Revealing the Mechanism of Thiopeptide Antibiotics at Atomistic Resolution

Implications for Rational Drug Design

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Antje Wolf
aus
Plauen

Bonn 2012

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. rer. nat. Martin Hofmann-Apitius

2. Gutachter: Prof. Dr. rer. nat. Michael Wiese

Tag der Promotion: 18.06.2012

Erscheinungsjahr: 2012

für meine Familie; meinen Ehemann Stefan und meine Tochter Marie

Abstract

For decades drug design has primarily focused on small molecules that bind to well-formed tight binding pockets, such as the catalytic centers of enzymes. Recently, there is increasing interest to design compounds that disrupt or stabilize biomacromolecular interfaces (e.g. protein–protein, protein–DNA, protein–RNA, protein–lipid interfaces). These non-traditional drug targets hold great therapeutic potential as they govern cellular pathways. In contrast to traditional drug targets, where computational methods are now routinely and productively used to complement experiments, the use of computer-based approaches for the study and design of interfacial modulators is still in its infancy.

The current thesis is a first detailed study into understanding the effects of modulators of a protein–RNA interface and developing computer-based approaches for their design. This work focuses on the 23S·L11 subunit of the ribosomal GTPase-associated region (GAR), a prototypic protein–RNA interface of high relevance in the development of novel antibacterials. The GAR is the target of naturally occuring thiopeptide antibiotics. These unique molecules are effective inhibitors of bacterial protein synthesis, but are currently unused in human antibacterial therapy due to their low aqueous solubility. Their mechanism of action is explored in the current thesis, enabling the design and proposition of new chemical scaffolds targeting their binding site.

The specific challenges associated with the 23S·L11·thiopeptide system, such as the inherent flexibility of the protein–RNA composite environment and the size and structural complexity of the thiopeptide ligands, are addressed by a combination of computational chemistry approaches at different levels of granularity and a steady feedback with experimental data to validate and improve the computational techniques. These approaches range from quantum mechanics for deriving optimized intramolecular parameters and partial atomic charges for the thiopeptide compounds, to molecular dynamics simulations accounting for the binding site's flexibility, to molecular docking studies for predicting the binding modes of different thiopeptides and derivatives. All-atom molecular dynamics simulations were conducted, providing a detailed understanding of the effect of thiopeptide binding at a previously unmet resolution. The findings of this work, coupled with previous experimental knowledge, strongly support the hypothesis that restricting the binding site's conformational flexibility is an important component of the thiopeptide antibiotics' mode of action. With the help of an MD-docking-MD workflow and an energy decomposition analysis crucial residues of the binding site and pharmacologically relevant moieties within the ligand structures could be identified. A

4D-pharmacophore model is presented that was derived from a refined 23S·L11·thiopeptide complex and additionally accounts for the dynamic stability of molecular interactions formed between the antibiotic and the ribosomal binding site as the fourth dimension.

The results of this thesis revealed, for the first time, a plausable description of the thiopeptide antibiotics' mode of action, down to the details of their pharmacologically relevant parts and provide a computational framework for the design of new ligands.

Contents

List of Abbreviations

All units used throughout this thesis are in accordance with the SI system. Abbrevations for amino acids and nucleotides follow the IUPAC conventions.

Å	Angström, 1 Å = 0.1 nm
aa-tRNA	aminoacyl-tRNA
ADT	AutoDockTools
CD	circular dichroism
cryo-EM	cryo-electron microscopy
CSD	Cambridge Structural Database
CTD	C-terminal domain
DCCM	dynamic cross-correlation map
DHA	dehydroalanine
DNA	deoxyribonucleic acid
EAR	exchange acceptance ratio
EF-4	elongation factor 4
EF-G	elongation factor G
EF-Tu	elongation factor thermo unstable
FF	force field
GA	genetic algorithm
GAR	GTPase-associated region
GDP	guanosine diphosphate
GTP	guanosine triphosphate
H-bond	hydrogen bond
H43/H44	helices 43 and 44
K_D	dissoziation constant
LGA	Lamarckian genetic algorithm
MC	micrococcin
MD	molecular dynamics
Me	methyl
MM	molecular mechanics

