UNIVERSITÄT ZU LÜBECK

From the Institut für Molekulare Medizin
Center for Structural and Cell Biology in Medicine
of the University of Lübeck
Director: Prof. Dr. Georg Sczakiel

A Computational and Systems Biology Approach to RNA interference in Homo sapiens

Dissertation
for Fulfilment of Requirements
for the Doctoral Degree
of the University of Lübeck
from the Department of Natural Sciences

Submitted by

Simon Dornseifer
from Olpe

Lübeck 2021

First referee: Professor Dr. Georg Sczakiel

Second referee: Professor Dr. Hauke Busch

Date of oral examination: 06.09.2021

Approved for printing. Lübeck, 20.09.2021

Acknowlegements

First of all, I thank my supervisors Professor Dr. Georg Sczakiel and Professor Dr. Thomas Martinetz for engaging me in their exciting projects and for many insightful and inspiring discussions. Furthermore, I would like to specially thank the late Professor Dr. Tobias Restle as well as Dr. Jens Christian Clausen for their productive engagement. I also would like to thank everybody at the Institut für Molekulare Medizin for their support and their scientific collaboration. Last but not least, I would like to thank Professor Pedro Mendes, PhD, Dr. Jürgen Pahle and the whole Computational Systems Biology group at the University of Manchester for many fruitful discussions on my project and the thorough introduction to the computational modelling of biochemical pathways.

Publications and Copyrights

Parts of this thesis were published:

- Parts of Sections 2.1, 4.5 and 6.1 (including Figures 4.14, 4.18, 4.21, 4.22, 4.23 and 4.24; Tables 4.12 and 4.13) were first published as *RNAi revised - Target mRNA-dependent enhancement of gene silencing.* Dornseifer, Willkomm, Kretschmer-Kazemi Far, Liebschwager, Beltsiou, Frank, Laufer, Martinetz, Sczakiel, Claussen, and Restle. NAR 2015, 43:22[40].

- Sections 2.2, 3.7, 5 and 6.2 (including Figures 3.2, 5.1, 5.3, 5.4, 5.5 and 5.6; Tables 5.1, 5.2, 5.3, 5.4, and 5.5; and Appendix A1 and 2) were first published as *Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA*. Dornseifer and Sczakiel. BMC Bioinformatics 2013, 14:122[39].

The schematic representation of Ago2 protein in Figures 1.1, 4.1, 4.6, 4.14 are based on an illustration by Andrea Deerberg, Sarah Willkomm and Tobias Restle.

Funding

This work was supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/1].

Deutsche Zusammenfassung

RNA-Interferenz (kurz RNAi) ist ein posttranskriptioneller Mechanismus zum Stilllegen von Genen (sogenanntes Gene-Silencing). Die beiden Schlüsselkomponenten des Mechanismus, eine kurze RNA (Ribonukleinsäure) und ein Argonaute-Protein, bilden zusammen den minimalen RNA-induced silencing complex (RISC). Die kurze RNA erkennt spezifisch eine target-mRNA und ermöglicht Argonaute die Genexpression des targets durch Zerstörung oder Translationsblockade der target-mRNA zu hemmen. Die Entdeckung von RNAi in den 1990er Jahren führte zu ihrer Anwendung in der Grundlagenforschung zur Aufklärung von Funktionen von Genen bzw. ihrer Proteine, sowie zur Entwicklung neuartiger Therapeutika gegen bisher nicht zugängliche targets. Dennoch sind viele mechanistische Details des RNAi-Mechanismus beim Menschen noch unbekannt. Bevor sich RNAi-basierte Therapeutika in der Humanmedizin etablieren können, müssen diese Details, welche oft mit rein experimentellen Methoden schwer zugänglich sind, ausreichend entschlüsselt werden.
Ziel dieser Studie war es, durch Kombination von Systembiologie und Bioinformatik, mechanistische und kinetische Einblicke in den RNAi-Mechanismus im menschlichen Organismus zu gewinnen. Die Studie bestand aus zwei Teilen, die durch unterschiedliche Herangehensweisen unterschiedliche Aspekte des RNAi-Prozesses abdeckten:

Die computer-gestützte Modellierung von RNAi liefert starke Hinweise auf eine target-mRNA-abhängige Beschleunigung von Gene-Silencing durch einen assoziativen Strangaustauchmechanismus

In diesem Teil der Studie wurden *in sillico* Modelle der siRNA-basierten RNAi in humanen Zellen entwickelt. Ziel war es, *in vitro* gemessene kinetische pre-steady state Daten mit einem quantitativen, zeitlich aufgelösten Verlauf auf zellulärer Ebene zu kombinieren um mechanistische Rückschlüsse zur RNAi im menschlichen Organismus zu ziehen.
Die Modelle wurden unter Verwendung von kontinuierlich-deterministischen und diskret-stochastischen Simulationen analysiert. Zur Validierung der Simulationsdaten wurden experimentelle Daten herangezogen. Die Beobachtung, dass *in vitro* die Produktfreisetzung durch Argonaute 2 in Anwesenheit eines target-Überschusses beschleunigt wurde, inspi-

rierte die Annahme eines assoziativen Mechanismus der RNA-Slicer-Reaktion, bei der eingehende target-mRNAs die Dissoziation gespaltener mRNA-Fragmente beschleunigen.
Dieses neuartige assoziative Modell ist kompatibel mit den hohen multiple-turnover Raten des RNAi-basierten Gen-Silencing in lebenden Zellen und erklärt die target-mRNA konzentrationsabhängige Beschleunigung der RNAi-Maschinerie. Das assoziative Modell zeigt, dass die Wirksamkeit einer siRNA oder miRNA vom Expressionsniveau seiner target-RNA abhängt: eine hohe target-Konzentration ermöglicht eine bessere Regulierbarkeit des RNAi-Mechanismus. Bei der Auswahl von targets und der Abschätzung von Gene-Silencings Aktivität, sollten targets deshalb auf Systemlevel, z.B. durch Berücksichtigung des Expressionsniveaus, charakterisiert werden. Die Existenz eines solchen assoziativen Reaktionszyklus hat somit auch Auswirkungen auf das Design von siRNA-basierten Wirkstoffen sowie von nukleinsäure-basierten Diagnostika wie z.B. Antagomirs gegen miRNA-Tumormarker.
Infolgedessen wird für die target-Charakterisierung eine neue Generation von Bioinformatik-Tools benötigt um die Eigenschaften von RNAi und seiner Komponenten auf Systemebene zu betrachten. Ein Schritt in diese Richtung wird im zweiten Teil dieser Studie beschrieben.

Computer-gestützte Identifizierung von biologisch funktionalen non-hairpin GC-Helices in humaner Argonaute mRNA

Im zweiten Teil dieser Studie wurde die potentielle Bildung von RNA-Duplex Motiven aus entfernt liegenden, komplementären Sequenzabschnitten auf langen einsträngigen RNAs betrachtet.
Ein neuentwickeltes Software-Tool identifiziert RNA-Duplexelemente bei frei wählbarer Motivlänge und Nukleoidzusammensetzung zwischen entfernt liegenden Sequenzabschnitten. Die Software baut auf RNA-Strukturvorhersagealgorithmen wie z.B. Mfold und Sfold auf.
Die Anwendung des Tools konnte zeigen, dass ein geringer Teil der menschlichen Gene, einschließlich der Argonaute-like Genfamilie, mRNAs kodiert, welche vermehrt GC-reiche non-hairpin Duplexelemente von 8 oder mehr Nukleoiden Länge beinhalten. Diese neuartigen RNA-Motive waren evolutionär konserviert und wurden vorwiegend in der $5'$-Region gefunden, was auf eine potentielle biologische Funktion der non-hairpin GC-Helices hinweist. Parallel zu dieser Doktorarbeit wurde eine experimentelle Studie durchgeführt, welche diese These stützt: es wurde eine Auswirkung der non-hairpin GC-Helices auf die posttranskriptionelle Regulation von Genexpression durch Einsatz eines Fusionstranskripts,

bestehend aus der 5′-Sequenze von humaner Ago2 mRNA und der kodierenden Sequenz von Renilla-Luciferase, nachgewiesen.

Abstract

RNA interference (RNAi) is a post-transcriptional gene silencing mechanism. Its key components, a small RNA and an Argonaute protein, form the minimal RNA-induced-silencing-complex (RISC). The small RNA's specific recognition of target mRNA allows Argonaute to inhibit gene expression via destruction or translational blockade of the target mRNA.
The discovery of RNAi in the 1990s gave rise to the development of new nucleic acid-based investigational technologies and potential therapeutics against previously unaccessible targets. However, mechanistic key details of RNAi in humans need to be deciphered, before such approaches take root in biomedicine and molecular therapy. The complex biological mechanism is often difficult to access experimentally.
The objective of this study was to gain mechanistic and kinetic insight into RNAi in the human system by combination of systems biology and bioinformatics. The doctoral study consisted of two parts, which covered different aspects of the biological process:

Computational modelling of RNAi supports target mRNA-dependent enhancement of gene silencing via associative strand-exchange

In this part of the study, *in silico*-based models of siRNA-mediated RNAi in human cells were developed in order to link *in vitro*-derived pre-steady state kinetic data with a quantitative and time-resolved understanding of RNAi on the cellular level.
The models were analysed using continuous-deterministic and discrete-stochastic simulation methods. Simulation results were validated by comparison to experimental findings.
The observation that product release by Argonaute 2 was accelerated in the presence of an excess of target RNA *in vitro* inspired the suggestion of an associative mechanism for the RNA slicer reaction where incoming target mRNAs actively promote dissociation of cleaved mRNA fragments. This novel associative model was compatible with high multiple turnover rates of RNAi-based gene silencing in living cells and accounted for target mRNA concentration-dependent enhancement of the RNAi machinery. The associative model indicated that the efficacy of an siRNA or miRNA depends on the expression level of its target

RNA in a way that high target levels allow better regulation via the RNAi mechanism. This suggests considering the target as a whole when choosing appropriate targets or when predicting gene silencing efficacy. The existence of such an associative reaction cycle in RNAi will have an impact on the design of nucleic acid-based diagnostics, such as antagomirs against miRNA tumour markers, as well as, siRNA-based therapeutics.

As a consequence, a new generation of RNAi bioinformatics tools is needed for the characterisation of system-level properties of RNAi and its components. A step in this direction was taken in the second part of this study.

Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA

In the second part of the study the potential formation of RNA duplex motifs by long-range RNA-RNA interactions of distantly located matching sequence elements of a single long-chain RNA was investigated.
A newly-developed software tool identifies consecutive RNA duplex elements at any given length and nucleotide content formed by distant sequences. The software tool takes into account local minimal energy states. It is built on top of RNA structure prediction algorithms like Mfold and Sfold.
It was found that a small ratio of human genes including the Argonaute (Ago)-like gene family encode mRNAs containing highly GC-rich non-hairpin duplex elements (GC-helices) of equal to or more than 8 base pairs in length. These newly identified RNA motifs were observed preferentially within the $5'$-region of mRNAs in an evolutionarily conserved fashion indicating their potential biological role. Parallel to this thesis, this view was supported experimentally by post-transcriptional regulation of gene expression of a fusion transcript containing $5'$-sequences of human mRNA^{Ago2} harbouring GC-helices and down-stream coding sequences of Renilla luciferase.

Contents

List of Figures

List of Tables

Chapter 1

Background: from biological process to cybernetic model

RNAi in short

RNA interference (RNAi), also known as gene silencing, is a common biological process in which small RNAs, aided by Argonaute protein, inhibit gene expression via destruction or translational blockade of mRNA, the so-called target. Recognition of target is highly specific due to base pairing between the RNAs. Compare[46, 61].
A first descriptive model of the siRNA-mediated RNAi process is provided in Fig. 1.1. Despite its simplicity, it already forms a realistic basis for the more complex formal models used for *in silico* simulation in this study.
Short double-stranded RNA fragments, the so-called short interfering RNA (siRNA) are loaded to Argonaute 2 (Ago2). Next, the siRNA's passenger strand is removed from the nucleoprotein complex, either by cleavage or by wrenching of the double-strand. The siRNA's guide strand is able to recognise and binds its target RNA via base pairing. The target RNA is cleaved via Ago2's endonuclease activity and cleavage products are released from the complex. From here, the remaining complex can process the next target RNA.

More information on RNAi, its late discovery, its evolution and its vital biological function follow in the next Section 1.1. The section will be completed with remarks on the current and the future medical and biotechnological utilisation of the RNAi process.
In Section 1.2 the mechanistic details, which are essential for the successful and safe medical usage of RNAi, will be brought together from biological studies.
Furthermore, without a proper understanding of the interplay between the mechanistic de-

tails and their quantitative parameters, any application of RNAi (e.g. in medicine or biotechnology) will be governed largely by trial and error. Therefore, experimental RNAi research, nowadays, is complemented by computational systems biology, which will be introduced in Section 1.3.

1.1 Introduction to RNAi

RNA interference (RNAi), also known as gene silencing, is a biological process in which small RNAs inhibit gene expression via destruction or translational blockade of mRNA, the so-called target. Recognition of targets is highly specific due to base pairing between the RNAs. Key components of the RNAi machinery, i.e. the RNA-binding Argonaute protein family are shared among eukaryotes, most archaea, and some bacteria.
In different organisms, RNAi has adapted different biological functions, such as gene regulation, antiviral immunity and preservation of genome integrity. In mammalians, it is an essential post-transcriptional gene silencing mechanism. As such it regulates at least 30% of human genes[87, 127], and thus is involved in controlling global processes like development and basic cellular processes: from cell growth via heterochromatin formation through to tissue differentiation. Its dysfunction is linked to cancer, cardiovascular disease and other degenerative disorders[121].
The discovery of RNA interference (RNAi) in the 1990s revolutionised the understanding of gene regulation. Since then, there has been great interest in RNAi, not least because of its potential in biomedicine and molecular therapy. Novel classes of RNAi-linked oligonucleotides, e.g. small interfering RNA (siRNA), are becoming powerful investigational tools and potential therapeutics that allow to selectively turn off any gene, even those previously unaccessible by conventional drugs[225].
Although overall RNAi is well understood, after nearly three decades of intensive research it has become clear that RNAi is a complex mechanism embedded in a complex cellular system. Its multitudinous, partly yet unknown, molecular details are often experimentally difficult to access or are even considered inaccessible, compare[44, 46, 61, 84, 129, 194]. Thus, computational support from bioinformatics in the field of RNAi has become common, e.g. for automated analysis of large-scale RNAi screen data[102] and for RNAi efficacy prediction[191, 209]. With the availability of powerful computer technology, cybernetic modelling of complex biological systems has emerged, and for the first time a systemic view of the RNAi system is possible. Pioneering, as well as, state of the art modelling approaches of RNAi are reviewed in Section 1.3.

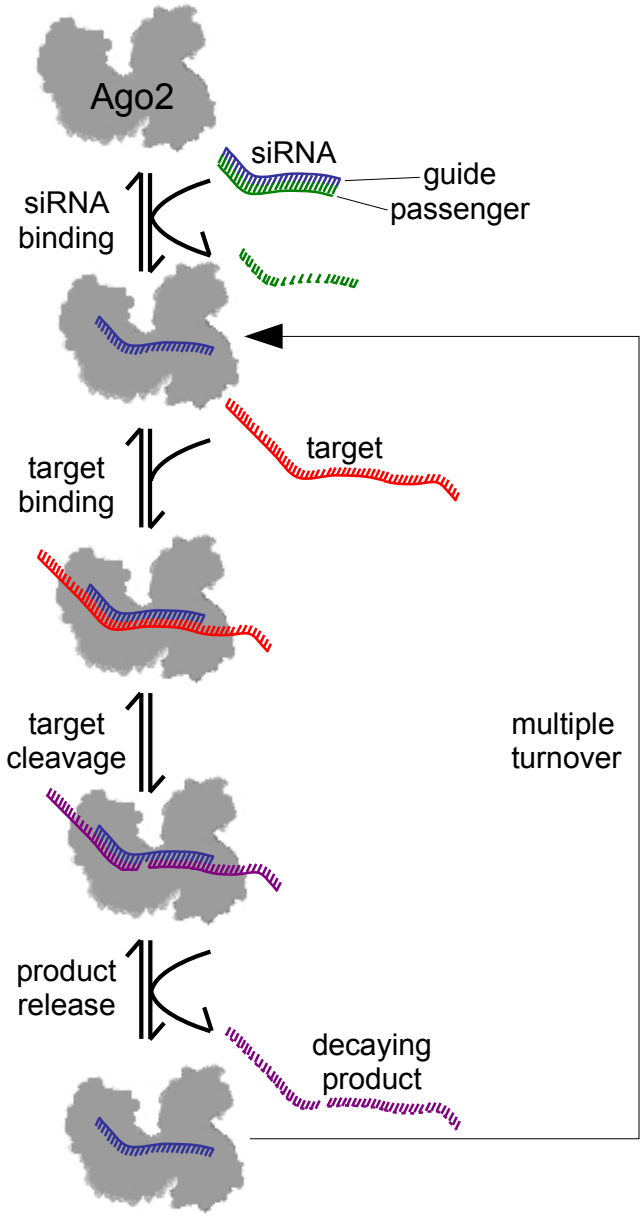


Figure 1.1 **Illustration of siRNA-mediated RNAi in mammalian cells.** Short double-stranded RNA fragments, the so-called short interfering RNA (siRNA) are loaded to Argonaute 2 (Ago2). Next, the siRNA's passenger strand is removed from the nucleoprotein complex, either by cleavage or by wrenching of the double-strand. The siRNA's guide strand is able to recognise and bind its target RNA via base pairing. The target RNA is cleaved via Ago2's endonuclease activity and cleavage products are released from the complex. From here, the remaining complex can process the next target RNA.

In this remaining chapter, the RNAi process is being further characterised and all relevant facts on RNAi are gleaned from historical milestones and current research.

1.1.1 Short history of RNAi discovery

In 1990, RNAi was accidentally discovered by Jorgensen et al. Their original intension was to express extra gene copies in petunias, aiming to increase the quantity of purple pigments and thus creating petunias of a deeper colour. Instead, the resulting transgenic flowers were white, lacking the desired purple pigments. That is, the very gene that should be over-expressed, instead got turned off by the extra copies. Jorgensen et al. named this effect co-suppression[152]. Similar seemingly paradox results were observed in other plants, as well as, in fungi. For instance, white transgenic colonies of Neurospora sitophila were observed in place of orange ones when adding extra copies of a pigment gene [180]. Furthermore, plants expressing small fragments of non-coding viral RNA, developed resistance to the virus from which the genetic fragments originated. Taken together, these observations revealed the existence of an unknown, potent mechanism of gene regulation [177].
Years later, in 1998, Mello and Fire observed a potent gene silencing effect after injecting ds RNA into Caenorhabditis elegans[50]. They were the first to identify the effector triggering the gene silencing phenomenon: neither mRNA nor as (antisense) RNA injections had an effect on target gene expression, but ds RNA successfully silenced the targeted gene. They named the mechanism RNA interference or, short, RNAi. For their work on RNAi, Fire and Mello were awarded the Nobel Prize in Physiology or Medicine in 2006.
It soon became evident that the mechanisms previously described as co-suppression, post-transcriptional gene silencing (PTGS), and quelling had to be related or even identical to RNAi: the introduced transgenes produced as transcripts, which in turns resulted in the formation of short ds RNA effectors for RNAi [210].
From these pioneering observations, researchers around the world began characterising the process at genetic and biochemical levels. Soon it was wondered whether the mechanism could be tamed for diverse uses, from basic science via biotechnology and agriculture through to human medicine[162].
In 1999 short RNAs were identified in plants [68]. In the same year RNAi was shown *in vitro* [210]. In the following, the puzzle pieces of 25 years of research on RNAi are brought together in a nutshell, to give a holistic picture of the process.

1.1.2 Evolution of RNAi

RNAi is widespread in plants and animals [195] with Argonaute proteins being identified as the key macro-molecules involved in the mechanism. Generally, Argonautes are found in all domains of life. While their function in bacteria and archaea is still not fully understood [195], their eukaryotic counterparts are well-characterised. In eukaryotes, Argonautes evolved into two clades with distinct functions. Members of the AGO clade mediate cytosolic gene silencing via RNAi via interaction with small effector RNAs, namely, siRNA and micros RNA (miRNA or miR). Whereas proteins of the PIWI clade interact with PIWI-interacting RNAs (piRNAs) to manage mobile genetic elements of the germ line [64]. Humans have four members of the AGO clade: Ago1, Ago2, Ago3 and Ago4. Of these, only Ago2 exhibits slicer activity, which enables the endonucleolytic cleavage of bound target RNA via siRNA-mediated RNAi [86].

Parsimony-based phylogenetic analysis, in combination with biochemical and genetic studies, have led to the identification of other conserved components of the RNAi machinery, i.e. Dicer and RdRP[195]. These studies further suggest that the last common ancestor of modern eukaryotes possessed RNAi-based mechanisms. The protein machinery of eukaryotic RNAi seems to have been combined from ancestral archaeal, bacterial and phage proteins involved in DNA repair and RNA processing[195].

The ancestral role of RNAi components may have been non-essential for unicellular life. However, they probably formed important defence responses against genomic parasites such as transposable elements and viruses[24]. Related functions such as histone modification may have already been present in the common ancestor of modern eukaryotes[16]. From a mechanistic perspective, the RNAi machinery in the eukaryotic ancestor may have been capable of both small-RNA-guided transcript degradation and translational repression [16, 24]. Both roles appear to be widespread among living eukaryotes and the diversification of function could account for the evolutionary conservation of duplicated AGO and PIWI proteins[16, 24].

The requirement for complex gene regulation to support multicellular organisms may have independently driven the elaboration of RNAi pathways in animals and plants as suggested by the independent origins of miRNA genes in these two kingdoms[4, 111, 149]. This suggests another explanation for the duplication of Argonaute and Dicer proteins in most multicellular organisms[147]. Conversely, the RNAi machinery has been entirely lost or extensively simplified in a number of unicellular eukaryotes with small genomes[24] and in many fungi[151].

RNA interference genes, as components of the antiviral innate immune system, are involved

in an evolutionary arms race with viral genes. For instance, many plant viruses have evolved mechanisms for suppressing the RNAi response in their host cells[123]. Furthermore, studies of evolutionary rates in Drosophila have shown that genes in the RNAi pathway are subject to strong directional selection and are among the fastest-evolving genes in the Drosophila genome[156].
The complex and not yet fully understood evolutionary processes of RNAi components point towards the mechanistic and functional complexity of modern day RNAi systems.

1.1.3 The functional role of RNAi

RNAi in Immunity

RNAi is an integral component of the immune response to viruses and other foreign genetic material, especially in plants[205]. RNAi in plants has a unique characteristic: gene silencing systemically spreads throughout the organism. It even can be transferred via grafting[160]. This systemic phenomenon allows the entire plant to respond quickly after an initial local virus infection[215].It may furthermore prevent the self-propagation of transposons[205]. Some plants also express endogenous siRNAs in response to infection by specific types of bacteria[91]. These effects may be part of a generalised response to pathogens[52]. In some animals, RNAi has also been shown to produce an immune response. For instance, in Drosophila melanogaster RNAi is known to be a part of the immune response to viruses[218, 231]; and in Caenorhabditis elegans RNAi is thought to have a similar function[122, 223]. Although the role of RNAi for antiviral immunity in vertebrates is unclear[31], recent studies suggest the existence of a functional antiviral RNAi pathway in mammalian cells[114, 131].

RNAi in gene regulation

There is no doubt about the importance of RNAi in gene regulation in mammalians and other organisms. Endogenously expressed miRNAs, whether of intronic or inter-genic origin, are important in translational repression[188]. In Homo sapiens, 5% of the genome is dedicated to encoding and producing more than 1,000 miRNAs that, in turn, regulate at least 30% of human genes[87, 127]. Thus, RNAi adds a whole new qualitative layer to the regulation of cellular networks.
RNAi controls vital processes such as cell growth, tissue differentiation, heterochromatin

formation, and cell proliferation. Furthermore, it regulates development, especially the timing of morphogenesis and the maintenance of undifferentiated or incompletely differentiated cell types such as stem cells[21].
Accordingly, RNAi dysfunction is linked to cardiovascular disease, neurological disorders, and multiple types of cancer[121]. In many organisms, including humans, miRNAs have been linked to the formation of tumours and dysregulation of the cell cycle. Here, miRNAs can function as both oncogenes and tumour suppressors[235].

1.1.4 Technological and medical applications of RNAi

Right after its discovery RNAi's potential medical applications sparked great interest among researchers. The silencing potential of the RNAi pathway can be selectively directed towards genes. Essentially, any target in the transcriptome, including previously inaccessible ones, can be silenced using small RNAs (i.e. siRNAs and miRNAs) with sequences complementary to the target transcript. The gene regulation via RNAi is fast compared to transcription factors whose expression additionally requires protein synthesis[197]. Thus, small synthetic RNAs have become indispensable investigational research tool. Furthermore, they possess a great potential of applications in medicine as therapies and drugs[225].

RNAi research and development

RNAi has become a valuable and widely used investigational research tool in living organisms, cell culture, as well as high throughput arrays. RNAi arrays and large, often genome-wide, libraries of small RNAs have revolutionised reverse genetics screening[7, 125, 198]. They make it possible to scan whole genomes for identification of genes (and their products) that have an effect on a specified phenotype. *In vivo*, short hairpin RNAs (shRNAs) are powerful tools for investigating gene function[76, 169].
Furthermore, RNAi has been used experimentally for applications in biotechnology, particularly in the engineering of food plants that produce lower levels of natural plant toxins or in inducing resistance to common plant viruses[230]. Such techniques take advantage of the systemic and heritable nature of the RNAi effect in plants[108, 200, 207].

RNAi in molecular medicine

The application of the RNAi pathways in molecular medicine are multifaceted. In general, both single-gene and genome-wide RNAi arrays using siRNAs have great potential for the discovery of new therapeutic targets for a variety of diseases. Mutations in miRNAs or

in their binding sites have been linked to cancer, cardiovascular disease and other degenerative disorders[121]. This presents an opportunity to target specific miRNAs for therapeutic benefit by the use of so-called antagomirs or blockmirs, i.e. oligonucleotides used as miRNA inhibitors[43]. The presence of specific miRNAs in body fluids are linked to several types of malignant tumours. Thus, circulating miRNAs have recently become promising biomarkers for non-invasive diagnosis in various types of cancer[14, 69, 227]. Application of synthetic siRNAs for silencing the expression of disease-causing genes is seen as a very promising new class of macromolecular drugs. Basically, siRNAs can silence any target in the transcriptome including previously inaccessible ("undruggable") targets such as intra-nuclear genes but also many proteins[225]. Furthermore, siRNAs are promising candidates for personalised medicine through their high target specificity and their rapid lead identification[25]. At the moment, siRNA drugs against a range of diseases are in clinical trial phase, from lung infections via liver cancer to age-related macular degeneration[27].

De facto, before RNAi-based therapies and drugs will be able to fully reach the market there are some challenges to overcome. Firstly, mechanistic details of the complex mammalian RNAi system have to be deciphered and quantified in order to safely apply the cellular mechanism for human medical applications. One important aspect is the needs for optimising efficacy (e.g. target knockdown and siRNA delivery), while avoiding possibly dangerous side-effects like off-target effects. Although overall RNAi is well understood, it remains a complex mechanism embedded in a complex cellular system with many yet unknown molecular details (compare [44, 46, 61, 84, 129, 194]).

1.2 Cellular mechanism of RNAi in the human system

First, the lowest common factors between the three RNAi pathways (i.e. siRNA-, miRNA-, and piRNA-mediated RNAi) in the human system are described. From there, the complexity of the human siRNA-mediated RNAi system is unfolded.
All three pathways share a common mode of action: The minimal effector is a ribonucleoprotein complex comprising an Argonaute-family protein and a small effector RNA that grants specificity via base-pairing interactions with its target mRNA. In the miRNA and siRNA pathways, this is known as the RNA-induced silencing complex (RISC) and it drives gene silencing via degradation or translational repression of the bound target mRNA[225].

1.2.1 Small effector RNA

The origins and functions of the three classes of effector RNA, miRNA (mircoRNA), siRNA (short interfering RNA), and piRNA (Piwi-interacting RNA), differ:

miRNAs originate from non-coding genomic regions. Before reaching maturity, an at least 1,000 nucleotides (nt) long primary transcript, a so-called pri-miRNA undergoes extensive post-transcriptional modification[183]. In the cell nucleus it is processed to a 65 to 70 nt stem-loop structure (pre-miRNA) by the microprocessor complex and transported to the cytoplasm[124]. Finally, it is cut into an imperfect duplex of about 22 nt length, i.e. the mature miRNA molecule that can integrate into the RISC complex by Dicer, a large endoribonuclease containing a helicase domain and a pair of internally dimerised RNase III domains[124].

In mammalians, some siRNAs may be endogenous but they mostly originate from exogenous, synthetic or viral sources[22]. In the case of endogenous siRNA, longer precursor double-stranded RNA activates the ribonuclease protein Dicer[12], which binds and cleaves the RNA to produce perfectly paired duplexes of 20 to 25 nt with phosphorylated $5'$-ends and hydroxylated $3'$-ends with two overhanging nucleotides (compare[23, 199, 213, 232]). At transfection of exogenous siRNA, the effect is only transient due to dilution of the siRNA concentration, especially in rapidly dividing cells[9]. For a permanent effect, genes for short hairpin RNA are introduced via transfection vectors, which can be transcribed and processed into functional siRNA by Dicer[208]. The introduction of longer exogenous double-stranded RNA or larger quantities of siRNA triggers innate immune response that cause undesired non-specific effects in mammalians cells[178, 221].

The key differences between miRNA and siRNA is the fact that miRNA double-strands contain mismatches, whereas siRNAs form perfect base-pairing. Furthermore, miRNAs contain more extended terminal loops compared to siRNAs. Functionally, miRNAs form incomplete base pairing to their targets. Thus, the same miRNA may block the translation of several different mRNAs with similar binding sites. In contrast, siRNA form perfect base-pairing to one specific target mRNA. The major mechanistic difference between the two pathways is the silencing of the target mRNA via degradation in the case of siRNA and translational repression in the case of miRNA[225].

In contrast, the key molecular players involved in the germ line-specific piRNA pathway

to large extent are still unidentified. Emerging structures of an involved endoribonuclease called Zucchini provide a first starting point for studying the piRNA-mediated pathway and its molecular mechanisms[82, 155, 167].

During RNAi the previously described small effector RNAs form complexes with proteins that will be described in the following.

1.2.2 RLC and pre-RISC

The three proteins Dicer, Argonaute, and dsRBP form a minimal RISC-loading complex (RLC). w RLC generates a small RNA by dicing its precursor RNA. The product, i.e. a double-stranded miRNA or siRNA is loaded to Argonaute, forming the transient pre-RISC[128]. The so-called guide strand of the duplex is bound to Argonaute: its $3'$-end and $5'$-phosphate are bound respectively to the PAZ and MID domains. The other strand, the so-called passenger strand is cleaved and discarded. This process is described in detail by Wilson and Doudna[225].

1.2.3 RISC

The RNA-induced silencing complex (RISC) consists of multiple proteins and RNA. Its most important components are an Argonaute protein and an siRNA or miRNA guide strand. RISC binds a single-stranded target RNA, typically an mRNA, with sequence complementarity to the Argonaute-bound guide strand. In the case of an siRNA guide strand, perfect complementarity between guide and target takes place. This siRNA-mediated pathway leads to the cleavage of the target RNA by Argonaute's catalytic activity. Whereas, in the case of the miRNA-mediated pathway, target-guide binding is not necessarily perfect complementary. This pathway leads to translation repression and/or deadenylation and decay of the target mRNA [18]. Even though additional proteins like e.g. GW182 are involved in the RNAi processes, the exact role of co-factors in the complex RNAi mechanism are not fully understood[44, 47]. RISC is described in detail by Wilson and Doudna[225].

1.2.4 Argonaute 2 protein

Argonaute is the key player of all RNAi pathways. It is responsible for guide strand recognition, target cleavage (e.g. in the case of hAgo2), and recruitment of further RNAi-related proteins[225]. Bound to a RNA guide strand, human Argonaute 2 forms the minimal func-

tional RISC capable of target cleavage activity[32]. Roughly speaking, Argonaute proteins form a dichotomous structure. Each lobe consists of two domains: the N-terminal and PAZ domains or the MID and PIWI domains, respectively[42, 189]. Functional information on the individual protein domains have been deducted from crystal structures of eubacterial, archaebacterial[229] Ago proteins and more importantly, hAgo2[42, 189]. Based on these initial details, biochemical studies were performed that enabled to relate steady-state and pre-steady-state kinetic data to conformational changes taking place during interaction of hAgo2 and guide stand (binary complex formation); interaction of hAgo2-guide and target RNA (ternary complex formation); followed by target cleavage and the release of product (i.e. the cleaved target)[32].

To sum up Deerberg, Willkomm and Restle[32]:

> Assemblies of both the binary and the ternary complex comprise at least three kinetically distinct binding steps, with the respective first steps representing a diffusion-limited collision complex formation. According to our current conception, the second and third step ... of binary complex formation represent anchoring of the guide strand 5′-end into the Mid binding pocket and binding of the guide strand 3′-end to the PAZ domain, respectively.
> ...
> The second and third step ... of ternary complex formation represent seed pairing within the 5′-region of the guide followed by the release of the guide strand 3′-end from the PAZ domain, which in turn is a prerequisite for extended hybridization of the two RNA molecules and eventually leads to a catalytically competent complex.

Their study provides direct evidence for the hypothesis, first suggested suggested by Haley & Zamore[67] and Rivas et al.[179], that product release is the rate limiting step during RNAi. This is, however, contradicting to Wee et al.[220].
The precise pre-steady state rate constants identified by Deerberg et al.[32] are an important break-through for model building and are used to parametrise the computational models during this study. They enable the set-up of detailed computational models and realistic simulations.

1.3 Computational systems biology of RNAi

Generally, biochemical pathways are intrinsically complex, not only because they encompass a large number of interacting components, but also because those interactions often are nonlinear [202]. Like many other nonlinear phenomena in nature, their behaviour can be unintuitive [97]. Thus, to describe and understand their mechanism and function, quantitative models can be useful [143]. The known details of biochemical pathways, e.g. the RNAi pathways, have arisen from the reductionist process of biochemistry with focus on studying isolated reactions [97]. The behaviour of the whole system, or more realistically, the impact of the interactions of its parts are the subject of systems biology[97].
Without a proper understanding of the interplay between the kinetics of the process and its parameters, any application of RNAi in medicine, biotechnology or else, will be governed largely by trial and error [13]. Deductively, the ability to specifically tailor and optimise the treatment for each particular system would save significant time and resources, especially given the high cost of synthetic siRNA molecules and the amount of material required for *in vivo* studies. Therefore, experimental RNAi research is complemented by computational systems biology.

1.3.1 An introduction to systems biology

Systems biology is an inter-disciplinary branch of life sciences that extends classical biology with methods from mathematics, physics and computer science [97]. The aim of systems biology is a quantitative understanding and prediction of the dynamic interactions between parts of a biological system by regarding the system as a whole [97]. The core of this holistic approach is the cybernetics modelling and computer simulation of complex metabolic, signalling, and genetic pathways. However, modelling and simulations are combined with laboratory experiments and literature research in an iterative process [97].

Development of the systems biology movement

Today, the practical use of modelling and computer simulation ranges from weather forecasting [126] and environmental models [45] via the design of industrial and logistic processes [107] through to social [130] and financial models [138]. Likewise, applying mathematical modelling to biology has a long but sparse history: In the 1930s the famous Lotka-

Volterra equations, a pair of nonlinear differential equations, were used in ecology to describe the dynamics of biological systems in which two species interact, one as a predator and the other as prey [119, 120, 216]. In 1952 one of the first numerical simulations on basis of differential equations used in cell biology was published by Nobel prize winners Alan Lloyd Hodgkin and Andrew Fielding Huxley: a model explaining the action potential propagating along axons of neuronal cells [78]. In 1960, Denis Noble presented a first computer model of the heart, which allowed to test new drugs, cardiac pacemakers and defibrillators *in sillico* [154]. In the same decade the name systems biology was coined for the emerging new discipline of modelling and computer simulation in biology [182]. In 1970 Jacques Monod started cybernetics, i.e. the the quantitative modelling of feedback-regulated mechanisms of enzyme kinetics on a molecular level [143]. At the same time, the theoretical basis for systems biology was provided by Ludwig von Bertalanffy with his general systems theory [117].
However, only recently has there been an increased interest in the application of modelling and simulation to biology. Particularly from year 2000 onwards, systems biology gained popularity as a movement in its own due to several parallel developments:

- the generation of big data in biology, including genomic, transcriptomic, metabolic and proteomic data, which are difficult to understand without the use of analytical tools [136],
- the widespread use of the Internet as means of shared use of big data in biology on an international scale [176],
- an increase in computing power which enables calculations and simulations to be performed that were not previously possible [136], and
- an increasing interest in *in silico* experiments due to ethical and financial constraints and other complications involved in human and animal experiments [157].

Understand and predict biological complexity

Biological processes often are complex processes that have developed during evolution and contain various level of organisation, ranging from populations and their individual organisms via tissues, organs and cells through to compartments, macro-molecules and atoms [97].
In this regard, relevant length and time scales can range over multiple magnitudes. On a molecular level, protein-RNA binding occurs in microseconds, transcription of a nucleotide

and translation of an amino acid takes milliseconds [192]. On a cellular level, metabolic processes and signalling may take minutes and the scope of cell cycles lies within hours [192]. Human life is measured in years and evolution occurs over billion of years [192]. Similarly length scales vary from nanometres for molecular bonds [115], 1 to 100 nm for macro-molecules, via cell sizes within a range of 0.5 to 500 μm, through to body sizes of several metres and even to global ecosystems on the scale of kilometres [97].
Taking this complexity into account, it becomes obvious that it is not always possible to explain a biological systems by first principle or predict its behaviour from intuition. It follows, that it is often difficult to foresee the global behaviour of a system from knowledge of its parts. Mathematical modelling and computational simulation can help to understand and predict behaviour and properties of complex systems in a systemic context [97]. Generally, there are several advantages, the use of modelling in biological studies provides [97]:

- conceptual clarification by formalisation of verbal hypotheses
- spot gaps in knowledge and understanding during model formulation
- computational simulation is cheap compared to experiments
- reduce ethical problems associated with animal experiments
- examine scenarios not accessible by experiment, e.g. follow dynamics of experimentally untraceable intermediates, impose conditions not feasible in experiment.

Formal mathematical models

A model is an abstract representation of real world objects and processes and explains certain features of the modelled system. It represents only specific aspects of reality, since its purpose is to answer particular questions, all other aspects of the real world system are neglected or simplified. Instead of striving to become a one to one copy, a model aspires to explain a system and to study the effects of different components, and to make predictions about behaviour [150]. This was summarised in the famous quote of British statistician George Box: “Essentially, all models are wrong, but some are useful”.
The kind of models dealt with in this study describe a biological process using mathematical concepts and formal language. The models consist of relationships (e.g. reaction steps) and variables. The relationships can be described by mathematical operators, such as differential equations. The variables are abstractions of system parameters of interest such as molecular concentrations of reactants, reaction compartment volumes and reaction rate constants. Compare [97, 150].

Model identification

Models can be classified into black-box and white-box approaches depending on whether a priori knowledge on the system of interest is available. A black-box model is used for system for which no a priori knowledge exists and thus, both, the functional form of relationships between variables and the quantitative parameters in those functions must be estimated during modelling. A white-box model, on the other hand, describes a systems for which all required information is available. It is preferable to use as much a priori information as possible to make the model more accurate [222]. De facto, a limited amount of information on the modelled system is available and a modelling approach somewhere between black-box and white-box model is chosen. Compare [97, 150].

Model complexity

Model complexity is a trade-off between simplicity and predictive power. Freely adapted from Occam's razor (or the law of parsimony): given several models with equal predictive power, the simplest model is the most desirable. While added complexity usually improves the predictive power of a model, it can make the model difficult to understand and may lead to an overfitting of the free model parameters to given data. Compare [97, 150].

Computer simulation

A computer simulation refers to the actual use of the model, usually by running computer programs that numerically solve mathematical equations defined by the model to emulate a real-life system. Thus, simulation refers to the process of producing data from a model given a set of initial conditions and can be seen as as an *in sillico* experiment. Simulations may produce a variety of output: data may be produced by stochastic or deterministic simulation methods; they may refer to the system's steady-state or be dynamic (i.e. time course data); they may be continuous or discrete. Compare [97, 150].

1.3.2 A review of modelling of RNAi

Bioinformatics approaches to support siRNA and miRNA efficacy prediction are plenty and, nowadays, well-established (i.e. [102, 191]). For a recent review, please refer to Tafer[209]. Compared with conventional bioinformatics, cybernetic modelling of the RNAi process is still in its infancy. The first pioneering approach using simple kinetic equations for reaction steps in the gene silencing process started in 2003[11], that is 5 years after small RNA was

identified as the effector of RNAi by Mello and Fire[50].
The publication frequency of systems biology approaches to RNAi has been low ever since, even though a holistic understanding of the process is strongly demanded not least to guarantee a safe and efficient application in human medicine. To understand this apparent contradiction one has to review and analyse the studies of modelling of RNAi. Many published models of RNAi were situated in organisms where RNAi differs mechanistically and functionally from that of mammalian systems, i.e. in none-coelomates. For instance, Bergstrom et al.[11] and Groenenboom et al.[60] targeted the amplification process during RNAi-based immune response in invertebrates. Often, systems biology approaches to RNAi in mammalian systems had very specialised and thus limited objectives. For instance, [133] thematised recursive RNAi (i.e. using RNAi for down-regulation RNAi-associated compounds). Most models were dedicated to the miRNA pathway, for example [57, 172, 214]. Even though siRNA- and miRNA-pathways are closely related and share a common concept, as well as most key components, there are important differences in the origin of the small effector RNAs, the binding modes to target mRNA and the mechanism of regulation.
Studies of modelling of RNAi in mammalian systems are summarised and commented on in the following. They are listed in chronological order.

In 2004 Arciero and co-workers [2] created a mathematical model to investigate tumour immune evasion and siRNA treatment. The RNAi part of the model was reduced to one parameter (i.e. siRNA) without any mentioning of RISC components. No mechanistic details of RNAi were touched in their study. The model and its parameters were not based on any experimental data. However, they provided first theoretical insights into how siRNA could be used in cancer treatment, i.e. returning an aggressive tumour to its passive state.

Raab and Stephanopoulos [172] (2004) investigated the shRNA-mediated dynamics of gene silencing. The focus of their study was on transcription, transport and processing of shRNA. RNAi itself was only touched in one of four equations, i.e. ordinary differential equations (ODEs). Model parameters including a condensed reaction rate constant for RNAi activity, were fitted to experimental time course data. With their study they were able to analyse how the efficiency of shRNA-plasmid transfection effects silencing potential. The RNAi process itself took a back seat in this study.

In 2006 Bartlett and Davis [8–10] were the first to connect high-quality kinetic data with a fairly detailed cybernetic model of siRNA delivery and siRNA-mediated RNAi in cell

culture. Furthermore, they extended their model to RNAi in living mice. Their model consisted of 7 ODEs for siRNA delivery and 5 equations for the RNAi mechanism with 29 parameters in total. The values for most parameters were harvested from literature. Where some of the reaction rate constants lay outside physiological feasible ranges. None-the-less, they determined the siRNA dosing schedule with regard to delivery *in vivo*, which could be of value for the prediction of the application of siRNA drugs. Their main finding for the cell culture model was that dilution due to cell division (instead of not intracellular siRNA half-life) governed the duration of gene silencing. These findings must be treated with some reservation given the obviously faulty reaction rate constants of siRNA degradation and of the siRNA cleavage reaction.

Larsson, Sander and Marks [104] (2010) shed light on the limiting roll of target mRNA turnover rates on RNAi and arguably were the first to touch upon the influence of target abundance on RNAi efficacy. Even though their approach was fairly simple, describing RNAi in a single equation with four model parameters, they were the firsts to suggest that mRNAs with high turnover rates could be more resistant to RNAi-based silencing. They explained this observation with the hypothesis that mRNA turnover rate limits siRNA and miRNA efficacy. Their model parameters were fitted during validation with IC_{50} data, but since the simple model was purely phenomenological, underlying biochemical reaction steps remain unknown and no mechanistic details could be inferred from this model.

Yet another study of miRNA-mediated RNAi was published by Vohradsky, Panek and Vomastek [214] in 2010. Their ODE model contained two linear, stoichiometric equations including saturative Michaelis-Menten kinetics for RNAi reactions with 6 parameters. These were fitted to 42 time course experiments. The study suggested a sigmoidal function of target mRNA time course with an initial period where the reaction to the miRNA input signal is delayed. This model was compared to a newly suggested mechanism: a “digital” U-shaped time course where the influence of the miRNA was stopped at a certain moment for target mRNA recovery.

Cuccato et al. [30] (2011) compared four different mathematical models of siRNA-mediated RNAi in mammalian cells and fitted their simulation output to their own experimental *in vitro* data. The compared models were relatively simple: Each model contained two ODEs with 5 to 6 free parameters. The compared models were a stoichiometric model, a stoichiometric model with cooperativity, a model with Michaelis-Menten-Kinetics for multiple

binding sites, and a model with standard Hill-kinetics for multiple binding sites, respectively. Cuccato et al. also accounted for multiple turnover or RISC recycling, which was compatible with our understanding of RNAi. However, the simplicity of the purely phenomenological models did not allow for inference on underlying biochemical reaction steps or mechanistic details on RNAi.

Another model of miRNA-mediated RNAi in mammalia was established by Sucheta, Gokhalea and Gadgil [57] in 2012. They used a simple system of four ODEs with a total of four lumped, dimensionless parameters. The parameter values were not determined experimentally and the model results were not experimentally validated. The model did not explicitly account for RISC assembly or the Argonaute protein itself. However, they were able to suggests a possible explanation for an "unexpected" increase in target protein levels when introducing miRNAs or siRNAs to a cell.

Most recently in 2013, Hausser et al. [74] published their mathematical modelling approach on miRNA-mediated RNAi. The model was based on a fairly complex system of ordinary differential equations (ODEs). Furthermore, simplified versions with analytical solutions of parts of the original model were provided. Model parameters were inferred from a diverse range of experimental data sets of low- and high-throughput studies, or were fitted directly to experimental data using the model. The model described time-dependent changes in mRNA, protein and ribosome density levels with a clear focus on miRNA expression level changes, e.g. via oscillations or other transcriptional changes. The second emphasis of the model was the binding of miRNA and Argonaute. In their study they identified two bottlenecks: the level of miRNA loading into Ago and the level of protein decay of the target. Since the authors did not find sufficient data for modelling the precise reactions that followed association of miRNA and Ago, the exact steps or mechanistic details responsible for the bottlenecks remain vague.

All in all it is important to keep in mind that a model never can be an exhaustive reflection of its real world process. Instead it is designed to examine certain aspects of the process within the real world to answer clearly defined objectives. The other way round holds also true: A certain process can be described in more than one way. The choice of mathematical models and numerical analyses methods depends on the intension of the modeller. Different modelling approaches highlight different aspects of the same system, in this case the RNAi system. As described above, none of the existing models of RNAi fulfils the re-

quirements completely, i.e. they do not provide insights into the mechanism and the kinetics of strand exchange between cleaved target and incoming target of siRNA-mediated RNAi. Nonetheless, the model structures and parameter values of some of the studies (most noteworthy Bartlett and Davis [8–10]) could be used as a rough starting point for the modelling approach of this doctoral study. Very reassuring were the compatibility of the results and conclusions with the findings of the studies of Larsson et al. [104] and Hausser et al. [74].

Chapter 2

Objective

The general objective of this study was to gain insight into RNAi in the human system by applying theoretical methods to aspects of the biological process that were otherwise inaccessible to experimental methods and whose behaviour was hard to predict, i.e. due to nonlinear effects. The study can be be divided into two parts, which covered different mechanistic aspects of the RNAi process and which were addressed by different methodologies from systems biology and bioinformatics respectively. The exact objectives of these two parts are described in the following sections.

2.1 *In sillico* modelling of RNAi in Homo sapiens

This part of the thesis focused on the mechanism and the kinetics of strand exchange between the cleaved target (i.e. the product) and the incoming target of siRNA-mediated RNAi in the humans system. The aim was to build, analyse and evaluate *in sillico*-based kinetic models which were able to quantitatively predict the kinetic behaviour of siRNA-mediated RNAi in living mammalian cells and to unravel underlying mechanistic details of the RNAi machinery. The models were based on precise quantitative kinetic data in order to gain a quantitative and time-resolved understanding of RNAi. They were analysed via numerical simulation techniques. Their credibilities were evaluated based on comparison to experimental observations on RNAi in mammalian cells.
De facto, before RNAi-based therapies and drugs will be able to fully reach the market there are some challenges to overcome. Firstly, mechanistic details of the complex mammalian RNAi system have to be deciphered and quantified in order to allow their safe application in human medicine. Important aspects, which drove this study were the needs for

- the prediction of the silencing potential of siRNAs with respect to the targeted mRNA
- understanding variations in efficacy of siRNA, e.g. in different cell types or tissues
- avoiding off-target effects and side-effects
- effective and specific delivery of small RNAs into target cells *in vivo*

Without a proper understanding of the interplay between the kinetics of the process and its parameters, any application of RNAi is governed largely by trial and error. The ability to specifically tailor and optimise the treatment for each particular system would save significant time and resources, especially given the high cost of synthetic siRNA molecules and the amount of material required for *in vivo* studies.

Formal model building

The aim of the model building step was the conceptual clarification of siRNA-mediated RNAi in Homo sapiens. Generally, gaps in knowledge and understanding were spotted during model formulation and identification. The main focus was on the comparison of two different theories on the mechanistics and kinetics of the strand exchange between the cleaved target (i.e. the product) and the incoming target. A first model, dubbed the dissociative model, was based on the current literature view of a dissociative strand exchange mechanism (e.g. [8, 11, 30, 60, 94, 104, 112, 133, 172, 197, 214]). It was compared to a new model, the so-called associative model, supported by novel experimental findings (see [32, 40]) for an associative stand exchange mechanism.

In this regard, the formal model building was divided into two sub-tasks:

1. Model formulation: The task was to build formal models for siRNA-mediated RNAi in Home sapiens with respect to the dissociative and the associative mechanisms. This included the model structure, i.e. the reaction steps that connect the conversion between molecular species and their intermediates; and farther the kinetic rate laws of these reaction steps. The model structure ideally is a trade-off between adequate modelling of the investigated mechanism and a minimum in model complexity.

2. Model identification: The tasks were the identification and assignment of model parameters. Mainly kinetic rate constants of reaction steps, volumes of reaction com-

partments, and initial concentrations of molecular species were identified and parametrised using recently published quantitative data and in collaboration with experimental scientists. Ideally, each model parameter should be assigned a precise quantitative value for a white-box modelling approach. Compare black-box versus white-box model, pp. 15.

Numerical analysis

Computational simulations ("*in sillico* experiments") were performed to examine scenarios not accessible by "wet lab" experiments, e.g. that followed dynamics of experimentally untraceable intermediates or that imposed conditions not feasible in experiment. Furthermore, simulations were used to perform parameter scans over large ranges of combinations of initial conditions, which would be too time consuming and expensive to examine experimentally.
The kinetic behaviour of the modelled RNAi system was numerically analysed to identify influential model parameters and reaction steps. Applied methodologies included systematic parameter estimation, sensitivity analyse, as well as structural and stoichiometric model analysis. Model optimisation and the variation of model complexity were applied to further identify relevant details of the modelled RNAi system via, e.g. condensing or extending reaction steps, reducing or increasing of mechanistic details, as well as exclusion or integration of larger parts of the reaction pathway.

Validation and evaluation

A crucial part of the study was the validation of whether or not a given formal model was able to describe the real life system accurately. Therefore, validation of the model structure and the appropriateness of its parameters was performed on base of empirical data from validation experiments that were executed by collaborating experimental scientists. Another important aim was the assessment of the model scope, i.e. to determine under which circumstances the model was valid and where its limit were. The two questions that were amid at were: *does the model adequately describe the properties of the system between data points?* and *is the model able to provide meaningful insights into events outside the observed data?*

During model evaluation, the goal was a decision between (or at least an educated statement about) the dissociative versus the associative mechanisms of ligand substitution during siRNA-mediated RNAi in Homo sapiens.

2.2 Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA

The second part of the thesis was dedicated to the computational examination of another indicator for the complexity of the RNAi mechanism in humans: the possibility of a potential post-transcriptional control step involved in regulating endogenous levels of human Ago2. For this, a novel bioinformatics tools was created for RNA motif prediction and was used to examine non-hairpin GC-helices in RNA libraries.

The ratio for this computational study was based on

- a study of Mallory et al. [132] that hinted the existence of a post-transcriptional control step in human Ago2
- preliminary experimental studies by Georg Sczakiel that indicated specific binding between RNA sequences surrounding the AUG start codon of the Ago2 mRNA and proteins involved in RNA interference.

Perfectly formed duplex elements in RNA occur within folding units, often as a part of hairpin motifs, which can be reliably predicted by various RNA folding algorithms. Double helices with consecutive Watson-Crick base-pairing may also be formed between distant RNA segments thereby facilitating long-range interactions of long-chain RNA that may be biologically functional.

The general aim was the investigation of potential formation of RNA duplex motifs by long-range RNA-RNA interactions of distantly located matching sequence elements of a single long-chain RNA. Special interest was on the analysis of highly GC-rich non-hairpin duplex elements (GC-helices) in mRNAs of the Argonaute-like gene family and the biological role of these GC-helices in the post-transcriptional regulation of human Ago2.
Initial computational RNA folding studies indicated the existence of two helices formed by consecutive GC base pairs within the $5'$-UTR and upstream coding sequences. The existence of these helices seemed to represent a rare case. Thus, the occurrence of these GC-helices in mRNAs was further characterised and it was studied whether the occurrence of GC-rich

helices could be biologically relevant. A novel bioinformatics tools was developed to perform systematic computational and phylogenetic studies about the observed GC-rich duplex motifs formed by distant RNA segments. The results were compared to experimental evidence in a mammalian cell system that supported the view that GC-rich duplex motifs formed by distant RNA segments, unlike typical hairpin elements, could bear biological functions.

Within the scope of this thesis were:

- the development of a software tool for RNA motif prediction, which accounts for primary sequence and secondary structure
- the identification and analysis of consecutive RNA duplex elements at variable length and nucleotide content formed by distant sequences in various RNA classes from various organisms
- the characterisation of highly GC-rich non-hairpin duplex elements (GC-helix) in the Argonaute (Ago)-like gene family and other gene families
- linking the occurrence of these GC-helices to post-transcriptional regulation of hAgo2.

Chapter 3

Materials and Methods

From a methodical point of view, the systems biology part of this study can be described as an iterative process to infer about the system of interest, i.e. siRNA-mediated RNAi in Homo sapiens (compare Fig. 3.1). It included the following sub-processes:

- formulating a clear objective
- analysing the system of interest using recent literature and state-of-the-art laboratory results from wet-lab scientists
- creating plausible formal models in accordance with the objective
- choosing appropriate modelling approaches and analysis methods
- designing validation routines in tandem with experimental biologists
- running computer simulations of the models
- visualising, analysing and interpreting the resulting data.

The according systems biology methodologies used in this study are described in detail in the following Sections 3.1 to 3.6.

While the last Section 3.7 is dedicated to the computational routines and the resulting bioinformatics tool developed within the scope of the second part of this thesis, the computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA.

