6.5c). These protons are attached to a double bond within a six-ring structure. In the case of Oseltamivir, the carboxylic acid part of the molecule was also relatively strongly affected. This could indicate a different binding mode of the mimetics as compared to sialoglycans. Due to the relatively low solubility of Zanamivir in aqueous solutions, titration experiments and the determination

of dissociation constants were not possible. However, for Oseltamivir, a dissociation constant of 4 mM was readily derived (Fig. 5.6.5b).

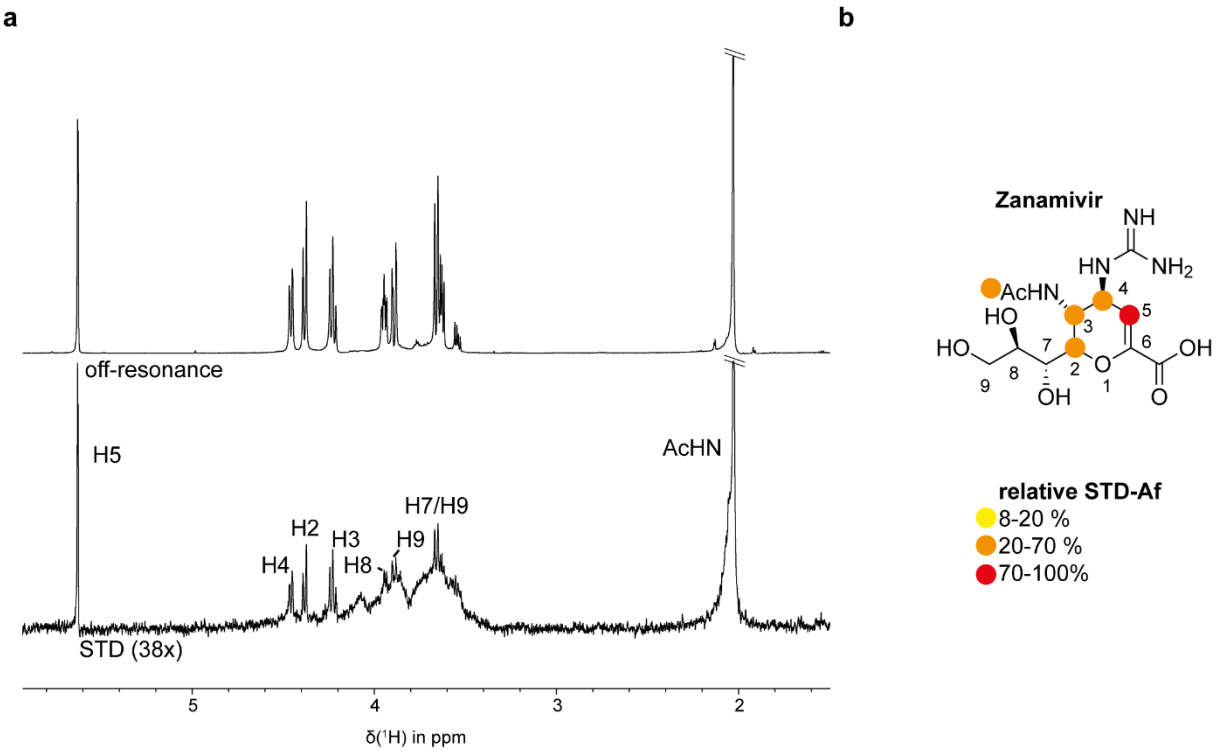


Fig. 5.6.4: Zanamivir binding to Wuhan SARS-CoV-2 Spike glycoprotein. (a) STD and off-resonance spectrum with the Wuhan spike protein show binding of Zanamivir. STD-Afs of H8, H9, and H7 were not considered because of the strong protein background in the respective spectral region (b). Spectra were acquired using 22 µM spike glycoprotein and 2 mM of Zanamivir on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K. See Fig. S 6.3 for negative controls. On-resonance frequencies were set to -4 ppm. Assignments are based on the literature (Duan et al., 2022). The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

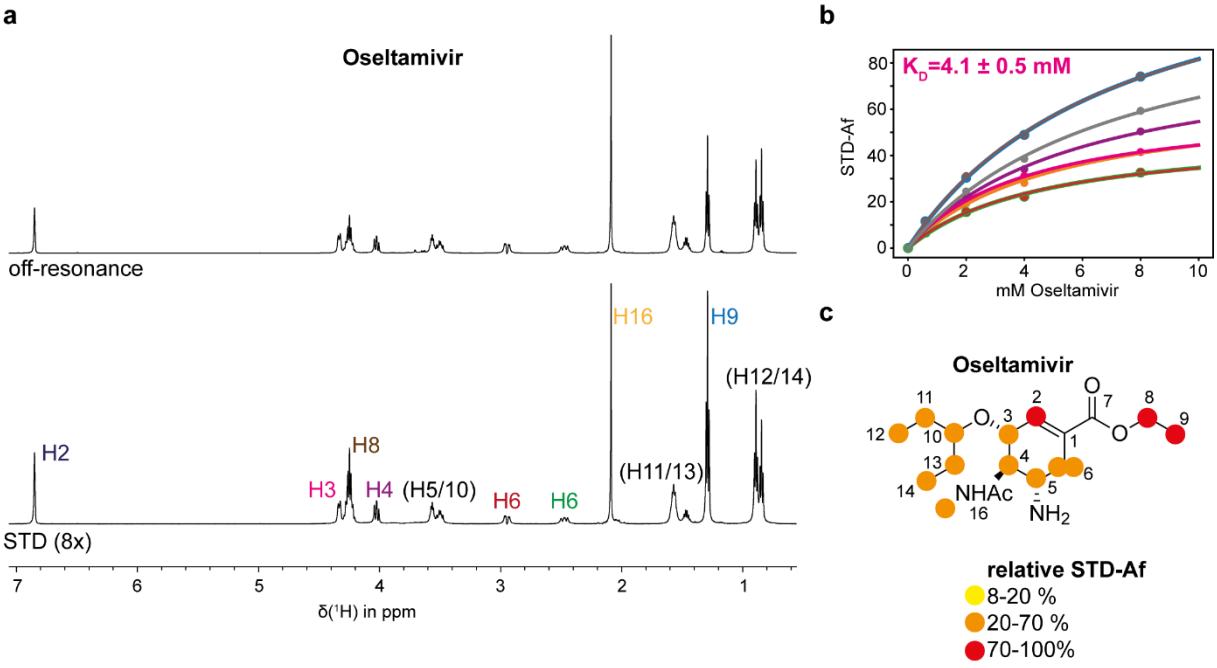


Fig. 5.6.5: STD NMR derived dissociation constant of Oseltamivir binding to Wuhan SARS-CoV2 spike glycoprotein. (a) shows off-resonance and STD spectra of Oseltamivir. Brackets indicate protons not used for deriving dissociation constants but for the epitope mapping shown in (c). (b) shows binding isotherms of the respective protons resulting in a low mM dissociation constant. See Tab. S 6.1 for individual fits. The color coding in (b) corresponds to the colors used in (a). (c) shows the STD-based epitope mapping. Spectra were acquired using 27 µM spike glycoprotein and 8 mM Oseltamivir on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K. On-resonance frequencies were set to -4 ppm. Assignments are based on the literature (D'Souza et al., 2009). See Fig. S 6.3 for negative controls. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

Titrating Oseltamivir to the spike protein in the presence of 30 mM 6'SL results in little or no decrease of 6'SL STD-Afs. In addition, no significant change of the dissociation constant of oseltamivir was observed (Fig. 5.6.6a). Thus, it can be confidently concluded that Oseltamivir does not or only weakly bind to the sialoglycan binding pocket. To learn more about the binding site of Zanamivir and Oseltamivir, the molecules were titrated to [U-^{15}N] or [U-^{2}H,^{15}N] labeled RBD proteins. The absence of CSPs shows that there is no significant binding to the RBD at mM concentrations (Fig. 5.6.6b and c). To further characterize the binding of Oseltamivir, ^{1}H spectra of Oseltamivir with and without spike protein were recorded (Fig. 5.6.6d). Strong line broadening effects of Oseltamivir signals were observed in the presence of the spike protein, indicating exchange between the free and the bound state. To highlight this striking finding for Oseltamivir, Neu5Ac, another spike ligand (see below), was added to the sample. No significant line broadening was observed. Remarkably, the addition of Lifitegrast partially reversed the line broadening of Oseltamivir signals, indicating a cross-talk between binding of Lifitegrast and Oseltamivir. This is also reflected in a decrease of Oseltamivir STD-Afs following the addition of Lifitegrast (Fig. 5.6.6e). Since Oseltamivir does not share the RBD binding site of Lifitegrast, this must mean that the two binding sites are allosterically connected or that there is another Lifitegrast binding site that is different from the one located at the RBD but similar to that of Oseltamivir.

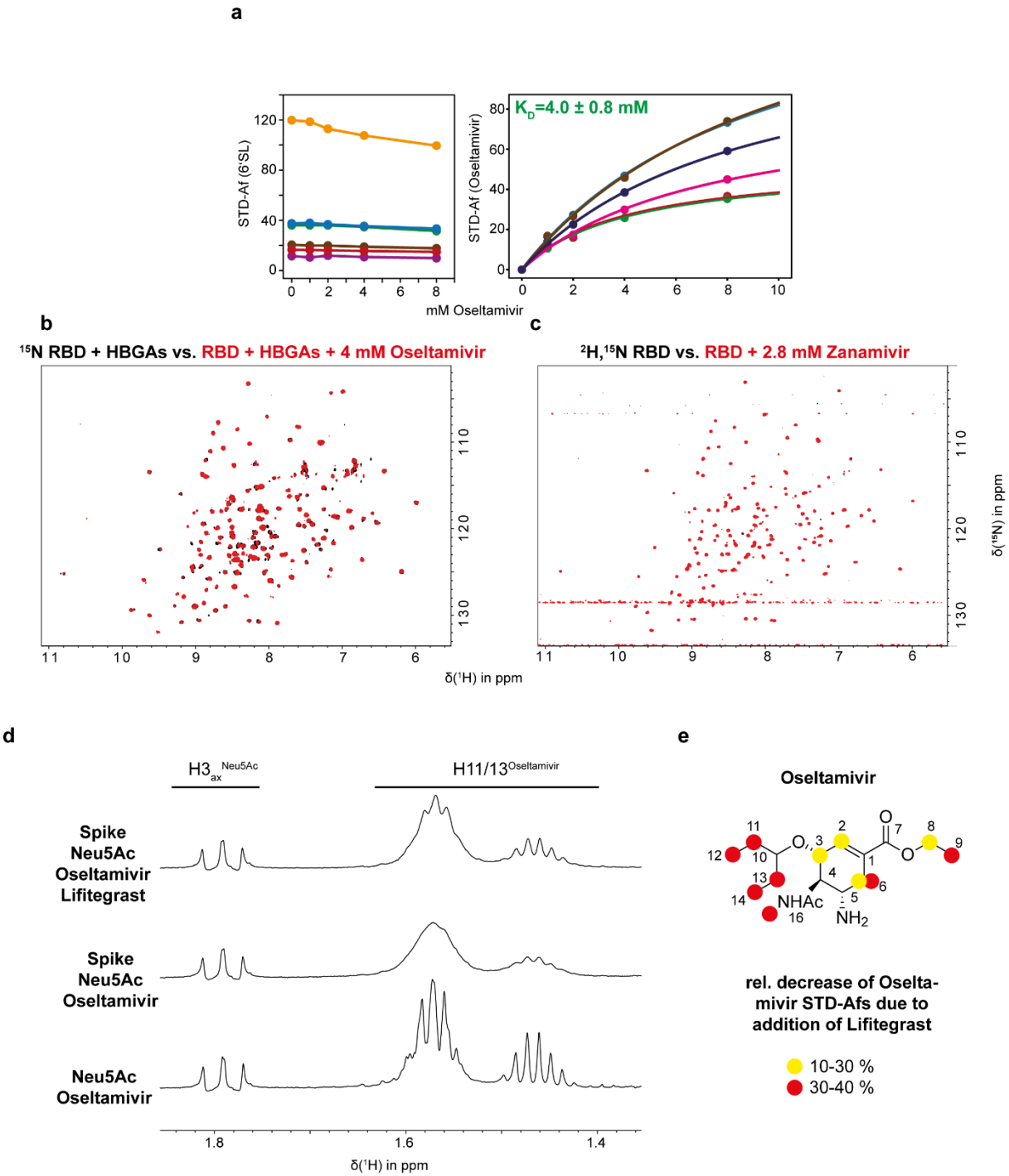


Fig. 5.6.6: Oseltamivir does not compete with sialoglycan binding but with Lifitegrast binding to Wuhan SARS-CoV-2 spike protein. (a) shows no or weak decrease of 6'SL STD-Af due to the addition of Oseltamivir. In addition, the dissociation constant of Oseltamivir in the presence of 30 mM 6'SL is not significantly altered as compared to the one obtained in absence of 6'SL (c.f. Fig. 5.6.5; the individual fits are provided in Tab. S 6.1). Binding isotherms are color-coded according to the respective labels in the STD spectra (c.f. Fig. 5.6.5 and Fig. 5.6.2). (b and c) Titrating 4 mM Oseltamivir and 2.8 mM Zanamivir to [U-^{15}N] or [U-^{2}H,^{15}N] labeled RBD did not result in CSPs in [^{1}H,^{15}N] TROSY HMQC experiments indicating no binding to this part of the spike protein. The sample in (b) was used before to test binding of human ABO(H) blood group antigens (HBGAs) to RBD (see Fig. 5.6.9). (d) 1D ^{1}H spectra clearly show line broadening of Oseltamivir signals due to the addition of the spike protein. These effects are not observed for other ligands such as Neu5Ac. Line broadening effects are reversible by adding 1 mM Lifitegrast, a ligand binding to the RBD. The competition is also visible in the reduction of Oseltamivir STD-Afs measured with the same samples (e). Spectra were acquired using 29 µM spike glycoprotein in (d), 60 µM

RBD in (c), 107 µM RBD in (b), and 2mM Neu5Ac and 2 mM Oseltamivir in (d) on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K for spike spectra or 298K for RBD spectra. Oseltamivir titration shown in (a) was carried out in the presence of 30 mM 6'SL at 279 K on a Bruker 600 MHz Avance III HD spectrometer with room temperature probe using 21 µM spike protein. On-resonance frequencies were set to -4 ppm. Water suppression for 1D spectra was carried out using excitation sculpting. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/). The data shown in (d) and (e) were added to the figure.

A recent AFM study describes that SARS-CoV-2 has an increased specificity for 9-O-acetylated sialic acids (Petitjean et al., 2022). The study reports a dissociation constant of 5 µM. To investigate a possible role of 9-O-acetylated Neu5Ac, STD spectra of the spike protein were acquired in the presence and absence of Neu5Ac and 9-O-Ac-Neu5Ac (Fig. 5.6.7). There is a weak but significant saturation transfer for both ligands, as can be seen from the comparison of STD spectra of the spike protein with STD spectra of the spike protein in the presence of the ligands (Fig. 5.6.7a, b and c). Further analysis of the acetyl signals of Neu5Ac (Fig. 5.6.7 b) shows stronger saturation transfer for the alpha form than for the beta form relative to the respective resonances in the off-resonance spectrum.

The acetyl region of 9-O-Ac-Neu5Ac shows seven signals with typical acetyl group chemical shifts (Fig. 5.6.7c). Three of them can be assigned to the 9-O-acetyl group of the beta-form and the 5-N-acetyl group of the alpha- and beta-form. The signal next to the 9-O-acetyl group of the beta-form could belong to the 9-O-acetyl group of the alpha-form. The presence of other peaks confirms impurities of the commercially purchased product with Neu5Ac (labeled with " * ") and acetate.

Epitope mapping (Fig. 5.6.7 d) was compromised by the strong protein background and must be interpreted with caution. Nevertheless, the epitope mapping of Neu5Ac again underlines the importance of the 5N-acetyl group and is consistent with binding epitopes obtained from 2D STD spectra of ^{13}C-labeled sialoglycans (Unione et al., 2022). Epitope mapping of 9-O-Ac-Neu5Ac was further impeded by the fact that most 9-O-Ac-Neu5Ac and Neu5Ac protons give similar chemical shifts. Only a subset of the signals can be used for ligand epitope mapping (without " * " label in Fig. 5.6.7c). One of these signals is that of the 9-O-Acetyl group. The saturation transfer is rather weak. This hints at a subordinate role of the 9-O-Acetyl group in the binding reaction. It is therefore difficult to imagine that the presence of the 9-O-Acetyl group would shift the dissociation constant from the mM range (Fig. 5.6.2) to the low µM range.

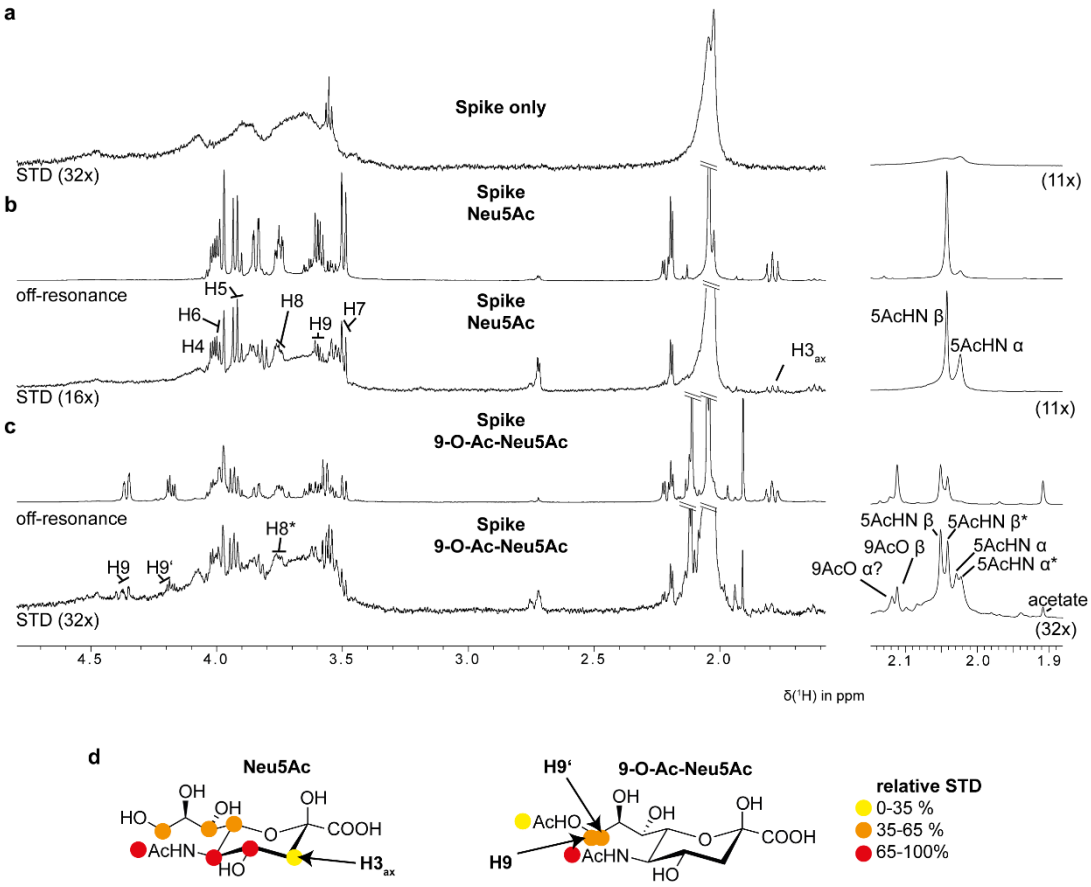


Fig. 5.6.7: Binding of Neu5Ac and 9-O-Ac-Neu5Ac to Wuhan SARS-CoV-2 spike protein. STD spectra of the spike protein (a), spike protein with Neu5Ac (b), and spike protein with 9-O-Ac-Neu5Ac (c). (b) Comparing 5AcHN proton STDs of Neu5Ac of the alpha and beta forms shows a much stronger relative saturation transfer for the alpha form. Note that in solution, the beta form is the dominant form. (c) Acetyl peaks of 9-O-Ac-Neu5Ac indicate saturation transfer to the 9-O and the 5-N-acetyl groups. However, 9-O-Ac-Neu5Ac is contaminated with Neu5Ac and acetate, impeding ligand epitope mapping. Signals solely corresponding to the Neu5Ac impurity are indicated with a " * ". Signals solely corresponding to 9-O-Ac-Neu5Ac are indicated by labels without a " * ". From peak intensities, the ratio is roughly 2:3 Neu5Ac:9-O-Ac-Neu5Ac. Free acetate was identified by chemical shift comparison with data banks. (d) shows crude epitope mappings of Neu5Ac and 9-O-Ac-Neu5Ac. As of the Neu5Ac impurity in the 9-O-Ac-Neu5Ac sample, epitope mapping was only performed for signals with sufficient isolation as indicated in (b) and (c). Obviously, the 9-O acetyl group only experiences relatively weak saturation transfer, indicating that it does not play a key role in the binding event. Spectra were acquired using 27 µM, 29 µM, and 27 µM spike glycoprotein (top to bottom) and 2 mM of Neu5Ac or 9-O-Ac-Neu5Ac on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K. On-resonance frequencies were set to -4 ppm. See Figs. S 6.1 and S 6.2 for negative controls. Assignments were obtained from the literature (Haverkamp et al., 1982).

The results presented here point to weak interactions between sialoglycans and spike protein. Interestingly, the binding site identified by the cryo- electron microscopy structural model of the spike in complex with a Neu5Ac derivate coincides with a mutational hotspot in several SARS-CoV-2 variants of concern (Harvey et al., 2021; McCallum et al., 2021a, 2021b; Buchanan et al., 2022; Coronavirus Resistance Database, Tzou et al., 2022). In particular, the Omicron BA1 subvariant is strongly affected at this site, as illustrated in Fig. 5.6.8a. Comparing the STD spectra of samples with the Wuhan spike

protein (Fig. 5.6.8b; referred to as “spike protein” or “Wuhan SARS-CoV-2 spike protein” before) and the Omicron BA1 spike protein (Fig. 5.6.8b c), there is almost no saturation transfer from the Omicron BA1 spike protein to 3’SL. However, the addition of Oseltamivir to the Omicron BA1 sample shows that the binding competence for Oseltamivir is still intact, which serves as a positive control (Fig. 5.6.8d). It is therefore quite possible that the mutations around the binding site interfere with the sialoglycan interaction.

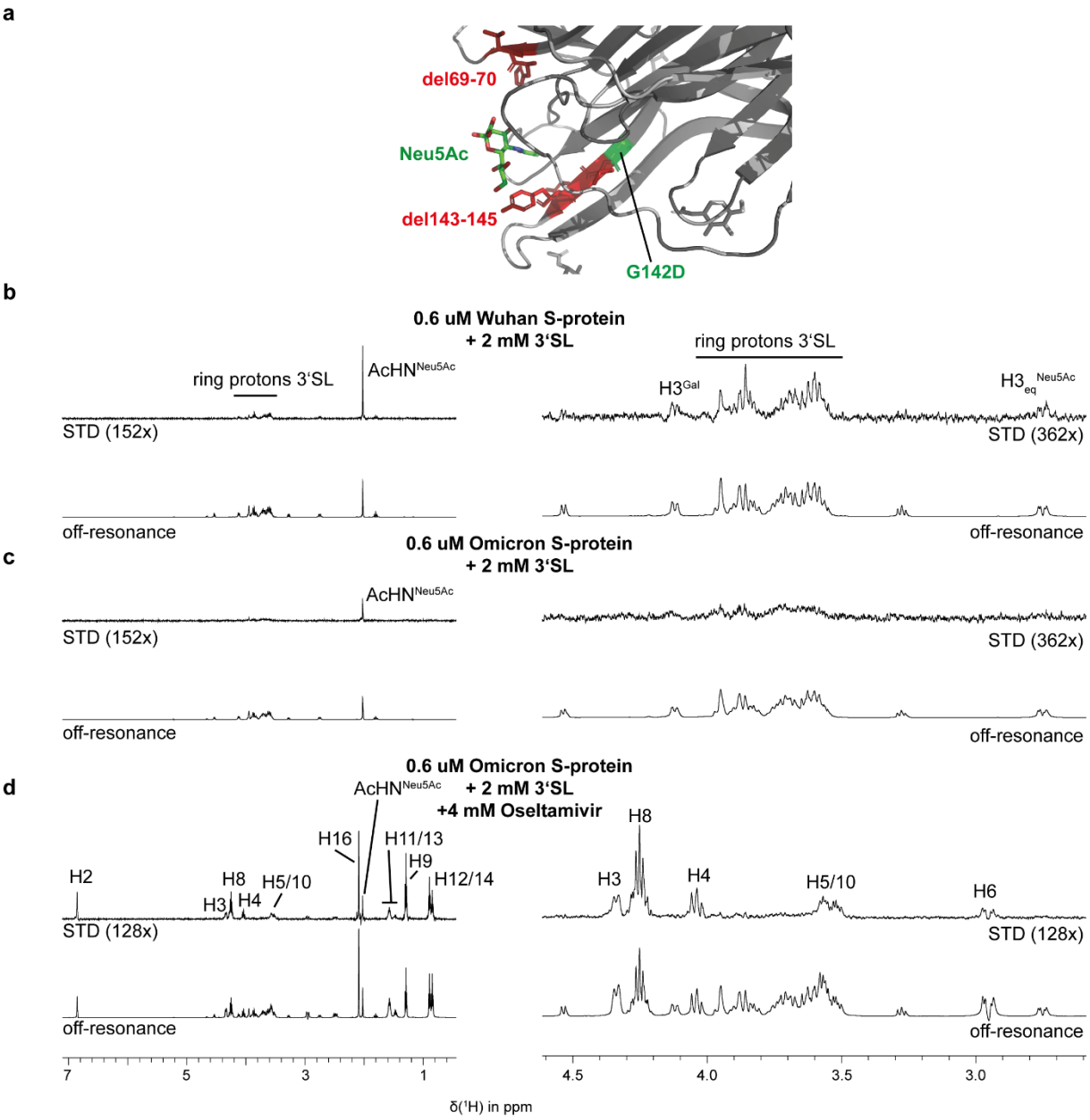


Fig. 5.6.8: SARS-CoV-2 Omicron BA1 and Wuhan spike protein differ in sialoglycan recognition. (a) shows Omicron BA1 mutations close to the sialic acid binding pocket compared to the Wuhan spike protein (pdb 7qur; Buchanan et al., 2022). Information on which restudies are mutated was obtained from the Coronavirus Resistance Database (Tzou et al., 2022). Deletions are highlighted in red. Substitutions are highlighted in green. STD NMR spectra with Wuhan (b) and Omicron BA1 spike protein (c) prove a loss of 3'SL binding to the Omicron BA1 spike. Oseltamivir, which does not bind to the sialic acid binding pocket, was used as a positive control (d). Spectra were acquired using on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K. On-resonance frequencies were set to -4 ppm. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

Next to sialoglycan binding, a potential role of human blood groups in susceptibility to Covid-19 is discussed in the literature (Wu et al., 2021; Goel et al., 2021; Nguyen et al., 2022). STD and [^{1}H,^{15}N] TROSY HSQC experiments with the Wuhan SARS-CoV-2 spike protein or the [U-^{15}N] labeled RBD were performed to gain deeper insights into HBGA binding.

The absence of clear CSPs of RBD resonances in [^{1}H,^{15}N] TROSY HSQC spectra in the presence of HBGAs unambiguously shows that there is no significant binding to the RBD (Fig. 5.6.9). Marginal signs of saturation transfer were observed in STD spectra with the spike protein and HBGAs under conditions where obvious saturation transfer effects were visible for the weakly binding sialoglycans (Fig. 5.6.10, c.f. Fig. 5.6.2). This indicates that there is no specific interaction between the spike protein and HBGAs.

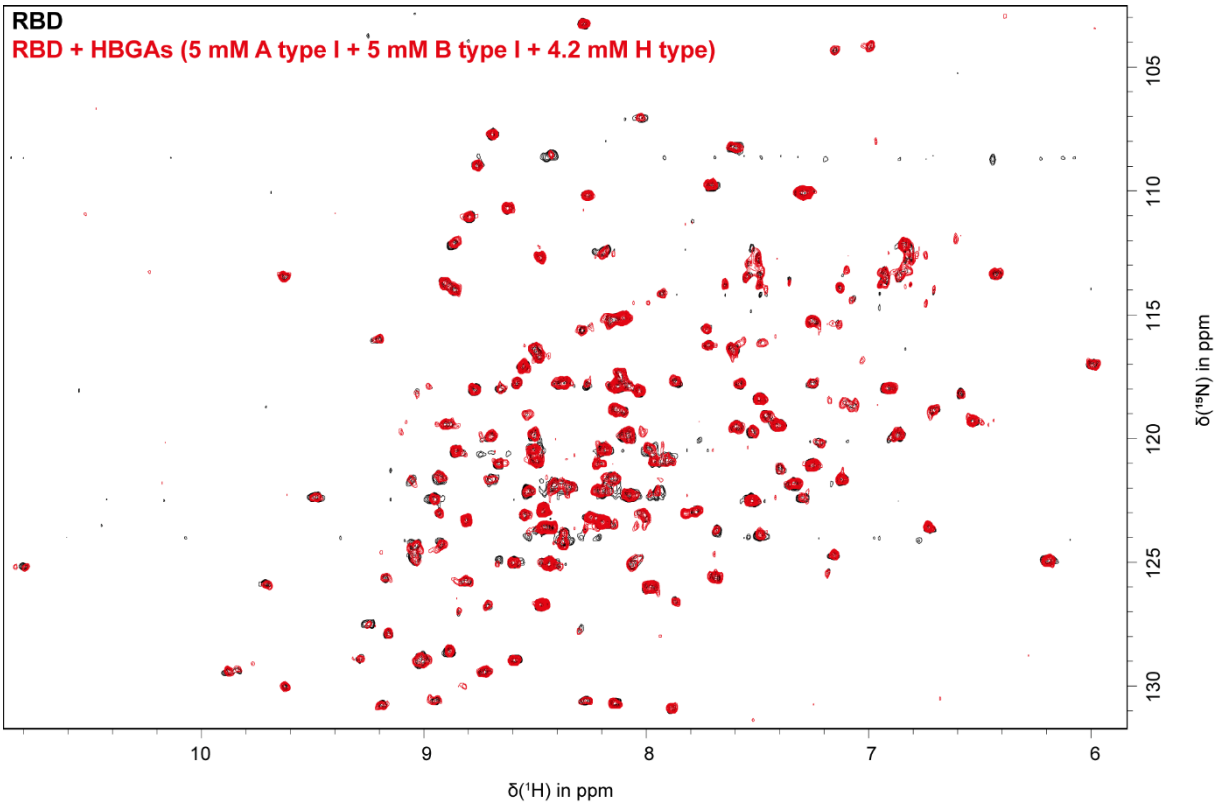


Fig. 5.6.9: The RBD does not bind human ABO(H) blood group antigens (HBGAs). The absence of significant CSPs proves that there is no binding of HBGAs to the RBD. Spectra were acquired using 107 µM [U-^{15}N] labeled RBD on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K. A type I: blood group A tetrasaccharide type I; B type I: blood group B tetrasaccharide type I; H type: blood group H disaccharide. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

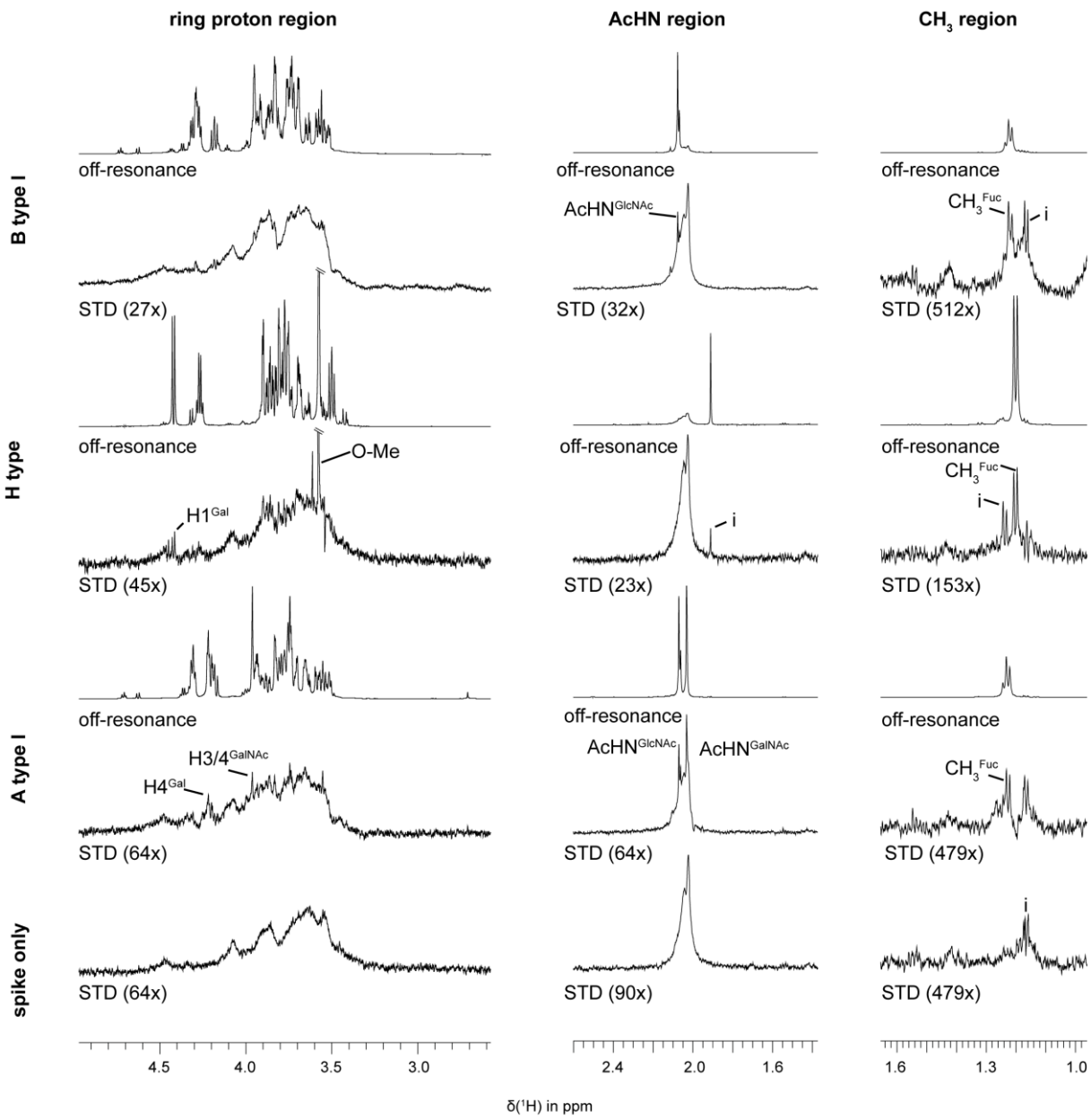


Fig. 5.6.10: Only weak saturation transfer to HBGAs indicate no specific interaction between human ABO(H) blood group antigens (HBGAs) and Wuhan SARS-CoV-2 spike protein. Spectra were acquired using 42 µM, 13 µM, 27 µM, and 27 µM spike glycoprotein (top to bottom) and 2 mM of the respective HBGA on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 279 K. On-resonance frequencies were set to -4 ppm. Negative controls are shown in Fig. S 6.4. A type I: blood group A tetrasaccharide type I; B type I: blood group B tetrasaccharide type I; H type: blood group H disaccharide. Assignments are based on the literature (Rademacher et al., 2008; Fiege et al., 2012). The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

6 Discussion

6.1 AMIGO

AMIGO utilizes the idea of *graph building blocks* derived here and creates structure and NOE-based graphs to solve the assignment problem within reasonable computation times. It is an implementation of the methyl walk method. AMIGO's accuracy is protein-dependent but sufficient, especially when pseudo contact shifts (PCSs) are included in the calculations (Fig. 5.1.11). It should be noted that the aim is not to provide a 100 % correct solution but to suggest methyl walks that the user can follow (cf. Fig. S 1.5). AMIGO helped to assign LmUGP, carrying almost 200 methyl groups with distinct resonances in [^{1}H,^{13}C] HMQC spectra (Mühlberg et al., 2022), and the MNV P-domain with about 100 methyl group resonances (c.f. chapter 5.2).

Other algorithms provide a "stand-alone" approach for solving the assignment problem with very high, but often not 100 % accuracy (Pritišanac et al., 2020). In the original publications, MAP-XSII and Flamengo were tested with 2 data sets (Chao et al., 2012, 2014; Xu and Matthews, 2013), MAGIC with 9 (Monneau et al., 2017), MethylFLYA with 5 (Pritišanac et al., 2019), and MAGMA with 8 data sets (Pritišanac et al., 2017). The relatively small number of test sets makes it difficult to assess the quality of the automated assignment routines. It must be assumed that the performance of all algorithms is to some extent protein-specific (Pritišanac et al., 2017, 2019). Moreover, the choice of parameters has shown to be crucial for the outcome (Xu and Matthews, 2013; Pritišanac et al., 2019). Given the complexity of the algorithms, it is not trivial to understand why a specific methyl group was assigned to a particular resonance. In contrast, AMIGO proposes methyl walks and therefore takes a different approach. The aim is to guide the user during the assignment process. In the manual assignment process, researchers compare measured NOE networks and possibly other experimental restraints such as PSCs with theoretical restraints based on a structural model. With the number of methyl groups, these systems become increasingly complex. Identification of methyl walks is most challenging in such systems and requires time-consuming data analysis. AMIGO can process all this data and suggest which theoretical and measured constraints can be combined to achieve assignments. The results are presented to the user in a very comprehensive way and the user can accept or reject the suggestions. AMIGO thus fills a gap in the automated assignment toolbox.

6.2 MNV P-domain [^{1}H,^{13}C] MILVA methyl group NMR assignment

Approximately 70 % MIL^{proS}V^{proS}A (referred to as MILVA) methyl groups of the MNV P-domain in complex with GCDCA were assigned in this study. Three different methods - NOEs, PCSs, and site-directed mutagenesis - were used for this purpose. Assignments based on the different techniques complemented and verified each other, resulting in a highly reliable outcome. With the present assignment, it was possible to study the dynamics, dimerization, and thermodynamics and kinetics of ligand binding of the MNV P-domain.

Interestingly, the assignment also revealed that MNV P-domains synthesized in *E. coli* are present in a mixture of Pro361 cis and trans isomers, with both forms being almost equally populated (Fig. 5.2.3). This feature has also been found for other proteins, but reports of it are relatively rare (Chazin et al., 1989; Roderer et al., 2015; Alderson et al., 2017). The loop carrying Pro361 constitutes the binding epitope for the CD300lf receptor, as can be seen in crystal structure models (pdbs 6e47, 6c6q, and 5or7; Nelson et al., 2018; Kilic et al., 2018). The residues within the loop migrate by up to 10 Å (cf. Fig.

5.2.3). This changes the conformation of the loop. The structural models of the P-domain in complex with CD300lf were exclusively found in the cis form. The affinity to the receptor was measured by isothermal titration calorimetry (ITC) with P-domains from *E. coli.* It is enhanced by metal ions and GCDCA. Nevertheless, according to these measurements, it does not exceed the low micromolar range, which is rather atypical for protein-receptor interactions (Nelson et al., 2018). Different receptor affinities for P-domain Pro361 cis/trans isomers would likely confound the analysis performed by Nelson et al., 2018. Future studies employing NMR spectroscopy with the methyl group labeled P-domain and the CD300lf receptor may provide more accurate values.

Notably, the assignment also confirmed the two metal ion binding sites of the MNV P-domain (Fig. S 2.4) as suggested by crystal structure models (pdb 6e47, 6c6q, Nelson et al., 2018).

6.3 MNV P-domain dimerization and GCDCA binding

MNV P-domain dimerization and GCDCA binding were analyzed applying the P-domain NMR assignment, [^{1}H,^{13}C] HMQC NMR experiments, and 2D line shape analysis with *TITAN*. The employed binding models were implemented in *TITAN* by Dr. Christopher A. Waudby (Institute of Structural and Molecular Biology, University College London) for this purpose, as documented in the corresponding publication (Creutznacher et al., 2022a). The study yielded a detailed understanding of the undelaying thermodynamics and kinetics. In addition, chemical shift perturbations (CSPs) show that GCDCA affects the entire P-domain and reveal subtle changes not visible in cryo-electron microscopy or X-ray crystallography-based structural models.

Recently, the off-rate of huNoV P-domain dimers was investigated using ion-exchange-chromatography (Creutznacher, 2020). 2D line shape analysis of MNV P-domain [^{1}H,^{13}C] HMQC spectra yielded an off-rate $k_{off,Dimerization}$ of 1.2 s^{-1} (Fig. 5.3.2), which differs by about six orders of magnitude from the rate obtained for huNoV P-domains. Creutznacher, 2020, showed that the relatively weak interactions between MNV P-domains can be compensated by GCDCA binding, triggering the formation of P-domain dimers. Interestingly, GCDCA also causes capsid contraction (Sherman et al., 2019), enhanced receptor affinity (Nelson et al., 2018), and escape from neutralizing antibodies (Williams et al., 2021a; Creutznacher et al., 2022a). All these biological and structural implications of GCDCA binding call for a deeper understanding of the interaction. 2D line shape analysis of [^{1}H,^{13}C] P-domain HMQC spectra in the presence of different GCDCA concentrations provided a thermodynamic and kinetic description of the reaction (Fig. 5.3.5). The dissociation constant of 11 µM is in good agreement with the one derived from ITC experiments (Nelson et al., 2018). However, the data presented here show a far more complex picture than simple binding and dissociation. As a matter of fact, the spectra can only be explained assuming at least three different P-domain states (Fig. 5.3.5a). A more complex model that assumes cooperativity even includes five P-domain states, from which three can be directly observed in the respective NMR spectra (Fig. 5.3.5b). Both models lead to a similar Gibb's free energy ΔG of approximately 56 KJ/mol. The five-state model is intuitively more realistic because it involves inter-subunit communication. Inter-subunit communication is expected as the GCDCA binding pocket is formed by both amino acid chains of the P-domain dimer. On the other hand, the fitted parameters are not very well defined. This suggests that the fitting problem is underdetermined. Furthermore, the additional states cannot be directly observed. These points, together with the simpler nature of the three-state model, make the three-state model the preferred model.

Analysis of chemical shift perturbations (CSPs) induced by GCDCA binding strongly suggests that the apo and GCDCA-bound states differ due to structural rearrangements (Fig. 5.3.4). Alanine chemical shifts report for changes in protein backbone (Godoy-Ruiz et al., 2010; Pederson et al., 2017). This is of particular interest considering the antibody binding site, i.e., the E'F' loop. Long-range CSPs observed in this region correlate well with diminished monoclonal antibody (mAb) binding (Williams et al., 2021a; Creutznacher et al., 2022a). Interestingly, the E'F'-loop was found in two different conformations in the presence and absence of GCDCA (Taube et al., 2010; Sherman et al., 2019; c.f. Fig. 3.1.2). The observation of methyl group CSPs as obtained for Ala380, Ala381, and Ala382 located in this loop therefore link these snapshots by directly showing a change of the chemical environment of this loop by titration of GCDCA. The NMR data provided here together with the structural models thus rationalize GCDCA induced immune evasion.