MM-GBSA ..... molecular mechanics generalized Born surface area
MM-PBSA ..... molecular mechanics Poisson-Boltzmann surface area
MPI-DO ..... Max Planck Institute for Molecular Physiology, Dortmund
MRSA ..... methicillin-resistent *Staphylococcus aureus*
NMR ..... nuclear magnetic resonance
NS ..... nosiheptide
NTD ..... N-terminal domain
PCA ..... principal component analysis
PDB ..... Protein Data Bank
QM ..... quantum mechanics
REMD ..... replica-exchange molecular dynamics
RMSD ..... root-mean-square deviation
RMSF ..... root-mean-square fluctuation
RNA ..... ribonucleic acid
RRMS ..... relative root-mean-square
rRNA ..... ribosomal RNA
SAR ..... structure-activity relationship
SASA ..... solvent-accessible surface area
SB ..... stalk base
sd ..... standard deviation
SF ..... scoring function
tRNA ..... transfer RNA
TS ..... thiostrepton
vdW ..... van der Waals

1. Introduction

1.1. The Antibacterial Challenge

Today, after more than 50 years of success, too few new antibiotics are produced to replace the ones that have become ineffective due to resistance [1]. To overcome resistance, new antibiotics have to be introduced that are effective against the resistant mutants. In the majority of cases, these "new" antibiotics are derivatives of a small set of well-known scaffolds [2]. However, the modifications which can be made to an old scaffold are finite, and most of the low-hanging fruits have been already picked [3]. Since the "Golden Age" of antibiotics discovery in the early 1960s, only three new classes of antibiotics entered the market (see Fig. 1.1): the oxazolidinones (linezolid), the lipopeptides (daptomycin) and the mutilins (retapamulin) [2]. Scientific challenges associated with the discovery of novel antibiotics, and a poor return on monetary investment, have led to the exit of many pharmaceutical companies from the field [3–6].

Resistance times become shorter. Typically, resistance occurs within just four years of approval of new drugs [7]; in the case of the recent antibiotic linezolid it occured already in its clinical trials [8]. The inappropriate and irrational use of antimicrobial medicines provides favourable conditions for resistant microorganisms to emerge, and spread, for example, when patients do not take the full course of a prescribed antimicrobial or when poor quality antimicrobials are used [9]. This high selective pressure bacteria are exposed to has led to an explosion of multidrug resistance, in both the community and hospital settings. One of the most dangerous strains, Methicillin-resistant *Staphylococcus aureus* (MRSA), is estimated to cause more than 100,000 infections each year [10] and ~19,000 deaths per year in the United States only (status 2007) [11]. Multidrug-resistant and extensively drug-resistant *Mycobacterium tuberculosis* (MDR-TB and XDR-TB) strains are another health threat, especially in the developing world [12]. According to recent statistics of the World Health Organization (WHO), one third of the world's total population is infected, of which approximately 1.7 million people die from tuberculosis each year [13]. As resistance is on the rise, the discovery of novel antibiotics is urgent.

Antibiotic candidates that act at new targets or via distinct mechanisms/biological pathways have the greatest potential to overcome resistance [14]. A rich source for new antibiotics are naturally occuring antimicrobials that have evolved over many years among competing microorganisms [8]. There are few marketed synthetic antibacterial agents not descended from natural products [15, 16].