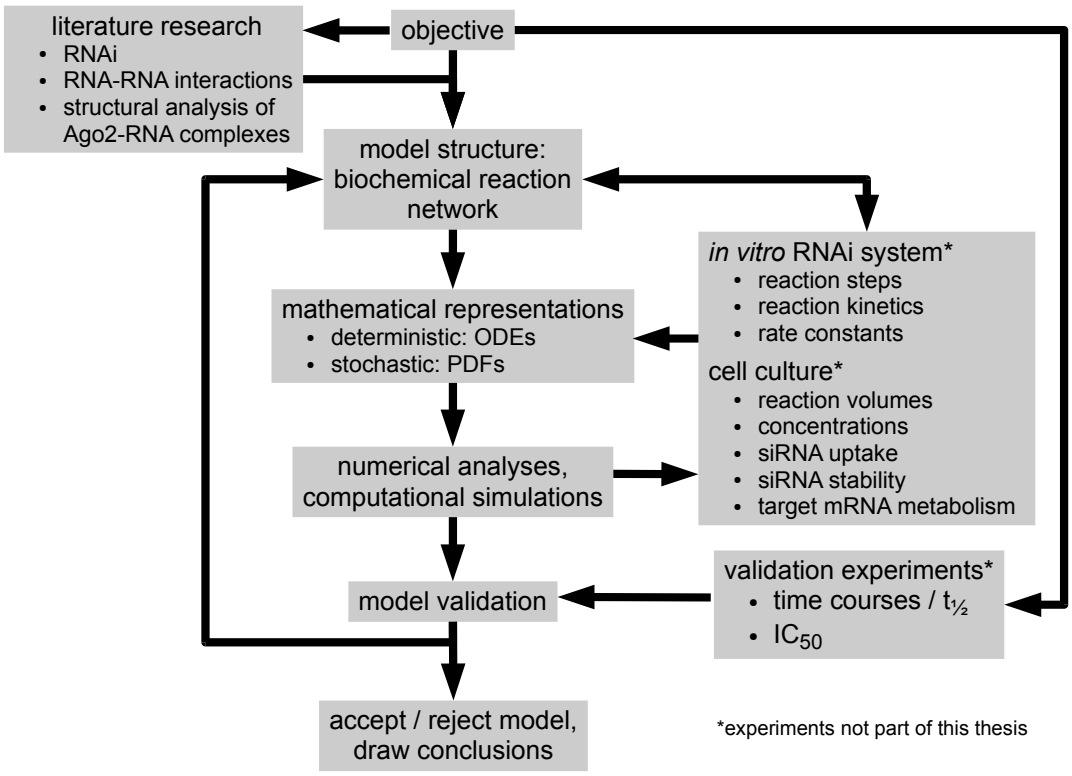


Figure 3.1 **From objective to conclusion: systemic modelling of RNAi.** This iterative process includes the formulation a clear objective and the analysis of the target system using recent literature and state-of-the-art laboratory results. Furthermore, it includes choosing an appropriate model (including model structure and a mathematical representation) as well as implementing that model for numerical analyses and computer simulation. Simulation output and results from validation experiments are used to decide about acceptance or rejection of the model.

3.1 Model building

3.1.1 Modelling of RNAi

From a modelling perspective, biochemical networks are a set of chemical species that can be converted into each other through chemical reactions.
In this study the RNAi process in a cellular context was formally described by a network of elementary reactions. Their biochemical rate equations connected experimentally derived *in vitro* reaction rate constants with physiological concentrations of reactants and cellular compartment volumes. The reaction mechanism was encoded in the reactions' rate laws, its parameters, and the way the reactions were linked to one another in metabolic pathways. The reactions translated in reaction equations.
In this thesis basic molecular entities, such as proteins or single strands of nucleic acids, are abbreviated by capital letter. Molecular complexes are indicated by combinations of capital letters. To distinguish between different complexes consisting of the same basic entities, dagger (†) and double dagger (‡) symbols are used. For instance, AS, AS†, and AS‡ are a collision complex between Ago (A) and ss RNA (S), an intermediate state of the ternary complex and the cleavage competent RISC, respectively.

This study included many types of computational and mathematical analyses, i.e. based on numerical algorithms, that can aid understanding of how the RNAi system works.

3.2 Kinetic analysis: time course simulations

Time course simulation results in time series of all model variables, such as molecular species concentrations, concentration fluxes and reaction fluxes, outputted in a sequence of data points, measured typically at successive points in time spaced at uniform time intervals. It is a powerful approach for characterising the transient or pre-steady state behaviour of chemical reaction system given explicit initial conditions.

3.2.1 Deterministic simulations – the ODE approach

ODE representation

Mathematically, a network of biochemical reactions can be transformed into explicit mathematical expressions, where each chemical species in the pathway is represented by an ODE

that describes the rate of change of that species along time given their initial values. The ODE is composed by an algebraic sum of terms that represent the rates of the reactions i that affect the chemical species. For a chemical species X:

$$\frac{dX}{dt} = \sum_i s_i \cdot v_i, \tag{3.1}$$

where s_i is a stoichiometry coefficient that is the concentration of species X consumed or produced in one cycle of reaction i. It has a positive sign if it is produced or negative if consumed. For reactions that do not produce or consume X the corresponding s_i is zero. The velocity v_i of each reaction is described by a rate law that depends on the concentrations of the reaction substrates, products, and modifiers. Rate laws are the subject of chemical and enzyme kinetics. Except in the case of first-order mass action kinetics, they are generally nonlinear.

Numerical Solutions

To examine the time-dependent concentration changes of reactants, intermediates and products over a given time interval, the system of ODEs was numerically integrated. Unless otherwise specified, the well-established deterministic LSODA solver[163] of the Fortran library ODEPACK[77] within the COPASI framework[80] was used. LSODA stands for Livermore Solver for Ordinary Differential Equations and solves the initial value problem of stiff and non-stiff systems of explicitly given ODEs, i.e. $dy/dt = f(t,y)$.
ODEs describing biochemical networks are often stiff, meaning that they contain very fast and very slow components and this poses a significant numerical problem. Methods such as forward Euler or Runge-Kutta are not appropriate when stiffness is present in the equations and can lead to completely spurious solutions because they accumulate truncation error.
In the stiff case, the LSODA solver treats the Jacobian matrix df/dy as either a dense (full) or a banded matrix, and as either user-supplied or internally approximated by difference quotients. It starts using the non-stiff Adams integration method (also known as predictor-corrector) before dynamically deciding whether to switch to a non-stiff approach based on backward differentiation formula (BDF), also known as the Gear method, for dense or banded Jacobians. Arising linear systems are then solved by direct methods, namely the Factor-solve approach with LU factorisation.
For all time course simulations in this study, relative and absolute tolerance were set to 10^{-6} and 10^{-12}, respectively. The maximum number of internal steps was fixed at 10^4, the option to integrate the reduced model was switched off. LSODA is a fast and reliable algorithm

and the solution for a $432,000$ sec (4 d) time course of the described ODE systems with 10 or 13 ODEs and 20 or 29 parameters for the dissociative and the associative models, respectively, was obtained within a few seconds on a MacBook Pro with 2.66 GHz Intel Core 2 Duo processor and 8 GB 1067 MHz DDR3 RAM, running COPASI 4.11 (Build 60, 64 Bit version).

3.2.2 Stochastic simulations – the PDF approach

If there exist reactions that contain small numbers of particles of each reactant during the time course, the assumption of continuous concentrations fails and consequently the underlying basis of the ODE representation also fails. Moreover, in such conditions, stochastic effects become more pronounced and may lead to dynamics that differ significantly from those that would result from the ODE approach, compare Section 3.2.3, pp. 32.

In the conditions described above, probability distribution functions were used for stochastic simulations in order to estimate when single reaction events happen and therefore track the number of particles of the molecular entities.

PDF representation

Stochastic models represent the number of particles of each molecular entity and use a reaction probability density function (PDF) to describe the timing of reaction events.

Transforming ODEs into PDFs

In this study models were defined as systems of ODEs and the COPASI framework, which is able to transparently switch between deterministic and stochastic simulations, was used to transform the equations to PDFs where needed. In the RNAi models kinetics were based on mass action kinetics, which allowed for a straightforward transformation of the reaction rates of the ODEs to corresponding reaction probabilities[56]. In the few cases where mass action kinetics of several reaction steps where lumped into more complex enzyme kinetic rate laws stochastic simulations were avoided even though several authors ([19, 175]) have suggested that it is justifiable to directly transform enzyme kinetic rate laws into PDFs and use them in stochastic simulations (methodological details can be found here: [79]).

Reversible reactions

While the forward and backward reaction rates can cancel each other out in deterministic simulations, this is not rue for stochastic simulations. Here each reaction event has to be considered separately. Thus, each reversible reaction step in the RNAi models had to be handled as separate forward and backward irreversible reactions for the use in stochastic simulations.

Numerical Solutions

There exist several mathematically equivalent algorithms for the stochastic simulation of biochemical networks. In this study a Monte Carlo simulation algorithm, known as the stochastic simulation algorithm (SSA) first reaction method, that simulates the stochastic dynamics of the system by sampling the PDFs was used[56]. Further methods are for example the Next Reaction Method[55] and a method by Cao & Petzhold[20]].

Probabilistic representation

It is important to stress that one simulation run according to this approach is only one realisation of a probabilistic representation, and thus provides a limited amount of information on its own. In this study stochastic simulations were repeated for a sufficient number of times in order to reveal the entire range of behaviour presented by the simulated system (i.e. to estimate a distribution for each molecular entity).

3.2.3 Hybrid simulations

When the simulations led to system states that contained some molecular entities with high particle numbers and some with low particle numbers, pure stochastic simulations were computationally too expensive, but a mere deterministic approach was not able to account for the low particle numbers other molecular entities. In these cases hybrid methods were able to simulate the models faster than pure stochastic methods, while still taking into account the random effects in reactions between molecular species with low particle numbers. There exist a number of hybrid methods to simulate biochemical systems (e.g. [1, 72, 95, 159, 170, 184, 185]). While they are all based on different combinations of mathematical methodologies, they are all aimed on partitioning the reaction network into subnetworks and use appropriate stochastic or deterministic simulation methods on each of those subnet-

works respectively.
A stochastic simulation algorithm, the Next Reaction Method by Gibson and Bruck [55] was combined with different algorithms for the numerical integration of ODEs from Section 3.2.1. The reaction networks in the RNAi models were dynamically separated into deterministic and stochastic subnetworks depending on the current particle numbers in the system. These reaction subnetworks were simulated in parallel with stochastic and deterministic algorithms, respectively.

3.2.4 Calculation of steady states

In a biological context, everything is in flow, where some processes are considerable faster than other, i.e. formation of chemical bounds happen in the range of femto seconds, while human growth happens over years. Only time scale separation leads to the perception of stationary behaviour. In fact, each steady state can be considered as quasi-steady behaviour in a larger, transient context, identified as transient states (or pre-steady situation), which can be best characterised with methods described in Section 3.2, pp. 29. However, the concept of steady state still is useful in modelling as it often reflects typical model behaviour by using simple mathematical concepts. In systems theory, a system in a steady state (also called stationary state or fixed point) has numerous properties that are unchanging in time. This means that for those properties X of the system, the partial derivative with respect to time t is zero. Steady states of a biochemical system are conditions when the concentrations of the chemical species X do not change.

$$\frac{\partial X}{\partial t} = \sum_i s_i \cdot v_i = 0, \tag{3.2}$$

If the steady state is such that the fluxes v_i are also zero, then the system is in chemical equilibrium, otherwise the fluxes are finite meaning that the concentrations do not change because the rates of synthesis balance with the rates of degradation for every chemical species. In this case, all state variables are constant in spite of ongoing processes that produce a flow through the system[201].

$$\sum_i v_i = 0, \tag{3.3}$$

$$\sum_i |v_i| \neq 0. \tag{3.4}$$

Often a steady state is achieved after sufficiently long time after a system is initiated, this is called asymptotic behaviour. Not all systems result in steady states, other asymptotic behaviour are oscillatory or chaotic regimes.
The steady states can be characterised with linear stability analysis and metabolic control analysis.

Numerical solutions

Steady states can be found using the Newton-Raphson method, a solver for nonlinear algebraic equations, which finds the roots of the right-hand side of the ODE (which must be zero by the definition of steady state). It is very fast, however, convergence is not guaranteed, i.e. it gets caught in a local minimum. It is unable to find multiple solutions (assuming there are more than one possible steady state). Furthermore, it cannot find unstable steady states. In these cases, forward or backward integration using LSODA can be used. This integration method usually is used to calculate a time course and is described in Section 3.2.1, p. 29. Hence, it attempts to find a steady state as it goes along a time course. In practice, first the Newton method is tried. If this does not converge, LSODA is used to integrate in time for a while and then the Newton method is tried again. This is repeated ten times, each time integrating $10\times$ further ahead of what was done earlier. If at the end a steady state is not found, the same strategy is applied again, but now integrating backward in time. Backward integration, if successful, will find an unstable steady state.

Newton-Raphson method

Here, the Newton-Raphson method was applied to solving the steady state solution of models by setting the rates of change to zero.

$$\bar{X}_{i+1} = \bar{X}_i - [\frac{\partial f(\bar{X})}{\partial \bar{X}}]^{-1} \cdot f(\bar{X}_i), \tag{3.5}$$

where $\bar{X}$ is a vector of variables to solve and $f(\bar{X})$ a vector of nonlinear equations of the type shown in Eq. (3.2). The matrix term, $\partial f(\bar{X})/\partial(\bar{X})$ is the Jacobian Matrix. Determining

whether the algorithm has reached convergence can be a accomplished by checking whether the relative error is less than a certain threshold ε. The algorithm can be summarised as:

1. random guess for initial values of species concentrations $\bar{X}_i$ with $i = 0$
2. compute $f(\bar{X}_i)$ (left-hand side of $d\bar{X}_i/dt$
3. compute matrix derivatives $\partial f(\bar{X}_i)/\partial(\bar{X}_i)$, i.e. $d(d\bar{X}_i/d\bar{X}_i)$, at current estimate for $\bar{X}_i$
4. compute inverse of matrix $\partial f(\bar{X}_i)/\partial(\bar{X}_i)$
5. for $i = i+1$, compute next guess $\bar{X}_i$
6. compute $f(\bar{X}_i)$
7. if $|f(\bar{X}_i)| < \varepsilon$, solution has been reached, else return to step 3.

3.2.5 Linear stability analysis

A stationary or quasi-stationary solution to a nonlinear system is called unstable if the linearisation of the system at this solution has the form

$$\frac{dr}{dt} = Ar, \tag{3.6}$$

where A is a linear operator whose spectrum contains points with positive real part. If there are no such eigenvalues, the solution is called stable. If at least one of the eigenvalues of A has a real part equal to zero then no conclusion can be made from the linear analysis. In these cases, nonlinear terms need to be considered. Compare [206] for a practical guide to stability analysis of biochemical systems.

3.2.6 Mass balance

The law of conservation of mass states that for any system mass cannot change quantity if it is not added or removed. Hence, the quantity of mass is "conserved" over time. The law implies that mass can neither be created nor destroyed, although it may be rearranged in space, or the entities associated with it may be changed in form. A mass balance, is an application of conservation of mass to the analysis of physical systems. By accounting for material entering and leaving a system, mass flows can be identified which might have been unknown, or difficult to measure without this technique. Mass balance is calculated using the algorithm described here: [211].

3.2.7 Elementary flux modes

For each reaction network there exists a unique set of elementary flux modes[161]. They correspond to the smallest possible reaction subnetworks to function in steady state[190, 204]. Elementary flux modes can be used to understand a reaction network's influence on a cellular level[204]. The analysis takes into account stoichiometrics and thermodynamics and allows the evaluation of whether a particular reaction route (in this study, e.g. competing ligand exchange pathways of RNAi) is viable for a set of initial conditions (i.e. initial molecular concentrations)[190]. In this study elementary flux modes were calculated using the COPASI-implementation of the METATOOL algorithm[164].

3.3 Sensitivity analyses

It is often difficult to predict the behaviour of the modelled biological system to perturbations from its reaction network structure alone and every model contains a number of parameters (e.g. reaction rate constants and initial concentrations) whose values are not all known exactly. Parameter changes influence the behaviour of the model. Sensitivity analysis quantifies how much the model behaviour depends on its parameters.
Sensitivity analyses were used at several ends in the study of RNAi models: In some cases the RNAi models' parameter values were inaccurate or even unknown, i.e. *in vivo* Ago2 concentration or the reaction rate constants for target cleavage. Sensitivity analysis was used evaluate how important it is to know a specific parameter value: the value of a parameter that influences the behaviour of the model significantly needs to be measured accurately; on the other hand, a rough estimate for a value may suffice if a parameter has little impact on the model behaviour. Furthermore, the robustness of RNAi against external perturbations, i.e. the transfection of siRNA or the target mRNA metabolism was estimated via sensitivity analyses. Likewise, it was determined which parameters influenced the RNAi efficacy.
In this study, two kind of sensitivity analysis are used and will be described in the following.

3.3.1 Metabolic control analysis

Metabolic control analysis (MCA) was used to analyse the concentrations and fluxes of intermediates that arise in the reaction networks of the modelled RNAi process during steady-state given a defined set of initial conditions.
MCA is able to describe how global reaction network dependent properties (control coefficients), depend on local properties (elasticity coefficients)[49, 186]. A control coefficient

is called a global quantity because it measures the steady state change in a system variable (e.g. a reaction pathway flux or a species concentration) in response to a relative change of another parameter (e.g. the steady-state reaction rate)[75, 89]. By way of illustration: a system in a steady state is perturbed by a small change in one of its parameters. After some time a new steady state will be reached in which potentially all the concentrations and fluxes in the system will have changed. The relative change of one of the reaction fluxes is the control coefficient of this reaction with respect to the perturbed parameter. An elasticity coefficient, on the other hand, is called a local quantity, because it quantifies the change of a local reaction rate to the change of another model parameter (e.g. a species concentration), while the rest of the network is not taken into account[49].
In this study MCA is based on the steady-state calculations described in Section 3.2.4, pp. 33).

3.3.2 Dynamic sensitivity analysis

Generic sensitivities

Response coefficients of parameters other than concentration or flux are called generic response coefficients. They have no known summation theorems. Hence they cannot be computed via a matrix method from elasticities, but are calculated numerically using finite differences. In practice a steady state is calculated (see Section 3.2.4, pp. 33); one of the parameter values is slightly changed, followed by calculating the new steady state. The ratios between the two results is used to estimate the differentials (the change applied is very small). These generic response coefficient computations allows the calculation of the sensitivity of a time course simulation. Compare Dornseifer et al.[40].

Multi-dimensional time-resolved sensitivities

Here, the objective was, to study the relationship between the global RNAi model behaviour and the properties of its parts, i.e. isolated reaction steps, as well as, initial concentrations and reaction volume. In this regard this analysis method is closely related to Metabolic Control Analysis (MCA)[75, 89]. However, MCA looks into steady state properties of reaction networks, whereas this analysis has been designed to comprehend the time-resolved properties of the modelled RNAi pathway.
Sensitivity coefficients ε_i^k quantified the sensitivity of a 24 h time course of target concentration to small parameter changes S_i as a function of basal target concentration $[T]_0 = C_k$.

The relative target concentration (in %) during the time course was differentiated with respect to the investigated parameters S_i using finite differences with delta factor 0.001 and delta minimum 10_{-12}. Investigated parameters S_i were kinetic rate constants k_{+1} to k_{+16} and k_{-1} to k_{-16}, initial concentrations of siRNA $[S]_0$, target $[T]_0$ and Ago2 $[A]_0$, as well as, cytoplasm volume V_{cyt}. Sensitivity coefficients were normalised between [-1, 1]. The sensitivity coefficients were compared with each other to determine which individual reaction steps controlled the change of target concentration during the investigated time course and which parameters were rate limiting under the investigated conditions. Compare Dornseifer et al.[40].

3.4 Optimisation scanning

To carry forward the idea of sensitivity analyses described in the previous section, it was sought for a way of exploring the space of complex multidimensional model behaviour using more realistic parameter ranges and not just calculations based on infinitesimal changes (as with MCA and dynamic sensitivity analysis). Optimisation methods (usually used in modelling to find conditions in which the model behaves in some desired way), offer a practical approach to explore complex models to answer specific constrained problems. For answering questions like *Given a set of known parameters, which conditions lead to a desired knockdown in target mRNA concentration?*, an objective function has to be defined along with parameter constrains. Optimisation is the search for minimum or maximum values of a so-called objective function. Because the RNAi models were composed of nonlinear functions, their variables may have several local minima or maxima. The objective was to find the global optimum, i.e. largest of all maxima or the smallest of all minima. In optimisation problems, it must also be specified which parameters of the model are allowed to change in order to meet the objective (the free parameters). Furthermore, for each parameter, an upper and a lower search limit should be provided to narrow the search space. Compare Hoops et al.[80].

3.4.1 Objective function

In the above example, the target mRNA concentration should be minimised. For this an appropriate objective function $S(P)$ was defined, along with free parameters h_i, i.e. those that were allowed to change (search space), parameter constrains c_i and d_i, to reasonably

limit the defined search space, and initial conditions e_i for all model parameters P.

$$S(P,h) = min(f(P,h_i)) \text{ with } c_i \leq h_i \geq d_i. \tag{3.7}$$

where $f(P,h_i)$ was the to be minimised variable (e.g. the target mRNA concentration at 3 h after siRNA transfection via Lipofectamine 2000) for each simulation outcome. This expression was optimised by means of an optimisation algorithm. For each optimisation problem several different algorithms were tried out and their results were compared.

3.4.2 Optimisation algorithms

Downhill simplex method

The downhill simplex method (also called amoeba method) can minimise a smooth and unimodal objective function in an n-dimensional space. The object function E(x, n) is the sum of least squares of the differences between experimental values and simulated values at the same positions. It depends on x-coordinates and free model parameters n. Each test point in this parameter space is arranged as a simplex, which is defined as a polytope of n + 1 vertices in n dimensions: for instance, a line segment, a triangle, or a tetrahedron in one-, two-, or three-dimensional space, respectively. The scale factor for the initial simplex is set to 10, where the edges' lengths of the polytope are inverse proportional to the scale factor. The behaviour of the object function is extrapolated at each vertex and the algorithm progresses by replacing one of these vertices creating a new simplex. That is, the worst point is replaced with the centroid of the remaining n points. If the new point is better than the best current point, the simplex is stretched out exponentially in the direction of the new point. Otherwise, stepping across a valley, the simplex is shrunken towards a better point. For the application of IC_{50} calculations, the maximum number of iterations was set to 200. Alternatively, the algorithm terminates if the variance of the values of the objective function at the vertices of the current simplex is smaller than 10^{-5}. Compare [153]. The method was chosen for its simplicity and relative robustness.

PRincipal Axis

Brent's PRincipal Axis (PRAXIS) algorithm directly searches for the minimum of nonlinear functions without calculating their derivatives. PRAXIS assumes quadratic form of the optimised function and repeatedly updates a set of conjugate search directions (the free parameters) one at a time. The algorithm is simple and fast but has its limits: it is not robust against the scaling of the objective function and thus may fail. Compare [54]. The

original Fortran code is available at http://www.netlib.org/opt/praxis. This method was tried for initial optimisations of the simpler models described in Section 4.1, but was not suited for more complex problems due above described limitations.

Genetic algorithm

Genetic algorithms (GAs) generate heuristic solutions, in general to global optimisation problems, by mimicking natural evolution based on concepts such as reproduction, inheritance, mutation, crossover, and selection[5, 6, 141, 142].

In this study, a population of 20 candidate solutions (e.g. an IC_{50} value or a time course), the so-called individuals or phenotypes, was evolved towards better solutions over 200 generations. Randomly chosen phenotypes between physiological meaningful boundaries were used as inputs. The random distribution was normal if these boundaries spun less than two orders of magnitude and exponential otherwise. For each candidate solution a set of free floating-point parameters (called "genes") were iteratively mutated and altered, e.g. by crossover over several generations. At each generation, reproduction occurred by pairing every individual with a random partner. This results in two offspring per mating by combining the two parent genomes, resulting in double the number of individuals. Each "gene" (model parameter) of an offspring was picked randomly from either of the two parents. Furthermore, each "genes" had a probability to mutate, that is, the parameter value changed slightly from the originally inherited value by adding to the parameter value a normally distributed random number with zero mean and a standard deviation of 10% of the original value. At the end of a generation, the phenotypes of each individual, i.e. parents and offspring, were randomly tested against a number of others, i.e. 20% of the population size. The number of individuals was reduced to the original number where only the 50% highest ranking candidate solutions were kept to start the next generation. To avoid local minima, when the fittest individual had not changed for the last 30 generations, the worst 50% of of the population were replaced by individuals with random "genes". Compare [5, 6, 141, 142].

Particle swarm

The particle swarm algorithm generate heuristic solutions to global optimisation problems, where only few assumptions about the problem are made and thus very large spaces of candidate solutions can be searched. In this study particle swarm was used as a substitute for GA in cases when no tight parameter boundaries could be decided for a particular optimisation problem. It originally was inspired by a school of fish searching for food. In general, a population (called a swarm) of candidate solutions (called particles) is moved around in

the search-space spanned by the free model parameters according to a few simple concepts: Each particle i has a position p_i and a velocity v_i within the j-dimensional parameter space. The particles track their so-far best objective value s_j and the corresponding position in parameter space. Dependent on its own best known position and the position of the best particle of a random subset of the swarm n_i, a new velocity is calculated:

$$v_i = c_0 \cdot v_i + c_1 \cdot r_1(s_{i,j} - p_{i,j}) + c_2 \cdot r_2(n_i - p_i), \tag{3.8}$$

where r_1 and r_2 are uniform random numbers in $]0,1[$; c_0, c_1 and c_2 are internal parameters to optimise the method's behaviour and have been calculated by an overlaying optimiser, according to[139]. Finally, the particles position is updated accordingly ($p_i = p_i + v_i$). The process is repeated a defined number of times. It is not guaranteed that a satisfactory solution will be discovered. Compare [92, 139].

Simulated annealing

This stochastic global optimisation algorithm was inspired by the way in which perfect crystals are formed by the annealing (i.e. to heat and allow to cool slowly) of metal or glass. At high temperatures the material's particles vibrate with wide amplitude, which allows for overcoming local optima and thus enables a wider search for the global optimum. As the temperature and vibration amplitudes decrease the system settles to the global optimum to form a perfect crystal[96]. Similarly, in simulated annealing the objective function $E(s)$ is a measure of the energy of the system with the goal to bring the system, from an (arbitrary) initial state s_0, to a state with minimum energy. In each iteration step, the free parameters are randomly changed to a close by position in parameter space and the objective function is re-evaluated. If its energy decreased, the new state is accepted. Otherwise, the new state is accepted with a certain probability according to a Boltzmann distribution. This means, the probability depends on a temperature value, where higher temperatures correspond to higher probabilities. Starting at a high initial temperature, with each iteration step, the temperature is reduced which reduces the probability of accepting a new system state. Compare [96]. In this study simulated annealing was used as an alternative to the GA and particle swarm methods.

Random number generator

All random numbers for the optimisation methods were generated using the Mersenne Twister random number generator [137].

3.5 Parameter estimation

Biological models depend on many parameters, but quite frequently the exact values of some of these parameters are unknown and have to be estimated from experimental data.

3.5.1 Statistical search

The global optimisation algorithms used for parameter estimation in this study (see pp. 39) were stochastic and thus did not always converge to the same solution. Therefore each parameter estimation task was repeated multiple times with at least three different global algorithms, before the best solution, e.g. the average of all runs, was chosen and further optimised with local optimisation algorithms, i.e. praxis or the simplex method.

3.5.2 Objective function

The optimisation algorithms were used for estimating parameter values that best fit a set of data from experimental observations, allowing mixtures of time course and steady-state data to be used simultaneously[80]. The objective function in the case of parameter estimation was given implicitly by a function that measures the distance between the model and the experimental data, in this study a sum of squares of residuals[88]. Depending on the scientific problem, different sets of model parameters P were chosen to be included as variables into the parameter estimation process. These variables were then adjusted to minimise a weighted sum of squares S(P):

$$S(P) = \sum_i \sum_j \sum_k \omega_{i,k}(x_{i,j,k} - y_{i,j,k}(P))^2, \tag{3.9}$$

where $x_{i,j,k}$ is the experimental data and $y_{i,j,k}(P)$ the corresponding modelled data. In complex biological systems, measured parameters like molecular species concentrations are likely to vary between independent experiments. Thus, the sums in Eq. (3.9) were calculated with respect to each variable i for each independent repetition j of each experiment k. Thus, it was possible to fit the models simultaneously to data from multiple steady-state and time course experiments. Weights $\omega_{i,k}$ were used to make all trajectories of each variable i within each experiment k have similar importance in the fit. In this study the weights were calculated using standard deviation to set larger weight to data containing less fluctuations.

$$\omega_{i,k} = \frac{1}{\sigma_{i,k}^2}, \tag{3.10}$$

where $\sigma^2_{i,k}$ is the variance of the trajectory of variable i between repetitions of experiment k. It was used in cases, in which fluctuation in experimental data was suspected to be the result of inaccuracy or insecurity. The minimisation of Eq. (3.9) was achieved with the optimisation algorithms described on pp. 39 and were the same methods used for optimisation scanning (Section 3.4).

3.6 Model validation and evaluation

A crucial part of the modelling process was the validation of whether or not the given mathematical models described the biological system, i.e. siRNA-mediated RNAi in humans accurately and sufficiently given specific scientific objectives, e.g. exploring kinetic and mechanistic details on the strand exchange between the target mRNA and the products. The validation of the models' accuracy and correctness involved several different processes as described in the following. Model building and validation was an iterative process leading to several generations of modelling approaches with different levels of complexity. Compare Fig. 3.1.

Methodologically closely related is the model evaluation process. The RNAi models were evaluated using data sets independent from model building, parametrisation and validation. During model evaluation the best model that represents the evaluation data set was analysed. The focus was on deciding between RNAi models with different reaction pathways of the strand exchange mechanism.

Fit to empirical data

The parametrised candidate models were fitted against verification data from experimental measurements independent of the training data (used for model building and parametrisation) according to the process in Section 3.5 with algorithms described in Section 3.4.2. An accurate model will closely match the verification data even though these data were not used to set the model's parameters.
This practice is comparable with cross-validation in statistics. Defining a metric to measure distances between observed and predicted data is a useful tool of assessing model fit. However, different model structures can produce similar model behaviour and thus similar output. It is therefore not sufficient to to examine model output alone to choose between alternative models and conclude mechanistic details of the examined system.

While it was rather straightforward to test the appropriateness of the parameters, it was more difficult to test the validity of the general mathematical form of a model. In general, few mathematical tools have been developed to test model structure and formalisms involving models based on differential equations[80, 97, 150]. However there exist a lot of practical modelling studies that extensively use differential equations. Examples are provided in Section 1.3.2.

Scope of the model

Assessing the scope of a model, that is, determining under which conditions a model is applicable, can be challenging. The models were assessed for events between observed training data points (interpolation) and events outside the boundaries of the known data, where the system was not accessible by experiment methods (extraposition). For instance, most kinetic laws of reaction steps and reaction rate constants used in the modelling approach of Section 4.5 were determined from well-known *in vitro* experimental studies[32, 40, 226] with all details available. Other parameters like compartment volumes, molecular species concentrations and kinetics of preceding modelling approaches were extracted from literature data (sources and details are provided in Tables 4.1 to 4.12 in the Results Chapter) with limited access to raw data and no influence on the experimental set-ups. Model behaviour had to be extrapolated from these heterogeneous, mostly *in vitro* sources to the more complex *in vivo* system, i.e. RNAi in the human organism. For this, several independent test data sets of time-resolved target knockdown in human cells were used for model validation and evaluation.

Phase space trajectories in a multidimensional space

Each of a model's parameters can be though of as a degree of freedom in phase space and it can be represented by an axis of a multidimensional space. Theoretically, each possible model state corresponds to one unique point in this phase space. The model system's time course simulation follows a path through the phase space. This path is also called phase space trajectory. At deterministic behaviour, exactly one path represents the phase space trajectory can be archived by one or more sets of initial conditions, while for nondeterministic systems the phase space trajectory is a set of paths given the initial conditions[1].

For a better understanding, a more intuitive way to represent the complex simulation data was sought after. It was decided to use contracted measurements for a meaningful un-

derstanding of the time-resolved model simulations: $t_{1/2}$, the time of half-maximal target knockdown, and IC_{50}. Dornseifer et al.[40] define the time of half-maximal target knockdown and IC_{50} as follows:

Time of half-maximal target knockdown ($t_{1/2}$)

$t_{1/2}$ was defined as the time of half-maximal target knockdown, where target concentration $[T]t_{1/2} = [T]_0/2$, starting with basal target concentration $[T]_0$ at the time of siRNA administration. It was used as a contracted measure of time-resolved target knockdown to relate model behaviour with the two model parameters basal target level and initial siRNA concentration, for instance for the use in in 3D-plots.

Extraction of IC_{50} values from experimentally concentration-response curves by nonlinear regression

IC_{50} is defined as the half maximal inhibitor concentration. In this study, it indicated the quantity of siRNA required to halve the basal target mRNA level within 24 h. The measure is used to compare the potency of siRNAs at different levels of target mRNA. Experimental reference values for IC_{50} were taken from Dornseifer et al.[40]. They were extracted from concentration-response curves by nonlinear regression.

Computation of IC_{50} values from models by parameter optimisation

In the case of models, IC_{50} values can be either extracted from dose-response curves relating the initial siRNA concentration in copies/cell to target gene activity in percent after 24 h time course simulation. A faster way of calculating a large number of IC_{50} for different initial conditions is via parameter optimisation algorithm. Here, the free parameter (IC_{50}) is set as the initial siRNA concentration $[siRNA]_0$ that minimises the expression $|[target]_{t=24h}$ (%) -50%| while all other parameters are kept fixed during a 24 h time course simulation. A simple and robust iterative algorithm (downhill simplex algorithm[153]) is used to solve the optimisation problem for a large number of initial conditions. In the few cases (< 1%), when no convergence was reached in reasonable computation time, the optimisation was repeated using a slower heuristic method (e.g. genetic algorithm or particle swarm).

3.7 RNA duplex motif prediction

For the prediction of RNA motifs it was decided to calculate and analyse the secondary structure space (rather than simply analysing primary sequences) for inverted repeats which may form duplex elements. It was aimed at minimising false helix predictions which may increase the significance of predicted helices compared to merely localising of reverse complementary sub-sequences. The methodology described here was published by Dornseifer and Sczakiel[39].

3.7.1 Automation of RNA secondary structure prediction

The software tool that was developed for this task was written in Python version 2.6.1 and, additionally, integrated the pyExcelerator library. The source code and a sample data set are available for academic users on request. In this study, Mfold version 2.3, Mfold 3.4 and Sfold 2.2[36, 37, 240, 241] were used for mRNA secondary structure prediction. However, generally most other RNA secondary structure prediction software can be used with the newly developed tool as well. Mfold 2.3 was chosen because it provided the best match for the structure-function relationship of antisense oligonucleotides (asON) and siRNA, respectively by Mfold version 2.3 compared to six alternative folding software tools[158, 238]. Considering that mRNAs usually serve as target RNA for asON and siRNA one might conclude that Mfold version 2.3 predicts local structures of mRNA in a reliable fashion[98].

The computational prediction of secondary structure of RNA was based on a large collection of folded sequence stretches differing in position, length or both along a given mRNA of interest. A sequence stretch is called *window* in the following. Collections of folded subsequences can be defined systematically by sliding a window at a given step width along the sequence to be analysed. Predicted local structural elements that occur at high frequency in a large number of folded windows (independent of neighbouring sequences turned out to have a greater chance of being meaningfully related to the biology of corresponding RNA elements: this approach of analysing RNA secondary structure computationally turned out to produce more reliable results compared with a single fold of the complete long-chain RNA sequence of interest. Further, in terms of calculation time this concept provided advantages as well (see Section 5.2.5, p. 150). In this study, runs with three different structure prediction methods were performed with varying window sizes between 500 and 1,200 nucleotides at a step width of 20 nucleotides for each mRNA sequence. Ten optimal and

suboptimal structures were considered per fold. For example, at a step width of 20 nt and a window size of 800 nt a RNA sequence of 3,500 nucleotides in length (approximately the length of hAgo2 mRNA) means 135 Mfold runs resulting in 1350 overlapping structures. Finally, the structures were scanned for helices using the definitions in Section 5.1.1, p. 140.

3.7.2 Automation of RNA motif analysis

The automated identification of duplex motives composed of distant segments is illustrated as flow chart in Fig. 3.2. First, a collection of mRNA sequences was downloaded from online resources like EMBL Nucleotide Sequence Database or NCBI Nucleotide. The generated sequence collection contained mRNA species of interest for which secondary structures were predicted in a systematic fashion as described in detail[99]: a long-chain sequence was sub-divided into overlapping sequence segments (windows) starting at position #1 and the window was slid along the complete sequence. Secondary structures were predicted and, depending on the step width of the shifting window and its length, a large number of differing structure predictions was produced. This output of secondary structures was parsed for user-defined duplex motifs and was used for the identification of species with specific characteristics such as GC-rich helices formed by distant sequence segments in a non-hairpin like fashion. The software tool allows to filter hits by setting parameters such as the base composition of the duplex, the range of distance between strands, and the range of duplex length.

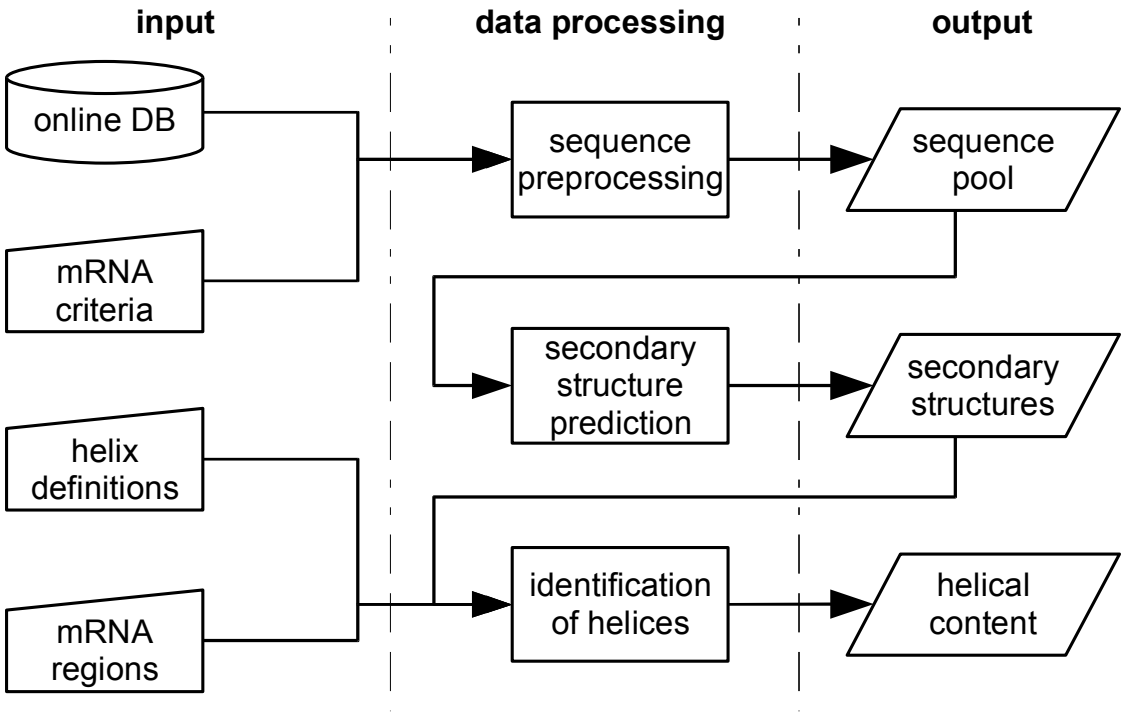


Figure 3.2 **Schematic depiction of the computational search for RNA motifs comprising duplexes.** This figure is based on Figures 1A of[39].

Chapter 4

Results: *in sillico* modelling of siRNA-mediated RNAi in mammalian systems

The objective of this part of the doctoral study is mathematical modelling and, based on that, computational simulation of RNA interference in the human system. The results presented in this Chapter span three distinct modelling approaches with different levels of complexity due to (i) different quantitative biological data available at the time of model building; due to (ii) the focus on and consideration of different details of the real-world RNAi process in the human system. Thus, applied analysis methods and validation processes vary between the presented modelling approaches.
However, all modelling approaches have in common that they regard two hypothetical, alternative biochemical reaction paths of ligand exchange during RNAi. The main objective is the identification of kinetic details during RNAi and, ideally, the results are able to clearly recommend one reaction paths over the other, given the available biological validation data. One reaction path (dissociative path) describes the current view on RNAi in the human system, as found in recent literature, whereas the other one (associative path) is based on new experimental findings by Rosel Kretschmer-Kazemi Far and Georg Sczakiel, as well as, by Sarah Willkomm, Andrea Deerberg, and Tobias Restle, whose experimental studies have preceded or accompanied this computational systems biology study, respectively[32, 40, 226].

The names of the two alternative models are based on the underlying ligand substitution mechanisms and mechanistic details will be provided in the following two sub-sections.

Dissociative model: RNAi with an assumed dissociative path for substrate binding and product release

In the biochemical ligand substitution mechanism, the departing ligands (here, cleaved RNA fragments) leave the complex in a slow step with positive entropy of activation, before another incoming ligand (i.e. target mRNA) can enter, compare Fig. 4.1A. Dissociative rates typically do not depend on the concentration of incoming ligand but on the slow dissociation of the departing ligand[226].

Associative model: RNAi with an assumed associative path for substrate binding and product release

Different from the dissociative mechanism, an associative mechanism involves the approach of the incoming ligand to the complex before departure of the leaving ligand. The first step of the associative mechanism typically is rate-determining and depends on the ligand concentration[226]. The assumption of an associative mechanism affects the modelling of target strand recognition by RISC. Here, hAgo2 supposedly facilitates the interaction of an incoming target strand with RISC-bound guide-product duplexes most likely via a strand invasion mechanism as first shown for *in vitro* RNA-RNA strand exchange by Wünsche & Sczakiel 2005[226]. Fig. 4.1B pictures the target-product-exchange mechanism, which is key to the multiple turnover cycle of the associative model.

The modelling of RNAi was an iterative process, comprising of three different modelling approaches. The implementation of the two alternative models differs between the three approaches. They will be described in the beginning of the three Sections 4.1, 4.2 and 4.5, where the last section contains the final, published versions of the dissociative and associative models.

4.1 Minimal models of dissociative and associative paths for single-stranded siRNA-mediated RNAi

Dissociative model

A minimalistic mathematical model of siRNA-mediated RNAi in mammalian cells was created in order to link experimental results with a quantitative and time-resolved understanding of RNAi. Mechanistically, it is based on the current literature view on RNAi, which is

discussed in Chapters 1 and 6. In a first pass, RISC, comprising hAgo2 (A) and possibly other co-factors, binds to free siRNA (S) to form complex AS. RISC co-factors, passenger strand cleavage and release during complex formation involving ds siRNA are omitted at this point for clarity. Essentially, the proposed pathway also holds true for ds siRNA. A free target target mRNA (T) enters the complex AS, eventually binding to the siRNA's guide strand (through formation of Watson-Crick base pairs) to form complex AST. Once the catalytically active complex is formed the system undergoes multiple rounds of target binding and cleavage, followed by a slower product release without falling apart: The bound target mRNA is cleaved by Ago2's endonuclease activity (target cleavage reaction) leading to complex ASQ. Subsequently, the remaining cleavage product (Q) are released from the complex. Thus, the target mRNA population is reduced by one and the free product strands are prone to degradation by cellular nucleases, leaving behind binary complex AS comprised of siRNA's guide strand bound to Ago2 which can bind the next target molecule and thus close the cycle of subsequent passes of target turnover. The reaction scheme for the dissociative model is shown in Fig. 4.1A.

$$\mathrm{A} + \mathrm{S} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_{+1}}{\rightleftharpoons}} \mathrm{AS} \tag{4.1}$$

$$\mathrm{AS} + \mathrm{T} \underset{\mathrm{k}_{-2}}{\overset{\mathrm{k}_{+2}}{\rightleftharpoons}} \mathrm{AST} \tag{4.2}$$

$$\mathrm{AST} \underset{\mathrm{k}_{-3}}{\overset{\mathrm{k}_{+3}}{\rightleftharpoons}} \mathrm{ASQ} \tag{4.3}$$

$$\mathrm{ASQ} \underset{\mathrm{k}_{-4}}{\overset{\mathrm{k}_{+4}}{\rightleftharpoons}} \mathrm{AS} + \mathrm{Q} \tag{4.4}$$

Associative model: accelerated product release via strand exchange between RISC-bound, cleaved mRNA and incoming target mRNA.

Cell culture experiments showed that gene silencing rates can be increased by rising the target mRNA concentration, at constant siRNA and hAgo2 concentrations. That is, IC_{50}, the half maximal inhibitory siRNA concentration after 24 h of knockdown, stays relatively constant over several magnitudes of basal target mRNA levels[40]. Additionally, it was shown *in vitro* that the release of the target RNA strand within a RISC bound RNA duplex is accelerated in the presence of free target RNA[32].

These experimental findings suggest that RNAi-based knockdown should be able to adapt to changes in target mRNA level over several magnitudes. The defiance of this adaptation in the dissociative model (compare Fig. 4.1A) motivates an alternative model. Different

from the dissociative mechanism, an associative mechanism involves the approach of the incoming ligand to the complex before departure of the leaving ligand. Here, an associative mechanism of target strand recognition by RISC during siRNA-mediated RNAi (dubbed associative model) is proposed. In this model, hAgo2 supposedly facilitates the interaction of an incoming target strand with RISC-bound guide-product duplexes most likely by a strand invasion mechanism as first shown for *in vitro* RNA-RNA strand exchange by Wünsche & Sczakiel 2005[226]. Fig. 4.1B pictures the target-product-exchange mechanism, which is key to the multiple turnover cycle of the associative model. However, before entering the cycle of multiple turnover, in a first pass, hAgo2 and possibly other co-factors, binds to free siRNA to form complex AS followed by binding of a first target to form complex AST. The first pass of siRNA binding and target binding, as well as, target cleavage are equivalent to the reaction steps of the dissociative mode. However, after formation of complex ASQ which comprises hAgo2, siRNA and cleaved product fragments, the two models differ. In contrast to the dissociative model, target cleavage is followed by an associative binding of another target (associative target binding) which may invade the hydrogen bonds between siRNA and the product, forming transient "quaternary" complex ASQT eventually replacing the product strands. The cycle is closed by the complete release of the replaced product fragments (product release), which leads to the formation of the next catalytically active complex AST. The released product fragments Q eventually are degraded by cellular nucleases. The remaining ternary complex can pass through the next, identical cycle of target cleavage, target association and product release. The reaction scheme for the associative model is shown in Fig. 4.1B.

Accordingly, in this associative model, Eq. (4.5) & (4.6) provide an alternative path of target binding and product release. These two reaction steps were later lumped into one reaction, since available literature data did not allow for extracting separate parameters for the single steps, resulting in the lumped reaction of Eq. (4.7).

$$\mathrm{ASQ} + \mathrm{T} \underset{\mathrm{k}_{-5a}}{\overset{\mathrm{k}_{+5a}}{\rightleftharpoons}} \mathrm{ASQT} \tag{4.5}$$

$$\mathrm{ASQ} \underset{\mathrm{k}_{-5b}}{\overset{\mathrm{k}_{+5b}}{\rightleftharpoons}} \mathrm{AS} + \mathrm{Q} \tag{4.6}$$

$$\mathrm{ASQ} + \mathrm{T} \underset{\mathrm{k}_{-5}}{\overset{\mathrm{k}_{+5}}{\rightleftharpoons}} \mathrm{AS} + \mathrm{Q} \tag{4.7}$$

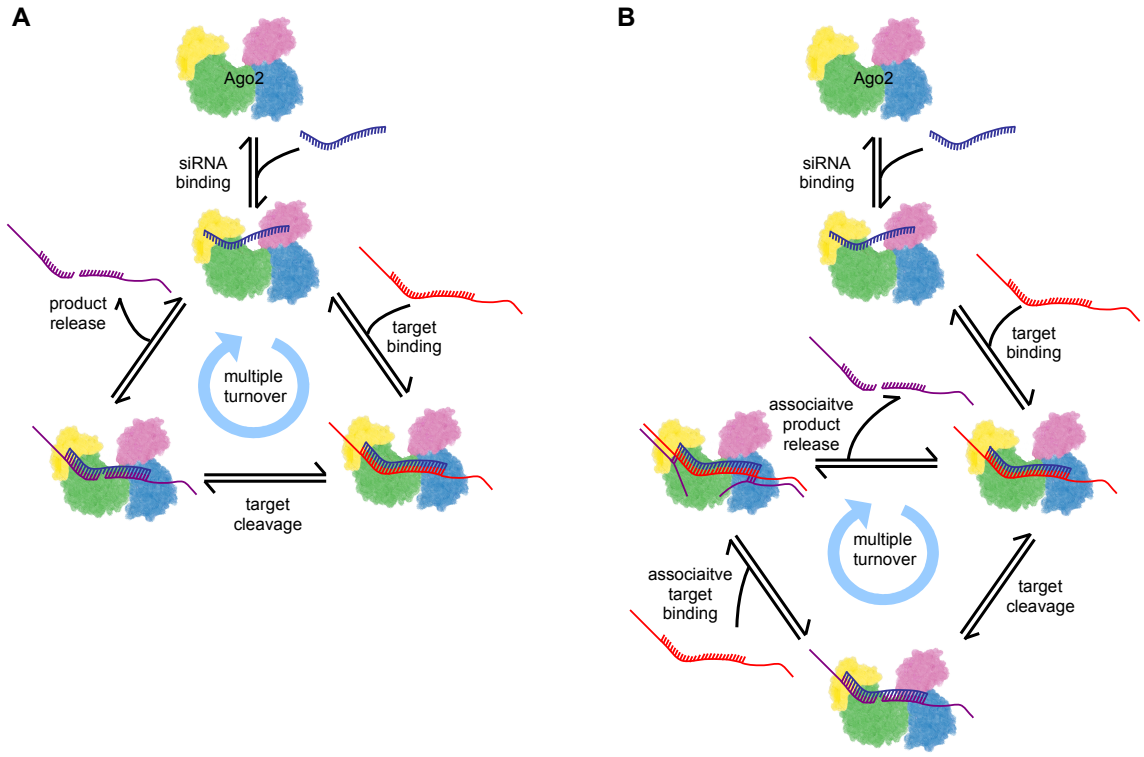


Figure 4.1 **Illustrations of multiple turnover cycles of two competing models of ss siRNA-mediated RNAi.** In these illustrations Ago2 is represented as a scheme of its electron density cloud with the four domains N, PAZ, Mid, and PIWI coloured individually. RNAs are indicated by lines and are as well colour-coded. The relative positions of protein and RNA substrates are indicated and the formation of Watson-Crick base pairs is illustrated by short connecting lines between RNA strands. **A)** Dissociative model. Once the catalytically active complex between ss guide siRNA and Ago2 is formed, the system undergoes multiple rounds of target binding and cleavage, followed by product release. **B)** Associative model. In contrast to the dissociative model, target cleavage is followed by the associative binding of the next target, which may invade the hydrogen bonds between siRNA and product, eventually replacing the product fragment, .which leads to the formation of the next catalytically active complex. For clarity, only the most relevant steps are shown.

4.1.1 Model parameterisation

Both the dissociative and the associative models were parameterised for RNAi-based knockdown of an arbitrary target mRNA in human cells in the presence of highly effective siRNA. Unfortunately, at the time of the study a conclusive set of parameters was not available, and the parameters had to be extracted from different literature sources, as well as, lab reports of in-house experiments. As the two models share reactions, i.e. Eq. (4.1) - (4.3), they also share these reactions' parameters.

Complex formation between Ago2 and siRNA was experimentally determined by Bartlett and Davis for siGL3 in the presence of its target mRNA in human hepatocytes[8]. The time-resolved change in binary complex resembles an exponential function. Thus, the time course data was exponentially fitted to estimate a reaction rate constant assuming a single reaction step with mass action kinetics for siRNA binding.

Complex formation between minimal programmed RISC, comprising of Ago2 and siRNA, was extracted from[29]. Cuccato, Polynikis and Bernardo numerically fitted the rate of mRNA-siRNA complex formation for four different ODE-based models to *in vitro* experimental data of knockdown of EGFP via ds siRNA. The values of three out of four models resulted in the same magnitude for the estimated reaction rate constant and, in this study, the average of the estimates of these three models is used for ternary complex formation.

The kinetic parameter of the catalytic step (target cleavage) is based on the pre-steady state RNase H rate of HIV1 reverse transcriptase[90], since data for Ago2's own endonuclease activity is not available yet.

The release of target mRNA from RISC was described in three different publications[15, 67, 135] and all three values are in the same magnitude. Their average was used in a modelling approach by[8]. In this study this reaction rate constant is used as a landmark for the kinetics of the release of cleaved target mRNA during dissociative target release.

Strand exchange between ligated sense and antisense RNA strands and free antisense RNA was measured in the presence of cetyltrimethylammoniumbromid (CTAB) to mimic the degree of freedom of RNAs in living cells[226]. The resulting exchange rates were used as estimates of the magnitude of exchange of product strand from the complex with siRNA-Ago2 via free target RNA. See Table 4.1 for a summary of all kinetic parameters.

The cytoplasm volume is calculated assuming, on average, a spheric cell shape with known cell radius r_{cell} and spherical nucleolus with radius r_{nuc}. Cell radii (r_{cell}) were measured for 100 ECV cells via light microscopy (Winfried Wünsche, unpublished data). Further values for cell and nuclei radii were taken from Fujioka, Terai, Itoh et al. for HeLa cells[53]. The average radius was set to $r_{\mathsf{cell}} = 3.85\ \mu$m or $8.1783\ \mu$m for HeLa and ECV cells, respect-

model	reaction	constant	units	value	ref.
both	siRNA binding	k_{+1} k_{-1}	$M^{-1}s^{-1}$ s^{-1}	10 $\simeq 0$	[8]
both	target binding	k_{+2} k_{-2}	$M^{-1}s^{-1}$ s^{-1}	1×10^6 $\simeq 0$	[29]
both	target cleavage	k_{+3} k_{-3}	s^{-1} s^{-1}	10 $\simeq 0$	[90]
disso	dissociative product release	k_{+4} k_{-4}	s^{-1} $M^{-1}s^{-1}$	3×10^{-5} $\simeq 0$	[8, 15, 67, 135]
asso	associative target product exchange	k_{+5} k_{-5}	$M^{-1}s^{-1}$ $M^{-1}s^{-1}$	1.17×10^5 5.07×10^5	[226]

Table 4.1 **Kinetic parameters of a minimal model of ss siRNA-mediated RNAi.** The models share many of their reaction rate constants, which is indicated with 'both' in column one. The tags 'disso' and 'asso' indicate a parameter specific for the dissociative or associative model, respectively. When $\simeq 0$ is given as value in column five for a backward reaction rate, the corresponding reaction is modelled as quasi irreversible.

ively. In both cases, the nucleus radius r_{nuc} was set to an average of 3.68 μm. This results in a cytoplasm volume of 0.91525 pl and 2.2912 pl for HeLa and ECV cells, respectively, see Eq. (4.8) and (4.9). Accordingly, the average cytoplasm volume of a human cell was set to $\sim$ 1.6 pl, when not stated otherwise. A human cell contains 1.4×10^5 to 1.7×10^5 Argonautes per cell[179, 217], where Ago2 is the most abundant of the four Ago-subtypes. According to unpublished data of Anke Detzer, there are 2.5×10^5 copies Ago2 per human cell. Non-kinetic parameters are summarised in Table 4.2.

$$V_{cyt} = V_{cell} - V_{nucl} \tag{4.8}$$

$$r_{cyt} = 5\sqrt[3]{6/\pi}\sqrt[3]{cbrt(V_{cyt})} \tag{4.9}$$

$$\tag{4.10}$$

4.1.2 Analyses of the minimal models

Scrutinising the models: Stoichiometric Analysis

Elementary flux modes could be identified for either of the two models. This implies, that the proposed models do not comprise smaller sub-networks that allow a metabolic reconstruction network to function in a steady state.

parameter	symbol	unit	value or boundaries
cytoplasm volume	V_{cyt}	l	1.6×10^{-12}
basal Ago2 concentration	$[A]_0$	copies/cell M	$200,000$ 2.08×10^{-7}
basal target concentration	$[T]_0$	copies/cell M	$(1,\ 250,000)$ $(1.04 \times 10^{-12},\ 2.59 \times 10^{-7})$
initial siRNA concentration	$[S]_0$	copies/cell M	$(1,\ 250,000)$ $(1.04 \times 10^{-12},\ 2.59 \times 10^{-7})$

Table 4.2 **Non-kinetic model parameters (ss siRNA-mediated RNAi).** The models share many of these non-kinetic parameters. Concentrations are given in copies per cell, as well as in M. Initial concentrations of reaction species that were not mentioned in the table were set to zero.