Another site of interest is the receptor binding site. The CSPs of Ala365 and Ile358 show that GCDCA alters the chemical environment of the respective loop. Assuming an underlaying conformational change might therefore explain the slightly enhanced CD300lf receptor affinity in the presence of GCDCA (Nelson et al., 2018).

It should be noted that GCDCA-induced dimerization at the P-domain level (Creutznacher, 2020) may be linked to contraction at the capsid level (Sherman et al., 2019). This point will be discussed in more detail later, as the rather loose connection so far becomes clearer with the discussion of data presented in chapters 5.4 and 5.5.

6.4 Metal ions control MNV P-domain dimerization and GCDCA binding

An important finding is that the P-domain can undergo a "global transition". The term "global transition" should be understood as a more subtle form of transition, that does not necessarily include large scale structural rearrangements (c.f. chapter 5.4.3). Such a global transition can be inferred from CSP patterns and binding kinetics of metal ions binding to the G'H' loop (c.f. chapter 5.4.4 and 5.4.5). The transition can be triggered by metal ions or GCDCA. It is associated with dimer formation as well as CSPs throughout the P-domain.

Binding of metal ions to the MNV P-domain has similar effects on dimer formation as binding of GCDCA. P-domain elution volumes in analytical size exclusion chromatography (SEC) experiments were shifted towards higher molecular weights (Fig. 5.4.2 and Creutznacher, 2020). At the same time, monomer resonances in [^{1}H,^{13}C] HMQC spectra of the P-domain gradually disappear. The two independent methods unambiguously show that metal ions trigger dimerization of the P-domain.

In addition to P-domain dimerization, metal ion and GCDCA-binding also correlate with escape from monoclonal antibodies (Williams et al., 2021a, 2021b; Creutznacher et al., 2022a). As discussed in chapter 6.3, the effects of GCDCA can be rationalized by GCDCA-induced CSPs and structural models. Metal ion-induced CSPs give a similar picture. They clearly show a change in the chemical environment of methyl groups in the E'F' loop, which is the antibody binding epitope (Fig. 5.4.3; Williams et al., 2021a). The CSPs may reflect structural rearrangements and thus explain reduced antibody binding. However, in contrast to the GCDCA-bound P-domain state, there is no structural model for the P-domain in which only metal ions are bound. Such a model could further confirm these conclusions. Yet, the CSPs of the E'F' loop would agree with differences in structural models of the apo P-domain and the P-domain in complex with metal ions and the CD300lf receptor (if the receptor has no additional effect on the structure, c.f. Fig. 3.1.2). Indeed, pseudo contact shifts (PCSs) indicate that the structure of the metal ion-bound P-domain in solution is in very good agreement with structural

models in the presence of metal ions and the CD300lf receptor, with and without GCDCA (pdbs 6c6q and 6e47, Nelson et al., 2018; c.f. Fig. 5.4.13). These structural models are also almost identical to the model in the presence of GCDCA alone (Sherman et al., 2019). Consequently, the data together with the available structural models suggest two P-domain conformations. One conformation is the apo form with the E'F' loop open (Taube et al., 2010). The other conformation can be induced either by GCDCA (Fig. 5.3.4 and Sherman et al. 2019) or by metal ions (Fig. 5.4.3 and Fig. 5.4.13). Here, the E'F' loop is closed. The observed CSPs link the snapshots provided by the structural models and complement the picture.

It has been reported that metal ions enhance the affinity to the CD300lf receptor (Nelson et al., 2018). Metal ions binding to the G'H' loop affect this site allosterically (Fig. 5.4.6), but it is also evident that the metal ion directly binding at the receptor epitope might influence receptor binding. In addition, structural models suggest that there is another metal ion which is complexed by the receptor-P-domain interface (c.f. Fig. 3.1.2). It is therefore not possible to derive solid conclusions about the mechanism behind the enhanced receptor affinity in the presence of metal ions.

Interestingly, as the metal ion bound P-domain and the metal ion and GCDCA bound P-domain appear to be structurally almost identical (see above), a synergetic effect in reduced antibody binding (personal communication, Prof. Dr. Stefan Taube, University of Lübeck) cannot be explained by structural differences. Neither CSPs nor PCSs or structural models point at such differences in the E'F' loop. However, the thermodynamic valley (see below) in which the closed P-domain is located is deeper if GCDCA is additionally bound (c.f. Tab. 5.4). This would lock the P-domain in the closed conformation. The thermodynamic valley can be characterised in part by methyl group order parameters of the P-domain (c.f. Fig. 5.4.14), which are linked directly to the conformational entropy of the protein (Frederick et al., 2007; Mittermaier and Kay, 2009). The increased flexibility in the presence of GCDCA at sites far from the antibody epitope would thus hinder antibody binding. At the very least, this clearly shows that quite complex mechanisms are possible and that three-dimensional structures provide only a very rough picture of the underlying mechanisms.

Finally, a model with five major P-domain states can be deduced (Fig. 6.4.1), i.e. the monomeric state (M), the dimeric state (D), the dimeric metal ion bound state (D'), the dimeric GCDCA bound state (DL), and the dimeric metal ion and GCDCA bound state (DL'):

The D', DL, and DL' states are structurally almost identical (Fig. 5.4.12; Fig. 5.4.13; Nelson et al., 2018; Sherman et al., 2019), but at least the DL' and the D' state differ in their entropic profiles (Fig. 5.4.14). The D' state and the DL' state exhibit increased receptor affinity (Nelson et al., 2018). Based on the structural similarities, it can be speculated that the DL state shares this feature. The three states are also characterized by diminished recognition of monoclonal antibodies (Williams et al., 2021a, 2021b; Creutznacher et al., 2022a; personal communication, Prof. Dr. Stefan Taube, University of Lübeck). These P-domain states are the result of GCDCA or metal ion-induced "global transitions" (c.f. chapter 5.4.4 and 5.4.5). Moreover, they are characterized by tight dimer interactions (Fig. 5.4.2 and Creutznacher, 2020). At the capsid level, the P-domains settle down on the shell (Sherman et al., 2019; Snowden et al., 2020; Song et al., 2020).

On the other hand, the D state can dissociate into monomers in solution (Creutznacher, 2020, and Fig. 5.3.2), has only weak receptor affinity (Nelson et al., 2018) and is recognized by monoclonal antibodies (Williams et al., 2021a; Creutznacher et al., 2022a). The capsid is found in the expanded conformation (Sherman et al., 2019; Song et al., 2020; Williams et al., 2021a).

[^{1}H,^{13}C] HMQC spectra of the P-domain at different ligand concentrations allow a complete thermodynamic and partial kinetic description of the transitions between the states (c.f. chapters 5.3 and 5.4). An important finding is the thermodynamic and kinetic cross-talk between metal ion and GCDCA binding. It is of note, that the order of binding events leads to almost identical Gibb's free energy balances (Fig. 6.4.1). This is expected from the laws of thermodynamics but has rarely been demonstrated for protein ligand systems at this level of complexity. The robustness of these results is underscored by the fact that the dissociation constants associated with individual binding steps have been obtained using different experimental techniques.

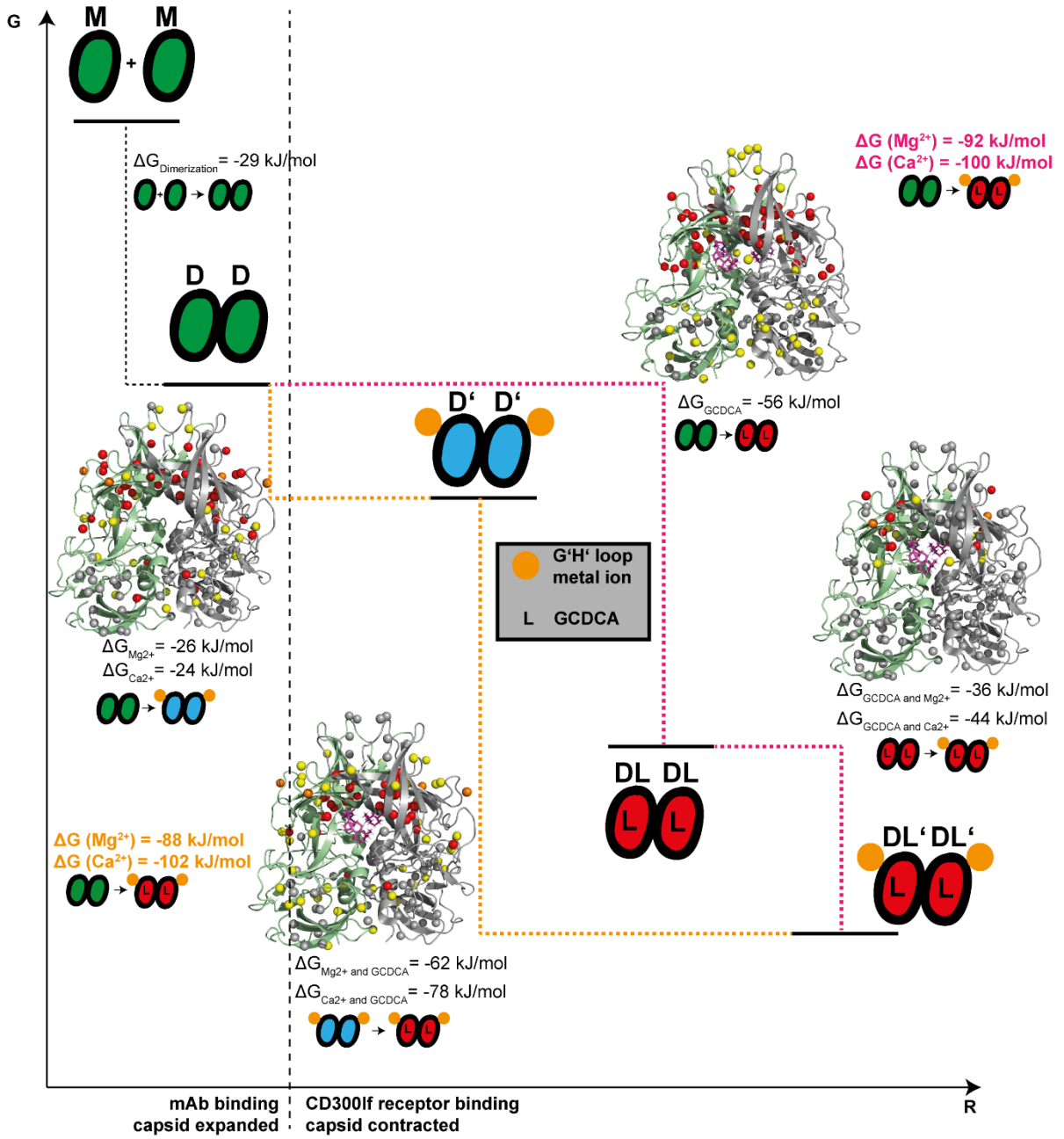


Fig. 6.4.1: A model for P-domain reacting to host factors containing five major states. M is the monomeric P-domain, D the dimeric P-domain, D’ the metal ion bound P-domain, DL the GCDCA bound P-domain, and DL’ the metal ion and GCDCA bound P-domain. Dotted lines illustrate the transitions between the major states. Corresponding CSP patterns are plotted on the structural models using the color code explained in Fig. 5.3.4 and resemble the data in the respective chapters of the main part of this thesis. If not stated otherwise, values for ΔG are calculated from the here derived dissociation constants for the WT (wildtype) P-domain. The ΔG value for Ca^{2+} binding to the GCDCA bound P-domain (DL -> DL’) was calculated from the dissociation constant derived by Creutznacher et al., 2021. References on mAb binding, CD300lf binding and capsid conformations are given in the text. The CSP plot of the transition from the DL to the DL’ state is generated from data of the D410A P-domain, to underline the local character of G’H’ loop metal ion binding to the GCDCA bound state. For the illustrations, a structural model by Nelson et al., 2018, was used (pdb 6e47).

6.5 Physiologically relevant pH ranges add another dimension to P-domain dimerization and metal ion binding

[^{1}H,^{13}C] HMQC peak intensities, SEC elution volumes, and 2D line shape analysis demonstrate a clear pH dependence of P-domain dimer formation (Fig. 5.5.1). The dissociation constant at a pH of 5.3 is almost one order of magnitude smaller than the one derived at pH 7.4. At pH values below 5, the monomer species is not sufficiently detectable in [^{1}H,^{13}C] HMQC experiments. Therefore, the exact dissociation constant cannot be derived using the 2D line shape analysis approach. Nevertheless, the fact that the P-domain monomer state is hardly populated in this pH range indicates even smaller dissociation constants. Interestingly, the G'H' loop is protonated between a pH of 4 and 6 (Fig. 5.5.5), which parallels with P-domain dimerization. Reconsidering the involvement of the G'H' loop in metal ion and GCDCA binding, it is tempting to assume that protonation of the G'H' loop and dimerization are linked. However, it should be noted that this link is not clear-cut as protonation happens simultaneously at many sites of the P-domain and not just at the G'H' loop.

The correlations between P-domain dimerization and GCDCA binding, G'H' loop metal ion binding, and G'H' loop protonation are obvious and match a model, where the G'H' loop constitutes the key platform for host factors to control P-domain dimerization (Creutznacher, 2020; Fig. 5.4.2; Fig. 5.5.1). In the capsid context, cryo-electron microscopy studies have revealed that these factors cause P-domains to contract and rest on the capsid shell (Sherman et al., 2019; Song et al., 2020; Snowden et al., 2020; Williams et al., 2021b). NMR-based P-domain experiments provide detailed insights into the molecular mechanisms of host factor P-domain interactions. In the absence of the shell domain, the P-domain alone is an artificial system. On the other hand, cryo-electron microscopy-derived structural models encompass the entire capsid but only provide snapshots of the mechanisms behind the observed changes. Obviously, NMR spectroscopy and cryo- electron microscopy complement and verify each other very well. It is difficult to prove that dimerization and capsid contraction are linked. However, it is tempting to speculate that dimerization and capsid contraction lean on the same driving force. P-domains are not well resolved in cryo- electron microscopy structures with the non-contracted capsid, indicating that there are no defined P-domain positions relative to the capsid shell. This observation can be explained by the potentially flexible linkers (Fig. 3.1.1) but would be further supported by the presence of P-domain monomers or partially dissociated dimers. In addition, host factor-induced dimerization would certainly reduce the P-domain's translational degrees of freedom. This would entropically favor the contraction of the capsid.

Another factor triggering capsid contraction could be further host factor-induced changes of the P-domain dimer itself. GCDCA and metal ion binding were accompanied by long-range CSPs consistent with differences in structural models of the P-domain (see above). In addition to methyl groups at the E'F' loop, methyl groups at the lower part of the P-domain were affected (Fig. 5.3.4; Fig. 5.4.3). Furthermore, methyl group order parameters hint at altered dynamics at these parts of the P-domain induced by GCDCA (Fig. 5.4.14; Fig. 5.5.7). Although it is difficult to directly interpret these alterations, the effects leading to these observations could contribute to the contraction of the capsid.

Metal ion binding shows a clear dependence on the pH (Fig. 5.5.2) as does the dimerization process. A pH of 4.0 is the upper pH limit in the gut of mice (McConnell et al., 2008). At this pH, significant metal ion binding to the G'H' loop can be ruled out by CSP experiments. At a pH of 5.3, which reflects the pH in the host's intestine, the dissociation constants of metal ion binding to the G'H' loop range from 1.7 mM to 7.8 mM for Ca^{2+} ions (c.f. chapter 5.4.5 and Tab. 5.7). At a pH of 7.4, which resembles the intracellular pH (Yoshida et al., 1993) and the pH in blood (Wolfe, 1961), a dissociation constant of 1

mM was derived. As explained before, the G'H' loop is protonated between a pH of 4 and 6. Apparently, metal ion binding competes with protonation of the G'H' loop. Since both factors are associated with dimerization, it appears that MNV has evolved different mechanisms that trigger dimer formation in different environments.

Interestingly, at pH 7.4, monomer, dimer, and metal ion bound dimer resonances were sufficiently resolved allowing 2D line shape analysis. The derived on-rate of $2.8*10^4$ $M^{-1}s^{-1}$ agrees with the estimate of the on-rate at a pH of 5.3 for Mg^{2+} ($k_{on,Mg2+} < 8*10^3$ $M^{-1}s^{-1}$, c.f. chapter 5.4.5). Such on-rates are clearly not controlled by mere diffusion and are evidence for a "global transition" (c.f. chapter 5.4.3) along with metal ion binding.

pH-induced dimerization enabled determination of order parameters in the presence and absence of GCDCA (Fig. 5.5.7). As observed for the effect of GCDCA on the metal ion-bound P-domain at a pH of 5.3 (c.f. Fig. 5.4.14), GCDCA redistributes the entropic profile of the P-domain. The order parameters reflect changes in pico- to nanosecond dynamics. Since it is not trivial to obtain such dynamics, it is difficult to put the results in context with the norovirus literature. However, for other proteins, such as calmodulin or the Zn efflux repressor CzrA, redistributions of entropy are associated with underlying allosteric networks (Lee et al., 2000; Capdevila et al., 2017). The idea of allosteric network-like communication between methyl groups within the P-domain is also suggested by methyl group chemical shifts, which move along similar vectors due to addition of different ligands (Fig. 5.4.15). The corresponding methyl groups are distributed throughout the protein. Such a network could transmit the effects of ligands to distant sites of the protein. This would explain the long-range CSPs and order parameter alterations of GCDCA or metal ion binding.

6.6 Glycan recognition of SARS-CoV-2 spike proteins

NMR spectroscopy was used to scrutinize glycan binding to the SARS-CoV-2 spike protein. Whereas sialoglycans bind with mM affinity, ABO(H) human blood group antigens (HBGAs) do not bind. The results are summarized in Fig. 6.6.1. The qualitative evidence for the direct interaction with sialoglycans (data not shown, see Creutznacher et al., 2022b, Fig. S12) was obtained in parallel with two other STD NMR studies (Unione et al., 2022; Buchanan et al., 2022). Subsequently, the interactions with sialoglycans were quantitatively characterized as shown in chapter 5.6.2.

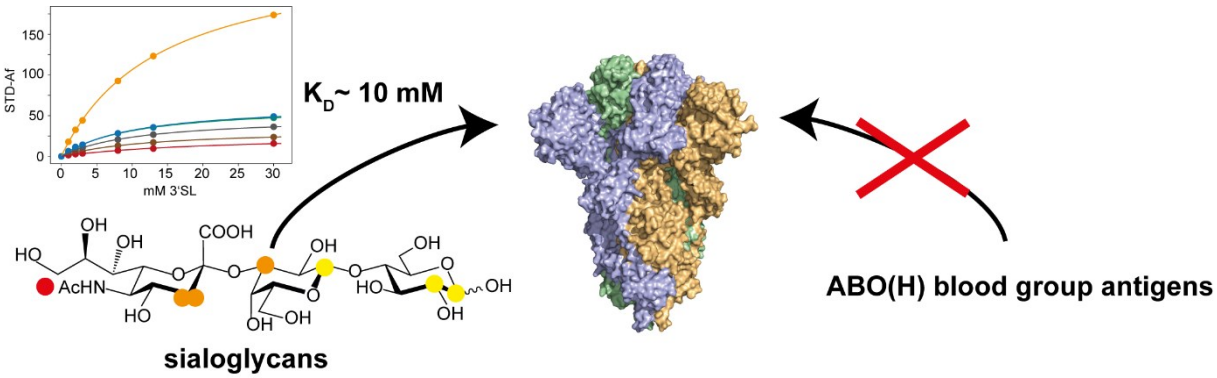


Fig. 6.6.1: Sialoglycans but not ABO(H) blood group antigens (HBGAs) interact with the SARS-CoV-2 spike protein. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

STD NMR yielded dissociation constants for the interaction between sialoglycans and the SARS-CoV-2 Wuhan spike protein. Ligand epitope mappings of 6'SL, 3'SL and Sialyl-Lewis^X show the importance of the sialic acid moiety, with the acetyl group being most pronounced. The epitope mappings agree with another STD NMR study employing ^{13}C labeled sialoglycans (Unione et al., 2022). The data fill an important gap in the biophysical characterization of SARS-CoV-2 spike-sialoglycan interactions.

The STD NMR method differs from the one used by Unione et al., 2022. In their approach, they synthesized 3'SLN and 6'SLN. These compounds differ from 3'SL and 6'SL by having an additional 2N-acetyl group on the glucose unit. By introducing ^{13}C nuclei in the Neu5Ac and galactose units 2D STD $^{1}H,^{13}C$ HSQC spectra could be acquired. This elegant approach enabled ligand epitope mapping with resolution for the ring atoms. This was not possible with the 1D STD approach presented here. However, the 3'SLN and 6'SLN used lack ^{13}C nuclei at the Neu5Ac acetyl group and the entire GlcNAc unit. From the 1D STD-derived ligand epitope mappings it becomes clear that the most pronounced STD effects at the Neu5Ac acetyl group were overlooked this way. Additionally, contributions from the glucose units were detected. Employing 1D STD methods does not require complex carbohydrate synthesis; therefore, it was also possible to test sialyl-Lewis^X, complementing the picture.

For interactions between RBD with sialoglycans and HBGAs, Nguyen et al., 2022, report dissociation constants between about 200 and 500 µM. The fact that the RBD does not bind sialoglycans has already been demonstrated by others (Unione et al., 2022; Creutznacher et al., 2022b). [$^{1}H,^{15}N$] TROSY HSQC experiments unambiguously show that this is also the case for HBGAs (Fig. 5.6.9). Specific interactions of HBGAs with other parts of the Wuhan SARS-CoV-2 spike also seem unlikely, as can be concluded from STD NMR experiments (Fig. 5.6.10). The hypothesis that HBGAs act as attachment factors (Wu et al., 2021) can therefore be rejected.

Another STD NMR study reported a dissociation constant of 32 µM, an off-rate of 0.2 s^{-1}, and an on-rate of 6300 $M^{-1}s^{-1}$ range for 3'SL (Buchanan et al., 2022). The study employed a new method for deriving dissociation constants from STD NMR experiments. The values obtained from this method agree well with values derived from SPR experiments used to validate the new method. However, in the SPR setup, the ligand instead of the protein was immobilized on the chip. In the past, this SPR setup has been shown to underestimate dissociation constants by several orders of magnitude when comparing them to NMR-derived dissociation constants (Sauter et al., 1989; Hanson et al., 1992). The discrepancy can be explained by rebinding and multivalency (Nieba et al., 1996). Without clean validation of the new method, the values derived by Buchanan et al., 2022, must be questioned critically. The obtained off-rate of 0.2 s^{-1} is a further inconsistency. For another protein ligand system, it was shown that such slow off-rates would lead to the absence of STD effects (Pérez-Victoria et al., 2009).

In contrast, the study presented here yielded STD-derived dissociation constants in the low mM regime. STD-Afs were calculated from a single saturation time. This method has its own hurdles and requires some considerations (Angulo et al., 2010). The choice of the saturation time and the magnitude of the saturation transfer can slightly affect the resulting dissociation constant. A larger effect can arise from the protein concentration used. At low ligand-to-protein ratios, rebinding can artificially increase the resulting dissociation constants. However, in the present case with low µM protein concentrations and mM ligand concentrations, these ratios are relatively high. Rebinding of individual ligand molecules therefore should become neglectable. In fact, the values for a protein-ligand example with a similar experimental setup and dissociation constant were almost identical to

those obtained from fast exchange CSPs (Creutznacher et al., 2020). Based on the cryo-electron microscopy structure model, a dissociation constant in the mM range would match well with other protein-sialic acid complexes for which dissociation constants and structural models exist (Tab. 6.1). Surprisingly, the number of contacts between protein and ligand would suggest even higher dissociation constants.

Tab. 6.1: Comparison of structural and thermodynamic data of sialic acid binding proteins. The literature review was carried out by Prof. Dr. Thomas Peters. Analysis of structural models was carried out by Dr. Martin Frank. The table is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

PDB	Protein	Ligand	Method	K_D in mM	Reference	Contacts (all/H-bonds/lipophilic)
1hgg	Influenza Hemagglutinin	3'SL α-D-Neu5Ac-OMe	NMR Ligand-CSP	3.2 2.8	Sauter et al., 1992	58/7/10
1kqr	Rotavirus VP8*	α-D-Neu5Ac-OMe, 3'SL, 6'SL	$[^1H,^{15}N]$ HSQC CSP	1.2	Dormitzer et al.; 2002a Dormitzer et al.; 2002b	51/5/12
3nxd	JC Polyoma VP1	NeuNAc-α2,6-Gal-β1,4-GlcNAc-β1,3-Gal-β1,4-Glc (LSTc) GD1a GM1 (3'SL)	X-ray; electron densities at different soaking concentrations	0.5 6 1.2	Ströh et al., 2015	59/4/12
1uxa	Adenovirus fiber knot Ad37	3'SL	X-ray; electron densities at different soaking concentrations	5	Burmeister et al., 2015	51/3/11
1uxb	Adenovirus fiber knot Ad19p	3'SL	X-ray; electron densities at different soaking concentrations	7	Burmeister et al., 2015	42/3/11
7qur	SARS-CoV-2	3'SL (derivative)	Cryo-electron microscopy	?	Buchanan et al., 2022	28/3/9

Experiments with the Omicron BA1 spike protein show that sialoglycan binding is a tuneable feature of SARS-CoV-2. This can be inferred from the substantial differences in saturation transfer to 3'SL from the Omicron BA1 spike protein and the Wuhan spike protein (Fig. 5.6.8). Interestingly, many amino acids around the sialoglycan binding site are deleted or substituted in Omicron BA1 spike proteins, as illustrated in Fig. 5.6.8a. Initially, these mutations were only linked to immune evasion, as the (back than unknown) sialoglycan binding site is also an important epitope for neutralizing antibodies (Harvey et al., 2021; McCallum et al., 2021a, 2021b). Now, these mutations may also be associated with altered sialoglycan binding. It seems as if SARS-CoV-2 can afford to switch off sialoglycan binding when evolutionary pressure from the host immune system becomes too great. One theory of sialoglycan utilization is a role as attachment factor (Sun, 2021). This theory comes from the fact that ACE2, the protein receptor, has only limited expression of ACE2 in respiratory tissues (Harmer et al, 2002; Lukassen et al, 2020; Hikmet et al, 2020). However, it should be noted that virological experiments do not reach clear conclusions about a possible role of sialic acids in infection (Chu et al., 2021; Saso et al., 2022). Whether this theory holds true is therefore questionable and only time will tell.

6.7 Outlook

In this thesis, the interactions of the MNV P-domain protein and the SARS-CoV-2 spike protein with host factors have been extensively studied using NMR spectroscopy. The results offer unique insights into complex mechanisms and characterize the thermodynamics and kinetics of the corresponding binding reactions.

It has been demonstrated that the dimerization of the MNV P-domain is triggered by the host factors GCDCA, metal ions, and acidic pH. However, it is likely that the individual host factors affect the formation of dimers to varying degree. Due to the low protein concentrations that would be required to detect the monomeric species, it was not possible to investigate thermodynamics and kinetics of dimerization in the presence of the above host factors using NMR spectroscopy. A biophysical method that is far more sensitive is Förster resonance energy transfer (FRET) fluorescence spectroscopy. First, it could be used to learn more about dimer stability in the presence of the above host factors at the P-domain level. It may be possible to determine the corresponding off-rates and dissociation constants of the dimerization. Second, FRET studies with the entire capsid could yield insights into the function of dimerization in the capsid context. The dynamics of both, contraction/expansion and the dimerization/monomerization, could be assessed. This would also reveal insights into capsid stability and thus into possible mechanisms of RNA release. Of note, the release of RNA is a step in norovirus infections that could be targeted for treatment. Such approaches already exist for other RNA viruses (e.g., Wald et al., 2019). Furthermore, [^{1}H,^{13}C] NMR experiments with the P-domain, antibodies and the CD300lf receptor may provide even deeper insights in how host-factors can control immune evasion and receptor binding.

The role of sialoglycans acting as host-factors for SARS-CoV-2 infection is still controversial. A systematic STD NMR study of glycan binding to SARS-CoV-2 spike proteins of different variants of concern or with single mutations could provide a deeper understanding. Such studies could also include other corona viruses such as MERS or OC43 and ancestral strains infecting bats. This may reveal the evolutionary development of sialoglycan binding and could provide unique insights in possible zoonotic mechanisms. This knowledge would be of great importance for the prevention of future pandemics. In addition, the further development of the molecules Lifitegrast and Oseltamivir using medicinal chemistry could lead to new anti-SARS-CoV-2 or even in pan-coronavirus drugs.

Bibliography

- Adler JL, Zickl R. Winter vomiting disease. J Infect Dis. 1969 Jun;119(6):668-73. doi: 10.1093/infdis/119.6.668. PMID: 5795109.
- Alderson TR, Benesch JLP, Baldwin AJ. Proline isomerization in the C-terminal region of HSP27. Cell Stress Chaperones. 2017 Jul;22(4):639-651. doi: 10.1007/s12192-017-0791-z. Epub 2017 May 25. PMID: 28547731; PMCID: PMC5465039.
- Amero C, Asunción Durá M, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin MJ, Vernet T, Franzetti B, Boisbouvier J. A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J Biomol NMR. 2011 Jul;50(3):229-36. doi: 10.1007/s10858-011-9513-5. Epub 2011 May 29. PMID: 21626214.
- Angulo J, Enríquez-Navas PM, Nieto PM. Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. Chemistry. 2010 Jul 12;16(26):7803-12. doi: 10.1002/chem.200903528. PMID: 20496354.
- Arai M, Ferreon JC, Wright PE. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. J Am Chem Soc. 2012 Feb 29;134(8):3792-803. doi: 10.1021/ja209936u. Epub 2012 Feb 15. PMID: 22280219; PMCID: PMC3290704.
- Arias A, Emmott E, Vashist S, Goodfellow I. Progress towards the prevention and treatment of norovirus infections. Future Microbiol. 2013 Nov;8(11):1475-87. doi: 10.2217/fmb.13.109. PMID: 24199805; PMCID: PMC3904215.
- Bartsch SM, Lopman BA, Ozawa S, Hall AJ, Lee BY. Global Economic Burden of Norovirus Gastroenteritis. PLoS One. 2016 Apr 26;11(4):e0151219. doi: 10.1371/journal.pone.0151219. PMID: 27115736; PMCID: PMC4846012.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2013 Jan;41(Database issue):D36-42. doi: 10.1093/nar/gks1195. Epub 2012 Nov 27. PMID: 23193287; PMCID: PMC3531190.
- Boswell ZK, Latham MP. Methyl-Based NMR Spectroscopy Methods for Uncovering Structural Dynamics in Large Proteins and Protein Complexes. Biochemistry. 2019 Jan 22;58(3):144-155. doi: 10.1021/acs.biochem.8b00953. Epub 2018 Oct 26. PMID: 30336000; PMCID: PMC6697110.
- Boswell ZK, Rahman S, Canny MD, Latham MP. A dynamic allosteric pathway underlies Rad50 ABC ATPase function in DNA repair. Sci Rep. 2018 Jan 26;8(1):1639. doi: 10.1038/s41598-018-19908-8. PMID: 29374232; PMCID: PMC5786021.
- Buchanan CJ, Gaunt B, Harrison PJ, Yang Y, Liu J, Khan A, Giltrap AM, Le Bas A, Ward PN, Gupta K, Dumoux M, Tan TK, Schimaski L, Daga S, Picchiotti N, Baldassarri M, Benetti E, Fallerini C, Fava F, Giliberti A, Koukos PI, Davy MJ, Lakshminarayanan A, Xue X, Papadakis G, Deimel LP, Casablancas-Antràs V, Claridge TDW, Bonvin AMJJ, Sattentau QJ, Furini S, Gori M, Huo J, Owens RJ, Schaffitzel C, Berger I, Renieri A; GEN-COVID Multicenter Study, Naismith JH, Baldwin AJ, Davis BG. Pathogen-sugar interactions revealed by universal saturation transfer analysis. Science. 2022 Jun 23:eabm3125. doi: 10.1126/science.abm3125. Epub ahead of print. PMID: 35737812.
- Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N. Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol. 2004 Jul;78(14):7727-36. doi: 10.1128/JVI.78.14.7727-7736.2004. PMID: 15220447; PMCID: PMC434083.

- Butterfoss GL, DeRose EF, Gabel SA, Perera L, Krahn JM, Mueller GA, Zheng X, London RE. Conformational dependence of 13C shielding and coupling constants for methionine methyl groups. J Biomol NMR. 2010 Sep;48(1):31-47. doi: 10.1007/s10858-010-9436-6. Epub 2010 Aug 24. PMID: 20734113; PMCID: PMC5598763.
- Cameroni E, Bowen JE, Rosen LE, Saliba C, Zepeda SK, Culap K, Pinto D, VanBlargan LA, De Marco A, di Iulio J, Zatta F, Kaiser H, Noack J, Farhat N, Czudnochowski N, Havenar-Daughton C, Sprouse KR, Dillen JR, Powell AE, Chen A, Maher C, Yin L, Sun D, Soriaga L, Bassi J, Silacci-Fregni C, Gustafsson C, Franko NM, Logue J, Iqbal NT, Mazzitelli I, Geffner J, Grifantini R, Chu H, Gori A, Riva A, Giannini O, Ceschi A, Ferrari P, Cippà PE, Franzetti-Pellanda A, Garzoni C, Halfmann PJ, Kawaoka Y, Hebner C, Purcell LA, Piccoli L, Pizzuto MS, Walls AC, Diamond MS, Telenti A, Virgin HW, Lanzavecchia A, Snell G, Veesler D, Corti D. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. Nature. 2022 Feb;602(7898):664-670. doi: 10.1038/s41586-021-04386-2. Epub 2021 Dec 23. PMID: 35016195.
- Capdevila DA, Braymer JJ, Edmonds KA, Wu H, Giedroc DP. Entropy redistribution controls allostery in a metalloregulatory protein. Proc Natl Acad Sci U S A. 2017 Apr 25;114(17):4424-4429. doi: 10.1073/pnas.1620665114. Epub 2017 Mar 27. PMID: 28348247; PMCID: PMC5410788.
- Cavanagh J, Fairbrother W, Palmer A, Skelton N. Protein NMR Spectroscopy: Principles and Practice, pages 532-553. Elsevier Science, 1995. ISBN 9780080515298
- Chao FA, Shi L, Masterson LR, Veglia G. FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. J Magn Reson. 2012 Jan;214(1):103-10. doi: 10.1016/j.jmr.2011.10.008. Epub 2011 Oct 20. PMID: 22134225; PMCID: PMC3487468.
- Chao FA, Kim J, Xia Y, Milligan M, Rowe N, Veglia G. FLAMEnGO 2.0: an enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances. J Magn Reson. 2014 Aug;245:17-23. doi: 10.1016/j.jmr.2014.04.012. Epub 2014 May 2. PMID: 24915505; PMCID: PMC4161213.
- Chazin WJ, Kördel J, Drakenberg T, Thulin E, Brodin P, Grundström T, Forsén S. Proline isomerism leads to multiple folded conformations of calbindin D9k: direct evidence from two-dimensional 1H NMR spectroscopy. Proc Natl Acad Sci U S A. 1989 Apr;86(7):2195-8. doi: 10.1073/pnas.86.7.2195. PMID: 2928325; PMCID: PMC286878.
- Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, Zhang Z, Fan P, Dong Y, Yang Y, Chen Z, Guo Y, Zhang J, Li Y, Song X, Chen Y, Xia L, Fu L, Hou L, Xu J, Yu C, Li J, Zhou Q, Chen W. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science. 2020 Aug 7;369(6504):650-655. doi: 10.1126/science.abc6952. Epub 2020 Jun 22. PMID: 32571838; PMCID: PMC7319273.
- Chu H, Hu B, Huang X, Chai Y, Zhou D, Wang Y, Shuai H, Yang D, Hou Y, Zhang X, Yuen TT, Cai JP, Zhang AJ, Zhou J, Yuan S, To KK, Chan IH, Sit KY, Foo DC, Wong IY, Ng AT, Cheung TT, Law SY, Au WK, Brindley MA, Chen Z, Kok KH, Chan JF, Yuen KY. Host and viral determinants for efficient SARS-CoV-2 infection of the human lung. Nat Commun. 2021 Jan 8;12(1):134. doi: 10.1038/s41467-020-20457-w. PMID: 33420022; PMCID: PMC7794309.
- Clore GM, Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev. 2009 Sep;109(9):4108-39. doi: 10.1021/cr900033p. PMID: 19522502; PMCID: PMC2825090.

- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020 Apr;5(4):536-544. doi: 10.1038/s41564-020-0695-z. Epub 2020 Mar 2. PMID: 32123347; PMCID: PMC7095448.
- Corti D, Purcell LA, Snell G, Veesler D. Tackling COVID-19 with neutralizing monoclonal antibodies. Cell. 2021 Jun 10;184(12):3086-3108. doi: 10.1016/j.cell.2021.05.005. Epub 2021 May 26. Erratum in: Cell. 2021 Aug 19;184(17):4593-4595. PMID: 34087172; PMCID: PMC8152891.
- Creutznacher R, Schulze E, Wallmann G, Peters T, Stein M, Mallagaray A. Chemical-Shift Perturbations Reflect Bile Acid Binding to Norovirus Coat Protein: Recognition Comes in Different Flavors. Chembiochem. 2020 Apr 1;21(7):1007-1021. doi: 10.1002/cbic.201900572. Epub 2019 Dec 5. PMID: 31644826; PMCID: PMC7186840.
- Creutznacher R. On the Interaction of Norovirus Capsid Proteins with Glycans and Bile Acids. Doctoral thesis, University of Lübeck, 2020.
- Creutznacher R, Maass T, Ogrissek P, Wallmann G, Feldmann C, Peters H, Lingemann M, Taube S, Peters T, Mallagaray A. NMR Experiments Shed New Light on Glycan Recognition by Human and Murine Norovirus Capsid Proteins. Viruses. 2021 Mar 5;13(3):416. doi: 10.3390/v13030416. PMID: 33807801; PMCID: PMC8001558.
- a: Creutznacher R, Maass T, Dülfer J, Feldmann C, Hartmann V, Lane MS, Knickmann J, Westermann LT, Thiede L, Smith TJ, Uetrecht C, Mallagaray A, Waudby CA, Taube S, Peters T. Distinct dissociation rates of murine and human norovirus P-domain dimers suggest a role of dimer stability in virus-host interactions. Commun Biol. 2022 Jun 9;5(1):563. doi: 10.1038/s42003-022-03497-4. PMID: 35680964; PMCID: PMC9184547.
- b: Creutznacher R, Maass T, Veselkova B, Ssebyatika G, Krey T, Empting M, Tautz N, Frank M, Kölbel K, Uetrecht C, Peters T. NMR Experiments Provide Insights into Ligand-Binding to the SARS-CoV-2 Spike Protein Receptor-Binding Domain. J Am Chem Soc. 2022 Jul 27;144(29):13060-13065. doi: 10.1021/jacs.2c05603. Epub 2022 Jul 13. PMID: 35830336.
- Croke RL, Patil SM, Quevreaux J, Kendall DA, Alexandrescu AT. NMR determination of pKa values in α-synuclein. Protein Sci. 2011 Feb;20(2):256-69. doi: 10.1002/pro.556. Epub 2010 Dec 13. PMID: 21280118; PMCID: PMC3048411.
- D'Souza C, Kanyalkar M, Joshi M, Coutinho E, Srivastava S. Probing molecular level interaction of oseltamivir with H5N1-NA and model membranes by molecular docking, multinuclear NMR and DSC methods. Biochim Biophys Acta. 2009 Feb;1788(2):484-94. doi: 10.1016/j.bbamem.2008.11.014. Epub 2008 Nov 28. PMID: 19100712.
- Day CJ, Bailly B, Guillon P, Dirr L, Jen FE, Spillings BL, Mak J, von Itzstein M, Haselhorst T, Jennings MP. Multidisciplinary Approaches Identify Compounds that Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2-ACE2 Receptor Interactions. mBio. 2021 Mar 30;12(2):e03681-20. doi: 10.1128/mBio.03681-20. PMID: 33785634; PMCID: PMC8092326.
- de Graaf M, van Beek J, Koopmans MP. Human norovirus transmission and evolution in a changing world. Nat Rev Microbiol. 2016 Jul;14(7):421-33. doi: 10.1038/nrmicro.2016.48. Epub 2016 May 23. PMID: 27211790.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995 Nov;6(3):277-93. doi: 10.1007/BF00197809. PMID: 8520220.

- DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA.
- Di Ciaula A, Garruti G, Lunardi Baccetto R, Molina-Molina E, Bonfrate L, Wang DQ, Portincasa P. Bile Acid Physiology. Ann Hepatol. 2017 Nov;16(Suppl. 1: s3-105.):s4-s14. doi: 10.5604/01.3001.0010.5493. PMID: 29080336.
- a: Dormitzer PR, Sun ZY, Wagner G, Harrison SC. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J. 2002 Mar 1;21(5):885-97. doi: 10.1093/emboj/21.5.885. PMID: 11867517; PMCID: PMC125907.
- b: Dormitzer PR, Sun ZY, Blixt O, Paulson JC, Wagner G, Harrison SC. Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8* core. J Virol. 2002 Oct;76(20):10512-7. doi: 10.1128/jvi.76.20.10512-10517.2002. PMID: 12239329; PMCID: PMC136543.
- Duan C, Liu W, Bi S, Chen L, Pan J, Zhou T, Lin K, Zhou W. J. Heterocycl. Synthesis and characterization of 4(*R*)-epimer impurities of zanamivir and laninamivir octanoate. Chem. 2022;59(3):569. doi: 10.1002/jhet.4402.
- Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, Neill FH, Blutt SE, Zeng XL, Qu L, Kou B, Opekun AR, Burrin D, Graham DY, Ramani S, Atmar RL, Estes MK. Replication of human noroviruses in stem cell-derived human enteroids. Science. 2016 Sep 23;353(6306):1387-1393. doi: 10.1126/science.aaf5211. Epub 2016 Aug 25. PMID: 27562956; PMCID: PMC5305121.
- Farmer BT 2nd, Constantine KL, Goldfarb V, Friedrichs MS, Wittekind M, Yanchunas J Jr, Robertson JG, Mueller L. Localizing the NADP+ binding site on the MurB enzyme by NMR. Nat Struct Biol. 1996 Dec;3(12):995-7. doi: 10.1038/nsb1296-995. PMID: 8946851.
- Fiege B, Rademacher C, Cartmell J, Kitov PI, Parra F, Peters T. Molecular details of the recognition of blood group antigens by a human norovirus as determined by STD NMR spectroscopy. Angew Chem Int Ed Engl. 2012 Jan 23;51(4):928-32. doi: 10.1002/anie.201105719. Epub 2011 Dec 13. PMID: 22170811.
- Flügge F, Peters T. Complete assignment of Ala, Ile, Leu, Met and Val methyl groups of human blood group A and B glycosyltransferases using lanthanide-induced pseudocontact shifts and methyl-methyl NOESY. J Biomol NMR. 2018 Apr;70(4):245-259. doi: 10.1007/s10858-018-0183-4. Epub 2018 Apr 26. PMID: 29700756.
- Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. Nature. 2007 Jul 19;448(7151):325-9. doi: 10.1038/nature05959. PMID: 17637663; PMCID: PMC4156320.
- Gans P, Hamelin O, Sounier R, Ayala I, Durá MA, Amero CD, Noirclerc-Savoye M, Franzetti B, Plevin MJ, Boisbouvier J. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew Chem Int Ed Engl. 2010 Mar 8;49(11):1958-62. doi: 10.1002/anie.200905660. PMID: 20157899.
- Gardner KH, Kay LE. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct. 1998;27:357-406. doi: 10.1146/annurev.biophys.27.1.357. PMID: 9646872.
- Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. N Engl J Med. 2009 Oct 29;361(18):1776-85. doi: 10.1056/NEJMra0804575. PMID: 19864676; PMCID: PMC3880795.