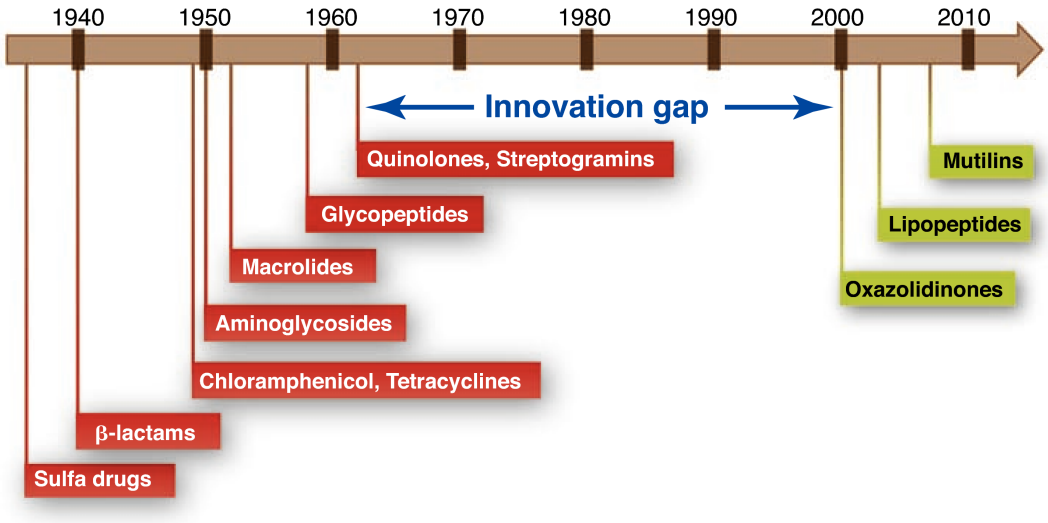


Figure 1.1.: Innovation gap of new antibiotics introductions. Taken from Ref. 2.

Containing 'priviledged structural scaffolds', they provide evolutionary tailored pharmacophores. One promising class of natural products are the thiopeptide antibiotics (see Section 1.4), that possess a remarkable activity against many severe bacterial pathogens but are currently unused in human antibacterial therapy due to their low aqueous solubility [17]. Their mechanism of action is exploited in the current thesis. In the following, we will take a closer look at one of the major targets for naturally occuring antibiotics, including the thiopeptides [18]: the bacterial ribosome.

1.2. The Bacterial Ribosome as a Major Target for Antibiotics

Bacterial ribosomes are supramolecular machines which consist of almost two thirds ribosomal RNA (rRNA) and 55 different proteins. They are composed of a large and small subunit (Fig. 1.2), called 50S and 30S, distinguished by their sedimentation constants. Each step during the bacterial protein synthesis – initiation, elongation and termination – is targeted by different classes of antibiotics, with the majority of them inhibiting a step associated with the elongation phase [19].

During the elongation phase, transfer RNAs (tRNAs) are moved through three different tRNA binding sites (A, P, E) in a cyclic fashion (see Fig. 1.2). First, a complex of aminoacyl-tRNA (aa-tRNA), elongation factor Tu (EF-Tu) and GTP is delivered to the aminoacyl (A) site. GTP on EF-Tu is hydrolyzed to GDP + P_i. This causes EF-Tu to undergo a large conformational change and leads to its dissociation from the complex, allowing for aa-tRNA accommodation. The ribosome is now in the PRE-state. A peptide-bond is formed between the amino acid attached to the tRNA at the A site and the one attached to the already present tRNA at the peptidyl (P) site. The peptide chain is thereby transferred to the tRNA at the A site, leaving the tRNA at the P site uncharged. The translocation reaction, shifting the peptidyl-tRNA from the A to the P site and the deacylated tRNA from the P to

the exit (E) site[1], is catalyzed by a second elongation factor, EF-G. Interestingly, EF-G is structurally very similar to the tRNA of the EF-Tu·GTP·tRNA complex [21]. Binding EF-G·GTP and the hydrolysis of GTP to GDP leads to a ratchet-like rotation of the 30S subunit relative to the 50S subunit [22, 23]. The ribosome is now in the POST-state, while the EF-G·GDP and deacylated tRNA dissociate and a new charged tRNA can be accommodated at the A site – thus completing the cycle.