Mass conservation analyses of dependent species starting with different combinations of initial species concentrations, led to the identification of the following three mass flows in Eq. (4.11) - (4.13) identical for both models. By accounting for material entering (i.e. siRNA S) and leaving (i.e. cleaved product Q) the system, mass conservation is true at all time for both models. Mass is neither created nor destroyed, although it is rearranged in space, and the entities associated with it (i.e. target mRNA, siRNA, Ago2) are changed in form.

$$T_0 = Q + AST + ASQ + T \tag{4.11}$$

$$A_0 = A + AS + AST + ASQ \tag{4.12}$$

$$S_0 = S + AS + AST + ASQ \tag{4.13}$$

The stoichiometries of the dissociative and associative models can be compactly represented by stoichiometry matrices in Eq. (4.14) & (4.15), respectively.

$$N_{diss} = \begin{array}{c} \\ T \\ Q \\ A \\ S \\ AS \\ AST \\ ASQ \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_2 & r_3 & r_4 \end{array} \\ \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \end{array} \tag{4.14}$$

$$N_{asso} = \begin{array}{c} \\ T \\ Q \\ A \\ S \\ AS \\ AST \\ ASQ \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_2 & r_3 & r_5 \end{array} \\ \begin{pmatrix} 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \end{array} \tag{4.15}$$

Reactions siRNA binding, target binding, target cleavage, dissociative release, and associative binding are abbreviated as r_1, r_2, r_3, r_4, and r_5, respectively. The corresponding link matrices of the dissociative and associative models are shown by Eq. (4.16) & (4.17), respectively.

$$L_{diss} = \begin{array}{c} \\ T \\ Q \\ A \\ S \\ AS \\ AST \\ ASQ \end{array} \begin{array}{c} \begin{array}{cccc} T & AS & AST & ASQ \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 \\ 0 & -1 & -1 & -1 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{4.16}$$

$$L_{asso} = \begin{array}{c} \\ T \\ Q \\ A \\ S \\ AS \\ AST \\ ASQ \end{array} \begin{array}{c} \begin{array}{cccc} T & AS & AST & ASQ \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 \\ 0 & -1 & -1 & -1 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{4.17}$$

Steady State Analysis

For the dissociative model one asymptotically stable steady state was found based on the eigenvalues of the Jacobian matrix in Eq. (4.18).

$$J_{diss} = \begin{array}{c} \\ AS \\ AST \\ T \\ ASQ \\ Q \\ A \\ S \end{array} \begin{array}{c} \begin{array}{ccccccc} AS & AST & T & ASQ & Q & A & S \end{array} \\ \begin{pmatrix} \simeq 0 & 0 & 0.00028 & -0.000311 & 0 & \simeq 0 & 2.08\times10^{-5} \\ \simeq 0 & -10 & 0 & 0.000311 & 0 & 0 & 0 \\ 0 & 10 & -0.00028 & 0 & 0 & 0 & 0 \\ \simeq 0 & 0 & 0 & -0.000311 & 0 & 0 & 0 \\ 0 & 0 & 0.00028 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \simeq 0 & -2.08\times10^{-5} \\ 0 & 0 & 0 & 0 & 0 & \simeq 0 & -2.08\times10^{-5} \end{pmatrix} \end{array} \tag{4.18}$$

The eigenvalues of J_{diss} are identified as 0, 0, 0, -2.07536e-05, -0.00028, -0.000311351, and -10, respectively. For the dissociative model, all reaction rates at steady state are numerically zero. The species concentrations for $[AS]_{ss} = [S]_0$, $[Q]_{ss} = [T]_0$ and $[A]_{ss} = [A]_0 - [S]_0$, while $[T]_{ss}$ and all other steady-state concentrations are numerically zero.

$$J_{asso} = \begin{array}{c} \\ AS \\ AST \\ T \\ ASQ \\ Q \\ A \\ S \end{array} \begin{array}{c} \begin{array}{ccccccc} AS & AST & T & ASQ & Q & A & S \end{array} \\ \begin{pmatrix} -10.0 & \simeq 0 & 0.000115 & \simeq 0 & \simeq 0 & 0 & 0 \\ 0 & \simeq 0 & -8.86\times10^{-5} & 0 & 0 & 0 & 2.08\times10^{-5} \\ 0.000157 & \simeq 0 & -0.000114 & \simeq 0 & \simeq 0 & 0 & 0 \\ 10.0 & 0 & -2.60\times10^{-5} & \simeq 0 & \simeq 0 & 0 & 0 \\ -0.000157 & 0 & 2.60\times10^{-5} & \simeq 0 & \simeq 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2.08\times10^{-5} \\ 0 & 0 & 0 & 0 & 0 & 0 & -2.08\times10^{-5} \end{pmatrix} \end{array} \tag{4.19}$$

For the associative model, the eigenvalues of J_{asso} are identified as 0, 0, 0, 0, -2.08 $\times10^{-5}$, -0.000114 and -10, respectively. In contrast to the dissociative model, the only steady state found for the associative model is of underdetermined stability based on the eigenvalues of the Jacobian matrix in Eq. (4.19). At steady state, all reaction rates are numerically zero. Two relations between the three non-zero species concentrations can be identified: $[AS]_{ss} + [ASQ]_{ss} = [S]_0$, $[ASQ]_{ss} + [Q]_{ss} = [T]_0$ and $[A]_{ss} = [A]_0 - [S]_0$. $[T]_{ss}$ and all other steady-state concentrations are numerically zero.

The models share stoichiometries and mass conservation properties, while they differ in

their internal behaviour at steady state. However, most importantly, they both lead to a steady state with $[T]_{ss} = 0$. Next, it will be investigated how the apparently similar steady states are dynamically developed given physiological meaningful initial conditions. Furthermore, their behaviour during interaction with cellular processes like RNA metabolism and siRNA delivery will be of interest.

Time course simulations

The dissociative model directly translates into a system of ODEs with Eq. (4.20) to (4.26).

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = -k_{+2}[T][AS] \tag{4.20}$$
$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = -k_{+1}[S][A] \tag{4.21}$$
$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = -k_{+1}[S][A] \tag{4.22}$$
$$\frac{\mathrm{d}[AS]}{\mathrm{d}t} = k_{+1}[S][A] + k_{+4}[ASQ] - k_{+2}[T][AS] \tag{4.23}$$
$$\frac{\mathrm{d}[AST]}{\mathrm{d}t} = k_{+2}[AS][T] - k_{+3}[AST] \tag{4.24}$$
$$\frac{\mathrm{d}[ASQ]}{\mathrm{d}t} = k_{+3}[AST] - k_{+4}[ASQ] \tag{4.25}$$
$$\frac{\mathrm{d}[Q]}{\mathrm{d}t} = k_{+4}[ASQ] \tag{4.26}$$

It can be reduce to 5 ODEs and one further Eq. (4.32):

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = -k_{+2}[T][AS] \tag{4.27}$$
$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = -k_{+1}[S][A] \tag{4.28}$$
$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = -k_{+1}[S][A] \tag{4.29}$$
$$\frac{\mathrm{d}[AST]}{\mathrm{d}t} = k_{+2}[AS][T] - k_{+3}[AST] \tag{4.30}$$
$$\frac{\mathrm{d}[ASQ]}{\mathrm{d}t} = k_{+3}[AST] - k_{+4}[ASQ] \tag{4.31}$$
$$[AS] = [A]_0 - [A] - [AST] - [ASQ] \tag{4.32}$$

In the same fashion, the reduced system of ODEs for the associative model is defined by Eq. (4.33) to (4.38) with constraint Eq. (4.39).

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = k_{-4}[Q][AST] - k_{+4}[ASQ][T] - k_{+2}[T][AS] \tag{4.33}$$

$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = -k_{+1}[S][A] \tag{4.34}$$

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = -k_{+1}[S][A] \tag{4.35}$$

$$\frac{\mathrm{d}[AS]}{\mathrm{d}t} = k_{+1}[S][A] - k_{+2}[AS][T] \tag{4.36}$$

$$\frac{\mathrm{d}[ASQ]}{\mathrm{d}t} = k_{+3}[AST] + k_{-4}[AST][Q] - k_{+4}[ASQ][T] \tag{4.37}$$

$$\frac{\mathrm{d}[Q]}{\mathrm{d}t} = k_{+4}[ASQ][T] - k_{-4}[AST][Q] \tag{4.38}$$

$$[AST] = [A]_0 - [A] - [AST] - [ASQ] \tag{4.39}$$

On the long term, target mRNA concentration during time course simulations tends towards zero for both models, as predicted by steady state analysis, and expected because target mRNA synthesis and siRNA degradation was omitted at this stage of modelling. Nonetheless, time course simulations starting with physiological meaningful initial conditions, i.e. regarding concentrations for hAgo2 and target mRNA, should lead to similar results as experimental observations. As a first ballpark figure, 15 to 300 copies of siRNA per cell were sufficient in cell culture to knock down target levels by 50% within 24 h after siRNA administration[106]. That is, IC_{50} = 15 to 300 copies/cell, which in HeLa cells with an estimated cytoplasm volume of 1.60×10^{-12} l corresponds to cellular concentrations of 0.02 to 0.3 nM. In this first modelling approach, neither the dissociative, nor the associative model can reach this standard during time course simulations.

However, provided initial concentrations of 3,000 copies/cell ($\sim$ 3 nM) target and 300 copies/cell (0.3 nM) siRNA, target drops below 20% for the associative path 24 h after reaction initialisation, while target concentration does not decrease below the 85% mark within the same period of time assuming the dissociative path. Compare Fig. 4.2.

In a series of simulations, the initial target mRNA concentration is varied between 30 and 290,000 copies/cell (i.e. 0.03 and 300 nM). Initial concentrations of siRNA and Ago2 are 300 and 200,000 copies/cell (i.e. 0.3 and $\sim$ 200 nM), respectively. Trivially, RNAi efficacy decreases with higher target concentrations. Given the dissociative model, the difference in RNAi efficacy at different target concentrations is more distinct compared to the associative model. While the RNAi machinery of the associative model tends to work slightly less ef-

fectively for target concentration between 0.3 and 3 nM, it is more effective in the presence of higher target concentrations. This results in a certain robustness of the time course of the relative target knockdown with respect to varying absolute target concentrations. Compare Fig. 4.3.

To sum up, neither model can achieve physiologically observed IC_{50}. However, in the case of slightly higher siRNA concentrations, the associative model allows for an adaptation of RNAi machinery to varying target levels. Looking at the trajectories of modelled intermediate complexes formed during RNAi, the similarities and differences between the two models become visible: For both models, siRNA binding slows down the RNAi machinery during the first minutes. This results in slow reduction of siRNA over the whole time period (blue lines in Fig. 4.4). Furthermore, the two models do not differ within the first 60 s of the time course simulation (Fig. 4.4A). Most importantly, the complex comprising the cleavage product builds up equally in both models, soon becoming the most abundant of the intermediates. This is due to the rate limiting effect of the slow product release step within the models. However, the target concentration reduces faster in the associative model (broken black lines) compared to the dissociative model (solid black lines). The only other trajectories that largely differ between the two models are binary complex between Ago2 and siRNA (AS, green lines) and ternary complex between Ago2, siRNA and target (AST, brown lines). AS is increased in the case of the dissociative, and AST is increased in the case of the associative model. The reaction step that connects these two complexes is target binding. This indicates that this reaction step is accelerated in the associative model under the used reaction conditions (i.e. a relatively high target concentration of 3,000 copies/cell, $\sim$ 3 nM). A simple look at the kinetics and their parameters would not reveal this behaviour. It would instead invite choosing the large difference in product release (compare Table 4.1) as the major factor that distinguishes both models.

Validation for minimal *in vitro* RNAi assay

For deciding whether numerical results obtained from one or both models are in fact acceptable descriptions of the real world situation, two different sets of experimental time course data are available. The first validation set is based on a minimal *in vitro* RNAi assay comprising of recombinant expressed and purified human Argonaute 2 (hAgo2), 21 nt long ss guide RNA including two $3'$ deoxynucleotides (as2b) and $3'$ ^{32}P-labeled target RNA (a 140 nt long *in vitro* transcript of ICAM-1 mRNA)[32]. For the validation experiment 3.7 μM hAgo2 and 100 nM ss guide RNA were pre-incubated in buffer for 10 min. It can be expected that due to aggregation only a fraction of the present hAgo2 is actually active[32].

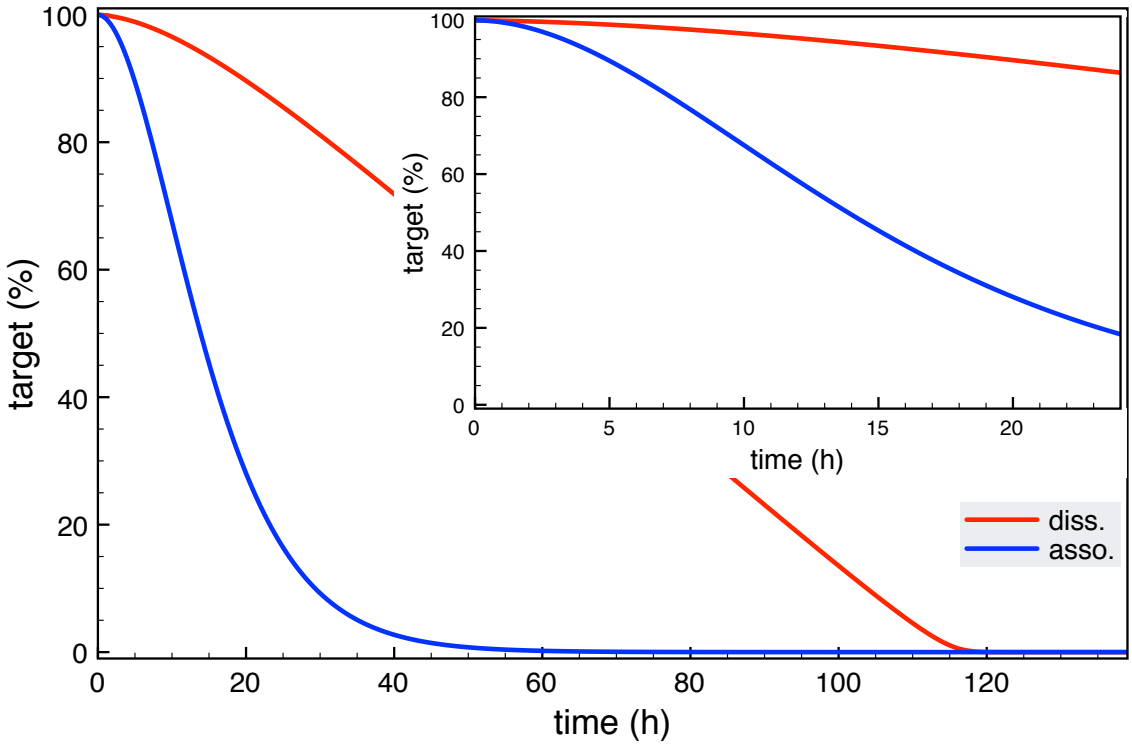
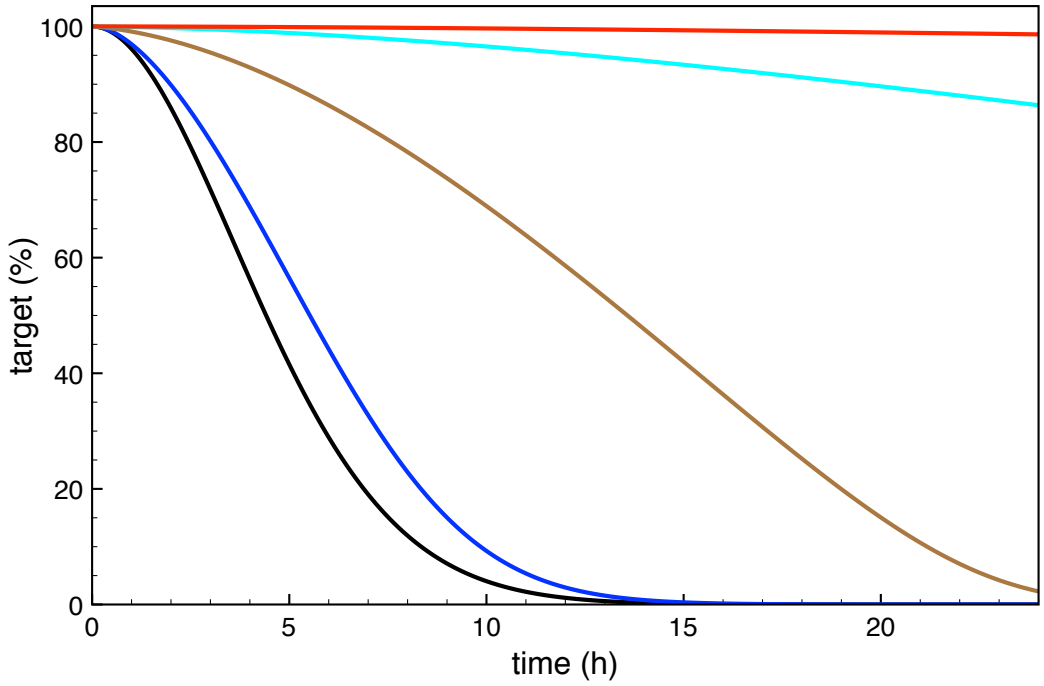
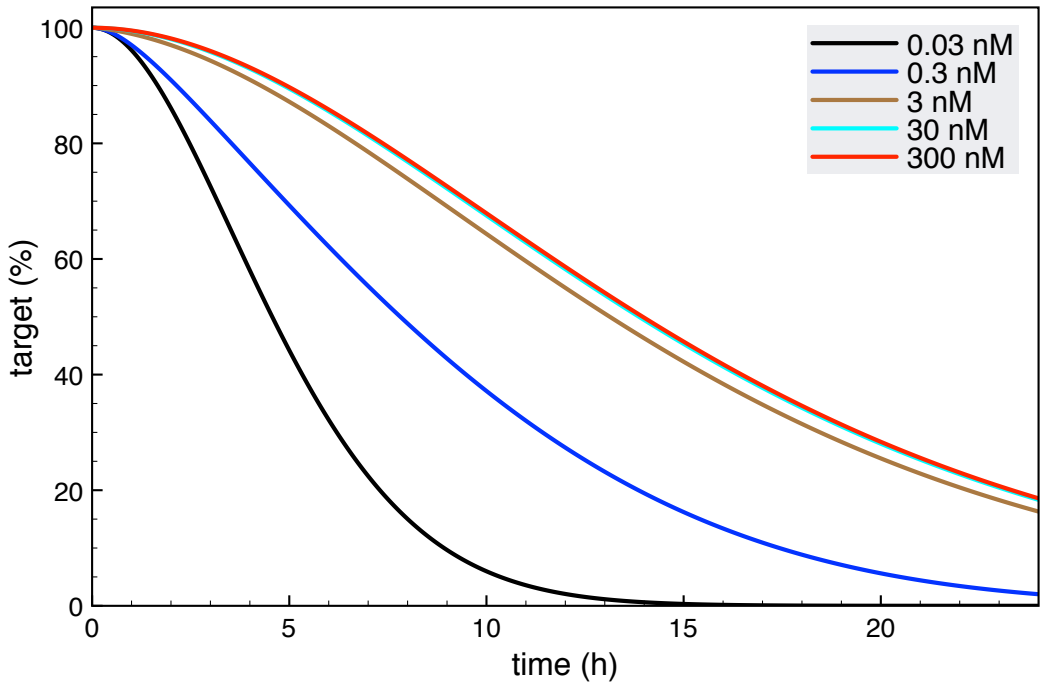


Figure 4.2 **Modelled time course of target mRNA turnover in the presence of ss siRNA.** Provided an initial target concentration of 3,000 copies/cell ($\sim$ 3 nM) and siRNA and Ago2 concentrations of 300 copies/cell (0.3 nM) and 200,000 copies/cell ($\sim$ 200 nM), respectively, target drops below 20% for the associative path 24 h after reaction initialisation. Target concentration does not decrease below 85% within the same period of time, assuming the dissociative path.

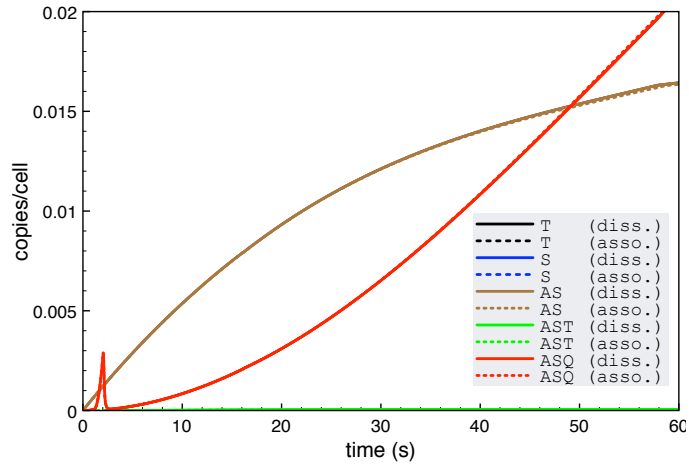


(a)

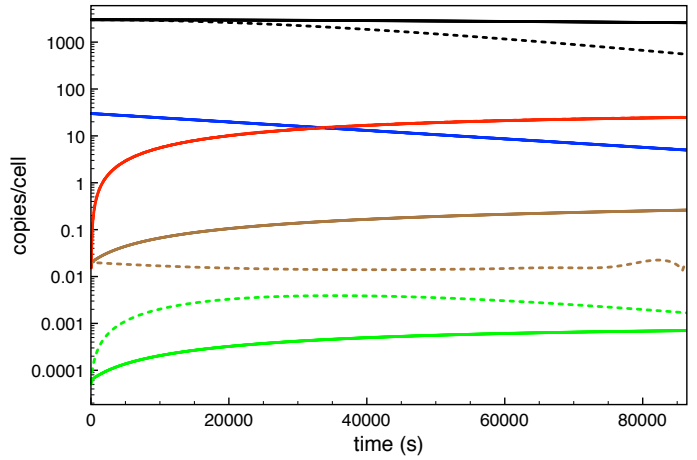


(b)

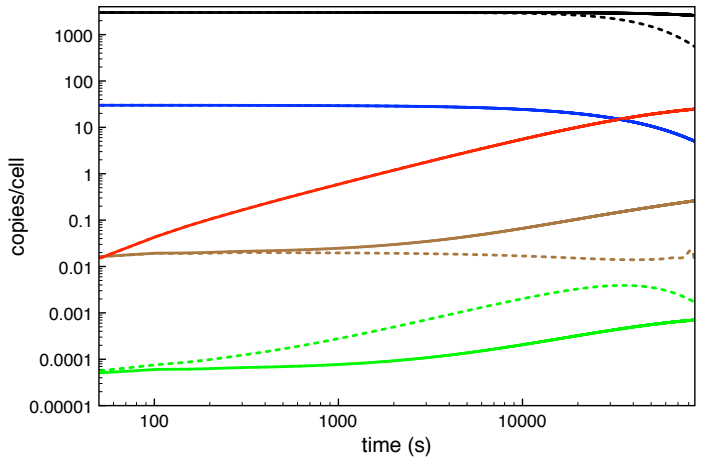
Figure 4.3 **The influence of target mRNA on time courses (of ss siRNA-mediated RNAi).** The initial target mRNA concentration is varied between 30 and 290,000 copies/cell (0.03 and $\sim$ 300 nM) at siRNA and Ago2 concentrations of 300 copies/cell (0.3 nM) and 200,000 copies/cell ($\sim$ 200 nM), respectively. Trivially, RNAi efficacy decreases with higher target concentrations. Given the dissociative model **(A)**, the difference in RNAi efficacy at different target concentrations is more distinct compared to the associative model **(B)**.



(a)



(b)



(c)

Figure 4.4 **Trajectories of intermediates (ss siRNA-mediated RNAi).** The trajectories of the modelled intermediate complexes are shown alongside with siRNA and target time courses. Simulations are started with initial target concentration of 3,000 copies/cell ($\sim$ 3 nM). Initial siRNA and Ago2 concentrations are 300 copies/cell (0.3 nM) and 200,000 copies/cell ($\sim$ 200 nM), respectively. **(A)** 60 s time course as linlog-plot; ASQ peak at 1 s is a numerical artefact; **(B)** 86,400 s (24 h) time course as linlog-plot; **(C)** same 86,400 s (24 h) time course as loglog-plot.

Eventually, the reaction was started by adding 2.5 nM target RNA. Target RNA abundance was monitored over time[32]. These concentrations are rather far away from physiological concentrations in human cells, compare Table 4.2. Given a cytoplasm volume of 1.6×10^{-12} l as reaction volume, 3700 nM (3.7 μM) hAgo2, 100 nM ss guide RNA, and 2.5 nM target RNA would correspond to approximately 3,500,000, 96,000, and 2,400 copies/cell, respectively.

However, an *in silico* model, other than a living cell, can be parameterised to any set of extreme conditions and at best should be able to provide meaningful, i.e. interpretable results. Thus, both mathematical models were parameterised to match the experimental conditions and their time course simulation results were compared to the corresponding experimental trajectory. Considering the simulated trajectories (black lines, Fig. 4.5), three things become obvious:

1. Both models perform similarly under the assumed *in vitro* conditions
2. Modelled knockdown has a slower onset compared to the data point at time 20 min
3. From there on, the modelled turnover is accelerated stronger than the experimental one, resulting in a complete turnover of target after only 50 min.

In contrast, the actual *in vitro* RNAi assay is not able to fully turnover all target RNA.

As described by Deerberg, Willkomm and Restle, the lack of complete turnover is due to progressing aggregation and inactivation of hAgo2[32]. It was estimated that less than 100 of the 3700 nM hAgo2 is active in the *in vitro* RNAi assay.

To account for the reduced compatibility between model and data, the models' initial hAgo2 concentration was fitted to the *in vitro* data. For both models 690 nM was the optimal concentration, resulting in both cases in a root mean square deviation (RMSD) of 7.9%. Again, both models lead to similar time course trajectories and a complete turnover of the target within the 2 h time course. Compare, blue lines, Fig. 4.5.

Considering estimations of the ratio between inactive and active Ago2[32], more realistic Ago2 concentrations between 1.5 nM and 100 nM have been tried for time course simulations with both models. Compare brown and cyan lines of Fig. 4.5, respectively. It becomes clear that neither of the two models get close to the experimental validation data under these conditions.

Two reasons were identified for the ongoing discrepancy between models and data: (i) the slower onset of target knockdown in the models indicates insufficient modelling of the siRNA binding kinetics. This will be addressed with changes in model structures in the

next Section 4.2, pp. 75. (ii) The complete turnover of target in simulations indicates an insufficient covering of the experimentally observed progressing aggregation and inactivation of hAgo2. The adaptation of the models to this aspect of the *in vitro* assay would require a new set of differential equations for which the kinetic parameters are unknown. Because this would not result in any information gain concerning the original model objectives and would require additional experimental data, this adaptation was abandoned. Another existing experimental data set was chosen instead to continue validation of the two models.

Validation for cytosol-extract-based RNAi assay

The second set of validation data is based on NHA-Ago2 assays in cytosolic HeLa cell extract (Carsten Geist, unpublished data), which was prepared according to[35, 140]. The cytosolic extract should contain all, or at least many, of the cytoplasm's components, although diluted by a factor $\frac{1}{3}$. It may be seen, at least to a limited extent, as a snap shot of the cytoplasm of living cells at the time of extraction. Thus it is a more realistic scenario for model validation compared to the previously described *in vitro* assay, as many known and unknown RNAi cofactors (as well as other influences) are likely to be present in extract.
The first data set (used as training set) consists of six sparse time series starting at different initial siRNA concentrations between 0.3 and 100 nM and constant concentrations of 2.4 nM target mRNA. The second set (the validation set) consists of a time series with 11 time points starting at 4 nM target and 100 nM siRNA.
Unfortunately, the initial Ago2 concentration in the assay is unknown. It is lower than *in vivo* due to dilution, extraction efficiency, and degradation during preparation and storage. It may vary between $\frac{1}{3}$ and $\frac{1}{10}$ of the original cellular concentration (communication with Juliane Neubert).
For model validation, time course simulations starting with the same initial conditions as present at the HeLa cytosol assay, should also lead to similar time courses as the experimental assay. However, as the Ago2 concentration in the assay remains vague, the initial Ago2 concentration is fitted to a training data set before agreement between models and experiments can be validated via the validation set. Table 4.3 shows the model fits to the training data with free parameter $[A]_0$ (initial Ago2 concentration). Table 4.4 shows the resulting best fit values for $[A]_0$. In this modelling approach, both the associative and the dissociative model best fit the set of experimental data of the HeLa cytosol-assay at an initial Ago2 concentration of 600 to 700 nM, which is four-fold the cellular hAgo2 concentration of 200 nM, i.e. 200,000 copies/cell. This is in contradiction to the observation that overall

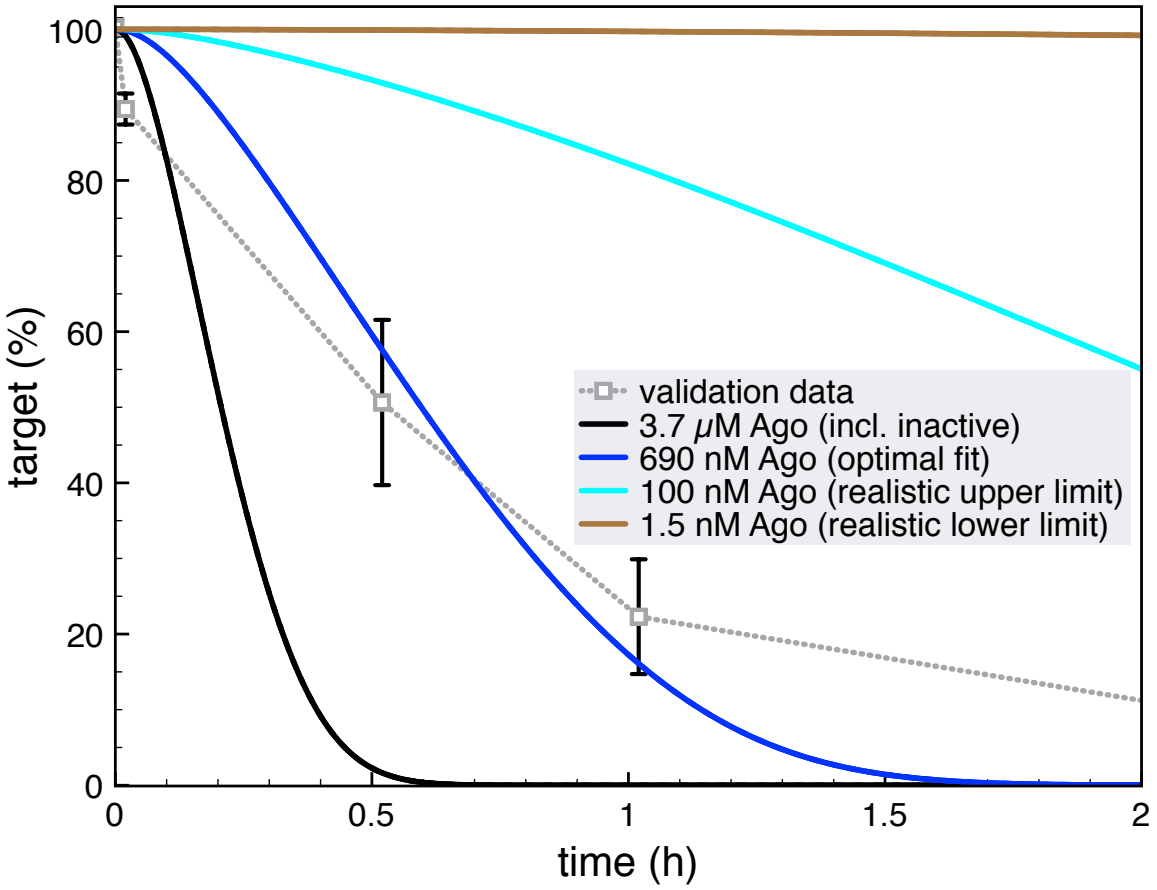


Figure 4.5 **Validation of the two competing minimal models of ss siRNA-mediated RNAi using *in vitro* assay.** The initial target concentration is 2.5 nM at an abundance of 100 nM siRNA. Both models result in the same trajectories under tested *in vitro* conditions. Reducing the assumed initial hAgo2 concentration from 3.7 μM (black lines, "incl. inactive") to 0.69μM (blue lines, "optimised fit") , the discrepancy between the models and the validation data can be reduced. This is in accordance with the aggregation and inactivation of hAgo2 observed in the RNAi assay. However, modelled knockdown has a slower onset compared to the original data point at time 20 min. Furthermore, the models do not account for the progression of the inactivation of hAgo2. Assuming 1.5 (cyan lines) to 100 nM (brown lines) active Ago2, none of the models achieves a trajectory close to the experimental data.

protein concentration in the extract is $\frac{1}{3}$ of the original cellular concentration. Furthermore, both models fit the data with an approximately similarly high objective value.

Because of the poor fits (high objective values and high RMSD), the validation is terminated at this point. The model fit to the validation data set is not shown, since the resemblance of the simulated time courses to the validation data was, as expected, poorer than that to the training data set. Swapping test and training data sets results in similar best fits and objective values (not shown).
To sum up: reoccurring issues with the experimental data are: (i) unknown concentrations of active hAgo2, for which a ballpark figure was estimated from the known physiological cellular concentration in HeLa cells and the protocol for the cytosolic extraction. (ii) Furthermore, the cytosol-extract-based RNAi assay (however, to a lesser degree than the *in vitro* assay) is affected by progressing inactivation of RNAi activity. Notwithstanding the lack of any target mRNA synthesis in extract, even high dosages of siRNA (up to 100 nM) do not result in complete turnover of target (even at low concentrations, e.g. 2.4 nM) in extract. There are different possible causes for this. Most likely siRNA is quickly degraded, while the more stable mRNA is degraded to a lesser degree. Furthermore, other essential components of the RNAi machinery, i.e. hAgo2 may be inactivated in the course of the assay.
Observed issues with the model fits are:

- There is a large variation between tuples of objective values and RMSD of the independent data sets.
- Overall objective values and RMSD are very high and simulations starting with the fitted parameter values neither fit the experimental validation data, nor even the training data.

Thus, the overall best fit for $[A]_0$ is an inconsistent compromise. Before fitting $[A]_0$, other issues with the model have to be resolved as described in the following section.

Systematic parameter estimation

As previous validation attempts using the literature-based parameter values (Tables 4.1 and 4.2) failed, a systematical parameter fit approach was launched. The objective was to mimic the available experimental data given the proposed model structures and physiological meaning full ranges for all model parameters. Model parameters comprise initial Ago2 concentration plus four or five reaction kinetic rate constants for the dissociative and associative model, respectively.

The fitting of only one parameter at a time does not require much repetition of the parameter estimation routine as most of the times the routine ends in similar results. Starting with only one free parameter, cytoplasm volume is added to the parameter space, since a change of reaction volume effects all other parameters simultaneously and thus can be a very effective way of model fitting.
The combination of k simultaneously fitted free parameters out of n parameters can be described by the binomial coefficient (n choose k). The more free parameters that are fitted simultaneously, the more likely the routine will end differently, i.e. get caught in a different local minimum. In this case, many repetitions have to be performed, ideally splitting the runs by using several different optimisation algorithms, like genetic algorithm, evolutionary programming, particle swarm, praxis etc. (see methods for details). To find the global minimum, one can agree on the one with the lowest objective value or lowest RMSD, or on a solution that occurs more often than all the others. In some cases there might be two or more solutions that occur with high frequency within the repetitions.
On average (depending on the algorithm used, the complexity of the optimisation problem, etc.) one routine runs for 285 s on a System with 2.66 GHz Intel Core 2 Duo, 8 GB 1,067 MHz DDR3 RAM and a steady state HDD. Due to combinatorial complexity, 58,300 routines wait to be executed, which would take 192.3 days using one machine. Since this time is not available, carefully selected scenarios are executed instead. Parameter estimation runs with exactly one free or with all free parameters were given first priority. These computations took 11 days. The remaining time of 10 days was shared between randomly chosen combinations. Compare Table 4.5.

The result of the multiple parameter estimation tasks can be summarised as follows:

1. In the case of the associative model, the combination of fitting the reaction rate constant of siRNA binding and either forward or backward reaction rate constant of associative target product exchange, fits the experimental data better than the previously shown fit with free initial Ago2 concentration. The objective value improves from 4822.31 to 3.13, while the RMSD improves from 35.59 to 0.95 (compare Tables 4.3 and 4.6).

2. Fitting the corresponding parameters of the dissociative model, i.e. the reaction rate constants of siRNA binding, of target binding, and of dissociative product release does improve the model fit compared to the fit with free initial Ago2 concentration. The objective value improves from 4980.9 to 12.61, while the RMSD improves from

41.76 to 1.24 (compare Tables 4.3 and 4.6).

3. The change of any other single parameter, tuple of, or triple of parameters cannot sufficiently fit either of the two models to the experimental data set. That is, in these cases the RMSD between simulated and experimental time courses is equal or higher than that during the previous fit with initial Ago2 concentration as single free parameter.

4. When fitting four or more free parameters at a time, the fit accuracy increases (e.g. lower RMSD than in 1 or 2). However, in these cases there is no single best parameter set, but there are infinitely many solutions, indicated by the large variation between the repetitions. Thus it is impossible to determine a meaningful average value for any parameter fit with four or more free parameters.

To sum up, given a sufficiently large number (≥ 4) of free model parameters, either of the models' structures are able to describe the experimental data. However, in these cases it is not possible to unambiguously determine an optimal parameter set. The best fit with an unambiguously determined set of optimal parameters is described in 1 and 2 for the associative and the dissociative model, respectively. Where the fit of the associative model is closer by a factor of four compared to the dissociative model.

Summary

At this stage of the modelling study, model parameters come from heterogenous sources. Most are literature-based. They do not provide comparable precision or their precision is inexplicable, with the result that parameters can only be seen as rough ballpark figures.
Given these literature parameters, the associative model resembles the validation data and the anticipated *in vivo* time course closer than its dissociative counterpart. Optimisation of model parameters can reduce the difference between experimental data and simulated data for both models. However, the optimised associative model still resembles the experimental data closer than the optimised dissociative model. The main difference between the two models is, that in the associative model, there are two different reactions for target binding, which both can be parameterised differently and follow different kinetic rules. That is, the target binding step participates in the first round of target turnover, while the associative target product exchange participates in multiple turnover and is coupled to the release of the cleavage product. Indeed, parameter estimation shows that different sets of reaction rate constants are preferred for these two reaction steps. Compare Table 4.7.
Model validation using experimental data based on time courses of *in vitro* or cytosol extract

RNAi respectively identify the parameters that most strongly influence the model outcome under the conditions during the validation experiments: kinetics of

- siRNA binding (both models)
- associative target product exchange (associative model)
- target binding (dissociative model)
- product release (dissociative model)

Target cleavage reaction kinetics (both models) and target binding kinetics (associative model) are comparable uninfluential on the simulation outcome.
The estimated values for forward reaction rate constants of associative target product exchange (k_{+5}) and target binding (k_{+2}) differ from literature values by a factor 2 and 3, respectively. Given the uncertainty that is associated with the literature values, the predicted values may be still considered close by.
Whereas, the difference between literature value and fitted value for dissociative product release (k_{+4} and k_{-4}) differs by a factor of 1.6×10^9, debunking the literature value as inapt for the reaction step. Compare Table 4.7.

initial concentrations		**target**		**diss. model**		**asso. model**	
target	**siRNA**	**at 1 h**	**at 2 h**	**objective**	**RMSD**	**objective**	**RMSD**
2.4 nM	0.3 nM	81.73%	60.43%	1818.78	26.57	1799.49	24.49
2.4 nM	1 nM	55.17%	24.67%	7034.5	49.58	7034.53	48.42
2.4 nM	3 nM	34.14%	10.80%	10946.01	90.24	10013.8	57.77
2.4 nM	10 nM	24.17%	6.59%	7714.7	51.13	7714.76	50.71
2.4 nM	30 nM	21.44%	5.18%	2314.53	27.1	2314.58	27.78
2.4 nM	100 nM	22.76%	4.90%	56.89	5.95	56.7	4.35
average				4980.9	41.76	4822.31	35.59

Table 4.3 **Model validation: objective values and RMSD of ss siRNA-mediated RNAi.** Shown are objective values and corresponding RMSD for each validation data set, as well as the resulting average values. Validation is based on 100 repetitions of model fits with free parameter $[A]_0$ (initial Ago2 concentration) to the training data based on cytosolic extract HeLa assays. Objective values and RMSD are similar between both models. There is a large variation between tuples of objective values and RMSD of the independent data sets.

model	lower bound (M)	upper bound (M)	start value	$[A]_0$ (M)	SD (M)
diss.	10^{-12}	10^{-4}	random	6.87×10^{-7}	1.92×10^{-7}
asso.				6.22×10^{-7}	1.43×10^{-7}

Table 4.4 **Model validation: fitted parameter (ss siRNA-mediated RNAi).** Validation is based on 100 repetitions of model fits with free parameter $[A]_0$ (initial Ago2 concentration) to the training data based on cytosolic extract HeLa assays. Shown are the resulting optimised values for free parameter $[A]_0$ (initial Ago2 concentration), as well as their standard deviation (SD).

n (free parameters)	model name	k (model parameters)	combinations	repetitions	calculations	computing time (s)	(h)	(d)
1	diss.	6	6	100	600	171,000	47.5	2.0
	asso.	7	7		700	199,500	55.4	2.3
2	diss.	5	10	500	5,000	1,425,000	395.8	16.5
	asso.	6	15		7,500	2,137,500	593.8	24.7
3	diss.	5	10	500	500	1,425,000	395.8	16.5
	asso.	6	20		10,000	2,850,000	791.7	33.0
4	diss.	5	4	500	2,000	5,700,000	1,583.3	66.0
	asso.	6	15		7,500	2,137,500	593.8	24.7
all	diss.	5	1	1000	1,000	285,000	79.2	3.3
	asso.	6	1		1,000	285,000	79.2	3.3
total time consumption					58,300	16,615,500	4,615.4	192.3

Table 4.5 **Systematic parameter estimation: combinatorics (ss siRNA-mediated RNAi).** As the number of free parameters increases, complexity and time consumption of the parameter estimation task increase exponentially.

initial concentrations		target		diss. model		asso. model	
target	**siRNA**	**at 1 h**	**at 2 h**	**objective**	**RMSD**	**objective**	**RMSD**
2.4 nM	0.3 nM	81.73%	60.43%	8.70	0.51	0.63	0.46
2.4 nM	1 nM	55.17%	24.67%	1.02	0.54	4.17	1.18
2.4 nM	3 nM	34.14%	10.80%	38.48	1.13	5.71	1.38
2.4 nM	10 nM	24.17%	6.59%	7.85	1.62	0.39	0.36
2.4 nM	30 nM	21.44%	5.18%	9.49	1.78	3.64	1.1
2.4 nM	100 nM	22.76%	4.90%	10.13	1.89	4.21	1.24
average				12.61	1.24	3.13	0.95

Table 4.6 **Parameter estimation results for the two competing models of ss siRNA-mediated RNAi.** The dissociative model is fitted with the three free parameters siRNA binding (k_{+1}) , target binding (k_{+2}) and dissociative release (k_{+4}). Their best fit values are 3.4×10^5 ($\pm 2.5 \times 10^4$) $M^{-1}s^{-1}$, 4.5×10^5 ($\pm 1.3 \times 10^4$) $M^{-1}s^{-1}$, and 4.8×10^4 ($\pm 2.5 \times 10^7$) $M^{-1}s^{-1}$, respectively. The associative model is fitted with two free parameters: siRNA binding (k_{+1}) and associative target binging (k_{+5}). Their best fit values are 3.3×10^5 ($\pm 1.4 \times 10^4$) $M^{-1}s^{-1}$ and a 3.5×10^5 ($\pm 2.2 \times 10^4$) $M^{-1}s^{-1}$, respectively. Values in brackets are standard deviation.

model	reaction	constant	units	lit.	diss.	asso.
both	siRNA binding	k_{+1}	$M^{-1}s^{-1}$	10	3.4×10^5 ($\pm 2.5 \times 10^4$)	3.3×10^5 ($\pm 1.4 \times 10^4$)
		k_{-1}	s^{-1}	$\simeq 0$	†	†
both	target binding	k_{+2}	$M^{-1}s^{-1}$	1×10^6	4.5×10^5 ($\pm 1.3 \times 10^4$)	
		k_{-2}	s^{-1}	$\simeq 0$	†	†
both	target cleavage	k_{+3}	s^{-1}	10	‡	‡
		k_{-3}	s^{-1}	$\simeq 0$	†	†
disso	dissociative product release	k_{+4}	s^{-1}	3×10^{-5}	4.8×10^4 ($\pm 2.5 \times 10^7$)	
		k_{-4}	$M^{-1}s^{-1}$	$\simeq 0$	†	-
asso	associative target product exchange	k_{+5}	$M^{-1}s^{-1}$	1.17×10^5	-	3.5×10^5 ($\pm 2.2 \times 10^5$)
		k_{-5}	$M^{-1}s^{-1}$	5.07×10^5	-	$*$

Table 4.7 **Optimised kinetic parameters of the minimal models of ss siRNA-mediated RNAi.** †: reaction modelled as quasi-irreversible, ‡: uninfluential on the simulation outcome, $*$: k_{-5} is kept constant, as it correlates with k_{+5}.

4.2 Extension towards ds siRNA and accounting for target siRNA affinity modes

Accounting for the results of the previously described and analysed minimalistic model approach, the model structures were revised aiming at better accounting for the RNAi process *in vivo* and a more homogenous parametrisation.
Firstly, the kinetics of siRNA binding, i.e. the absence of precise reaction rate constants, as modelled in the previous section, were identified as a reason for an insufficient adaptation of the onset of the simulation trajectories to the time courses of the validation data sets. Compare also, pp. 61.
Secondly, two distinct active RISC (binary complex between Ago2 and siRNA) are introduced for single and multiple turnover with distinct target siRNA affinities. This addresses the poorly adaptability of the models in Section 4.1, described at pp. 66.
The simultaneous introduction of these two new concepts bears the advantage of becoming independent of a heterogenous set of model parameters (compare Table 4.1). Instead, model structure and the kinetics of complex formation can be fully implemented accounting for only one source: Wünsche and Sczakiel (2005)[226]. This allows parametrisation of a specific experimental set-up and adds homogeneity to the models. That is, RNAi-based knockdown of ICAM^{-1} mRNA in human ECV305-cells in the presence of highly effective (in terms of target knockdown efficiency) siRNA si2b, as well as ineffective siRNA si1. This will be further described in Section 4.2.3, pp. 79.

The modified model structures can be described as follows:

4.2.1 Dissociative model: a basic model of ds siRNA-mediated RNAi

In a first step, ds siRNA (GP) binds to Ago2 (A). The siRNA consists of a guide (G) and a complementary passenger (P) stand. In the next step the passenger strand is replaced by target mRNA (T). The details of this step are omitted for simplicity. But it can be assumed, that after cleavage of the passenger strand, the hydrogen bonds between cleavage product and guide strand are replaced by hydrogen bonds between the incoming target strand with the bound guide strand. In contrast to mRNA, the short passenger strand is vulnerable for degradation by cellular nucleases. None-the-less, the passenger-target strand exchange reaction is reversible[226]. Thus, the resulting catalytically active complex between guide, target and Ago2 after the described first-pass reaction is dubbed low affinity RISC (AGT) to

distinguish it from a second catalytically active complex which occurs further downstream during the multiple turnover cycle of the biochemical network. The target mRNA is cleaved by Ago2's endonuclease activity (target cleavag) and, subsequently, the remaining cleavage product (Q) is released from the complex (AGQ, product release). Thus the target mRNA population is reduced by one and the resulting product strands are prone to degradation by cellular nucleases, leaving behind a binary complex between guide and Ago2 (AG). The forward reaction of product release, the product's dissociation, is the slowest step in the dissociative model. However, once the binary complex is formed, a new target can bind (target binding) to form a ternary complex between guide, target and Ago2, the so-called high affinity RISC ($AGT^{\dagger}$). All subsequent passes occur via this high affinity RISC. A schematic depiction reaction pathway of the dissociative model is provided in Fig. 4.6A. Target mRNA synthesis and non-RNAi-based degradation are neglected at this stage of modelling.
The reactions for the dissociative model translate in the following rate Eq. (4.40) to (4.46), using the same symbols as in the text: Molecular entities are abbreviated as: A (Ago2), T (target RNA), G (siRNA's guide strand), P (siRNA's passenger strand), Q (cleavage product). To distinguish between different complexes composed of the same entities daggers are used, i.e. AGT versus $AGT^{\dagger}$ for low-affinity and high affinity RISC, respectively.

$$AGP + T \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} AGT + P \tag{4.40}$$

$$AGT \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} AGQ \tag{4.41}$$

$$AGQ \underset{k_{-3}}{\overset{k_{+3}}{\rightleftharpoons}} AG + Q \tag{4.42}$$

$$AG + T \underset{k_{-4}}{\overset{k_{+4}}{\rightleftharpoons}} AGT^{\dagger} \tag{4.43}$$

$$AGT^{\dagger} \underset{k_{-5}}{\overset{k_{+5}}{\rightleftharpoons}} AGQ \tag{4.44}$$

$$Q \underset{k_{-6}}{\overset{k_{+6}}{\rightleftharpoons}} Q^{\ddagger} \tag{4.45}$$

$$P \underset{k_{-7}}{\overset{k_{+7}}{\rightleftharpoons}} P^{\ddagger} \tag{4.46}$$

4.2.2 Associative model: accelerated product release via strand exchange between RISC-bound product and an associating target RNA

The first pass cycle is identical between dissociative and associative models: siRNA binding is followed by passenger-target strand exchange reactions, which results in a low affinity RISC (AGT). The bound target can be sliced in the target cleavage reaction.
However, when entering the cycle of multiple turnover, the associative model differs: the complex between Ago2, guide and cleaved product (AGQ) binds the next target, before the cleaved mRNA fragments (product, Q) are released from complex. This new target replaces the cleavage product by invading the hydrogen bonds between siRNA and the product. The resulting catalytically active complex ($AGT^{\dagger}$) is characterised by a high affinity between target and guide strand, analogous to the dissociative path. All subsequent passes, occur via this high affinity RISC. The associative mechanism allows the RNAi machinery to undergo multiple rounds of target turnover, without the bottle neck of a slow product release step. A schematic depiction reaction pathway of the associative model is provided in Fig. 4.6B.
In the associative model Eq. (4.43) and (4.44) are replaced by Eq. (4.47).

$$\mathrm{AGP} + \mathrm{T} \underset{\mathrm{k}_{-8}}{\overset{\mathrm{k}_{+8}}{\rightleftharpoons}} \mathrm{AGT} + \mathrm{P} \tag{4.47}$$

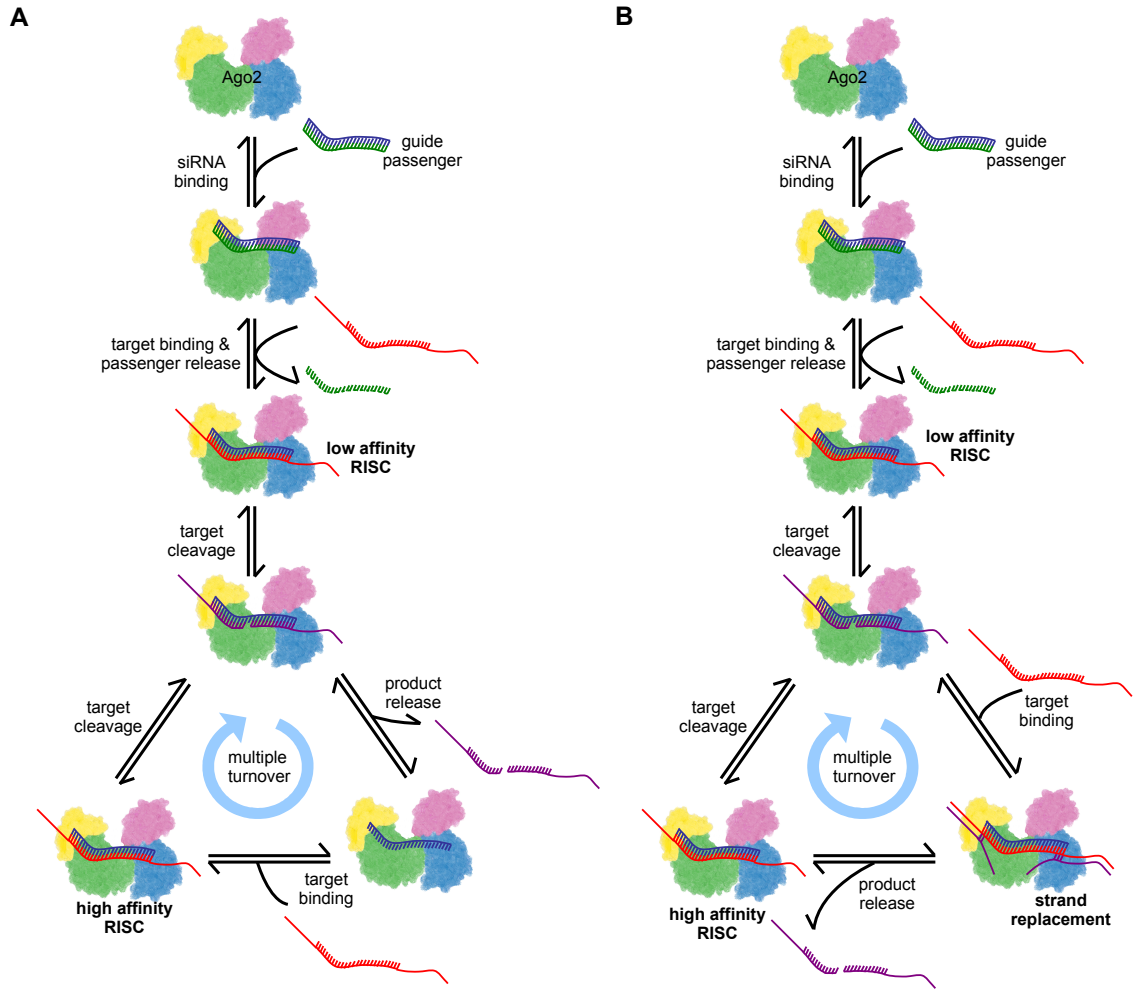


Figure 4.6 **Illustrations of the dissociative and associative models accounting for ds siRNA and target-RISC affinity.** Ago2 is schematically represented based on its electron density cloud. RNA strands are indicated by colour-coded lines. Description continues on the next page.