- Godoy-Ruiz R, Guo C, Tugarinov V. Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. J Am Chem Soc. 2010 Dec 29;132(51):18340-50. doi: 10.1021/ja1083656. Epub 2010 Dec 7. PMID: 21138300.
- Goel R, Bloch EM, Pirenne F, Al-Riyami AZ, Crowe E, Dau L, Land K, Townsend M, Jecko T, Rahimi-Levene N, Patidar G, Josephson CD, Arora S, Vermeulen M, Vrielink H, Montemayor C, Oreh A, Hindawi S, van den Berg K, Serrano K, So-Osman C, Wood E, Devine DV, Spitalnik SL; ISBT COVID-19 Working Group. ABO blood group and COVID-19: a review on behalf of the ISBT COVID-19 Working Group. Vox Sang. 2021 Sep;116(8):849-861. doi: 10.1111/vox.13076. Epub 2021 Feb 12. PMID: 33578447; PMCID: PMC8014128.
- Günther UL, Schaffhausen B. NMRKIN: simulating line shapes from two-dimensional spectra of proteins upon ligand binding. J Biomol NMR. 2002 Mar;22(3):201-9. doi: 10.1023/a:1014985726029. PMID: 11991351.
- Hansen DF, Neudecker P, Kay LE. Determination of isoleucine side-chain conformations in ground and excited states of proteins from chemical shifts. J Am Chem Soc. 2010 Jun 9;132(22):7589-91. doi: 10.1021/ja102090z. PMID: 20465253.
- Hansen DF, Kay LE. Determining valine side-chain rotamer conformations in proteins from methyl 13C chemical shifts: application to the 360 kDa half-proteasome. J Am Chem Soc. 2011 Jun 1;133(21):8272-81. doi: 10.1021/ja2014532. Epub 2011 May 5. PMID: 21545099.
- Hanson JE, Sauter NK, Skehel JJ, Wiley DC. Proton nuclear magnetic resonance studies of the binding of sialosides to intact influenza virus. Virology. 1992 Aug;189(2):525-33. doi: 10.1016/0042-6822(92)90576-b. PMID: 1641979.
- Harmer D, Gilbert M, Borman R, Clark KL. Quantitative mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme. FEBS Lett. 2002 Dec 4;532(1-2):107-10. doi: 10.1016/s0014-5793(02)03640-2. PMID: 12459472.
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A; COVID-19 Genomics UK (COG-UK) Consortium, Peacock SJ, Robertson DL. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021 Jul;19(7):409-424. doi: 10.1038/s41579-021-00573-0. Epub 2021 Jun 1. PMID: 34075212; PMCID: PMC8167834.
- Haselhorst T, Weimar T, Peters T. Molecular recognition of sialyl Lewis(x) and related saccharides by two lectins. J Am Chem Soc. 2001 Oct 31;123(43):10705-14. doi: 10.1021/ja011156h. PMID: 11674003.
- Haverkamp J, van Halbeek H, Dorland L, Vliegenthart JF, Pfeil R, Schauer R. High-resolution 1H-NMR spectroscopy of free and glycosidically linked O-acetylated sialic acids. Eur J Biochem. 1982 Feb;122(2):305-11. doi: 10.1111/j.1432-1033.1982.tb05881.x. PMID: 7060578.
- Hikmet F, Méar L, Edvinsson Å, Micke P, Uhlén M, Lindskog C. The protein expression profile of ACE2 in human tissues. Mol Syst Biol. 2020 Jul;16(7):e9610. doi: 10.15252/msb.20209610. PMID: 32715618; PMCID: PMC7383091.
- Hofmann AF. The Function Of Bile Salts In Fat Absorption. The Solvent Properties Of Dilute Micellar Solutions Of Conjugated Bile Salts. Biochem J. 1963 Oct;89(1):57-68. doi: 10.1042/bj0890057. PMID: 14097367; PMCID: PMC1202272.
- Hyberts SG, Robson SA, Wagner G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. J Biomol NMR. 2013 Feb;55(2):167-78. doi: 10.1007/s10858-012-9698-2. Epub 2012 Dec 29. PMID: 23274692; PMCID: PMC3570699.

- Hyberts SG, Arthanari H, Robson SA, Wagner G. Perspectives in magnetic resonance: NMR in the post-FFT era. J Magn Reson. 2014 Apr;241:60-73. doi: 10.1016/j.jmr.2013.11.014. PMID: 24656081; PMCID: PMC3964371.
- Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. Nat Rev Mol Cell Biol. 2022 Jan;23(1):3-20. doi: 10.1038/s41580-021-00418-x. Epub 2021 Oct 5. PMID: 34611326; PMCID: PMC8491763.
- Jackson RJ, Cooper KL, Tappenden P, Rees A, Simpson EL, Read RC, Nicholson KG. Oseltamivir, zanamivir and amantadine in the prevention of influenza: a systematic review. J Infect. 2011 Jan;62(1):14-25. doi: 10.1016/j.jinf.2010.10.003. Epub 2010 Oct 13. PMID: 20950645.
- Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF. Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). J Mol Biol. 2009 Dec 11;394(4):644-52. doi: 10.1016/j.jmb.2009.09.052. Epub 2009 Sep 26. PMID: 19786035.
- John M, Schmitz C, Park AY, Dixon NE, Huber T, Otting G. Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. J Am Chem Soc. 2007 Nov 7;129(44):13749-57. doi: 10.1021/ja0744753. Epub 2007 Oct 12. PMID: 17929923.
- Jones MK, Grau KR, Costantini V, Kolawole AO, de Graaf M, Freiden P, Graves CL, Koopmans M, Wallet SM, Tibbetts SA, Schultz-Cherry S, Wobus CE, Vinjé J, Karst SM. Human norovirus culture in B cells. Nat Protoc. 2015 Dec;10(12):1939-47. doi: 10.1038/nprot.2015.121. Epub 2015 Oct 29. PMID: 26513671; PMCID: PMC4689599.
- Katpally U, Wobus CE, Dryden K, Virgin HW 4th, Smith TJ. Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles. J Virol. 2008 Mar;82(5):2079-88. doi: 10.1128/JVI.02200-07. Epub 2007 Dec 19. PMID: 18094184; PMCID: PMC2258928.
- Katpally U, Voss NR, Cavazza T, Taube S, Rubin JR, Young VL, Stuckey J, Ward VK, Virgin HW 4th, Wobus CE, Smith TJ. High-resolution cryo-electron microscopy structures of murine norovirus 1 and rabbit hemorrhagic disease virus reveal marked flexibility in the receptor binding domains. J Virol. 2010 Jun;84(11):5836-41. doi: 10.1128/JVI.00314-10. Epub 2010 Mar 24. PMID: 20335264; PMCID: PMC2876586.
- Keeler J. Understanding NMR Spectroscopy. John Wiley & Sons Ltd, 2nd edition, pages 241-317, 2010. ISBN: 978-0-470-74608-0
- Kilic T, Koromyslova A, Malak V, Hansman GS. Atomic Structure of the Murine Norovirus Protruding Domain and Soluble CD300lf Receptor Complex. J Virol. 2018 May 14;92(11):e00413-18. doi: 10.1128/JVI.00413-18. PMID: 29563286; PMCID: PMC5952153.
- Kolawole AO, Smith HQ, Svoboda SA, Lewis MS, Sherman MB, Lynch GC, Pettitt BM, Smith TJ, Wobus CE. Norovirus Escape from Broadly Neutralizing Antibodies Is Limited to Allostery-Like Mechanisms. mSphere. 2017 Oct 18;2(5):e00334-17. doi: 10.1128/mSphere.00334-17. PMID: 29062895; PMCID: PMC5646240.
- Koromyslova AD, Devant JM, Kilic T, Sabin CD, Malak V, Hansman GS. Nanobody-Mediated Neutralization Reveals an Achilles Heel for Norovirus. J Virol. 2020 Jun 16;94(13):e00660-20. doi: 10.1128/JVI.00660-20. PMID: 32321816; PMCID: PMC7307151.
- Kramer F, Deshmukh M, Kessler H, Glaser S. Residual dipolar coupling constants: An elementary derivation of key equations. Concepts Magn Reson. 2004 Mar;21A:10-21. doi: 10.1002/cmr.a.20003.

- Krezel A, Bal W. A formula for correlating pKa values determined in D2O and H2O. J Inorg Biochem. 2004 Jan;98(1):161-6. doi: 10.1016/j.jinorgbio.2003.10.001. PMID: 14659645.
- Krishna NR, Jayalakshmi V. Quantitative Analysis of STD-NMR Spectra of Reversibly Forming Ligand-Receptor Complexes. Top Curr Chem. 2008;273:15-54. doi: 10.1007/128_2007_144. PMID: 23605458.
- Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol. 2000 Jan;7(1):72-7. doi: 10.1038/71280. PMID: 10625431.
- Lee D, Hilty C, Wider G, Wüthrich K. Effective rotational correlation times of proteins from NMR relaxation interference. J Magn Reson. 2006 Jan;178(1):72-6. doi: 10.1016/j.jmr.2005.08.014. Epub 2005 Sep 26. PMID: 16188473.
- Lescanne M, Skinner SP, Blok A, Timmer M, Cerofolini L, Fragai M, Luchinat C, Ubbink M. Methyl group assignment using pseudocontact shifts with PARAssign. J Biomol NMR. 2017 Dec;69(4):183-195. doi: 10.1007/s10858-017-0136-3. Epub 2017 Nov 27. PMID: 29181729; PMCID: PMC5736784.
- Levitt MH. Spin Dynamics: Basics of Nuclear Magnetic Resonance. John Wiley & Sons Ltd, 2nd edition, pages 599-664, 509-539, 2008. ISBN: 978-0-470-51117-6
- Li B, Wang L, Ge H, Zhang X, Ren P, Guo Y, Chen W, Li J, Zhu W, Chen W, Zhu L, Bai F. Identification of Potential Binding Sites of Sialic Acids on the RBD Domain of SARS-CoV-2 Spike Protein. Front Chem. 2021 Jul 22;9:659764. doi: 10.3389/fchem.2021.659764. PMID: 34368076; PMCID: PMC8341434.
- Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T, Winter H, Meister M, Veith C, Boots AW, Hennig BP, Kreuter M, Conrad C, Eils R. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J. 2020 May 18;39(10):e105114. doi: 10.15252/embj.20105114. Epub 2020 Apr 14. PMID: 32246845; PMCID: PMC7232010.
- Lüttgen H, Robelek R, Mühlberger R, Diercks T, Schuster SC, Köhler P, Kessler H, Bacher A, Richter G. Transcriptional regulation by antitermination. Interaction of RNA with NusB protein and NusB/NusE protein complex of Escherichia coli. J Mol Biol. 2002 Mar 1;316(4):875-85. doi: 10.1006/jmbi.2001.5388. PMID: 11884128.
- Maaß T. Stable isotope labeling schemes for the assignment of NMR spectra of murine Norovirus P-domain proteins. Master's thesis, University of Lübeck, 2019.
- a: Maass T, Westermann LT, Creutznacher R, Mallagaray A, Dülfer J, Uetrecht C, Peters T. Assignment of Ala, Ile, Leu^{proS}, Met, and Val^{proS} methyl groups of the protruding domain of murine norovirus capsid protein VP1 using methyl-methyl NOEs, site directed mutagenesis, and pseudocontact shifts. Biomol NMR Assign. 2022 Apr;16(1):97-107. doi: 10.1007/s12104-022-10066-7. Epub 2022 Jan 20. PMID: 35050443; PMCID: PMC9068638.
- b: Maass T, Ssebyatika G, Brückner M, Breckwoldt L, Krey T, Mallagaray A, Peters T, Frank M, Creutznacher R. Binding of Glycans to the SARS CoV-2 Spike Protein, an Open Question: NMR Data on Binding Site Localization, Affinity, and Selectivity. Chemistry. 2022 Sep 26:e202202614. doi: 10.1002/chem.202202614. Epub ahead of print. PMID: 36161798; PMCID: PMC9537997.
- Mallagaray A, Creutznacher R, Dülfer J, Mayer PHO, Grimm LL, Orduña JM, Trabjerg E, Stehle T, Rand KD, Blaum BS, Uetrecht C, Peters T. A post-translational modification of human Norovirus

capsid protein attenuates glycan binding. Nat Commun. 2019 Mar 21;10(1):1320. doi: 10.1038/s41467-019-09251-5. PMID: 30899001; PMCID: PMC6428809.

- Mayer M, Meyer B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angew Chem Int Ed Engl. 1999 Jun 14;38(12):1784-1788. doi: 10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q. PMID: 29711196.
- Mayer M, Meyer B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc. 2001 Jun 27;123(25):6108-17. doi: 10.1021/ja0100120. PMID: 11414845.
- a: McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, Tortorici MA, Navarro MJ, Silacci-Fregni C, Saliba C, Sprouse KR, Agostini M, Pinto D, Culap K, Bianchi S, Jaconi S, Cameroni E, Bowen JE, Tilles SW, Pizzuto MS, Guastalla SB, Bona G, Pellanda AF, Garzoni C, Van Voorhis WC, Rosen LE, Snell G, Telenti A, Virgin HW, Piccoli L, Corti D, Veesler D. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. Science. 2021 Aug 6;373(6555):648-654. doi: 10.1126/science.abi7994. Epub 2021 Jul 1. PMID: 34210893.
- b: McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, Beltramello M, Chen A, Liu Z, Zatta F, Zepeda S, di Iulio J, Bowen JE, Montiel-Ruiz M, Zhou J, Rosen LE, Bianchi S, Guarino B, Fregni CS, Abdelnabi R, Foo SC, Rothlauf PW, Bloyet LM, Benigni F, Cameroni E, Neyts J, Riva A, Snell G, Telenti A, Whelan SPJ, Virgin HW, Corti D, Pizzuto MS, Veesler D. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. 2021 Apr 29;184(9):2332-2347.e16. doi: 10.1016/j.cell.2021.03.028. Epub 2021 Mar 16. PMID: 33761326; PMCID: PMC7962585.
- McConnell EL, Basit AW, Murdan S. Measurements of rat and mouse gastrointestinal pH, fluid and lymphoid tissue, and implications for in-vivo experiments. J Pharm Pharmacol. 2008 Jan;60(1):63-70. doi: 10.1211/jpp.60.1.0008. PMID: 18088506.
- McConnell HM. Reaction Rates by Nuclear Magnetic Resonance. J Chem Phys. 1958; 28:430–431. doi: 10.1063/1.1744152.
- McFadden N, Bailey D, Carrara G, Benson A, Chaudhry Y, Shortland A, Heeney J, Yarovinsky F, Simmonds P, Macdonald A, Goodfellow I. Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4. PLoS Pathog. 2011 Dec;7(12):e1002413. doi: 10.1371/journal.ppat.1002413. Epub 2011 Dec 8. PMID: 22174679; PMCID: PMC3234229.
- Meyer B, Peters T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem Int Ed Engl. 2003 Feb 24;42(8):864-90. doi: 10.1002/anie.200390233. PMID: 12596167.
- Mittermaier AK, Kay LE. Observing biological dynamics at atomic resolution using NMR. Trends Biochem Sci. 2009 Dec;34(12):601-11. doi: 10.1016/j.tibs.2009.07.004. Epub 2009 Oct 19. PMID: 19846313.
- Moe SM. Calcium Homeostasis in Health and in Kidney Disease. Compr Physiol. 2016 Sep 15;6(4):1781-1800. doi: 10.1002/cphy.c150052. PMID: 27783859.
- Monneau YR, Rossi P, Bhaumik A, Huang C, Jiang Y, Saleh T, Xie T, Xing Q, Kalodimos CG. Automatic methyl assignment in large proteins by the MAGIC algorithm. J Biomol NMR. 2017 Dec;69(4):215-227. doi: 10.1007/s10858-017-0149-y. Epub 2017 Nov 2. PMID: 29098507; PMCID: PMC5764113.

- Mühlberg L, Alarcin T, Maass T, Creutznacher R, Küchler R, Mallagaray A. Ligand-induced structural transitions combined with paramagnetic ions facilitate unambiguous NMR assignments of methyl groups in large proteins. J Biomol NMR. 2022 Jun;76(3):59-74. doi: 10.1007/s10858-022-00394-0. Epub 2022 Apr 10. PMID: 35397749; PMCID: PMC9247001.
- Mulder FA. Leucine side-chain conformation and dynamics in proteins from 13C NMR chemical shifts. Chembiochem. 2009 Jun 15;10(9):1477-9. doi: 10.1002/cbic.200900086. PMID: 19466705.
- Müller-Hermes C, Creutznacher R, Mallagaray A. Complete assignment of Ala, Ile, LeuProS, Met and ValProS methyl groups of the protruding domain from human norovirus GII.4 Saga. Biomol NMR Assign. 2020 Apr;14(1):123-130. doi: 10.1007/s12104-020-09932-z. Epub 2020 Jan 28. PMID: 31993958; PMCID: PMC7069894.
- Nakamura T, Bubb WA, Saito T, Arai I, Urashima T. An NMR study of the lactonization of alpha-N-acetylneuraminyl-(2 --> 3)-lactose. Carbohydr Res. 2000 Nov 3;329(2):471-6. doi: 10.1016/s0008-6215(00)00202-0. PMID: 11117332.
- Neffe AT, Bilang M, Grüneberg I, Meyer B. Rational optimization of the binding affinity of CD4 targeting peptidomimetics with potential anti HIV activity. J Med Chem. 2007 Jul 26;50(15):3482-8. doi: 10.1021/jm070206b. Epub 2007 Jun 28. PMID: 17602545.
- Nelson CA, Wilen CB, Dai YN, Orchard RC, Kim AS, Stegeman RA, Hsieh LL, Smith TJ, Virgin HW, Fremont DH. Structural basis for murine norovirus engagement of bile acids and the CD300lf receptor. Proc Natl Acad Sci U S A. 2018 Sep 25;115(39):E9201-E9210. doi: 10.1073/pnas.1805797115. Epub 2018 Sep 7. PMID: 30194229; PMCID: PMC6166816.
- Neuhaus D. Nuclear Overhauser Effect. eMagRes (eds R.K. Harris and R.L. Wasylishen). 2011 Jun 15. doi: 10.1002/9780470034590.emrstm0350.pub2
- Nguyen L, McCord KA, Bui DT, Bouwman KM, Kitova EN, Elaish M, Kumawat D, Daskhan GC, Tomris I, Han L, Chopra P, Yang TJ, Willows SD, Mason AL, Mahal LK, Lowary TL, West LJ, Hsu SD, Hobman T, Tompkins SM, Boons GJ, de Vries RP, Macauley MS, Klassen JS. Sialic acid-containing glycolipids mediate binding and viral entry of SARS-CoV-2. Nat Chem Biol. 2022 Jan;18(1):81-90. doi: 10.1038/s41589-021-00924-1. Epub 2021 Nov 9. PMID: 34754101.
- Nieba L, Krebber A, Plückthun A. Competition BIAcore for measuring true affinities: large differences from values determined from binding kinetics. Anal Biochem. 1996 Feb 15;234(2):155-65. doi: 10.1006/abio.1996.0067. PMID: 8714593.
- Orton HW, Huber T, Otting G. Paramagpy: software for fitting magnetic susceptibility tensors using paramagnetic effects measured in NMR spectra. Magn Reson. 2020 Feb 14;1:1-12. doi: 10.5194/mr-1-1-2020.
- Otting G. Protein NMR using paramagnetic ions. Annu Rev Biophys. 2010;39:387-405. doi: 10.1146/annurev.biophys.093008.131321. PMID: 20462377.
- Patel R, Kaki M, Potluri VS, Kahar P, Khanna D. A comprehensive review of SARS-CoV-2 vaccines: Pfizer, Moderna & Johnson & Johnson. Hum Vaccin Immunother. 2022 Dec 31;18(1):2002083. doi: 10.1080/21645515.2021.2002083. Epub 2022 Feb 7. PMID: 35130825; PMCID: PMC8862159.
- Pederson K, Chalmers GR, Gao Q, Elnatan D, Ramelot TA, Ma LC, Montelione GT, Kennedy MA, Agard DA, Prestegard JH. NMR characterization of HtpG, the E. coli Hsp90, using sparse labeling with 13C-methyl alanine. J Biomol NMR. 2017 Jul;68(3):225-236. doi: 10.1007/s10858-017-0123-8. Epub 2017 Jun 26. PMID: 28653216; PMCID: PMC5546222.

- Pérez-Victoria I, Kemper S, Patel MK, Edwards JM, Errey JC, Primavesi LF, Paul MJ, Claridge TD, Davis BG. Saturation transfer difference NMR reveals functionally essential kinetic differences for a sugar-binding repressor protein. Chem Commun (Camb). 2009 Oct 21;(39):5862-4. doi: 10.1039/b913489a. Epub 2009 Aug 19. PMID: 19787122.
- Pervushin K, Riek R, Wider G, Wüthrich 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 U S A. 1997 Nov 11;94(23):12366-71. doi: 10.1073/pnas.94.23.12366. PMID: 9356455; PMCID: PMC24947.
- Petitjean SJL, Chen W, Koehler M, Jimmidi R, Yang J, Mohammed D, Juniku B, Stanifer ML, Boulant S, Vincent SP, Alsteens D. Multivalent 9-O-Acetylated-sialic acid glycoclusters as potent inhibitors for SARS-CoV-2 infection. Nat Commun. 2022 May 10;13(1):2564. doi: 10.1038/s41467-022-30313-8. Erratum in: Nat Commun. 2022 Jun 24;13(1):3611. PMID: 35538121; PMCID: PMC9091252.
- Pires SM, Fischer-Walker CL, Lanata CF, Devleesschauwer B, Hall AJ, Kirk MD, Duarte AS, Black RE, Angulo FJ. Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food. PLoS One. 2015 Dec 3;10(12):e0142927. doi: 10.1371/journal.pone.0142927. PMID: 26632843; PMCID: PMC4668836.
- Pritišanac I, Degiacomi MT, Alderson TR, Carneiro MG, Ab E, Siegal G, Baldwin AJ. Automatic Assignment of Methyl-NMR Spectra of Supramolecular Machines Using Graph Theory. J Am Chem Soc. 2017 Jul 19;139(28):9523-9533. doi: 10.1021/jacs.6b11358. Epub 2017 Jul 10. PMID: 28691806.
- Pritišanac I, Würz JM, Alderson TR, Güntert P. Automatic structure-based NMR methyl resonance assignment in large proteins. Nat Commun. 2019 Oct 29;10(1):4922. doi: 10.1038/s41467-019-12837-8. PMID: 31664028; PMCID: PMC6820720.
- Pritišanac I, Alderson TR, Güntert P. Automated assignment of methyl NMR spectra from large proteins. Prog Nucl Magn Reson Spectrosc. 2020 Jun-Aug;118-119:54-73. doi: 10.1016/j.pnmrs.2020.04.001. Epub 2020 Apr 24. PMID: 32883449.
- Proudfoot A, Frank AO, Ruggiu F, Mamo M, Lingel A. Facilitating unambiguous NMR assignments and enabling higher probe density through selective labeling of all methyl containing amino acids. J Biomol NMR. 2016 May;65(1):15-27. doi: 10.1007/s10858-016-0032-2. Epub 2016 Apr 29. PMID: 27130242.
- Rademacher C, Krishna NR, Palcic M, Parra F, Peters T. NMR experiments reveal the molecular basis of receptor recognition by a calicivirus. J Am Chem Soc. 2008 Mar 19;130(11):3669-75. doi: 10.1021/ja710854r. Epub 2008 Feb 27. PMID: 18302385.
- Robilotti E, Deresinski S, Pinsky BA. Norovirus. Clin Microbiol Rev. 2015 Jan;28(1):134-64. doi: 10.1128/CMR.00075-14. PMID: 25567225; PMCID: PMC4284304.
- Roderer DJ, Schärer MA, Rubini M, Glockshuber R. Acceleration of protein folding by four orders of magnitude through a single amino acid substitution. Sci Rep. 2015 Jun 30;5:11840. doi: 10.1038/srep11840. PMID: 26121966; PMCID: PMC4485320.
- Romani A, Scarpa A. Regulation of cell magnesium. Arch Biochem Biophys. 1992 Oct;298(1):1-12. doi: 10.1016/0003-9861(92)90086-c. PMID: 1524417.
- Saso W, Yamasaki M, Nakakita SI, Fukushi S, Tsuchimoto K, Watanabe N, Sriwilaijaroen N, Kanie O, Muramatsu M, Takahashi Y, Matano T, Takeda M, Suzuki Y, Watashi K. Significant role of

host sialylated glycans in the infection and spread of severe acute respiratory syndrome coronavirus 2. PLoS Pathog. 2022 Jun 14;18(6):e1010590. doi: 10.1371/journal.ppat.1010590. PMID: 35700214; PMCID: PMC9197039.

- Sauter NK, Bednarski MD, Wurzburg BA, Hanson JE, Whitesides GM, Skehel JJ, Wiley DC. Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. Biochemistry. 1989 Oct 17;28(21):8388-96. doi: 10.1021/bi00447a018. PMID: 2605190.
- Sauter NK, Hanson JE, Glick GD, Brown JH, Crowther RL, Park SJ, Skehel JJ, Wiley DC. Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. Biochemistry. 1992 Oct 13;31(40):9609-21. doi: 10.1021/bi00155a013. PMID: 1327122.
- Schmitz C, Stanton-Cook MJ, Su XC, Otting G, Huber T. Numbat: an interactive software tool for fitting Deltachi-tensors to molecular coordinates using pseudocontact shifts. J Biomol NMR. 2008 Jul;41(3):179-89. doi: 10.1007/s10858-008-9249-z. Epub 2008 Jun 24. PMID: 18574699.
- Schoenle MV, Li Y, Yuan M, Clarkson MW, Wilson IA, Peti W, Page R. NMR Based SARS-CoV-2 Antibody Screening. J Am Chem Soc. 2021 Jun 2;143(21):7930-7934. doi: 10.1021/jacs.1c03945. Epub 2021 May 21. PMID: 34018723; PMCID: PMC8171107.
- Schütz S, Sprangers R. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. Prog Nucl Magn Reson Spectrosc. 2020 Feb;116:56-84. doi: 10.1016/j.pnmrs.2019.09.004. Epub 2019 Sep 30. PMID: 32130959.
- Sereda TJ, Mant CT, Quinn AM, Hodges RS. Effect of the alpha-amino group on peptide retention behaviour in reversed-phase chromatography. Determination of the pK(a) values of the alpha-amino group of 19 different N-terminal amino acid residues. J Chromatogr. 1993 Aug 27;646(1):17-30. doi: 10.1016/s0021-9673(99)87003-4. PMID: 8408425.
- Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020 May;581(7807):221-224. doi: 10.1038/s41586-020-2179-y. Epub 2020 Mar 30. PMID: 32225175; PMCID: PMC7328981.
- Sherman MB, Williams AN, Smith HQ, Nelson C, Wilen CB, Fremont DH, Virgin HW, Smith TJ. Bile Salts Alter the Mouse Norovirus Capsid Conformation: Possible Implications for Cell Attachment and Immune Evasion. J Virol. 2019 Sep 12;93(19):e00970-19. doi: 10.1128/JVI.00970-19. PMID: 31341042; PMCID: PMC6744230.
- Sherman MB, Williams AN, Smith HQ, Pettitt BM, Wobus CE, Smith TJ. Structural Studies on the Shapeshifting Murine Norovirus. Viruses. 2021 Oct 26;13(11):2162. doi: 10.3390/v13112162. PMID: 34834968; PMCID: PMC8621758.
- Siemons L, Uluca-Yazgi B, Pritchard RB, McCarthy S, Heise H, Hansen DF. Determining isoleucine side-chain rotamer-sampling in proteins from 13C chemical shift. Chem Commun (Camb). 2019 Dec 7;55(94):14107-14110. doi: 10.1039/c9cc06496f. Epub 2019 Oct 23. PMID: 31642826; PMCID: PMC7138115.
- Singh BK, Leuthold MM, Hansman GS. Human noroviruses' fondness for histo-blood group antigens. J Virol. 2015 Feb;89(4):2024-40. doi: 10.1128/JVI.02968-14. Epub 2014 Nov 26. PMID: 25428879; PMCID: PMC4338890.
- Skjaerven L, Reuter N, Martinez A. Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: a computational approach. Future Med Chem. 2011 Dec;3(16):2079-100. doi: 10.4155/fmc.11.159. PMID: 22098354.

- Snowden JS, Hurdiss DL, Adeyemi OO, Ranson NA, Herod MR, Stonehouse NJ. Dynamics in the murine norovirus capsid revealed by high-resolution cryo-EM. PLoS Biol. 2020 Mar 31;18(3):e3000649. doi: 10.1371/journal.pbio.3000649. PMID: 32231352; PMCID: PMC7108717.
- Song C, Takai-Todaka R, Miki M, Haga K, Fujimoto A, Ishiyama R, Oikawa K, Yokoyama M, Miyazaki N, Iwasaki K, Murakami K, Katayama K, Murata K. Dynamic rotation of the protruding domain enhances the infectivity of norovirus. PLoS Pathog. 2020 Jul 2;16(7):e1008619. doi: 10.1371/journal.ppat.1008619. PMID: 32614892; PMCID: PMC7331980.
- Sprangers R, Gribun A, Hwang PM, Houry WA, Kay LE. Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. Proc Natl Acad Sci U S A. 2005 Nov 15;102(46):16678-83. doi: 10.1073/pnas.0507370102. Epub 2005 Nov 1. PMID: 16263929; PMCID: PMC1283831.
- Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007 Feb 8;445(7128):618-22. doi: 10.1038/nature05512. Epub 2007 Jan 21. PMID: 17237764.
- Sprangers R, Velyvis A, Kay LE. Solution NMR of supramolecular complexes: providing new insights into function. Nat Methods. 2007 Sep;4(9):697-703. doi: 10.1038/nmeth1080. PMID: 17762877.
- Ströh LJ, Maginnis MS, Blaum BS, Nelson CD, Neu U, Gee GV, O'Hara BA, Motamedi N, DiMaio D, Atwood WJ, Stehle T. The Greater Affinity of JC Polyomavirus Capsid for α2,6-Linked Lactoseries Tetrasaccharide c than for Other Sialylated Glycans Is a Major Determinant of Infectivity. J Virol. 2015 Jun;89(12):6364-75. doi: 10.1128/JVI.00489-15. Epub 2015 Apr 8. PMID: 25855729; PMCID: PMC4474300.
- Sun H, Kay LE, Tugarinov V. An optimized relaxation-based coherence transfer NMR experiment for the measurement of side-chain order in methyl-protonated, highly deuterated proteins. J Phys Chem B. 2011 Dec 15;115(49):14878-84. doi: 10.1021/jp209049k. Epub 2011 Nov 15. PMID: 22040035.
- Sun XL. The role of cell surface sialic acids for SARS-CoV-2 infection. Glycobiology. 2021 Nov 18;31(10):1245-1253. doi: 10.1093/glycob/cwab032. PMID: 33909065; PMCID: PMC8600286.
- Taube S, Rubin JR, Katpally U, Smith TJ, Kendall A, Stuckey JA, Wobus CE. High-resolution x-ray structure and functional analysis of the murine norovirus 1 capsid protein protruding domain. J Virol. 2010 Jun;84(11):5695-705. doi: 10.1128/JVI.00316-10. Epub 2010 Mar 24. PMID: 20335262; PMCID: PMC2876589.
- Thackray LB, Wobus CE, Chachu KA, Liu B, Alegre ER, Henderson KS, Kelley ST, Virgin HW 4th. Murine noroviruses comprising a single genogroup exhibit biological diversity despite limited sequence divergence. J Virol. 2007 Oct;81(19):10460-73. doi: 10.1128/JVI.00783-07. Epub 2007 Jul 25. PMID: 17652401; PMCID: PMC2045448.
- Toelzer C, Gupta K, Yadav SKN, Borucu U, Davidson AD, Kavanagh Williamson M, Shoemark DK, Garzoni F, Staufer O, Milligan R, Capin J, Mulholland AJ, Spatz J, Fitzgerald D, Berger I, Schaffitzel C. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. Science. 2020 Nov 6;370(6517):725-730. doi: 10.1126/science.abd3255. Epub 2020 Sep 21. PMID: 32958580; PMCID: PMC8050947.
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and

protein complexes. J Am Chem Soc. 2003 Aug 27;125(34):10420-8. doi: 10.1021/ja030153x. PMID: 12926967.

- Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. J Am Chem Soc. 2005 Mar 2;127(8):2767-75. doi: 10.1021/ja044032o. PMID: 15725035.
- Tugarinov V, Sprangers R, Kay LE. Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. J Am Chem Soc. 2007 Feb 14;129(6):1743-50. doi: 10.1021/ja067827z. Epub 2007 Jan 24. PMID: 17249677.
- Tugarinov V, Kay LE. Estimating side-chain order in [U-2H;13CH3]-labeled high molecular weight proteins from analysis of HMQC/HSQC spectra. J Phys Chem B. 2013 Apr 4;117(13):3571-7. doi: 10.1021/jp401088c. Epub 2013 Mar 25. PMID: 23458382.
- Tzou PL, Tao K, Pond SLK, Shafer RW. Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons. PLoS One. 2022 Mar 9;17(3):e0261045. doi: 10.1371/journal.pone.0261045. PMID: 35263335; PMCID: PMC8906623.
- Unione L, Moure MJ, Lenza MP, Oyenarte I, Ereño-Orbea J, Ardá A, Jiménez-Barbero J. The SARS-CoV-2 Spike Glycoprotein Directly Binds Exogeneous Sialic Acids: A NMR View. Angew Chem Int Ed Engl. 2022 Apr 25;61(18):e202201432. doi: 10.1002/anie.202201432. Epub 2022 Mar 7. PMID: 35191576; PMCID: PMC9074024.
- Vashist S, Bailey D, Putics A, Goodfellow I. Model systems for the study of human norovirus Biology. Future Virol. 2009 Jul;4(4):353-367. doi: 10.2217/fvl.09.18. PMID: 21516251; PMCID: PMC3079900.
- Venditti V, Fawzi NL, Clore GM. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. J Biomol NMR. 2011 Nov;51(3):319-28. doi: 10.1007/s10858-011-9559-4. Epub 2011 Sep 4. PMID: 21935714; PMCID: PMC3212433.
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins. 2005 Jun 1;59(4):687-96. doi: 10.1002/prot.20449. PMID: 15815974.
- Wald J, Pasin M, Richter M, Walther C, Mathai N, Kirchmair J, Makarov VA, Goessweiner-Mohr N, Marlovits TC, Zanella I, Real-Hohn A, Verdaguer N, Blaas D, Schmidtke M. Cryo-EM structure of pleconaril-resistant rhinovirus-B5 complexed to the antiviral OBR-5-340 reveals unexpected binding site. Proc Natl Acad Sci U S A. 2019 Sep 17;116(38):19109-19115. doi: 10.1073/pnas.1904732116. Epub 2019 Aug 28. PMID: 31462495; PMCID: PMC6754595
- Walker IH, Hsieh PC, Riggs PD. Mutations in maltose-binding protein that alter affinity and solubility properties. Appl Microbiol Biotechnol. 2010 Sep;88(1):187-97. doi: 10.1007/s00253-010-2696-y. Epub 2010 Jun 10. PMID: 20535468; PMCID: PMC2940430.
- Wang S, McDonnell EH, Sedor FA, Toffaletti JG. pH effects on measurements of ionized calcium and ionized magnesium in blood. Arch Pathol Lab Med. 2002 Aug;126(8):947-50. doi: 10.5858/2002-126-0947-PEOMOI. PMID: 12171493.
- Waudby CA, Ramos A, Cabrita LD, Christodoulou J. Two-Dimensional NMR Lineshape Analysis. Sci Rep. 2016 Apr 25;6:24826. doi: 10.1038/srep24826. PMID: 27109776; PMCID: PMC4843008.
- Welling SH, Hubálek F, Jacobsen J, Brayden DJ, Rahbek UL, Buckley ST. The role of citric acid in oral peptide and protein formulations: relationship between calcium chelation and proteolysis

inhibition. Eur J Pharm Biopharm. 2014 Apr;86(3):544-51. doi: 10.1016/j.ejpb.2013.12.017. Epub 2013 Dec 31. PMID: 24384069.

- Wen J, Zhou P, Wu J. Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl-methyl NOESY for large proteins. J Magn Reson. 2012 May;218:128-32. doi: 10.1016/j.jmr.2012.02.021. Epub 2012 Mar 9. PMID: 22464875; PMCID: PMC3625671.
- Westermann LT. MD Simulations and NMR experiments of the P-domain of a murine norovirus. Bachelor's thesis, University of Lübeck, 2020.
- Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Lambson BE, de Oliveira T, Vermeulen M, van der Berg K, Rossouw T, Boswell M, Ueckermann V, Meiring S, von Gottberg A, Cohen C, Morris L, Bhiman JN, Moore PL. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat Med. 2021 Apr;27(4):622-625. doi: 10.1038/s41591-021-01285-x. Epub 2021 Mar 2. PMID: 33654292.
- Willett BJ, Grove J, MacLean OA, Wilkie C, De Lorenzo G, Furnon W, Cantoni D, Scott S, Logan N, Ashraf S, Manali M, Szemiel A, Cowton V, Vink E, Harvey WT, Davis C, Asamaphan P, Smollett K, Tong L, Orton R, Hughes J, Holland P, Silva V, Pascall DJ, Puxty K, da Silva Filipe A, Yebra G, Shaaban S, Holden MTG, Pinto RM, Gunson R, Templeton K, Murcia PR, Patel AH, Klenerman P, Dunachie S; PITCH Consortium; COVID-19 Genomics UK (COG-UK) Consortium, Haughney J, Robertson DL, Palmarini M, Ray S, Thomson EC. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. Nat Microbiol. 2022 Jul 7. doi: 10.1038/s41564-022-01143-7. Epub ahead of print. PMID: 35798890.
- a: Williams AN, Sherman MB, Smith HQ, Taube S, Pettitt BM, Wobus CE, Smith TJ. A Norovirus Uses Bile Salts To Escape Antibody Recognition While Enhancing Receptor Binding. J Virol. 2021 Jun 10;95(13):e0017621. doi: 10.1128/JVI.00176-21. Epub 2021 Jun 10. PMID: 33827952; PMCID: PMC8315966.
- b: Williams AN, Sherman MB, Smith HQ, Taube S, Pettitt BM, Wobus CE, Smith TJ. Multiple Signals in the Gut Contract the Mouse Norovirus Capsid To Block Antibody Binding While Enhancing Receptor Affinity. J Virol. 2021 Oct 27;95(22):e0147121. doi: 10.1128/JVI.01471-21. Epub 2021 Sep 1. PMID: 34468172; PMCID: PMC8549501.
- Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc. 2013 Aug;73:1-16. doi: 10.1016/j.pnmrs.2013.02.001. Epub 2013 Mar 21. Erratum in: Prog Nucl Magn Reson Spectrosc. 2014 Jul;80:64. PMID: 23962882.
- Wolfe HG. Selection for Blood-Ph in the House Mouse. Genetics. 1961 Jan;46(1):55-75. doi: 10.1093/genetics/46.1.55. PMID: 17248035; PMCID: PMC1210130.
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe. 2020 Mar 11;27(3):325-328. doi: 10.1016/j.chom.2020.02.001. Epub 2020 Feb 7. PMID: 32035028; PMCID: PMC7154514.
- Wu SC, Arthur CM, Wang J, Verkerke H, Josephson CD, Kalman D, Roback JD, Cummings RD, Stowell SR. The SARS-CoV-2 receptor-binding domain preferentially recognizes blood group A. Blood Adv. 2021 Mar 9;5(5):1305-1309. doi: 10.1182/bloodadvances.2020003259. PMID: 33656534; PMCID: PMC7929867.
- Xu Y, Matthews S. MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins. J Biomol NMR. 2013 Feb;55(2):179-87. doi: 10.1007/s10858-012-9700-z. Epub 2013 Jan 5. PMID: 23292498.

- Yoshida T, Takahashi Y, Takashima S. Effect of low extracellular Ca2+ on growth, spreading area, cytoplasmic Ca2+ concentration, and intracellular pH in normal and transformed human fibroblasts. J Cell Physiol. 1993 Feb;154(2):301-9. doi: 10.1002/jcp.1041540213. PMID: 8425911.

Supplementary Information on “An NMR study into virus-host interactions of murine Norovirus capsid and SARS-CoV-2 spike proteins”

1 AMIGO

1.1 How to run AMIGO, adjust parameters, and provide input files

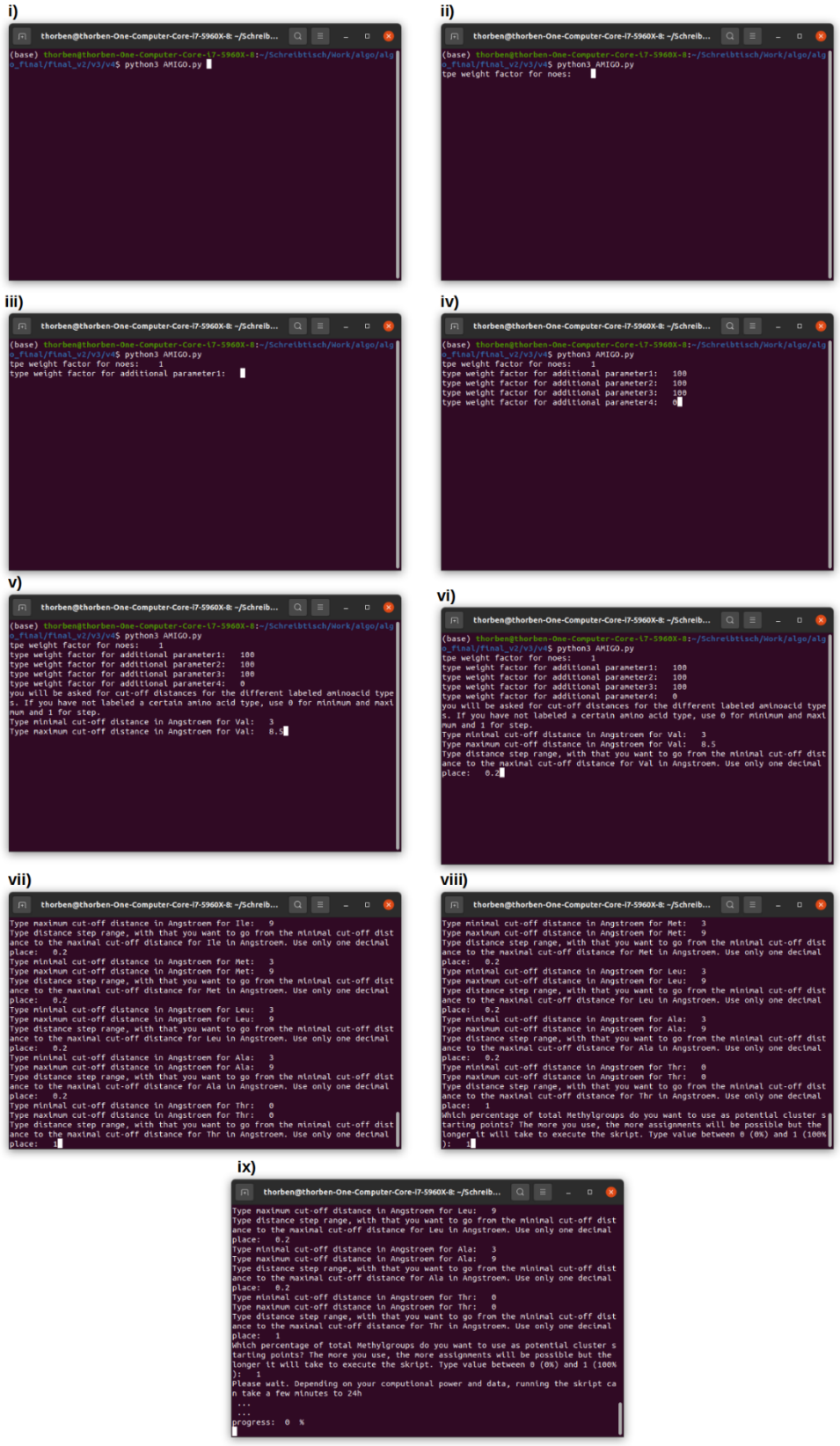


Fig. S 1.1: How to run AMIGO - AMIGO's interactive dialog with the user.