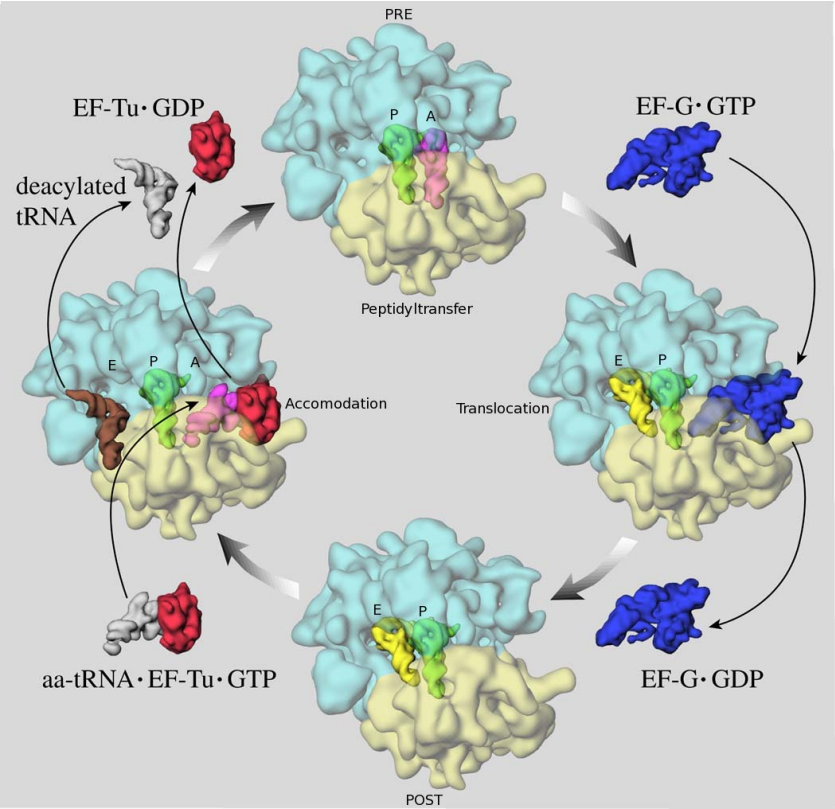


Figure 1.2.: Elongation cycle on the bacterial ribosome. Color coding: 50S subunit cyan, 30S subunit pale yellow, elongation factors EF-Tu red, EF-G blue, tRNAs are colored according to their state: free or complexed with EF-Tu gray, at A site magenta, at P site green, at E site yellow, exiting brown. Picture modified after Ref. 24.

Antibiotics, such as edeines, or pactamycin, disrupt the beginning of protein biosynthesis by inhibiting the formation of the initiation complex [19, 25]. Several other antibacterial classes disrupt the elongation cycle at different stages by preventing tRNA accomodation (tetracyclines), peptide bond formation (oxazolidinones), or elongation of the nascent polypeptide chain (macrolides). The thiopeptide antibiotics thiostrepton and micrococcin (see Section 1.4) disturb the elongation process by interferring with elongation factors, such as EF-G or EF-Tu.

Antibiotics are not only important for treating bacterial infections but also contributed a lot to today's understanding of ribosomal function. Structures of the ribosome or ribosomal subunits with

[1] This is a oversimplified description. Actually, there exist hybrid A/P and P/E states [20].

and without bound antibiotics – as determined by X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM) – provided a deep understanding of its structure and function. This pioneering work was recently honored with the 2009 Nobel prize in Chemistry, awarded to Venky Ramakrishnan, Tom Steitz und Ada Yonath.

As a result of these efforts, three functionally important sites at the ribosome could be identified: 1) The decoding site at the 30S subunit, deciphering the anti-codon of the incoming aminoacyl-tRNA, 2) the peptidyltransferase center at the 50S subunit, the active center of the ribosome, catalyzing the peptide bond formation, and 3) the GTPase-associated region (GAR), which binds translation factors and stimulates their GTPase activity. The latter is the target site of thiopeptide antibiotics and subject of the current thesis.

1.3. The GTPase-associated Region as an Unused Target in Human Antibacterial Therapy

The GTPase-associated region (GAR) on the 70S ribosome plays a central role in peptide elongation by providing a docking site for elongation factors and by coordinating GTP hydrolysis during protein synthesis [26–28]. Structurally, the GAR consists of three components: 1) the sarcin-ricin loop located in helix 95 of the 23S rRNA, 2) the ribosomal stalk, composed of the ribosomal proteins L10 and L7/L12, and 3) the stalk base, comprising L11 and the helices H43/H44 on the 23S rRNA [29]. Unlike other ribosomal active sites, the GAR is equally rich in rRNA and proteins. The ribosomal protein L11 is a highly conserved two-domain protein, whose C-terminal domain (L11-CTD) binds the compactly folded RNA tightly, whereas its N-terminal domain (L11-NTD) makes only limited contacts with RNA [30]. The detection of different orientations for L11-NTD strongly suggests that the dynamical behavior of L11-NTD is important for its function during the ribosomal cycle by allowing interaction with elongation and initiation factors [28, 31].