Fig. 4.6, **continued. Illustrations of the dissociative and associative models accounting for ds siRNA and target-RISC affinity.**
A) Dissociative model. In a first step, ds siRNA binds to Ago2. Ds siRNA consists of a guide and a complementary passenger stand. In the next step the passenger strand is replaced by target mRNA. The details of this step are omitted for simplicity. The resulting catalytically active complex consists of Ago2, a guide and a target mRNA. It is dubbed low affinity RISC to distinguish it from a second catalytically active complex, which occurs further downstream during the multiple turnover cycle of the biochemical network. That is, after the bound target is cleaved and released from RISC, a new free target enters the complex. This catalytically active RISC is characterised by a higher affinity between target and guide strand, allowing the system to undergo multiple rounds of target binding and cleavage, followed by a slower product release step, without falling apart. **B) Associative model.** During the first turnover cycle, the associative model follows the same path as its dissociative counterpart. However, when entering the cycle of multiple turnover, the complex between Ago2, guide and cleaved mRNA binds a next target before the cleaved mRNA fragments (dubbed product) are released from the complex. The new target replaces the cleavage product by invading the hydrogen bonds between siRNA and the product. The resulting catalytically active complex is characterised by a higher affinity between target and guide strand, analogous to the dissociative path. The associative mechanism allows the RNAi machinery to undergo multiple rounds of target turnover, without the bottle neck of a slow product release step.

4.2.3 Model parameterisation

The parameterisation of the modelling of ds siRNA-mediated RNAi considers the relationship between biological effectiveness of siRNA and strand exchange kinetics between siRNA and the target. Both, the dissociative and the associative models were parameterised for RNAi-based knockdown of ICAM^{-1} mRNA in human ECV305-cells in the presence of highly effective siRNA si2b, as well as ineffective siRNA si1 (see Table 4.8 for parameters). Highly effective and inactive siRNA differ from each other in their target knockdown efficiency, which can be for instance measured as IC_{50} value. The apparent IC_{50} value is the siRNA dose needed for 50% effect, which is per definition the reduction of the target mRNA level at 24 h after dose. In this regard, the biologically highly effective si2b has a low IC_{50} value of 0.24 nM compared to the ineffective si1 with an IC_{50} value above 500 nM, targeting ICAM^{-1} mRNA in ECV cells after being transfected via Lipofectamine 2000. The original

model	reaction step	constant	units	si2b	si1
both	passenger-target strand exchange	k_{+1}	$M^{-1}s^{-1}$	5.07e+03	1.75e+01
		k_{-1}	$M^{-1}s^{-1}$	1.17e+05	1.58e+08
both	target cleavage	k_{+2}	s^{-1}	1.00e+01	1.00e+01
disso	target binding	k_{+3}	$M^{-1}s^{-1}$	5.07e+03	5.07e+03
disso	product release	k_{+4}	s^{-1}	1.00e-05	1.00e-05
asso	product-target strand exchange	k_{+5}	$M^{-1}s^{-1}$	5.07e+03	5.07e+03
both	product degradation	k_{+6}	s^{-1}	1.00e+02	1.00e+02
both	passenger degradation	k_{+7}	s^{-1}	1.00e+01	1.00e+01

Table 4.8 **Kinetic parameters of two competing models of ds siRNA-mediated RNAi.** The two models share many of their parameters, indicated with *both* in column one, the tags *asso* and *disso* indicate a parameter specific for the associative or dissociative model, respectively. Where no value for a backward reaction rate is given, the reaction is modelled as quasi irreversible. If constants are siRNA-specific, a value is provided for a biologically active, e.g. si2b, and a inactive, e.g. si1, siRNA.

experiment is described in Kretschmer-Kazemi Far and Sczakiel (2003)[100]. Wünsche and Sczakiel (2005) showed that the kinetics of the strand exchange between double stranded siRNA and target mRNA differs whether biologically highly effective siRNA or a less effective one is involved in the reaction[226]. Strand exchange was measured in the presence of Cetyltrimethylammoniumbromid (CTAB) to mimic the degree of freedom of RNAs in living cells. The observed equilibrium constant was 400-fold more favourable for si2b compared to si1[226]. See also reaction rate constant pair k_{+1} and k_{-1} in Table 4.8. This connection between biological effectiveness of an siRNA (IC_{50} values) and the strand exchange kinetics between siRNA and its target RNA (k_+ / k_- reaction rate constant ratio) should be reproducible by a valid model. Furthermore, the cytoplasm volume of an average ECV cell is set to 9.0×10^{-13} l, compare Section 4.1.1, pp. 54.

4.2.4 Analyses

Scrutinising the models: Stoichiometric Analysis

Elementary flux modes could be identified in either of the two proposed models of ds siRNA-mediated RNAi. This implies, that neither the dissociative nor the associative model comprise smaller sub-networks that allow a metabolic reconstruction network to function in steady state.

Mass conservation analyses of dependent species starting with different combinations of

initial species concentrations led to the identification of one mass flow in each of the two models: Eq. (4.48) in the case of the dissociative and Eq. (4.49) in the case of the associative model. It can be observed that mass conservation holds true at all times for both models. That is, mass is neither created nor destroyed, although it is rearranged in space, and the entities associated with it (e.g. target mRNA) are changed in form.

$$AGP_0 = AG + AGP + AGQ + AGT + AGT^{\dagger} \tag{4.48}$$

$$AGP_0 = AGP + AGQ + AGT + AGT^{\dagger} \tag{4.49}$$

$$N_{diss} = \begin{array}{c} \\ T \\ Q \\ P \\ AG \\ AGP \\ AGT \\ AGT^{\dagger} \\ AGQ \end{array} \begin{array}{c} \begin{array}{ccccccc} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{4.50}$$

Reactions passenger-target strand exchange, target cleavage (single turnover), target release, dissociative target binding, target cleavage (multiple turnover), cleavage product degradation, and passenger degradation are abbreviated as r_1, r_2, r_3, r_4, r_5, r_6, and r_7, respectively.

$$N_{asso} = \begin{array}{c} \\ T \\ Q \\ P \\ AGP \\ AGT \\ AGT^{\dagger} \\ AGQ \end{array} \begin{array}{c} \begin{array}{cccccc} r_1 & r_2 & r_5 & r_6 & r_7 & r_8 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & -1 \end{pmatrix} \end{array} \tag{4.51}$$

Reactions passenger-target strand exchange, target cleavage (single turnover), product-target strand exchange, target cleavage (multiple turnover), cleavage product degradation, and passenger degradation are abbreviated as r_1, r_2, r_5, r_6, r_7, and r_8, respectively. The corresponding reduced link matrices of the dissociative and associative models are shown by Eq. (4.52)

& (4.53), respectively.

$$L_{diss} = \begin{array}{c} \\ T \\ Q \\ P \\ AG \\ AGP \\ AGT \\ AGT^{\dagger} \\ AGQ \end{array} \begin{array}{c} \begin{array}{ccccccc} T & Q & P & AG & AGT & AGT^{\dagger} & AGQ \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{4.52}$$

$$L_{asso} = \begin{array}{c} \\ T \\ Q \\ P \\ AGP \\ AGT \\ AGT^{\dagger} \\ AGQ \end{array} \begin{array}{c} \begin{array}{cccccc} T & Q & P & AGT & AGT^{\dagger} & AGQ \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{4.53}$$

Steady State Analysis

For the dissociative model one asymptotically stable steady state was found based on the eigenvalues of the Jacobian matrix in Eq. (4.54).

$$J_{diss} = \begin{array}{c} \\ AGP \\ T \\ AGT \\ Q \\ P \\ AGQ \\ AG \\ AGT^{\dagger} \end{array} \begin{array}{c} \begin{array}{cccccccc} AGP & T & AGT & Q & P & AGQ & AG & AGT^{\dagger} \end{array} \\ \begin{pmatrix} -10^{-5} & 0 & 0 & 0 & 100 & 0 & 100 & 0 \\ 0 & -10^{-6} & 0 & 0 & \simeq 0 & \simeq 0 & 0 & \simeq 0 \\ 10^{-5} & 0 & -100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10^{-6} & 0 & -10 & \simeq 0 & 0 & 0 & \simeq 0 \\ 0 & 10^{-6} & 0 & 0 & -100 & 0 & 0 & \simeq 0 \\ -10^{-5} & \simeq 0 & 0 & 0 & 0 & \simeq 0 & 0 & 0 \\ 0 & \simeq 0 & 0 & 0 & 0 & \simeq 0 & -100 & 0 \\ 0 & -10^{-6} & 0 & 0 & \simeq 0 & 0 & 0 & \simeq 0 \end{pmatrix} \end{array} \tag{4.54}$$

The eigenvalues of J_{diss} are identified as 0, 0, -1.60×10^{-6}, -1.00×10^{-6}, -10, -100, -100, and -100, respectively. For the dissociative model, all reaction rates at steady state are numerically zero.

$$J_{asso} = \begin{array}{c} \\ AGP \\ T \\ AGT \\ Q \\ AGT^{\dagger} \\ P \\ AGQ \end{array} \begin{array}{c} \begin{array}{ccccccc} AGP & T & AGT & Q & AGT^{\dagger} & P & AGQ \end{array} \\ \begin{pmatrix} \simeq 0 & 0 & 0 & 100 & \simeq 0 & 100 & 0 \\ 0 & -10 & 0 & \simeq 0 & 1.60 \times 10^{-6} & 0 & \simeq 0 \\ \simeq 0 & 0 & -100 & 0 & \simeq 0 & 0 & 0 \\ 0 & 0 & 0 & -100 & 1.60 \times 10^{-6} & 0 & \simeq 0 \\ \simeq 0 & 0 & 0 & \simeq 0 & -1.60 \times 10^{-6} & 0 & \simeq 0 \\ \simeq 0 & 0 & 0 & 0 & \simeq 0 & -100 & 0 \\ 0 & 0 & 0 & \simeq 0 & -1.60 \times 10^{-6} & 0 & \simeq 0 \end{pmatrix} \end{array} \tag{4.55}$$

The eigenvalues of J_{asso} are identified as 0, 0, -1.60×10^{-6}, -10, -100, -100, and -100, respectively. In contrast to the dissociative model, the only steady state found for the associative model is of undetermined stability based on the eigenvalues of the Jacobian matrix in Eq. (4.55). At steady state all reaction rates are numerically zero.

To sum up, looking at stoichiometries and mass conservation properties, the two models differ in their internal behaviour at steady state. However, most importantly, they both lead to a steady state with $[T]_{ss} = 0$.

Next, it will be investigated how the apparently similar steady states are dynamically developed given physiological meaningful initial conditions.

Time course simulations

The models' networks of biochemical rate equations are mathematically described by a set of ordinary differential equations (ODEs). To examine the concentration changes of reactants, intermediates and product over time the systems of ODEs were numerically integrated using using the LSODA[163] algorithm under COPASI[80].

The rate equations of the dissociative model, i.e. Eq. (4.40) to (4.45), were transformed into

the system of ODEs defined in Eq. (4.56) to (4.62) with the constraint in Eq. (4.63).

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = k_{-1}[P][AGT] - k_{+1}[T][AGP] - k_{+1}[T][AG] \tag{4.56}$$
$$\frac{\mathrm{d}[P]}{\mathrm{d}t} = k_{+1}[T][AGP] - k_{-1}[P][AGT] - k_{+6}[P] \tag{4.57}$$
$$\frac{\mathrm{d}[AGP]}{\mathrm{d}t} = k_{-1}[P][AGT] - k_{+1}[T][AGP] \tag{4.58}$$
$$\frac{\mathrm{d}[AGT]}{\mathrm{d}t} = k_{+1}[T][AGP] - k_{-1}[P][AGT] - k_{+2}[AGT] \tag{4.59}$$
$$\frac{\mathrm{d}[AGQ]}{\mathrm{d}t} = k_{+2}[AGT] + k_{+2}[AGT]^{\dagger} - k_{+3}[AGQ] \tag{4.60}$$
$$\frac{\mathrm{d}[AG]}{\mathrm{d}t} = k_{+3}[AGQ] - k_{+1}[T][AG] \tag{4.61}$$
$$\frac{\mathrm{d}[AGT]^{\dagger}}{\mathrm{d}t} = k_{+1}[T][AG] - k_{+2}[AGT]^{\dagger} \tag{4.62}$$
$$[AGP]_0 = [AGP] + [AGT] + [AGQ] + [AGT]^{\dagger} + [AG] \tag{4.63}$$

In the same way, the system of ODEs for the associative model is defined as Eq. (4.64) to (4.70) with constraint in Eq. (4.71).

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = k_{-1}[P][AGT] - k_{+1}[T][AGP] - k_{+1}[T][AG] \tag{4.64}$$
$$\frac{\mathrm{d}[P]}{\mathrm{d}t} = k_{+1}[T][AGP] - k_{-1}[P][AGT] - k_{+6}[P] \tag{4.65}$$
$$\frac{\mathrm{d}[AGP]}{\mathrm{d}t} = k_{-1}[P][AGT] - k_{+1}[T][AGP] \tag{4.66}$$
$$\frac{\mathrm{d}[AGT]}{\mathrm{d}t} = k_{+1}[T][AGP] - k_{-1}[P][AGT] - k_{+2}[AGT] \tag{4.67}$$
$$\frac{\mathrm{d}[AGQ]}{\mathrm{d}t} = k_{+2}[AGT] + k_{+2}[AGT]^{\dagger} - k_{+3}[AGQ] \tag{4.68}$$
$$\frac{\mathrm{d}[AG]}{\mathrm{d}t} = k_{+3}[AGQ] - k_{+1}[T][AG] \tag{4.69}$$
$$\frac{\mathrm{d}[AGT]^{\dagger}}{\mathrm{d}t} = k_{+1}[T][AG] - k_{+2}[AGT]^{\dagger} \tag{4.70}$$
$$[AGP]_0 = [AGP] + [AGT] + [AGQ] + [AGT]^{\dagger} + [AG] \tag{4.71}$$

As predicted by steady state analysis and in accordance with the omission of target mRNA synthesis and siRNA degradation, on the long term, target mRNA concentration during time course simulations tends towards zero for both models.

Nonetheless, time course simulations starting with physiological meaningful initial con-

ditions, i.e. concentrations for hAgo2 and target mRNA, should lead to similar results as experimentally observed. As a first ballpark figure, 15 to 300 copies of siRNA per cell were sufficient in cell culture to knock down target levels by 50% within 24 h after siRNA administration[106]. In ECV cells with an estimated cytoplasm volume of 9.00×10^{-13}, this corresponds to cellular concentrations of 0.03 to 0.6 nM.

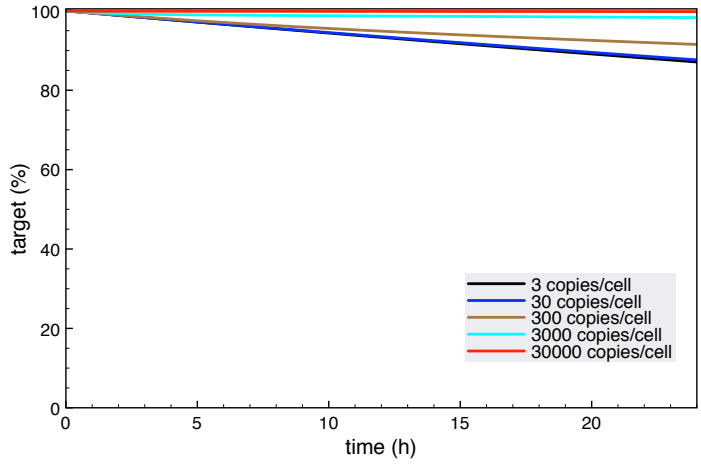
Neither the dissociative, nor the associative model of ds siRNA-mediated RNAi can reach 50% knockdown, given ≤ 30 pre-incubated siRNA-Ago2 complexes, during time course simulations of varying target mRNA concentrations between 3 and $30,000$ copies/cell, compare Fig. 4.7.

Providing slightly higher concentrations of pre-incubated siRNA-Ago2 complexes (100 to 300 copies/cell, i.e. 0.1 to 0.3 nM), target populations as high as $30,000$ copies/cell (300 nM) are reduced by 80% for the associative path 24 h after reaction initialisation. This could not be reached using the model with dissociative reaction path, where target concentration was not reduced by 15% within the same period of time. Compare Fig. 4.8.

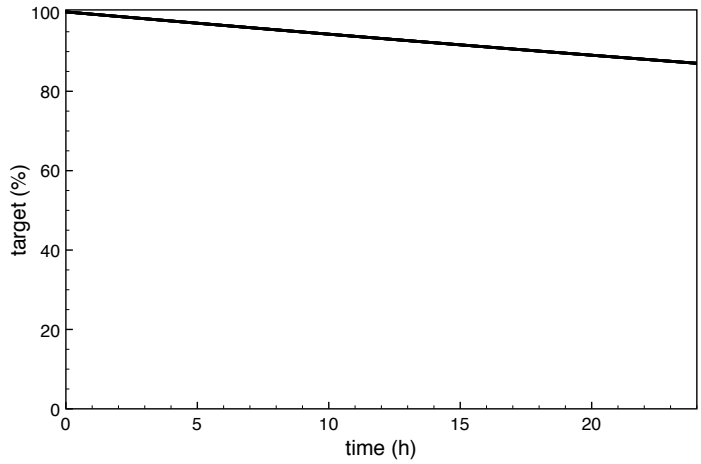
During the parameter scan task, the initial target mRNA concentration is varied between 300 and $30,000$ copies/cell (0.3 and 300 nM), while the concentration of pre-incubated siRNA-Ago2 complex varies between 30 and 300 (0.03 and 0.3 nM). Trivially, RNAi efficacy decreases with higher target concentrations and increases with higher concentrations of pre-incubated siRNA-Ago2 complex. Given the dissociative model, physiological meaningful concentrations of pre-incubated siRNA-Ago2 complex are not sufficient to knock down target concentration below 50% of its basal level for basal target levels of $3,000$ or $30,000$ copies/cell (Fig. 4.8A and B). Further reduction of the absolute level of basal target to 300 copies/cell leads to an efficient target turnover of 50% within 24 h (Fig. 4.8C). 300 copies/cell is within a physiological meaningful range for human mRNA. It is important to mention that target synthesis was not considered at this point of modelling. The associative path allows for efficient target turnover even for high target concentrations up to $30,000$ copies/cell, given physiological meaningful concentrations of pre-incubated siRNA-Ago2 complex (Fig. 4.8D).

Validation via *in vitro* assay

As for the previous modelling approach (Section 4.1.2, pp. 61), the validation data set based on a minimal *in vitro* RNAi assay is used. To sum up, 3.7 μM hAgo2 and 100 nM ss guide RNA were pre-incubated in buffer for 10 min. It can be expected that due to aggregation only a fraction of the present hAgo2 is actually active[32]. Eventually, the reaction was started by adding 2.5 nM target RNA and target RNA abundance was monitored over time[32].

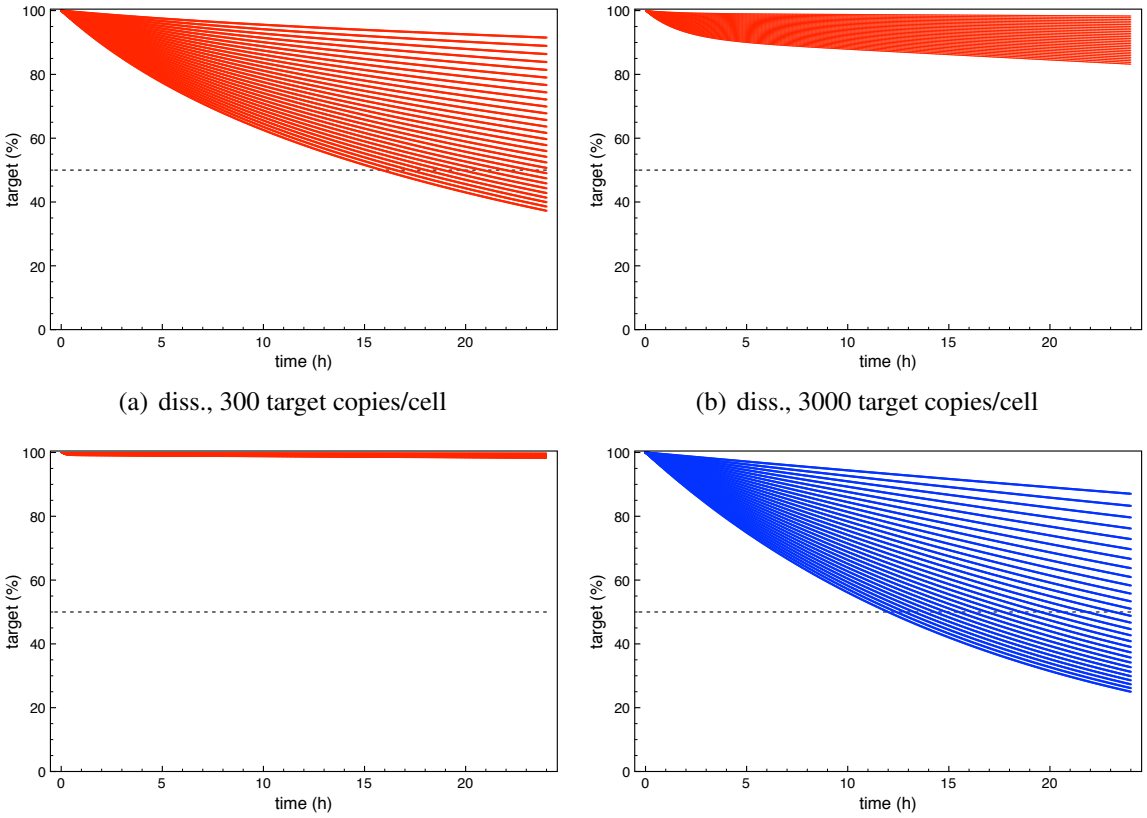


(a)



(b)

Figure 4.7 **The influence of target mRNA on time courses of ds siRNA-mediated RNAi.** The initial target mRNA concentration is varied between 3 and 30,000 copies/cell (0.03 and 300 nM) at pre-incubated siRNA-Ago2 complex concentrations of 30 copies/cell (0.3 nM). RNAi efficacy is low for both models. Compared to the dissociative model **(A)**, the variance in RNAi efficacy at different target concentrations is relatively constant for the associative model **(B)**.



(a) diss., 300 target copies/cell

(b) diss., 3000 target copies/cell

(c) diss., 30000 target copies/cell

(d) asso., 30000 target copies/cell

Figure 4.8 **Parameter scan of time course simulations (ds siRNA-mediated RNAi).** The initial target mRNA concentration is varied between 300 and 30,000 copies/cell (0.3 and 300 nM), while the concentration of pre-incubated siRNA-Ago2 complex is varied between 30 and 300 (0.03 and 0.3 nM). Trivially, RNAi efficacy decreases with higher target concentrations and increases with higher concentrations of pre-incubated siRNA-Ago2 complex. **A)** and **B)** Given the dissociative model, physiological meaningful concentrations of pre-incubated siRNA-Ago2 complex are not sufficient to knock down the target concentration below 50% of its basal level for basal target levels of 3,000 or 30,000 copies/cell. **C)** Further reduction of the absolute level of basal target to 300 copies/cell leads to an efficient target turnover of 50% within 24 h. 300 copies/cell is within a physiological meaningful range for human mRNA. It is important to mention that target synthesis was not considered at this point of modelling. **D)** The associative path allows for efficient target turnover even for high target concentrations up to 30,000 copies/cell, given physiological meaningful concentrations of pre-incubated siRNA-Ago2 complex.

These concentrations are rather far away from physiological concentrations in human cells, compare Table 4.2. However, an *in silico* model, other than a living cell, can be parameterised to any set of extreme conditions and at best should be able to provide meaningful, i.e. interpretable results. Thus, the mathematical models were parameterised to match the experimental conditions and their time course simulation results were compared to the corresponding experimental trajectory.
Both the dissociative and the associative models require 300 pre-incubated Ago2-siRNA complexes to quantitatively and qualitatively match the experimental data. This is a decade more than expected from experimental observations. Furthermore, there is no difference in model behaviour at these experimental conditions. None-the-less, the time course simulations of the new model approach quantitatively and qualitatively resemble the validation data closer than those of the previous approach. Compare Fig. 4.5 (previous) and Fig. 4.9 (current). At a first glance this seems paradoxical, as the first model approach and the *in vitro* assay are based on ss siRNA, while the new approach considers ds siRNA. Then again, the introduction of pre-incubation of Ago2 and siRNA during the new approach closely resembles the *in vitro*. The modelled time course of the target is falling faster than the one observed for the *in vitro* assay. This is in accordance with the aggregation and inactivation of hAgo2 observed in the assay, for which the models do not account. Modelled knockdown has a slower onset compared to the original data point at a time of 20 min, which might be due to cleavage and release of the guide strand during the modelled ds siRNA-mediated RNAi, which is not necessary during the ss siRNA-based *in vitro* assay. In the following paragraphs the two models of ds siRNA-mediated RNAi are compared to experimental cell culture data.

Associative model explains experimental observed adaptation of the RNAi machinery to changing levels of the target

To identify the relation between the effectiveness of siRNA-mediated target reduction and target abundance, ECV305 cells were transfected with different amounts of plasmid coding for Renilla luciferase which was used as target RNA (Wünsche & Sczakiel, unpublished). At the same time, the cells were transfected via Lipofectamine 2000 with siSCR3, an siRNA against Renilla luciferase. In ECV305 cells, siSCR3 is a highly effective siRNA against RL, reflected by low IC_{50} values. Concentration-response curves and observed IC_{50} values are provided for knockdown of RL mRNA 24 h after transfection with siSCR3. siSCR3 concentration and IC_{50} values are measured in nM of transfected siRNA and converted to copies/cell according to[106] for comparability with the models.

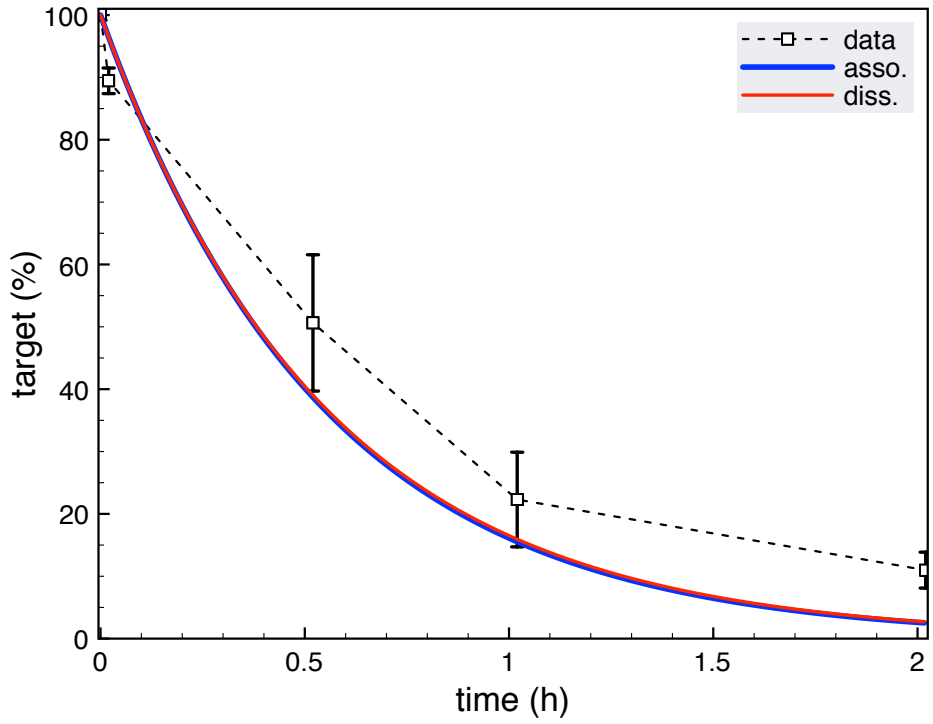


Figure 4.9 **Validation of the two competing minimal models of ds siRNA-mediated RNAi using *in vitro* siRNA-based assay.** The turnover is started by mixing pre-incubated siRNA-Ago2 complex with target RNA. Both, the dissociative and the associative models, result in the same trajectories (red and blue lines) under these *in vitro* conditions. For information on the experimental reference data, please refer to Section 4.1.2 (Validation for minimal *in vitro* RNAi assay), pp. 61.

While changing the concentration of transfected plasmid for Renilla luciferase and thus increasing target gene activity over five magnitudes, IC_{50} values stayed relatively constant i.e. 0.0495, 0.0338 and 0.0640 nM for 0.15, 1.5 and 15 ng/μl Renilla luciferase plasmid, respectively for highly effective siRNA (Fig. 4.10, Table 4.9). Deductively, the RNAi machinery is able to adapt to changing levels of the target without the need of larger amounts of siRNA.

It was tested whether this behaviour can also be observed in the two proposed models to conclude, which of the modelled mechanisms, the associative or dissociative, is likely to take place during RNAi in mammalian cells? As described above, both models were parameterised for RNAi-based knockdown of $ICAM^{-1}$ mRNA in human ECV cells in the presence of highly effective siRNA si2b. Concentration-response curves were plotted from time-course simulations with varying basal target mRNA concentrations between 2.5 and 2.5×10^6 mRNAs/cell. Similar simulations were run for ineffective siRNA si1 against $ICAM^{-1}$.

In the associative model, ineffective siRNAs leads to slowing down target turnover for low basal target concentrations. Here, IC_{50} value is reversely proportionally to target concentration. The lower the basal target concentration, the larger the IC_{50} values. Only at target concentration $\geq 1 \times 10^5$ copies/cell IC_{50} values of $\leq 1 \times 10^3$ siRNAs per cell are reached. In the case of the dissociative model, IC_{50} values proportionally depend on the basal target concentration. In the presence of effective siRNA, only low target concentrations (of $\leq 2.5 \times 10^2$ copies/cell) lead to IC_{50} values of $\leq 1 \times 10^3$ siRNAs/cell. For higher basal target concentrations, IC_{50} values increase proportionally to the basal target concentration. IC_{50} values in the presence of ineffective siRNA are equally high ($\geq 3 \times 10^4$ copies/cell) for all basal target concentrations $\leq 2.5 \times 10^4$ copies/cell and even higher in the case of higher mRNA concentrations.

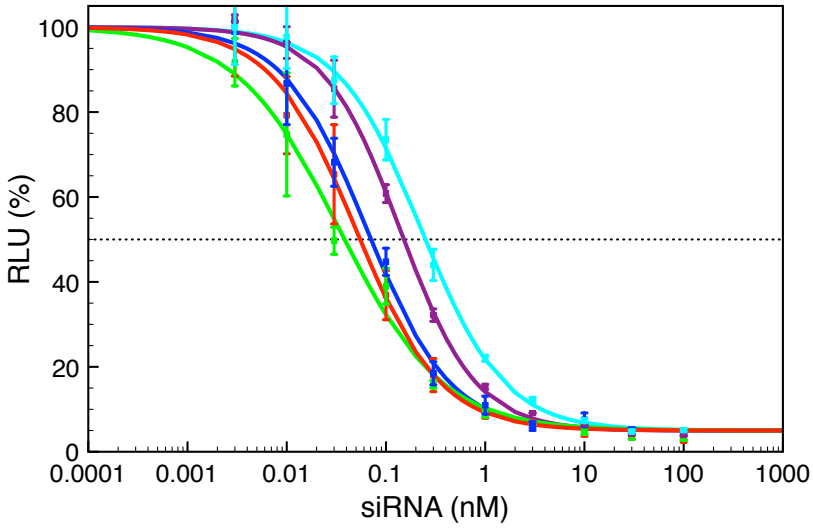
In the case of the dissociative model, target turnover depends on the basal target concentration. In the presence of affective siRNA, only low target concentrations (of $\leq 2.5 \times 10^2$ molecules per cell) lead to low IC_{50} values of $\leq 1 \times 10^3$ siRNAs per cell. For larger basal target concentrations IC_{50} values increase proportional to the basal target concentration (Fig. 4.11A). Turnover in the presence of si1 is equally slow (IC_{50} values of $\geq 3 \times 10^4$ siRNAs per cell) for all basal target concentrations $\leq 2.5 \times 10^4$ targets/cell and even slower in the case of higher concentrations (Fig. 4.11B). The dissociative model cannot explain the experimental observation. In the case of the associative model, in the presence of affective siRNA, little variation is observed between dose-concentration curves regardless of basal mRNA level. Experimental observations, as well as, simulations using the associat-

ive model show constant IC_{50} values for highly effective siRNA at all levels of basal target mRNA levels which comprise 5 magnitudes in the case of experiments (Fig. 4.10) and 6 magnitudes in the case of modelling (Fig. 4.11). For an overview, please refer to Table 4.10, suggesting that RISC is able to adapt to variations in the absolute concentration of target mRNA according to the associative model (compare also Fig. 4.6B). Ineffective siRNA slows down target turnover for low basal target concentrations. In contrast to the dissociative model, here, the IC_{50} values are reversely proportional to target concentration. The lower the basal target concentration, the larger the IC_{50} values. Only if target concentration is $\geq 1 \times 10^5$ molecules per cell low IC_{50} values of $\leq 1 \times 10^3$ siRNAs per cell are reached (Fig. 4.11D).

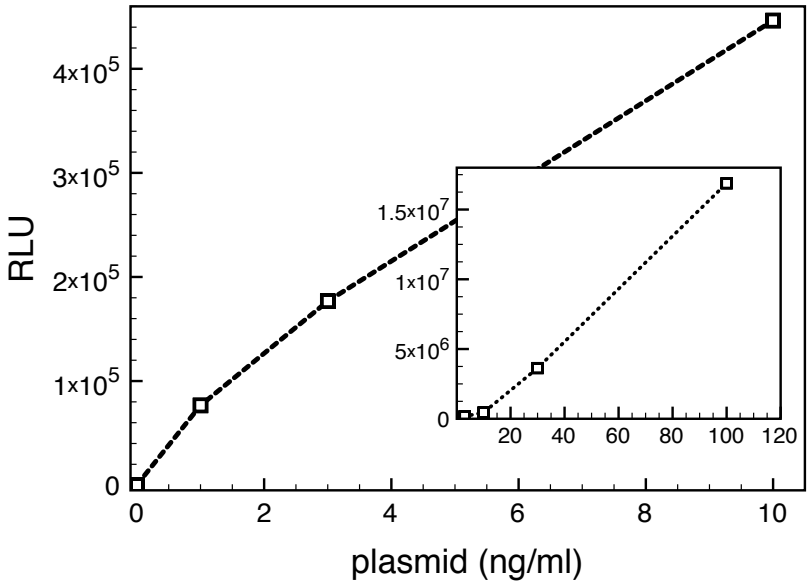
The associative model accounts for target concentration-dependent acceleration of product release from RISC and is compatible with the high multiple turnover rate of RNAi-based gene silencing in living cells. The experimental observations can be entirely explained by the associative model using measured model parameters. In contrast, the model structure of the dissociatve pathway is not sufficiently consistent with the data and this can not be improved by simply changing the model parameters. Thus, an associative mechanism of target recognition by a double-strand of guide and cleaved target possibly facilitated by Ago2 during siRNA-mediated RNAi was proposed. This model is further supported by Wünsche and Sczakiel[226] who quantified the increased kinetics of RNA-RNA recognition that occurred via facilitated strand exchange between RNA duplexes and target RNA strands *in vitro*.

Degrees of freedom: Cytoplasm volume and reaction kinetics

The cytoplasm volumes of 0.91525 pl and 2.2912 pl for HeLa and ECV cells, respectively (p. 55), are the starting point for examining the degrees of freedom in which target mRNA and RISC (containing siRNA and Argonaute) meet for reaction. When comparing the measured cytoplasm volume (V_{cyt}) and the estimated volume (V_{cyt}), $n = \frac{\log(750/\pi)}{\log(r_{cyt})} V_{cyt}^{\ddagger}$ gives the degree of freedom of the reactions. The degrees of freedom for HeLa and ECV cells is calculated as 3.18 and 2.82, respectively; these estimated three dimensions of freedom and the similarity between observed and estimated cytoplasm volume, suggest a diffusion controlled reaction during the formation of collision complexes between Ago2 and nucleic acids.

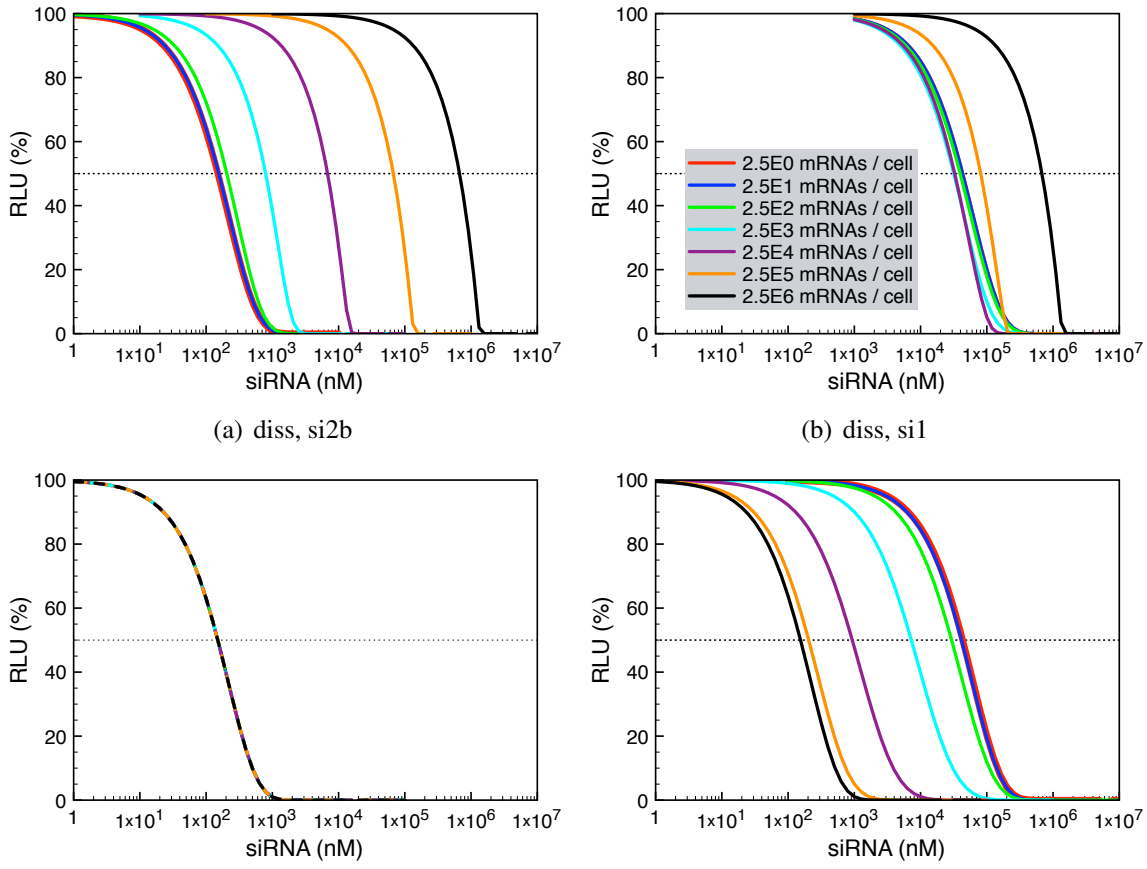


(a)



(b)

Figure 4.10 **Observed IC_{50} values of siRNA-mediated RNAi** are widely independent from target mRNA expression level. **A) Semi-log plot of concentration response curves for different target expression levels.** Response is given as target knockdown in percent and is defined as the reduction of RL activity 24 h after siRNA transfection in relation to the basal RL activity. Target expression can be regulated by the amount of transfected RL plasmid, which was 0.15, 1.5, 15, 50, and 250 ng/μl in the case of the red, green, blue, purple and cyan plot, respectively. Target knockdown is plotted against the concentration of siRNA siSCR in nM which was transfected via Lipofectamine 2000. Data, indicated by crosses, was measured in triplicates and error bars indicate standard deviation. Lines indicate data fits to a sigmoid function (x) which was used to estimate IC_{50} values and Hill slopes in Table 4.10. **B) Basal RL activity** (in absence of siSCR) is measured in RLU and plotted against transfected amount of RL plasmid to demonstrate the linear dependency between target gene product level and transfected amount of plasmid in the range of 0.1 to 15 ng/μl.



(a) diss, si2b (b) diss, si1

(c) asso, si2b (d) asso, si1

Figure 4.11 **Simulated concentration-response curves based on two competing models of ds siRNA-mediated RNAi.** Analogous to the experimental data (Fig. 4.10), simulation results are provided as a semi-log plot for different target expression levels. Response is given as target concentration 24 h after siRNA transfection (% basal target concentration). **A)** In the case of the dissociative model, target turnover depends on the basal target concentration. In the presence of affective siRNA si2b, only low target concentrations (of $\leq 2.5 \times 10^2$ copies/cell) lead to IC_{50} values of $\leq 10^3$ siRNAs/cell. For larger basal target concentrations, IC_{50} values increase proportional to the basal target concentration. **B)** Turnover in the presence of si1 is equally slow (IC_{50} of $\geq 3 \times 10^4$ siRNAs/cell) for all basal target concentrations $\leq 2.5 \times 10^4$ targets/cell and even slower in the case of higher concentrations. **C)** In the case of the associative model, in the presence of affective siRNA, little variation is observed between dose-concentration curves regardless of basal mRNA level, suggesting that RISC is able to adapt to variations in the absolute concentration of target mRNA. **D)** Ineffective siRNA slows down target turnover for low basal target concentrations. In contrast to the dissociative model, here, the IC_{50} is reversely proportional to target concentration. The lower the basal target concentration, the larger the IC_{50} values.

target plasmid ($ng/\mu l$)	IC_{50} (nM)
0.15	0.049522
1.50	0.033777
15.00	0.063975
50.00	0.137999
250.00	0.226013
IC_{50}-ratio	4.56

Table 4.9 **Experimentally observed IC_{50} values at different basal target mRNA levels.** Observed IC_{50} values are provided for knockdown of RL mRNA 24 h after LF-2000 transfection of siSCR in ECV-cells and were measured in nM and converted in copies/cell. IC_{50} was obtained from sigmoid concentration-effect curves of experiments performed in triplicates (see Fig. 4.10A). In this system, siSCR is a highly effective siRNA reflected by low IC_{50} values. The IC_{50} is relatively constant for all levels of basal target mRNA levels which comprise 5 magnitudes. This suggesting that RISC is able to adapt to variations in the absolute concentration of target mRNA.

basal	**Associative**		**Dissociative**	
target	si2b	si1	si2b	si1
2.5	149	43300	150	43387
2.5×10^1	146	41500	155	42980
2.5×10^2	148	29500	205	39656
2.5×10^3	149	7250	793	31746
2.5×10^4	151	931	6834	32833
2.5×10^5	151	205	67175	80951
2.5×10^6	155	155	669831	679485
IC_{50}-ratio	1.01	0.005	433	1.88

Table 4.10 **Comparison of modelled IC_{50} at different basal target mRNA levels(ds siRNA-mediated RNAi).** IC_{50} values are provided for knockdown of target mRNA 24 h after transfection of siRNA. Models are parameterised for knockdown of $ICAM^{-1}$ in ECV304 cells after LF-2000-based transfection of either highly effective si2b or ineffective si1. All concentrations are given in copies/cell. Similar to the experimental observations in Table 4.9, simulations based on the associative model show constant IC_{50} values for highly effective siRNA at all levels of basal target mRNA levels which comprise 6 magnitudes. In the associative model, ineffective siRNAs slow down target turnover for low basal target concentrations. Here, IC_{50} is reversely proportional to target concentration. The lower the basal target concentration, the larger the IC_{50} values. Only at target concentration $\geq 10^5$ copies/cell IC_{50} values of $\leq 10^3$ siRNAs per cell are reached. In the case of the dissociative model, IC_{50} values proportionally depend on the basal target concentration. In the presence of effective siRNA, only low target concentrations (of $\leq 2.5 \times 10^2$ copies/cell) lead to IC_{50} values of $\leq 10^3$ siRNAs/cell. For higher basal target concentrations, IC_{50} values increase proportionally to the basal target concentration. IC_{50} values in the presence of ineffective siRNA are equally high ($\geq 3 \times 10^4$ copies/cell) for all basal target concentrations $\leq 2.5 \times 10^4$ copies/cell and even higher in the case of higher mRNA concentrations. Suggesting that RISC is able to adapt to variations in the absolute concentration of target mRNA in the presence of highly effective siRNA for the associative model, but not for the dissociative model.

4.3 RNA metabolism

The sequence match between effector RNA (i.e. siRNA or miRNA) and target mRNA, i.e. the binding site, cannot fully explain RNAi knockdown efficacy, e.g. [59, 101]. Larsson et al.[104] propose mRNA metabolism as a key factor for RNAi knockdown efficacy. Thus, the modelling of mRNA metabolism was included into this study.

4.3.1 Transcriptional bursting

The previous way of dealing with mRNA concentration in the models, i.e. a lack of any mRNA synthesis after initialisation of relatively high concentrations (500 - 30,000 copies / cell) of target mRNA best describes the situation after a transcriptional burst[26, 58, 173]. Transcription can be irregular, with strong periods of activity, interspersed by long periods of inactivity, leading to bursts (or pulses) of activity. In its most simple form, a gene exists in two states, one where activity is negligible and one where there is a certain probability of activation in which transcription occurs[174]. While most eukaryotic genes are likely to exist in more than two states, e.g. more than twenty post-translational modifications of nucleosomes are known, the simple two step model may provides a reasonable intellectual framework for understanding the effect of transcriptional bursting on RNAi[234].

4.3.2 Homogenous synthesis

However, it is likely that not all eukaryotic genes show transcriptional bursting[234]. These genes are effectively always active, with a simple probability describing the numbers of RNAs generated. This may also be true for plasmid-located genes. In the following, an alternative way of dealing with mRNA transcription is introduced. A reductionist's model of the metabolism of an mRNA species can be described by a synthesis reaction rate constant and a degradation reaction rate constant, which, taken together, build up a steady state target mRNA concentration in the cell. Different pairs of constants can result in the same steady state concentration, i.e. two high constants or two low constants. However, these pairs differ in their kinetic properties, such as recovery of steady state after knockdown, e.g. by RNAi. The synthesis rate also influences the kinetics of siRNA-mediated RNAi itself. Hence, it is not sufficient to know a target's steady state concentration, but also the kinetic rate constants that lead to that steady state. This adds one further reaction step

steady state copies/cell	number of mRNA species in Caco2	in human tissues
> 500	61	55
50 - 500	554	578
5 - 50	6,023	6,160
≤ 5	62,400	127,342
total	69,381	134,135

Table 4.11 **Steady state concentrations of mRNA species in human cells** according to[212].

$$\mathrm{T} \underset{\mathrm{k}_{-1}^{degr}}{\overset{\mathrm{k}_{+1}^{degr}}{\rightleftharpoons}} \mathrm{T}^{\dagger}$$

and a constant synthesis rate k_{synt} to the model, where $k_{+1}^{degr} = \frac{\ln(2)}{\lambda}$ and $k_{-1}^{degr} \simeq 0$, λ is the mRNA's half time. The time-resolved behaviour of the target RNA metabolism can be modelled by Eq. (4.72).

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = k_{synt} - k_{+1}^{degr}[T] + \tag{4.72}$$

Steady state concentrations of mRNA species, not prone to transcriptional bursting, were described[71, 73] and revised[212], leading to the distribution summarised in Table 4.11. While degradation and synthesis rates of target mRNA used in previous modelling approaches are inexplicable[10–12, 172], there exist other reliable literature sources: [228] distinguish between transcription factor transcripts and housekeeping transcripts with respective half times of < 2 h and > 10 h and an overall median of 10 h[228]. Formation of target mRNA concentration in cytoplasm roughly depends on transcription, post-transcriptional processing and transportation to cytoplasm[70]. The concept of *metabolic factor* f_{met} was introduced to connect the synthesis rate constant directly to the desired steady state concentration and the literature-based degradation rate constant, where $k_{synt} = \frac{k_{+1}^{degr}}{f_{met}}$ and $f_{met} = \frac{1.50553 \times 10^{12}}{[T]_{steady\_state}}$. Compare Fig. 4.12.

4.4 Delivery of small RNAs

All cell culture-based experimental data used subsequently as reference for the models is based on assays using Lipfectamine 2000 for siRNA delivery. Lipofectamine 2000 (Invitrogen) is a common transfection reagent of RNA, including siRNA, or plasmid DNA into *in vitro* cell cultures by lipofection. It contains lipid subunits that can form liposomes in an

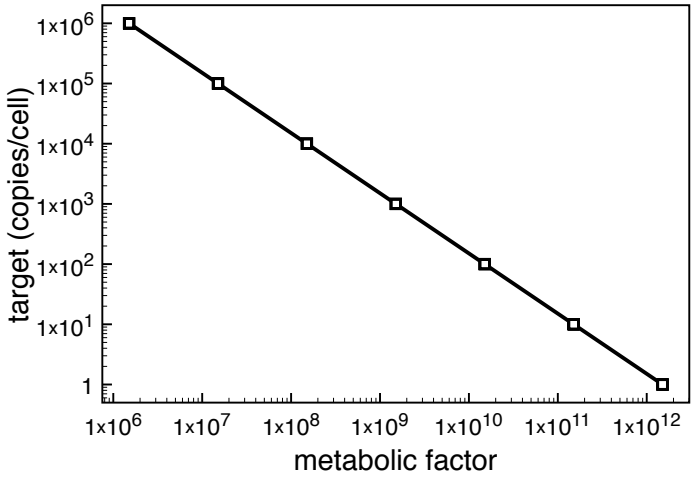


Figure 4.12 **Relation between metabolic factor and steady state target concentration.** The concept of *metabolic factor* f_{met} was introduced to connect the synthesis rate constant directly to the desired steady state concentration and the literature-based degradation rate constant, where $k_{synt} = \frac{k_{+1}^{degr}}{f_{met}}$ and $f_{met} = \frac{1.50553 \times 10^{12}}{[T]_{steady\_state}}$.

aqueous environment, which entrap the transfection materials: the cationic liposome formulation forms a complex with negatively charged nucleic acid molecules. This allows them to overcome the electrostatic repulsion of the cell membrane. In cell culture assays, siRNA up to 100 nM is transfected efficiently within the first 30 min and barely any uptake occurs 4 h after transfection. Experiments are not part of this thesis; the data was published in[40, 106] The delivery of siRNA is modelled using two rate laws. The first one accounts for the Liposome-facilitated spacial re-location of siRNA between the extracellular compartment (i.e. the transfection buffer) and the cell. Compare Eq. (4.73). Mind that the intracellular siRNA is not readily available to the RNAi machinery. Therefore, a second rate law accounts for the re-location of, so-called, trapped siRNA to freely available siRNA in the cell's cytoplasm. Compare Eq. (4.74).

$$\mathrm{S}^{\mathrm{extra}} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_{+1}}{\rightleftharpoons}} \mathrm{S}^{\mathrm{trap}} \tag{4.73}$$

$$\mathrm{S}^{\mathrm{trap}} \underset{\mathrm{k}_{-4}}{\overset{\mathrm{k}_{+4}}{\rightleftharpoons}} \mathrm{S}^{\mathrm{free}} \tag{4.74}$$

The biochemical equations translate into two ODEs, which can be analytically solved (compare Section 3.2.1, pp. 29). Their reaction rate constants are fitted to time-resolved experimental data from[40, 106]. There are two observations from the experimental data: (i) 20 min after transfection the uptake from the extracellular compartment is saturated; (ii) 4 h

after transfection, 4% of the delivered siRNA is bioavailable, i.e. free to interact with the RNAi machinery. The resulting parameters are: for the first delivery step, keeping the backward reaction rate to zero, the data fits a first order rate law with a reaction rate constant of $0.0723(\pm 0.0010)$ s^{-1} and a limit of $311,472(\pm 1304)$ copies/cell. The experimental data for the second step (again, the backwards reaction is set to zero) fits a first order rate law with a reaction rate constant of $7.99 \times 10^{-5}(\pm 3.85 \times 10^{-5})$ s^{-1} and a limit of $242(\pm 17,06)$ copies/cell. Values in brackets provide the standard errors of the fit. Compare Fig. 4.13a and b.

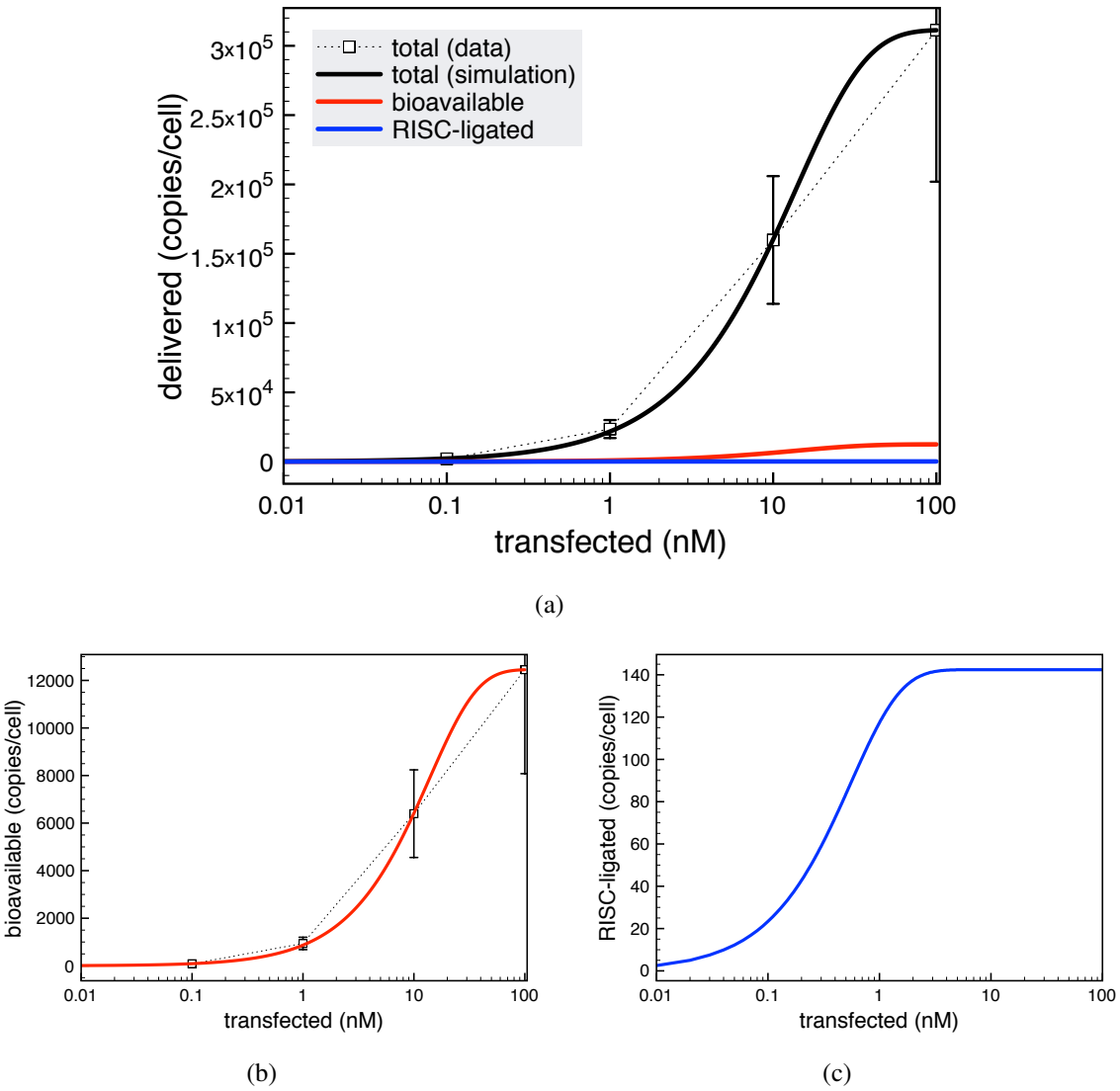


Figure 4.13 **Liposome-facilitated delivery of small RNAs.** 4 h after transfection of various siRNA concentrations via Lipofectamine 2000[40, 106]. **A)** Total delivered siRNA (black line) compared to bioavailable (red line) and RISC-bound (blue line) siRNA; **B)** Bioavailable siRNA: data and fit; **C)** modelled RISC-bound siRNA at 4 h (both models).