(i) shows how the script is executed. In (ii-iv), weighting factors are submitted by the user. If there is no set of additional parameters (i.e. restraints) available or you don't want to use it for the calculation, use "0". In (v) the minimal and maximal cut-off distances for valine methyl groups are chosen in Å. In (vi) the step with that AMIGO goes from the minimal to the maximal cut-off distance for valine methyl groups is chosen in Å. Note that this is important for creating the set of structure based building blocks. If you have not labeled a certain amino acid type, use "0" as minimal and maximal cut-off distance and "1" as step (vii). The last parameter required for running AMIGO is the number of potential starting points that should be considered by AMIGO. In the example (viii and ix), "1" is chosen corresponding to 100 % of the methyl groups (i.e. *building blocks*). Lower values can be used to save calculation time according to Note S 1.1 (e.g., 0.1 for only using the 10% "most unique" methyl groups).

Note S 1.1: Calculating the *uniqueness score* for sorting potential methyl walk starting points

The *uniqueness score* is calculated as explained in chapter 5.1.2 (Eq. 29-31). NOE building blocks are sorted according to their uniqueness score. Note that a resonance with a low uniqueness score (very "unique") is more likely to be assigned correctly as a resonance with a high uniqueness score ("not so unique").

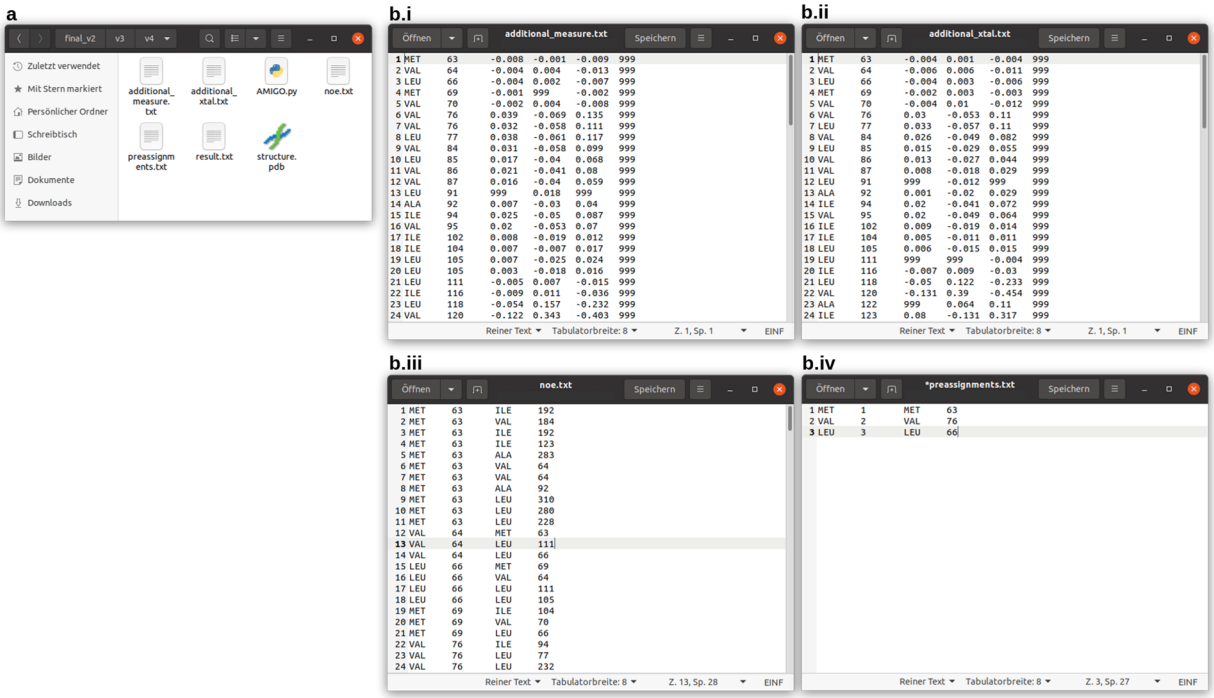


Fig. S 1.2: Input files for AMIGO.

AMIGO is implemented in *python3*. Place all necessary files in a single folder (a). They must have exactly the same name as shown in (a). additional_measure.txt is a text file containing additional experimental restraints (b.i, e.g. PCSs). The first column contains the amino acid type of a certain methyl group resonance. The second column contains the peak ID which is typically an integer. The third, fourths, and sixths column contains the respective additional experimental parameters (i.e. restraints). Here, column three contains PCSs induced by Ce^{3+}, column four PCSs induced by Eu^{3+}, and column five PCSs induced by Sa^{3+}. additional_xtal.txt (b.ii) is a text file containing theoretically determined additional restraints corresponding to the measured ones shown in (b.i). The first column contains the amino acid type of a certain methyl group from the structural model, the second one the corresponding residue number. Columns three, four, five, and six are for the theoretically determined additional restraints. The order of the additional restraints must match the order in additional_measure.txt (b.i). In this example there is no fourths set of additional restraints, so column six only contains “999” (b.i and b.ii). Also use “999” in case you haven’t measured or theoretically determined an additional restraint for a certain methyl group (b.i and b.ii). Thereby, AMIGO replaces the value with a punishment value (Eq. 27). If you don’t have additional restraints at all, use a single line containing “0 0 0 0 0 0” as the only content of additional_measure.txt and additional_xtal.txt. (b.iii) shows the text file noe.txt which is an NOE list. The first column contains the amnio acid type of a certain donor methyl group resonance, the second one its peak ID. The third and fourth columns contain the amino acid type and peak ID of the acceptor methyl group resonance. The list must be **sorted** using column two to allow AMIGO to create *NOE graph-based building blocks* (c.f. chapter 5.1.2). E.g., sorting can be done using Excel. AMIGO gives an error message if the last line of this file is empty, so there **must not be a line break at the end of the list**. (b.iv) shows the text file preassignments.txt. The first and second columns contain the amino acid type and peak ID of a certain methyl group resonance. The third and fourth columns contain the amino acid type and residue number of the methyl group the resonance is assigned to. If you don’t have pre-assignments, leave this text file empty. See also Note S 1.3 for more information about utilization of pre-assignments.

Note S 1.2: Using an ensemble of structural models as input files.

Multiple structures can be considered by AMIGO by aligning the structures and providing the coordinates to AMIGO in a single .pdb file.

Note S 1.3: Include pre-assignments in AMIGO

Pre-assignments can be included in the calculations as shown in Fig. S 1.2 b iv). AMIGO will then in a first step only consider the pre-assigned resonances as starting points to create the methyl walks. Afterwards, all remaining methyl group resonances are considered.

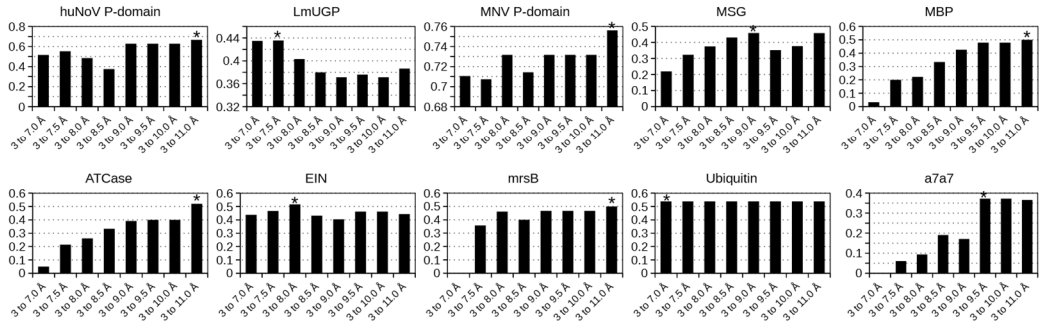


Fig. S 1.3: Test runs for determination of optimal cut-off distance range for the benchmark set without additional restraints.

The range of cut-off distances was optimized for each protein. Ratios of assigned methyl groups to methyl groups were the structure building block and the NOE building block matched perfectly are plotted on the y-axis. The highest value for each protein is highlighted with an asterisk. The actual runs were performed using the highlighted ranges (see also 4.1.1). The step size from the minimal to maximal cut-off distance was 0.2 Å.

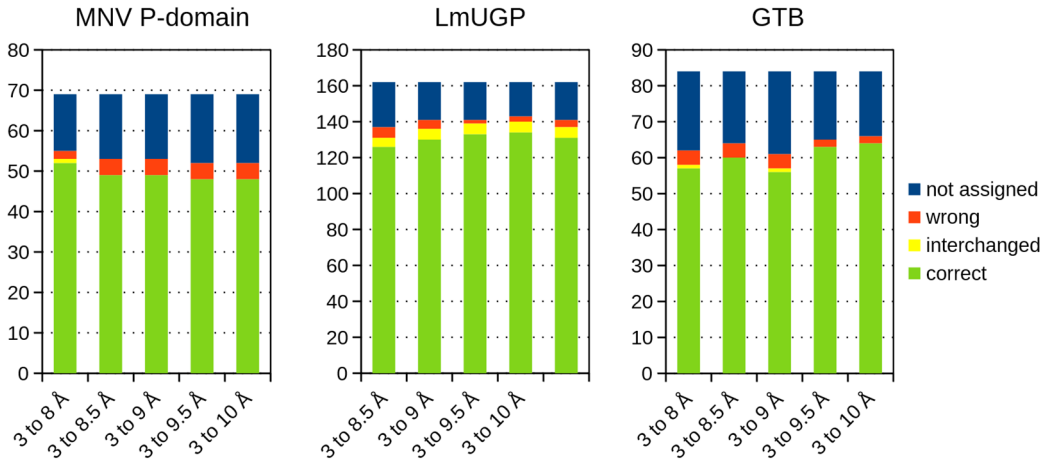
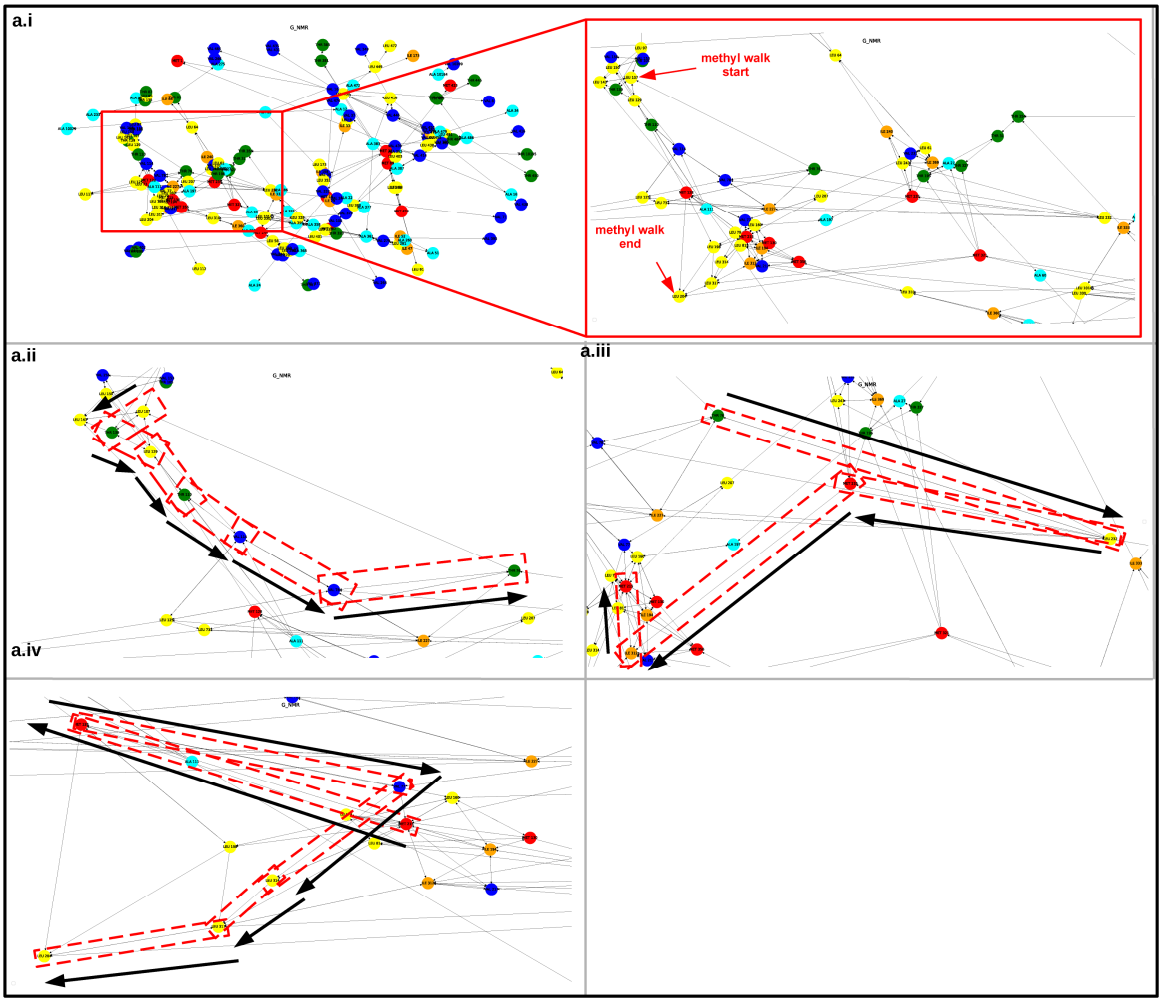


Fig. S 1.4: Running AMIGO using additional restraints makes it robust to variation of cut-off distance ranges.

The number of methyl group assignments is plotted on the y-axis. Using additional restraints makes the ratio more robust against the alteration of the cut-off distance range (see chapters 5.1.3, 4.1.2, and 4.1.3 for more explanations).

1.2 Outputs generated by AMIGO to guide the assignment process



b

result.txt

```
570 was extened with methyl group resonance  :     LEU 85  with NOEs to resonances THR 135 / THR 87 /
571 and thereby assigned to pdb methyl group :     LEU 85 with expected NOEs to   THR 87 / THR 135 /  applying a cut-off distance of 6.6000000000000005 Angstroem
572
573
574
575 methyl walk (NMR//pdb)                  :     start => LEU 107 // LEU 107 => LEU 97 // LEU 97 => ALA 94 // ALA 94 => THR 135 // THR 135 => LEU 85 // LEU 85
576 was extened with methyl group resonance  :     THR 87  with NOEs to resonances LEU 85 /
577 and thereby assigned to pdb methyl group :     THR 87 with expected NOEs to   LEU 85 /  applying a cut-off distance of 4.800000000000001 Angstroem
578
579
580
581 methyl walk (NMR//pdb)                  :     start => LEU 107 // LEU 107 => THR 139 // THR 139 => LEU 129 // LEU 129 => THR 110 // THR 110 => VAL 114 // VAL 114 => VAL 78 // VAL 78 => THR 76 //
    THR 76 => LEU 232 // LEU 232 => ILE 227 // ILE 227 => LEU 125 // LEU 117
582 was extened with methyl group resonance  :     LEU 73  with NOEs to resonances ILE 227 / THR 76 / LEU 117 /
583 and thereby assigned to pdb methyl group :     LEU 73 with expected NOEs to   THR 76 / LEU 117 / ILE 227 /  applying a cut-off distance of 8.6 Angstroem
584 additional experimental restrains are [additional restrain 1, 2, 3 and 4]    :[-0.013, -0.009, 0.001, 999.0]
585 additional theoretical restrains are [additional restrain 1, 2, 3 and 4]     :[-0.01233, -0.00949, 0.00258, 999.0]
586
587
588
589 methyl walk (NMR//pdb)                  :     start => LEU 107 // LEU 107 => THR 139 // THR 139 => LEU 129 // LEU 129 => THR 110 // THR 110 => VAL 114 // VAL 114 => VAL 78 // VAL 78 => THR 76 //
    THR 76 => LEU 232 // LEU 232 => MET 235 // MET 235 => ILE 312 // ILE 312 => MET 215 // MET 215 => MET 128 // MET 128 => VAL 77 // VAL 77
590 was extened with methyl group resonance  :     LEU 314  with NOEs to resonances MET 215 / LEU 317 /
591 and thereby assigned to pdb methyl group :     LEU 314 with expected NOEs to   MET 215 / LEU 317 /  applying a cut-off distance of 4.800000000000001 Angstroem
592
593
594
595 methyl walk (NMR//pdb)                  :     start => LEU 107 // LEU 107 => THR 139 // THR 139 => LEU 129 // LEU 129 => THR 110 // THR 110 => VAL 114 // VAL 114 => VAL 78 // VAL 78 => THR 76 //
    THR 76 => LEU 232 // LEU 232 => MET 235 // MET 235 => ILE 312 // ILE 312 => MET 215 // MET 215 => MET 128 // MET 128 => VAL 77 // VAL 77 => LEU 314 // LEU 314
596 was extened with methyl group resonance  :     LEU 317  with NOEs to resonances ILE 312 / LEU 204 /
597 and thereby assigned to pdb methyl group :     LEU 317 with expected NOEs to   LEU 204 / ILE 312 / LEU 314 / MET 321 /  applying a cut-off distance of 6.2 Angstroem
598 additional experimental restrains are [additional restrain 1, 2, 3 and 4]    :[999.0, 0.031, 0.01, 999.0]
599 additional theoretical restrains are [additional restrain 1, 2, 3 and 4]     :[999.0, 0.0258, 0.00357, 999.0]
600
601
602
603 methyl walk (NMR//pdb)                  :     start => LEU 107 // LEU 107 => THR 139 // THR 139 => LEU 129 // LEU 129 => THR 110 // THR 110 => VAL 114 // VAL 114 => VAL 78 // VAL 78 => THR 76 //
    THR 76 => LEU 232 // LEU 232 => MET 235 // MET 235 => ILE 312 // ILE 312 => MET 215 // MET 215 => MET 128 // MET 128 => VAL 158 // VAL 158 => VAL 153 // VAL 153 => LEU 150 // LEU 150
604 was extened with methyl group resonance  :     LEU 143  with NOEs to resonances LEU 107 /
605 and thereby assigned to pdb methyl group :     LEU 143 with expected NOEs to   LEU 107 / ALA 111 /  applying a cut-off distance of 5.600000000000005 Angstroem
606 additional experimental restrains are [additional restrain 1, 2, 3 and 4]    :[999.0, 0.017, 0.017, 999.0]
607 additional theoretical restrains are [additional restrain 1, 2, 3 and 4]     :[999.0, 0.01813, 0.01836, 999.0]
608
609
610
611 methyl walk (NMR//pdb)                  :     start => LEU 107 // LEU 107 => THR 139 // THR 139 => LEU 129 // LEU 129 => THR 110 // THR 110 => VAL 114 // VAL 114 => VAL 78 // VAL 78 => THR 76 //
    THR 76 => LEU 232 // LEU 232 => MET 235 // MET 235 => ILE 312 // ILE 312 => MET 215 // MET 215 => MET 128 // MET 128 => VAL 77 // VAL 77 => LEU 314 // LEU 314 => LEU 317 // LEU 317
612 was extened with methyl group resonance  :     LEU 204  with NOEs to resonances LEU 317 /
613 and thereby assigned to pdb methyl group :     LEU 204 with expected NOEs to   LEU 317 /  applying a cut-off distance of 3.8000000000000003 Angstroem
614 additional experimental restrains are [additional restrain 1, 2, 3 and 4]    :[999.0, 0.021, 0.008, 999.0]
615 additional theoretical restrains are [additional restrain 1, 2, 3 and 4]     :[999.0, 0.02028, 0.00536, 999.0]
616
617 amount of assignments in cluster                : 84
618 amount of perfectly matching building block pairs : 36
619 Cluster Score  :      193.2877503645464 of 606.0
```

Methyl walk used to assign Leu204

Fig. S 1.5: Using AMIGO output for guiding methyl walk assisted assignment.

Experimental (a) and structure-based (c, figure continued on the next page) NOE networks are plotted by AMIGO. (b) shows the text file created by AMIGO for the LmUGP data set. The red box highlights, how Leu204 was assigned. The methyl group resonance showed an NOE to a methyl group resonance corresponding to another leucine methyl group. Applying a cut-off distance of 3.8 Å, an NOE to another leucine methyl group (LEU 317) is expected based on the high-resolution structural model (pdb 4m2a). Theoretically and experimentally determined additional restraints are in very well agreement. Note that the value "999" is a placeholder for values that are not available (c.f. Fig. S 1.2). The methyl walk can be followed and comprehended by the user in the reconstructed *NOE-based* (a.i to a.iv) and s*tructure-based graphs* (c.i to c.v) plotted by AMIGO. Red arrows in (b) highlight further values given by AMIGO. These values are the total number of methyl group resonances assigned within in a cluster based on a certain methyl walk, the number of assignments where the respective NMR building block exactly matched the structure-based building block (important for parameter optimization, see chapter 4.1.1), and the *total score* of the respective cluster (see above and chapter 5.1.2).

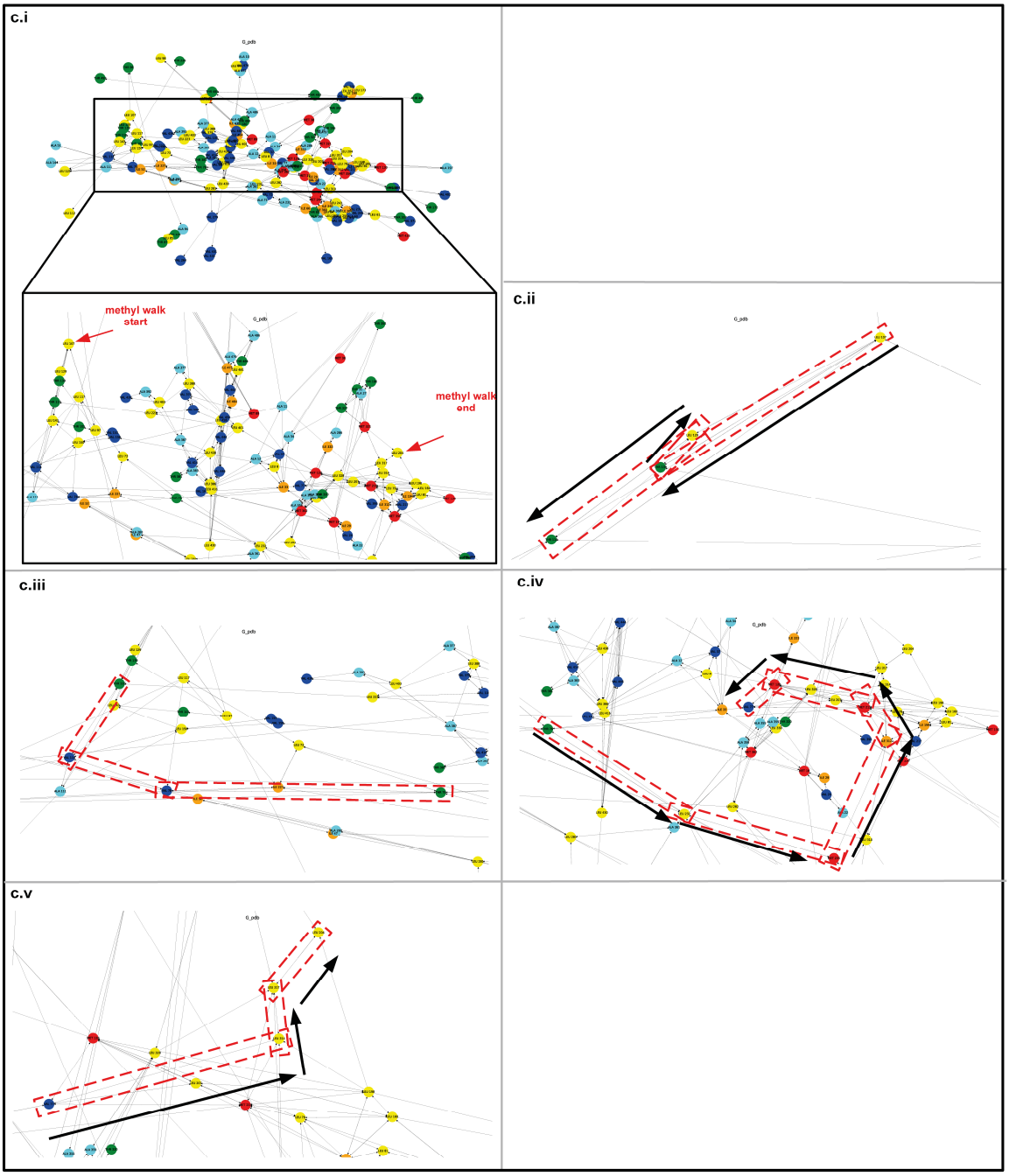


Fig. S 1.5 continued

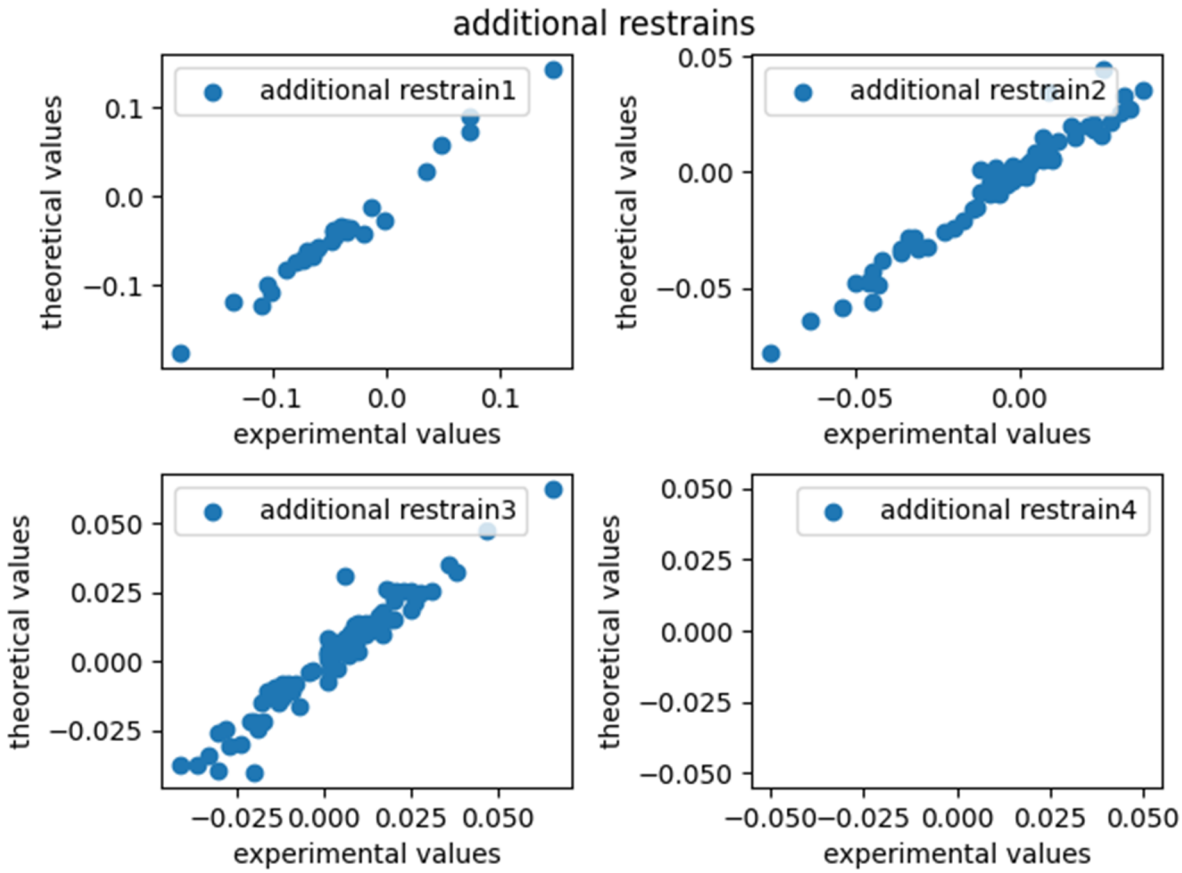


Fig. S 1.6: Example for assignment validation using experimentally and theoretically determined additional restraints (PCSs) for the LmUGP data set.

For good assignment solutions, experimental and theoretical values should follow a linear correlation. The figures are automatically generated by AMIGO. Note that the weighting factor for "additional restrain1" was set to 0. Thereby, the values were not included in the calculation (c.f. chapter 4.1.2). However, plotting the data shows that measured and theoretically determined values are in good agreement, validating the assignment independently.

1.3 Reference assignments

Tab. S 1.1: Reference assignments for the tested data sets.

protein	references for data and assignment	pdb id of structural model
mrsB	Lange et al., 2012; Pritišanac et al., 2017	3e0o (Kim et al., 2009)
Ubiquitin	Chao et al., 2012, 2014; Pritišanac et al., 2017	1ubq (Vijay-Kumar et al., 1987)
EIN	Venditti et al., 2011; Pritišanac et al., 2017	1eza (Garrett et al., 1997)
MBP	Lange et al., 2012; Pritišanac et al., 2017	1ez9 (Duan and Quiocho, 2002)
msg	Tugarinov et al., 2005; Tugarinov and Kay, 2003; Pritišanac et al., 2017	1y8b (Tugarinov et al., 2005)
ATCase	Velyvis et al., 2009; Pritišanac et al., 2017	1d09 (Jin et al., 1999)
a7a7	Tugarinov et al., 2007; Pritišanac et al., 2017	1yau (Förster et al., 2005)
HuNoV P-domain	Müller-Hermes et al., 2020**	4x06 (Singh et al., 2015)*
MNV P-domain	Maass et al., 2022a; chapter 5.2	6e47 (Nelson et al., 2018)
LmUGP	Mühlberg et al., 2022***	4m2a (Führing et al., 2013)
GTB	Flügge and Peters, 2018; Flügge, 2018	2rit (Alfaro et al., 2008)****

*structure was minimized using Maestro by Dr. Alvaro Mallagaray
**NOE lists and assignments were provided by Dr. Alvaro Mallagaray
***Assignments, NOE lists, and PCSs lists were provided by Lars Mühlberg and Dr. Alvaro Mallagaray prior to publication. The data could have been subject to change.
****structure was modified by Dr. Friedemann Flügge as described in Flügge and Peters, 2018

2 CW1 MNV P-domain [^{1}H,^{13}C] MIL^{proS}V^{proS}A methyl group NMR assignment

2.1 Assignments based on the 4D HMQC-NOESY-HMQC experiment

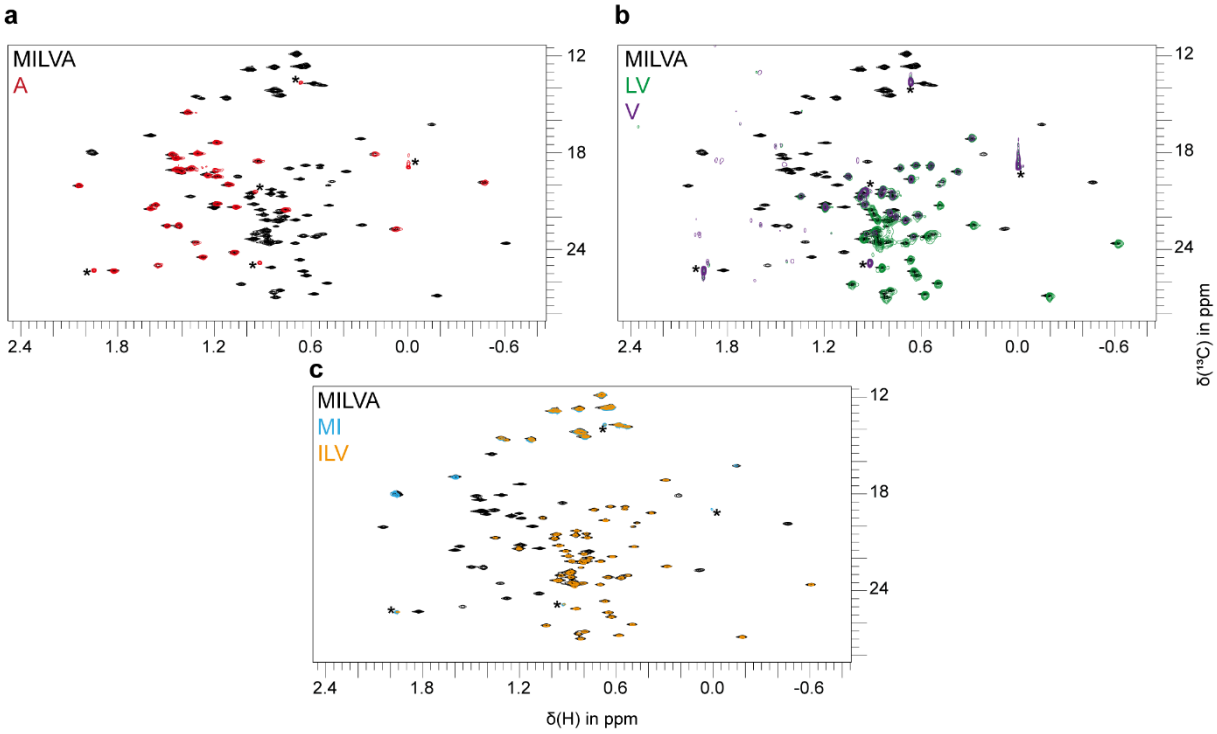


Fig. S 2.1: **Overlay of [^{1}H,^{13}C] HMQC spectra of methyl group labeled MNV P-domain samples.**

Overlay of spectra of a MILVA labeled MNV P-domain (500 µM) (a-c) with A-labeled (7 µM) (a), with LV-labeled (4 µM) and V-labeled (2 µM) (b), and an ILV-labeled (50 µM) and a MI-labeled (40 µM) MNV P-domains (c). The overlay allows counting the number of methyl group resonances for each amino acid type: 4 Met, 16 Ile, 28 Leu, 26 Val, and 29 Ala. All spectra except the one for the MI-labeled sample were recorded on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K. The MI-labeled sample was measured on a Bruker 500 MHz Avance III spectrometer equipped with a TCI cryogenic probe. Asterisks mark signals from GCDCA and acetate. All samples contained saturating amounts of GCDCA. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/). The spectra of A, LV, V, and MI labelled P-domains were generated during my master thesis (Maaß, 2019).

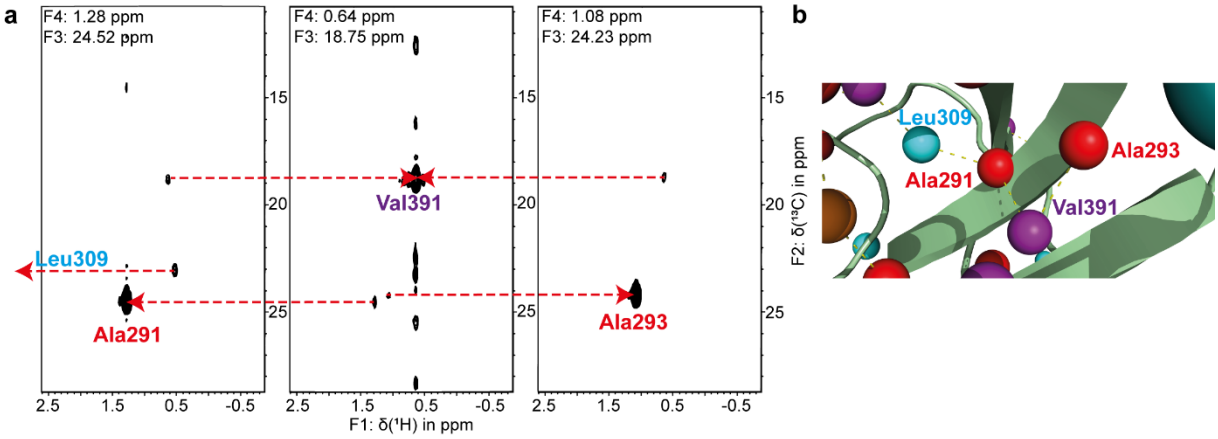


Fig. S 2.2: Assignment of Val391 using methyl-methyl NOEs.

Unambiguous assignment of Val391 exclusively based on methyl-methyl NOEs was essential for assigning Val352, where two peaks disappeared upon V352L single point mutation, one of the peaks allocated to the methyl group of Val352 and the other peak belonging to the Val391 methyl group (cf. Fig. S 2.3i). The methyl walk procedure and experimental details are explained and given in the legend of Fig. 5.2.2. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

2.2 Assignments by site directed mutagenesis and metal ion titration reveals Pro361 cis/trans configuration

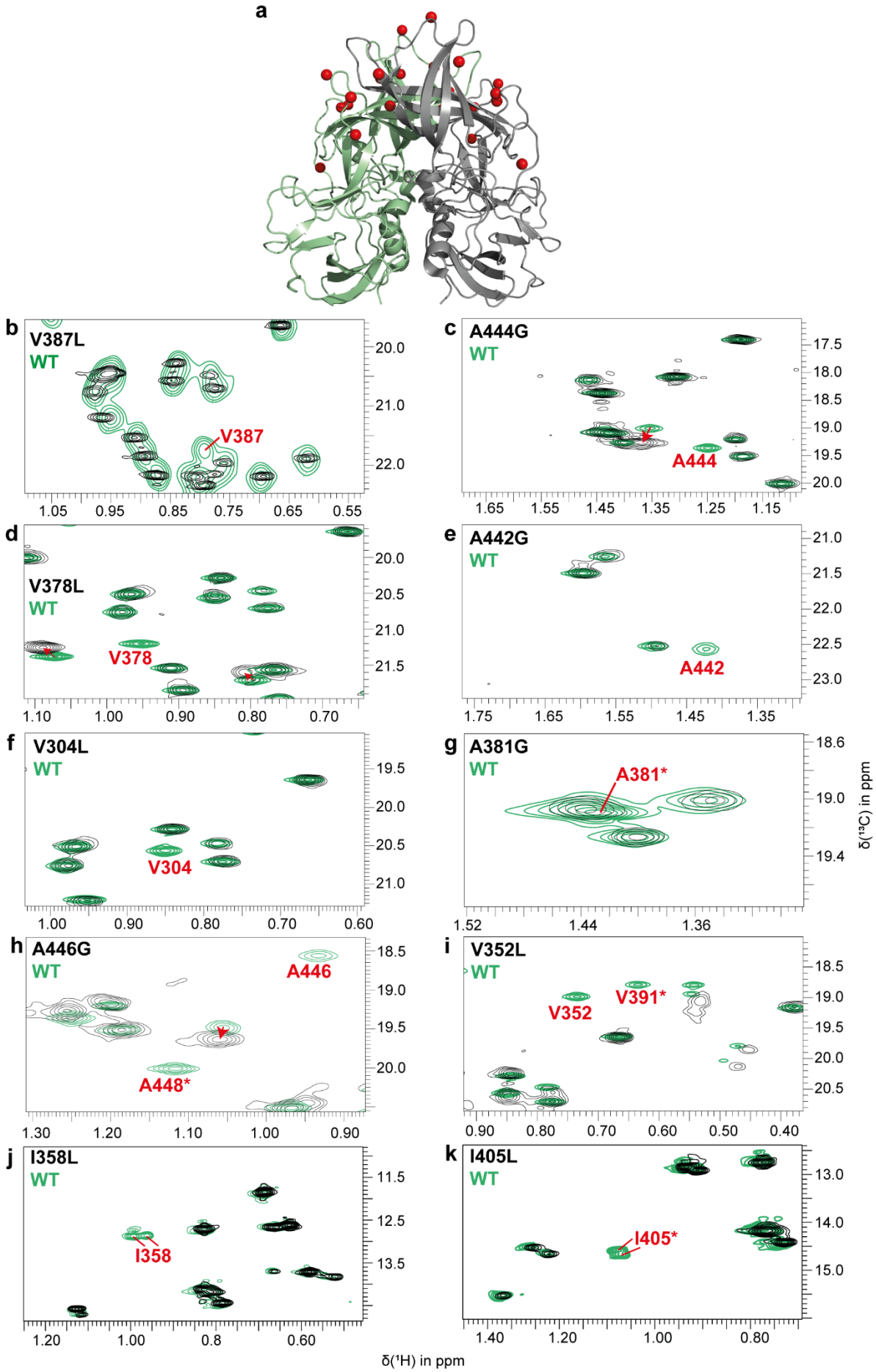

a
b
V387L
WT
V387
c
A444G
WT
A444
d
V378L
WT
V378
e
A442G
WT
A442
f
V304L
WT
V304
g
A381G
WT
A381*
h
A446G
WT
A446
A448*
i
V352L
WT
V352
V391*
j
I358L
WT
I358
k
I405L
WT
I405*
δ(¹³C) in ppm
δ(¹H) in ppm

Fig. S 2.3: [^{1}H,^{13}C] HMQC spectra of the MNV P-domain and selected point mutants.