The thiopeptide antibiotics (see Section 1.4), such as thiostrepton (TS), nosiheptide (NS) and micrococcin (MC), bind to the SB where they make contacts with the 23S rRNA and the N-terminal domain of L11 [32]. Figure 1.3 depicts thiostrepton bound to the L11·23S rRNA subunit and its location on the 50S ribosome.

Based on their X-ray studies of the 50S subunit [32] combined with a cryo-EM reconstruction of the 70S·EF-G·GDPNP[2] complex [33], Harms et al. proposed the inhibitory mechanisms of thiostrepton and micrococcin as schematically depicted in Figure 1.4: Binding of EF-G to the stalk base results in a widening of the cleft formed by the 23S rRNA and the N-terminal domain of L11 (SB “in” state, Fig. 1.4B), allowing EF-G’s domain V to protrude into the cleft. Secondly, L11-NTD interacts with the C-terminal domain of L7 and positions it ideally to contact the G’-subdomain of EF-G, leading to P_i release (Fig. 1.4D). Binding of TS to the SB interfers with both “switches” (Fig. 1.4E) by restricting

[2] GDPNP is a non-hydrolyzable GTP analog that was used to stabilize the state for visualization.

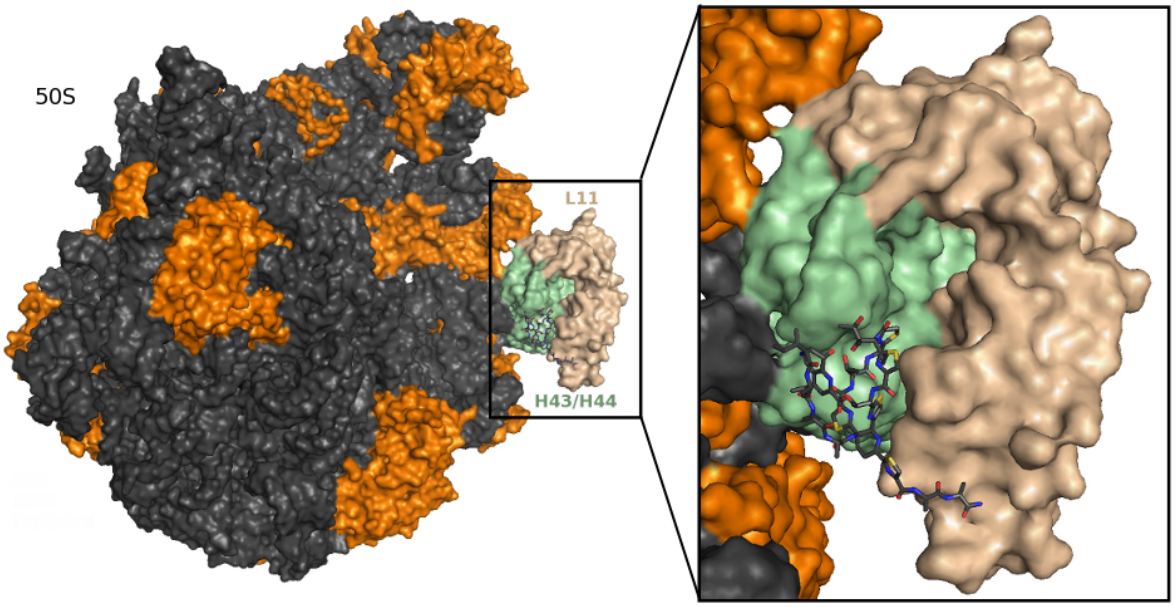


Figure 1.3.: Depiction of the 50S subunit with bound thiostrepton (sticks representation), based on PDB entry 3CF5 [32]. RNA is colored in gray, proteins in orange. The stalk base (SB), consisting of L11 and H43/H44 23S rRNA is highlighted in wheat and pale green, respectively.

the conformational flexibility of L11-NTD that is necessary for widening the cleft and subsequent binding of EF-G. Harms et al. conclude further that TS binding additionally precludes interaction with L7-CTD and thus reduces P_i release. The same is true for nosiheptide. In contrast, micrococcin seems to stabilize the L11–L7 interaction, leading to a constant stimulation of P_i release (Fig. 1.4F). It should be kept in mind that the schema proposed in Figure 1.4 is a hypothesis that fits well to experimental observations. An exact understanding of the dynamics of this mechanism at atomic resolution, however, is still missing.