4.5 A refined model based on precise quantitative data

4.5.1 Parametrisation

In the meantime, the implementation of pre-steady-state kinetic techniques for characterisation of siRNA binding, target RNA recognition, sequence-specific cleavage and product release by recombinant human Ago2 led to precise quantitative kinetic data of RNAi[32]. In combination with available X-ray structures of bacterial Argonaute and human Argonaute 2 it was possible to match the kinetic data to conformational changes along the RNAi pathway and propose a comprehensive minimal mechanistic model describing fundamental steps during RNAi[32]. These new data enable the mathematical formulation of unprecedented detailed models of RNAi. The kinetic parameter of the catalytic step (target cleavage) is based on the pre-steady state RNase H rate of HIV1 reverse transcriptase[90], since data for Ago2's own endonuclease activity is not available yet. All kinetic parameters of the RNAi model are listed in Table 4.12.

Reaction compartment, i.e. the cytoplasm of a mammalian cell, as well as, Argonaute 2 concentration, the basal levels of target mRNAs and siRNA concentrations are based on literature data and are consistent with in house experiments. Parameters of cellular processes directly interacting with RNAi components, like RNA metabolism and the delivery of siRNA into cells are as well taken from literature and in house studies. These parameters and their references are provided in Table 4.13.

info	reaction step	k	model		value	unit
			dis	**ass**		
siRNA binding	binary complex formation 1	k_{+1}	+	+	6.0×10^7	$M^{-1}s^{-1}$
		k_{-1}	+	+	6.2	s^{-1}
	binary complex formation 2	k_{+2}	+	+	0.3	s^{-1}
		k_{-2}	+	+	0.2	s^{-1}
	binary complex formation 3	k_{+3}	+	+	0.01	s^{-1}
		k_{-3}	+	+	0.007	s^{-1}
target binding	ternary complex formation 1	k_{+4}	+	+	3.2×10^8	$M^{-1}s^{-1}$
		k_{-4}	+	+	2	s^{-1}
	ternary complex formation 2	k_{+5}	+	+	0.01	s^{-1}
		k_{-5}	+	+	0.002	s^{-1}
	ternary complex formation 3	k_{+6}	+	+	0.003	s^{-1}
		k_{-6}	+	+	0.0002	s^{-1}
	target cleavage	k_{+7}	+	+	10	s^{-1}
		k_{-7}	+	+	-	s^{-1}
dissociative path	product release 1	k_{+8}	+	-	1.2	s^{-1}
		k_{-8}	+	-	-	s^{-1}
	product release 2	k_{+9}	+	-	0.002	s^{-1}
		k_{-9}	+	-	-	s^{-1}
	product release 3	k_{+10}	+	-	0.0002	s^{-1}
		k_{-10}	+	-	-	s^{-1}
associative path	quaternary complex formation 1	k_{+11}	-	+	3.2×10^8	$M^{-1}s^{-1}$
		k_{-11}	-	+	2	s^{-1}
	quaternary complex formation 2	k_{+12}	-	+	0.01	s^{-1}
		k_{-12}	-	+	0.002	s^{-1}
	quaternary complex formation 3	k_{+13}	-	+	0.003	s^{-1}
		k_{-13}	-	+	0.0002	s^{-1}
	product release 1	k_{+14}	-	+	5	s^{-1}
		k_{-14}	-	+	-	s^{-1}
	product release 2	k_{+15}	-	+	0.03	s^{-1}
		k_{-15}	-	+	-	s^{-1}
	product release 3	k_{+16}	-	+	0.006	s^{-1}
		k_{-16}	-	+	-	s^{-1}

Table 4.12 **Kinetic rate constants based on precise quantitative kinetic data (used for the refined RNAi models).** For each elementary reaction step, the forward and backward rate constant pair is provided. Columns 'dis' and 'ass' indicate a parameter pair's use in the dissociative and associative model, respectively. Description continues on the next page.

Table 4.12, **continued. Kinetic rate constants based on precise quantitative kinetic data.** If both forward and backward reaction rate constants are provided, reaction steps are modelled as reversible. If the data is not available, reaction steps are modelled as quasi-irreversible, as in the case of the target cleavage step. The reaction rate constants for complex formations (siRNA binding, target binding are based on pre-steady-state binding data for hAgo2/RNA binary and ternary complex formation in Table 1 of[32]. Similarly, complex formation between target RNA and hAgo2 in the presence of excess target RNA was measured to account for associative target binding. Reaction rate constants for dissociative and associative target release are based on complex dissolution rate constants in the absence or presence of excess target RNA, respectively (Table S1 of[32]). The constant of target cleavage is based on the pre-steady state RNase H rate of HIV1 reverse transcriptase[90] since data for Ago2's own endonuclease activity is currently not available. This table is based on Supplementary Table S1 of[40].

4.5.2 Model structure

Dissociative model of siRNA-mediated RNAi in mammalian systems

A new computational model of siRNA-mediated RNAi in mammalian cells was created in order to link experimental results with a quantitative and time-resolved understanding of RNAi. Mechanistically, it is based on the current literature view on RNAi, featuring a dissociative ligand substitution mechanism between incoming target mRNA and cleavage product. Strand selection of ds siRNA is neglected for simplification. Thus, the model is a direct extension of the first dissociative model, introduced in Section 4.1, pp. 50.

In a first pass, RISC, comprising hAgo2 (A) and possibly other co-factors, binds to free siRNA (S) in a three phase reaction to form complex AS (siRNA binding I-III). RISC co-factors and the passenger strand cleavage and release during complex formation involving ds siRNA are omitted in the model for clarity, but essentially the proposed pathway also holds true for ds siRNA. A free target mRNA (T) enters the complex AS, again in a three phase reaction, eventually binding complementary to the siRNA's guide strand to form complex AST (target binding I-III). Once the catalytically active complex is formed the system undergoes multiple rounds of target binding and cleavage, every time followed by a slower product release without falling apart: the bound target mRNA is cleaved by Ago2's endonuclease activity (target cleavage reaction) leading to complex ASQ. Subsequently, the remaining cleavage product (Q) is released from the complex in a three step reaction (product dissociation I-III). Thus the target mRNA population is reduced by one and the free product

parameter	S	C	value	unit	ref
cytoplasm volume	V_{cyto}	-	2.5×10^{-12}	l	[28, 33, 53, 116, 144, 171]
initial siRNA	$[S]_0$	V_{cyto}	30 4.98×10^{-5}	copies/cell nM	sample value[#]
basal hAgo2 level	$[A]_0$	V_{cyto}	$200,000$ 132.84	copies/cell nM	[85, 217]
basal target mRNA level	$[T]_0$	V_{cyto}	300 0.2	copies/cell nM	sample value[*]
target degradation rate	k_{deg}	-	6.40×10^{-5}	s^{-1}	[224, 228]
target synthesis rate	v_{synth}	-	1.27×10^{-14}	Ms^{-1}	[212]

Table 4.13 **Reaction compartment volume and species concentrations used for the refined RNAi models.** Column 'S' provides the parameters' symbols. In the case of concentration, column 'C' indicates the compartment for which it is given. Calculated from[217], a human cell contains 1.4×10^5 to 1.7×10^5 Argonautes per cell, where Ago2 is the most abundant of the four Ago-subtypes. According to an in-house study[33], there are 2.5×10^5 copies Ago2 per human cell. The basal target mRNA level and initial siRNA concentrations are varied depending on model objective or modelled experimental set up. This table is based on Supplementary Table S2 of[40].

[*]Physiological mRNA level can vary between 1 and 10,000 particles per human cell, compare[173, 237]. For instance, the basal Renilla luciferase mRNA level is between 300 and 500 copies/cell for stably transfected ECV304 cells[106].

[#]Small RNAs, involved in RNAi occur in 5 - 500 copies per human cell[110] and the number of active siRNA for half maximal target suppression can be reached at 10 - 20 copies/cell[105, 106].

strands are prone to degradation by cellular nucleases, leaving behind binary complex AS comprising of siRNA's guide strand bound to Ago2 which can bind the next target molecule and thus close the cycle of subsequent passes of target turnover. The reaction scheme for the dissociative model is shown in Fig. 4.14A.

Associative model: accelerated product release via strand exchange between RISC-bound, cleaved RNA and incoming target RNA.

The same argumentation for the introduction of an alternative model with associative mechanism for target binding and release of cleavage product during multiple turnover of target mRNA, described in Section 4.1, p. 51, holds true for its introduction during the extended modelling approach, explained here: cell culture experiments revealed that gene silencing rates can be increased by rising the target mRNA concentration, at constant siRNA and hAgo2 concentrations. That is, IC_{50}, the half maximal inhibitory siRNA concentration after 24 h of knockdown, stays relatively constant over several magnitudes of basal target mRNA levels. Additionally, it was shown *in vitro* that the release of the target RNA strand within a RISC bound RNA duplex is accelerated in the presence of free target RNA[32]. These experimental findings suggest that RNAi-based knockdown should be able to adapt to changes in target mRNA level over several magnitudes. The defiance of this adaptation in the dissociative model (compare Fig. 4.18E&F) motivates an alternative model. Different from the dissociative mechanism, an associative mechanism involves the approach of the incoming ligand to the complex before departure of the leaving ligand. The first step of the associative mechanism typically is rate-determining and depends on the ligand concentration. An associative mechanism of target strand recognition by RISC during siRNA-mediated RNAi (dubbed associative model) is proposed. In this model, hAgo2 supposedly facilitates the interaction of an incoming target strand with RISC-bound guide-product duplexes most likely by a strand invasion mechanism as first shown for *in vitro* RNA-RNA strand exchange by Wünsche & Sczakiel 2005[226]. Fig. 4.14B pictures the target-product-exchange mechanism, which is key to the multiple turnover cycle of the associative model. However, before entering the cycle of multiple turnover, in a first pass, hAgo2 and possibly other co-factors, binds to free siRNA in a three phase reaction to form complex AS (siRNA binding I-III) followed by a three stepped binding of a first target to form complex AST. The first pass of siRNA binding and target binding, as well as, target cleavage are equivalent to the reaction steps of the dissociative model. However, after formation of complex ASQ which consits of hAgo2, siRNA and cleaved product fragments, the two models differ. In contrast to the dissociative model, target cleavage is followed by an associative binding of another target

(associative target binding I-III) which may invade the hydrogen bonds between siRNA and product, forming transient "quaternary" complex ASQT eventually replacing the product strands. The cycle is closed by the complete release of the replaced product fragments in three reaction steps (product release I-III), which leads to the formation of the next catalytically active complex AST. The released product fragments Q eventually are degraded by cellular nucleases. The remaining ternary complex can pass through the next, identical cycle of target cleavage, target association and product release. The reaction scheme for the associative model is shown in Fig. 4.14B. Analogous to the dissociative model, the reaction rate constants for complex formation are based on pre-steady-state binding data of[32]. Complex formation between target RNA and hAgo2 in the presence of excess target RNA was measured to account for associative target binding. The reaction rate constants for associative target release are based on target dissociation rate constants in the presence of excess target RNA (SI Appendix Table S1 of[32]).

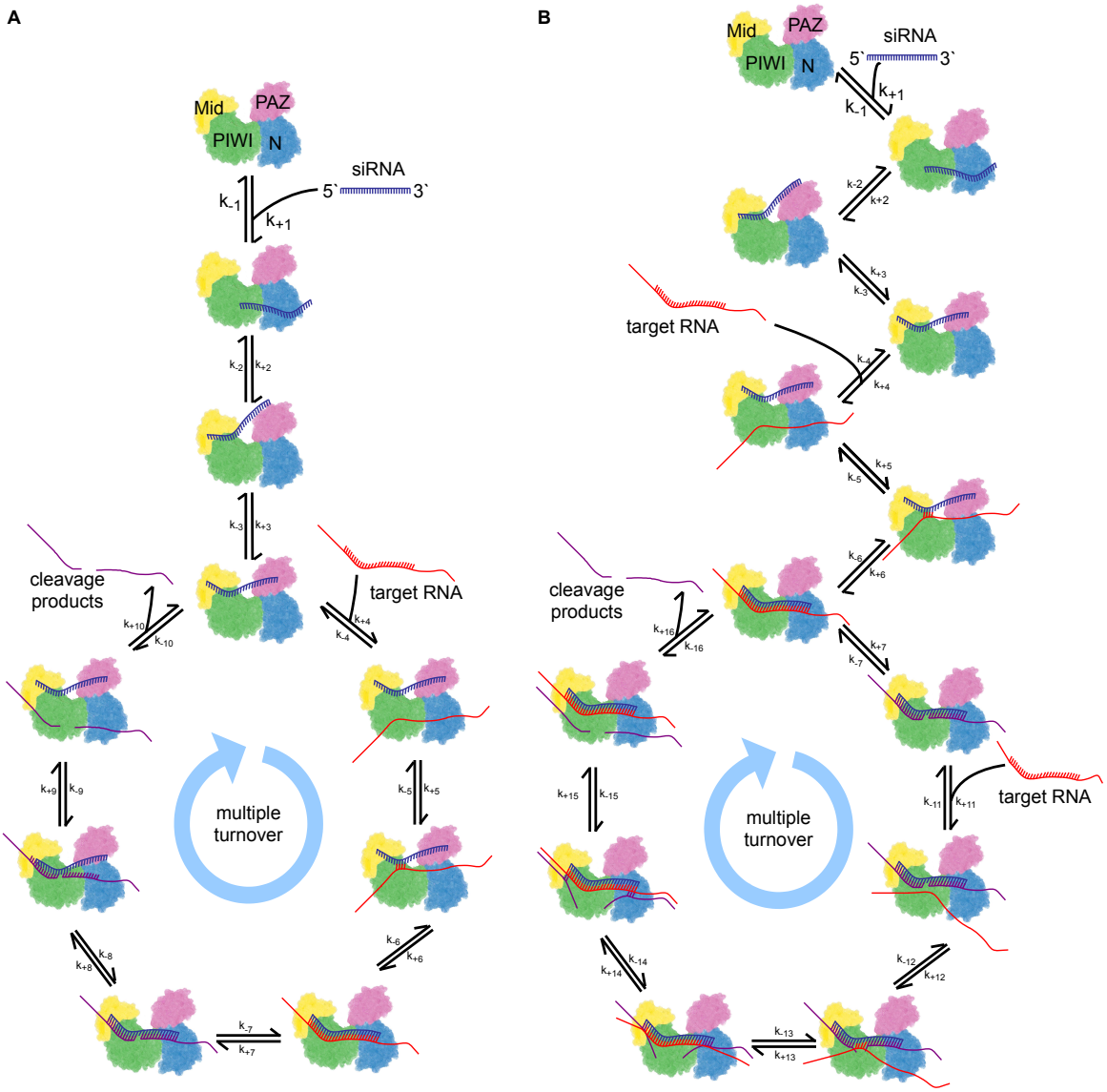


Figure 4.14 **Illustrations of the two refined RNAi models accounting for precise quantitative kinetic data. A)** Dissociative model. **B)** Associative model. Ago2 is represented as a scheme of its electron density cloud with the four domains N, PAZ, Mid, and PIWI coloured individually. RNAs are indicated by lines and are as well colour-coded. The relative positions of protein and RNA substrates are indicated and the formation of hydrogen bonds is illustrated by short connection lines between RNA strands. RISC co-factors and the passenger strand cleavage and release during complex formation involving ds siRNA are omitted for clarity but essentially, the proposed pathway also holds true for ds siRNA. The two illustrations start with complex formation between ss siRNA and Ago2. From here the two models diverge essentially as described in detail below. Description continues on the next page.

Fig. 4.14, **continued. Illustrations of the two refined RNAi models accounting for precise quantitative kinetic data.**
The reactions for the dissociative model match rate equations in Eq. (4.75) to (4.84). The reactions for the associative model match rate equations in Eq. (4.75) to (4.81) and (4.85) to (4.90). The corresponding reaction rate constants for each individual step are given in Table 4.12. **A) Dissociative model.** Once the catalytically active complex between ss siRNA and Ago2 is formed, the system undergoes multiple rounds of target binding and cleavage, followed by product release without falling apart. **B) Associative model.** Before entering the cycle of multiple turnover, the complex between ss siRNA and Ago2 has to bind a first target. This triggers multiple rounds of target turnover. In contrast to the dissociative model, target cleavage is followed by the associative binding of another target which may invade the hydrogen bonds between siRNA and product, eventually replacing the product strands. The cycle is closed by the complete release of the replaced product fragments, leading to the formation of the next catalytically active complex. This figure is based on Supplementary Figure S1 of[40].

In this section the RNAi process in mammalian cells is formally described by a network of elementary reactions. Their biochemical rate equations connect experimentally derived *in vitro* reaction rate constants with physiological concentrations of reactants and cell compartment volumes. The reaction mechanism is encoded in the reactions' rate laws, its parameters (Table 4.12), and the way the reactions are linked to one another in the two alternative metabolic pathways of Fig. 4.14. The reactions for the dissociative model translate in the reactions of Eq. (4.75) to (4.84). Both models share the reactions of Eq. (4.75) - (4.81). However, the dissociative model's reactions of Eq. (4.82) - (4.84) are replaced by the reactions of Eq. (4.85) - (4.90) in the case of the associative model. Molecular entities are abbreviated as: A (Ago), T (target mRNA), S (ss siRNA), Q (product fragments, resulting from cleavage of S). To distinguish between different complexes consisting of the same basic entities, dagger (†) and double dagger (‡) symbols may be used. For instance, AST, AST†, and AST‡ are a collision complex between Ago and target RNA, an intermediate

state of ternary complex and cleavage competent RISC, respectively.

$$A + S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} AS \tag{4.75}$$

$$AS \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} AS^{\dagger} \tag{4.76}$$

$$AS^{\dagger} \underset{k_{-3}}{\overset{k_{+3}}{\rightleftharpoons}} AS^{\ddagger} \tag{4.77}$$

$$AS^{\ddagger} + T \underset{k_{-4}}{\overset{k_{+4}}{\rightleftharpoons}} AST \tag{4.78}$$

$$AST \underset{k_{-5}}{\overset{k_{+5}}{\rightleftharpoons}} AST^{\dagger} \tag{4.79}$$

$$AST^{\dagger} \underset{k_{-6}}{\overset{k_{+6}}{\rightleftharpoons}} AST^{\ddagger} \tag{4.80}$$

$$AST^{\ddagger} \underset{k_{-7}}{\overset{k_{+7}}{\rightleftharpoons}} ASQ \tag{4.81}$$

$$ASQ \underset{k_{-8}}{\overset{k_{+8}}{\rightleftharpoons}} ASQ^{\dagger} \tag{4.82}$$

$$ASQ^{\dagger} \underset{k_{-9}}{\overset{k_{+9}}{\rightleftharpoons}} ASQ^{\ddagger} \tag{4.83}$$

$$ASQ^{\ddagger} \underset{k_{-10}}{\overset{k_{+10}}{\rightleftharpoons}} AS^{\ddagger} + Q \tag{4.84}$$

$$ASQ^{\ddagger} + T \underset{k_{-11}}{\overset{k_{+11}}{\rightleftharpoons}} ASQT \tag{4.85}$$

$$ASQT \underset{k_{-12}}{\overset{k_{+12}}{\rightleftharpoons}} ASQT^{\dagger} \tag{4.86}$$

$$ASQT^{\dagger} \underset{k_{-13}}{\overset{k_{+13}}{\rightleftharpoons}} ASQT^{\ddagger} \tag{4.87}$$

$$ASQT^{\ddagger} \underset{k_{-14}}{\overset{k_{+14}}{\rightleftharpoons}} ASTQ \tag{4.88}$$

$$ASTQ \underset{k_{-15}}{\overset{k_{+15}}{\rightleftharpoons}} ASTQ^{\dagger} \tag{4.89}$$

$$ASTQ^{\dagger} \underset{k_{-16}}{\overset{k_{+16}}{\rightleftharpoons}} AST^{\ddagger} + Q \tag{4.90}$$

4.5.3 Analyses

Scrutinising the models: Stoichiometric Analysis

Elementary flux modes could be identified for either of the two models. This implies, that the proposed models do not comprise smaller sub-networks that allow a metabolic reconstruction network to function in steady state.
Mass conservation analyses of dependent species starting with different combinations of initial species concentrations, led to the identification of three distinguishable mass flows in each of the two models: Eq. (4.91) - (4.93) in the case of the dissociative and Eq. (4.94) - (4.96) in the case of the associative model. It can be observed that mass conservation holds true at all time for both models. That is, mass is neither created nor destroyed, although it is rearranged in space, and the entities associated with it (i.e. target mRNA, Ago2 and siRNA) are changed in form.

$$T_0 = T + AST^{\dagger} + ASQ + ASQ^{\ddagger} + AST + ASQ^{\dagger} + AST^{\ddagger} + Q \tag{4.91}$$

$$A_0 = A + AS^{\ddagger} + AS + AST^{\dagger} + ASQ + ASQ^{\ddagger} + AST + AS^{\dagger} + ASQ^{\dagger} + AST^{\ddagger} \tag{4.92}$$

$$S_0 = G + AS^{\ddagger} + AS + AST^{\dagger} + ASQ + ASQ^{\ddagger} + AST + AS^{\dagger} + ASQ^{\dagger} + AST^{\ddagger} \tag{4.93}$$

$$\begin{aligned} ASQT_0^{\dagger} = {} & ASQT^{\dagger} + 0.5AST^{\ddagger} + ASQT + ASQT^{\ddagger} + 0.5AST^{\dagger} + AST\,Q^{\dagger} \\ & + 0.5ASQ + AST\,Q + 0.5AST + 0.5t + 0.5Q \end{aligned} \tag{4.94}$$

$$\begin{aligned} A_0 = {} & A + 0.5AST^{\ddagger} + AS + AS^{\ddagger} + 0.5AST^{\dagger} + 0.5ASQ + AS^{\dagger} + 0.5AST \\ & - 0.5t - 0.5Q \end{aligned} \tag{4.95}$$

$$\begin{aligned} S_0 = {} & g + 0.5AST^{\ddagger} + AS + AS^{\ddagger} + 0.5AST^{\dagger} + 0.5ASQ + AS^{\dagger} + 0.5AST \\ & - 0.5t - 0.5Q \end{aligned} \tag{4.96}$$

$$
N_{diss} = \begin{array}{c} \\ T \\ S \\ A \\ Q \\ AS \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ AST^{\dagger} \\ AST^{\ddagger} \\ ASQ \\ ASQ^{\dagger} \\ ASQ^{\ddagger} \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 & r_8 & r_9 & r_{10} \end{array} \\
\left(\begin{array}{cccccccccc}
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{array} \right)
\end{array}
\tag{4.97}
$$

Eq. (4.97) shows the stoichiometric matrix for the dissociative model with reactions siRNA binding 1 (r_1), siRNA binding 2 (r_2), siRNA binding 3 (r_3), target binding 1 (r_4), target binding 2 (r_5), target binding 3 (r_6), target cleavage (r_7), dissociative product release 1 (r_8), dissociative product release 2 (r_9), and dissociative product release 3 (r_{10}). Likewise Eq. (4.98) shows the stoichiometric matrix for the associative model with reactions siRNA binding 1 (r_1), siRNA binding 2 (r_2), siRNA binding 3 (r_3), target binding 1 (r_4), target binding 2 (r_5), target binding 3 (r_6), target cleavage (r_7), associative target binding 1 (r_{11}), associative target binding 2 (r_{12}), associative target binding 3 (r_{13}), associative product release 1 (r_{14}), associative product release 2 (r_{15}), associative product release 3 (r_{16}). The corresponding reduced link matrices of the dissociative and associative models are given by Eq. (4.99) and (4.100), respectively.

$$
N_{asso} = \begin{array}{c}
\begin{array}{c} S \\ A \\ Q \\ AS \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ AST^{\dagger} \\ AST^{\ddagger} \\ ASQ \\ ASQT \\ ASQT^{\dagger} \\ ASQT^{\ddagger} \\ ASTQ \\ ASTQ^{\dagger} \end{array}
\begin{array}{c}
\begin{array}{ccccccccccccccc}
r_1^{bin} & r_2^{bin} & r_3^{bin} & r_1^{ter} & r_2^{ter} & r_3^{ter} & r^{cleav} & r_1^{ass\_qua} & r_2^{ass\_qua} & r_3^{ass\_qua} & r_1^{ass\_rel} & r_2^{ass\_rel} & r_3^{ass\_rel} & \end{array} \\
\begin{pmatrix}
t & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0
\end{pmatrix}
\end{array}
\end{array}
\tag{4.98}
$$

$$L_{diss} = \begin{array}{c} \\ S \\ A \\ Q \\ AS \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ AST^{\dagger} \\ AST^{\ddagger} \\ ASQ \\ ASQ^{\dagger} \\ ASQ^{\ddagger} \end{array} \begin{array}{c} \begin{array}{cccccccccccc} T & Q & AS & AS^{\dagger} & AS^{\ddagger} & AST & AST^{\dagger} & AST^{\ddagger} & ASQ & ASQ^{\dagger} & ASQ^{\ddagger} \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & \\ 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & \\ 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \end{pmatrix} \end{array} \tag{4.99}$$

Steady State Analysis

For the dissociative model one asymptotically stable steady state was found based on the eigenvalues of the Jacobian matrix of Eq. (4.101). The eigenvalues of J_{diss} are identified as 0, 0, 0, -9.67×10^{-6}, -0.0002, -0.0021, -0.00499648, -0.0140762, -0.433315, -1.2, -2.01324, -100, and -1002.52. For the dissociative model, all reaction rates at steady state are numerically zero. The eigenvalues of J_{asso} are identified as 0, 0, 0, 0, -1.29×10^{-5}, -0.0036, -0.00498999, -0.00499863, -0.0140762, -0.02, -0.433315, -2.01001, -2.01431, -13.8002, -100, -1002.52; In contrast to the dissociative model, the only steady state found for the associative model is of undetermined stability based on the eigenvalues of the Jacobian matrix of Eq. (4.102). At steady state, all reaction rates are numerically zero. Looking at their stoichiometries, mass conservation properties, the two models differ in their internal behaviour at steady state. However, most importantly, they both lead to a steady state with $[T]_{ss} = 0$. Next, it will be investigated how the apparently similar steady states are dynamically developed given physiological meaningful initial conditions.

$$
L_{asso} = \begin{array}{c}
\begin{array}{cccccccccccccccc}
T & Q & AS & AS^{\dagger} & AS^{\ddagger} & AST & AST^{\dagger} & AST^{\ddagger} & ASQ & ASQT & ASQT^{\ddagger} & ASTQ & ASTQ^{\dagger} \end{array} \\
\left(\begin{array}{cccccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0.5 & -1 & -1 & -1 & -0.5 & -0.5 & -0.5 & -0.5 & 0 & 0 & 0 & 0 \\
0.5 & 0.5 & -1 & -1 & -1 & -0.5 & -0.5 & -0.5 & -0.5 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right)
\begin{array}{c} S \\ A \\ Q \\ AS \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ AST^{\dagger} \\ AST^{\ddagger} \\ ASQ \\ ASQT \\ ASQT^{\dagger} \\ ASQT^{\ddagger} \\ ASTQ \\ ASTQ^{\dagger} \end{array}
\end{array}
\tag{4.100}
$$

$$
J_{diss} = \begin{array}{c} \\ A \\ AS \\ S \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ T \\ AST^{\dagger} \\ AST^{\ddagger} \\ Q \\ ASQ \\ ASQ^{\dagger} \\ ASQ^{\ddagger} \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} A & AS & S & AS^{\dagger} & AS^{\ddagger} & AST & T & AST^{\dagger} & AST^{\ddagger} \end{array} \\
\begin{pmatrix}
-0.007 & 0 & 0 & 0 & 0.0002 & 2 & 0.012 & 0 & -0 \\
0 & -6.46 & 0 & 0 & 0 & 0 & 0.17 & 0 & -0 \\
0 & 0 & -0.005 & 0 & 0 & 0.01 & 0 & 0 & 0.0002 \\
0 & 0 & 0 & -1.2 & 0 & 0 & 0 & 0 & 100 \\
0 & 0 & 0 & 0 & -0.0002 & 0 & 0 & 0.0021 & -0 \\
0 & 0 & 0.002 & 0 & 0 & -2.01 & 0 & 0 & -0 \\
0.007 & 0.26 & 0 & 0 & 0 & 0 & -0.182 & 0 & -0 \\
0 & 0 & 0 & 1.2 & 0 & 0 & 0 & -0.0021 & -0 \\
0 & 0 & 0.003 & 0 & 0 & 0 & 0 & 0 & -100 \\
0 & 0 & 0 & 0 & 0.0002 & 0 & 0 & 0 & -0 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & -0 \\
0 & 6.2 & 0 & 0 & 0 & 0 & 0 & 0 & -0 \\
0 & 6.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

$$
J_{asso}\ (continued) = \begin{array}{c} \\ A \\ AS \\ S \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ T \\ AST^{\dagger} \\ AST^{\ddagger} \\ Q \\ ASQ \\ ASQ^{\dagger} \\ ASQ^{\ddagger} \end{array}
\begin{array}{c}
\begin{array}{cccc} Q & ASQ & ASQ^{\dagger} & ASQ^{\ddagger} \end{array} \\
\begin{pmatrix}
0 & -0.0032 & 0 & -0 \\
0 & 0 & 996 & 7.21 \times 10^{-6} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & -0 \\
0 & 0 & 0 & -0 \\
0 & 0.0032 & 0 & -0 \\
0 & 0 & 0 & -0 \\
0 & 0 & 0 & -0 \\
0 & 0 & 0 & -0 \\
0 & 0 & 0 & -0 \\
0 & -0.0032 & 0 & -0 \\
0 & 0 & -996 & -7.21 \times 10^{-6} \\
0 & 0 & -996 & -7.21 \times 10^{-6}
\end{pmatrix}
\end{array}
\tag{4.101}
$$

$$J_{asso} = \begin{array}{c} \\ A \\ AS \\ S \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ t \\ AST^{\dagger} \\ AST^{\ddagger} \\ Q \\ ASQ \\ ASQT \\ ASQT^{\dagger} \\ ASQT^{\ddagger} \\ ASTQ \\ ASTQ^{\dagger} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} A & AS & S & AS^{\dagger} & AS^{\ddagger} & AST & T & AST^{\dagger} & AST^{\ddagger} & Q & ASQ \end{array} \\ \begin{pmatrix} -100 & 0 & 0 & 0 & 0 & 0.003 & 0.0036 & 0 & 0 & 0 & 0 \\ 0 & -6.46 & 0 & 0 & 0 & 0 & 0 & 0 & 0.17 & 0 & 0 \\ 0 & 0 & -0.007 & 0 & 0 & 0 & 0 & 0 & 0.012 & 0 & 2 \\ 0 & 0 & 0 & -2.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -13.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0002 & 0 & 0 & 0 & 0 & -0.005 & 0 & 0 & 0 & 0 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.0036 & 0 & 0 & 0.02 & 0 \\ 100 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & -0 \\ 0 & 0.26 & 0.007 & 0 & 0 & 0 & 0 & 0 & -0.182 & 0 & 0 \\ 0 & 0 & 0 & 0 & 13.8 & 0 & 0 & 0 & 0 & -0.02 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.002 & 0 & 0 & 0 & 0 & -2.01 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0036 & 0 & 0 & 0 & -0 \\ 0 & 6.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0 \\ 0 & 0 & 0 & 0.01 & 0.0002 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$J_{asso}\ (continued) \quad \begin{array}{c} \\ A \\ AS \\ S \\ AS^{\dagger} \\ AS^{\ddagger} \\ AST \\ t \\ AST^{\dagger} \\ AST^{\ddagger} \\ Q \\ ASQ \\ ASQT \\ ASQT^{\dagger} \\ ASQT^{\ddagger} \\ ASTQ \\ ASTQ^{\dagger} \end{array} \begin{array}{c} \begin{array}{ccccc} ASQT & ASQT^{\dagger} & ASQT^{\ddagger} & ASTQ & ASTQ^{\dagger} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 996 & 0 & 4.73 \times 10^{-6} \\ -0.0021 & 0 & 0 & 0 & 0 \\ 0.0022 & 0 & 0 & 0.002 & 0 \\ 0 & 0 & 0 & 0.003 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.0022 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0021 & 0 & 0 & 0 & 0 \\ -0.0043 & 0 & 0 & 0 & -0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -996 & 0 & -4.73 \times 10^{-6} \\ 0 & 0 & 0 & -0.005 & 0 \\ 0 & 0 & -996 & 0 & -4.73 \times 10^{-6} \end{pmatrix} \end{array} \tag{4.102}$$

Time Course Simulations

Analogous to the reactions of Eq. (4.75) - (4.84), the system of ODEs for the dissociative model is defined by Eq. (4.103) - (4.115). All rate laws of these elementary reactions are modelled as mass action whereas the catalytic step of Eq. (4.81) and the three product release steps of Eq. (4.82) - (4.84) are considered quasi-irreversible in accordance with their reaction rate constants (Table 4.12).

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_{-1}[AS] - k_{+1}[A][S] \tag{4.103}$$

$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = k_{-1}[AS] - k_{+1}[A][S] \tag{4.104}$$

$$\frac{\mathrm{d}[AS]}{\mathrm{d}t} = k_{+1}[A][S] + k_{-2}[AS^{\dagger}] - k_{-1}[AS] - k_{+2}[AS] \tag{4.105}$$

$$\frac{\mathrm{d}[AS^{\dagger}]}{\mathrm{d}t} = k_{+2}[AS] + k_{-3}[AS^{\ddagger}] - k_{-2}[AS^{\dagger}] - k_{+3}[AS] \tag{4.106}$$

$$\frac{\mathrm{d}[AS^{\ddagger}]}{\mathrm{d}t} = k_{+3}[AS^{\dagger}] + k_{-4}[AST] + k_{+10}[ASQ^{\ddagger}] - k_{-3}[AS^{\ddagger}] - k_{+4}[AS^{\ddagger}][T] \tag{4.107}$$

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = k_{-4}[AST] - k_{+4}[AS^{\ddagger}][T] \tag{4.108}$$

$$\frac{\mathrm{d}[AST]}{\mathrm{d}t} = k_{+4}[AS^{\ddagger}][T] + k_{-5}[AST^{\dagger}] - k_{-4}[AST][T] - k_{+5}[AST] \tag{4.109}$$

$$\frac{\mathrm{d}[AST^{\dagger}]}{\mathrm{d}t} = k_{+5}[AST] + k_{-6}[AST^{\ddagger}] - k_{-5}[AST^{\dagger}][T] - k_{+6}[AST] \tag{4.110}$$

$$\frac{\mathrm{d}[AST^{\ddagger}]}{\mathrm{d}t} = k_{+6}[AST^{\dagger}] - k_{-6}[AST^{\ddagger}][T] - k_{+7}[AST] \tag{4.111}$$

$$\frac{\mathrm{d}[ASQ]}{\mathrm{d}t} = k_{+7}[AST^{\ddagger}] - k_{+8}[ASQ] \tag{4.112}$$

$$\frac{\mathrm{d}[ASQ^{\dagger}]}{\mathrm{d}t} = k_{+8}[ASQ] - k_{+9}[ASQ] \tag{4.113}$$

$$\frac{\mathrm{d}[ASQ^{\ddagger}]}{\mathrm{d}t} = k_{+9}[ASQ^{\dagger}] - k_{+10}[ASQ] \tag{4.114}$$

$$\frac{\mathrm{d}[Q]}{\mathrm{d}t} = k_{+10}[ASQ] \tag{4.115}$$

The system of ODEs for the associative model is defined as Eq. (4.116) - (4.131), where all rate laws of are modelled as mass action. The catalytic step of Eq. (4.81) and the three associative product release steps of Eq. (4.88) - (4.90) are considered quasi-irreversible in

accordance with their reaction rate constants (Table 4.12).

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_{-1}[AS] - k_{+1}[A][S] \tag{4.116}$$

$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = k_{-1}[AS] - k_{+1}[A][S] \tag{4.117}$$

$$\frac{\mathrm{d}[AS]}{\mathrm{d}t} = k_{+1}[A][S] + k_{-2}[AS^{\dagger}] - k_{-1}[AS] - k_{+2}[AS] \tag{4.118}$$

$$\frac{\mathrm{d}[AS^{\dagger}]}{\mathrm{d}t} = k_{+2}[AS] + k_{-3}[AS^{\ddagger}] - k_{-2}[AS^{\dagger}] - k_{+3}[AS] \tag{4.119}$$

$$\frac{\mathrm{d}[AS^{\ddagger}]}{\mathrm{d}t} = k_{+3}[AS^{\dagger}] + k_{-4}[AST] - k_{-3}[AS^{\ddagger}] - k_{+4}[AS^{\ddagger}][T] \tag{4.120}$$

$$\frac{\mathrm{d}[T]}{\mathrm{d}t} = k_{-4}[AST] + k_{-11}[ASQT] - k_{+4}[AS^{\ddagger}][T] - k_{+11}[ASQ][T] \tag{4.121}$$

$$\frac{\mathrm{d}[AST]}{\mathrm{d}t} = k_{+4}[AS^{\ddagger}][T] + k_{-5}[AST^{\dagger}] - k_{-4}[AST] - k_{+5}[AST] \tag{4.122}$$

$$\frac{\mathrm{d}[AST^{\dagger}]}{\mathrm{d}t} = k_{+5}[AST] + k_{-6}[AST^{\ddagger}] - k_{-5}[AST^{\dagger}] - k_{+6}[AST] \tag{4.123}$$

$$\frac{\mathrm{d}[AST^{\ddagger}]}{\mathrm{d}t} = k_{+6}[AST^{\dagger}] + k_{+16}[ASTQ^{\ddagger}] - k_{-6}[AST^{\ddagger}][T] - k_{+7}[AST] \tag{4.124}$$

$$\frac{\mathrm{d}[ASQ]}{\mathrm{d}t} = k_{+7}[AST^{\ddagger}] + k_{-11}[ASQT] - k_{+11}[ASQ][T] \tag{4.125}$$

$$\frac{\mathrm{d}[ASQT]}{\mathrm{d}t} = k_{+11}[ASQ][T] + k_{-12}[ASQT^{\dagger}] - k_{-11}[ASQT] - k_{+12}[ASQT] \tag{4.126}$$

$$\frac{\mathrm{d}[ASQT^{\dagger}]}{\mathrm{d}t} = k_{+12}[ASQT] + k_{-13}[ASQT^{\ddagger}] - k_{-12}[ASQT^{\dagger}] - k_{+13}[ASQT] \tag{4.127}$$

$$\frac{\mathrm{d}[ASQT^{\ddagger}]}{\mathrm{d}t} = k_{+13}[ASQT^{\dagger}] - k_{-13}[ASQT^{\ddagger}] - k_{+14}[ASQT] \tag{4.128}$$

$$\frac{\mathrm{d}[ASTQ]}{\mathrm{d}t} = k_{+14}[ASQT^{\ddagger}] - k_{+15}[ASTQ] \tag{4.129}$$

$$\frac{\mathrm{d}[ASTQ^{\dagger}]}{\mathrm{d}t} = k_{+15}[ASTQ] - k_{+16}[ASTQ] \tag{4.130}$$

$$\frac{\mathrm{d}[Q]}{\mathrm{d}t} = k_{+16}[ASTQ] \tag{4.131}$$

Neglecting target mRNA synthesis, on the long term, target mRNA concentration during time course simulations tends towards zero for both models, as described by the steady state analysis above and similar to the previous modelling approaches (described in Sections 4.1 and 4.2, respectively). Nonetheless, the revised models comprise more precise kinetic rate constants based on precise quantitative kinetic data (from Table 4.12). If the model structure has been selected well and if the modelled biological mechanism holds true, then time

course simulations starting with physiological meaningful initial conditions, i.e. concentrations for hAgo2 and target mRNA, should lead to similar results as observed experimentally.

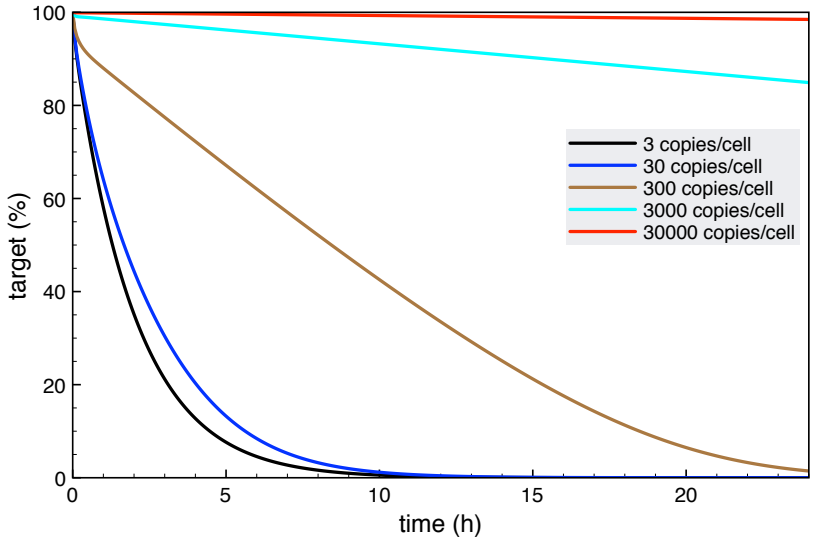
The initial target mRNA concentration is varied between 0.03 and 300 nM (i.e. 3 to 30,000 copies/cell) at concentrations of 0.3 nM siRNA and 130 nM Ago2. Given the dissociative model, the difference in RNAi efficacy at different target concentrations is strongly pronounced (Fig. 4.15A). Furthermore, there is a great similarity in the time courses to the previous modelling approach of ss siRNA-mediated RNAi (Section 4.1, Fig. 4.3A). Whereas in the case of the associative model the time courses resulting from different initial target concentrations differ noticeable between the two modelling approaches (i.e. previous model: Fig. 4.3B & revised model: Fig. 4.15B). Most strikingly, RNAi efficacy in the revised model stays constantly high over three decimal powers of target concentrations.

Trajectories of intermediates

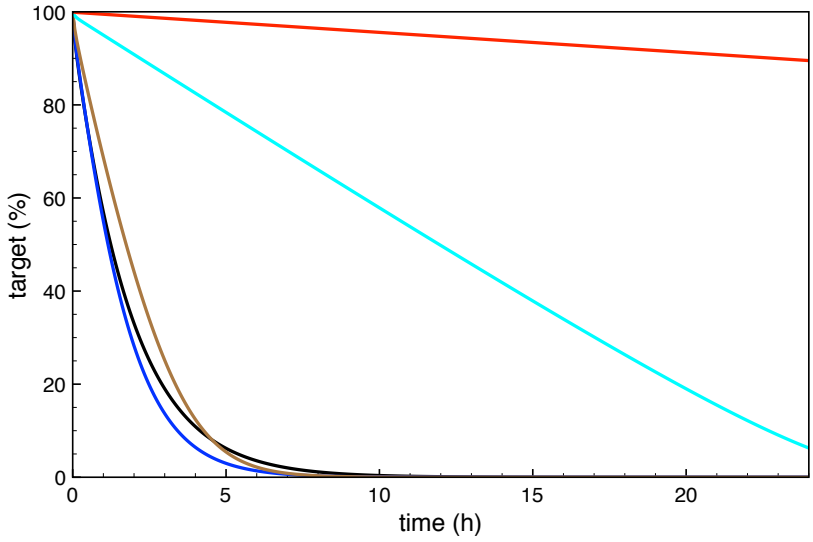
When looking at the trajectories of modelled intermediate complexes formed during RNAi, the similarities and differences between the two competing models become visible:

Dissociative model

Fig. 4.16 shows the time courses of intermediate complexes of three simulations with initial target concentration of 300 (A-C), 3,000 (D-F) and 30,000 (G-I) copies/cell, respectively. The initial concentrations of siRNA and Ago2 were set to 300 and 200,000 copies/cell, respectively. Each row accounts for one of the three time course simulations. The time scale of each run is separated: the 1^{st} column shows time courses between zero and 30 s, the 2^{nd} column shows zero to 600 s, and the 3^{rd} shows zero to 3,500 s.
All three simulations start with fast occurring collisions of siRNA and Ago2, where reaction step target binding I is rate limiting only within fractions of a second. This results in a peak of collision complex AS and its conversion into intermediate $AS^{\dagger}$ within the first 5 s. AS and $AS^{\dagger}$ dam up due to a slower third step of siRNA binding and saturation of the RNAi machinery with siRNA due to slower downstream reaction steps. The subsequently formed $AS^{\ddagger}$ intermediate is consumed quickly due to fast collision with free target, resulting in only slow accumulation of this intermediate. The slow increase later during the time course is due to a lack of sufficient amounts of free target necessary for the next reaction step, indicating an excess in free siRNA over free target. For instance, for the simulation with an initial 300 target copies/cell, this occurs faster and is more distinct compared to simulations starting with higher target concentrations. The fast occurring collision between



(a)



(b)

Figure 4.15 **The influence of target mRNA on time courses (revised models of siRNA-mediated RNAi).** The initial target mRNA concentration is varied between 0.03 and 300 nM ($\sim$ 3 to 30,000 copies/cell) at concentrations of 0.3 nM siRNA and 130 nM Ago2. Trivially, RNAi efficacy decreases with higher target concentrations. **(a)** Given the dissociative model, the difference in RNAi efficacy at different target concentrations is more distinct. **(b)** Whereas in the case of the associative model RNAi efficacy stays constantly high over three decimal powers of target concentrations.

RISC (i.e. $AS^{\ddagger}$) and free target builds up a surplus of collision complex AST within the first 120 s of the simulation, which is followed by another surplus of intermediate $AST^{\dagger}$ with a ~ 200 s lag, indicating that target binding II is rate limiting only within the first 200 s of the simulations. The slow target binding step III limits the RNAi cycle for 200 s to 580, which is extended for lower initial target concentrations (e.g. to 720 s in the case of 300 target copies/cell). There is no accumulation of intermediates $AST^{\ddagger}$ and ASQ during the time courses. Conclusively, target cleavage and subsequent target release step I are not rate limiting at any point of the investigated simulations. Intermediate $ASQ^{\dagger}$ dams up quickly from about 80 to 100 s, while accumulation of $ASQ^{\ddagger}$ is delayed by 120 s. Starting from around $1,000$ s, this accumulation of $ASQ^{\ddagger}$ via slow product release step III dominates the rest of the time courses independent of initial target concentration.
To sum up, $ASQ^{\ddagger}$ continuously increases until reaching its maximum concentration at steady state. The concentrations of all other intermediates peak early (as described in detail above) and decrease towards their steady state concentrations. This leads to $ASQ^{\ddagger}$ being the dominant intermediate from 800 to 900 s after simulation initialisation due to a rate limiting effect of the slow reaction step product release III after about 900 s (15 min).
There is little qualitative or quantitative difference between between the trajectories stating with $3,000$ or $30,000$ target copies/cell, save for the already mentioned $AS^{\ddagger}$, which tends to accumulate in the former case due to a lack of excess of free target (i.e. at low target concentrations).

Associative model

Analogous, Fig. 4.17 shows the time courses of intermediate complexes for the associative model.
The initial reaction steps, covering siRNA binding and target binding, tend to have similar effects on the overall time courses of RNAi intermediates in both models. Please refer to the previous paragraph for details. All simulations start with fast occurring collisions of siRNA and Ago2, where reaction step target binding I is rate limiting only within fractions of a second. AS and $AS^{\dagger}$ dam up due to a slower third step of siRNA binding and saturation of the RNAi machinery with siRNA due to slower downstream reaction steps. The subsequent formed $AS^{\ddagger}$ intermediate is consumed quickly due to fast collision with free target, resulting in only slow accumulation of this intermediate. The slow increase later during the time course is due to a lack of sufficient amounts of free target necessary for the next reaction step, indicating an excess in free siRNA over free target. For instance, for the simulation with an initial 300 target copies/cell, this occurs faster and is more distinct compared to

simulations starting with higher target concentrations. The fast occurring collision between RISC (i.e. $AS^{\ddagger}$) and free target builds up a surplus of collision complex AST within the first 120 s of the simulation, which is followed by another surplus of intermediate $AST^{\dagger}$ with a ~ 200 s lag, indicating that target binding II is rate limiting only within the first 200 s of the simulations. The slow target binding step III limits the RNAi cycle for 200 s to 580, which is extended for lower initial target concentrations (e.g. up to 720 s in the case of 300 target copies/cell). There is no accumulation of intermediate $AST^{\ddagger}$ during the time courses. Conclusively, target cleavage is not rate limiting at any point of the investigated simulations. At low initial target concentrations (e.g. 300 copies/cell or lower), ASQ is becoming the most prominent intermediate for time points greater than $1,000$ s. This is due to a lack of a supply of free target for associative target binding step I. At sufficiently high initial concentrations of free target the accumulation of $AST^{\ddagger}$ is avoided and associative target binding step I is not rate limiting at any point of the investigated simulations. The accumulation of $ASQT^{\dagger}$ is anti-proportional to initial target concentration, indicating a direct link between the adaptation of RNAi efficacy to varying levels of target mRNA and the coalition complex formation between target and RISC comprising cleaved target, i.e. ASQ. This target-concentration-dependent acceleration during coalition complex formation is limited by a slow third step of associative target binding. In comparison with the dissociative model, product release is at no time the main limiting factor during the investigated simulations. Even though a prominent accumulation of $ASTQ^{\dagger}$ can be observed throughout all simulations, it is shadowed by a higher concentration of $ASQT^{\dagger}$ (due to a slow associative target binding step III), and for lower initial target concentrations, additionally by ASQ (due to a slow associative target binding step I, in terms due to the low initial target concentration itself).

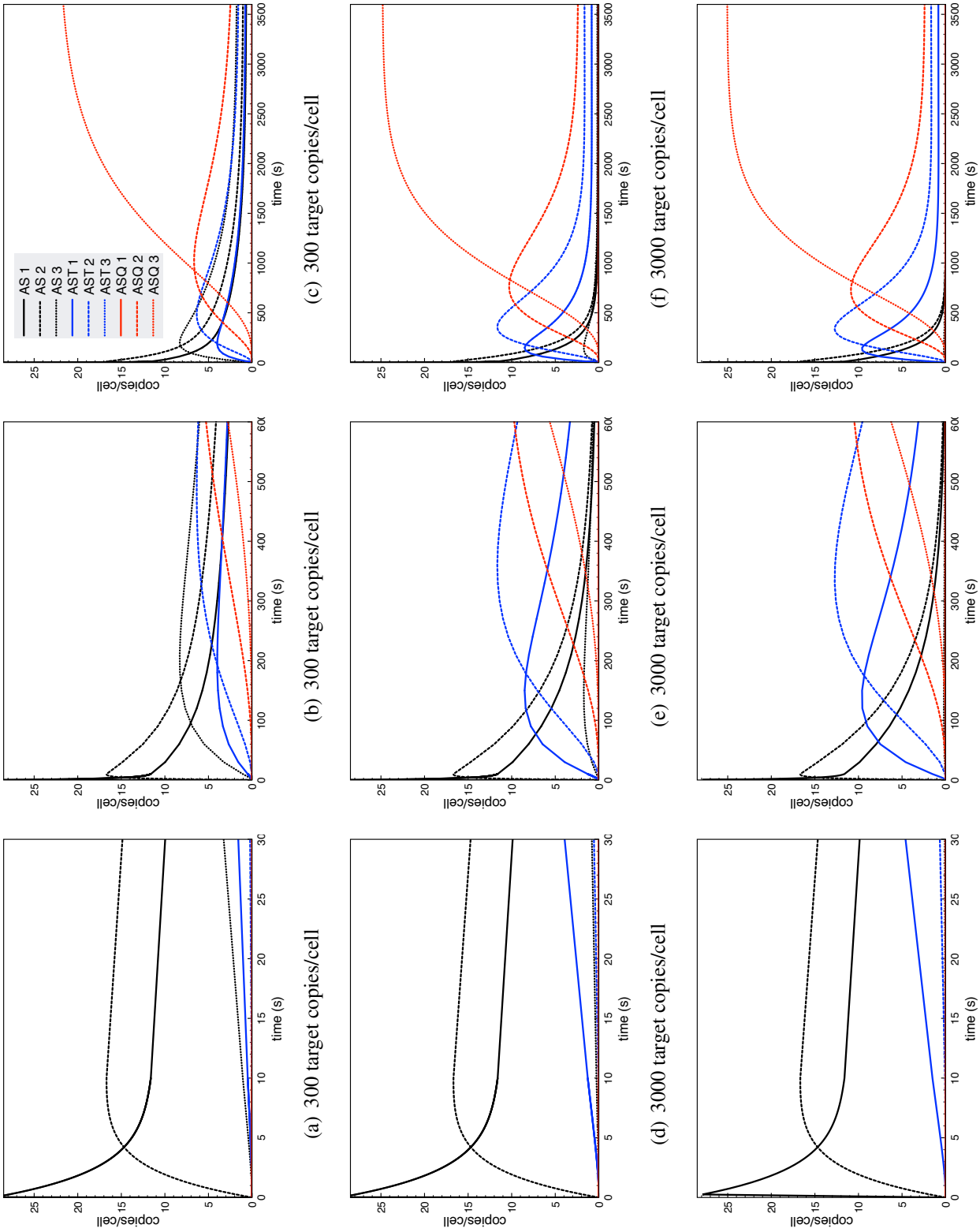
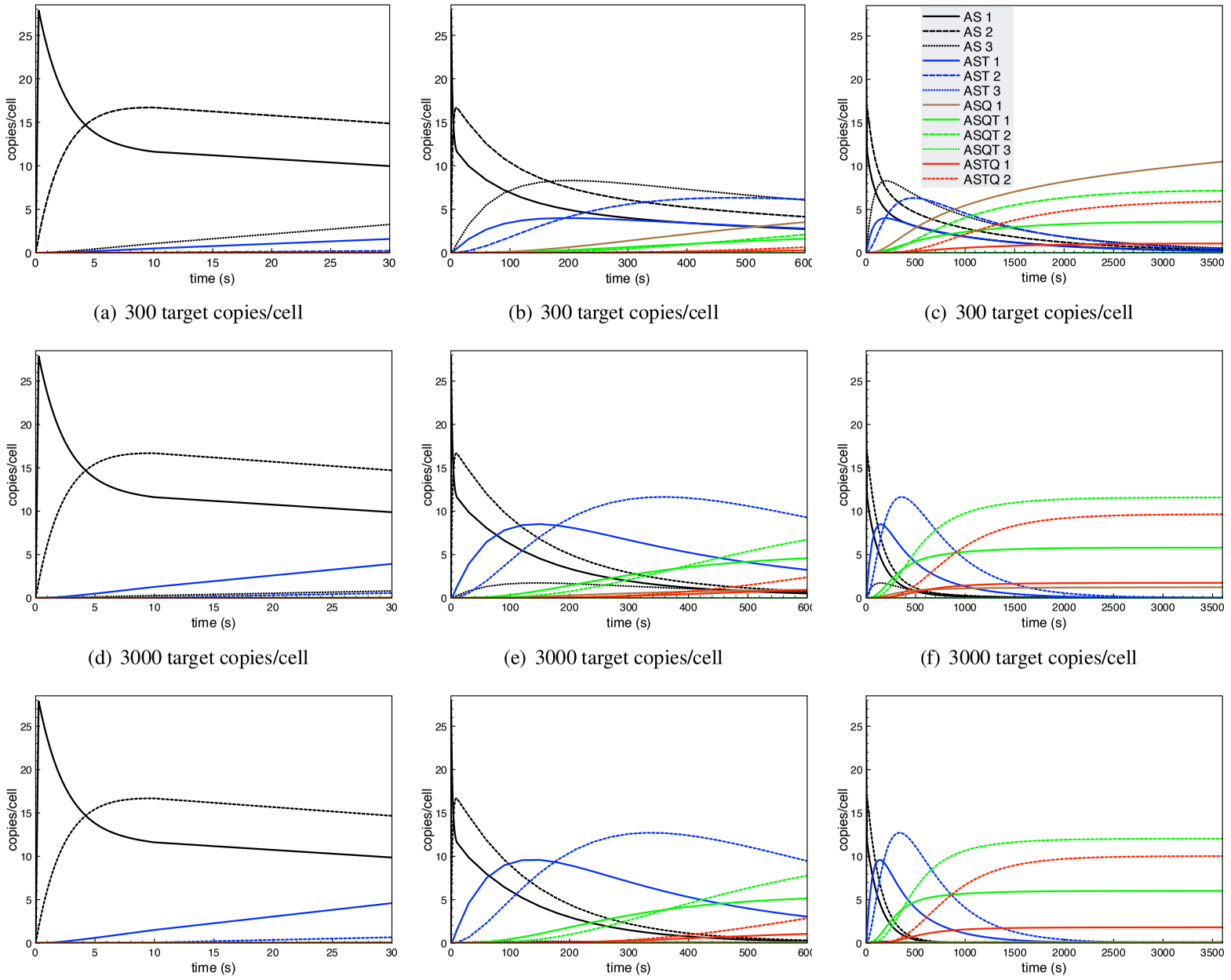


Figure 4.16 **Trajectories of intermediates for the dissociative model.** Initial target: **A-C)** 300, **D-F)** 3,000, **G-I)** 30,000 copies/cell; siRNA: 300 copies/cell; Ago2: 200,000 copies/cell. Time scale: 1^{st} column: 0 - 30 s, 2^{nd}: 0 - 600 s, 3^{rd}: 0 - 3,500 s.