Overlay of methyl [^{1}H,^{13}C] HMQC of WT (wild type) MNV P-domain (green) with spectra of point mutants (black) yields an assignment of the mutated side chains. All spectra were obtained in the presence of saturating amounts of GCDCA to shift the monomer-dimer equilibrium exclusively towards the dimers. (a) Single point mutations mapped onto the crystal structural model of MNV P-dimers (pdb 6e47). (b-f)°Comparison of spectra yielded the unambiguous assignment of methyl group resonances of V387, A444, V378, A442, and V304. (g) The methyl group of A381 had already been assigned via the methyl walk (see Fig. 5.2.2). Therefore, the rather similar cross peak patterns of the WT P-domain and of its A381G mutant imply overlap with another resonance. (h, i) For A446G and V352L point mutants two resonances disappeared for each mutant. In both cases, the additional resonance disappearing had been assigned using methyl walks and PCSs (see also Fig. S 2.7 for assignment of A446 and A448 and Fig. S 2.2 for assignment of Val391), i.e., in the case of A446G the second peak disappearing belongs to A448, and in the case of V352L the second peak is allocated to V391. Most likely, these mutations are associated with altered protein dynamics leading to exchange broadening and thus disappearance of the respective resonances. Therefore, the remaining disappearing resonances were assigned to A446 and V352. (j) For I358L two resonances disappeared upon mutation. I358 is in the loop that changes its orientation depending on whether P361 is *cis* or *trans*. Therefore, both resonances were assigned to I358, reflecting the presence of a P361 *cis-trans* mixture (c.f. Fig. 5.2.3d, showing two peaks disappearing upon mutation of A365 in the same loop). (k) I405 is located in a loop adjacent to the one containing A365 and had already been assigned using the methyl walk. The corresponding cross peak consists of two peaks, both disappearing in the I405L mutant and thus indicating that this loop is also affected by the P361 *cis-trans* isomerism. All peaks labeled with stars had been assigned using methyl walks or PCSs. Red arrows assist backtracking signals showing CSPs upon mutation. (b) The concentration of LV-labeled WT P-domain was 4 µM. The concentration of all other WT P-domain samples was 500 µM (cf. Fig. 5.2.2). P-domain concentrations of mutants were: 9 µM LV-labeled V387L (b); a mixture of 115 µM each of A-labeled A442G, LV-labelled V304L, and I-labelled I405L (e,f,k); a mixture of 60 µM each of LV-labeled V352L and A-labeled A446G (h,i); a mixture of 75 µM each of A-labeled A444G and I-labeled I358L (c,j); a mixture of 67 µM each of LV-labelled V378L and A-labeled A381G (d,g). All spectra were acquired at 298 K on a Bruker Avance III HD 600 MHz NMR spectrometer equipped with a cryogenic probe. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

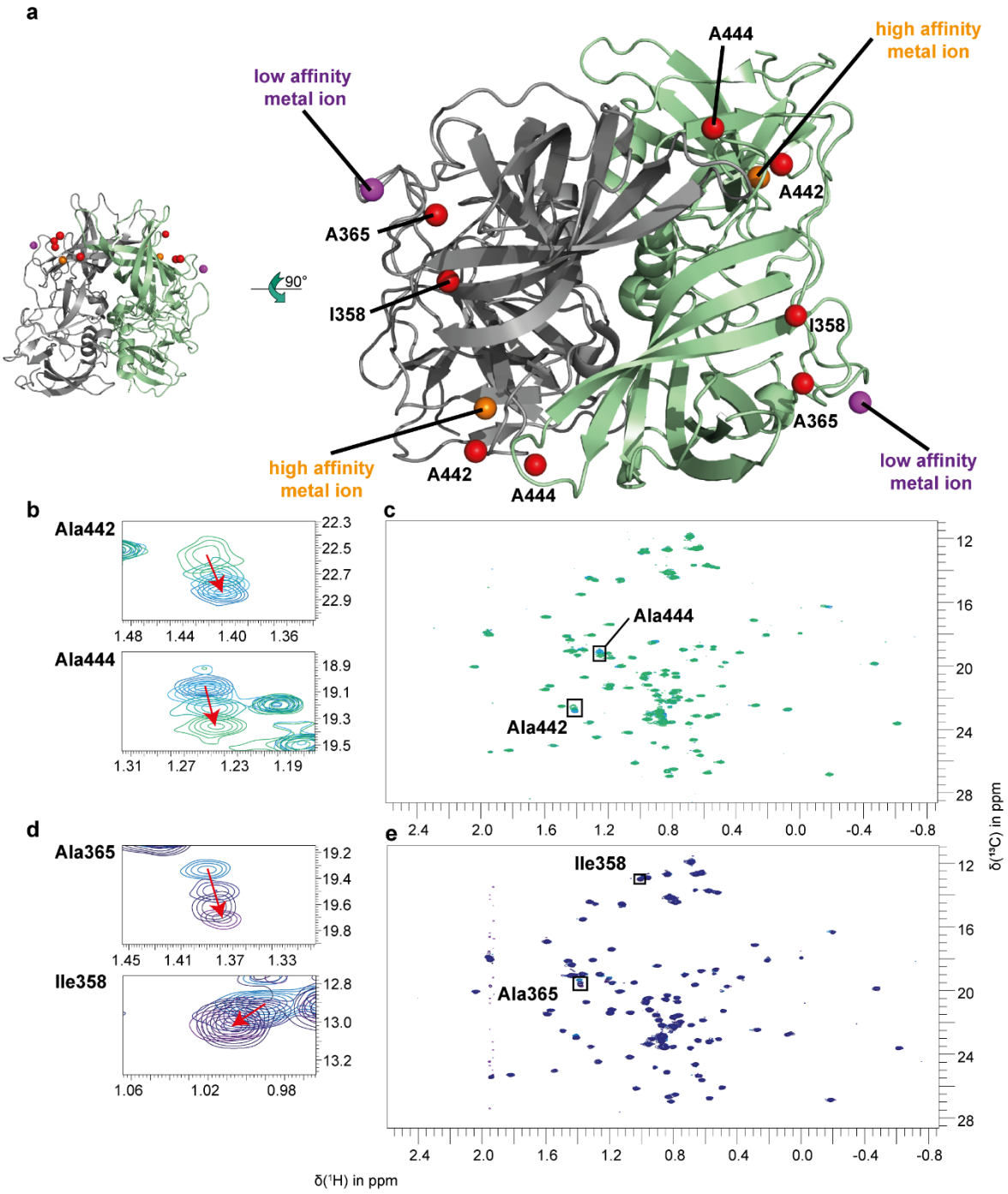


Fig. S 2.4: $CaCl_2$ titration attests two Ca^{2+} binding sites per monomer.

Titration of a P-domain sample with $CaCl_2$ exposes two distinct metal ion binding sites. (b,c) One set of peaks is characterized by CSPs going into saturation at about 550 µM $CaCl_2$ (green - blue, 0 - 550 µM $CaCl_2$). (e,d) A second set of peaks approaches saturation at about 20 mM $CaCl_2$ (blue - magenta, 0.55 - 20 mM $CaCl_2$). A comparison with crystal structure data (a) (pdb 6e47; Nelson et al., 2018) reveals that Ala444 and Ala442 (b,c) belong to the first, higher-affinity binding site. Ala365 and Ile358 (e,d) belong to the lower-affinity site. (a) Location of the methyl groups in the crystal structure relative to the position of the two metal ions. Spectra were acquired with 38 µM MILVA-labeled P-domain at 298

K on a 600 MHz spectrometer with cryo probe in the presence of saturating amounts of GCDCA. For assignment of resonances see Fig. 5.2.3, Fig. S 2.3 , and Fig. S 2.5. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

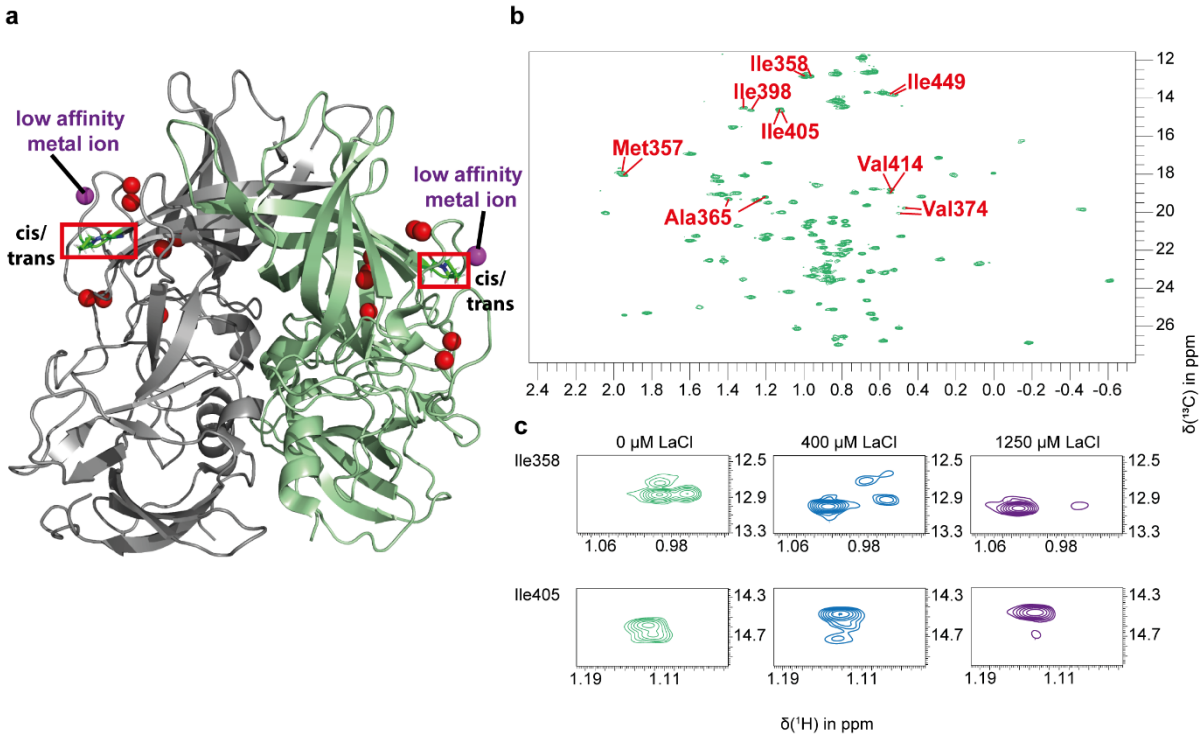


Fig. S 2.5: Titration with La^{3+} discloses two distinct forms of the MNV P-domain.

Titration with $LaCl_3$ results in different behavior (c) of methyl groups that exhibit double cross peaks in [^{1}H,^{13}C] HMQC spectra (cf. Fig. S 2.3) due to the presence of a mixture of *cis-trans* isomers of P361. (a) Methyl groups giving rising to double cross peaks mapped on to the crystal structure (pdb 6e47, Nelson et al., 2018). (b) [^{1}H,^{13}C] HMQC spectrum of MNV P-domain with resonances showing up as double cross peaks labeled. P-domain concentration was 38 µM (b, c) and spectra were acquired at 298 K on 600 MHz spectrometer with cryo probe in the presence of saturating amounts of GCDCA. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

2.3 Assignments using pseudo contact shifts

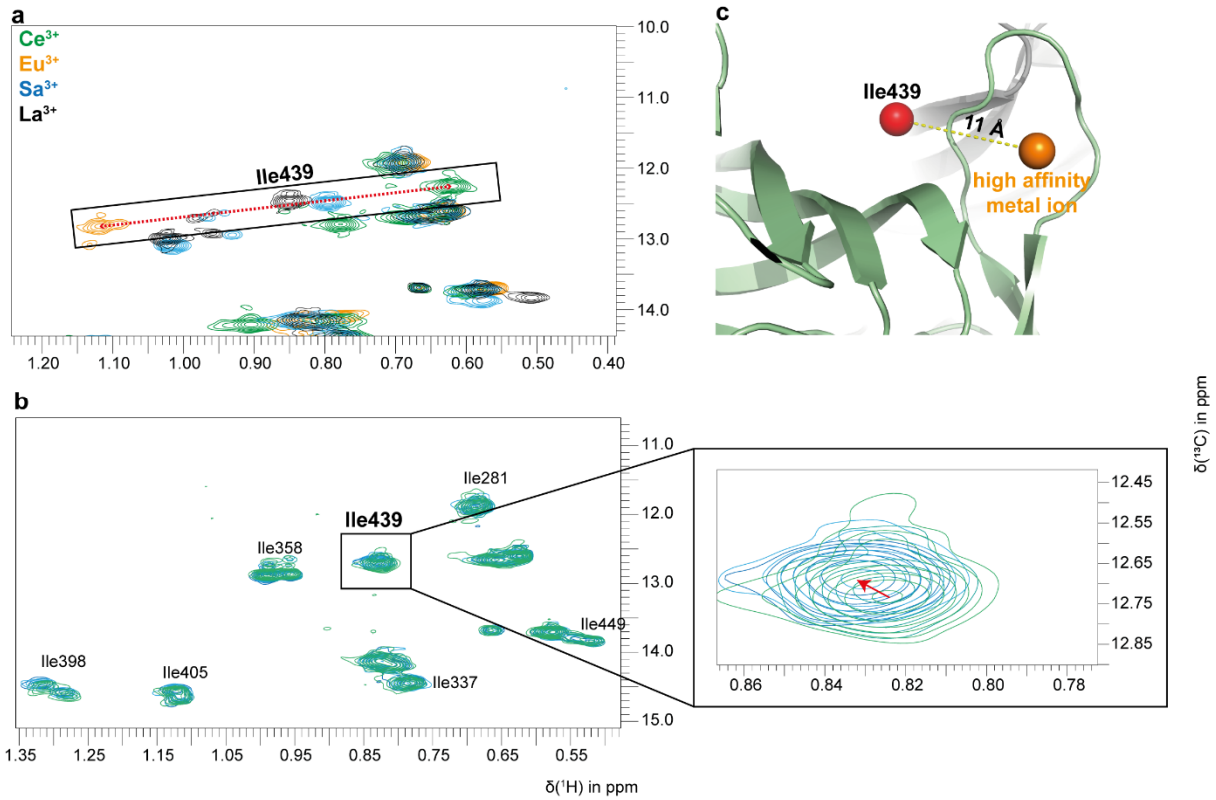


Fig. S 2.6: Assignment of Ile439.

(a) Section of a [^{1}H,^{13}C] HMQC spectrum of MILVA-labeled MNV P-domain. The box encompasses PCSs of an unassigned Ile residue in the presence of 450 µM $EuCl_3$ (yellow), 400 µM $SaCl_3$ (blue), or 400 µM $CeCl_3$ (green). 400 µM $LaCl_3$ was used as diamagnetic reference (black). (b) The unassigned Ile signal also shifts upon titration with $CaCl_2$ (cf. Fig. S 2.4) and belongs to the higher affinity metal ion binding site. (c) Within a sphere with a radius of 20 Å around the metal ion (orange) occupying the higher affinity site there is only one Ile residue that has not yet been assigned, Ile 439. Experimental conditions are as given in the legend to Fig. 5.2.4 and Fig. S 2.4. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

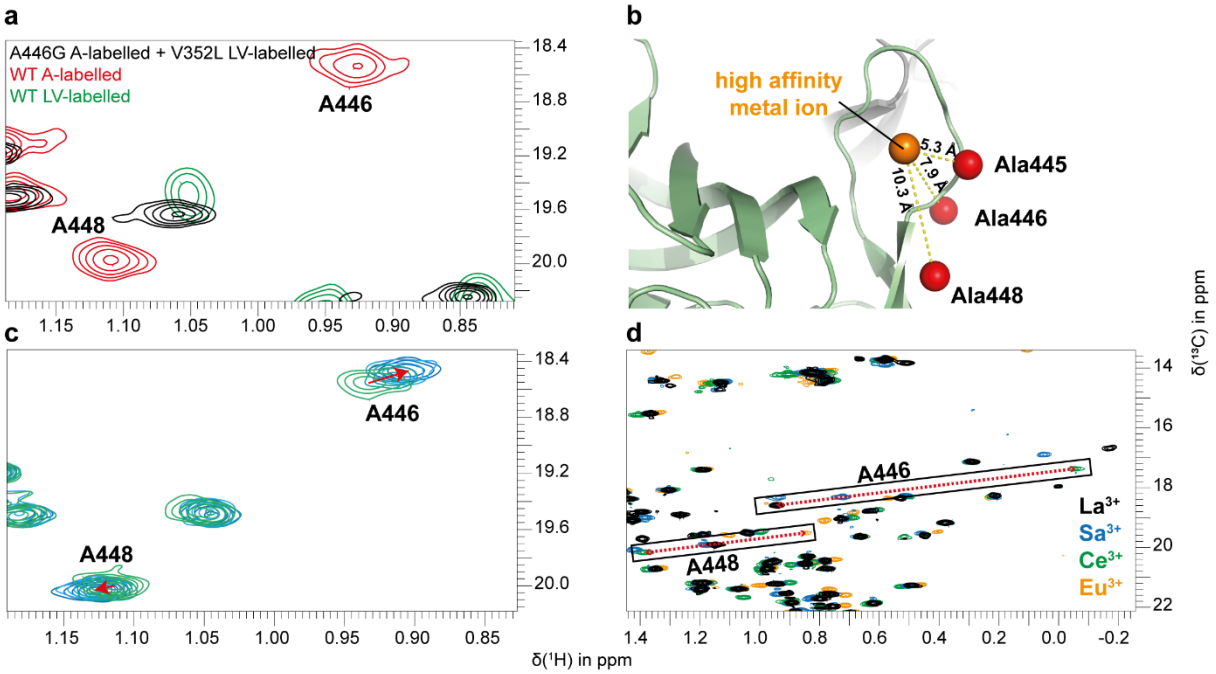
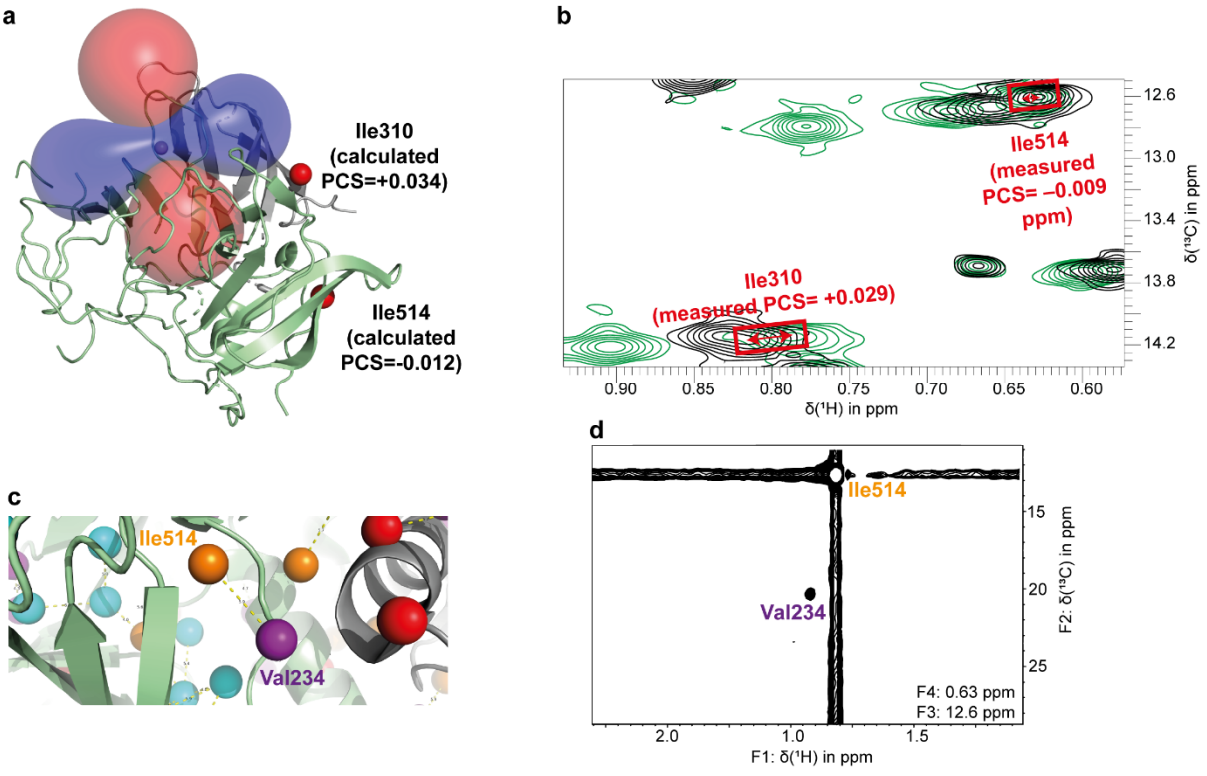


Fig. S 2.7: **Assignment of Ala446 and Ala448.**

(a) [^{1}H,^{13}C] HMQC spectra of A-labelled mutant A446G and of LV-labelled mutant V352L overlaid with the corresponding A- and LV-labeled spectra of WT (wildtype) P-domains. For the conclusions here, only the A-labeled spectra are relevant. The LV-labeled sample has been measured together with the A-labeled sample to save measurement time. Upon the A446G mutation two resonances disappear, preventing an unambiguous assignment. Both peaks belong to the lower affinity metal ion binding site shown in the crystal structure model in (b) (pdb 6e47; Nelson et al., 2018). Remaining unassigned methyl groups at this site are highlighted as red spheres. (c) Shows the shifts of the two peaks upon titration with $CaCl_2$ (cf. Fig. S 2.4). (d) Comparison of experimental and theoretical PCSs allows to discriminate the two peaks, leading to the assignment shown. Calculated PCSs induced by Ce^{3+} (second round of refinement, cf. Tab. 5.1) for Ala446, Ala448 and Ala445 are 0.776, -0.237 and 1.651, respectively. The two resonances disappearing due to mutation show PCSs of -0.237 and 0.997, allocating them to Ala446 and Ala448. Sample conditions in (a) were 4 µM LV-labelled wildtype P-domain, 7 µM A-labelled wildtype P-domain and a mixture of 60 µM A-labelled A446G and LV-labelled V352L P-domain. For the experimental conditions in (c) and (d) see the legend of Fig. S 2.4 and Fig. 5.2.4. All spectra were acquired in the presence of saturating amounts of GCDCA. Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).



(a)

Fig. S 2.8: Assignment of Ile514, Ile310 and Val234.

(a) Isosurfaces of the alignment tensor in the presence of Ce^{3+} from the second round of refinement (See Fig. 5.2.5 and Tab. 5.1). Calculated PCSs (based on crystal structure pdb 6e47) are annotated with the only remaining unassigned Ile residues, Ile310 and Ile514. (b) Comparison to experimental PCSs (c.f. Fig. 5.2.4) of two unassigned Ile peaks yields an unambiguous assignment of Ile310 and Ile514. Supporting evidence is found from methyl-methyl NOEs. As methyl groups of Ile514 and Val234 are in close vicinity to each other (c), an NOE is expected and indeed found experimentally in the corresponding F1-F2 plane of the 4D HMQC-NOESY-HMQC spectrum (cf. Fig. 5.2.2) positioned at the methyl resonances of Ile514 in F3 and F4 (d). Figure and legend are adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

Tab. S 2.1: Parameters of alignment tensors of Ce^{3+}, Eu^{3+} and Sa^{3+} using different structural processing approaches (see Fig. 5.2.6).

The table is adapted from Maass et al., 2022a, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

Metal ion		Ce^{3+}			Eu^{3+}			Sa^{3+}			Mg^{2+} (pdb 6e47)
Processing conditions		$\|d_1\text{-}d_2\|$>0 Å (Fig. 5.2.6 (b.i))	$\|d_1\text{-}d_2\|$>2.5 Å (Fig. 5.2.6 (c.i))	$\|d_1\text{-}d_2\|$>5 Å (Fig. 5.2.6 (d.i))	$\|d_1\text{-}d_2\|$>0 Å (Fig. 5.2.6 (b.i))	$\|d_1\text{-}d_2\|$>2.5 Å (Fig. 5.2.6 (c.i))	$\|d_1\text{-}d_2\|$>5 Å (Fig. 5.2.6 (d.i))	$\|d_1\text{-}d_2\|$>0 Å (Fig. 5.2.6 (b.i))	$\|d_1\text{-}d_2\|$>2.5 Å (Fig. 5.2.6 (c.i))	$\|d_1\text{-}d_2\|$>5 Å (Fig. 5.2.6 (d.i))	-
ΔX_{ax} in 10^{-32} m^3		-1.60 ±0.05	-1.60 ±0.03	-1.58±0.03	1.17±0.05	1.19±0.05	1.29±0.10	-0.29±0.03	-0.29±0.04	-0.29±0.05	-
ΔX_{rh} in 10^{-32} m^3		-0.74 ±0.04	-0.76 ±0.03	-0.83±0.04	0.66±0.14	0.68±0.14	0.67±0.11	-0.15±0.02	-0.16±0.03	-0.16±0.03	-
Coordinates of origin in Angstroem	X	-16.3 ±0.5	-16.4 ±0.6	-16.6 ±0.5	-15.6 ±0.7	-15.6 ±0.8	-16.7 ±1.2	-16.5 ±0.5	-16.5 ±0.5	-16.5 ±0.4	-14.9
	Y	2.9 ±0.4	2.9 ±0.4	2.8 ±0.3	3.8 ±1.6	3.8 ±1.6	3.2 ±1.6	2.9 ±0.5	2.9 ±0.5	2.9 ±0.5	2.0
	z	-45.3 ±0.7	-45.2 ±0.7	-45.1 ±0.6	-44.6 ±1.1	-44.7 ±1.1	-45.2 ±1.1	-45.0 ±1.2	-45.1 ±1.2	-45.0 ±1.1	-46.3
Orinentation of principal axis of tensor in °	α	109.6 ±0.3	109.3 ±0.4	108.0 ±0.3	107.6 ±1.23	107.5 ±1.2	100.0 ±3.0	109.9 ±5.0	109.8 ±6.4	109.3 ±6.9	-
	β	66.8 ±2.4	67.4 ±1.8	70.5 ±2.2	79.1 ±5.6	77.4 ±5.8	72.7 ±3.8	71.4 ±3.7	71.6 ±3.6	72.8 ±3.5	-
	γ	155.9 ±2.0	156.4 ±1.9	157.7 ±2.0	169.8 ±47.8	169.7 ±42.6	170.7 ±22.1	164.3 ±10.2	164.9 ±17.6	165.4 ±12.9	-
Number of methyl groups included		61	52	47	54	45	41	64	55	50	-
Q-factor		0.10	0.09	0.07	0.25	0.20	0.16	0.16	0.16	0.14	-

3 P-domain GCDCA interactions

3.1 Line shape analysis

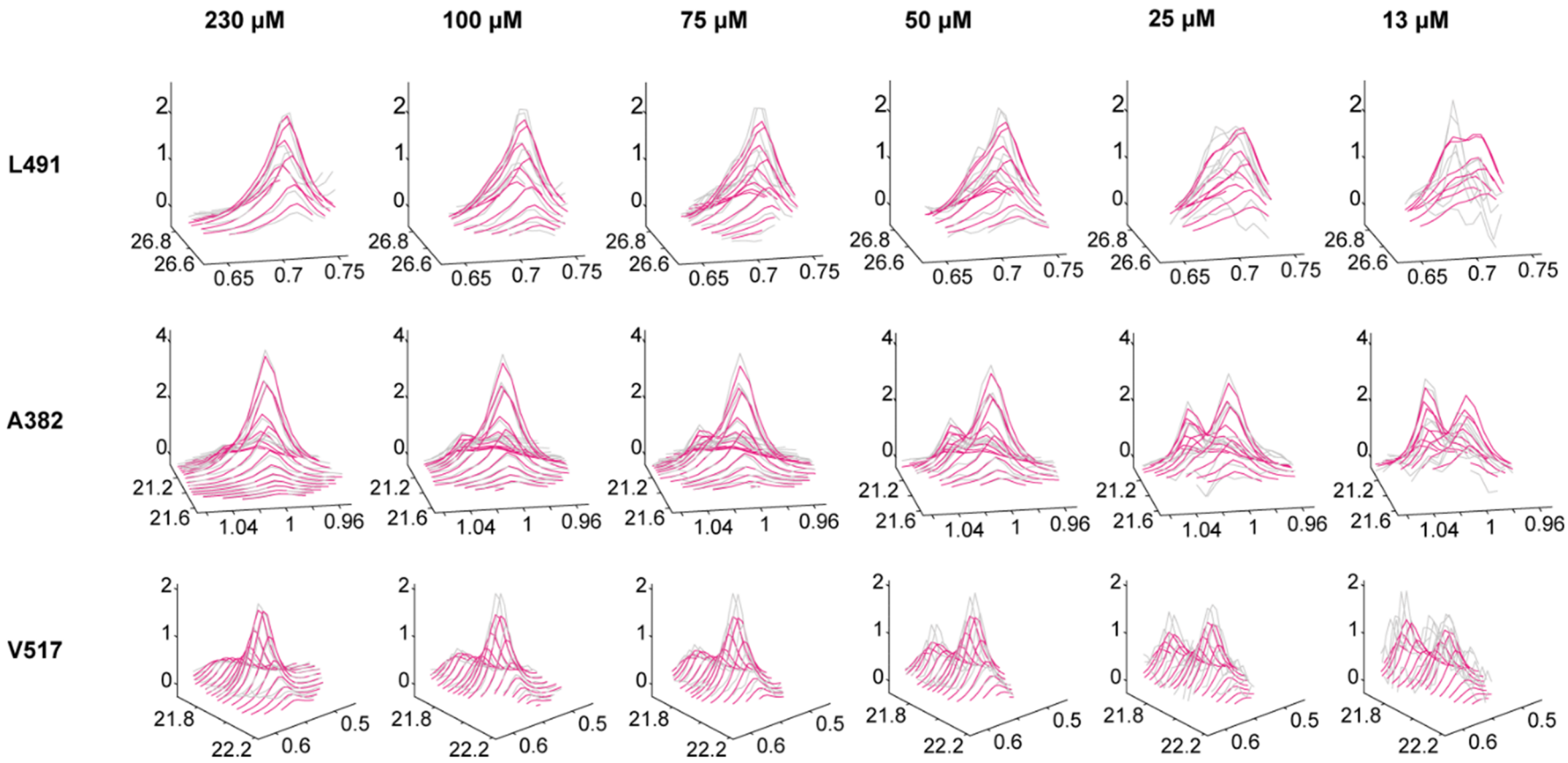


Fig. S 3.1: 2D line shape analysis of [^{1}H,^{13}C] HMQC spectra of MILVA-labeled MNV CW1 P-domain at varying protein concentrations.

Experimental and simulated spectral regions are shown in grey and magenta, respectively. The overlays reflect the good quality of the fit. The analysis results in a dissociation constant $K_{D,Dimer}$ of 6.9 ± 0.6 µM and a dissociation rate constant $k_{off,Dimer}$ of 1.25 ± 0.21 s^{-1}. See Fig. 5.3.2 for more information. Figure and legend are adapted from Creutznacher et al., 2022, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

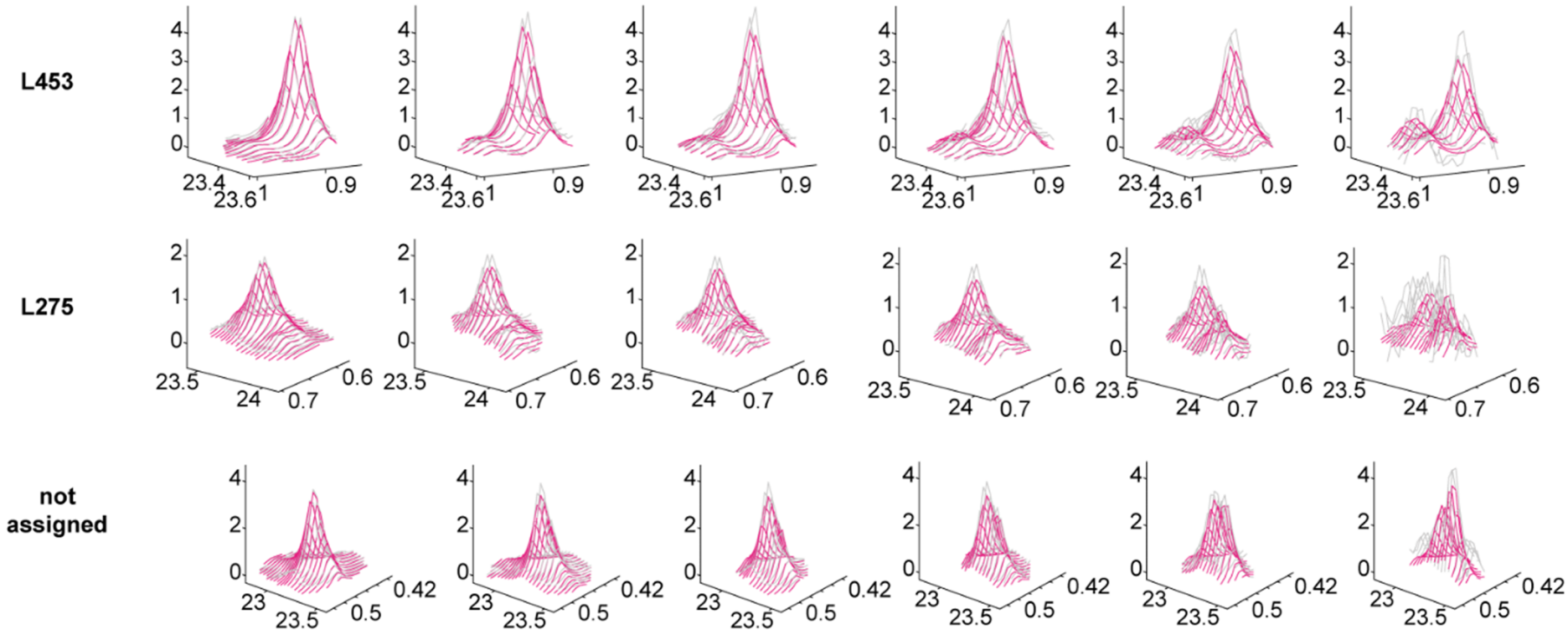


Fig. S 3.1 continued

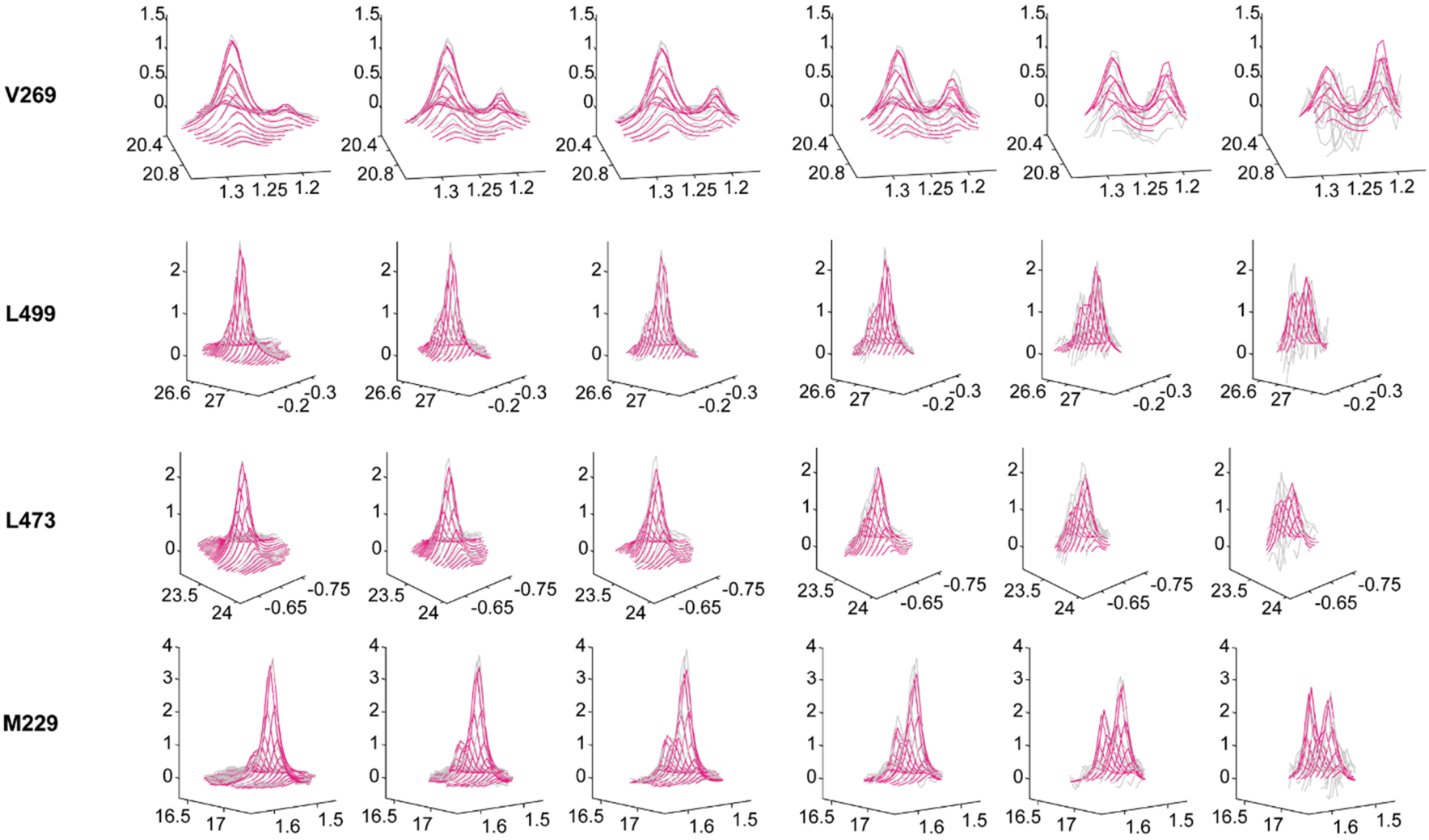
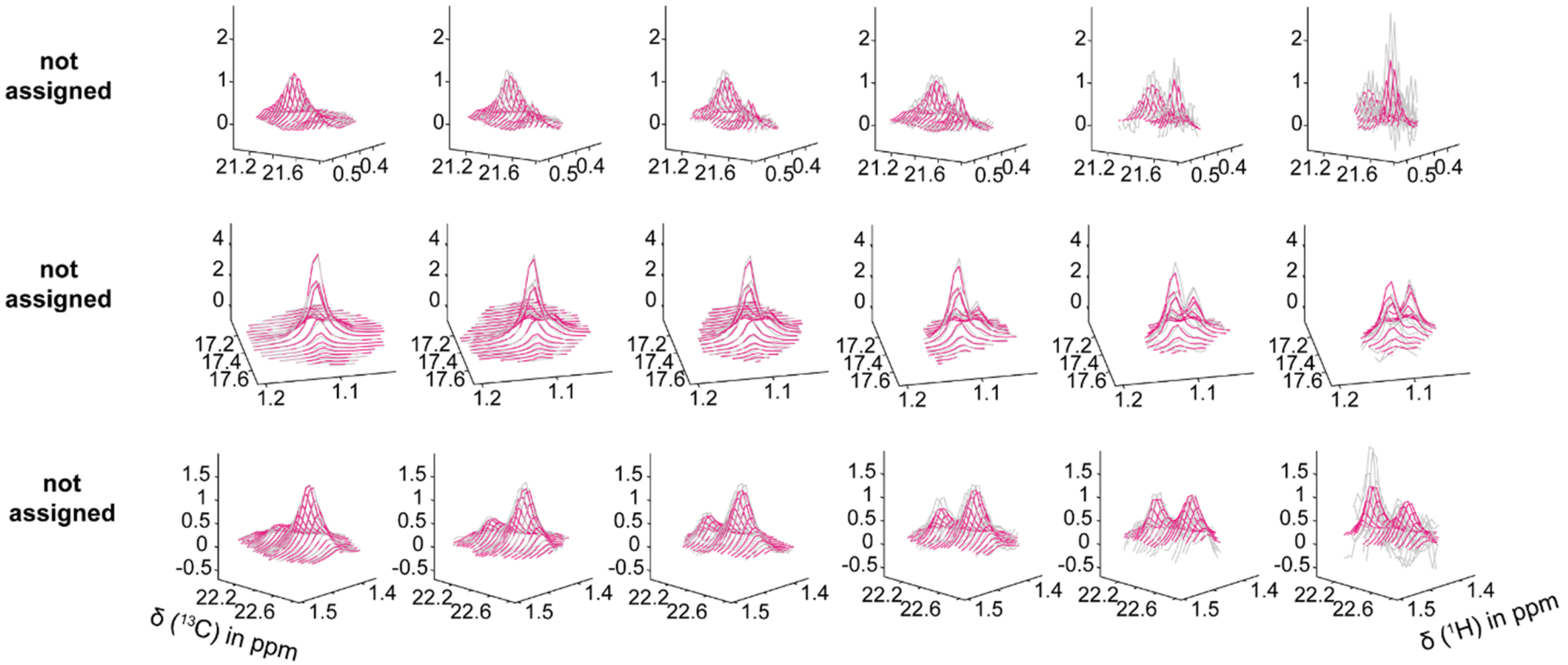


Fig. S 3.1 continued


not assigned
not assigned
not assigned
δ (13C) in ppm
δ (1H) in ppm

3.2 CSP analysis

Tab. S 3.1: Chemical shifts of apo and GCDCA-bound P-domain.