(a) 300 target copies/cell (b) 300 target copies/cell (c) 300 target copies/cell

(d) 3000 target copies/cell (e) 3000 target copies/cell (f) 3000 target copies/cell

(g) 30000 target copies/cell (h) 30000 target copies/cell (i) 30000 target copies/cell

Figure 4.17 **Trajectories of intermediates for the associative model.** Initial target: **A-C)** 300, **D-F)** 3,000, **G-I)** 30,000 copies/cell; siRNA: 300 copies/cell; Ago2: 200,000 copies/cell. Time scale: 1^{st} column: 0 - 30 s, 2^{nd}: 0 - 600 s, 3^{rd}: 0 - 3,500 s.

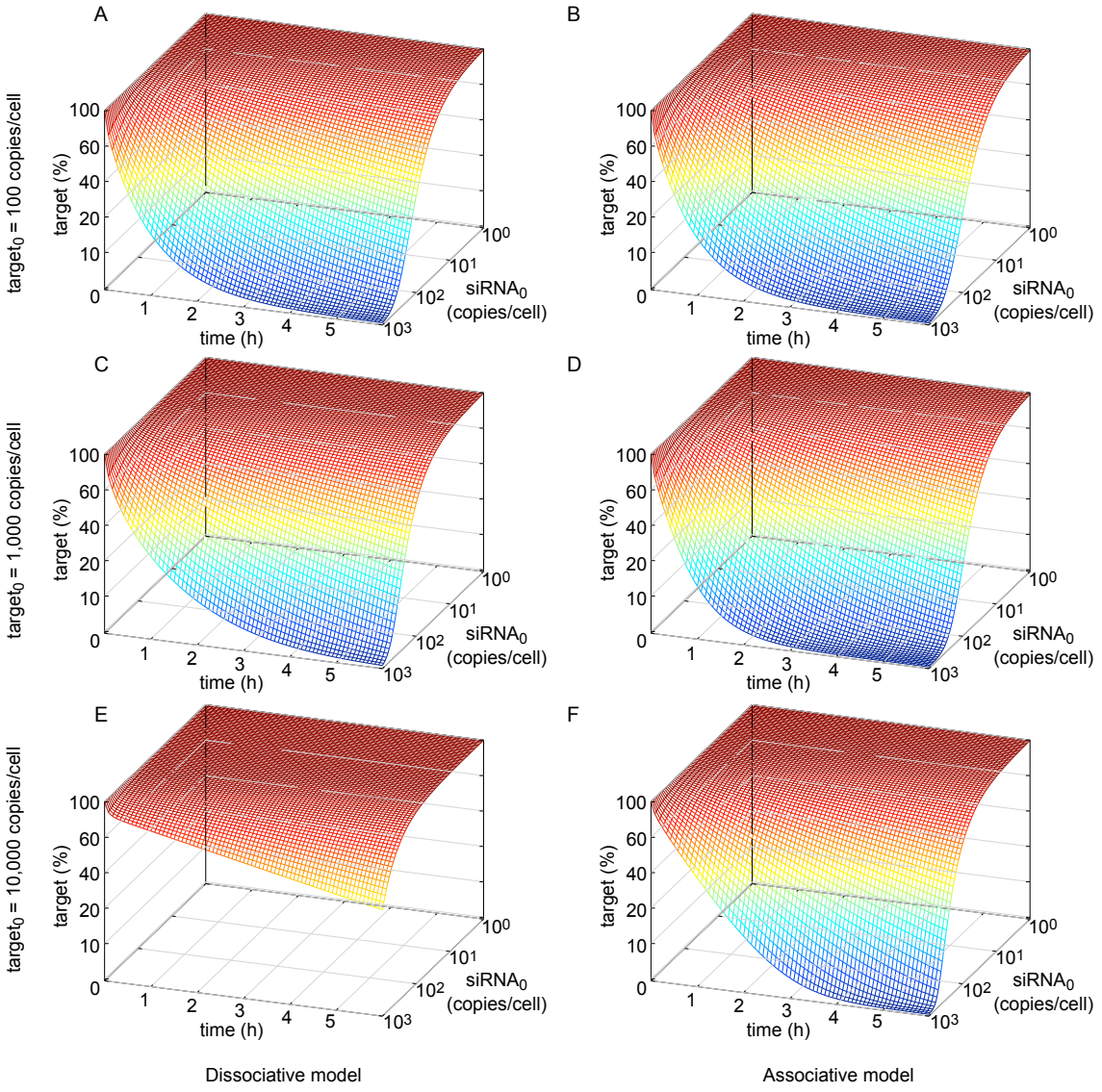


Figure 4.18 **Time resolved siRNA-mediated target knockdown with respect to initial siRNA concentration (revised models).** In these 3D-plots, target concentration (% of basal target level) is plotted against reaction time (h) and logarithmically-scaled initial siRNA concentration (copies/cell). Simulation results of the dissociative (1^{st} column and associative (2^{nd}) models are presented in parallel. Row-wise, absolute basal target levels increase by 10-fold from 100 to 1,000 to 10,000 copies/cell. A difference in time-dependent model behaviour between dissociative and associative models is recognisable: At lower target level, i.e. 100 copies/cell, the RNAi machinery in **A)** the dissociative and **B)** the associative model responds likewise to several magnitudes of siRNA with respect to time resolved target knockdown. **C)** However, in the case of the dissociative model, the RNAi machinery requires increased siRNA concentrations to maintain the rate of target knockdown at a higher target level, i.e. 1,000 copies/cell. Description continues on the next page.

Fig. 4.18, **continued. Time resolved siRNA-mediated target knockdown with respect to initial siRNA concentration (revised models).**
D) For the associative model, the relative time courses of target knockdown are far more similar for different levels of basal target concentrations. **E)** In the case of the dissociative model, the RNAi machinery requires high siRNA concentrations > 100 copies/cell to achieve target knockdown at a basal target level of $10,000$ copies/cell. **F)** For the associative model, an effect can be observed with siRNA concentration of 10 copies/cell. RNA levels in mammalian cells vary in magnitude between 1 and 100,000 copies/cell[237] and thus, RNAi-based knockdown must be able to adapt to several magnitudes of target mRNA levels as in the associative model. This figure is based on Figure 3 of[40].

4.5.4 Model validation using minimal *in vitro* RNAi assay

Both models are able to quantitatively match the experimentally observed time course. Thereby, they require different concentrations of Ago2 to do so, and even then their trajectories qualitatively look rather different. While the dissociative model better matches late time points, the associative model closely fits the early pre-steady-state kinetics below 1 h. Its deviation from the 2 h data point can be explained by the increasing inactivation of Ago2 within the *in vitro* assay, which is not addressed in any of the models. However, an ultimate decision for (or against) one of the models cannot be drawn, as the concentration of active Ago2 is unknown but estimated to be rather low (< 1 nM), both models may pass the validation. Compare Fig. 4.19.

4.5.5 Model validation using cytosol-extract-based RNAi assay

As described on p. 66, the second set of validation data is based on NHA-Ago2 assays in cytosolic HeLa cell extract (Carsten Geist, unpublished data), which was prepared according to[35, 140]. The cytosolic extract should contain all, or at least many, of the cytoplasm's components, although diluted by a factor $\frac{1}{3}$. It may be seen, at least to a limited extent, as a snap shot of the cytoplasm of living cell at the time of extraction. Thus it is a more realistic scenario for model validation compared to the previously described *in vitro* assay, as many known and unknown RNAi cofactors (as well as other influences) are likely to be present in extract. The first data set (used as training set) consists of six sparse time series starting at different initial concentrations siRNA between 0.3 and 100 nM and constant concentrations of 2.4 nM target mRNA. The second set (the validation set) consists of a time series with 11 time points starting at 4 nM target and 100 nM siRNA. The initial

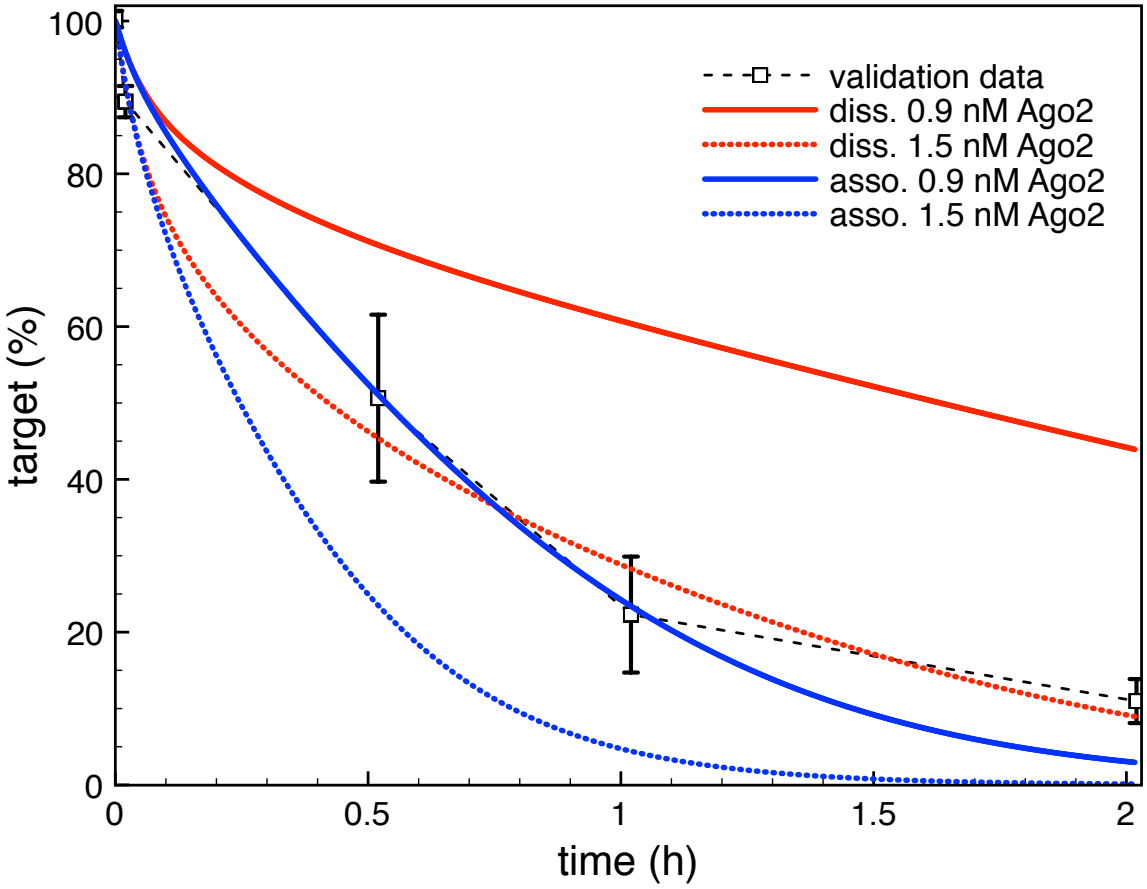


Figure 4.19 **Model validation using *in vitro* ss siRNA-based assay (revised models).** The initial target concentration is 2.5 nM at an abundance of 100 nM siRNA.

Ago2 concentration in the assay is not known. It is lower than *in vivo* due to dilution, extraction efficiency, or degradation during preparation and storage. It may vary between $\frac{1}{3}$ and $\frac{1}{10}$ of the original cellular concentration (communication with Juliane Neubert). Time course simulations starting with the same initial concentrations for hAgo2, target RNA and siRNA used in experiments, should lead to similar time courses as in the HeLa cytosol assay. Table 4.14 shows the model fits to the training data with free parameter $[A]_0$ (initial Ago2 concentration). Table 4.15 shows the resulting best fit values for $[A]_0$. As with the first model approach (Section 4.1.2, pp. 66), the best fit values for the unknown parameter of initial Ago2 concentration are very similar between the dissociative and the associative models. In contrast to the earlier modelling approach, this parameter value lies within 1.13 and 3.52 $\times 10^{-9}$ M, which is a more realistic ballpark figure with respect to pre-knowledge on the experimental set up. That is, the Ago2 concentration in cytosolic extract must be smaller than the cellular hAgo2 concentration in HeLa, which is 1.33×10^{-7} M or 200,000 copies/cell. However, this alone is not a confirmation of the modelling approach. It points towards the approach's physiological meaningfulness. Not least because the predicted range of Ago2 concentration is about two magnitudes smaller than the cellular concentration, where a reduction of between $\frac{1}{3}$ and $\frac{1}{10}$ was expected (communication with Juliane Neubert). Comparing the objective values between the current modelling approach and the one of Section 4.1.2, Table 4.3, clearly shows an improvement with the changes introduced with the new modelling approach (Table 4.14). Furthermore, in the new approach there is a clear difference in objective value between dissociative and associative models with a factor of 7.6 in favour of the associative model. Save for the last experiment (starting with 100 nM siRNA), all objective values and their RMSD are relatively equally distributed between the experiments. The fits to the experiments are separately visualised in Fig. 4.20.

4.5.6 RNAi in living mammalian cells suggests an alternative model structure

In a first attempt to test the model's relevance in living mammalian cells, time-course simulations of target knockdown were compared to cell culture experiments. In Fig. 4.21A, $t_{1/2}$ spans the surface of the parameter space of the two initial concentrations, siRNA $[S]_0$ and target $[T]_0$. In general, $t_{1/2}$ is proportional to $[T]_0$ and anti-proportional to $[S]_0$. It is also evident that $t_{1/2}$ is very sensitive to changes in $[S]_0$ and less sensitive to changes in $[T]_0$. Fig. 4.21B shows a cross section of the 3D-plot in Fig. 4.21A, at a basal target concentration

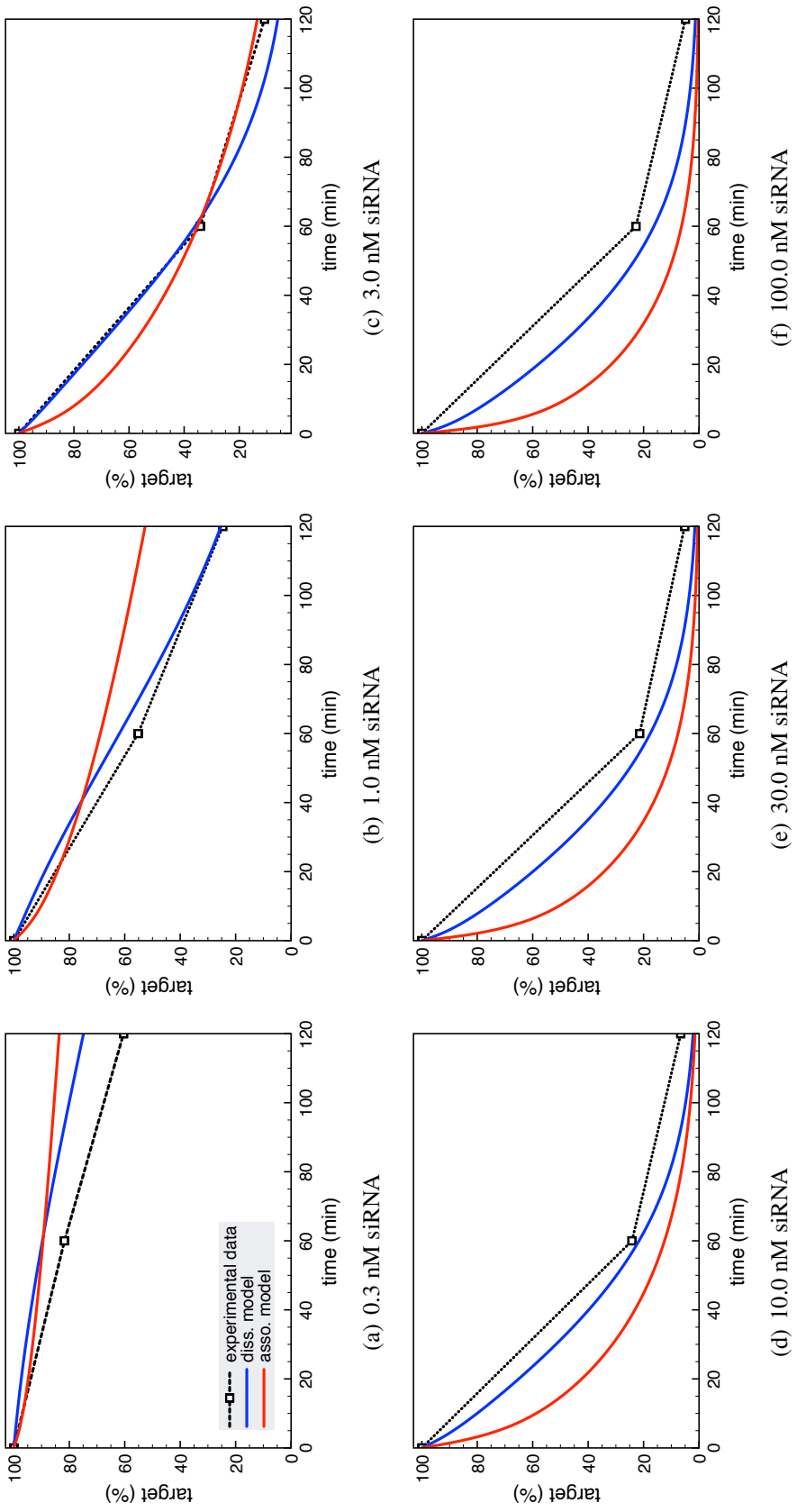


Figure 4.20 **Model validation using cytosol-extract-based RNAi assay (revised models).** Each sub-figure shows the time-resolved target knockdown given varying initial siRNA concentrations between **A**) 0.3 nM and **F**) 100 nM. Time course simulation of the two competing models (blue and red lines for the dissociative and the associative models, respectively) are compared to experimentally measurements (dashed black lines).

initial concentrations		target		diss. model		asso. model	
target	siRNA	at 1 h	at 2 h	objective	RMSD	objective	RMSD
2.4 nM	0.3 nM	81.73%	60.43%	597.8	14.12	29.96	3.16
2.4 nM	1 nM	55.17%	24.67%	968.79	17.97	46.01	3.92
2.4 nM	3 nM	34.14%	10.80%	7.96	1.63	26.55	2.97
2.4 nM	10 nM	24.17%	6.59%	163.84	7.39	26.27	2.96
2.4 nM	30 nM	21.44%	5.18%	214.06	8.45	28.4	3.08
2.4 nM	100 nM	22.76%	4.90%	91.30	7.14	111.49	14.62
average				340.63	9.45	44.78	5.12

Table 4.14 **Model validation: objective values and RMSD (revised models).** Shown are objective values and corresponding RMSD for each validation data set, as well as, the resulting average values. Validation is based on 100 repetitions of model fits with free parameter $[A]_0$ (initial Ago2 concentration) to the training data based on cytosolic extract HeLa assays. For models, objective values and RMSD are close by. There is a large variation between tuples of objective values and RMSD of the independent data sets.

Model	lower bound	upper bound	start value	value	S.D.
diss.	10^{-12}	10^{-4}	random	3.52×10^{-9}	5.60×10^{-10}
asso.				1.13×10^{-9}	3.62×10^{-11}

Table 4.15 **Model validation: fitted parameter (revised models).** Shown are the predicted value for free parameter $[A]_0$ (initial Ago2 concentration) to the training data based on 100 repetitions of model fit to cytosolic extract HeLa assays. For both models, the resulting fitted values are close by.

of $4,000$ copies/cell, which corresponds to the average basal target mRNA concentration in the control experiment. A first discrepancy between dissociative model and experimentally determined values (grey squares) was observed. It was not possible to eliminate the discrepancy between model and experiment by merely adjusting the model parameters within physiologically meaningful boundaries. For example, a two-fold increase of the measured *in vitro* constants of the two most sensitive parameters (including the rate-limiting step), which may be justified by facilitating co-factors *in vivo*, was not sufficient to solve the discrepancy.

4.5.7 The associative model closely resembles siRNA-mediated target knockdown in cell culture

Time-course simulations were repeated with the associative model and $t_{1/2}$ of target knockdown was again compared to the experimental data. In Fig. 4.21A, $t_{1/2}$ of dissociative and

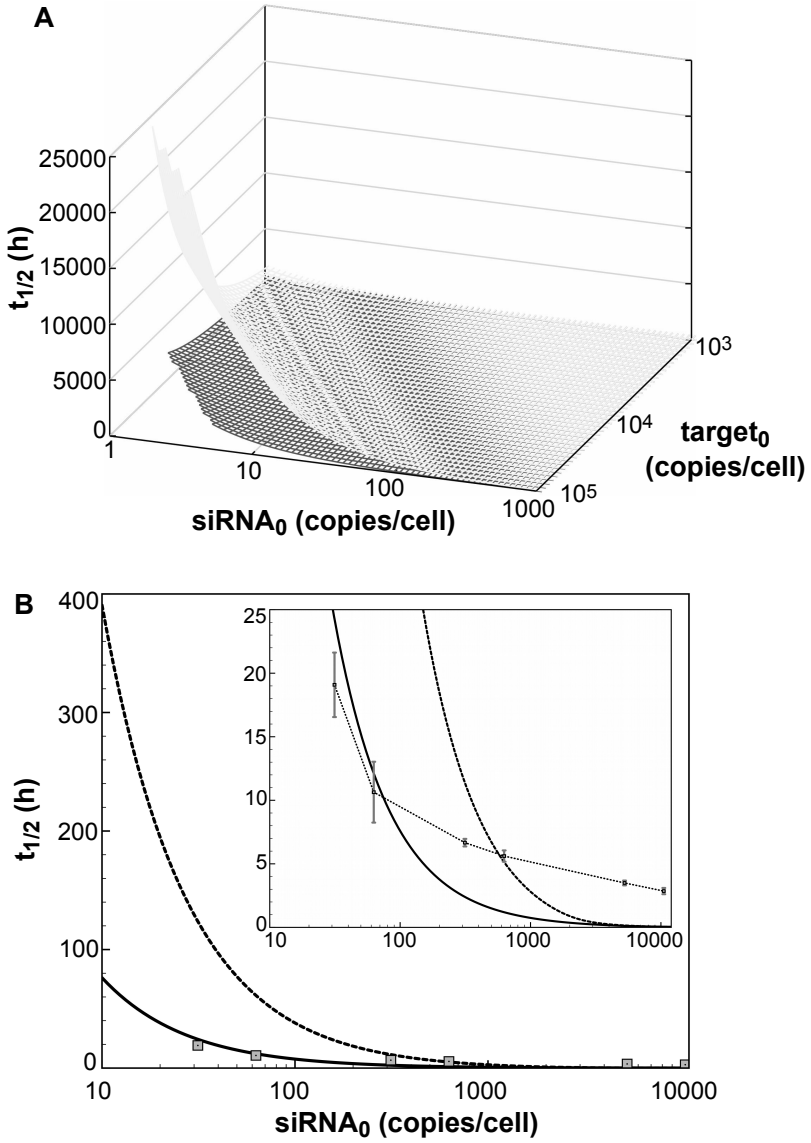
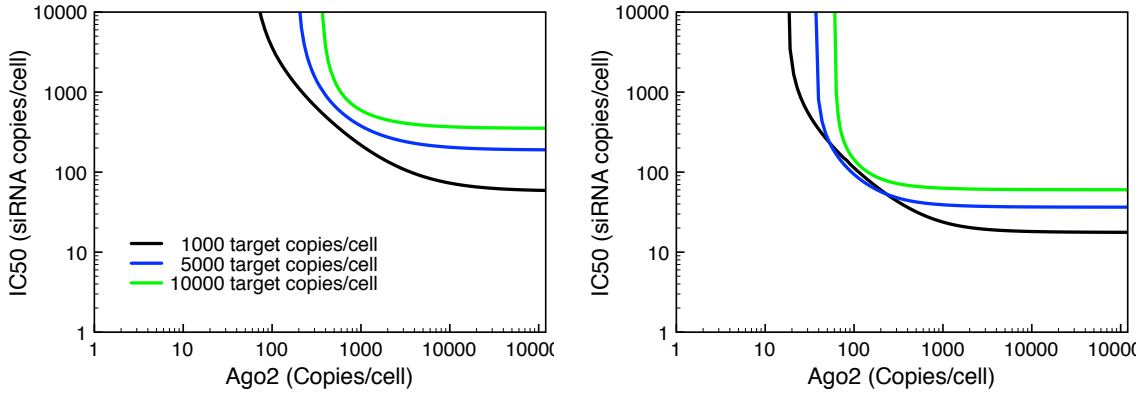


Figure 4.21 **$t_{1/2}$: the associative model closely resembles siRNA-mediated target knockdown in cell culture (revised models). A)** $t_{1/2}$, the time of half-maximal target knockdown, is plotted against logarithmically-scaled initial siRNA concentration and logarithmically-scaled basal target concentration. $t_{1/2}$ of dissociative and the associative model span the light grey and the black surface over the parameter space, respectively. For a combination of large $siRNA_0$ and low $target_0$, both surfaces superimpose due to similarities in $t_{1/2}$. They increasingly diverge with higher $target_0$ concentration and to a lesser extent with lower $siRNA_0$ concentration. **B)** shows a cross section of the 3D-plot at a basal target concentration of $4,000$ copies/cell. For decreasing siRNA concentrations the difference in $t_{1/2}$ between the dissociative (broken line) and the associative (solid line) model increases progressively. Experimentally determined $t_{1/2}$ values (grey squares with S.D.) of LF-transfected siR206, which loosely correspond to 31 and 314 bioavailable siRNA copies/cell, respectively, closely match the behaviour of the associative model. This figure is based on Figure 2 of[40].

the associative models span the grey and the black surface, respectively, over the parameter space of the two initial concentrations, $target_0$ and $siRNA_0$. For both models, $t_{1/2}$ is proportional to $target_0$ and anti-proportional to $siRNA_0$. Here, it is also evident that in both models $t_{1/2}$ is very sensitive to changes in $siRNA_0$ and less sensitive to changes in $target_0$. For a combination of large $siRNA_0$ and low $target_0$, both surfaces superimpose due to similarities in $t_{1/2}$. They increasingly diverge with $target_0$ concentration and to a lesser extent with $siRNA_0$ concentration. Experimentally determined $t_{1/2}$ values (grey squares in Fig. 4.21B), which correspond to 31 and 314 siRNA copies/cell, respectively, closely match the behaviour of the associative model.

4.5.8 Time resolved siRNA-mediated target knockdown with respect to initial siRNA concentration at three different levels of basal target mRNA

After administration of siRNA, target mRNA concentration and, as a consequence, gene activity decrease within a defined time span until reaching minima. In cell culture, recovery of mRNA concentration back to basal levels takes several (2-4) days and is likely to be controlled by dilution of siRNA or loaded RISC due to cell division. In the following, the time-resolved knockdown of target mRNA concentration as a function of administered siRNA dose, as well as, of basal mRNA level is analysed on a short term time scale of 6 h. Ago2 concentrations are 250,000 - 500,000 copies/cell[179, 217] and do not limit model performance down to 10,000 or 50,000 copies/cell for the associative and dissociative model, respectively (Fig. 4.22). Thus initial Ago2 concentration was not of concern during the following simulations, where a number of time course simulations with different combinations of administered siRNA dose and basal mRNA level is performed for both models. Fig. 4.18 shows the time resolved on-set of target knockdown within the first 6 h after siRNA administration as a function of initial siRNA concentration, where relative target concentration is given as function of reaction time and initial siRNA concentration. At lower target levels, i.e. 100 copies/cell, the RNAi machineries of dissociative and associative model, respond likewise to several magnitudes of siRNA (Fig. 4.18A&B). However, in the case of the dissociative model, the RNAi machinery requires increased siRNA concentrations to maintain the rate of target knockdown at higher target level, i.e. 1,000 copies/cell (Fig. 4.18C&D). For the associative model the relative time courses of target knockdown are far more similar for different levels of basal target concentrations.This difference in time-dependent model behaviour between dissociative and associative models becomes more recognisable at even



(a) Dissociative model

(b) Associative model

Figure 4.22 **Relation between Ago2 concentration and IC_{50} (revised models). A)** For the dissociative and **B)** the associative models, logarithmically-scaled IC_{50}, the half maximal inhibitory siRNA concentration after 24 h of knockdown, is plotted against logarithmically-scaled Ago2 concentration. Basal target concentrations are $1,000$ (black), $5,000$ (blue) and $10,000$ (green lines) copies/cell, respectively. This figure is based on Supplementary Figure S4 of[40].

higher target levels of 10,000 copies/cell (Fig. 4.18E&F). RNA levels in mammalian cells vary in magnitude between 1 and 10,000 copies/cell[212, 237] and thus, RNAi-based knockdown should be able to adapt to several magnitudes of target mRNA levels, achieved by the associative, but not the dissociative model.

4.5.9 Adaptation of the RNAi machinery against variations in target mRNA level indicated by IC_{50} values

To further investigate the adaptiveness of the RNAi system to changes in target mRNA levels, IC_{50} is investigated over several magnitudes of basal target mRNA levels. For the dissociative (broken blue line) and the associative (solid black line) model, IC_{50} is plotted against basal target mRNA concentration and compared to experimentally derived IC_{50} (red squares), compare Fig. 4.23. In the case of experiment, IC_{50} is relatively constant over three magnitudes of transfected target plasmid (0.15 - 150 ng/μl). This indicates a robustness of the RNAi machinery against a certain extent of variation in target mRNA level. For the dissociative model, the IC_{50} stays relatively constant for small target concentrations < 50 copies/cell, then increases rapidly. For the associative model, IC_{50} slowly decreases with increasing target concentration until it reaches its optimum between 200 and 300 copies/cell, from where it starts increasing. All in all, the associative model can adapt more efficiently

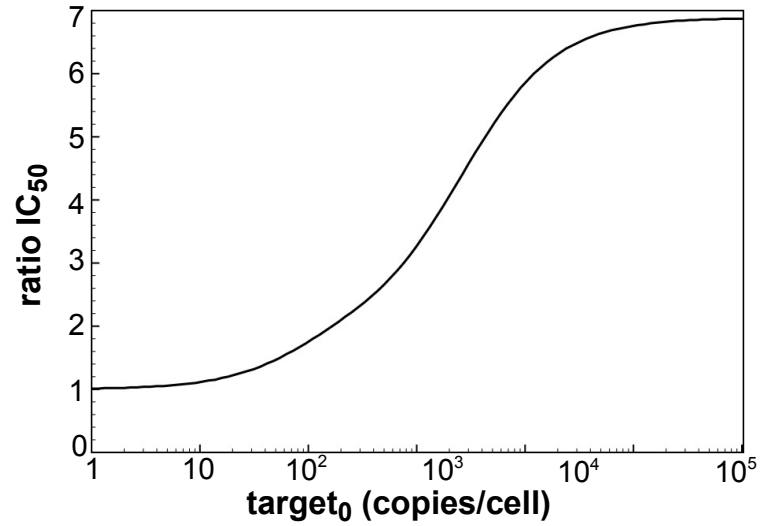
to variations in target mRNA level and thus resembles the experimental data closer than the dissociative model.

4.5.10 Sensitivity of time-resolved target knockdown to model parameters

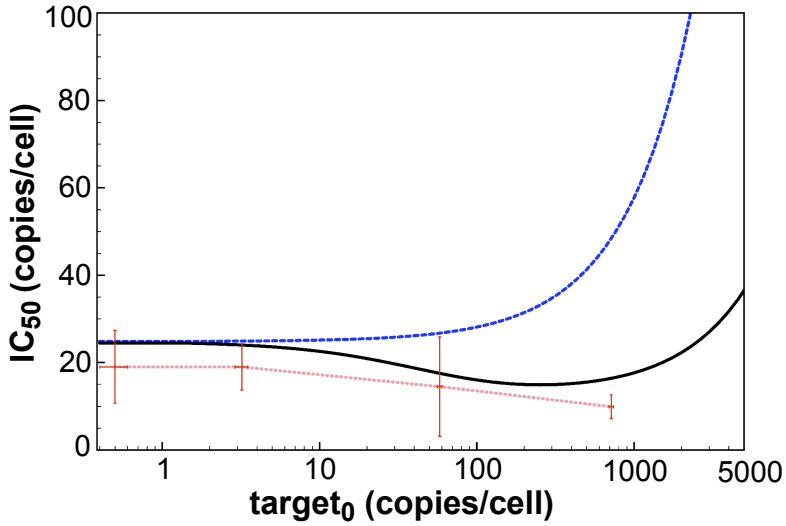
The sensitivity analysis describes how small deviations of a specific model parameter changes the overall behaviour of the model. If a parameter is found not to affect the system considerably, a rough guess of its value may be sufficient. However, if on the other hand a parameter influences the behaviour of the model substantively, steps must be taken to determine its value more accurately. Furthermore, sensitivity analysis gives information about which parameters should be changed (or kept as constant as possible) to achieve a specific effect (or avoid a specific effect). Additionally, the robustness of a system against external influences, i.e. transfection of siRNA, target levels, target metabolism etc. can be determined via sensitivity analyses.
More precisely, sensitivity coefficients ε_i^k quantify the sensitivity of target knockdown efficacy to small parameter changes as a function of basal target concentration over a 24 h time course. Investigated parameters are the kinetic rate constants, initial species concentrations and cytoplasm volume. The results of the sensitivity analysis are visualised in Fig. 4.24.

The dissociative model is sensitive to changes in reaction rate constants of binary complex formation (k_{+2}, k_{-2}, k_{+3}, and k_{-3}), as well as ternary complex formation (k_{+4}, k_{-4}, k_{+5}, k_{-5}, and k_{+6}). These parameters limit knockdown efficacy for low target concentrations up to $1,000$ or $2,000$ copies/cell (for 30 and 300 siRNAs/cell, respectively). At higher target concentrations the knockdown efficacy is practically zero (no target knock-down after 24 h); furthermore the sensitivities are zero for all parameters. Thus, physiological meaningful changes of any reaction rate constant cannot lead to the rescue of knockdown efficacy.

The associative model is sensitive to changes in reaction rate constants of binary complex formation (k_{+2}, k_{-2}, k_{+3}, and k_{-3}), as well as ternary complex formation (k_{+4}, k_{-4}, k_{+5}, k_{-5}, and k_{+6}). These parameters limit knockdown efficacy for low target concentrations up to $1,000$ or $2,000$ copies/cell (for 30 or 300 siRNAs/cell, respectively). At higher target concentrations the knockdown efficacy is practically zero (no target knock-down after 24 h). Thus, changes of any reaction rate constant will not lead to rescue of the knockdown efficacy. Just before this breakdown in efficacy occurs, the system becomes sensitive



(a)



(b)

Figure 4.23 **IC_{50}: adaptation of the RNAi machinery against variations in target mRNA level (revised models). A)** The ration between IC_{50}, the half maximal inhibitory siRNA concentration after 24 h of knockdown, of the dissociative and the associative model is plotted against logarithmically-scaled basal target mRNA concentration. **B)** For the dissociative (broken blue line) and the associative (black line) models, IC_{50} is plotted against logarithmically-scaled basal target mRNA concentration. In the case of the experimental data (red squares with S.D.), the IC_{50} is constant over four magnitudes of transfected target plasmid (0.15 - 150 ng/μl). IC_{50} of the dissociative model increases rapidly with increasing target0. For the associative model, IC_{50} slowly decreases with $[target]_0$ until it reaches its optimum between 200 and 300 copies/cell. From there it rapidly increases. All in all, the associative model can adapt more efficiently to variation in target mRNA level between 1 - 10,000 copies/cell and thus resembles the experimental data closer than the dissociative model does. This figure is based on Figures 4 and 5 of[40].

to k_{+10} (i.e. product release), which can be identified as the limiting step / bottleneck for knockdown efficacy when increasing target concentration. In comparing the two models, it becomes evident the associative model is far less sensitive to changes of binary or ternary complex formation at low target concentrations. In contrast to the dissociative model, the knockdown efficacy is sustained even for higher target concentrations. Here, product release does not become the limiting step. However, association/dissociation of target into the reaction complex (i.e. quaternary complex formation; k_{+11}, k_{-11}, k_{+12}, k_{-12} and k_{+13}) eventually limits efficacy when the RNAi machinery becomes over-saturated with target.

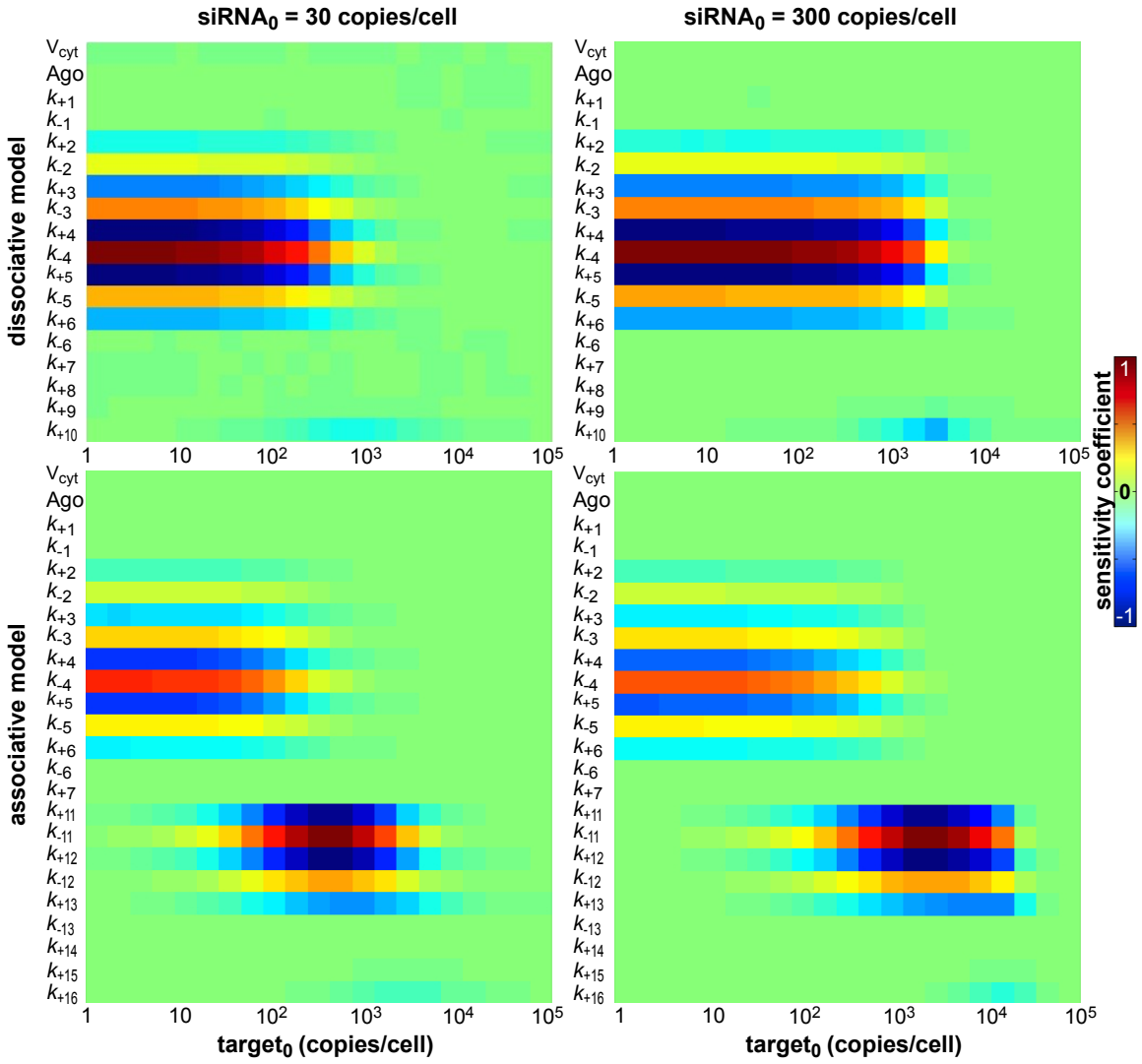


Figure 4.24 **Sensitivity of time-resolved target knockdown (revised models).** Sensitivity coefficients ε_i^k quantify the sensitivity of a 24 h time course of target concentration after siRNA-transfection of respectively 30 or 300 copies/cell to small parameter changes S_i (rows) as a function of basal target concentration C_k (columns). Investigated parameters Si are cytoplasm volume V_{cyt}, Ago2 concentration, kinetic rate constants k_{+1} to k_{+16} and k_{-1} to k_{-16}. As defined in the colour bar, sensitivity coefficients are normalised between [-1, 1] and colour-coded. Sensitivities are shown for the dissociative (upper row) and the associative (lower row) model. This figure is based on Supplementary Figure S6 of[40].

Chapter 5

Results: Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA

Post-transcriptional regulation of gene expression and viral functions, i.e. regulation on the level of gene-specific RNAs usually involve structural and functional domains of cellular mRNAs or viral transcripts. For example, the TAR element of the human immunodeficiency virus type 1 (HIV-1), a sequence stretch of approximately 60 nucleotides within the $5'$-region of genomic HIV-1 RNA as well as the $5'$-region of viral mRNAs, adopt a thermodynamically stable stem loop structure that is functionally involved in the regulation of the elongation of transcription[203]. Furthermore, internal ribosomal entry sites, mostly found in viral sequences, form defined structural elements that are necessary for the translation of protein-coding transcripts (for a review see: [41, 165]. On the level of cellular genes, an extra-stable local folding unit of the $5'$-untranslated region (UTR) of the murine p53 mRNA seems to be involved in the post-transcriptional control of its level of protein translation[145]. In summary, structural and functional RNA modules, often located upstream of the translational initiation site or overlapping with it, are known to be involved in the regulation of protein biosynthesis. Such regulatory local RNA folding units may interact with modulators such as complementary RNA sequences in the case of microRNA or antisense RNA-mediated control of gene expression (for review see:[148]), RNA-binding proteins (for review see:[38]), or low-molecular metabolites like in the case of ribo-switches[236]. In most cases regulatory RNA units are composed of consecutively neighbouring sequence segments. Only in rare cases long-range interactions seem to be involved in the formation of regulatory local structures like in IRES elements (e.g.[62, 181])

or certain classes of catalytic RNA (e.g. [134]). Particularly simple regulatory modules of limited length such as the TAR element, the RRE element, or packaging signals of the human immunodeficiency virus type 1 (HIV-1) often adopt structural domains that coincide with local minimal energy states[219, 233]. Studies on the biological role of Argonaute (Ago) genes in human cells indicate a potential post-transcriptional control step involved in regulating endogenous levels of the family member Ago2[132]. Further, preliminary experimental studies (not part of this doctoral thesis) in our laboratory indicate specific binding of RNA sequences surrounding the AUG start codon of the Ago2 mRNA with proteins involved in RNA interference (Sczakiel, unpublished data). This inspired the analysis of the $5'$-UTR of mRNA^{Ago2} with regard to potential RNA cis-elements. Initial computational RNA folding studies indicated the existence of two helices formed by consecutive GC base pairs within the $5'$-UTR and upstream coding sequences which seemed to represent a rare case. Here it was studied whether the occurrence of GC-rich helices could be biologically relevant. Systematic computational and phylogenetic studies as well as experimental evidence in a mammalian cell system support the view that GC-rich duplex motifs formed by distant RNA segments, unlike typical hairpin elements, could bear biological functions.

5.1 Characterisation of nucleotide composition of long-range duplexes in mRNA

5.1.1 Helix definitions

In this part of the study the focus was on predicted duplex elements composed of distant sequence segments and set definitions such that they could be distinguished from usual hairpin motifs. Hence two prerequisites were defined: a minimal distance of 40 nt between two helix strands and a minimal number of 16 bp within this inter-segment region to differentiate helices from hairpin elements as it requires RNA structure at a position where a loop is found in hairpin elements. First screens indicated that these conditions result in a very small number of predicted long-range duplexes at a duplex length of greater than or equal to 8 bp. It should be noted that the software tool created in this study also accepts any other kind of duplex settings including (i) the base composition, (ii) different characteristics of the inter-segment sequence which may be longer, shorter or adopt other structures, (iii) and the number of consecutive base pairs included in the duplex of interest. For a more detailed description see Table 5.1.

parameter	description	co-domain	duplex	helix	stem	AU-duplex	AU-helix	GC-duplex	GC-helix	GC-stem
min_dist	minimal length of inter-segment sequence	$\mathbb{N}^0$ (nt)	0	40	ULTD	0	40	0	40	0
min_linker_pair	minimal number of bp in inter-segment sequence	$\mathbb{N}_0$ (bp)	0	6	0	0	6	0	6	0
max_dist	maximal length of inter-segment sequence	$\mathbb{N}^0$ (nt)	ULTD	ULTD	39	ULTD	ULTD	ULTD	ULTD	39
max_linker_pair	maximal number of bp in inter-segment sequence	$\mathbb{N}^0$ (bp)	ULTD	ULTD	5	ULTD	ULTD	ULTD	ULTD	5
min_len	minimal number of consecutive bp with duplex of interest	$\mathbb{N}$ (bp)	8	8	8	8	8	8	8	8
anti_parallel	duplex formation between anti-parallel strands valid?	Boolean	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
max_AU	maximal number of AU bp allowed in duplex	$\mathbb{N}^0$ (nt)	ULTD	ULTD	ULTD	ULTD	ULTD	0	0	0
max_GU	maximal number of GU bp allowed in duplex	$\mathbb{N}^0$ (nt)	ULTD	ULTD	ULTD	0	0	0	0	0
max_GC	maximal number of GC bp allowed in duplex	$\mathbb{N}^0$ (nt)	ULTD	ULTD	ULTD	0	0	ULTD	ULTD	ULTD

Table 5.1 **Parameters for duplex definition.** The software tool provides seven free parameters to specify duplex classes. It is possible to restrict base composition, different characteristics of the inter-segment sequence which may be longer, shorter or adopt other structures, and the number of consecutive base pairs with the duplex of interest. Columns 1-3 describe these parameters; columns 4-11 define duplex classes as used in this study. This table is based on Supplementary Table S1 of[39].

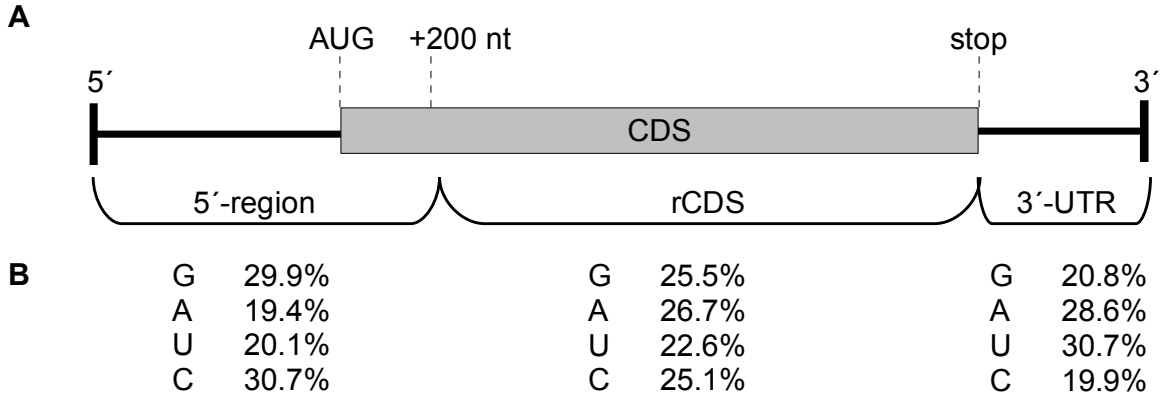


Figure 5.1 **Schematic depiction of mRNA sub-regions. A)** Sub-regions of mRNA and **B)** their nucleotide composition. This figure is based on Figures 1B&C of[39].

5.1.2 Search for GC-rich double helices

5.1.3 Region definition of mRNA

In this study, mRNAs were subdivided into three functional regions (see Fig. 5.1 A), the 5′-region which includes the 5′-UTR and 200 nt of the upstream coding sequences, the remaining coding region (rCDS) which spans the remaining CDS, and the 3′-UTR.

5.1.4 Selection of human and non-human RNAs

Analyses were restricted to mRNAs with intact 5′-UTR, CDS, and 3′-UTR. Human sequences were chosen randomly from the databases. Furthermore, 10 sequences, coding for members of the human Argonaute protein family were included. They can be subdivided into five PIWI-like (PIWI1, PIWI2 transcript variant 1, PIWI2 transcript variant 2, PIWI3, PIWI4) and five AGO-like (Ago1, Ago2 , Ago3 transcript variant 1, Ago3 transcript variant 2 , Ago4) mRNAs[187]. In addition, 49 orthologous non-human Argonaute mRNAs were investigated. The sequences originate from the 13 organisms Pan troglodytes, Canis lupus familiaris, Bos taurus, Oryctolagus cuniculus, Drosophila melanogaster, Mus musculus, Rattus norvegicus, Gallus gallus, Macaca mulatta, Mustel aputorius furo, Sus scrofa, Xenos puslaevis and Danio rerio. Many non-human sequences found in databases were incomplete or labelled 'predicted' to the time of this work and were thus excluded. 12 predicted Argonaute sequences from Pan troglodytes and one predicted sequence from Canis lupus familiaris were included. Additionally, 28 non-coding RNA transcripts from *Homo sapiens* and Mus musculus were investigated. A collection of random sequences was gen-

erated by re-shuffling human mRNAs while conserving the nucleotide frequencies of their 5′-UTRs, CDS's, and 3′-UTRs. All sequences used in this study are listed in appendix A, p. 167.

5.1.5 Evolutionary conservation

During comparative analyses the structure of paralogs and orthologs of hAgo2 mRNA were compared. Their duplex content was characterised with respect to (i) frequency, (ii) location within RNA sequences, (iii) duplex length, (iv) nucleotide composition, and (v) structure of inter-segment sequences.

5.2 Structural and functional analysis of Argonaute-class mRNAs

5.2.1 Double helices composed of distant RNA segments

Although long RNA strands may be extensively folded, consecutively base-paired double helices longer than 4 to 6 bp located outside of hairpin elements (including short helices that are packed together in the overall architecture of RNA) occur rarely in resolved and predicted folding units[17]. The occurrence of non-hairpin duplex elements including GC-rich or AU-rich helices in a collection of human mRNA sequences was studied by using the newly developed software tool. The collection of human RNA sequences was representative because the sequences were chosen randomly and include mRNA from transcription factors, metabolic regulators, receptors, enzymes, ion channels, carrier proteins, glycoproteins, structural proteins, heat shock proteins as well as yet uncharacterised cDNA sequences. Firstly, sequences were retrieved (sequences are listed in appendix A, p. 167). Then, secondary structure predictions were performed as described in Section 3.7.2, p. 47 using a window size of 800 nt and a step width of 20 nt. Finally, the results were screened for duplex elements, i.e. helices according to the definitions described in Section 5.1.1, p. 140. Human mRNAs vary in total length which is also true for their domains, i.e. the 5′-UTR, CDS, and 3′-UTR. To account for these variations in length, helix frequency per 100 nt was used as measuring unit. When non-hairpin duplexes composed of 8 or more consecutive base pairs were considered, a frequency of 1.774 duplexes per 100 nt was found in the examined mRNA population. In 21 of the evaluated structures (32%) one or more GC-helices were found. The overall GC-helix frequency was 0.01 helices / 100 nt. In com-

parison, AU-helices occur at about the same frequency and were found in 23 of the scanned structures (35%). In summary, the sequences analysed in this study showed an average length of non-hairpin double helices of 9 ±1.5 bp (S.D.). For GC-helices an average distance between the two complementary strand segments of 180 ± 177 nt was observed. For AU-helices this value was 231 ± 181 nt and for all helices regardless of base composition it was 327.63 ± 227 nt. These characteristics were similar between human and non-human mRNAs. The highest degree of consistency at the inter-species comparison was seen in the 5′-region, which by itself is very rich in GC-helices (compare Section 5.2.3, p. 145). For AU-helices this agreement was less pronounced. Non-coding RNA transcripts and shuffled human mRNAs were taken as two separate control groups to test whether the abundance of predicted GC- and AU-duplexes in the examined mRNA collection was non-random. No GC-duplexes of 8 or more consecutive base pairs were found in any examined non-coding RNA transcript. The GC-helix frequency in the shuffled mRNAs was 0.002 and thus lower by a factor of 5 compared to the human mRNA population. In contrast, AU-duplexes were more abundant in non-coding RNA compared to human mRNA, while their frequency in shuffled mRNA sequences was in the same magnitude as in the mRNA population. Please refer to Tables 5.2 and 5.3 for GC- and AU-duplex motifs, respectively.

5.2.2 Helix location within the mRNA sequence

Next it was analysed whether GC-helices are equally distributed along mRNA. mRNAs were sub-divided into three domains (Fig. 5.1 A and Section 5.1.3, p. 142). The nucleotide composition of these domains of the human mRNA sequences studied here is summarised in Fig. 5.1 B. Systematic computational analyses of the human mRNA sequences and their predicted structures reveal that GC-helices predominantly occur within the 5′-region of mRNAs (0.084 per 100 nt versus 0.006 in the rCDS and 0.002 in the 3′-UTR, respectively) while AU-helix frequency is much lower and AU-helices predominantly occur within the 3′-UTR (Fig. 5.3 A and B). The location of predicted GC-helices in relation to the start codon can be found in Additional file 3. In contrast, the occurrence of duplex elements with mixed base compositions seems to be more balanced between the three regions (1.738, 1.915 and 1.565 per 100 nt for 5′-region, rCDS, and 3′-UTR, respectively; Fig. 5.3 C). Replacing the 5′-region by 5′-UTR and hence rCDS by CDS (i.e. the complete coding region) does not alter these figures substantially. The uneven distribution of GC- and AU-helices between the three defined regions might be related to a certain extent to the base composition of these regions (Fig. 5.1 B) but it seems to be unlikely that this fully explains the sharp differences

of their local occurrence. To test this, GC- and AU-duplex motifs were searched for in randomly shuffled RNA sequences. Here, GC-duplex frequencies were lower than found in the mRNA population and their location within the three sequence domains was balanced in the collection of shuffled mRNA. The preferred occurrence of GC duplexes in the $5'$-region of mRNAs was not observed in the case of shuffled mRNAs (Table 5.2). Also AU-duplexes occurred in a balanced fashion within the three sequence domains in contrast to the mRNA sequences, where AU-duplexes were preferably located in the $3'$-UTR (Tabel 5.3). It should be mentioned that the position of GC-rich helices did not coincide with CpG islands on the level of chromosomal DNA. In the following the focus was on double helices predominantly or exclusively composed of GC base pairs.