Chemical shifts of resonances where a transfer of assignment was possible are tabulated along with the resulting CSPs. CSPs were calculated as Euclidian distances $\Delta\nu_{Eucl}$ and as distances $\Delta\nu_H$ and $\Delta\nu_C$ in the ^1H and ^{13}C dimensions, respectively. Light grey rows signify resonance signals where a transfer of assignment was not possible due to signal crowding or CSPs could not be determined due to line broadening. Red rows highlight peaks that are well isolated (no spectral crowding) in the GCDCA-bound form. Nonetheless, no corresponding signal of the apo-form can be identified, suggesting that the apo-signal is either outside a Euclidian distance range of ca. 20 Hz, or it is broadened due to unfavorable exchange dynamics between monomers and dimers in the apo-form. The classification of resonance signals is illustrated in Fig. 5.3.3. See Fig. 5.3.4 for more details. The table and the legend are adapted from Creutznacher et al., 2022, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

	Chemical shifts δ /ppm						
	Apo		**GCDCA bound**				
Amino acid	^1H	^{13}C	^1H	^{13}C	Total CSP in Hz	^1H CSP in Hz	^{13}C CSP in Hz
229Met	1.61	16.9	1.59	16.9	8.9	8.8	1.3
230Val	0.69	22.2	0.69	22.2	1.8	0.2	1.8
232Leu	0.3	22.5	0.28	22.5	n.d.	n.d.	n.d.
234Val	0.84	20.2	0.84	20.3	10.7	1.7	10.5
235Ile	0.58	13.7	0.58	13.7	2.9	2.9	0.1
252Leu	0.49	26.1	0.5	26.1	10.1	6.3	7.9
254Val	0.38	19.2	0.38	19.2	2.3	0.3	2.3
269Val	1.34	20.7	1.35	20.7	4.1	2.8	3
271Val	0.63	23.1	0.65	23.2	13.2	11.2	6.9
275Leu	0.7	23.6	0.7	23.6	4.3	0.7	4.3
281Ile	0.69	12	0.69	11.9	18.6	4.4	18
291Ala	1.21	24.6	1.28	24.5	40.7	36.9	17.3
293Ala	n.d.	n.d.	1.08	24.2	>20	n.d.	n.d.
294Ala	n.d.	n.d	1.19	21.2	n.d.	n.d.	n.d.
304Val	0.85	20.6	0.85	20.6	8.9	1.4	8.8
305Ala	0.08	22.4	0.08	22.7	50.5	1.2	50.4
309Leu	n.d.	n.d.	0.52	23.1	n.d.	n.d.	n.d.
310Ile	0.81	14.2	0.81	14.2	9.1	2.4	8.8

323Ala	0.22	18	0.21	18.1	4.9	3.9	3
324Ala	1.16	19.5	1.18	19.5	16.3	15	6.3
335Leu	0.99	26.2	1.03	26.1	23.3	21.8	8.2
337Ile	n.d.	n.d.	0.79	14.4	>20	n.d.	n.d.
339Val	n.d.	n.d.	1.05	19.5	>20	n.d.	n.d.
352Val	n.d.	n.d.	0.73	19	>20	n.d.	n.d.
*357Met	1.94	18	1.95	18	11.5	7.3	8.9
*357Met	1.95	17.9	1.96	17.9	12.1	5	11
*358Ile	0.95	12.5	0.96	12.9	53.1	3.6	53
*358 Ile	0.95	12.5	0.99	12.9	54.5	20.7	50.5
359Leu	n.d.	n.d.	0.77	22.2	n.d.	n.d.	n.d.
*365Ala	n.d.	n.d.	1.4	19.3	n.d.	n.d.	n.d.
*365Ala	1.22	19.1	1.2	19.2	27.7	14.8	23.4
*374Val	0.48	19.9	0.47	19.8	n.d.	n.d.	n.d.
*374Val	0.48	19.9	0.49	20	n.d.	n.d.	n.d.
376Ala	1.54	24.8	1.55	25	27.1	3.3	26.9
378Val	n.d.	n.d.	0.95	21.2	>20	n.d.	n.d.
380Ala	1.61	21.4	1.6	21.5	13.8	8.3	11
381Ala	1.43	19.1	1.43	19.1	7.1	0.3	7.1
382Ala	1.07	21.3	1.07	21.4	10.4	0.4	10.3
384Leu	n.d.	n.d.	0.8	22.3	n.d.	n.d.	n.d.
387Val	n.d.	n.d.	0.8	21.7	n.d.	n.d.	n.d.
391Val	n.d.	n.d.	0.63	18.8	>20	n.d.	n.d.
393Ala	-0.33	20.3	-0.47	19.8	103.3	80.3	65
*398Ile	n.d.	n.d.	1.31	14.5	>20	n.d.	n.d.
*398Ile	n.d.	n.d.	1.27	14.7	>20	n.d.	n.d.
*405Ile	1.1	14.8	1.12	14.7	25.5	8.2	24.1
*405Ile	1.11	14.8	1.12	14.6	27.7	8.2	26.4
*414Val	0.53	19	0.54	18.8	33.6	6.1	33.1
*414Val	0.54	19.2	0.54	18.9	36.7	0.4	36.7

416Leu	n.d.	n.d.	0.8	23.5	n.d.	n.d.	n.d.
420Ile	0.83	14	0.83	14.1	14.4	0.8	14.4
428Val	0.98	20.8	0.98	20.8	5.7	1.6	5.5
429Leu	0.83	26.6	0.83	26.7	3.7	0.8	3.6
430Leu	0.82	27	0.82	27	5.3	4.5	2.8
436Met	0.01	16.9	-0.14	16.3	136.4	91.2	101.4
439Ile	n.d.	n.d.	0.82	12.7	>20	n.d.	n.d.
442Ala	n.d.	n.d.	1.42	22.5	>20	n.d.	n.d.
444Ala	n.d.	n.d.	1.25	19.4	>20	n.d.	n.d.
446Ala	n.d.	n.d.	0.93	18.6	>20	n.d.	n.d.
448Ala	1.15	19.9	1.11	20	26.8	19.9	17.9
*449Ile	0.46	13.7	0.54	13.8	52.3	49.5	16.8
*449Ile	0.43	13.7	0.52	13.8	58.3	54.2	21.4
452Ala	2.04	20	2.04	20.1	9.3	1.6	9.1
453Leu	0.96	23.4	0.95	23.4	3.6	1.3	3.4
462Ala	1.46	18.2	1.46	18.1	13.8	4.4	13
472Ala	0.76	21.6	0.77	21.6	7.3	4.7	5.6
473Leu	-0.6	23.6	-0.61	23.6	3.7	3.2	1.9
474Leu	0.65	25.4	0.65	25.4	1.8	1.8	0
475Leu	0.63	25.6	0.63	25.6	3	2.9	0.9
491Leu	0.78	26.6	0.79	26.5	7.6	7.1	2.5
497Ile	0.66	12.7	0.66	12.7	2.5	0.9	2.3
499Leu	-0.17	26.8	-0.18	26.9	9.6	7.5	6.1
507Leu	0.59	26.8	0.58	26.8	4.8	3.9	2.8
514Ile	0.63	12.6	0.63	12.6	6.2	1.7	5.9
517Val	0.63	21.9	0.62	21.9	8.5	3.5	7.8
518Val	0.78	20.7	0.78	20.7	3	1.3	2.8
521Val	0.29	17.1	0.29	17.1	1.3	1.3	0.1
530Val	1.2	21.4	1.2	21.4	4.4	2	3.9

For amino acids labeled with * we observed two resonances due to a mixture of E- and Z-isomers of Pro 361 (c.f. chapter 5.2.2)

3.3 Reversibility of GCDCA effects

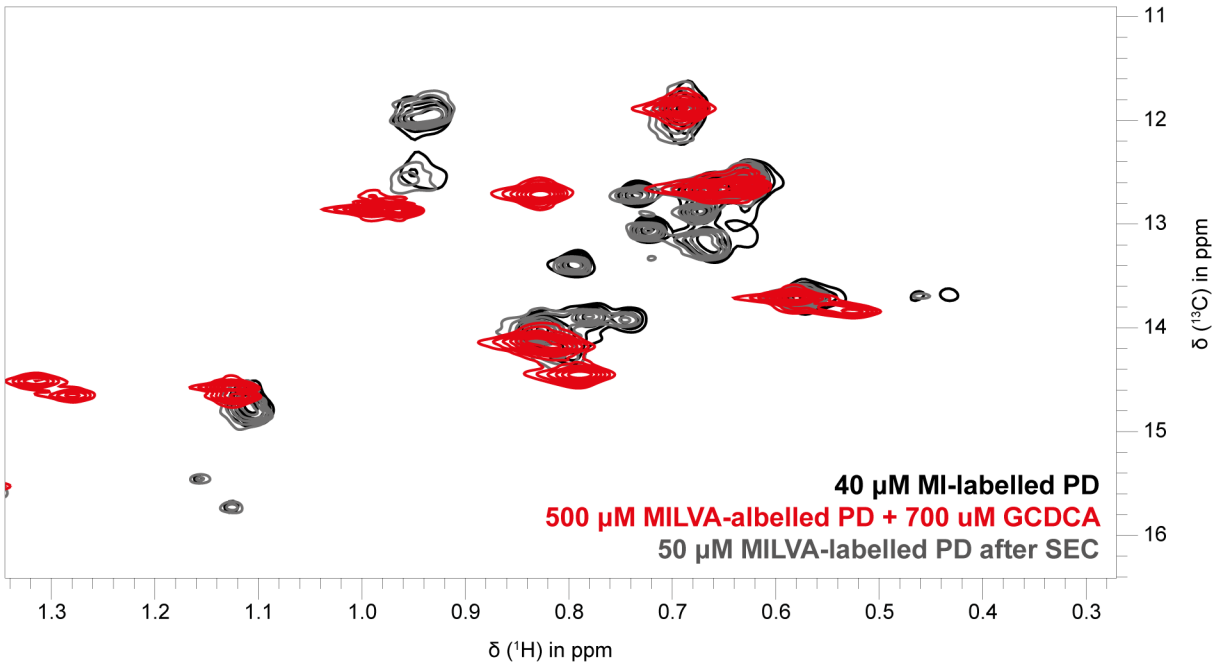


Fig. S 3.2: **MNV P-domain proteins titrated with GCDCA revert to their apo state during size exclusion chromatography (SEC) as can be seen in $^{1}H,^{13}C$ HMQC spectra.**

The black spectrum shows the isoleucine region of MI-labelled apo P-domain (see Fig. S 2.1 for details on acquisition). The red spectrum shows the isoleucine region of MILVA-labelled P-domain in the presence of saturating amounts of GCDCA (see Fig. 5.2.1 for details on acquisition). The protein was subjected to size exclusion chromatography (SEC) and another spectrum was acquired (grey, the spectrum was recorded on a Bruker 600 MHz Avance III HD spectrometer equipped with a TCI cryogenic probe at 298 K). The resonances of the apo P-domain show the same fingerprint as the re-purified protein. SEC was carried out as explained in chapter 4.2.4. Figure and legend are adapted from Creutznacher et al., 2022, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

4 P-domain metal ion interactions

4.1 Assignment transfer to metal ion bound P-domains

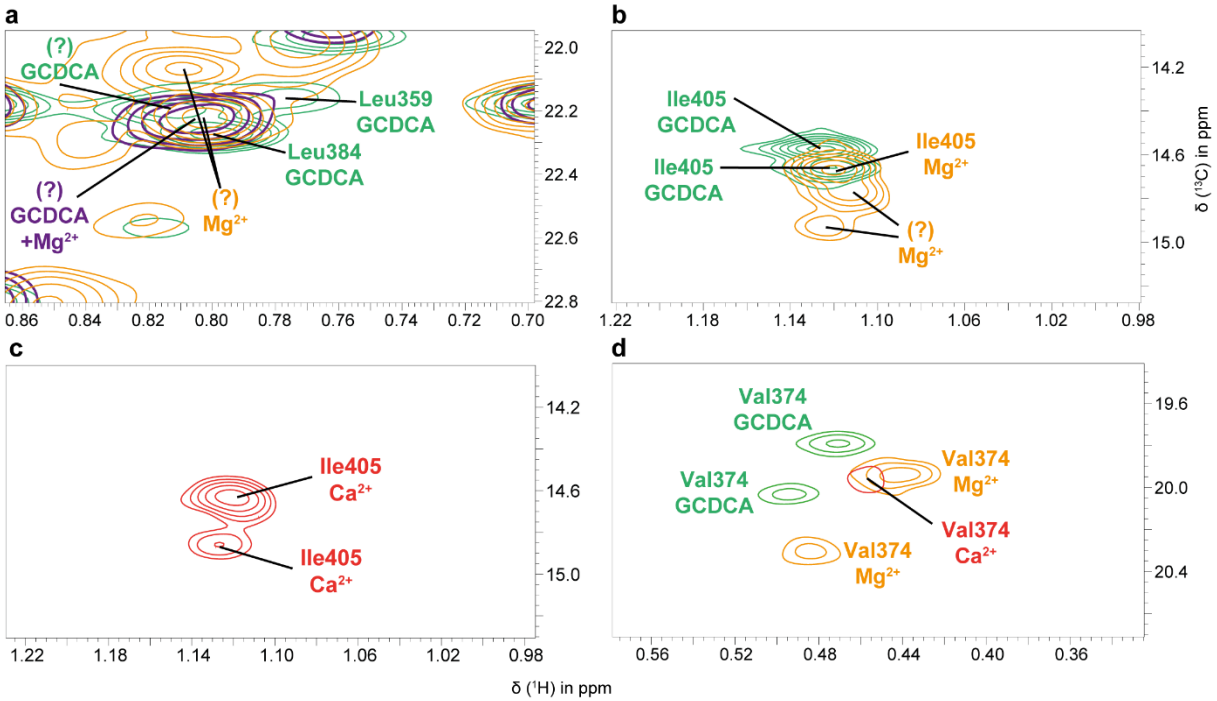


Fig. S 4.1: Resonances in [^{1}H,^{13}C] HMQC spectra of the P-domain for that no assignment transfer is possible.

Transfer of assignment from the GCDCA bound state (green) to the GCDCA and Mg^{2+} or Ca^{2+} bound states (purple) or to the Mg^{2+} or Ca^{2+} bound states (orange) is not possible for Leu384 and Leu359 (a, exemplarily shown for Mg^{2+}) and one of the two Ile405 resonances for the Mg^{2+} state (b). However, for Ca^{2+}, both Ile405 peaks are observable and not split into a third peak (c). (d) For the Ca^{2+} bound state, one of the two resonances corresponding to Val374 is broadened beyond detection. For acquisition of spectra, see Figs. 5.3.4, 5.4.3, 5.4.10, and S 4.4. Note that the presence of two resonances corresponding to a single methyl group is caused by Pro361 cis/trans isomers as explained in chapter 5.2. Unassigned peaks are labelled with a "?".

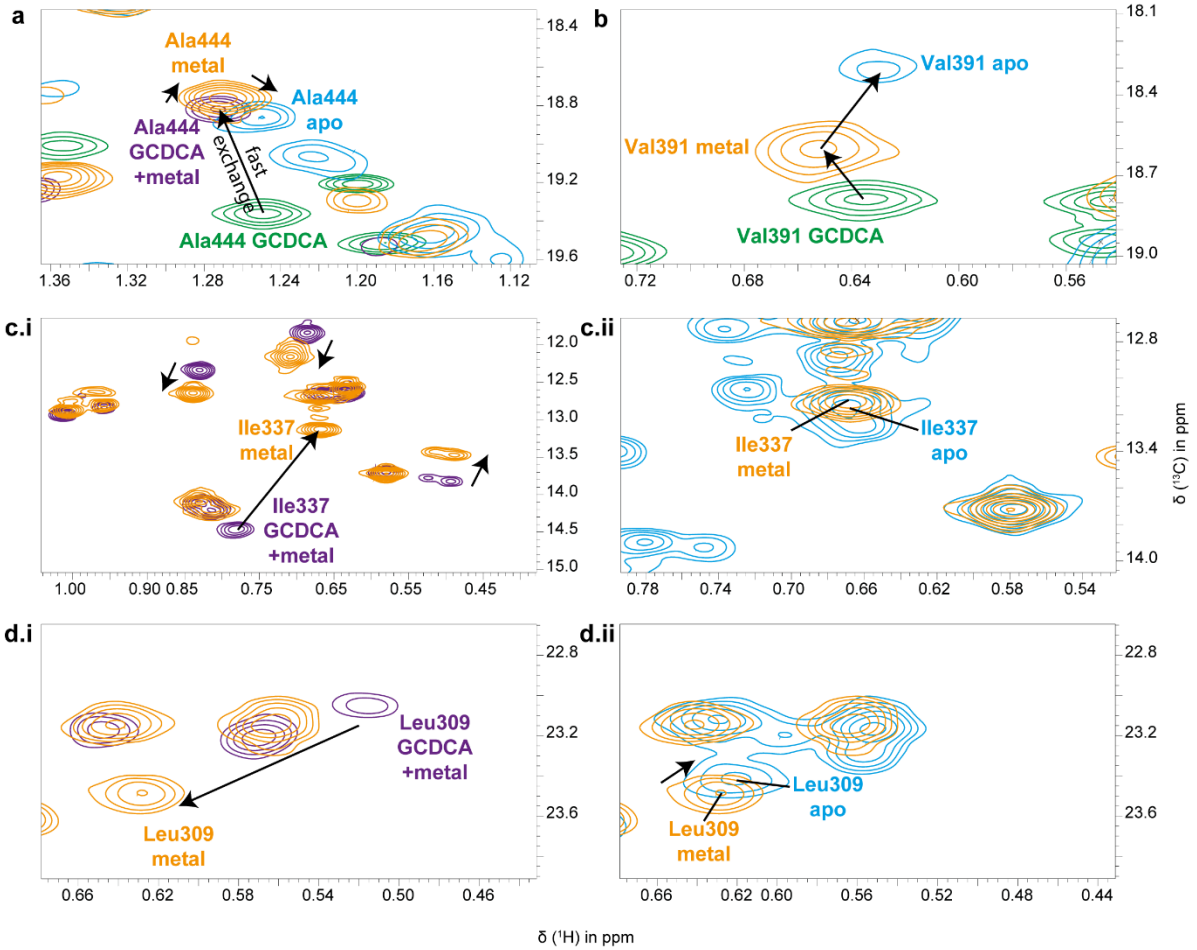


Fig. S 4.2: Comparing [^{1}H,^{13}C] HMQC spectra of metal ion bound, GCDCA bound, and metal ion and GCDCA bound P-domain with apo P-domain spectra enables assignment of Ala444 (a), Val391 (b), Ile337 (c), and Leu309 (d) apo P-domain resonances.

The transfer from the metal ion bound P-domain state is shown with spectra in presence of Mg^{2+}. For acquisition of spectra and further information, see Fig. S 4.1.

4.2 Metal ion binding to MNV P-domain

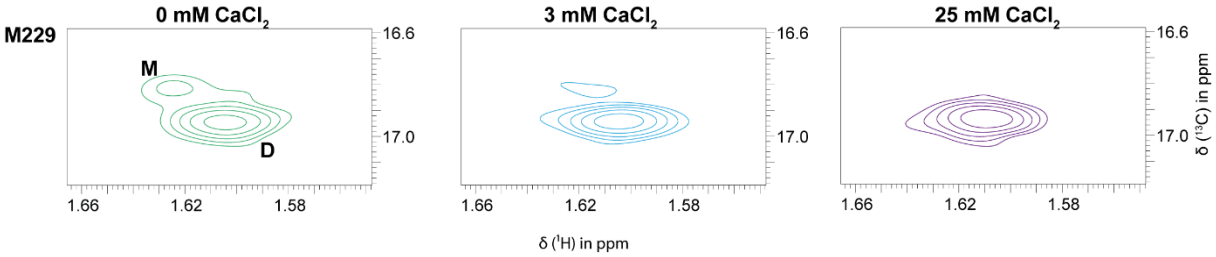


Fig. S 4.3: Ca^{2+} dependent depletion of monomer peaks in [^{1}H^{13}C] HMQC spectra.

Upon addition of $CaCl_2$ to the P-domain, monomer resonances (M) disappear, whereas dimer resonances (D) remain. For acquisition of spectra, see Fig. S 4.4.

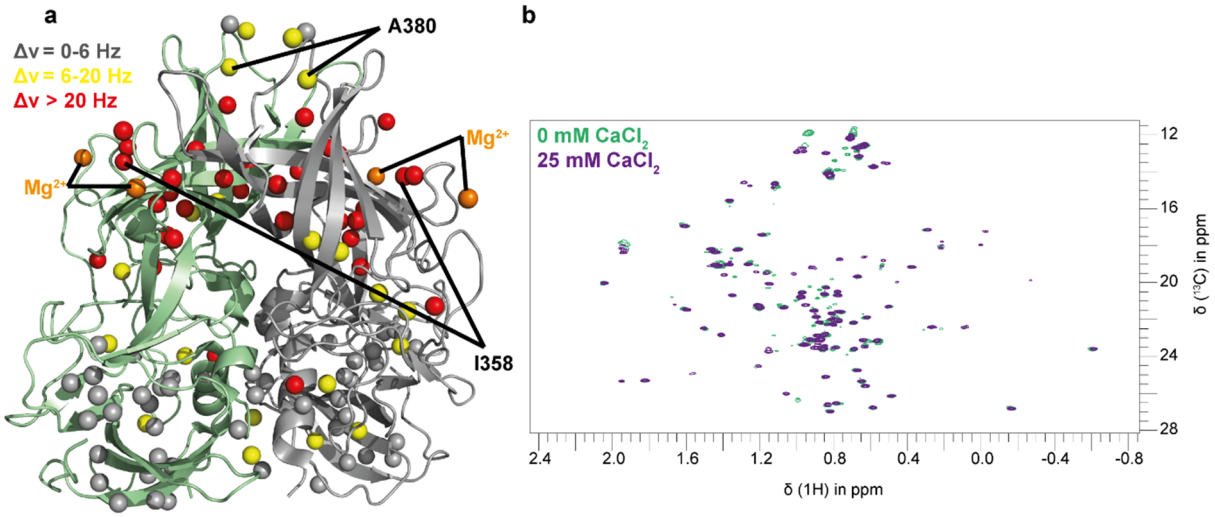


Fig. S 4.4: CSP pattern due to $CaCl_2$ titration is highly similar to CSP pattern due to $MgCl_2$ titration (a, b; see Fig. 5.4.3 for $MgCl_2$ induced CSPs).

[^{1}H,^{13}C] HMQC Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 38 uM MILVA labelled MNV CW1 P-domain. For the illustration, a structural model of Nelson et al., 2018 (pdb 6e47) was used. In this model, Mg^{2+} instead of Ca^{2+} ions are present. See also Tab. S 4.2.

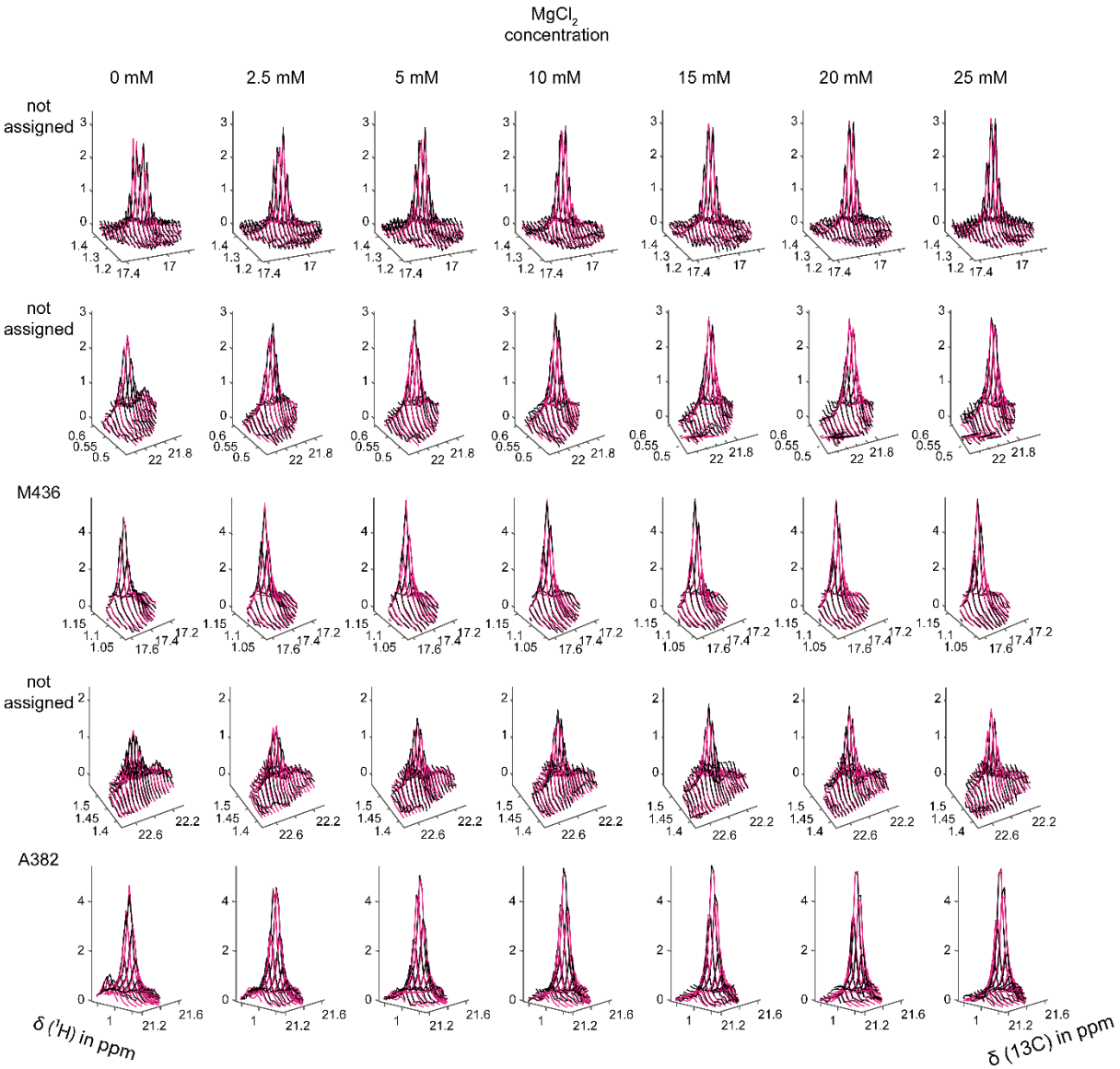


Fig. S 4.5: *TITAN* 2D line shape analyses of Mg^{2+} binding to the binding site at G'H' loop in D410A P-domain in absence of GCDCA.

For details on acquisition of spectra, see Fig. 5.4.6. The analysis yielded a $K_{D,Mg'}$ of 3.8 +- 0.2 mM and a $k_{off\ Mg}$ of 336 +- 692 1/s. Black and magenta spectra correspond to the measured and simulated spectra. The overlay reflects the good quality of the fit.

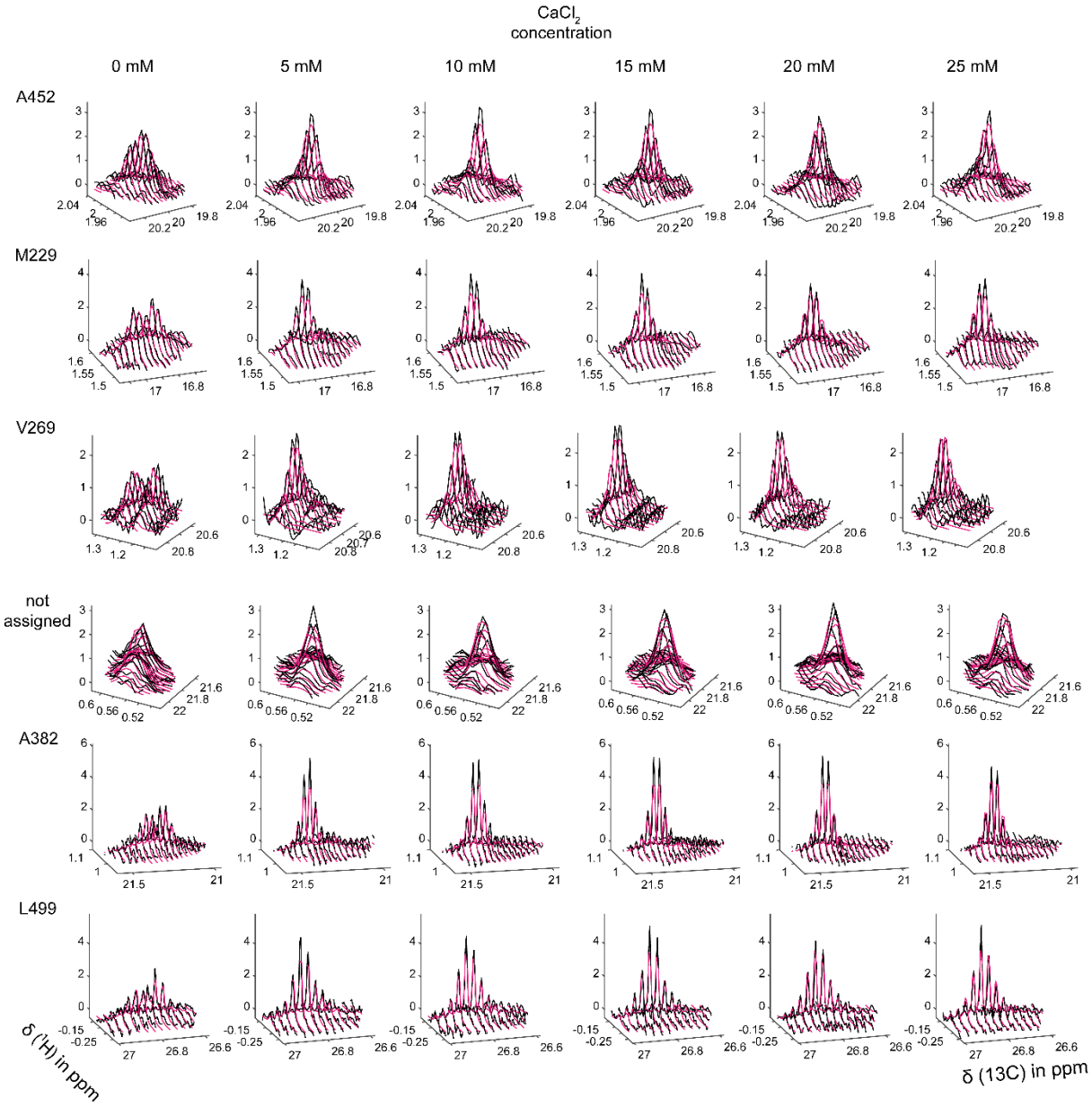


Fig. S 4.6: *TITAN* 2D line shape analyses of Ca^{2+} binding to the binding site at G'H' loop in D410A P-domain in absence of GCDCA.

For details on acquisition of spectra, see Fig. 5.4.9. The analysis yielded a $K_{D,Ca}$ of 1.7 +- 0.2 mM and a $k_{off,Ca}$ of 3052 +- 2998 1/s. Black and magenta spectra correspond to the measured and simulated spectra. The overlay reflects the good quality of the fit.

Note S 4.1: Estimating Mg^{2+} dissociation constants in presence of Ca^{2+} contaminated GCDCA.

At a GCDCA concentration of 0.300 mM, it can be estimated that 0.14 mM Ca^{2+} ions are present (Creutznacher et al., 2021).

Using the Chen-Prusoff equation (Cheng and Prusoff, 1973),

$$K_D = \frac{[L] * K_I}{IC_{50} - K_I}$$

Eq. S 4.1

With IC_{50} (0.62 mM) being the apparent dissociation constant K_I of an inhibitor (Mg^{2+}) in presence of a ligand (Ca^{2+}) with the concentration [L] (0.14 mM) and a dissociation constant K_D (0.138 mM, Creutznacher et al., 2021), the K_I can be readily calculated. The resulting dissociation constant of Mg^{2+} is $K_I = K_{D;Mg}$ = 0.31 mM .The values differs from the value obtained neglecting the Ca^{2+} contamination by a factor of ca. 2.

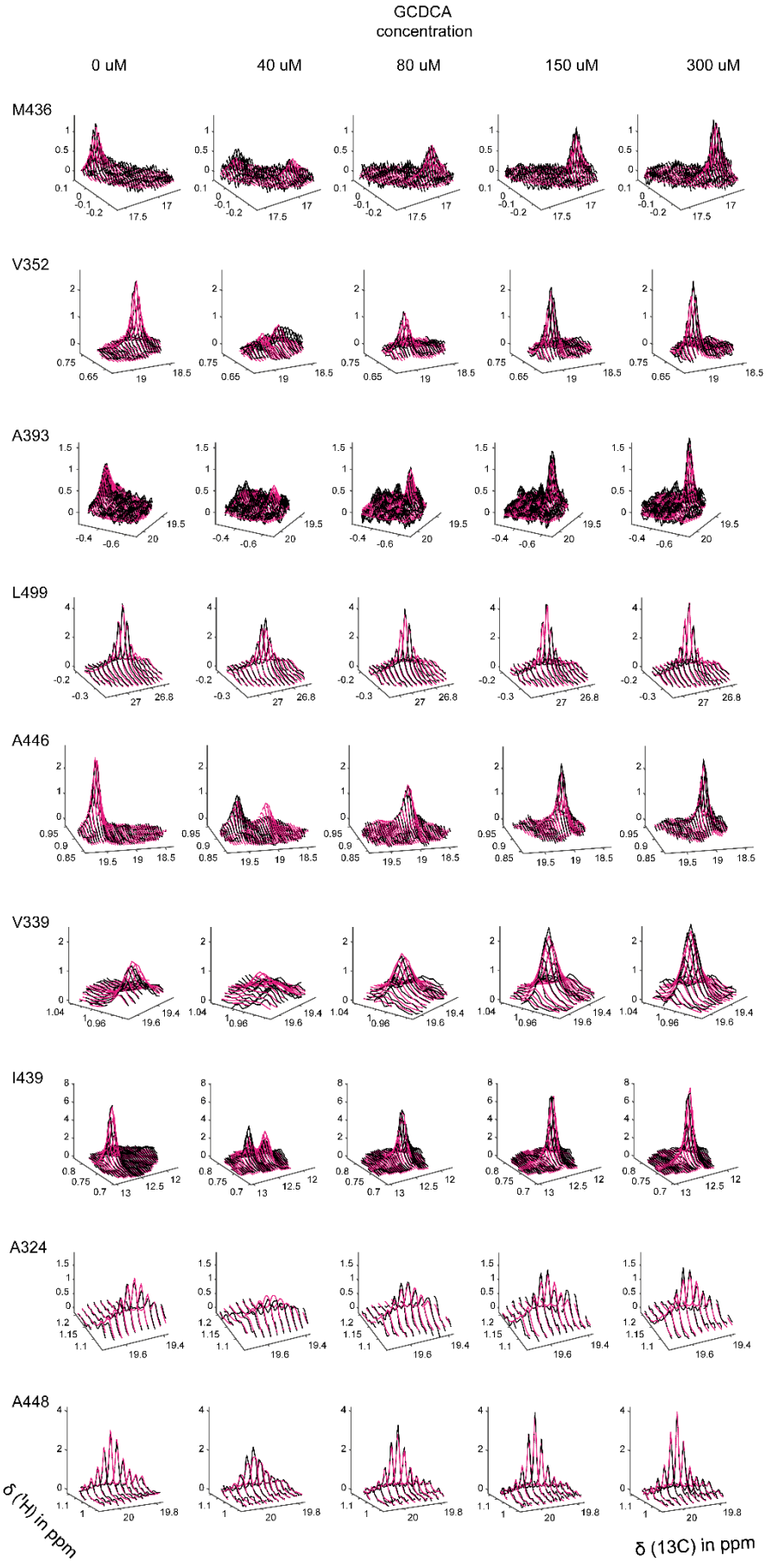


Fig. S 4.7: *TITAN* analysis of GCDCA binding to Mg^{2+} bound P-domain.

Resulting K_D and k_{off} values are given in Fig. 5.4.10 and in the main text. For details on acquisition of spectra, see Fig. 5.4.10. Black and magenta spectra correspond to the measured and simulated spectra. The overlay reflects the good quality of the fit.

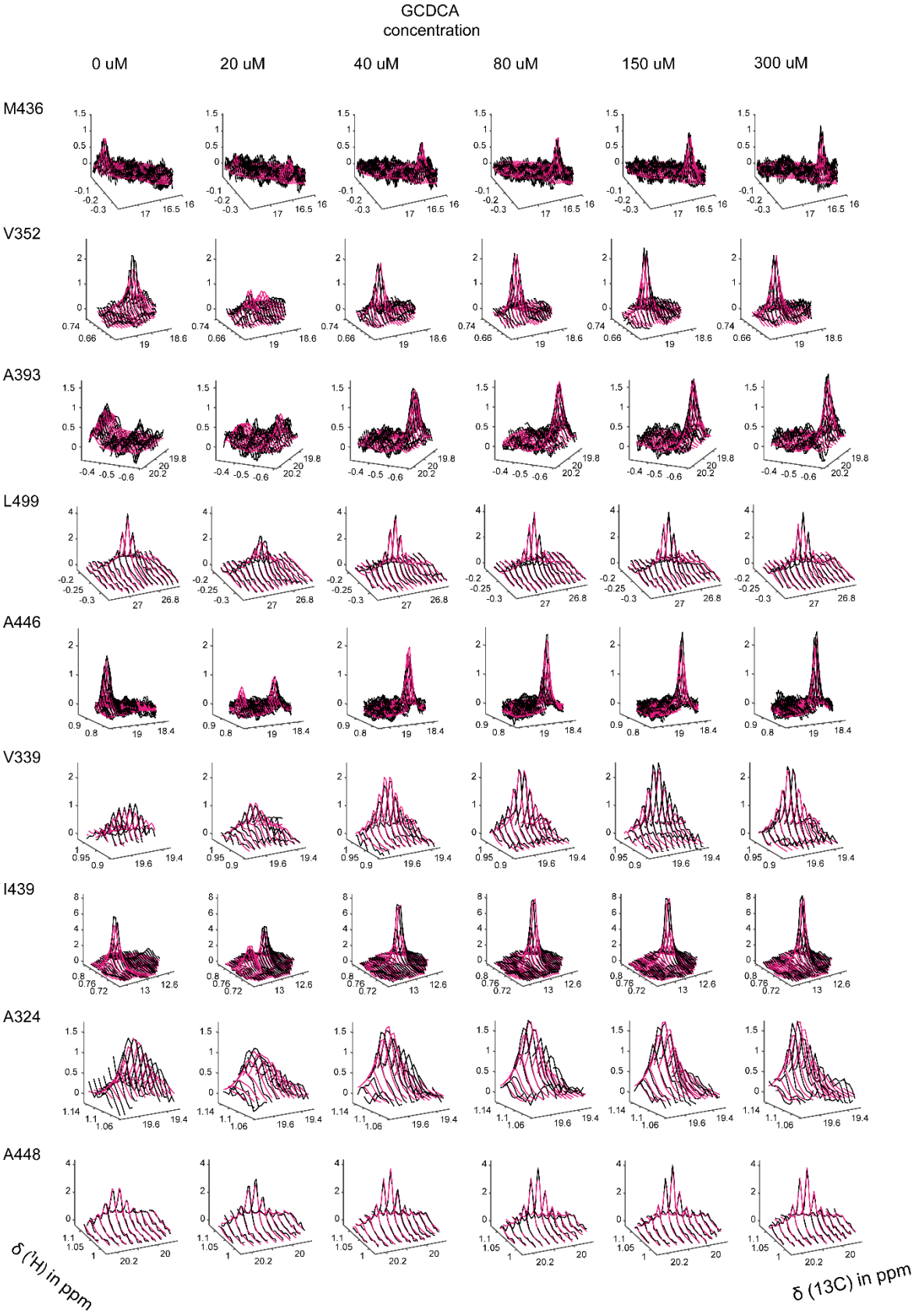


Fig. S 4.8: *TITAN* analysis of GCDCA binding to Ca^{2+} bound P-domain.

Resulting K_D and k_{off} values are given in Fig. 5.4.10 and in the main text. For details on acquisition of spectra, see Fig. S 4.9. Black and magenta spectra correspond to the measured and simulated spectra. The overlay reflects the good quality of the fit.

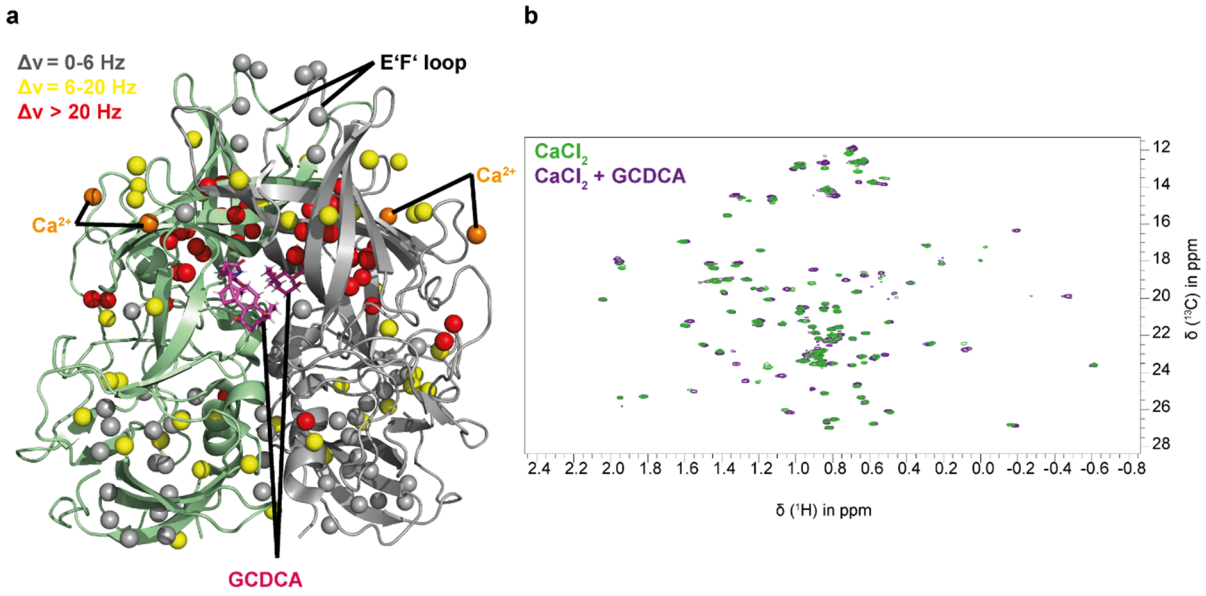


Fig. S 4.9: GCDCA CSPs on Ca bound WT P-domain.

(b) $^{1}H,^{13}C$ HMQC spectra of the P-domain in the presence of $CaCl_2$ with and without 300 µM GCDCA were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 39 uM MILVA labelled MNV CW1 P-domain at 25 mM $CaCl_2$. For the illustration in (a), a structural model of Nelson et al., 2018 (pdb 6e47) was used. In this model, Mg^{2+} instead of Ca^{2+} ions are present, but the ions are labelled with "Ca^{2+}" here.

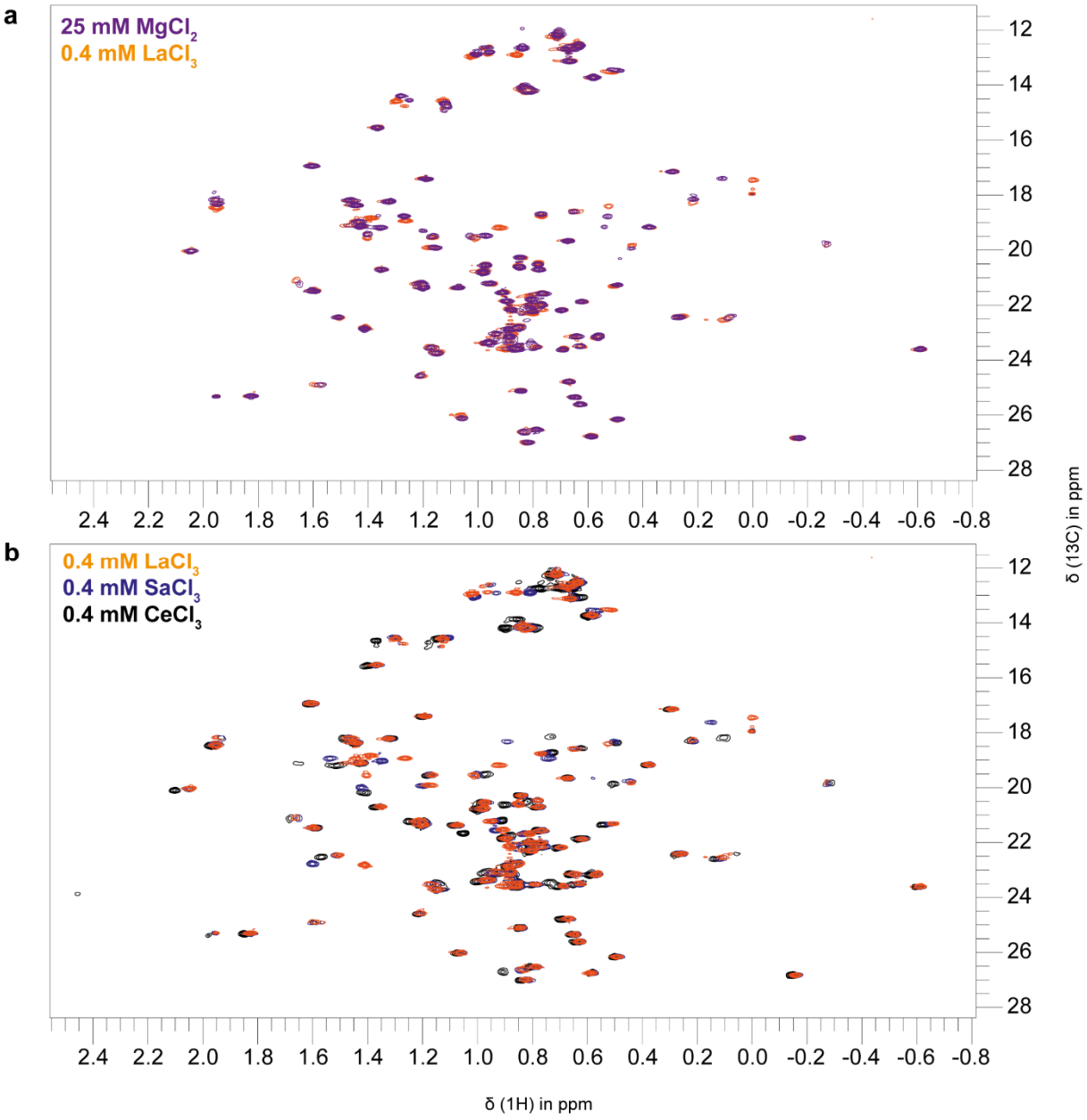


Fig. S 4.10: P-domain Spectra with paramagnetic and diamagnetic metals.

[^{1}H,^{13}C] HMQC spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 38 uM MILVA labelled MNV CW1 P-domain. For sample conditions corresponding to the spectrum in presence of $MgCl_2$, see Fig. 5.4.3.

4.3 Order Parameters of metal ion-bound P-domain

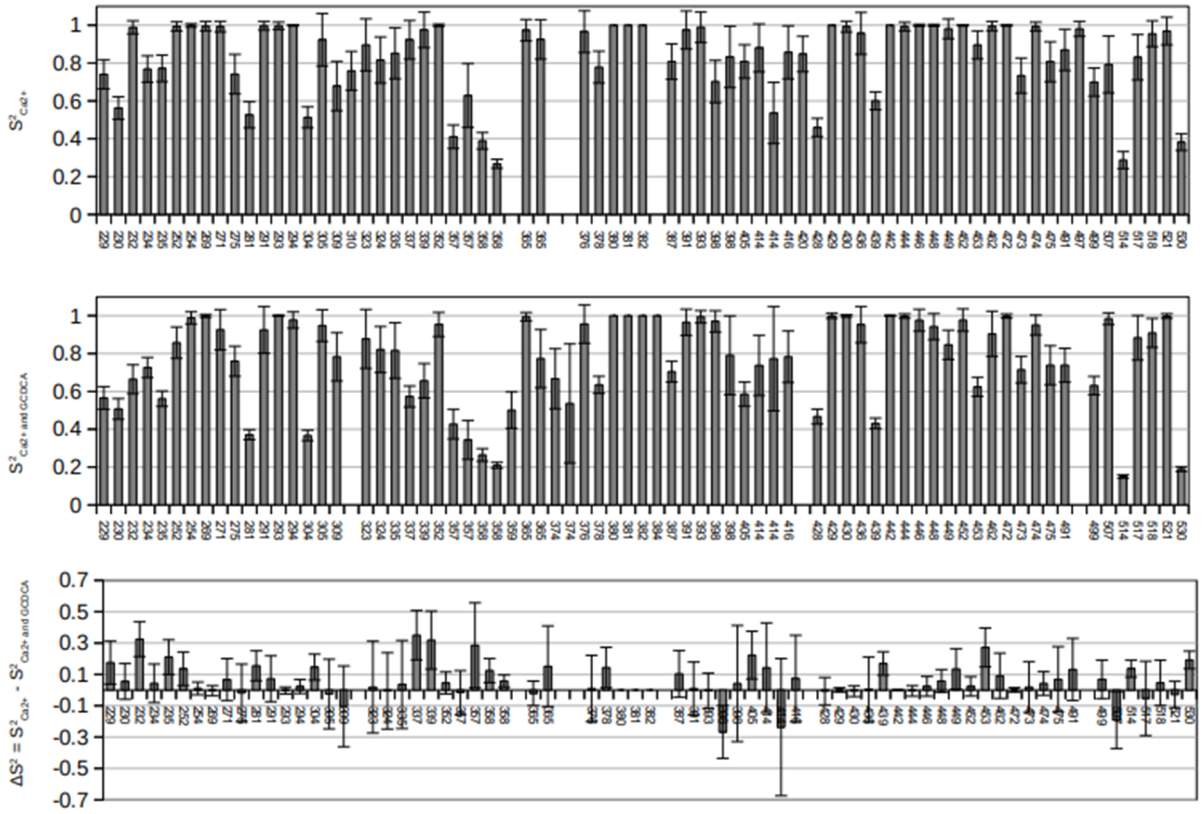


Fig. S 4.11: Methyl group order parameters S^2 of Ca^{2+} bound P-domain and Ca^{2+} and GCDCA bound P-domain.

See Fig. 5.4.14 for more information.

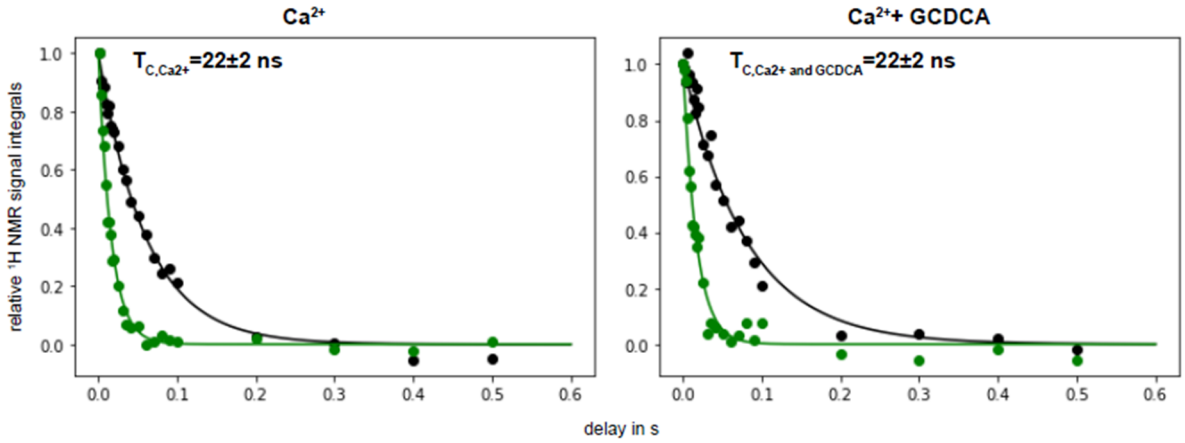


Fig. S 4.12: TRACT experiments of P-domain at pH 5.3 in the presence of Ca^{2+} or in the presence of Ca^{2+} and GCDCA.

Sample conditions, experimental details, and details on the analysis are given in 4.3.3 and 4.4.6.

4.4 CSP analysis of metal ion binding to apo or GCDCA-bound P-domains

Tab. S 4.1: Chemical shifts and CSPs between the apo WT P-domain and Mg^{2+} bound WT P-domain.

See Fig. 5.4.3 for more details. Also see Tab. S 3.1 for information about color coding.

Amino acid	Apo P-domain		Mg^{2+} bound P-domain		^{1}H CSP in Hz	^{13}C CSP in Hz	total CSP in Hz
	^{1}H in ppm	^{13}C in ppm	^{1}H in ppm	^{13}C in ppm			
229Met	1.61	16.9	1.6	16.9	1.5	2.1	2.6
230Val	0.7	22.2	0.7	22.2	0.7	1.6	1.7
232Leu	0.31	22.5	0.27	22.4	23.8	13	27.1
234Val	0.84	20.2	0.85	20.3	1.7	7.3	7.5
235Ile	0.58	13.7	0.58	13.7	1	0.3	1
252Leu	0.49	26.1	0.49	26.2	3.2	2.3	4
254Val	0.38	19.2	0.38	19.2	0.6	1.9	2
269Val	1.34	20.7	1.35	20.7	5	0.5	5
271Val	0.63	23.1	0.64	23.1	7.1	3.9	8.1
275Leu	0.7	23.7	0.69	23.6	2.6	5.3	6
281Ile	0.7	12	0.71	12.2	7.9	25.1	26.3
291Ala	1.22	24.6	1.21	24.6	4	1.9	4.5
293Ala	1.15	23.8	1.15	23.7	n.d.	n.d.	n.d.
294Ala	n.d.	n.d.	1.21	21.2	n.d.	n.d.	n.d.
304Val	0.85	20.6	0.85	20.6	0.6	3.7	3.7
305Ala	0.08	22.4	0.09	22.5	n.d.	n.d.	n.d.
309Leu	0.62	23.4	0.63	23.5	4.6	10.7	11.6
310Ile	n.d.	n.d.	0.81	14.2	n.d.	n.d.	n.d.
323Ala	0.22	18	0.21	18.1	3.1	18.9	19.1
324Ala	1.16	19.5	1.17	19.5	3.4	6	6.9
335Leu	1	26.4	1.06	26.1	35.9	37.1	51.6
337Ile	0.67	13.2	0.67	13.1	1.3	5	5.2
339Val	n.d.	n.d.	1.03	19.5	n.d.	n.d.	>20
352Val	n.d.	n.d.	0.77	18.7	n.d.	n.d.	>20
357Met	1.94	18	1.95	18.3	6.2	48.7	49.1

357Met	1.95	17.8	1.96	18.2	2.5	47.3	47.4
358Ile	0.96	12.6	1.01	12.9	29	40.3	49.7
358Ile	0.96	12.6	0.96	12.8	0.1	29	29
365Ala	n.d.	n.d.	1.4	19.4	n.d.	n.d.	>20
365Ala	1.22	19.1	1.2	19.3	14.4	34.8	37.6
374Val	0.47	19.9	0.44	19.9	15.9	11.9	19.8
374Val	0.49	20.1	0.48	20.3	2.9	37.8	37.9
376Ala	1.54	24.9	1.57	24.9	n.d.	n.d.	n.d.
378Val	n.d.	n.d.	0.96	21.2	n.d.	n.d.	>20
380Ala	1.61	21.4	1.6	21.5	7.5	10.5	12.9
381Ala	1.43	19.1	1.43	19.1	0	2.5	2.5
382Ala	1.07	21.3	1.07	21.4	0.4	7.8	7.8
387Val	n.d.	n.d.	0.81	21.8	n.d.	n.d.	n.d.
391Val	0.63	18.3	0.65	18.6	13.5	43.6	45.6
393Ala	-0.28	19.8	-0.27	19.7	n.d.	n.d.	n.d.
398Ile	n.d.	n.d.	1.28	14.4	n.d.	n.d.	>20
398Ile	n.d.	n.d.	1.25	14.5	n.d.	n.d.	>20
405Ile	1.11	14.7	1.12	14.7	4.3	11.8	12.5
414Val	0.54	19	0.53	18.8	4.1	36.6	36.8
414Val	0.55	19.2	0.54	19.2	3.1	6.1	6.9
416Leu	n.d.	n.d.	0.81	23.53	n.d.	n.d.	n.d.
420Ile	0.83	14	0.83	14.1	2.2	8.6	8.8
428Val	0.98	20.8	0.98	20.8	0.7	3.2	3.3
429Leu	0.83	26.6	0.83	26.6	1.1	2	2.3
430Leu	0.82	27	0.82	27	0.8	3.2	3.3
436Met	0.01	16.9	0.11	17.4	58.4	71.4	92.2
439Ile	n.d.	n.d.	0.84	12.6	n.d.	n.d.	>20
442Ala	n.d.	n.d.	1.41	22.9	n.d.	n.d.	>20
444Ala	1.25	18.9	1.27	18.8	12.1	14.9	19.2
446Ala	n.d.	n.d.	0.97	19.5	n.d.	n.d.	>20

448Ala	1.14	19.9	1.16	19.9	7.3	3.7	8.1
449Ile	0.44	13.7	0.49	13.5	31.4	32.6	45.3
449Ile	0.46	13.7	0.51	13.4	30.7	38.2	49
452Ala	2.04	20	2.04	20	0	3.5	3.5
453Leu	0.96	23.4	0.96	23.4	0.9	1.1	1.4
462Ala	1.46	18.2	1.47	18.2	4.4	2.4	5
472Ala	0.76	21.6	0.76	21.6	2.4	2.1	3.2
473Leu	-0.6	23.6	-0.61	23.6	3.2	4.3	5.4
474Leu	0.65	25.4	0.65	25.4	0.4	0.8	0.9
475Leu	0.63	25.6	0.63	25.6	2.8	3.1	4.1
491Leu	0.78	26.6	0.79	26.5	2.2	4.5	5
497Ile	0.66	12.7	0.67	12.7	3.3	1.7	3.7
499Leu	-0.17	26.8	-0.17	26.8	1.5	0.1	1.5
507Leu	0.59	26.8	0.59	26.8	0.3	1	1
514Ile	0.63	12.6	0.64	12.6	5	1.6	5.2
517Val	0.63	21.9	0.62	21.9	3.2	7.9	8.5
518Val	0.78	20.7	0.78	20.7	0.4	2.5	2.5
521Val	0.29	17.1	0.29	17.1	1.2	0.3	1.2
530Val	1.2	21.4	1.2	21.4	0.9	3.3	3.4

*missing resonances of the Mg^{2+} bound P-domain in comparison to GCDCA bound P-domain correspond to one the two Ile405 resonances, and the Leu359 and Leu384 resonances (c.f. Fig. S 4.1).

Tab. S 4.2: Chemical shifts and CSPs between the apo WT P-domain and Ca^{2+} bound WT P-domain.

See Fig. S 4.4 for more details. Also see Tab. S 3.1 for information about color coding.

Amino acid	Apo P-domain		Ca^{2+} bound P-domain		^{1}H CSP in Hz	^{13}C CSP in Hz	total CSP in Hz
	^{1}H in ppm	^{13}C in ppm	^{1}H in ppm	^{13}C in ppm			
229Met	1.61	16.9	1.60	16.9	2.8	1.5	3.2
230Val	0.69	22.2	0.69	22.2	0.0	2.1	2.1
232Leu	0.26	22.4	0.31	22.6	28.1	21.9	35.6
234Val	0.84	20.3	0.84	20.2	2.5	7.2	7.7
235Ile	0.58	13.7	0.58	13.7	1.1	2.9	3.1
252Leu	0.49	26.1	0.48	26.1	4.3	3.1	5.3
254Val	0.38	19.2	0.38	19.2	0.5	2.1	2.2
269Val	1.35	20.7	1.35	20.7	2.5	0.4	2.5
271Val	0.64	23.1	0.63	23.1	6.7	1.7	6.9
275Leu	0.69	23.6	0.70	23.7	3.8	10.3	10.9
281Ile	0.71	12.2	0.69	12.0	9.8	32.9	34.3
291Ala	1.21	24.5	1.22	24.6	6.6	9.9	11.9
293Ala	1.15	23.7	1.15	23.7	n.d.	n.d.	n.d.
294Ala	n.d.	n.d.	1.21	21.2	n.d.	n.d.	n.d.
304Val	0.85	20.6	0.85	20.7	1.0	6.4	6.5
305Ala	0.09	22.5	0.09	22.4	n.d.	n.d.	n.d.
309Leu	0.63	23.5	0.63	23.4	1.1	13.3	13.4
310Ile	n.d.	n.d.	0.82	14.2	n.d.	n.d.	n.d.
323Ala	0.21	18.1	0.22	18.0	2.0	16.2	16.4
324Ala	1.17	19.5	1.16	19.5	2.7	7.8	8.3
335Leu	1.06	26.0	0.99	26.3	38.6	45.2	59.4
337Ile	0.66	13.2	0.66	13.3	n.d.	n.d.	n.d.
339Val	n.d.	n.d.	1.02	19.5	n.d.	n.d.	>20
352Val	n.d.	n.d.	0.77	18.7	n.d.	n.d.	>20
357Met	1.94	18.4	1.93	18.0	4.3	60.4	60.5
357Met	1.95	18.2	1.95	17.8	3.6	51.0	51.1
358Ile	0.96	12.9	0.95	12.6	3.9	44.5	44.7
358Ile	1.00	12.9	0.95	12.6	29.2	53.8	61.2
365Ala	n.d.	n.d.	1.39	19.6	n.d.	n.d.	>20
365Ala	1.22	19.5	1.22	19.1	3.5	60.6	60.7
374Val	n.d.	n.d.	0.46	20.0	n.d.	n.d.	n.d.
376Ala	1.57	24.9	1.55	24.9	n.d.	n.d.	n.d.
378Val	n.d.	n.d.	0.95	21.2	n.d.	n.d.	>20
380Ala	1.60	21.5	1.61	21.4	8.9	8.2	12.1
381Ala	1.43	19.1	1.43	19.1	0.0	3.0	3.0
382Ala	1.07	21.4	1.07	21.3	1.4	9.8	9.9
387Val	n.d.	n.d.	0.81	21.7	n.d.	n.d.	n.d.
391Val	0.63	18.6	0.63	18.3	1.8	48.1	48.1
393Ala	n.d.	n.d.	-0.27	19.9	n.d.	n.d.	n.d.
398Ile	n.d.	n.d.	1.25	14.7	n.d.	n.d.	>20
398Ile	n.d.	n.d.	1.29	14.6	n.d.	n.d.	>20
405Ile	1.12	14.6	1.11	14.7	n.d.	n.d.	n.d.
405Ile	n.d.	n.d.	1.13	14.9	n.d.	n.d.	n.d.
414Val	0.53	18.8	0.53	19.0	2.4	38.3	38.4
414Val	0.54	19.1	0.54	19.3	n.d.	n.d.	n.d.

416Leu	n.d.	n.d.	0.79	23.5	n.d.	n.d.	n.d.
420Ile	0.83	14.1	0.83	14.0	2.6	11.3	11.6
428Val	0.98	20.8	0.98	20.8	0.2	1.8	1.8
429Leu	0.83	26.6	0.83	26.7	1.2	5.3	5.4
430Leu	0.82	27.0	0.82	27.0	0.4	5.1	5.1
436Met	-0.03	17.2	0.00	16.9	17.3	46.8	49.9
439Ile	n.d.	n.d.	0.84	13.0	n.d.	n.d.	>20
442Ala	n.d.	n.d.	1.41	22.9	n.d.	n.d.	>20
444Ala	1.26	19.0	1.25	18.9	8.8	22.4	24.0
446Ala	n.d.	n.d.	0.92	19.2	n.d.	n.d.	>20
448Ala	n.d.	n.d.	1.15	20.1	n.d.	n.d.	n.d.
449Ile	0.53	13.5	0.46	13.7	41.1	26.5	48.9
449Ile	0.51	13.5	0.44	13.7	42.3	21.9	47.6
452Ala	2.04	20.0	2.04	20.0	0.3	4.7	4.7
453Leu	0.96	23.4	0.96	23.4	1.5	1.0	1.9
462Ala	1.46	18.2	1.46	18.2	2.8	0.6	2.9
472Ala	0.76	21.6	0.76	21.6	1.7	2.8	3.3
473Leu	-0.61	23.6	-0.61	23.6	2.8	3.7	4.6
474Leu	0.65	25.3	0.65	25.4	0.8	0.8	1.1
475Leu	0.63	25.6	0.63	25.6	2.1	1.6	2.7
491Leu	0.78	26.5	0.78	26.6	1.0	4.3	4.4
497Ile	0.67	12.7	0.66	12.7	3.7	3.2	5.0
499Leu	-0.16	26.8	-0.17	26.8	2.8	0.1	2.8
507Leu	0.58	26.8	0.59	26.8	0.8	1.2	1.4
514Ile	0.63	12.5	0.63	12.6	4.0	2.0	4.5
517Val	0.62	21.9	0.62	21.9	3.7	5.5	6.6
518Val	0.78	20.7	0.78	20.7	0.3	0.8	0.8
521Val	0.29	17.1	0.29	17.1	1.9	0.3	2.0
530Val	1.20	21.4	1.20	21.4	0.8	0.4	0.9

*missing resonances of the Ca^{2+} bound state in comparison to GCDCA bound P-domain correspond to Leu359 and Leu384. Furthermore, one of the two Val374 resonances is missing (see also Fig. S 4.1).

Tab. S 4.3: Chemical shifts and CSPs between the apo D410A P-domain and Mg^{2+} bound D410A P-domain.

See Fig. 5.4.6 for more information. Also see Tab. S 3.1 for information about color coding.

Amino acid	Apo P-domain		Ca^{2+} bound P-domain		^{1}H CSP in Hz	^{13}C CSP in Hz	total CSP in Hz
	^{1}H in ppm	^{13}C in ppm	^{1}H in ppm	^{13}C in ppm			
229Met	1.6	16.9	1.6	16.9	1	2	2.3
230Val	0.69	22.2	0.69	22.2	0.6	1.4	1.5
232Leu	0.31	22.5	0.27	22.4	n.d.	n.d.	n.d.
234Val	0.84	20.2	0.85	20.3	1.6	4.2	4.5
235Ile	0.58	13.6	0.58	13.6	0.6	0.3	0.6
252Leu	0.48	26.1	0.49	26.2	4.5	3	5.4
254Val	0.38	19.2	0.38	19.2	0.6	1.5	1.6
269Val	1.35	20.7	1.35	20.7	3.2	1.3	3.5
271Val	0.63	23.1	0.64	23.2	6.5	2.6	7
275Leu	0.7	23.7	0.69	23.6	2.4	5.6	6.1
281Ile	0.69	11.8	0.71	12	7.6	24.2	25.3
291Ala	1.21	24.5	1.21	24.5	n.d.	n.d.	n.d.
293Ala	1.16	23.8	1.15	23.7	n.d.	n.d.	n.d.
294Ala	n.d.	n.d.	1.21	21.2	n.d.	n.d.	n.d.
304Val	0.84	20.7	0.84	20.6	1	4.3	4.4
305Ala	0.1	22.4	0.09	22.4	n.d.	n.d.	n.d.
309Leu	0.63	23.4	0.63	23.5	1.8	10	10.2
323Ala	0.23	18	0.23	18	3	1.8	3.5
324Ala	1.17	19.5	1.17	19.5	0.5	0.2	0.5
335Leu	1.01	26.3	1.06	26.1	34.1	33.2	47.6
337Ile	n.d.	n.d.	0.66	13	n.d.	n.d.	n.d.
339Val	n.d.	n.d.	1.03	19.5	n.d.	n.d.	>20
352Val	n.d.	n.d.	0.77	18.7	n.d.	n.d.	>20
357Met	1.94	17.9	1.94	18.2	1.3	45.8	45.9

357Met	1.95	18	1.95	18.2	2.5	35.3	35.4
358Ile	0.96	12.5	0.97	12.7	n.d.	n.d.	n.d.
358Ile	0.96	12.5	0.99	12.6	n.d.	n.d.	>20
365Ala	1.37	17.1	1.38	17.1	5.3	3.5	6.3
365Ala	1.1	18.3	1.1	18.5	0.8	25.2	25.2
374Val	0.49	19.7	0.47	19.8	13.3	19.4	23.6
374Val	0.48	20	0.47	20.3	3.2	48	48.1
376Ala	1.56	24.8	1.58	24.8	n.d.	n.d.	n.d.
378Val	n.d.	n.d.	0.96	21.2	n.d.	n.d.	>20
380Ala	1.61	21.4	1.6	21.5	7.4	11.5	13.7
381Ala	1.43	19.1	1.43	19.1	0.4	2.7	2.7
382Ala	1.07	21.3	1.07	21.4	0	8.6	8.6
387Val	n.d.	n.d.	0.81	21.8	n.d.	n.d.	n.d.
391Val	0.62	18.3	0.64	18.6	13.5	51.5	53.2
393Ala	-0.28	19.9	-0.27	19.7	n.d.	n.d.	n.d.
398Ile	n.d.	n.d.	1.25	14.4	n.d.	n.d.	>20
398Ile	n.d.	n.d.	1.29	14.3	n.d.	n.d.	>20
405Ile	1.11	14.7	1.12	14.7	2.1	0.4	2.2
414Val	0.55	19.5	0.54	19.4	1.2	15.8	15.8
416Leu	n.d.	n.d.	0.8	23.6	n.d.	n.d.	n.d.
420Ile	0.83	13.9	0.83	14	1.1	7.1	7.2
428Val	0.98	20.8	0.98	20.8	0.7	4.4	4.5
429Leu	0.83	26.6	0.83	26.6	1.2	4	4.2
430Leu	0.82	27	0.82	27	0.6	3.5	3.5
436Met	0.02	16.9	0.11	17.4	53.2	71.5	89.1
439Ile	n.d.	n.d.	0.84	12.5	n.d.	n.d.	>20
442Ala	n.d.	n.d.	1.41	22.9	n.d.	n.d.	>20
444Ala	1.25	18.9	1.27	18.8	12.5	15.8	20.1
446Ala	n.d.	n.d.	0.97	19.5	n.d.	n.d.	>20
448Ala	1.13	19.9	1.15	19.9	9.4	2.4	9.7

449Ile	n.d.	n.d.	0.5	13.4	n.d.	n.d.	>20
449Ile	n.d.	n.d.	0.49	13.3	n.d.	n.d.	>20
452Ala	2.04	20	2.04	20	0.3	3.8	3.8
453Leu	0.96	23.4	0.96	23.4	1.3	0.2	1.3
462Ala	1.46	18.2	1.47	18.2	4.2	2.1	4.7
472Ala	0.76	21.6	0.76	21.6	0.9	2.1	2.3
473Leu	-0.61	23.6	-0.61	23.6	3.1	5.1	6
474Leu	0.65	25.4	0.65	25.4	0.8	0.3	0.9
475Leu	0.63	25.6	0.63	25.6	2.8	2.1	3.5
491Leu	0.78	26.6	0.79	26.5	1	4.8	4.9
497Ile	0.66	12.6	0.67	12.6	3.2	1.3	3.4
499Leu	-0.17	26.8	-0.17	26.8	0.5	0.1	0.5
507Leu	0.59	26.8	0.59	26.8	0	1.6	1.6
514Ile	0.63	12.4	0.63	12.5	4.6	3	5.5
517Val	0.62	21.9	0.62	21.9	2.4	7.9	8.2
518Val	0.78	20.7	0.78	20.7	0	2	2
521Val	0.29	17.1	0.29	17.1	1.5	0.3	1.5
530Val	1.2	21.4	1.2	21.4	1.2	3.1	3.4

Tab. S 4.4: Chemical shifts and CSPs between the Mg^{2+} bound WT (wildtype) P-domain and Mg^{2+} and GCDCA bound WT P-domain.

See Fig. 5.4.10 for more information. Also see Tab. S 3.1 for information about color coding.

Amino acid	Mg^{2+} bound P-domain		GCDCA and Mg^{2+} bound P-domain		^{1}H CSP in Hz	^{13}C CSP in Hz	total CSP in Hz
	^{1}H in ppm	^{13}C in ppm	^{1}H in ppm	^{13}C in ppm			
229Met	1.60	16.9	1.59	16.9	9.4	2.1	9.7
230Val	0.70	22.2	0.69	22.2	1.5	0.2	1.5
232Leu	0.27	22.4	0.28	22.5	8.0	7.8	11.2
234Val	0.85	20.3	0.85	20.3	0.4	1.4	1.4

235Ile	0.58	13.7	0.58	13.7	0.1	2.4	2.4
252Leu	0.49	26.2	0.50	26.1	2.6	10.4	10.8
254Val	0.38	19.2	0.38	19.2	0.9	1.7	1.9
269Val	1.35	20.7	1.35	20.7	2.0	2.6	3.3
271Val	0.64	23.1	0.65	23.2	3.8	4.6	5.9
275Leu	0.69	23.6	0.70	23.6	3.3	2.6	4.2
281Ile	0.71	12.2	0.69	11.8	13.8	47.9	49.8
291Ala	1.21	24.6	1.28	24.5	40.7	11.1	42.2
293Ala	1.15	23.7	1.07	24.2	46.5	69.2	83.4
294Ala	1.21	21.2	1.19	21.2	14.3	3.6	14.7
304Val	0.85	20.6	0.85	20.5	1.0	9.0	9.1
305Ala	0.09	22.5	0.08	22.8	2.9	45.6	45.7
309Leu	0.63	23.5	0.52	23.1	67.2	65.5	93.8
310Ile	0.81	14.2	0.81	14.2	0.3	2.6	2.6
323Ala	0.21	18.1	0.21	18.2	2.0	1.1	2.3
324Ala	1.17	19.5	1.19	19.5	13.9	5.4	14.9
335Leu	1.06	26.1	1.03	26.2	16.0	18.1	24.2
337Ile	0.67	13.1	0.78	14.5	67.0	201.4	212.3
339Val	1.03	19.5	1.06	19.5	18.7	6.4	19.8
352Val	0.77	18.7	0.73	19.0	23.4	47.5	53.0
357Met	1.95	18.3	1.96	18.0	5.6	43.1	43.4
357Met	1.96	18.2	1.97	17.9	5.4	39.3	39.7
358Ile	0.96	12.8	0.96	12.8	0.3	6.1	6.1
358Ile	1.01	12.9	1.01	12.9	4.2	9.4	10.3
359Leu	n.d.	n.d.	0.77	22.1	n.d.	n.d.	n.d.
365Ala	1.40	19.4	1.40	19.5	4.8	16.0	16.7
365Ala	1.20	19.3	1.20	19.3	1.7	1.8	2.5
374Val	0.44	19.9	0.45	19.7	7.4	35.1	35.8
374Val	0.48	20.3	0.49	20.0	3.3	40.5	40.7
376Ala	1.57	24.9	1.56	25.0	8.4	13.5	15.9

378Val	0.96	21.2	0.95	21.2	4.4	1.3	4.6
380Ala	1.60	21.5	1.60	21.5	1.2	1.8	2.2
381Ala	1.43	19.1	1.43	19.1	0.9	2.2	2.4
382Ala	1.07	21.4	1.07	21.4	0.1	3.1	3.1
387Val	0.81	21.8	0.80	21.8	6.5	7.5	9.9
391Val	0.65	18.6	0.64	18.8	6.5	28.8	29.5
393Ala	-0.27	19.7	-0.48	19.7	124.9	4.6	125.0
398Ile	1.28	14.4	1.30	14.3	13.2	13.8	19.1
398Ile	1.25	14.5	1.27	14.4	12.4	14.7	19.3
405Ile	1.12	14.7	1.13	14.6	3.0	14.8	15.1
405Ile	n.d.	n.d.	1.12	14.7	n.d.	n.d.	n.d.
414Val	0.53	18.8	0.53	18.6	2.9	23.7	23.9
414Val	0.54	19.2	0.54	18.9	2.6	30.3	30.4
416Leu	0.79	23.5	0.80	23.5	5.9	1.9	6.2
420Ile	0.83	14.1	0.83	14.1	1.3	3.3	3.6
428Val	0.98	20.8	0.98	20.8	1.9	6.5	6.8
429Leu	0.83	26.6	0.83	26.7	0.9	7.4	7.5
430Leu	0.82	27.0	0.82	27.0	1.6	2.9	3.3
436Met	0.11	17.4	-0.07	16.9	111.4	76.7	135.3
439Ile	0.84	12.6	0.83	12.3	5.2	46.1	46.4
442Ala	1.41	22.9	1.42	22.9	3.9	8.1	9.0
444Ala	1.27	18.8	1.27	18.8	1.5	8.8	8.9
446Ala	0.97	19.5	0.98	19.0	5.6	78.5	78.7
448Ala	1.16	19.9	1.13	19.9	16.8	1.1	16.8
449Ile	0.49	13.5	0.49	13.8	3.0	53.9	53.9
449Ile	0.51	13.4	0.52	13.8	5.8	53.0	53.3
452Ala	2.05	20.0	2.04	20.1	1.7	4.2	4.6
453Leu	0.96	23.4	0.95	23.4	2.8	2.0	3.4
462Ala	1.47	18.2	1.46	18.1	0.8	12.7	12.7
472Ala	0.76	21.6	0.76	21.6	0.1	1.8	1.8

473Leu	-0.61	23.6	-0.61	23.6	0.2	2.7	2.7
474Leu	0.65	25.4	0.65	25.3	2.3	0.4	2.3
475Leu	0.63	25.6	0.62	25.6	1.0	0.5	1.1
491Leu	0.79	26.5	0.79	26.5	3.8	0.9	3.9
497Ile	0.67	12.7	0.66	12.7	6.9	4.6	8.3
499Leu	-0.17	26.8	-0.19	26.9	11.0	5.3	12.2
507Leu	0.59	26.8	0.58	26.8	4.3	2.0	4.7
514Ile	0.64	12.6	0.63	12.6	2.9	7.2	7.7
517Val	0.62	21.9	0.62	21.9	0.1	1.0	1.0
518Val	0.78	20.7	0.78	20.7	0.9	1.3	1.6
521Val	0.29	17.1	0.29	17.1	1.0	0.0	1.0
530Val	1.20	21.4	1.20	21.4	0.9	6.0	6.1

*the only missing resonance of the Mg^{2+} and GCDCA bound form compared to the GCDCA bound form is Leu387.

Tab. S 4.5: Chemical shifts and CSPs between the Ca^{2+} bound WT P-domain and Ca^{2+} and GCDCA bound WT P-domain.

See Fig. S 4.9 for more information. Also see Tab. S 3.1 for information about color coding.

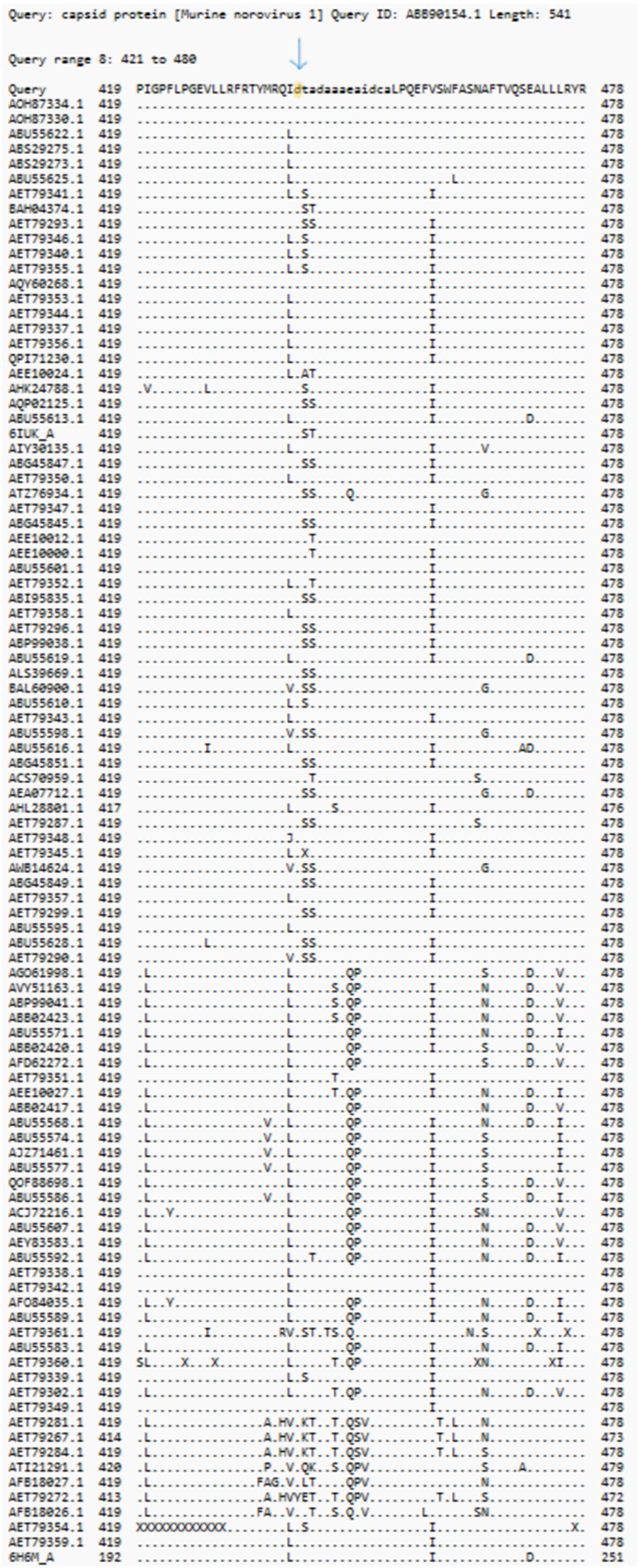
Amino acid	Mg^{2+} bound P-domain		GCDCA and Mg^{2+} bound P-domain		^{1}H CSP in Hz	^{13}C CSP in Hz	total CSP in Hz
	^{1}H in ppm	^{13}C in ppm	^{1}H in ppm	^{13}C in ppm			
229Met	1.61	16.9	1.60	16.9	8.1	1.6	8.2
230Val	0.69	22.2	0.69	22.2	0.1	1.1	1.2
232Leu	0.26	22.4	0.28	22.5	7.7	9.1	11.9
234Val	0.84	20.3	0.84	20.3	0.3	1.0	1.0
235Ile	0.58	13.7	0.58	13.7	0.9	0.1	0.9
252Leu	0.49	26.1	0.50	26.1	4.5	9.1	10.1
254Val	0.38	19.2	0.38	19.2	2.3	0.1	2.3
269Val	1.35	20.7	1.35	20.7	1.4	4.3	4.6
271Val	0.64	23.1	0.65	23.1	4.4	1.7	4.7
275Leu	0.69	23.6	0.70	23.6	4.2	2.6	4.9
281Ile	0.71	12.2	0.69	11.9	12.6	37.9	40.0
291Ala	1.21	24.5	1.27	24.5	39.7	6.4	40.2
293Ala	1.15	23.7	1.08	24.2	45.3	65.3	79.5
294Ala	1.21	21.2	1.19	21.2	11.9	3.0	12.3
304Val	0.85	20.6	0.85	20.6	0.1	7.6	7.6
305Ala	0.09	22.5	0.09	22.8	1.7	46.9	46.9
309Leu	0.63	23.5	0.52	23.0	67.3	63.9	92.8
310Ile	0.82	14.2	0.82	14.2	0.1	2.8	2.8
323Ala	0.21	18.1	0.21	18.1	0.5	1.1	1.2
324Ala	1.17	19.5	1.19	19.5	14.5	3.1	14.8
335Leu	1.06	26.0	1.03	26.2	14.0	19.7	24.2
337Ile	0.66	13.2	0.78	14.4	73.3	194.7	208.1
339Val	1.02	19.5	1.05	19.5	19.1	4.9	19.7
352Val	0.77	18.7	0.73	19.0	22.7	41.5	47.3

357Met	1.94	18.4	1.95	18.1	7.9	44.6	45.3
357Met	1.95	18.2	1.96	17.9	5.2	40.5	40.8
358Ile	0.96	12.9	0.96	12.9	0.4	7.7	7.7
358Ile	1.00	12.9	1.01	13.0	5.2	11.0	12.2
359Leu	n.d.	n.d.	0.77	22.2	n.d.	n.d.	n.d.
365Ala	1.39	19.6	1.39	19.7	3.9	13.8	14.4
365Ala	1.22	19.5	1.23	19.5	3.0	1.9	3.5
374Val	0.46	20.0	0.47	19.7	5.6	36.8	37.3
374Val	n.d.	n.d.	0.50	20.0	n.d.	n.d.	n.d.
376Ala	1.57	24.9	1.55	25.0	10.5	9.4	14.1
378Val	0.95	21.2	0.95	21.2	0.8	0.3	0.8
380Ala	1.60	21.5	1.59	21.5	1.3	0.4	1.4
381Ala	1.43	19.1	1.43	19.1	0.6	3.0	3.1
382Ala	1.07	21.4	1.07	21.4	0.9	2.7	2.9
384Leu	n.d.	n.d.	0.80	22.3	n.d.	n.d.	n.d.
387Val	0.81	21.7	0.80	21.7	6.2	2.7	6.8
391Val	0.63	18.6	0.63	18.8	2.6	23.1	23.2
393Ala	-0.27	19.9	-0.47	19.9	118.0	0.4	118.0
398Ile	1.25	14.7	1.33	14.4	42.1	48.7	64.3
398Ile	1.29	14.6	1.29	14.6	4.6	2.2	5.1
405Ile	1.12	14.6	1.13	14.5	5.1	16.4	17.2
405Ile	1.13	14.9	1.14	14.7	5.9	24.5	25.2
414Val	0.53	18.8	0.54	18.6	3.9	21.8	22.1
414Val	0.54	19.1	0.54	18.9	3.9	29.2	29.5
416Leu	0.79	23.5	0.81	23.5	7.9	1.4	8.0
420Ile	0.83	14.1	0.83	14.2	0.8	7.3	7.3
428Val	0.98	20.8	0.98	20.7	0.6	8.1	8.1
429Leu	0.83	26.6	0.83	26.7	1.2	10.3	10.4
430Leu	0.82	27.0	0.82	27.0	1.5	3.5	3.8
436Met	-0.03	17.2	-0.20	16.3	101.9	134.4	168.7

439Ile	0.84	13.0	0.84	12.7	0.8	43.6	43.6
442Ala	1.41	22.9	1.41	23.0	0.5	16.5	16.5
444Ala	1.26	19.0	1.26	19.0	0.4	5.4	5.4
446Ala	0.92	19.2	0.90	18.4	7.6	108.9	109.2
448Ala	1.15	20.1	1.14	20.1	7.3	1.8	7.5
449Ile	0.53	13.5	0.55	13.8	13.6	45.9	47.9
449Ile	0.51	13.5	0.52	13.9	8.5	48.2	49.0
452Ala	2.04	20.0	2.04	20.1	0.8	4.5	4.6
453Leu	0.96	23.4	0.95	23.4	2.7	1.8	3.2
462Ala	1.46	18.2	1.46	18.1	0.1	10.3	10.3
472Ala	0.76	21.6	0.76	21.6	0.4	0.3	0.5
473Leu	-0.61	23.6	-0.61	23.6	0.9	1.8	2.0
474Leu	0.65	25.3	0.65	25.4	0.5	0.9	1.1
475Leu	0.63	25.6	0.62	25.6	0.5	1.3	1.4
491Leu	0.78	26.5	0.79	26.5	3.2	2.1	3.8
497Ile	0.67	12.7	0.66	12.7	7.0	6.2	9.4
499Leu	-0.16	26.8	-0.19	26.9	12.5	5.4	13.6
507Leu	0.58	26.8	0.58	26.8	3.8	2.1	4.4
514Ile	0.63	12.5	0.63	12.6	2.4	9.6	9.8
517Val	0.62	21.9	0.62	21.9	0.5	1.4	1.4
518Val	0.78	20.7	0.78	20.7	0.6	0.8	1.0
521Val	0.29	17.1	0.29	17.1	0.3	2.2	2.2
530Val	1.20	21.4	1.20	21.4	0.4	5.8	5.9

*no missing resonances of Ca^{2+} and GCDCA bound form compared to the GCDCA bound form.