5.2.3 GC-helix abundance and gene family

Conserved RNA structure elements may be involved in biological functions[34]. Structural and related functional conservation[48, 81] can be supported by comparative analysis between classes of genes and by studying evolutionary conservation. Here it became obvious that GC-helices occur in the $5'$-UTR regions of the three human AGO-like mRNAs Ago1, Ago2, and Ago4 while being completely absent in human PIWI-like mRNAs. This finding is also true for other vertebrates studied here. Hence, the occurrence of GC-helices and their location within the $5'$-UTR of mRNAs of the Argonaute gene family might be biologically meaningful.

5.2.4 Phylogenetic analysis of GC-helices in Argonautes

The accumulation of GC-helices was further studied for orthologous Argonaute sequences. This study was restricted to 49 sequences from 13 different organisms (sequences are listed in Appendix A, p. 167) since many non-human sequences found in online databases were not suited for this study because of limitations described in the Methods section. Out of 49 orthologous Argonaute sequences, 10 were GC-helix positive. The ten hits share two criteria: they are of vertebrate origin and their gene class is AGO-like Argonautes: Bos taurus Ago1 and Ago2 (two variants), Mus musculus Ago2 (two variants) and Ago4, Rattus norvegicus Ago4, Pan troglodytes two Ago-like (Transcriptome Shotgun Assembly), as well as, Gallus gallus Ago4. As there exist only few complete non-human Argonaute mRNA sequence files, additionally 13 predicted sequences were evaluated. However, since they were not determined experimentally, but are estimated from homologue sequences by GNOMON, these results were separated from the rest: GC-helices were found in four out

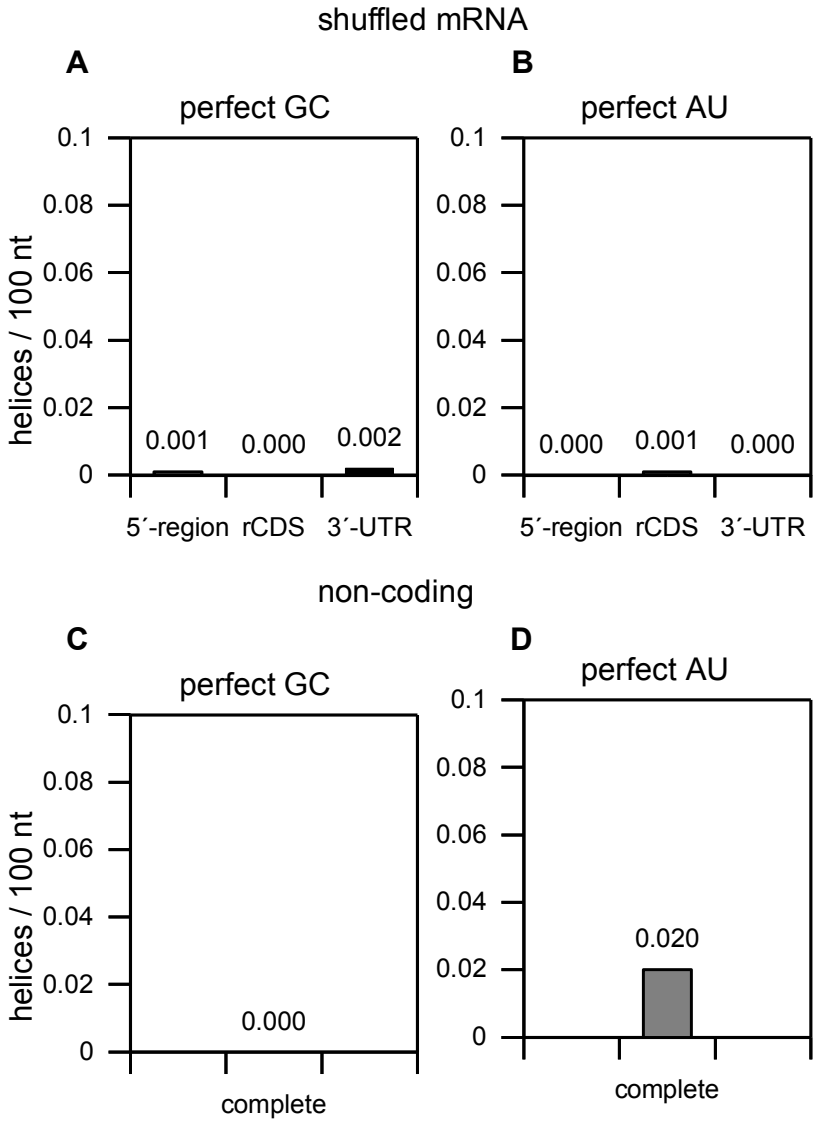


Figure 5.2 **Duplex motifs in RNA control libraries.** The occurrence and nucleotide composition of long-range duplexes in RNA control libraries.

RNA library and motif length (nts)	5′-region			rCDS			3′-UTR		
	sliding win (mfold2.3)	mfold3.4	sfold2.2	sliding win (mfold2.3)	mfold3.4	sfold2.2	sliding win (mfold2.3)	mfold3.4	sfold2.2
human mRNA, n=12	0.01 (4)	0.01 (2)	0.01 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
human mRNA, n=11	0.02 (9)	0.01 (4)	0.01 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
human mRNA, n=10	0.02 (9)	0.01 (4)	0.01 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
human mRNA, n=9	0.05 (18)	0.03 (10)	0.02 (8)	0.00 (1)	0.00 (1)	0.00 (1)	0 (0)	0 (0)	0 (0)
human mRNA, n=8	0.1 (37)	0.07 (28)	0.06 (20)	0.01 (13)	0.00 (7)	0.00 (6)	0.00 (2)	0.00 (1)	0.00 (1)
human mRNA, n=7	0.2 (75)	0.18 (67)	0.12 (42)	0.02 (37)	0.01 (19)	0.01 (13)	0.00 (4)	0.00 (3)	0.00 (3)
human mRNA, n=6	0.48 (181)	0.42 (159)	0.25 (91)	0.08 (146)	0.04 (70)	0.03 (51)	0.02 (22)	0.01 (9)	0.01 (9)
human mRNA, n=5	1.16 (438)	1.04 (391)	0.55 (197)	0.34 (591)	0.13 (231)	0.10 (167)	0.13 (125)	0.05 (51)	0.04 (34)
shuffled mRNA, n=9	0.00 (2)	0.00 (1)	n/a	0 (0)	0 (0)	n/a	0 (0)	0 (0)	n/a
shuffled mRNA, n=8	0.01 (4)	0.01 (4)	n/a	0.00 (1)	0.00 (1)	n/a	0.00 (4)	0.00 (1)	n/a
shuffled mRNA, n=7	0.02 (13)	0.03 (20)	n/a	0.01 (21)	0.01 (13)	n/a	0.02 (23)	0.01 (10)	n/a
shuffled mRNA, n=6	0.08 (60)	0.09 (69)	n/a	0.08 (122)	0.04 (66)	n/a	0.06 (81)	0.03 (40)	n/a
shuffled mRNA, n=5	0.36 (288)	0.40 (321)	n/a	0.37 (604)	0.18 (289)	n/a	0.31 (403)	0.12 (151)	n/a

RNA library and motif length (nts)	complete sequence		
	sliding win (mfold2.3)	mfold3.4	sfold2.2
non-coding, n=8	0 (0)	0 (0)	0 (0)
non-coding, n=7	0.00 (2)	0.00 (1)	0.00 (1)
non-coding, n=6	0.01 (23)	0.02 (14)	0.02 (7)
non-coding, n=5	0.05 (160)	0.12 (79)	0.08 (24)

Table 5.2 **Distribution of predicted RNA duplex motifs: GC-duplexes of various length.** RNA sequences were classified in five groups human Ago-like Argonautes, none-human Argonautes (Ago- and PIWI-like), human mRNAs, non-coding RNAs, and shuffled mRNAs (for accession numbers of RNA sequences see Appendix A1). Each sequence (except for non-coding RNAs) is sub-divided into three regions **(i)** 5‘-region (5‘-UTR + 200 nt), **(ii)** rCDS (remaining coding sequence), and **(iii)** 3‘-UTR (see also Figure 5.1B). Analysis of secondary structure motif abundance is based on three secondary structure prediction approaches **(i)** sliding window (window size = 800 nt, step width = 20 nt, based on Mfold 2.3), **(ii)** Mfold 3.4, and **(iii)** Sfold 2.2. Abundance of structural motifs is given in motifs per 100 nt. Numbers in brackets indicate absolute number of helices found in each group. Absolute numbers are not directly comparable among the three approaches, since not all sequences could be folded by Sfold 2.2 and Mfold 3.4, respectively, due to length restrictions for input sequences on the online servers. Parameters for definition of the GC-duplex motif ($\geq$ n uninterrupted GC pairs) is provided in Table 5.1. This table is based on Supplementary Table S4 (p. S10) of[39].

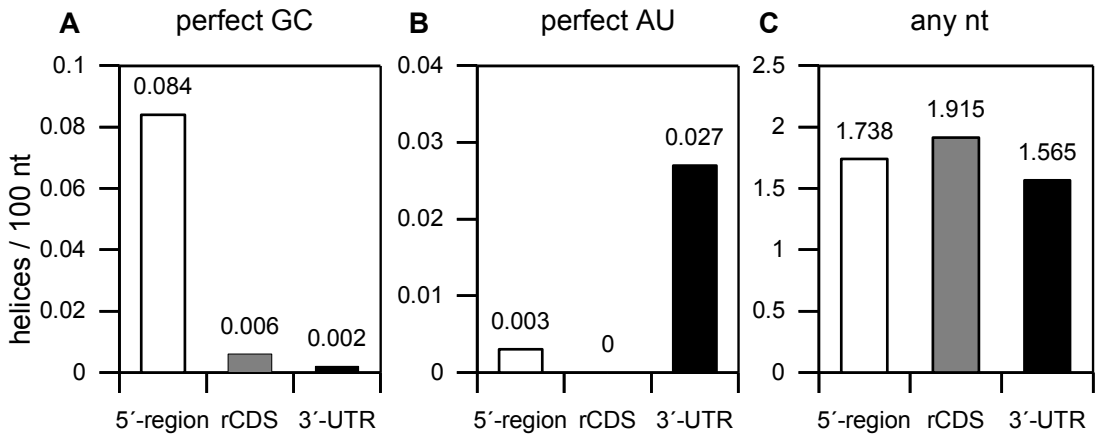


Figure 5.3 **Duplex motifs as a function of mRNA region.** The occurrence and nucleotide composition of long-range duplexes as a function of mRNA region. This figure is based on Figures 1D-F of[39].

RNA library and motif length (nts)	5′-region sliding win (mfold2.3)	5′-region mfold3.4	5′-region sfold2.2	rCDS sliding win (mfold2.3)	rCDS mfold3.4	rCDS sfold2.2	3′-UTR sliding win (mfold2.3)	3′-UTR mfold3.4	3′-UTR sfold2.2
human mRNA, n=12	0 (0)	0.01 (3)	0.01 (4)	0 (0)	0.00 (4)	0 (0)	0.01 (10)	0.01 (9)	0.00 (4)
human mRNA, n=11	0 (0)	0.01 (3)	0.01 (4)	0 (0)	0.00 (4)	0 (0)	0.01 (15)	0.01 (13)	0.01 (7)
human mRNA, n=10	0 (0)	0.01 (3)	0.01 (4)	0 (0)	0.00 (4)	0 (0)	0.02 (20)	0.02 (18)	0.01 (11)
human mRNA, n=9	0.00 (1)	0.01 (3)	0.01 (4)	0 (0)	0.00 (4)	0 (0)	0.03 (35)	0.03 (29)	0.02 (17)
human mRNA, n=8	0.01 (2)	0.02 (7)	0.01 (5)	0 (0)	0.00 (7)	0 (0)	0.06 (71)	0.05 (50)	0.03 (28)
human mRNA, n=7	0.01 (3)	0.02 (9)	0.01 (5)	0.01 (10)	0.01 (19)	0.00 (4)	0.13 (157)	0.08 (92)	0.05 (48)
human mRNA, n=6	0.03 (12)	0.05 (20)	0.03 (10)	0.03 (44)	0.03 (50)	0.01 (18)	0.31 (378)	0.16 (180)	0.10 (98)
human mRNA, n=5	0.10 (38)	0.18 (68)	0.06 (23)	0.10 (174)	0.11 (189)	0.04 (63)	0.71 (860)	0.33 (370)	0.19 (182)
shuffled mRNA, n=9	0 (0)	0 (0)	n/a	0 (0)	0 (0)	n/a	0 (0)	0 (0)	n/a
shuffled mRNA, n=8	0 (0)	0.01 (4)	n/a	0.00 (2)	0.00 (1)	n/a	0 (0)	0 (0)	n/a
shuffled mRNA, n=7	0.00 (3)	0.01 (10)	n/a	0.01 (10)	0.00 (2)	n/a	0 (1)	0 (0)	n/a
shuffled mRNA, n=6	0.02 (14)	0.03 (21)	n/a	0.03 (41)	0.01 (15)	n/a	0.01 (13)	0.01 (9)	n/a
shuffled mRNA, n=5	0.11 (84)	0.10 (82)	n/a	0.12 (201)	0.04 (69)	n/a	0.09 (117)	0.04 (48)	n/a

RNA library and motif length (nts)	complete sequence sliding win (mfold2.3)	complete sequence mfold3.4	complete sequence sfold2.2
non-coding, n= 12	0.00 (4)	0.00 (2)	0.00 (1)
non-coding, n= 9	0.01 (34)	0.01 (6)	0.01 (2)
non-coding, n= 8	0.03 (104)	0.03 (20)	0.01 (3)
non-coding, n=7	0.09 (281)	0.07 (43)	0.01 (4)
non-coding, n=6	0.22 (281)	0.13 (87)	0.04 (11)
non-coding, n=5	0.60 (1955)	0.41 (270)	0.10 (32)

Table 5.3 **Distribution of predicted RNA duplex motifs: AU-duplexes of various length.** Please refer to the legend of Table 5.2 for more information. Parameters for definition of the AU-duplex motif ($\geq$ n uninterrupted AU pairs) is provided in Table 5.1. This table is based on Supplementary Table S4 (p. S11) of[39].

RNA library	5′-region			rCDS			3′-UTR		
	sliding win (mfold2.3)	mfold3.4	sfold2.2	sliding win (mfold2.3)	mfold3.4	sfold2.2	sliding win (mfold2.3)	mfold3.4	sfold2.2
hAgo	0.24 (7)	0.07 (2)	0.03 (1)	0 (0)	0.01 (1)	0.01 (1)	0 (0)	0 (0)	0 (0)
none-h Argo	0.07 (12)	0.06 (11)	0.04 (7)	0.00 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
human mRNA	0.08 (32)	0.06 (23)	0.04 (15)	0.01 (11)	0.00 (6)	0.00 (5)	0.00 (2)	0.00 (1)	0.00 (1)
shuffled mRNA	0.01 (4)	0.01 (4)	n/a	0.00 (1)	0.00 (1)	n/a	0.00 (4)	0.00 (1)	n/a

RNA library	complete sequence		
	sliding win (mfold2.3)	mfold3.4	sfold2.2
non-coding	0 (0)	0 (0)	0 (0)

Table 5.4 **Distribution of predicted RNA motif GC-helix.** RNA sequences were classified in five groups human Ago-like Argonautes, none-human Argonautes (Ago- and PIWI-like), human mRNAs, non-coding RNAs, and shuffled mRNAs (for accession numbers of RNA sequences see Appendix A1). Each sequence (except for non-coding RNAs) is sub-divided into three regions **(i)** 5‘-region (5‘-UTR + 200 nt), **(ii)** rCDS (remaining coding sequence), and **(iii)** 3‘-UTR (see also Figure 5.1B). Analysis of secondary structure motif abundance is based on three secondary structure prediction approaches **(i)** sliding window (window size = 800 nt, step width = 20 nt, based on Mfold 2.3), **(ii)** Mfold 3.4, and **(iii)** Sfold 2.2. Abundance of structural motifs is given in motifs per 100 nt. Numbers in brackets indicate absolute number of helices found in each group. Absolute numbers are not directly comparable among the three approaches, since not all sequences could be folded by Sfold 2.2 and Mfold 3.4, respectively, due to length restrictions for input sequences on the online servers. Parameters for definition of the GC helix motif ($\geq$ 8 uninterrupted GC pairs, strand distance $\geq$ 40 nt, $\geq$ 16 nt in linker region involved in base pairings) is provided in Table 5.1. This table is based on Supplementary Table S4 (p. S9) of[39].

RNA library	5′-region			rCDS			3′-UTR		
	sliding win (mfold2.3)	mfold3.4	sfold2.2	sliding win (mfold2.3)	mfold3.4	sfold2.2	sliding win (mfold2.3)	mfold3.4	sfold2.2
hAgo	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0.03 (3)	0.02 (2)	0.03 (2)
none-h Argo	0.01 (1)	0.01 (1)	0.01 (1)	0.00 (3)	0.00 (5)	0.00(4)	0.02 (10)	0.00 (2)	0.00(6)
human mRNA	0.00 (1)	0.02 (7)	0.01 (5)	0 (0)	0.00 (7)	0 (0)	0.03 (35)	0.03 (33)	0.02 (17)
shuffled mRNA	0 (0)	0.01 (4)	n/a	0.00 (1)	0 (0)	n/a	0 (0)	0 (0)	n/a

RNA library	complete sequence		
	sliding win (mfold2.3)	mfold3.4	sfold2.2
non-coding	0.02 (70)	0.03 (15)	0.02 (2)

Table 5.5 **Distribution of predicted RNA motif AU-helix.** Please refer to the legend of Table 5.4 for more information. Parameters for the definition of the AU helix motif ($\geq$ 8 uninterrupted AU pairs, strand distance $\geq$ 40 nt, $\geq$ 16 nt in linker region involved in base pairings) is provided in Table 5.1. This table is based on Supplementary Table S4 (p. S9) of[39].

of twelve 5′-UTR-regions of predicted AGO-like mRNAs in Pan troglodytes.

5.2.5 Compatibility with different secondary structure prediction algorithms

Compared to the sliding window approach based on Mfold 2.3 (as described in Section 3.7, p. 46), Sfold 2.2 and Mfold 3.4 predicted 48% and 31% less GC-duplex motifs of 8 or more base pairs in the investigated mRNA collection, respectively (compare Tabel 5.2). However, the qualitative finding of this study was independent of the kind of secondary structure prediction algorithm. All three compared algorithms predicted that 5′-regions of human and non-human mRNAs are rich in GC-helices compared to rCDS's and 3′-UTRs, while AU-helices are most abundant in 3′-UTRs. In technical terms, on a PC with Intel®Core™2 Duo CPU E4500 @ 2.20GHz and 2 GB memory, running openSUSE 11.1 (x86 64) folding of a 9,000 nt sequence by Mfold 2.3 took on average 206 min. At a step width of 20 nt and a window size of 800 nt a RNA sequence of 9,000 nucleotides in length means 410 Mfold runs which all together took about 60 min on the same system. This is in accordance with the fact that the underlying Mfold algorithm computes in time proportional to the cube of the folded length of sequence. Because Mfold 3.4 and Sfold 2.2 computations were performed on online servers, their performance was not directly comparable. However, in the case of Sfold 2.2 calculations of 9,000 nt sequences took several days (mainly due to queuing) and in the case of Mfold 3.4 the results were ready for download after 43 min.

5.2.6 A GC-helix within the 5′-region of the Ago2 mRNA is involved in post-transcriptional regulation of Ago2 gene expression

In order to shed more light on the existence of the predicted GC-helix in the 5′-region of mRNA^{Ago2} the local folding potential was calculated. This parameter is often correlated with biologically functional and stably folded domains [98, 109]. The local energy minimum related to the folding unit between positions 20 and 80 (minimum at position 20 at a window size of 60 nt; (Fig. 5.4) coincided with the 9 bp GC-helix depicted in Fig. 5.5. In order to test whether this GC-helix could be involved in regulation of Ago2 gene expression in functional terms, the role of this element in gene expression studies in cell culture as well as its characteristics in protein binding studies *in vitro* was analysed. The conduct of the following experiments were performed by experimental collaborators under the guidance of Georg Sczakiel and were not part of this doctoral study:

Firstly, recombinant plasmids containing the Renilla luciferase open reading frame fused downstream to the 5′-UTR of mRNA^{Ago2} termed pRL-Ago2 were cloned (Fig. 5.6 A). For control purposes parental luciferase-harbouring plasmid (pRL) as well as a variant of pRL-Ago2 in which the GC-helices were destroyed by nucleotide exchanges was used (please see Appendix A, p. 172 for the plasmids' sequences). In order to study possible effects of over-expression of Ago2 on the 5′-UTR^{Ago2}-luciferase fusion mRNA, either of these plasmids was co-transfected together with an established recombinant eukaryotic expression plasmid for Ago2 (pNHA-Ago2;[166]) into human ECV304 cells. The results shown in Fig. 5.6 B indicate an up-regulation of the 5′-UTR^{Ago2}-luciferase fusion by over-expressed Ago2 protein but not in the presence of the parental Ago2-negative plasmid termed 'contr'. Further, basal expression levels of the 5′-UTR^{Ago2}-luciferase fusion seem to be higher than the levels of controls (Fig. 5.6 B, open bars) which is consistent, to a certain extent, with the assumption that endogenous Ago2 increases its expression via the 5′-UTR^{Ago2}.
Secondly, binding studies were performed with HeLa cell extracts and *in vitro* transcribed Ago2 mRNA 5'-sequences or control sequences depicted in Fig. 5.6 A. These studies indicated increased binding of proteins involved in RNA interference including Ago2 itself with the GC-duplex-containing sequences of Ago2 mRNA. This observation was consistent with the involvement of the GC-duplex of the Ago2 mRNA in post-transcriptional control of gene expression (Sczakiel, unpublished data).

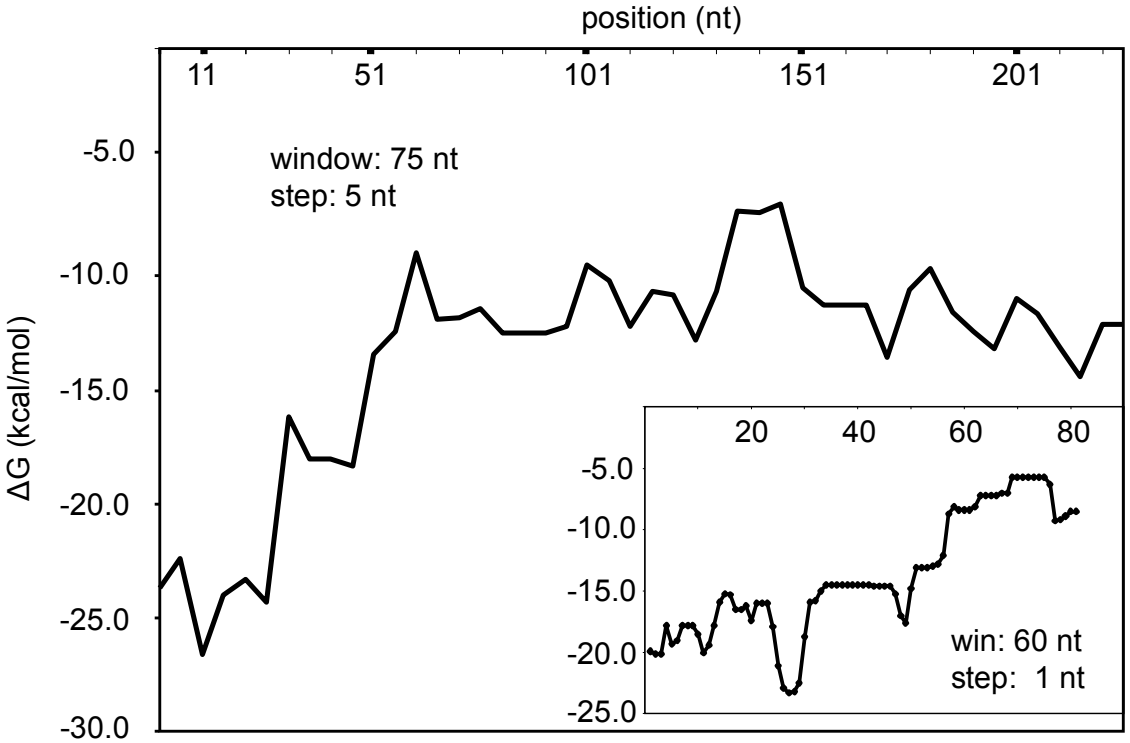


Figure 5.4 **The folding potential of the 5′-region of 5′-UTR^{Ago2}.** This figure is based on Figure 2A of[39].

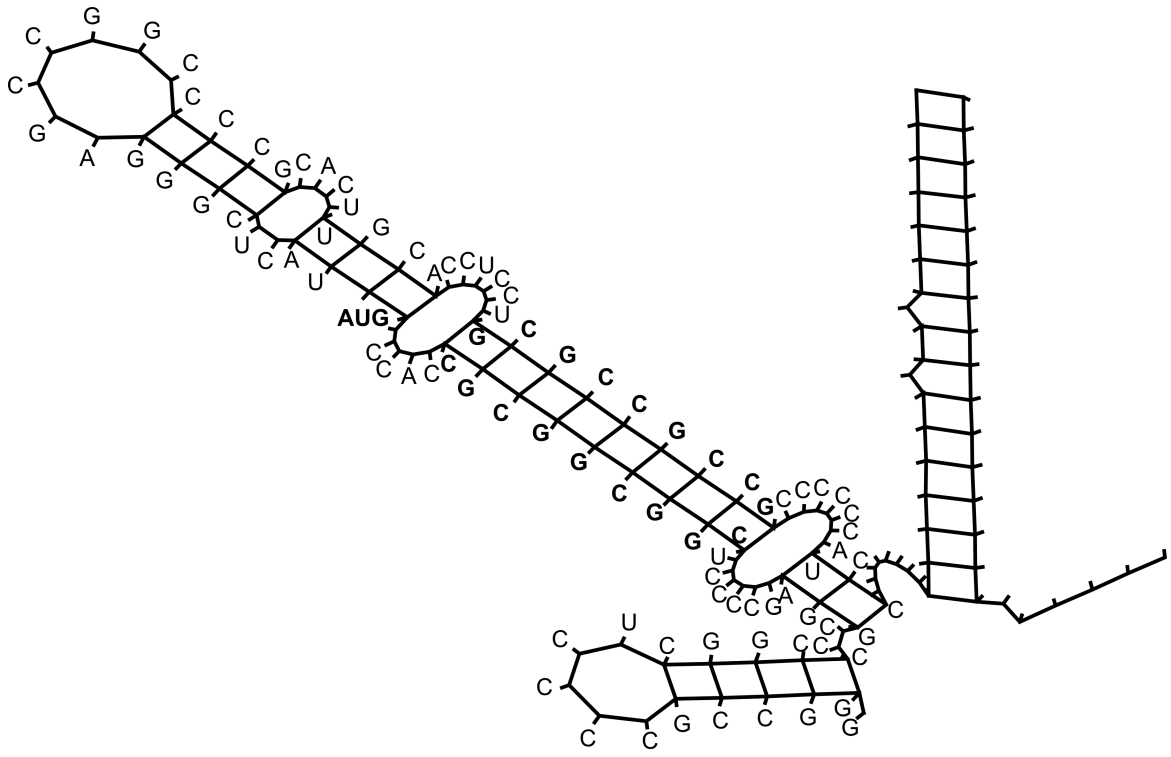


Figure 5.5 **The predicted 2D structure of the 5‘-region of 5′-UTR^{Ago2} suggests two GC-helices.** This figure is based on Figure 2B of[39].

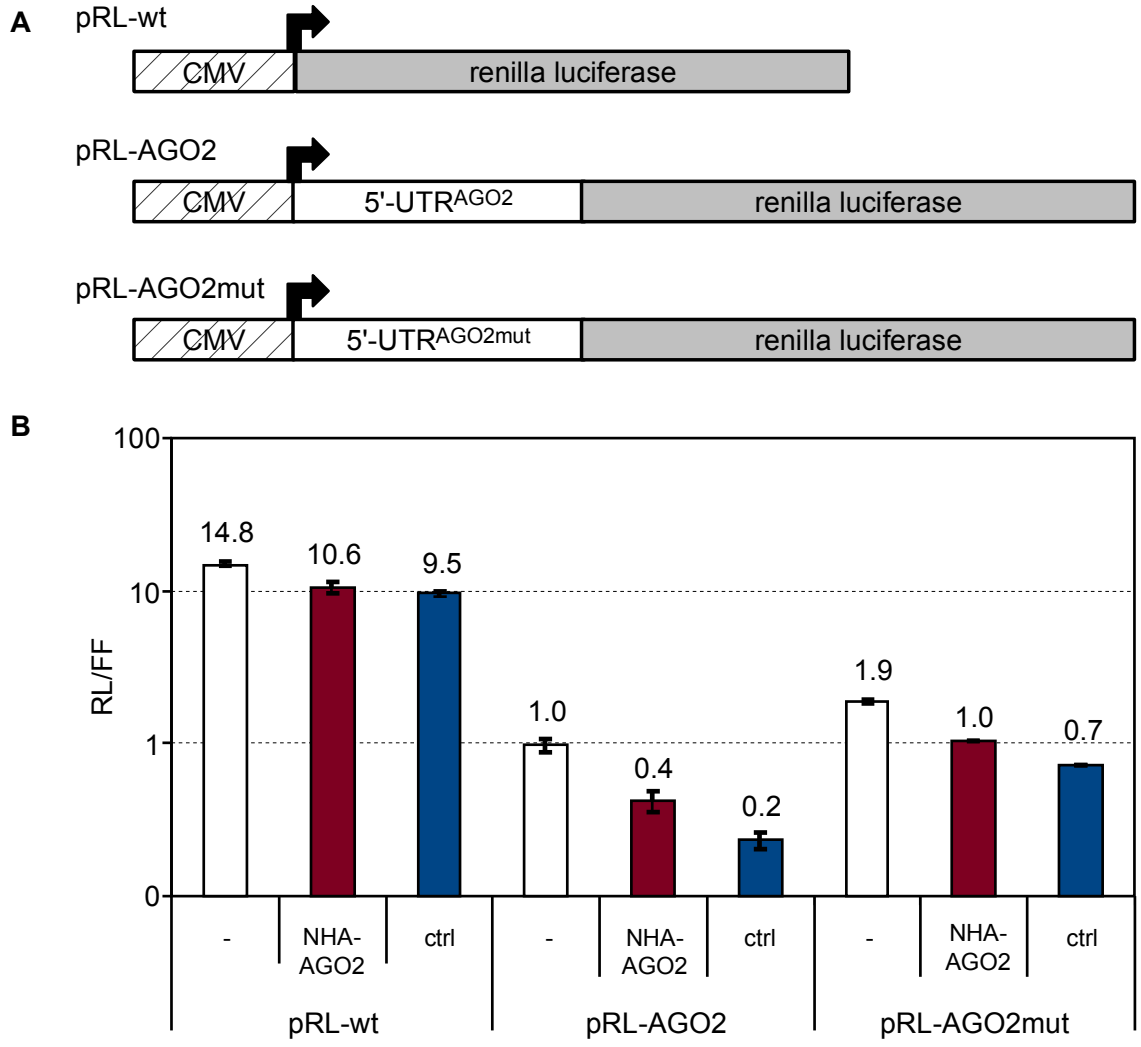


Figure 5.6 **Experimental evidence for the involvement of the GC-helix in the regulation of Ago2 gene expression:** a luciferase gene containing the mRNA^{Ago2} 5′-region (pRL-Ago2) but not a derivative mutated in the GC-helix region nor the wt gene indicate Ago2-mediated up-regulation of gene expression (NHA-Ago2). This figure is based on Figure 2C&D of[39].

Chapter 6

Discussion and Conclusion

This study was able provided insights into RNAi in the human system by applying theoretical methods to aspects of the biological process that is otherwise inaccessible to experimental methods and whose behaviour is hard to predict, i.e. due to nonlinear effects.

6.1 *In sillico* modelling of RNAi in Homo sapiens

The aim of this study was to fit different models of the siRNA-mediated RNAi in the human system to empirical observations, but also to extrapolate the system's behaviour beyond the given data to situations that are difficult to access experimentally. The evaluation of the models provided an understanding of the RNAi process beyond what is already known from direct investigation of the experimental observations.

Experimental findings that were performed both before and in parallel to this doctoral study (with reciprocal influences), suggest that:

- gene silencing rates can be increased by raising the target mRNA concentration, at constant siRNA and hAgo2 concentrations[40]
- the release of the target RNA strand within a RISC bound RNA duplex is accelerated in the presence of free target RNA[32].

On basis of these findings, the following questions were raised and addressed by this modelling study:

- Is it possible to model the adaptability of RNAi-based knockdown to changes in target mRNA level over several magnitudes?
- How can target mRNA-dependent enhancement of gene silencing via RNAi in Homo sapiens be explained mechanistically?
- What implications does an associative strand-exchange mechanism (between products and next target mRNA) have in terms of RNAi efficacy?
- How does target mRNA metabolism influence RNAi efficacy?
- How does siRNA delivery influence RNAi efficacy?
- What implications does an associative strand-exchange mechanism have on the development and the application of RNAi-based diagnostics and therapeutics?

In order to answer these questions three differential equation-based modelling approaches with different levels of complexities and slightly different focuses were built, analysed and evaluated in order to investigate target mRNA-dependent enhancement of gene silencing via associative strand-exchange versus the dissociative strand-exchange mechanism suggested by current literature.

6.1.1 A minimalistic modelling approach to the dissociative and associative strand-exchange pathways of single-stranded siRNA-mediated RNAi

At first, two minimalistic mathematical models of siRNA-mediated RNAi were created in order to link empirical observations with a quantitative and time-resolved understanding of the RNAi mechanism. Mechanistic details like RISC co-factors, passenger strand cleavage and passenger strand release during complex formation involving ds siRNA, target mRNA metabolism and siRNA transfection were omitted intentionally. This was decided to reduce model complexity to a minimum while keeping the focus on the modelling objective, i.e. the strand-exchange mechanism between products and next target mRNA. Some of these aspects were included in later modelling approaches (i.e. passenger strand cleavage and passenger strand release during complex formation involving ds siRNA) or modelled separately (i.e. target mRNA metabolism and siRNA update). Other details were determined as essential but had to be omitted due to a lack of reliable physical data in this first modelling approach: rate laws and rate constants of the modelled biochemical reaction steps. These

limitation were already faced by other previously discussed RNAi modelling studies.
It was not possible to find a set of unambiguous values for the missing and inaccurate model parameters by parameter estimation methods due to combinatorial explosion as there were still too many unknown quantities in the system.
Nonetheless, the minimalistic modelling approach was able to describe the available time-resolved experimental data (i.e. knockdown of target gene activity for different initial siRNA concentrations) fairly well. Time course simulations were identified as a meaningful tool to ensure comparability between experiments and models. The time courses based on the associative model resembled the experimental data (cytosol-extract-based RNAi assay) slightly closer than those based on the dissociative model. The difference was not enough to to fully settle whether the associative or the dissociative strand-exchange mechanism better describes the experimental observations.

6.1.2 An extension towards ds siRNA and accounting for target siRNA affinity modes

The structures of the previously discussed minimalistic models were revised aiming at better accounting for the RNAi process *in vivo* and a more homogenous parametrisation: no precise reaction rate constants existed for the kinetics of ss siRNA binding; while siRNA binding was identified as one of the main reasons for an insufficient adaptation of the onset of the simulation trajectories to the time courses of the validation data sets. This led to an extension of the RNAi models towards ds siRNA binding with more homogenous of model parameters from one single source.
Two different sets of parameters were available: one for a biological active siRNA with high target affinity, as well as, one for an inactive siRNA with low target affinity. An analysis of the degrees of freedom of the available model parameters suggested that the formation of collision complexes between RNA molecules and RISC was diffusion controlled (as opposed to the involvement of an active delivery mechanisms). Thus, the introduction of further complexity into the modelling regarding collision complex formation could be omitted.
Besides time course simulations, the IC_{50} (which is very common for describing inhibitory potentials of biological or biochemical functions) was identified as a key quantitative measure for comparability between experiments and models. The associative model had a higher robustness against variations in target mRNA concentration in a range between 1 and $30,000$ mRNA/cell. Especially at higher concentrations (i.e. above $2,000$ mRNA/cell), the

associative model showed more realistic target knockdown rates within physiological meaningful parameter ranges compared to the dissociative model. This led to the conclusion that the associative model explained experimentally observed robustness (or adaptation) of the RNAi machinery to higher levels of target mRNA.

6.1.3 An excursion towards target mRNA metabolism

The previously discussed modelling approaches excluded the effects of RNA metabolism on the knockout efficacy of RNAi, while transcription of eukaryotic genes is fairly complicated. Two important modi of RNA metabolism were identified and modelled: transcriptional bursting (long periods of inactivity are followed by bursts or pulses of activity) and homogenous synthesis.
The previous modelling approaches assumed a target mRNA concentration without taking into account synthesis or degradation rates. Thus, they were compatible to the knockdown of mRNA species that are transcribed in a burst of activity and of those that are homogeneously transcribed but have a relatively high half time of several hours (i.e. housekeeping transcripts). To be able to work against homogeneously transcribed transcription factor transcripts with low half times, the RNAi machinery required a fast and efficient turnover mechanism compatible with the associative pathway but not with the dissociative one.

6.1.4 An excursion towards the delivery of small RNAs

It was possible to model the cellular uptake of siRNAs via the common transfection reagent Lipofectamine 2000 with an analytical solvable two step reaction kinetic. The first step accounted for the Liposome-facilitated spacial re-location of siRNA between an extracellular compartment (i.e. the transfection buffer) and the cell; a second step had to be added, possibly accounting for the re-location of, so-called, trapped siRNA to freely available siRNA in the cell's cytoplasm. From here, diffusion controlled collision between RNA and RISC components could occur as described in the RNAi models.

6.1.5 Drawing up an interim balance

Both of the above discussed RNAi modelling approaches indicated an associative strand exchange mechanism between the cleaved product fragments and the next incoming target mRNA over an dissociative mechanism:

- as opposed to the dissociative mode, the associative model was able to describe the experimentally observed time courses of the siRNA-mediated knockdown of target gene activity for different initial siRNA concentrations
- the associative model described experimentally determined concentration response curves and IC_{50} values of the siRNA-mediated knockdown of target gene activity at a range of different basal target mRNA concentrations
- the associative model was able to explain the knockdown of target mRNA of a wide range in half times and of different modi of RNA metabolism
- an extension of the model was able to account for siRNA uptake via the common transfection reagent Lipofectamine 2000.

However, the simulations based on the model approaches described the experimental validation data only qualitatively, not quantitatively: the associative model provided meaningful results within physically meaningful parameter ranges; while in the case of the dissociative model, this was not possible. Several shortcoming of the modelling approaches were identified:

- due to the lack of reliable parameter values, the models remained minimalistic in terms of model structure, e.g. intermediate steps of binding between RNA and RISC and the release of cleaved RNA product had to be omitted due to the lack of available parameters
- the available parameters lacked in precision and were a patchwork from heterogeneous sources.

6.1.6 A refined model based on precise quantitative data

New, precise quantitative kinetic data of the fundamental reaction steps of siRNA-mediated RNAi provided the opportunity for a revised modelling approach of the two competing models of the dissociative and associative strand-exchange pathways, respectively. Like the first minimalistic modelling approach, the new approach targeted single-stranded siRNA-mediated RNAi. Mechanistic details like RISC co-factors, passenger strand cleavage and passenger strand release during complex formation involving ds siRNA, target mRNA metabolism and siRNA transfection were omitted to reduce model complexity. However, the new quantitative kinetic data made it possible to precisely model rate laws and parametrise rate

constants of the biochemical reaction steps of the core RNAi process. All reaction step were modelled as elementary reaction equilibrium with either a pair of forward and backward reaction rate constants, or in the case of quasi-irreversible steps, with a forward rate constant. Thus, the new modelling approach overcame the main shortcomings of previous attempts (both, from literature and the ones build during this study).
The associative model involved the association of an incoming ligand with the enzyme / substrate complex before departure of the leaving ligand. The model was based on the theory that Ago2 facilitates the interaction of an incoming target strand with a ternary Ago2 / guide / product complex by a strand invasion mechanism. The model accounted for the experimentally observed target concentration-dependent acceleration of product release from the ternary complex and was compatible with the high multiple turnover rate of RNAi-based gene silencing in living cells. The experimental observations for all available sets of *in vivo* validation and evaluation experiments, i.e.

- time courses and $t_{1/2}$, (the time of half-maximal knockdown) of the siRNA-mediated knockdown of target gene activity for different initial siRNA concentrations
- concentration response curves and IC_{50} values of the siRNA-mediated knockdown of target gene activity at different basal target mRNA concentrations

could be entirely explained by the associative model parametrised with independently measured *in vitro* reaction rate constants, physiological species concentrations and cellular reactions volumes. In contrast, the model structure of the dissociative pathway was not sufficiently consistent with the experimental data and the fit could not be improved by merely adjusting the model parameters within physiological meaningful boundaries. For example, a 20-fold increase of the measured *in vitro* constants of the two most sensitive parameters (i.e. product release III k_{+10} and target binding III k_{+6}), which may be justified by facilitating cofactors *in vivo*, was not sufficient to solve the discrepancy. The discrepancy to the dissociative model was observed for all experimental validation data sets.

6.1.7 Implications of an associative strand exchange mechanism

Thus, this study proposes an associative mechanism of target recognition by a duplex of guide and cleaved target possibly facilitated by Ago2 during siRNA-mediated RNAi. The associative mechanism is further compatible with Wünsche & Sczakiel 2005[226] who quantified the increased kinetics of RNA-RNA recognition that occurred via facilitated strand exchange between RNA duplexes and target RNA strands *in vitro*.

To further verify the general validity of these findings, it is proposed to parametrise the two competing models for further combinations of target mRNA and effector RNA (i.e. siRNA). The pairs should be selected in a way that guarantees a wide variation of knockdown efficacy, binding site sequences and structural motifs compared to the already investigated pair of target mRNA and siRNA.
siRNAs typically induce strong silencing by cleavage of their perfectly complementary targets, whereas miRNAs are guided by imperfect complementarity to induce deadenylation and subsequent degradation. Nevertheless, the mode of action is similar enough to suggest that an associative mechanism could also be present in miRNA-mediated RNAi. This would haven e.g. implications on the use of biomarker miRNAs.

6.1.8 Effect of an associative pathway on target selection

siRNA-base drugs are currently being developed and tested for therapeutic use, e.g. for lowering LDL-cholesterol by targeting of APOB and PCSK9 transcripts[51, 239]. Questions like *What determines the cellular response to an siRNA or miRNA stimulus?* and *How can the response be quantitatively predicted?* are sought after for improved and secure application of RNAi in medicine and life science. It is generally known that properties of the sequence match between effector RNA (i.e. siRNA or miRNA) and target mRNA, i.e. the binding site, are essential for RNAi efficacy. However, there are many examples, where the RNA binding site is not sufficient to describe or predict RNAi knockdown[59, 101]. Recently, further factors which possibly effect the outcome of RNAi efficacy were identified, such as: RNA-binding protein (RBP) motifs[83], local RNA sequence context[59], local RNA structure[93], cellular target RNA levels[3] and the basic dynamics of target turnover[104]. However, even after taking these factors into account Larsson et al.[104] still failed to quantitatively predict RNAi efficacy by several magnitudes. The acceptance of the associative RNAi model and its analyses strongly support Larsson's et al.[104] hypothesis that siRNA efficacy is correlated with target transcript stability. However, Larsson et al.[104] used a phenomenological model to account for additive and synergistic (nonelinear) interactions between mRNA turnover by RNAi and other mRNA decay.

The approach of this doctoral study took a step forward and explicitly modelled the underlying biochemical reaction steps of RNAi (which was impossible with a mere phenomenological approach). The comparison between simulations and experimental IC_{50} studies, suggested that the associative mechanism interlinks target mRNA loading with acceleration of

unloading of cleaved product fragments from RISC at multiple turnover conditions. These data also showed that the abundance of target mRNA rather than the single factor of a synergistic combination of RNAi-unrelated target mRNA turnover (as suggested by Larsson's et al. model[104]) was a, previously unknown, key factor of RNAi efficacy. Nonetheless, the experimental part of Larsson's et al.[104] (i.e. the analysis of qPCR-based siRNA validation data) supports the findings that a high abundance of target transcript leads to an improved RNAi efficacy of the target, while it is in contradiction with the results of two earlier studies[65, 101]. Furthermore, this implies that low abundant mRNAs are more resistant to siRNA off-target effects. For typical target concentrations of 1–100 copies/cell the increase of siRNA concentration from 30 to 300 siRNAs/cell does not lead to better target knock-down efficacy. Thus, it is not worth risking off-target and the side effects associated with transfection of high amounts of siRNA (compare: sensitivity analysis, 4.24.

6.1.9 Overcoming transcriptional bursting related resistance of microbes and cancer cells

Transcriptional bursting, as opposed to homogenous transcription, accounts for the high variability, also called transcriptional noise, in gene expression between isogenic cell populations[174]. This transcriptional variability in turn can have a great (negative) impact on cellular behaviour[26]. On the other die, in certain contexts, e.g. the survival of microbes under rapidly changing stressful conditions, a wide transcriptional variability may be essential[118]. For cells to be able to deal with transcriptional bursting, it must be possible to selectively and efficiently regulate transcriptional variability by post-transcriptional processes such as RNAi. Variability also impacts the effectiveness of medical treatment against antigens; e.g. bacterial resistance against antibiotics may be caused by non-genetic transcriptional variability [113, 146]. Likewise, variability in gene expression may contribute to the resistance of sub-populations of cancer cells to chemotherapy[196]. Robustness of RNAi drugs to transcriptional variation as shown in IC_{50} experiments and modelled by the associative model, may be an important property of RNAi-based drugs to overcome microbial and cancer cell resistance to conventional drugs.

6.1.10 Suggestions for target-prediction

Recent work has suggested a move away from the binary distinction between mRNAs as targets or non-targets, towards a more context-dependent and quantitative approach[3, 168, 193]. The work in this thesis supports this view, in which an mRNA in one cell type may

not be targetable by an effector RNA to the same extent in another cell type, depending on the turnover rate of the mRNA in that context. This suggests that a new generation of bioinformatics tools for siRNA (and in the broad sense, miRNA) target prediction should be developed using system-level properties such as target transcript abundance and mRNA decay rates.

6.2 Conclusion of the computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA

The regulation of the level of gene-specific RNAs, both in post-transcriptional gene regulation and in viral functions, usually comprises structural elements and functional domains of cellular mRNAs or viral transcripts, respectively. Structural and functional RNA modules, located upstream of the translational initiation site (e.g. the AUG codon) or overlapping with it, are often involved in the regulation of protein biosynthesis. Such regulatory local RNA folding elements may interact with modulators such as complementary RNA sequences in the case of microRNA or antisense RNA-mediated control of gene expression (for review see:[148]), RNA-binding proteins (for review see:[38]), or low-molecular metabolites like in the case of ribo-switches[236]. In most cases regulatory RNA elements are composed of consecutively neighbouring sequence segments. Only in rare cases long-range interactions seem to be involved in the formation of regulatory local structures like in IRES elements (e.g. [63, 181]) or certain classes of catalytic RNA (e.g. [134]). Particularly simple regulatory modules of limited length such as the TAR element, the RRE element, or packaging signals of the human immunodeficiency virus type 1 (HIV-1) often adopt structural domains that coincide with local minimal energy states[219, 233].

This doctoral study provides a software tool to identify consecutive RNA structural elements at any given length and nucleotide content formed by distant sequences taking into account local minimal energy states. The tool was used to investigate duplex RNA elements formed by distantly located segments ($\geq$ 40 nt) of complementary RNA in a non-hairpin fashion. Within a systematic study, coding and non-coding RNA sequences of human and non-human origin were investigated and analysed. The software tool identified structural elements with prominent non-hairpin GC-helices within the 5′-region of mRNAs of many different species. Phylogenetic analyses indicated an evolutionary conserved occurrence of

these structural elements in mRNAs transcribed from certain genes and gene families, including the Argonaute genes of vertebrates.

Studies on the biological role of Argonaute genes in human cells indicate a potential post-transcriptional control step involved in regulating endogenous levels of the family member Ago2[132]. Further, preliminary experimental studies by George Sczakiel indicated specific binding of RNA sequences surrounding the AUG start codon of the Ago2 mRNA with proteins involved in RNA interference[39].

It was proposed that these evolutionarily conserved structural elements have a potential biological role. In conjunction with the above mentioned preliminary biological observations, validation experiments indicated a potential role of GC-rich helices for post-transcriptional regulatory processes within the mRNAs of the Argonaute gene family.

A short excursion towards binding site prediction in nucleotide sequences

The here developed software tool can be furthermore helpful to generally predict structural elements in nucleotide sequences possibly preventing or facilitating nucleotide binding sites. It has been successfully applied to diagnostic tools such as PCR (e.g. [103]) and the prediction of miRNA binding sites (e.g. [66]).

6.3 Linking of identification of an associative RNAi pathway and identification of new regulatory elements in Ago2 mRNA

Based on the revised modelling approach, the associative model of siRNA-mediated RNAi described the experimentally observed time courses, IC_{50} values and $t_{1/2}$ values reasonably well. It was possible to select the newly proposed associative strand exchange mechanism over the hitherto well-established theory of a dissociative mechanism. Numerical simulations and analysis of the model gave new insides into the kinetics of the RNAi pathway in the human system. Nonetheless, a discrepancy remains: all simulations showed a slightly lower RNAi efficacy compared to the experimental observations. This hints towards the existence of yet-to-be-identified details within (or connected to) the RNAi mechanism.

In the second part of the thesis, a bioinformatics tool for RNA motif prediction was developed and successfully used (in complementation to experimental methods) for the identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA. These findings further showed that RNAi in the human systems is a complex process with many unknowns yet to discover.

Both of the two very different computational approaches to RNAi showed that additional unravelling of the RNAi process is essential before a safe introduction of RNAi-based therapeutics will become possible. Furthermore, the development of RNAi-based diagnostics and other tools will largely benefit from a better understanding of the mechanistic details. The two studies also showed that computational methods can be a useful complement to experimental biology for the investigation of RNAi.
There are several conceivable advancements for the here developed RNAi modelling approaches and bioinformatics tool for RNA motif prediction. The incorporation of sequential and structural information about RNA binding sites in combination with the newly proposed associative model could bring kinetic modelling of RNAi to a whole new level. This would be a further step towards a more precise prediction of RNAi efficacy as it would account for concrete pairs of target mRNA and effector RNA (i.e. siRNA or miRNA). It is proposed to select a set of target-effector pairs with a wide range of different knockdown efficacies and diverse features (with an affect on RNAi efficacy), such as: binding sites, local sequence context[59], local RNA structure[93], RNA-binding protein (RBP) motifs[83], total target abundance[3] and basic dynamics of target turnover[104]. In this context, the bioinformatics tool for for structural motif prediction (developed within the scope of this doctoral study) is recommended as it provides flexible configuration options and the option for process automatisation. Experimental set-ups for the determination of knockdown efficacies by time courses, IC_{50} values and $t_{1/2}$ values are described e.g. in Dornseifer et al.[40]. To be able to computationally model the RNAi knockdown of a target-effector pair, new *in vitro*-derived pre-steady state kinetic data is needed for parametrisation of the biochemical reaction steps. The experiments necessary to extract the data are fairly time consuming[32]. To avoid this bottleneck, it is proposed to replace the experimental calculation of some of the rate constants with predicted values from machine learning approaches. It is suggested to either predict parameters that are experimentally expensive to measure or those that were identified as less influential on the simulation outcome (see: sensitivity analysis results, pp. 134), i.e. where inaccuracies are tolerable. If successful, the new combination of kinetic modelling and machine learning will be able to predict time-resolved RNAi efficacy based

on targeted mRNA and effector RNA.

Appendix A

Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA – additional data

A.1 RNA sequences used in this study

name	origin	Accession
ACIN1	Homo sapiens	NM_014977
ALDH7A1	Homo sapiens	NM_001182
API5-1	Homo sapiens	NM_001142930
APP-2	Homo sapiens	NM_201413
ATG13-1	Homo sapiens	NM_001142673
ATP1A1-4	Homo sapiens	NM_001160234
BCKDHB-1	Homo sapiens	NM_183050
CCDC120-1	Homo sapiens	NM_001163321
CD36	Homo sapiens	NM_001001548
COL2A-1	Homo sapiens	NM_001844
COX2	Homo sapiens	NM_000963
CP-2	Homo sapiens	NM_000096
DMD-Dp71b	Homo sapiens	NM_004016

Continued on next page

Table A.1		– continued from previous page
name	origin	Accession
E2F3	Homo sapiens	NM_001949
FGFR2-7	Homo sapiens	NM_001144917
FGFR3-1	Homo sapiens	NM_000142
GRM7-1	Homo sapiens	NM_000844
HIV-1-pNL4-3	Homo sapiens	AF324493
ICAM1	Homo sapiens	P05362 (SwissProt)
INPPL1	Homo sapiens	NM_001567
IPO13	Homo sapiens	NM_014652
ITG-AV	Homo sapiens	NM_002210
ITGAM	Homo sapiens	NM_001145808
KTN-1	Homo sapiens	NM_182926
LAMB3	Homo sapiens	NM_001127641
LAP2-1	Homo sapiens	NM_001017963
MANEA	Homo sapiens	BC146671
MYB	Homo sapiens	M15024
MYB-1	Homo sapiens	NM_001130173
NHE5	Homo sapiens	NM_004594
NOS3-1	Homo sapiens	NM_000603
PGY1	Homo sapiens	M14758
PLOD2	Homo sapiens	NM_000935
PRKD1	Homo sapiens	NM_002742
PROM1	Homo sapiens	NM_006017
RAD4-2	Homo sapiens	NM_001145769
RAF1	Homo sapiens	XM_051583
Rab11-FIP3-1	Homo sapiens	NM_014700
SAPS3-2	Homo sapiens	NM_001164163
SCIDA-2	Homo sapiens	NM_022487
SELE	Homo sapiens	XM_057446
SLC25A12	Homo sapiens	NM_003705
SLC6A4	Homo sapiens	NM+AF8-001045
SLCO2B1	Homo sapiens	NM_007256
STOX2	Homo sapiens	BC146754

Continued on next page

Table A.1		– continued from previous page
name	origin	Accession
STXBP1-1	Homo sapiens	NM_003165
TCF7L2-2	Homo sapiens	NM_030756
TIAL1-2	Homo sapiens	NM_001033925
TMEM2	Homo sapiens	BC146780
TRP3-1	Homo sapiens	NM_001130698
UNC51-2	Homo sapiens	NM_001142610
USP19	Homo sapiens	BC146752
WFS1	Homo sapiens	NM_001145853
ZNF828	Homo sapiens	NM_001164145
hAGO1 (EIF2C1)	Homo sapiens	NM_012199
hAGO2 (EIF2C2)	Homo sapiens	NM_012154
hAGO3-1 (EIF2C3)	Homo sapiens	NM_024852
hAGO3-2 (EIF2C3)	Homo sapiens	NM_177422
hAGO4 (EIF2C4)	Homo sapiens	NM_017629
phosphodiesterase	Homo sapiens	L20971
PIWIL1 (HIWI)	Homo sapiens	NM_004764
PIWIL2-1 (HILI-1)	Homo sapiens	NM_001135721
PIWIL2-2 (HILI-2)	Homo sapiens	NM_018068
PIWIL3	Homo sapiens	NM_001008496
PIWIL4 (HIWI2, HILI2)	Homo sapiens	NM_152431
DCR1-T1	Homo sapiens	NM_030621
DCR1-T2	Homo sapiens	NM_177438
RNASEN-T1	Homo sapiens	NM_013235
RNASEN-T2	Homo sapiens	NM_001100412
TP53-ISOA-T1	Homo sapiens	NM_000546
TP53-ISOA-T2	Homo sapiens	NM_001126112
TP53-ISOB-T3	Homo sapiens	NM_001126114
TP53-ISOC-T4	Homo sapiens	NM_001126113
TP53-ISOD-T5	Homo sapiens	NM_001126115
TP53-ISOE-T6	Homo sapiens	NM_001126116
TP53-ISOF-T7	Homo sapiens	NM_001126117
Ago3b	Danio rerio	NM_001160028

Continued on next page

Table A.1		– continued from previous page
name	origin	Accession
Ago4	Danio rerio	NM_001252559
ZILI (piwil2)	Danio rerio	NM_001080199
ZIWI (piwil1)	Danio rerio	NM_183338
AGO1	Rattus norvegicus	NM_001191765
AGO2	Rattus norvegicus	NM_001271193
AGO4 (Eif2c4)	Rattus norvegicus	NM_001106686
PIWIL1	Rattus norvegicus	NM_001108853
PIWIL2-2	Rattus norvegicus	NM_001107276
MILI (PIWIL2)	Mus musculus	NM_021308
MIWI (PIWIL1)	Mus musculus	NM_021311
MIWI2 (PIWIL4)	Mus musculus	AB258534
AGO1clone (Eif2c1, meIF2C1)	Mus musculus	AK080954
AGO2 (Eif2c2, GERp95, meIF2C2)	Mus musculus	NM_153178
AGO2 (alternative)	Mus musculus	NM_153178
AGO3 (Eif2c3)	Mus musculus	NM_153402
AGO4 (Eif2c4)	Mus musculus	NM_153177
AGO1-A	Drosophila melanogaster	NM_166020
AGO1-B	Drosophila melanogaster	NM_079010
AGO1-C	Drosophila melanogaster	NM_166021
AGO2-B	Drosophila melanogaster	NM_140518
AGO2-C	Drosophila melanogaster	NM_168626
AGO2-C (alternative)	Drosophila melanogaster	NM_168626
AGO2-E	Drosophila melanogaster	NM_001274953
AGO3-D	Drosophila melanogaster	NM_001043162
AGO3-E	Drosophila melanogaster	NM_001043164
AGO3-F	Drosophila melanogaster	NM_001043163
AUB-A	Drosophila melanogaster	NM_057386
AUB-C	Drosophila melanogaster	NM_001103674
PIWI	Drosophila melanogaster	NM_057527
AGO2 (EIF2C2)	Oryctolagus cuniculus	NM_001082710
PIWIL2-t2 (predicted)	Canis lupus familiaris	XM_853957
EIF2C1-t1 (predicted)	Pan troglodytes	XM_001167251

Continued on next page

Table A.1		– continued from previous page
name	origin	Accession
Ago (TSA)	Pan troglodytes	GABF01000457
Ago (TSA)	Pan troglodytes	GABD01009992
EIF2C1-t2 (predicted)	Pan troglodytes	XM_001167312
EIF2C1-t3 (predicted)	Pan troglodytes	XM_001167349
EIF2C1-t4 (predicted)	Pan troglodytes	XM_513312
EIF2C2-t1 (predicted)	Pan troglodytes	XM_001142838
EIF2C2-t2 (predicted)	Pan troglodytes	XM_001143064
EIF2C3-t1 (predicted)	Pan troglodytes	XM_001167488
EIF2C3-t2 (predicted)	Pan troglodytes	XM_001167515
EIF2C3-t3 (predicted)	Pan troglodytes	XM_001167651
EIF2C3-t4 (predicted)	Pan troglodytes	XM_001167675
EIF2C3-t5 (predicted)	Pan troglodytes	XM_524664
Ago1	Bos taurus	NM_001205899
Ago2	Bos taurus	BT030588
Ago2 (alternative)	Bos taurus	BC151491
Ago3	Gallus gallus	NM_001030900
Ago4	Gallus gallus	NM_001039276
Ago (TSA)	Macaca mulatta	JU475189
Ago (TSA)	Mustela putorius furo	JP009204
Ago2	Oryctolagus cuniculus	NM_001082710
Ago1	Sus scrofa	NM_001194976
Ago2	Sus scrofa	NM_001194975
Ago3	Sus scrofa	NM_001194974
Ago4	Sus scrofa	NM_001194972
Ago1 (clone)	Xenopus laevis	BC077863
Ago2	Xenopus laevis	NM_001093519
Ago2 (clone)	Xenopus laevis	CR761661
Ago4	Xenopus laevis	NM_001096105
HOTAIR-t1	Homo sapiens	NR_047517
HOTAIR-t2	Homo sapiens	NR_003716
JPX	Homo sapiens	NR_024582
PCAT1	Homo sapiens	HQ605084

Continued on next page

Table A.1		– continued from previous page
name	origin	Accession
PCAT14	Homo sapiens	HQ605085
SUZ12	Homo sapiens	NM_015355
Tsix	Homo sapiens	NR_003255
lincRNA-ST8SIA3	Homo sapiens	HQ315778
intergenic non-protein coding RNA	Homo sapiens	JX088243
Tsix	Homo sapiens	NR_002844
Xist-t1	Mus musculus	NR_001463
Xist-t2	Mus musculus	NR_001570
AIR-t1	Mus musculus	DQ275619
AIR-t1b	Mus musculus	DQ275617
AIR-t2	Mus musculus	DQ275620
AIR-t3	Mus musculus	DQ275618
lgf2r-as-Airn-t1	Mus musculus	NR_027773
lgf2r-as-Airn-t2	Mus musculus	NR_027772
lgf2r-as-Airn-t3	Mus musculus	DQ220013
lgf2r-as-Airn-t4	Mus musculus	NR_027784
lgf2r-as-t1	Mus musculus	DQ220010
lgf2r-as-t2	Mus musculus	DQ220012
lgf2r-as-t3	Mus musculus	NR_002853
lgf2r-as-t4	Mus musculus	DQ220011
lgf2r-as-t5	Mus musculus	DQ220014
Xist	Bos taurus	NR_001464

Table A.1 **List of mRNA sequences included in RNA motif prediction.** mRNAs with intact 3′-UTR, CDS, and 5′-UTR. 77 sequences of human origin, including 10 mRNAs of the Argonaute gene family. 49 sequences of non-human Argonaute orthologs plus 13 predicted ones. 28 non-coding RNA transcripts from Homo sapiens and Mus musculus. All sequences are freely available online at resources like EMBL Nucleotide Sequence Database and NCBI Nucleotide. This Table is based on Supplementary Table S2B of[39].

A.2 Sequences of pRL-Ago2 and pRL-Ago2^{mut} plasmids

This Section is based on Supplementary Table S3 of[39].

Plasmid pRL-AGO2 (pRL-CMV_Ago2_244_SacI_NheI)

AGATCTTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATA
CGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTGATTATTGACT
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAG
GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCA
AGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGAT
AGCGGTTTGACTCACGGGGATTT**CCAAGTCTCCACCCCATTGACG**TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT**GACGCAAATGGGCGGTAGG**CGTGTACGGTGGGAGGTCTATATAAG
CA**gagctc**GTTTAGTGAACCGGGCCGCCCCTCGGCCCCGGAGCCCCTCGGCGGCGCCACCATGTACTCGGGAGCCGGCCC
CGCACTTGCACCTCCTGCGCCGCCGCCCCCCATCCAAGGATATGCCTTCAAGCCTCCACCTAGACCCGACTTTGGGACCCT
CCGGGAGAACAATCAAATTACAGGCCAATTTCTTCGAAATGGACATCCCCAAAATTGACATCTATCATTATGAATTGGAT
ATCAAGCCAGAGAAGTGCCCGAGGggtggcggaggtggcggattt**gctagc**CACCATGACTTCGAAAGTTTATGATCCAGA
ACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGAATGTTCTTGATTCATTTATTAATT
ATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATTTTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACAT
GTTGTGCCACATATTGAGCCAGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAA
TGGTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTACCAAAGAAGATCATTT
TTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTATGAGCATCAAGATAAGATCAAAGCAATAGTTCAC
GCTGAAAGTGTAGTAGATGTGATTGAATCATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGA
AGAAGGAGAAAAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAAGTTAGAAC
CAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTCCAACATTATCATGGCCTCGTGAA
ATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTACAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGA
TTTACCAAAAATGTTTATTGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGAAAATATATCAAATCGTTC
GTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGAGCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTT
GGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCAT
TATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTT
TTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTG
TTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTG
GATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCT
ATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGG
TTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGG
GGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGC
CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGAC
TGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGG
AGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAA
CTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGAT
CCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTT
TGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCT
TCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC
CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGG
TCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTG
CTGGCCTTTTGCTCACATGGCTCGAC

Plasmid pRL-AGO2mut (pRL-CMV_Ago2_244_SacI_NheI_D1/D2mut)

AGATCTTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATA
CGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTGATTATTGACT
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAG
GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCA
AGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGAT
AGCGGTTTGACTCACGGGGATTT**CCAAGTCTCCACCCCATTGACG**TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT**GACGCAAATGGGCGGTAGG**CGTGTACGGTGGGAGGTCTATATAAG
CA**gagctc**GTTTAGTGAACCG**AAAA**GCCCCTCGGCCCCGGAGCCCCT**AAAA**GGCGCCACCATGTACTCGGGAGCCGGCCC
CGCACTTGCACCTCCTGCGCCGCCGCCCCCCATCCAAGGATATGCCTTCAAGCCTCCACCTAGACCCGACTTTGGGACCT
CCGGGAGAACAATCAAATTACAGGCCAATTTCTTCGAAATGGACATCCCCAAAATTGACATCTATCATTATGAATTGGAT
ATCAAGCCAGAGAAGTGCCCGAGGggtggcggaggtggcggattt**gctagc**CACCATGACTTCGAAAGTTTATGATCCAGA
ACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGAATGTTCTTGATTCATTTATTAATT
ATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATTTTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACAT
GTTGTGCCACATATTGAGCCAGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAA
TGGTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTACCAAAGAAGATCATTT
TTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTATGAGCATCAAGATAAGATCAAAGCAATAGTTCAC
GCTGAAAGTGTAGTAGATGTGATTGAATCATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGA
AGAAGGAGAAAAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAAGTTAGAAC
CAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTCCAACATTATCATGGCCTCGTGAA
ATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTACAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGA
TTTACCAAAAATGTTTATTGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGAAAATATATCAAATCGTTC
GTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGAGCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTT
GGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCAT
TATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTT
TTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTG
TTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTG
GATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCT
ATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGG
TTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG
AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGG
GGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGC
CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGAC
TGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGG
AGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAA
CTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGAT
CCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTT
TGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCT
TCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC
CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGG
TCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTG
CTGGCCTTTTGCTCACATGGCTCGAC

colour coding:

blue.....	CMV promoter
brown background.....	TATA-Box / start of transcription for CMV
lowercase underlined...	cloning site SacI/NheI
UPPERCASE UNDELINED....	sequence primer
red background.....	Ago2 (244nt)
red underlined.....	mutations Ago2
red background, white..	AUG Ago2
yellow background.....	ORF Renilla-Luciferase
purple background.....	glycin linker

Appendix B

Nomenclature

2D	two-dimensional
3D	three-dimensional
$3'$	three prime: DNA or RNA strand tail end
$5'$	five prime: DNA or RNA strand head end
AGO	Argonaute
Ago2	Argonaute 2
APOB	apolipo protein B
asON	antisense oligonucleotide
asso	associative
AT	adenine-thymine Watson-Crick base pair
AUG	the most common start codon
BDF	backwards differential formula
bp	base pair
C	concentration
cDNA	complementary DNA
CDS	coding sequence, the coding region of a gene
CMV	Human cytomegalovirus
COPASI	Complex Pathway Simulator
CPU	central processing unit
crtl	control
CTAB	cetyltrimethylammonium bromide
cyt	cytoplasm
d	day(s)
diss	dissociative

DNA	desoxyribonucleic acids
ds	double-stranded
ε	sensitivity coefficient
ECV	an endothelial cell line
e.g.	for example
Eq.	equation
EGFP	enhanced green fluorescent protein
Fig.	figure
GA	Genetic algorithm
GB	giga byte
GC	guanine-cytosine Watson-Crick base pair
GHz	giga hertz
h	hour(s)
hAgo	human Argonaute
HDD	hard disk drive
HeLa	an immortal cell line
HIV	human immunodeficiency viruses
IC_{50}	half maximal inhibitory concentration
ICAM	intercellular adhesion molecule
i.e.	that is to say
IRES	internal ribosomal entry site
λ	half-time
k	reaction rate constant
kcal	kilo calories
LDL	low-density lipoprotein
LF	Lipofectamine
log	decimal logarithm
LSODA	livermore solver for ordinary differential equations
LU	lower–upper (decomposition or factorization)
m	meter(s)
M	Mole
MCA	metabolic control analysis
n/a	not applicable
min	minute(s)
miRNA	microRNA

MHz	mega hertz
μm	micro meter(s)
μM	micro Mole
μl	micro litre
mRNA	messenger RNA
mut	mutation
$\mathbb{N}$	natural numbers
$\mathbb{N}^0$	natural numbers and zero
ng	nano gram
nm	nano meter(s)
NCBI	National Center for Biotechnology Information
nM	nano Mole
nt(s)	nucleotide(s)
ODE	ordinary differential equation
p53	tumor protein 53
p.	page
PAZ	PIWI, AGO and Zwille (domain)
PCR	polymerase chain reaction
PCSK9	proprotein convertase subtilisin/kexin type 9
PDE	partial differential equations
PDF	probability density function
piRNA	piwi-interacting RNA
PIWI	P-element induced wimpy testis
pl	pico litre
pp.	pages
PRAXIS	PRincipal Axis (algorithm)
PTGS	post-transcriptional gene silencing
qPCR	quantitative polymerase chain reaction
RAM	random-access memory
RBP	RNA-binding protein
rCDS	remaining coding region
RISC	RNA-induced-silencing-complex
RL	renilla luciferase
RLU	relative light unit
RMSD	root-mean-square deviation

RNA	ribonucleic acid
RNAi	RNA interference
RNAs	ribonucleic acids
RRE	(HIV-1) Rev response element
s	second(s)
SD	standard deviation
ss	single-stranded
si	small interfering
si1	a small interfering RNA
si2b	a small interfering RNA
siRNA	small interfering RNA
siSCR	a small interfering RNA
shRNA	short hairpin RNA
SSA	stochastic simulation algorithm
$t_{1/2}$	time of half-maximal target knockdown
TAR	trans-activation response (element)
ULTD	unlimited
UTR	untranslated region
V	volume
wt	wild type

Bibliography

[1] Alfonsi, A., Cancès, E., Turinici, G., Di Ventura, B., & Huisinga, W. (2005). Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. *ESAIM: Proceedings*, 14, 1–13.

[2] Arciero, J. C., Jackson, T. L., & Kirschner, D. E. (2004). A mathematical model of tumor-immune evasion and siRNA treatment. *Discrete and Continuous Dynamical Systems*, 4(1), 39–58.

[3] Arvey, A., Larsson, E., Sander, C., Leslie, C. S., & Marks, D. S. (2010). Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular systems biology*, 6(363), 363.

[4] Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals.

[5] Bäck, T., Fogel, D. B., & Michalewicz, Z. (1997). *Handbook of Evolutionary Computation*, volume 2. Inst of Physics.

[6] Bäck, T. & Schwefel, H.-P. (1993). An Overview of Evolutionary Algorithms for Parameter Optimization.

[7] Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., Gupta, P. B., Wadlow, R. C., Le, H., Hoersch, S., Wittner, B. S., Ramaswamy, S., Livingston, D. M., Sabatini, D. M., Meyerson, M., Thomas, R. K., Lander, E. S., Mesirov, J. P., Root, D. E., Gilliland, D. G., Jacks, T., & Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, 462, 108–112.

[8] Bartlett, D. W. & Davis, M. E. (2006). Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. *Nucleic acids research*, 34(1), 322–33.

[9] Bartlett, D. W. & Davis, M. E. (2007). Effect of siRNA nuclease stability on the in vitro and in vivo kinetics of siRNA-mediated gene silencing. *Biotechnology and bioengineering*, 97(4), 909–21.

[10] Bartlett, D. W. & Davis, M. E. (2008). Impact of tumor-specific targeting and dosing schedule on tumor growth inhibition after intravenous administration of siRNA-containing nanoparticles. *Biotechnology and bioengineering*, 99(4), 975–85.

[11] Bergstrom, C. T., McKittrick, E., & Antia, R. (2003). Mathematical models of RNA silencing: unidirectional amplification limits accidental self-directed reactions. *Proceedings of the National Academy of Sciences of the United States of America*, 100(20), 11511–6.

[12] Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818), 363–6.

[13] Bramsen, J. B. & Hansen, T. B. (2013). RNA Interference Pathways and Therapeutic Exploitation. In K. A. Howard (Ed.), *RNA Interference from Biology to Therapeutics* (pp. 1–29). New York: Springer US, 1 edition.

[14] Brase, J. C., Johannes, M., Schlomm, T., Fälth, M., Haese, A., Steuber, T., Beissbarth, T., Kuner, R., & Sültmann, H. (2011). Circulating miRNAs are correlated with tumor progression in prostate cancer. *International journal of cancer. Journal international du cancer*, 128(3), 608–16.

[15] Brown, K. M., Chu, C.-Y., & Rana, T. M. (2005). Target accessibility dictates the potency of human RISC. *Nature structural & molecular biology*, 12(5), 469–470.

[16] Buchon, N. & Vaury, C. (2006). RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity*, 96(2), 195–202.

[17] Butcher, S. E. & Pyle, A. M. (2011). The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of chemical research*, 44(12), 1302–11.

[18] Cannell, I. G., Kong, Y. W., & Bushell, M. (2008). How do microRNAs regulate gene expression? *Biochemical Society transactions*, 36(Pt 6), 1224–31.

[19] Cao, Y., Gillespie, D., & Petzold, L. (2005). Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *Journal of Computational Physics*, 206, 395–411.

[20] Cao, Y., Li, H., & Petzold, L. (2004). Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *The Journal of chemical physics*, 121, 4059–4067.

[21] Carrington, J. C. & Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science (New York, N.Y.)*, 301(5631), 336–8.

[22] Carthew, R. W. & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136, 642–655.

[23] Castanotto, D. & Rossi, J. J. (2009). The promises and pitfalls of RNA-interference-based therapeutics. *Nature*, 457(7228), 426–33.

[24] Cerutti, H. & Casas-Mollano, J. A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. *Current genetics*, 50(2), 81–99.

[25] Chang, K. & Hannon, G. J. (2011). Tools for studying and using small RNAs : from pathways to functions to therapies. *Nature Reviews Genetics*, (pp. 11111).

[26] Chubb, J. R., Trcek, T., Shenoy, S. M., & Singer, R. H. (2006). Transcriptional pulsing of a developmental gene. *Current biology : CB*, 16(10), 1018–25.

[27] Coelho, T., Adams, D., Silva, A., Lozeron, P., Hawkins, P. N., Mant, T., Perez, J., Chiesa, J., Warrington, S., Tranter, E., Munisamy, M., Falzone, R., Harrop, J., Cehelsky, J., Bettencourt, B. R., Geissler, M., Butler, J. S., Sehgal, A., Meyers, R. E., Chen, Q., Borland, T., Hutabarat, R. M., Clausen, V. a., Alvarez, R., Fitzgerald, K., Gamba-Vitalo, C., Nochur, S. V., Vaishnaw, A. K., Sah, D. W. Y., Gollob, J. a., & Suhr, O. B. (2013). Safety and efficacy of RNAi therapy for transthyretin amyloidosis. *The New England journal of medicine*, 369, 819–29.

[28] Cohen, L. S. & Studzinski, G. P. (1967). Correlation between cell enlargement and nucleic acid and protein content of HeLa cells in unbalanced growth produced by inhibitors of DNA synthesis. *Journal of cellular physiology*, 69(3), 331–9.

[29] Cuccato, G., Polynikis, A., & Bernardo, D. (2011a). Supplementary Material for Modeling RNA interference in mammalian cells. *BMC systems biology*, 5(1), 1–3.

[30] Cuccato, G., Polynikis, A., Siciliano, V., Graziano, M., di Bernardo, M., & di Bernardo, D. (2011b). Modeling RNA interference in mammalian cells. *BMC systems biology*, 5(1), 19.

[31] Cullen, B. R. (2006). Is RNA interference involved in intrinsic antiviral immunity in mammals? *Nature immunology*, 7(6), 563–7.

[32] Deerberg, A., Willkomm, S., & Restle, T. (2013). Minimal mechanistic model of siRNA-dependent target RNA slicing by recombinant human Argonaute 2 protein. *Proceedings of the National Academy of Sciences of the United States of America*.

[33] Detzer, A. (2010). *Small interfering RNA (siRNA): Zellulare Einschleusung und Wirkmechanismen*. Phd thesis, Universität zu Lübeck.

[34] Dickmeis, T. & Müller, F. (2005). The identification and functional characterisation of conserved regulatory elements in developmental genes. *Briefings in functional genomics & proteomics*, 3(4), 332–50.

[35] Dignam, J. D., Lebovitz, R. M., & Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research*, 11, 1475–1489.

[36] Ding, Y. (2003). A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Research*, 31(24), 7280–7301.

[37] Ding, Y., Chan, C. Y., & Lawrence, C. E. (2005). RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA (New York, N.Y.)*, 11(8), 1157–66.

[38] Doetsch, M., Schroeder, R., & Fürtig, B. (2011). Transient RNA-protein interactions in RNA folding. *The FEBS journal*, 278(10), 1634–42.

[39] Dornseifer, S. & Sczakiel, G. (2013). Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA. *BMC bioinformatics*, 14(1), 122.

[40] Dornseifer, S., Willkomm, S., Far, R.-K., Liebschwager, J., Beltsiou, F., Frank, K., Laufer, S., Martinetz, T., Sczakiel, G., Claussen, J., & Restle, T. (2015). RNAi revised - Target mRNA-dependent enhancement of gene silencing. *Nucleic Acids Research*, 43(22).

[41] Doudna, J. & Sarnow, P. (2007). Translation initiation by viral internal ribosome entry sites. In *Translational control in biology and medicine* (pp. 129–153). Cold Spring Harbor Laboratory Press.

[42] Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., & Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. *Cell*, 150(1), 100–10.

[43] Elmén, J., Lindow, M., Schütz, S., Lawrence, M., Petri, A., Obad, S., Lindholm, M., Hedtjärn, M., Hansen, H. F., Berger, U., Gullans, S., Kearney, P., Sarnow, P., Straarup, E. M., & Kauppinen, S. (2008). LNA-mediated microRNA silencing in non-human primates. *Nature*, 452, 896–899.

[44] Eulalio, A., Huntzinger, E., & Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nature structural & molecular biology*, 15, 346–353.

[45] Evans, M. R., Bithell, M., Cornell, S. J., Dall, S. R. X., Diaz, S., Emmott, S., Ernande, B., Grimm, V., Hodgson, D. J., Lewis, S. L., Mace, G. M., Morecroft, M., Moustakas, A., Murphy, E., Newbold, T., Norris, K. J., Petchey, O., Smith, M., Travis, J. M. J., & Benton, T. G. (2013). Predictive systems ecology. *Proceedings of the Royal Society B-Biological Sciences*, 280(1771), 9.

[46] Fabian, M. R. & Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature structural & molecular biology*, 19(6), 586–93.

[47] Fabian, M. R., Sonenberg, N., & Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry*, 79, 351–379.

[48] Fang, G., Bhardwaj, N., Robilotto, R., & Gerstein, M. B. (2010). Getting started in gene orthology and functional analysis. *PLoS computational biology*, 6(3), e1000703.

[49] Fell, D. (1997). *Understanding the Control of Metabolism*. London: Portland Press.

[50] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669), 806–11.

[51] Frank-Kamenetsky, M., Grefhorst, A., Anderson, N. N., Racie, T. S., Bramlage, B., Akinc, A., Butler, D., Charisse, K., Dorkin, R., Fan, Y., Gamba-Vitalo, C., Hadwiger, P., Jayaraman, M., John, M., Jayaprakash, K. N., Maier, M., Nechev, L., Rajeev, K. G.,

Read, T., Röhl, I., Soutschek, J., Tan, P., Wong, J., Wang, G., Zimmermann, T., De Fougerolles, A., Vornlocher, H. P., Langer, R., Anderson, D. G., Manoharan, M., Koteliansky, V., Horton, J. D., & Fitzgerald, K. (2008). Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), 11915–11920.

[52] Fritz, J. H., Girardin, S. E., & Philpott, D. J. (2006). Innate immune defense through RNA interference. *Science's STKE : signal transduction knowledge environment*, 2006(339), pe27.

[53] Fujioka, A., Terai, K., Itoh, R. E., Aoki, K., Nakamura, T., Kuroda, S., Nishida, E., & Matsuda, M. (2006). Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *The Journal of biological chemistry*, 281(13), 8917–26.

[54] Gegenfurtner, K. R. (1992). PRAXIS: Brent's algorithm for function minimization.

[55] Gibson, M. a. & Bruck, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104, 1876–1889.

[56] Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4), 403–434.

[57] Gokhale, S. a. & Gadgil, C. J. (2012). Analysis of miRNA regulation suggests an explanation for 'unexpected' increase in target protein levels. *Molecular bioSystems*, 8(3), 760–5.

[58] Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6), 1025–36.

[59] Grimson, A., Farh, K. K. H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*, 27, 91–105.

[60] Groenenboom, M. C., Athanasius, M. F., & Hogeweg, P. (2005). The RNA silencing pathway: the bits and pieces that matter. *PLoS computational biology*, 1(2), 155–65.

[61] Grosshans, H. & Chatterjee, S. (2010). MicroRNAses and the regulated degradation of mature animal miRNAs. *Advances in experimental medicine and biology*, 700, 140–155.

[62] Guan, B.-J., Su, Y.-P., Wu, H.-Y., & Brian, D. A. (2012a). Genetic Evidence of a Long-Range RNA-RNA Interaction between the Genomic 5' -Untranslated Region and Nonstructural Protein 1 Coding Region in Murine and Bovine Coronaviruses. *Journal of virology*, (pp. JVI.06265–11–).

[63] Guan, B.-J., Su, Y.-P., Wu, H.-Y., & Brian, D. A. (2012b). Genetic Evidence of a Long-Range RNA-RNA Interaction between the Genomic 5' Untranslated Region and the Nonstructural Protein 1 Coding Region in Murine and Bovine Coronaviruses. *Journal of Virology*, 86(8), 4631–4643.

[64] Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., & Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science (New York, N.Y.)*, 315, 1587–1590.

[65] Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466, 835–840.

[66] Haas, U., Sczakiel, G., & Laufer, S. D. (2012). MicroRNA-mediated regulation of gene expression is affected by disease-associated SNPs within the 3'-UTR via altered RNA structure. *RNA Biology*, 9(6), 924–937.

[67] Haley, B. & Zamore, P. D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nature structural & molecular biology*, 11(7), 599–606.

[68] Hamilton, A. J. & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science (New York, N.Y.)*, 286, 950–952.

[69] Hanke, M., Hoefig, K., Merz, H., Feller, A. C., Kausch, I., Jocham, D., Warnecke, J. M., & Sczakiel, G. (2010). A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. *Urologic oncology*, 28(6), 655–61.

[70] Hargrove, J. L. (1993). Microcomputer-assisted kinetic modeling of mammalian gene expression. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 7(12), 1163–70.

[71] Hargrove, J. L. & Schmidt, F. H. (1989). The role of mRNA and protein stability in gene expression. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 3(12), 2360–70.

[72] Haseltine, E. L. & Rawlings, J. B. (2002). Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *The Journal of Chemical Physics*, 117, 6959–6969.

[73] Hastie, N. D. & Bishop, J. O. (1976). The expression of three abundance classes of messenger RNA in mouse tissues. *Cell*, 9(4), 761–774.

[74] Hausser, J., Syed, A. P., Selevsek, N., van Nimwegen, E., Jaskiewicz, L., Aebersold, R., & Zavolan, M. (2013). Timescales and bottlenecks in miRNA-dependent gene regulation. *Molecular systems biology*, 9, 711.

[75] Heinrich, R. & Rapoport, T. A. (1974). A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *European journal of biochemistry / FEBS*, 42(1), 89–95.

[76] Hemann, M. T., Fridman, J. S., Zilfou, J. T., Hernando, E., Paddison, P. J., Cordon-Cardo, C., Hannon, G. J., & Lowe, S. W. (2003). An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. *Nature genetics*, 33, 396–400.

[77] Hindmarsh, A. (1983). ODEPACK, a systematized collection of ODE solvers. In S. RS (Ed.), *IMACS Transactions on Scientific Computation In IMACS Transactions on Scientific Computation, Vol. 1* (pp. 55–64). Amsterdam, North-Holland: Scientific Computing, 1 edition.

[78] Hodgkin, A. L. & Huxley, A. F. (1990). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bulletin of Mathematical Biology*, 52(1-2), 25–71.

[79] Hofmeyr, J. H. (1995). Metabolic regulation: a control analytic perspective. *Journal of bioenergetics and biomembranes*, 27, 479–490.

[80] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., & Kummer, U. (2006). COPASI–a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)*, 22(24), 3067–74.

[81] Huang, H., Winter, E. E., Wang, H., Weinstock, K. G., Xing, H., Goodstadt, L., Stenson, P. D., Cooper, D. N., Smith, D., Albà, M. M., Ponting, C. P., & Fechtel, K. (2004). Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome biology*, 5(7), R47.

[82] Ipsaro, J. J., Joshua-Tor, L., Haase, A. D., Hannon, G. J., & Knott, S. R. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis.

[83] Jacobsen, A., Wen, J., Marks, D. S., & Krogh, A. (2010). Signatures of RNA binding proteins globally coupled to effective microRNA target sites. *Genome Research*, 20(8), 1010–1019.

[84] Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J. C., Fritzler, M. J., & Chan, E. K. L. (2005). Disruption of GW bodies impairs mammalian RNA interference. *Nature cell biology*, 7(12), 1267–74.

[85] Janas, M. M., Wang, B., Harris, A. S., Aguiar, M., Shaffer, J. M., Subrahmanyam, Y. V. B. K., Behlke, M. A., Wucherpfennig, K. W., Gygi, S. P., Gagnon, E., & Novina, C. D. (2012). Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. *RNA (New York, N.Y.)*, 18(11), 2041–55.

[86] Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., Minna, J. D., & Corey, D. R. (2006). Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature structural & molecular biology*, 13, 787–792.

[87] Jinek, M. & Doudna, J. A. (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature*, 457, 405–412.

[88] Johnson, M. L. & Faunt, L. M. (1992). Parameter estimation by least-squares methods. *Methods in enzymology*, 210, 1–37.

[89] Kacser, H. & Burns, J. A. (1973). The control of flux. *Symposia of the Society for Experimental Biology*, 27, 65–104.

[90] Kati, W. M., Johnson, K. a., Jerva, L. F., & Anderson, K. S. (1992). Mechanism and fidelity of HIV reverse transcriptase. *The Journal of biological chemistry*, 267(36), 25988–97.

[91] Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J.-K., Staskawicz, B. J., & Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), 18002–7.

[92] Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4 (pp. 1942–1948 vol.4).

[93] Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics*, 39(10), 1278–1284.

[94] Khanin, R. & Higham, D. J. (2007). A minimal mathematical model of post-transcriptional gene regulation by by microRNAs*. *University of Strathclyde Mathematics Research Report 33*, 33, 1–9.

[95] Kiehl, T. R., Mattheyses, R. M., & Simmons, M. K. (2004). Hybrid simulation of cellular behavior. *Bioinformatics (Oxford, England)*, 20, 316–322.

[96] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science (New York, N.Y.)*, 220, 671–680.

[97] Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., & Herwig, R. (2009). *Systems Biology: A Textbook*. Wiley.

[98] Kretschmer-Kazemi Far, R., Leppert, J., Frank, K., & Sczakiel, G. (2005). Technical improvements in the computational target search for antisense oligonucleotides. *Oligonucleotides*, 15(3), 223–33.

[99] Kretschmer-Kazemi Far, R., Nedbal, W., & Sczakiel, G. (2001). Concepts to automate the theoretical design of effective antisense oligonucleotides. *Bioinformatics (Oxford, England)*, 17(11), 1058–61.

[100] Kretschmer-Kazemi Far, R. & Sczakiel, G. (2003). The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Research*, 31(15), 4417–4424.

[101] Krueger, U., Bergauer, T., Kaufmann, B., Wolter, I., Pilk, S., Heider-Fabian, M., Kirch, S., Artz-Oppitz, C., Isselhorst, M., & Konrad, J. (2007). Insights into effective RNAi gained from large-scale siRNA validation screening. *Oligonucleotides*, 17, 237–250.

[102] Kumar, P., Goh, G., Wongphayak, S., Moreau, D., & Bard, F. (2013). ScreenSifter: analysis and visualization of RNAi screening data. *BMC bioinformatics*, 14, 290.

[103] Lammermann, K., Vogel, H., & Traut, W. (2016). The mitochondrial genome of the Mediterranean flour moth, Ephestia kuehniella (Lepidoptera: Pyralidae), and identification of invading mitochondrial sequences (numts) in the W chromosome. *European Journal of Entomology*, 113(1), 482–488.

[104] Larsson, E., Sander, C., & Marks, D. (2010). mRNA turnover rate limits siRNA and microRNA efficacy. *Molecular systems biology*, 6(433), 433.

[105] Laufer, S., Detzer, A., Sczakiel, G., & Restle, T. (2010). . In V. A. Erdmann & J. Barciszewski (Eds.), *RNA Technologies and Their Applications* (pp. 289–58). Springer.

[106] Laufer, S. U. o. L. (2008). *Quantitative Untersuchung der Aufnahme und des biologischen Effekts von oligomeren Nukleinsaurewirkstoffen nach Peptid-vermitteltem Transport in Saugerzellen*. Phd thesis, University of Lübeck.

[107] Law, A. M. & Kelton, W. D. (2014). *Simulation, Modeling and Analysis*, volume 5. New York: McGraw-Hill Science.

[108] Le, L. Q., Lorenz, Y., Scheurer, S., Fötisch, K., Enrique, E., Bartra, J., Biemelt, S., Vieths, S., & Sonnewald, U. (2006). Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (Lyc e 3) expression. *Plant biotechnology journal*, 4(2), 231–42.

[109] Le, S. Y., Malim, M. H., Cullen, B. R., & Maizel, J. V. (1990). A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses. *Nucleic acids research*, 18(6), 1613–23.

[110] Lee, E. J., Baek, M., Gusev, Y., Brackett, D. J., Nuovo, G. J., & Schmittgen, T. D. (2008). Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA (New York, N.Y.)*, 14(1), 35–42.

[111] Lee, H.-C., Li, L., Gu, W., Xue, Z., Crosthwaite, S. K., Pertsemlidis, A., Lewis, Z. A., Freitag, M., Selker, E. U., Mello, C. C., & Liu, Y. (2010). Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Molecular cell*, 38, 803–814.

[112] Levine, E., Ben Jacob, E., & Levine, H. (2007). Target-specific and global effectors in gene regulation by MicroRNA. *Biophysical journal*, 93(11), L52–4.

[113] Lewis, K. (2010). Persister cells. *Annual review of microbiology*, 64, 357–72.

[114] Li, Y., Lu, J., Han, Y., Fan, X., & Ding, S.-W. (2013). RNA Interference Functions as an Antiviral Immunity Mechanism in Mammals. *Science (New York, N.Y.)*, 342(6155), 231–234.

[115] Lide, D. R. & Baysinger, G. (2005). CRC Handbook of Chemistry and Physics. *Journal of the American Chemical Society*, 127(12), 4542–4542.

[116] Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., & Bartel, D. P. (2003). The microRNAs of Caenorhabditis elegans. *Genes & development*, 17(8), 991–1008.

[117] Lopreato, J. & von Bertalanffy, L. (1970). General System Theory: Foundations, Development, Applications.

[118] Losick, R. & Desplan, C. (2008). Stochasticity and cell fate. *Science (New York, N.Y.)*, 320(5872), 65–8.

[119] Lotka, A. J. (1910). Contribution to the Theory of Periodic Reaction. *The Journal of Physical Chemistry A*, 14(3), 271–274.

[120] Lotka, A. J. (1920). Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proceedings of the National Academy of Sciences of the United States of America*, 6(7), 410–415.

[121] Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., & Cui, Q. (2008). An analysis of human microRNA and disease associations. *PloS one*, 3, e3420.

[122] Lu, R., Maduro, M., Li, F., Li, H. W., Broitman-Maduro, G., Li, W. X., & Ding, S. W. (2005). Animal virus replication and RNAi-mediated antiviral silencing in Caenorhabditis elegans. *Nature*, 436(7053), 1040–3.

[123] Lucy, A. P., Guo, H. S., Li, W. X., & Ding, S. W. (2000). Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. *The EMBO journal*, 19(7), 1672–80.

[124] Lund, E. & Dahlberg, J. E. (2006). Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harbor symposia on quantitative biology*, 71, 59–66.

[125] Luo, J., Emanuele, M. J., Li, D., Creighton, C. J., Schlabach, M. R., Westbrook, T. F., Wong, K.-K., & Elledge, S. J. (2009). A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell*, 137, 835–848.

[126] Lynch, P. (2008). The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, 227(7), 3431–3444.

[127] Macfarlane, L.-A. & Murphy, P. R. (2010). MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics*, 11(7), 537–61.

[128] MacRae, I. J. & Doudna, J. A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current opinion in structural biology*, 17, 138–145.

[129] MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V., & Doudna, J. a. (2008). In vitro reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2), 512–7.

[130] Macy, M. W. & Willer, R. (2002). : Computational Sociology and Agent-Based Modeling.

[131] Maillard, P. V., Ciaudo, C., Marchais, A., Li, Y., Jay, F., Ding, S. W., & Voinnet, O. (2013). Antiviral RNA Interference in Mammalian Cells. *Science (New York, N.Y.)*, 342(6155), 235–238.

[132] Mallory, A. C. & Vaucheret, H. (2009). ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. *EMBO reports*, 10(5), 521–6.

[133] Marshall, W. F. (2008). Modeling recursive RNA interference. *PLoS computational biology*, 4(9), e1000183.

[134] Martick, M. & Scott, W. G. (2006). Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126(2), 309–20.

[135] Martinez, J. & Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes and Development*, 18(9), 975–980.

[136] Marx, V. (2013). Biology: The big challenges of big data. *Nature*, 498(7453), 255–260.

[137] Matsumoto, M. & Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator.

[138] McAdam, P. (1998). Handbook of computational economics.

[139] Meissner, M., Schmuker, M., & Schneider, G. (2006). Optimized Particle Swarm Optimization (OPSO) and its application to artificial neural network training. *BMC bioinformatics*, 7, 125.

[140] Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, 15, 185–197.

[141] Michalewicz, Z. (1996). *Genetic algorithms + data structures = evolution programs*, volume 1. Springer Verlag, 3 edition.

[142] Mitchell, M. (1998). *An Introduction to Genetic Algorithms (Complex Adaptive Systems)*. The MIT Press.

[143] Monod, J. (1971). *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. New York: Knopf, Alfred A., 1 edition.

[144] Moran, U., Phillips, R., & Milo, R. (2010). SnapShot: key numbers in biology. *Cell*, 141(7), 1262–1262.e1.

[145] Mosner, J., Mummenbrauer, T., Bauer, C., Sczakiel, G., Grosse, F., & Deppert, W. (1995). Negative feedback regulation of wild-type p53 biosynthesis. *The EMBO journal*, 14(18), 4442–9.

[146] Moyed, H. S. & Bertrand, K. P. (1983). hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *Journal of bacteriology*, 155(2), 768–75.

[147] Mukherjee, K., Campos, H., & Kolaczkowski, B. (2013). Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Molecular biology and evolution*, 30(3), 627–41.

[148] Munroe, S. H. (2004). Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns. *Journal of cellular biochemistry*, 93(4), 664–71.

[149] Murphy, D., Dancis, B., & Brown, J. R. (2008). The evolution of core proteins involved in microRNA biogenesis. *BMC evolutionary biology*, 8, 92.

[150] Murray, J. D. (2002). *Mathematical Biology I. An introduction*, volume 17. Interdisciplinary Applied Mathematics.

[151] Nakayashiki, H. (2005). RNA silencing in fungi: mechanisms and applications. *FEBS letters*, 579, 5950–5957.

[152] Napoli, C., Lemieux, C., & Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell*, 2, 279–289.

[153] Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7, 308–313.

[154] Noble, D. (1960). Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations. *Nature*, 188, 495–497.

[155] Noland, C. L., Ma, E., & Doudna, J. A. (2011). siRNA repositioning for guide strand selection by human Dicer complexes. *Molecular cell*, 43, 110–121.

[156] Obbard, D. J., Jiggins, F. M., Halligan, D. L., & Little, T. J. (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Current biology : CB*, 16(6), 580–5.

[157] O'Malley, M. A., Calvert, J., & Dupré, J. (2007). The study of socioethical issues in systems biology. *The American journal of bioethics : AJOB*, 7(4), 67–78.

[158] Overhoff, M., Alken, M., Far, R. K.-K., Lemaitre, M., Lebleu, B., Sczakiel, G., & Robbins, I. (2005). Local RNA Target Structure Influences siRNA Efficacy: A Systematic Global Analysis. *Journal of Molecular Biology*, 348(4), 871–881.

[159] Pahle, J. (2002). *EINE HYBRIDMETHODE ZUR SIMULATION BIOCHEMISCHER PROZESSE*. Diploma thesis, Universitaet Karlsruhe and European Media Lab (EML) Heidelberg.

[160] Palauqui, J. C., Elmayan, T., Pollien, J. M., & Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *The EMBO journal*, 16(15), 4738–45.

[161] Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., & Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends in biotechnology*, 22(8), 400–5.

[162] Pereira, T. C. & Lopes-Cendes, I. (2013). Medical applications of RNA interference (RNAi). *BMC Proceedings*, 7(2), K21.

[163] Petzold, L. (1983). Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1), 136–148.

[164] Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J. C., Montero, F., & Schuster, S. (1999). METATOOL: for studying metabolic networks. *Bioinformatics (Oxford, England)*, 15, 251–257.

[165] Pfingsten, J. S. & Kieft, J. S. (2008). RNA structure-based ribosome recruitment: lessons from the Dicistroviridae intergenic region IRESes. *RNA (New York, N.Y.)*, 14(7), 1255–63.

[166] Pillai, R. S., Artus, C. G., & Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA (New York, N.Y.)*, 10(10), 1518–25.

[167] Pillai, R. S., Reuter, M., Voigt, F., Schulz, E. C., Kasaruho, A., & Barabas, O. (2012). Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc.

[168] Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., & Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, 465(7301), 1033–1038.

[169] Premsrirut, P. K., Dow, L. E., Kim, S. Y., Camiolo, M., Malone, C. D., Miething, C., Scuoppo, C., Zuber, J., Dickins, R. A., Kogan, S. C., Shroyer, K. R., Sordella, R., Hannon, G. J., & Lowe, S. W. (2011). A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell*, 145, 145–158.

[170] Puchałka, J. & Kierzek, A. M. (2004). Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophysical journal*, 86, 1357–1372.

[171] Puck, T. T., Marcus, P. I., & Cieciura, S. J. (1956). Clonal growth of mammalian cells in vitro; growth characteristics of colonies from single HeLa cells with and without a feeder layer. *The Journal of experimental medicine*, 103(2), 273–83.

[172] Raab, R. M. & Stephanopoulos, G. (2004). Dynamics of gene silencing by RNA interference. *Biotechnology and bioengineering*, 88(1), 121–32.

[173] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS biology*, 4(10), e309.

[174] Raj, A. & van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2), 216–26.

[175] Rao, C. V. & Arkin, A. P. (2003). Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *The Journal of Chemical Physics*, 118, 4999–5010.

[176] Rashbass, J. (1996). Molecular biology and the Internet. *Clin Mol Pathol.*, 49(3), M136–M139.

[177] Ratcliff, F., Harrison, B. D., & Baulcombe, D. C. (1997). A similarity between viral defense and gene silencing in plants. *Science (New York, N.Y.)*, 276, 1558–1560.

[178] Reynolds, A., Anderson, E. M., Vermeulen, A., Fedorov, Y., Robinson, K., Leake, D., Karpilow, J., Marshall, W. S., & Khvorova, A. (2006). Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA (New York, N.Y.)*, 12(6), 988–993.

[179] Rivas, F. V., Tolia, N. H., Song, J.-J., Aragon, J. P., Liu, J., Hannon, G. J., & Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature structural & molecular biology*, 12(4), 340–9.

[180] Romano, N. & Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular microbiology*, 6, 3343–3353.

[181] Romero-López, C. & Berzal-Herranz, A. (2009). A long-range RNA-RNA interaction between the 5' and 3' ends of the HCV genome. *RNA (New York, N.Y.)*, 15(9), 1740–52.

[182] Rosen, R. (1968). Systems Theory and Biology. Proceedings of the 3rd Systems Symposium, Cleveland, Ohio, Oct. 1966. M. D. Mesarovicacute, Ed. Springer-Verlag, New York, 1968. xii + 403 pp., illus. $16.

[183] Saini, H. K., Griffiths-Jones, S., & Enright, A. J. (2007). Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(45), 17719–24.

[184] Salis, H. & Kaznessis, Y. (2005). Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *The Journal of chemical physics*, 122, 54103.

[185] Salis, H., Sotiropoulos, V., & Kaznessis, Y. N. (2006). Multiscale Hy3S: hybrid stochastic simulation for supercomputers. *BMC bioinformatics*, 7, 93.

[186] Salter, M., Knowles, R. G., & Pogson, C. I. (1994). Metabolic control. *Essays in biochemistry*, 28, 1–12.

[187] Sasaki, T., Shiohama, A., Minoshima, S., & Shimizu, N. (2003). Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics*, 82(3), 323–30.

[188] Saumet, A. & Lecellier, C.-H. (2006). Anti-viral RNA silencing: do we look like plants? *Retrovirology*, 3, 3.

[189] Schirle, N. T. & MacRae, I. J. (2012). The Crystal Structure of Human Argonaute2. *Science*.

[190] Schuster, S., Fell, D. A., & Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature biotechnology*, 18(3), 326–32.

[191] Sciabola, S., Cao, Q., Orozco, M., Faustino, I., & Stanton, R. V. (2013). Improved nucleic acid descriptors for siRNA efficacy prediction. *Nucleic acids research*, 41(3), 1383–94.

[192] Secrier, M. & Schneider, R. (2013). Visualizing time-related data in biology, a review. *Briefings in Bioinformatics*, 15(5), 771–782.

[193] Seitz, H. (2009). Redefining MicroRNA Targets. *Current Biology*, 19(10), 870–873.

[194] Sen, G. L. & Blau, H. M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature cell biology*, 7(6), 633–6.

[195] Shabalina, S. A. & Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. *Trends in ecology & evolution (Personal edition)*, 23, 578–587.

[196] Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M. A., Wong, K.-K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., & Settleman, J. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1), 69–80.

[197] Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., & Margalit, H. (2007). Regulation of gene expression by small non-coding RNAs: a quantitative view. *Molecular systems biology*, 3(138), 138.

[198] Silva, J. M., Marran, K., Parker, J. S., Silva, J., Golding, M., Schlabach, M. R., Elledge, S. J., Hannon, G. J., & Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science (New York, N.Y.)*, 319, 617–620.

[199] Siomi, H. & Siomi, M. C. (2009). On the road to reading the RNA-interference code. *Nature*, 457(7228), 396–404.

[200] Siritunga, D. & Sayre, R. T. (2003). Generation of cyanogen-free transgenic cassava. *Planta*, 217(3), 367–73.

[201] Smith, J. M., Van Ness, H. C., & Abbott, M. M. (2005). *Introduction to Chemical Engineering Thermodynamics*, volume 27. Chemical Engineering.

[202] Sornette, D. & Werner, M. J. (2009). *Encyclopedia of Complexity and Systems Science*. Springer.

[203] Soto-Rifo, R., Limousin, T., Rubilar, P. S., Ricci, E. P., Décimo, D., Moncorgé, O., Trabaud, M.-A., André, P., Cimarelli, A., & Ohlmann, T. (2012). Different effects of the TAR structure on HIV-1 and HIV-2 genomic RNA translation. *Nucleic acids research*, 40(6), 2653–67.

[204] Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., & Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912), 190–3.

[205] Stram, Y. & Kuzntzova, L. (2006). Inhibition of viruses by RNA interference. *Virus genes*, 32(3), 299–306.

[206] Stucki, J. W. (1978). Stability analysis of biochemical systems–a practical guide. *Progress in biophysics and molecular biology*, 33(2), 99–187.

[207] Sunilkumar, G., Campbell, L. M., Puckhaber, L., Stipanovic, R. D., & Rathore, K. S. (2006). Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proceedings of the National Academy of Sciences of the United States of America*, 103(48), 18054–9.

[208] Svoboda, P., Stein, P., & Schultz, R. M. (2001). RNAi in mouse oocytes and preimplantation embryos: effectiveness of hairpin dsRNA. *Biochemical and biophysical research communications*, 287(5), 1099–1104.

[209] Tafer, H. (2014). Bioinformatics of siRNA design. *Methods in molecular biology (Clifton, N.J.)*, 1097, 477–90.

[210] Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., & Sharp, P. A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. *Genes and Development*, 13, 3191–3197.

[211] Vallabhajosyula, R. R., Chickarmane, V., & Sauro, H. M. (2006). Conservation analysis of large biochemical networks. *Bioinformatics (Oxford, England)*, 22, 346–353.

[212] Velculescu, V. E., Madden, S. L., Zhang, L., Lash, A. E., Yu, J., Rago, C., Lal, A., Wang, C. J., Beaudry, G. A., Ciriello, K. M., Cook, B. P., Dufault, M. R., Ferguson, A. T., Gao, Y., He, T. C., Hermeking, H., Hiraldo, S. K., Hwang, P. M., Lopez, M. A., Luderer, H. F., Mathews, B., Petroziello, J. M., Polyak, K., Zawel, L., & Kinzler, K. W. (1999). Analysis of human transcriptomes. *Nature genetics*, 23(4), 387–8.

[213] Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W. S., Karpilow, J., & Khvorova, A. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA (New York, N.Y.)*, 11(5), 674–82.

[214] Vohradsky, J., Panek, J., & Vomastek, T. (2010). Numerical modelling of microRNA-mediated mRNA decay identifies novel mechanism of microRNA controlled mRNA downregulation. *Nucleic acids research*, 38(14), 4579–85.

[215] Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends in Genetics*, 17(8), 449–459.

[216] Volterra, V. (1926). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma*, 2(2), 31–113.

[217] Wang, D., Zhang, Z., O'Loughlin, E., Lee, T., Houel, S., O'Carroll, D., Tarakhovsky, A., Ahn, N. G., & Yi, R. (2012). Quantitative functions of Argonaute proteins in mammalian development. *Genes & Development*, 26(7), 693–704.

[218] Wang, X.-H., Aliyari, R., Li, W.-X., Li, H.-W., Kim, K., Carthew, R., Atkinson, P., & Ding, S.-W. (2006). RNA interference directs innate immunity against viruses in adult Drosophila. *Science (New York, N.Y.)*, 312(5772), 452–4.

[219] Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Swanstrom, R., Burch, C. L., & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, 460(7256), 711–6.

[220] Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., & Zamore, P. D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*, 151(5), 1055–67.

[221] Whitehead, K. A., Dahlman, J. E., Langer, R. S., & Anderson, D. G. (2011). Silencing or Stimulation? siRNA Delivery and the Immune System.

[222] Wiberg, D. (1983). Dynamic system identification: Experiment design and data analysis.

[223] Wilkins, C., Dishongh, R., Moore, S. C., Whitt, M. A., Chow, M., & Machaca, K. (2005). RNA interference is an antiviral defence mechanism in Caenorhabditis elegans. *Nature*, 436(7053), 1044–7.

[224] Wilsbacher, L. D., Yamazaki, S., Herzog, E. D., Song, E.-J., Radcliffe, L. A., Abe, M., Block, G., Spitznagel, E., Menaker, M., & Takahashi, J. S. (2002). Photic and circadian expression of luciferase in mPeriod1-luc transgenic mice invivo. *Proceedings of the National Academy of Sciences of the United States of America*, 99(1), 489–94.

[225] Wilson, R. C. & Doudna, J. A. (2013). Molecular mechanisms of RNA interference. *Annual review of biophysics*, 42, 217–39.

[226] Wünsche, W. & Sczakiel, G. (2005). The activity of siRNA in mammalian cells is related to the kinetics of siRNA-target recognition in vitro: mechanistic implications. *Journal of molecular biology*, 345(2), 203–9.

[227] Xue, Y., Abou Tayoun, A. N., Abo, K. M., Pipas, J. M., Gordon, S. R., Gardner, T. B., Barth, R. J., Suriawinata, A. A., & Tsongalis, G. J. (2013). MicroRNAs as diagnostic markers for pancreatic ductal adenocarcinoma and its precursor, pancreatic intraepithelial neoplasm. *Cancer genetics*, 206(6), 217–21.

[228] Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., & Darnell, J. E. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome research*, 13(8), 1863–72.

[229] Yuan, Y.-R., Pei, Y., Ma, J.-B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.-Y., Dauter, Z., Tuschl, T., & Patel, D. J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Molecular cell*, 19(3), 405–19.

[230] Zadeh, A. H. & Foster, G. D. (2004). Transgenic resistance to tobacco ringspot virus. *Acta virologica*, 48(3), 145–52.

[231] Zambon, R. A., Vakharia, V. N., & Wu, L. P. (2006). RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster. *Cellular microbiology*, 8(5), 880–9.

[232] Zamore, P. D., Tuschl, T., Sharp, P. A., & Bartel, D. P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101(1), 25–33.

[233] Zeffman, A., Hassard, S., Varani, G., & Lever, A. (2000). The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *Journal of molecular biology*, 297(4), 877–93.

[234] Zenklusen, D., Larson, D. R., & Singer, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology*, 15(12), 1263–71.

[235] Zhang, B., Pan, X., Cobb, G. P., & Anderson, T. A. (2007). microRNAs as oncogenes and tumor suppressors. *Developmental biology*, 302(1), 1–12.

[236] Zhang, J., Lau, M. W., & Ferré-D'Amaré, A. R. (2010). Ribozymes and riboswitches: modulation of RNA function by small molecules. *Biochemistry*, 49(43), 9123–31.

[237] Zhang, L. (1997). Gene Expression Profiles in Normal and Cancer Cells. *Science*, 276(5316), 1268–1272.

[238] Zietz, K. (2004). *Theoretische Analyse der Struktur-Funktionsbeziehung von Antisense Oligonukleotiden und siRNA*. Bachelor thesis, Universität zu Lübeck.

[239] Zimmermann, T. S., Lee, A. C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M. N., Harborth, J., Heyes, J. A., Jeffs, L. B., John, M., Judge, A. D., Lam, K., McClintock, K., Nechev, L. V., Palmer, L. R., Racie, T., Röhl, I., Seiffert, S., Shanmugam, S., Sood, V., Soutschek, J., Toudjarska, I., Wheat, A. J., Yaworski, E., Zedalis, W., Koteliansky, V., Manoharan, M., Vornlocher, H. P., & MacLachlan, I. (2006). RNAi-mediated gene silencing in non-human primates. *Nature*, 441(1), 111–114.

[240] Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science (New York, N.Y.)*, 244(4900), 48–52.

[241] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415.