4.5 Alignment of MNV P-domain amino acid sequences (aa 419-478)



```
Query: capsid protein [Murine norovirus 1] Query ID: ABB90154.1 Length: 541

Query range 8: 421 to 480

Query       419  PIGPFLPGEVLLRFRTYMRQIdtadaaaeaidcaLPQEFVSWFASNAFTVQSEALLLRYR  478
AOH87334.1  419  ..............................................................  478
AOH87330.1  419  ..............................................................  478
ABU55622.1  419  ..................L...........................................  478
ABS29275.1  419  ..................L...........................................  478
ABS29273.1  419  ..................L...........................................  478
ABU55625.1  419  ..................L.....................L.....................  478
AET79341.1  419  ..................L.S.................I.....................  478
BAH04374.1  419  ....................ST.....................................  478
AET79293.1  419  ....................SS.................I.....................  478
AET79346.1  419  ..................L.S.................I.....................  478
AET79340.1  419  ..................L.S.................I.....................  478
AET79355.1  419  ..................L.S.................I.....................  478
AQV60268.1  419  ..............................................I.....................  478
AET79353.1  419  ..................L.................I.....................  478
AET79344.1  419  ..................L.................I.....................  478
AET79337.1  419  ..................L.................I.....................  478
AET79356.1  419  ..................L.................I.....................  478
QPI71230.1  419  ..................L.................I.....................  478
AEE10024.1  419  ..................L.AT..............................................  478
AHK24788.1  419  .V.......L.............S.................I.....................  478
AQP02125.1  419  ....................SS.................I.....................  478
ABU55613.1  419  ..................L.................I...........D.......  478
6IUK_A      419  ....................ST.....................................  478
AIV30135.1  419  ..................L.................I.....V.............  478
ABG45847.1  419  ....................SS.................I.....................  478
AET79350.1  419  ..................L.................I.....................  478
ATZ76934.1  419  ....................SS....Q.................G.............  478
AET79347.1  419  ..............................................I.....................  478
ABG45845.1  419  ....................SS.................I.....................  478
AEE10012.1  419  ....................T.....................................  478
AEE10000.1  419  ....................T.................I.....................  478
ABU55601.1  419  ..............................................I.....................  478
AET79352.1  419  ..................L..T.................I.....................  478
ABI95835.1  419  ....................SS.................I.....................  478
AET79358.1  419  ..................L.................I.....................  478
AET79296.1  419  ......................SS.............I.....................  478
ABP99038.1  419  ....................SS.................I.....................  478
ABU55619.1  419  ..................L.................I...........D.......  478
ALS39669.1  419  ....................SS.....................................  478
BAL60900.1  419  ..................V.SS.....................G.............  478
ABU55610.1  419  ..................L.S.....................................  478
AET79343.1  419  ..................L.................I.....................  478
ABU55598.1  419  ..................V.SS.....................G.............  478
ABU55616.1  419  .........I.........L.................I...........AD.......  478
ABG45851.1  419  ....................SS.................I.....................  478
ACS70959.1  419  ....................T.....................S...............  478
AEA07712.1  419  ....................SS.....................G.....D.......  478
AHL28801.1  417  ..................L.....S...........I.....................  476
AET79287.1  419  ....................SS.....................S...............  478
AET79348.1  419  ....................J.................I.....................  478
AET79345.1  419  ..................L.X.................I.....................  478
AWB14624.1  419  ..................V.SS.....................G.............  478
ABG45849.1  419  ....................SS.................I.....................  478
AET79357.1  419  ..................L.................I.....................  478
AET79299.1  419  ....................SS.................I.....................  478
ABU55595.1  419  ..................L.....................................  478
ABU55628.1  419  .........L.............SS.................I.....................  478
AET79290.1  419  ..................V.SS.................I.....................  478
AGO61998.1  419  .L..................L.......QP.................S.....D...V...  478
AVY51163.1  419  .L..................L.....S.QP.........I.....N.....D...V...  478
ABP99041.1  419  .L..................L.....S.QP.........I.....N.....D...V...  478
ABB02423.1  419  .L..................L.....S.QP.........I.....N.....D...V...  478
ABU55571.1  419  .L..................L.......QP.........I.....N.....D...I...  478
ABB02420.1  419  .L..................L.......QP.........I.....S.....D...V...  478
AFD62272.1  419  .L..................L.......QP.................S.....D...V...  478
AET79351.1  419  ..................L.....T.............I.....................  478
AEE10027.1  419  .L..................L.....T.QP.........I.....N.....D...I...  478
ABB02417.1  419  .L..................L.......QP.................N.....D...V...  478
ABU55568.1  419  .L................V..L.......QP.........I.....N.....D...I...  478
ABU55574.1  419  .L................V..L.......QP.........I.....S.........I...  478
AJJ71461.1  419  .L................V..L.......QP.........I.....S.........I...  478
ABU55577.1  419  .L................V..L.......QP.........I.....S.........I...  478
QOF88698.1  419  .L..................L.......QP.........I.....S.....D...V...  478
ABU55586.1  419  .L................V..L.......QP.........I.....S.....D...I...  478
ACJ72216.1  419  .L..Y...............L.......QP.........I.....SN........V...  478
ABU55607.1  419  .L..................L.......QP.........I.....N.....D...V...  478
AEY83583.1  419  .L..................L.......QP.........I.....N.....D...V...  478
ABU55592.1  419  .L..................L..T....QP.........I.....N.....D...I...  478
AET79338.1  419  ..................L.....................I.....................  478
AET79342.1  419  ..................L.....................I.....................  478
AFO84035.1  419  .L..Y...............L.......QP.........I.....N.....D...I...  478
ABU55589.1  419  .L..................L.......QP.........I.....N.....D...I...  478
AET79361.1  419  .........I.........RV.ST.TS.Q.................N.S......X...X..  478
ABU55583.1  419  .L..................L.......QP.........I.....N.....D...I...  478
AET79360.1  419  SL....X...X.........L.....T.QP.........I.....XN........XI...  478
AET79339.1  419  ..................L.S.....................I.....................  478
AET79302.1  419  .L..................L.....T.QP.........I.....N.....D...V...  478
AET79349.1  419  ..............................................I.....................  478
AET79281.1  419  .L................A.HV.KT..T.QSV.........T.L...N.............  478
AET79267.1  414  .L................A.HV.KT..T.QSV.........T.L...N.............  473
AET79284.1  419  .L................A.HV.KT..T.QSV.........T.L...S.............  478
ATI21291.1  420  .L................P..V.QK..S.QPV.................S....A.......  479
AFB18027.1  419  .L................FAG.V.LT....QPV.................N.............  478
AET79272.1  413  .L................A.HVYET..T.QPV.........T.L...S.............  472
AFB18026.1  419  .L................FA..V..T..S.Q.V.......L......SN............  478
AET79354.1  419  XXXXXXXXXXX.........L.S.................I...................X.  478
AET79359.1  419  ..................L.....................I.....................  478
6H6M_A      192  ..................L.....................I...........D.......  251
```

Fig. S 4.13: Alignment of MNV P-domain amino acid sequences using the BLAST tool of the GenBank (Benson et al., 2013).

Asp440 of the P-domain is highly conserved in MNV strains. Asp440 is highlighted with a blue arrow.

5 P-domain pH interactions

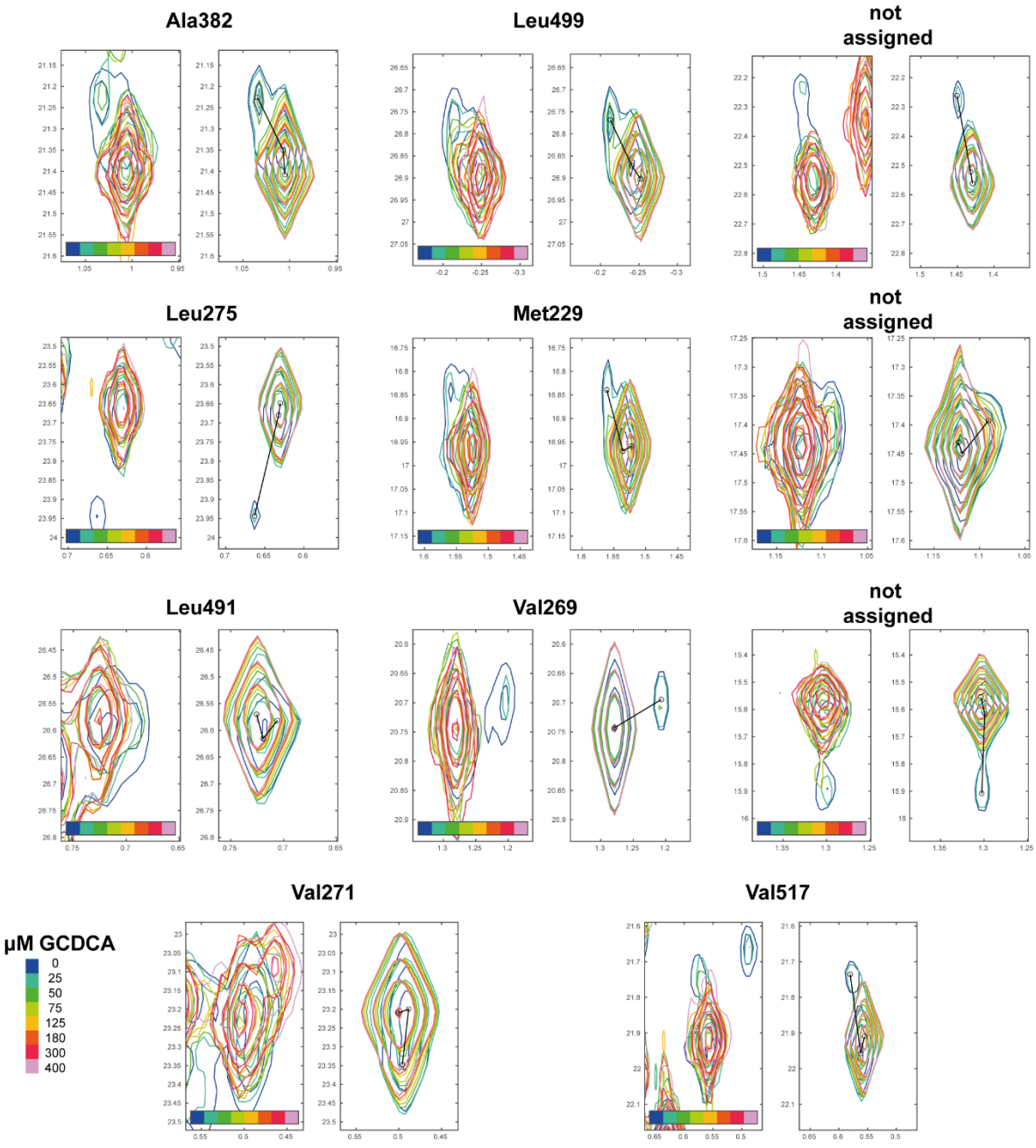


Fig. S 5.1: 2D lineshape *TITAN* analysis of GCDCA binding to P-domain at a pH_{corr} of 5.3 in citric acid buffer.

Spectra were acquired at 298 K on a 600 MHz spectrometer with cryo probe using 53 uM MILVA labelled MNV CW1 P-domain in 20 mM deuterated citric acid, 100 mM NaCl, pH_{corr} 5.3. The analysis results in a $K_{D,GCDCA}$ of 15 +-1 uM and a $k_{off,GCDCA}$ of 17 +-4 s^{-1}. The obtained dissociation constant and

off-rate translate into an on-rate $k_{on,GCDCA}$ of $1.1*10^6$ $M^{-1}s^{-1}$. The left panels show the measured spectra and the right panels show the simulated spectra.

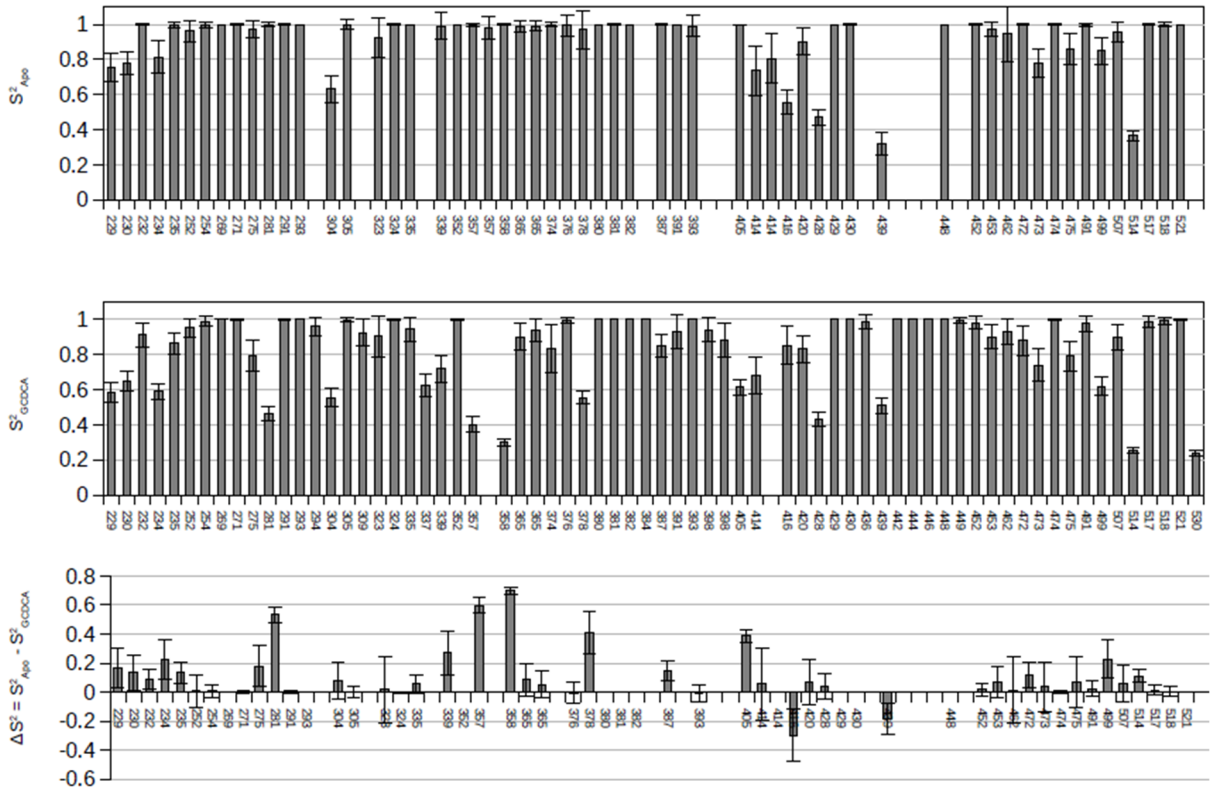


Fig. S 5.2: Methyl group order parameters S^2 of apo P-domain and GCDCA bound P-domain.

See Fig. 5.5.7 for more information. Methyl groups for that two peaks are observed due to Pro361 cis/trans isomerization were treated as follows: For Met357 and Val414, there was only one peak for the GCDCA bound state. The apo states order parameters were almost identical and were averaged in the diagram for comparison. For Ala365, both peaks were visible in both states and the assignment was transferred. The respective order parameters were compared. For Ile358, only one resonance was visible for the apo state. The Ile358 peak of the bound state being closer to the apo peak was used for comparison. Val374 was not included for the analysis shown in the main part as it was not clear which of the two apo peaks corresponded to which of the two bound state peaks. For Ile405 there was only one peak for the apo and bound P-domain forms, respectively. The corresponding order parameters were compared. For Ile449 there was no resonance in the apo state and, hence, it was excluded from the analysis. Ile398 was excluded from the analyses as no peaks were identified.

6 SARS-CoV-2 Carbohydrate interactions

Tab. S 6.1: Dissociation constants derived from STD-Afs

Ligand	Proton	K_D in mM*
3'SL	H1Gal	12 ± 1
	H3Gal	10.9 ± 0.6
	H2Glc	21 ± 3
	H3eqNeu5Ac	**10.1 ± 0.4**
	NAcNeu5Ac	13.8 ± 0.2
	H3axNeu5Ac	10.4 ± 0.4
6'SL	H1Gal	17 ± 2
	H1Glc	17 ± 4
	H2Glc	25 ± 4
	H3eqNeu5Ac	**16 ± 1**
	NAcNeu5Ac	23.9 ± 0.5
	H3axNeu5Ac	16 ± 1
Sialyl-Lewis^X	CH3Fuc	20 ± 10
	H3eqNeu5Ac	**14 ± 2**
	NAcNeu5Ac	15.8 ± 0.8
	H3axNeu5Ac	15 ± 2
Oseltamivir	H9	7.4 ± 0.6
	H16	5.1 ± 0.7
	H6	4.7 ± 0.8
	H6	4.5 ± 0.7
	H4	5.9 ± 0.7
	H8	7.2 ± 0.6
	H3	**4.1 ± 0.5**
	H2	7.4 ± 0.8
Oseltamivir in presence of 30 mM 6'SL	H9	10 ± 1
	H6	**4.0 ± 0.8**
	H6	4.2 ± 0.8
	H8	11 ± 2
	H3	7 ± 1
	H2	9 ± 1

*errors were rounded on the first digit not being zero. The dissociation constant was rounded accordingly.

6.1 Negative controls of STD NMR experiments

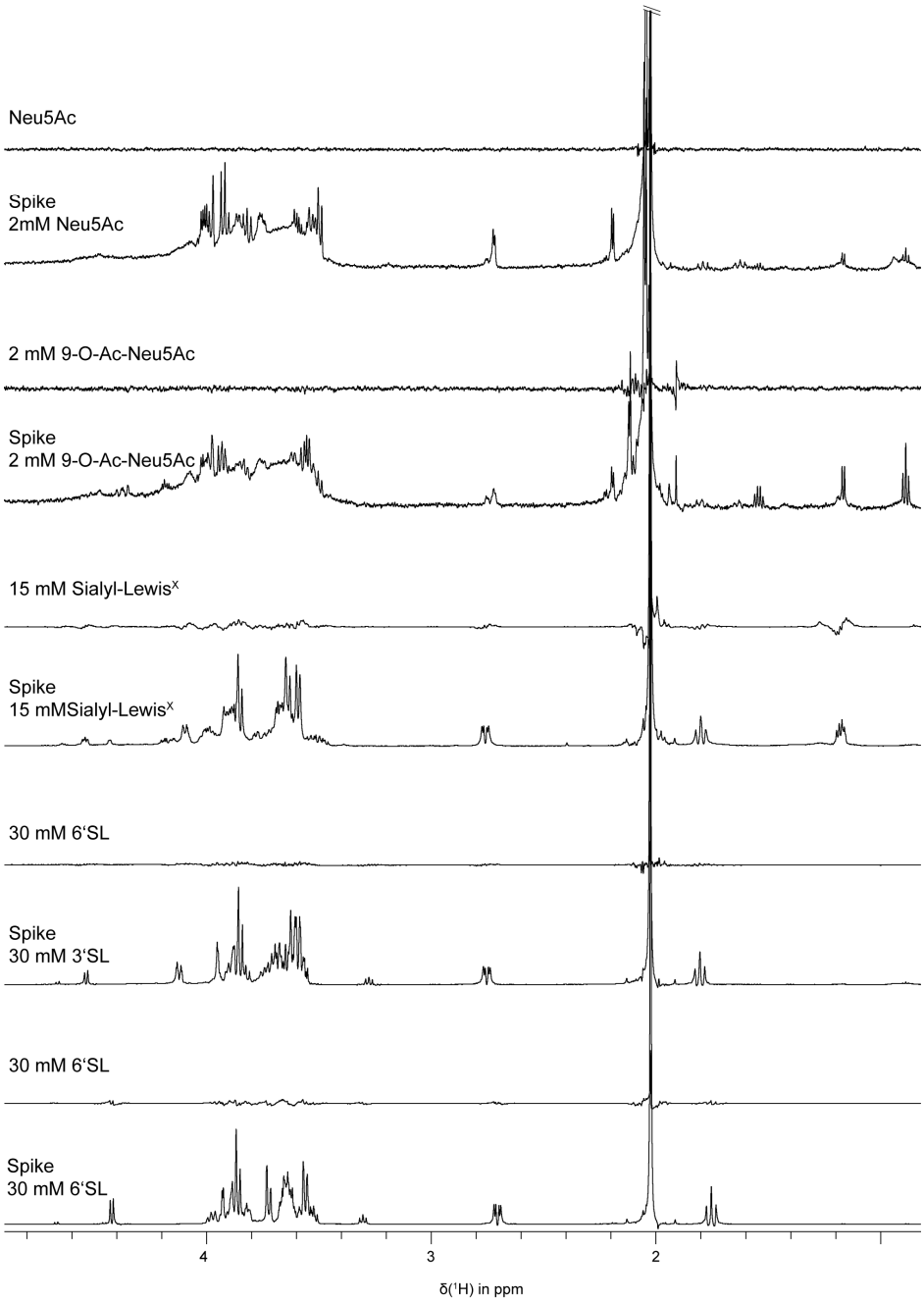


Fig. S 6.1: **Negative controls for STD experiments with sialoglycans shown in Figs. 5.6.2 and 5.6.7.**

Spectra of negative controls were acquired using the same ligand concentrations, receiver gains, and number of scans as for the respective samples containing the spike protein. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/). The Sialyl-Lewis^X data, the Neu5Ac data, and the 9-O-Ac-Neu5Ac data were added to the figure.

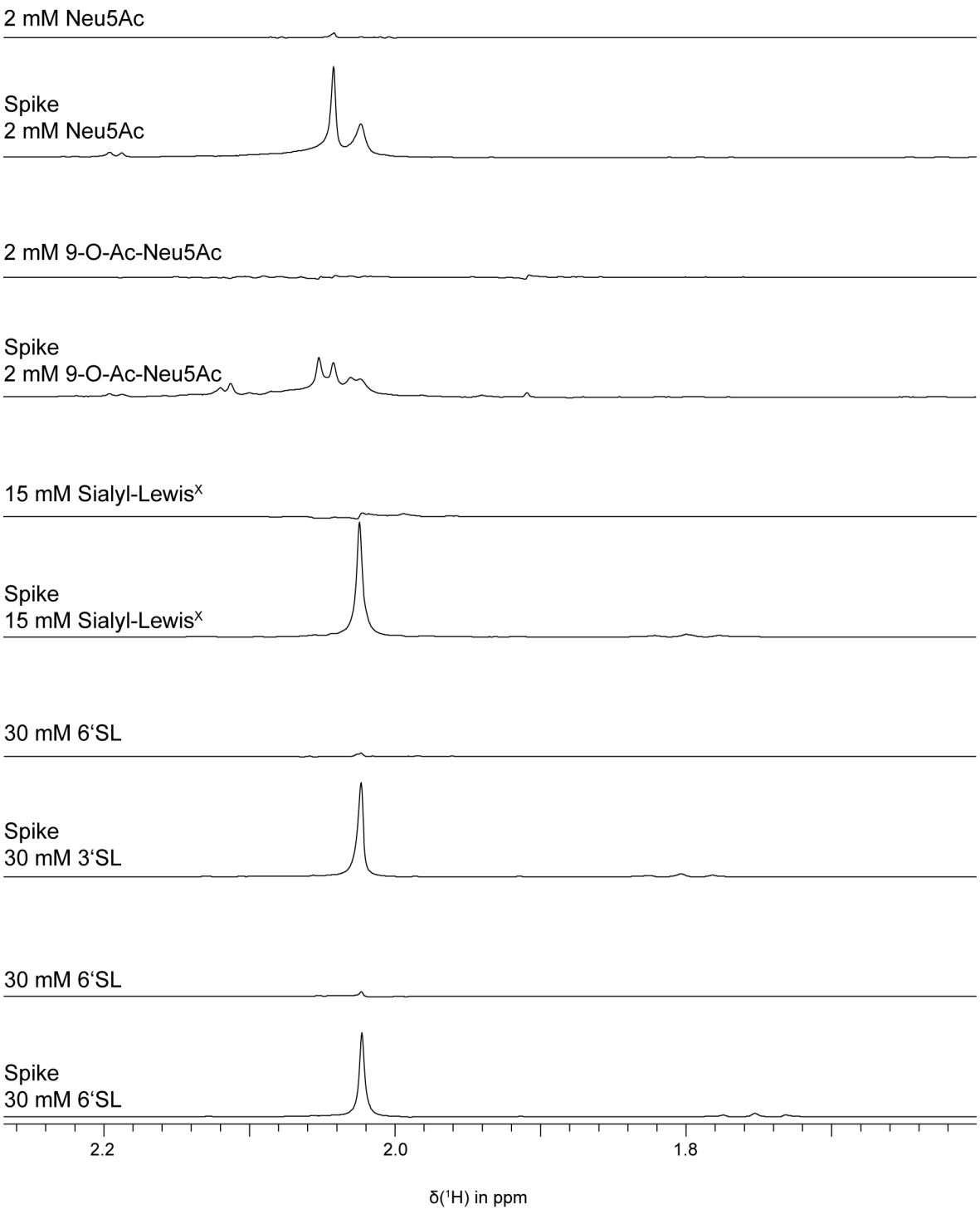


Fig. S 6.2: Negative controls for STD experiments with sialoglycans with emphasis on the spectral region of acetyl groups shown in Figs. 5.6.2 and 5.6.7.

Spectra of negative controls were acquired using the same ligand concentrations, receiver gains, and number of scans as for the respective samples containing the spike protein. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/). The Sialyl-Lewis^X data, the Neu5Ac data, and the 9-O-Ac-Neu5Ac data were added to the figure.

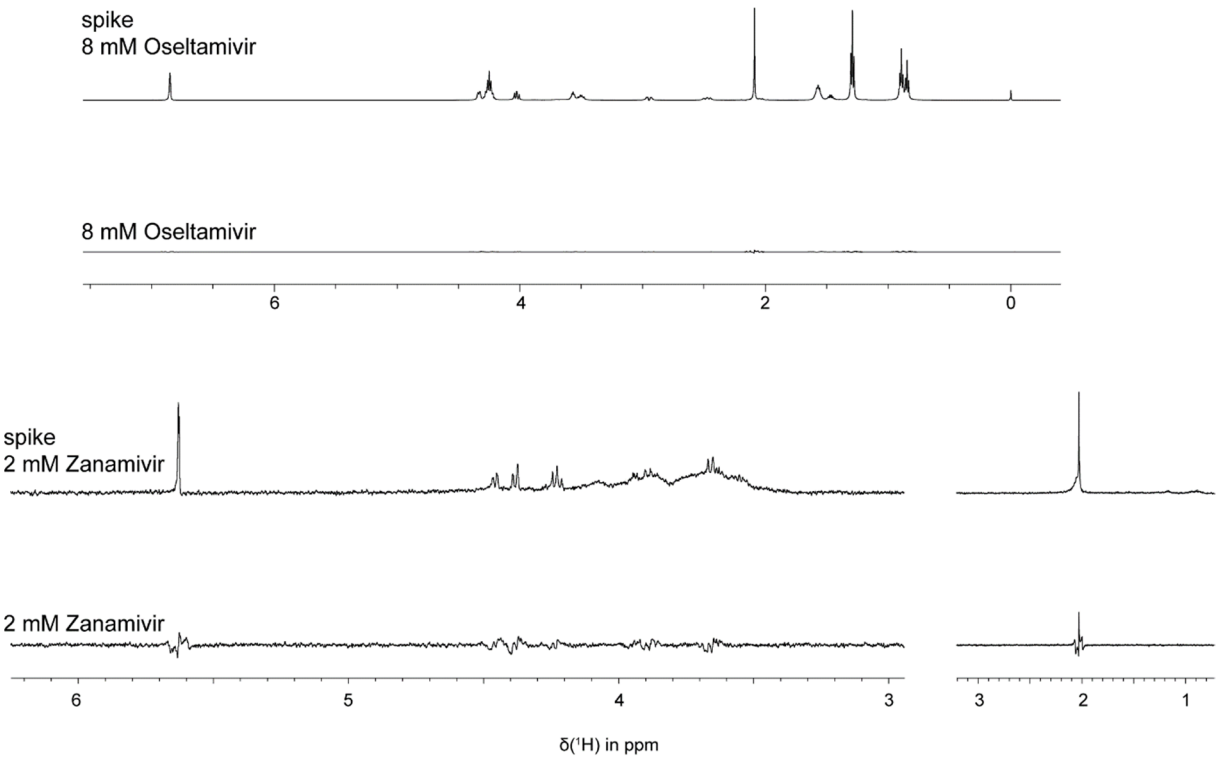


Fig. S 6.3: Negative controls for STD experiments with Oseltamivir and Relenza shown in Figs. 5.6.4 and 5.6.5.

Spectra of negative controls were acquired using the same ligand concentrations, receiver gains, and number of scans as for the respective samples containing the spike protein. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

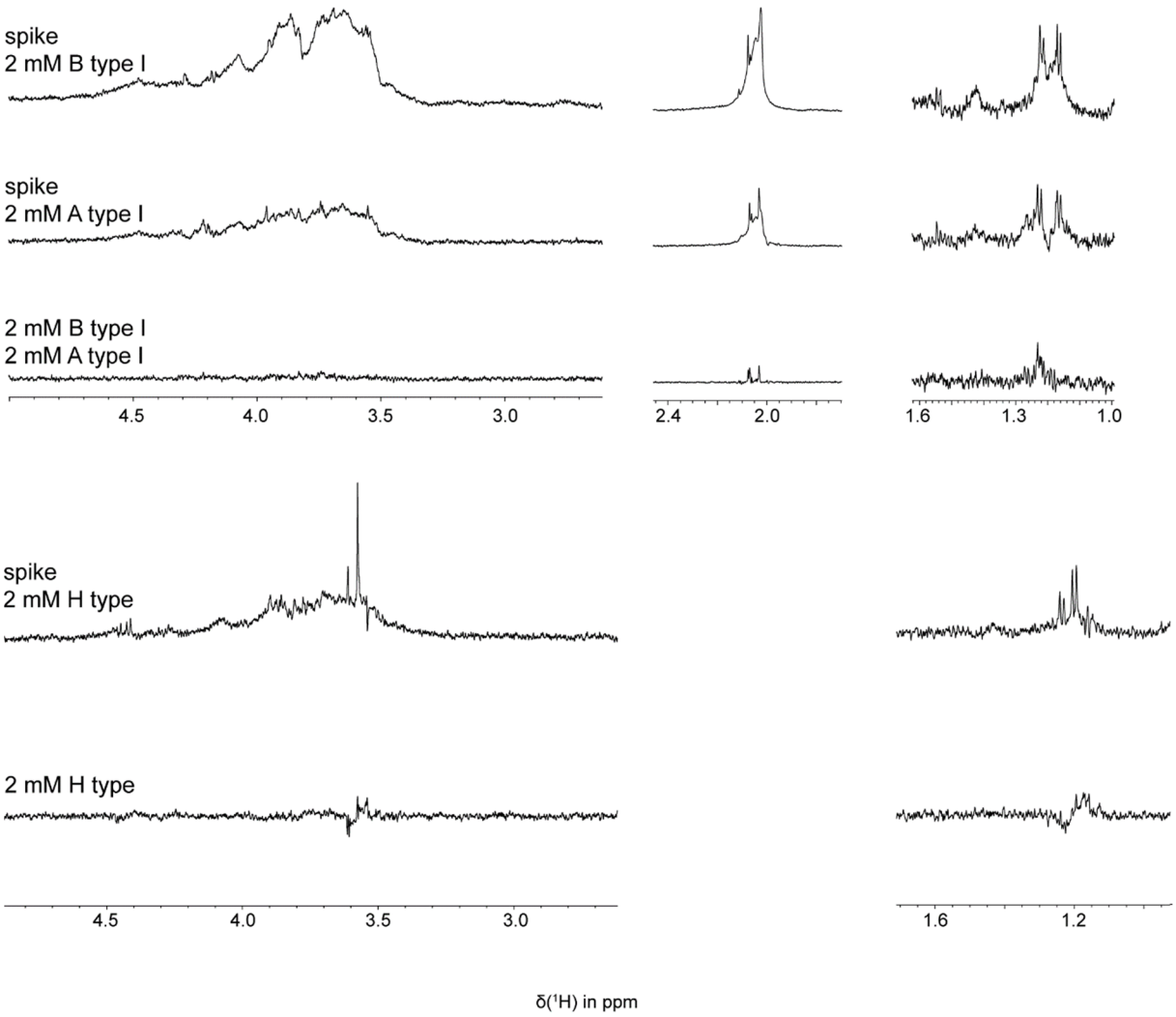


Fig. S 6.4: Negative controls for STD experiments with human ABO(H) blood group antigens shown in Fig. 5.6.10.

Spectra of negative controls were acquired using the same ligand concentrations, receiver gains, and number of scans as for the respective samples containing the spike protein. The figure is adapted from Maass et al., 2022b, under creative common attributions licence (link to the Creative Common licence: http://creativecommons.org/licenses/by/4.0/).

6.2 SARS-CoV-2 spike protein sequences

VNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYF
ASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEY
VSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLT
PGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVR
FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDE
VRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGV
EGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYS
TGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSIIAYTMSLGAENSVAYSNNSIAIPT
NFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIK
DFGGFNFSQILPDPSKPSKRSPIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYT

SALLAGTITSGWTFGAGPALQIPFPMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTPSALGKLQDV
VNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM
SECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFV
TQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEI
DRLNEVAKNLNESLIDLQELGKYEQ

Fig. S 6.5: Wuhan SARS-CoV-2 spike protein amino acid sequence.

VNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHVISGTNGTKRFDNPVLPFNDGVYFASI
EKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDHKNNKSWMESEFRVYSSANNCTFEYVSQPFL
MDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVREPEDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSS
SGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN
LCPFDEVFNATRFASVYAWNRKRISNCVADYSVLYNLAPFFTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAP
GQTGNIADYNYKLPDDFTGCVIAWNSNKLDSKVSGNYNYLYRLFRKSNLKPFERDISTEIYQAGNKPCNGVAGFNCY
FPLKSYSFRPTYGVGHQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLKGTGVLTESNKKFLPFQQFG
RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVF
QTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTKSHGSASSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVT
TEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLKRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKYFGGFN
FSQILPDPSKPSKRSPIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFKGLTVLPPLLTDEMIAQYTSALLAGTI
TSGWTFGAGPALQIPFPMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTPSALGKLQDVVNHNAQ
ALNTLVKQLSSKFGAISSVLNDIFSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS
KRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYE
PQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVA
KNLNESLIDLQELGKYEQ

Fig. S 6.6: Omicron BA1 SARS-CoV-2 spike protein amino acid sequence.

7 Bibliography

- Alfaro JA, Zheng RB, Persson M, Letts JA, Polakowski R, Bai Y, Borisova SN, Seto NO, Lowary TL, Palcic MM, Evans SV. ABO(H) blood group A and B glycosyltransferases recognize substrate via specific conformational changes. J Biol Chem. 2008 Apr 11;283(15):10097-108. doi: 10.1074/jbc.M708669200. Epub 2008 Jan 11. PMID: 18192272.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2013 Jan;41(Database issue):D36-42. doi: 10.1093/nar/gks1195. Epub 2012 Nov 27. PMID: 23193287; PMCID: PMC3531190.
- Chao FA, Shi L, Masterson LR, Veglia G. FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. J Magn Reson. 2012 Jan;214(1):103-10. doi: 10.1016/j.jmr.2011.10.008. Epub 2011 Oct 20. PMID: 22134225; PMCID: PMC3487468.
- Chao FA, Kim J, Xia Y, Milligan M, Rowe N, Veglia G. FLAMEnGO 2.0: an enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances. J Magn Reson. 2014 Aug;245:17-23. doi: 10.1016/j.jmr.2014.04.012. Epub 2014 May 2. PMID: 24915505; PMCID: PMC4161213.
- Cheng Y, Prusoff WH. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol. 1973 Dec 1;22(23):3099-108. doi: 10.1016/0006-2952(73)90196-2. PMID: 4202581.

- Creutznacher R, Maass T, Ogrissek P, Wallmann G, Feldmann C, Peters H, Lingemann M, Taube S, Peters T, Mallagaray A. NMR Experiments Shed New Light on Glycan Recognition by Human and Murine Norovirus Capsid Proteins. Viruses. 2021 Mar 5;13(3):416. doi: 10.3390/v13030416. PMID: 33807801; PMCID: PMC8001558.
- Creutznacher R, Maass T, Dülfer J, Feldmann C, Hartmann V, Lane MS, Knickmann J, Westermann LT, Thiede L, Smith TJ, Uetrecht C, Mallagaray A, Waudby CA, Taube S, Peters T. Distinct dissociation rates of murine and human norovirus P-domain dimers suggest a role of dimer stability in virus-host interactions. Commun Biol. 2022 Jun 9;5(1):563. doi: 10.1038/s42003-022-03497-4. PMID: 35680964; PMCID: PMC9184547.
- Duan X, Quiocho FA. Structural evidence for a dominant role of nonpolar interactions in the binding of a transport/chemosensory receptor to its highly polar ligands. Biochemistry. 2002 Jan 22;41(3):706-12. doi: 10.1021/bi015784n. PMID: 11790091.
- Flügge F, Peters T. Complete assignment of Ala, Ile, Leu, Met and Val methyl groups of human blood group A and B glycosyltransferases using lanthanide-induced pseudocontact shifts and methyl-methyl NOESY. J Biomol NMR. 2018 Apr;70(4):245-259. doi: 10.1007/s10858-018-0183-4. Epub 2018 Apr 26. PMID: 29700756.
- Flügge F. Investigation into Substrate Binding and Structural Dynamics of Human Blood Group A and B Glycosyltransferases by NMR spectroscopy. Doctoral thesis, University of Lübeck, 2018.
- Förster A, Masters EI, Whitby FG, Robinson H, Hill CP. The 1.9 A structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. Mol Cell. 2005 May 27;18(5):589-99. doi: 10.1016/j.molcel.2005.04.016. PMID: 15916965.
- Führing J, Cramer JZ, Routier FH, Lamerz AC, Brauch P, Gerardy-Schahn R, Federov R. Catalytic Mechanism and Allosteric Regulation of UDP-Glucose Pyrophosphorylase from Leishmania major. ACS Catal. 2013 Nov 4;3:2976–2985. doi: 10.1021/cs4007777.
- Garrett DS, Seok YJ, Liao DI, Peterkofsky A, Gronenborn AM, Clore GM. Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. Biochemistry. 1997 Mar 4;36(9):2517-30. doi: 10.1021/bi962924y. PMID: 9054557.
- Jin L, Stec B, Lipscomb WN, Kantrowitz ER. Insights into the mechanisms of catalysis and heterotropic regulation of Escherichia coli aspartate transcarbamoylase based upon a structure of the enzyme complexed with the bisubstrate analogue N-phosphonacetyl-L-aspartate at 2.1 A. Proteins. 1999 Dec 1;37(4):729-42. PMID: 10651286.
- Kim YK, Shin YJ, Lee WH, Kim HY, Hwang KY. Structural and kinetic analysis of an MsrA-MsrB fusion protein from Streptococcus pneumoniae. Mol Microbiol. 2009 May;72(3):699-709. doi: 10.1111/j.1365-2958.2009.06680.x. Epub 2009 Apr 7. PMID: 19400786; PMCID: PMC2713860.
- Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. Proc Natl Acad Sci U S A. 2012 Jul 3;109(27):10873-8. doi: 10.1073/pnas.1203013109. Epub 2012 Jun 25. PMID: 22733734; PMCID: PMC3390869.
- Maaß T. Stable isotope labeling schemes for the assignment of NMR spectra of murine Norovirus P-domain proteins. Master's thesis, University of Lübeck, 2019.
- a: Maass T, Westermann LT, Creutznacher R, Mallagaray A, Dülfer J, Uetrecht C, Peters T. Assignment of Ala, Ile, Leu^{proS}, Met, and Val^{proS} methyl groups of the protruding domain of murine norovirus capsid protein VP1 using methyl-methyl NOEs, site directed mutagenesis, and pseudocontact shifts. Biomol NMR Assign. 2022 Apr;16(1):97-107. doi: 10.1007/s12104-022-10066-7. Epub 2022 Jan 20. PMID: 35050443; PMCID: PMC9068638.

- b: Maass T, Ssebyatika G, Brückner M, Breckwoldt L, Krey T, Mallagaray A, Peters T, Frank M, Creutznacher R. Binding of Glycans to the SARS CoV-2 Spike Protein, an Open Question: NMR Data on Binding Site Localization, Affinity, and Selectivity. Chemistry. 2022 Sep 26:e202202614. doi: 10.1002/chem.202202614. Epub ahead of print. PMID: 36161798; PMCID: PMC9537997.
- Mühlberg L, Alarcin T, Maass T, Creutznacher R, Küchler R, Mallagaray A. Ligand-induced structural transitions combined with paramagnetic ions facilitate unambiguous NMR assignments of methyl groups in large proteins. J Biomol NMR. 2022 Jun;76(3):59-74. doi: 10.1007/s10858-022-00394-0. Epub 2022 Apr 10. PMID: 35397749; PMCID: PMC9247001.
- Müller-Hermes C, Creutznacher R, Mallagaray A. Complete assignment of Ala, Ile, LeuProS, Met and ValProS methyl groups of the protruding domain from human norovirus GII.4 Saga. Biomol NMR Assign. 2020 Apr;14(1):123-130. doi: 10.1007/s12104-020-09932-z. Epub 2020 Jan 28. PMID: 31993958; PMCID: PMC7069894.
- Nelson CA, Wilen CB, Dai YN, Orchard RC, Kim AS, Stegeman RA, Hsieh LL, Smith TJ, Virgin HW, Fremont DH. Structural basis for murine norovirus engagement of bile acids and the CD300lf receptor. Proc Natl Acad Sci U S A. 2018 Sep 25;115(39):E9201-E9210. doi: 10.1073/pnas.1805797115. Epub 2018 Sep 7. PMID: 30194229; PMCID: PMC6166816.
- Pritišanac I, Degiacomi MT, Alderson TR, Carneiro MG, Ab E, Siegal G, Baldwin AJ. Automatic Assignment of Methyl-NMR Spectra of Supramolecular Machines Using Graph Theory. J Am Chem Soc. 2017 Jul 19;139(28):9523-9533. doi: 10.1021/jacs.6b11358. Epub 2017 Jul 10. PMID: 28691806.
- Singh BK, Leuthold MM, Hansman GS. Human noroviruses' fondness for histo-blood group antigens. J Virol. 2015 Feb;89(4):2024-40. doi: 10.1128/JVI.02968-14. Epub 2014 Nov 26. PMID: 25428879; PMCID: PMC4338890.
- Tugarinov V, Kay LE. 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. 2003 Nov 12;125(45):13868-78. doi: 10.1021/ja030345s. PMID: 14599227.
- Tugarinov V, Choy WY, Orekhov VY, Kay LE. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci U S A. 2005 Jan 18;102(3):622-7. doi: 10.1073/pnas.0407792102. Epub 2005 Jan 6. PMID: 15637152; PMCID: PMC545550.
- Tugarinov V, Sprangers R, Kay LE. Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. J Am Chem Soc. 2007 Feb 14;129(6):1743-50. doi: 10.1021/ja067827z. Epub 2007 Jan 24. PMID: 17249677.
- Velyvis A, Schachman HK, Kay LE. Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase. J Am Chem Soc. 2009 Nov 18;131(45):16534-43. doi: 10.1021/ja906978r. PMID: 19860411.
- Venditti V, Fawzi NL, Clore GM. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. J Biomol NMR. 2011 Nov;51(3):319-28. doi: 10.1007/s10858-011-9559-4. Epub 2011 Sep 4. PMID: 21935714; PMCID: PMC3212433.
- Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol. 1987 Apr 5;194(3):531-44. doi: 10.1016/0022-2836(87)90679-6. PMID: 3041007.