The thiopeptides, which are unused as antibacterial agents in human therapy due to their low aqueous solubility, are the only class of antibiotics targeting the GTPase-associated region. This makes the GAR an attractive target for drug design as cross-reactions with existing antibiotics are not expected. Additionally, eukaryotic ribosomes possess natural resistance to thiostrepton, because of their different residue composition at their active site [29].

1.4. Thiopeptide Antibiotics

1.4.1. Structure and Biological Activity

The thiopeptide antibiotics are a large class of macrocyclic, sulfur-containing peptides, produced as secondary metabolites by Gram-positive bacteria, which largely belong to the genus *Streptomyces* [17]. The more than 80 members share common structural characteristics: they all possess

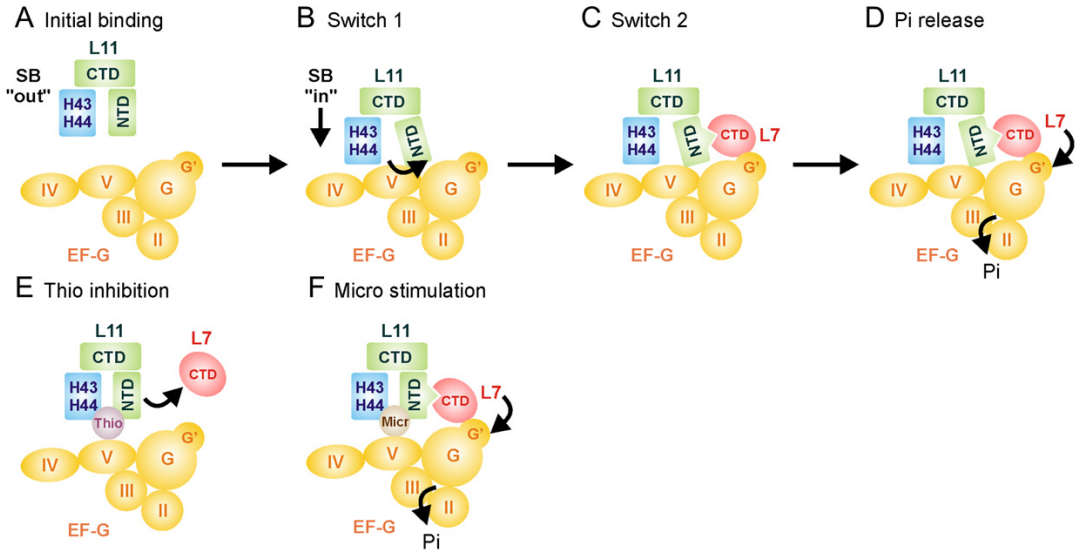


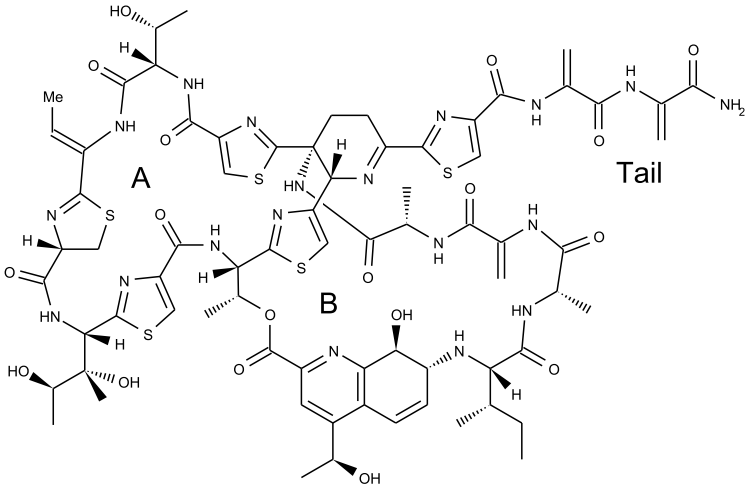
Figure 1.4.: Schematic for EF-G action during elongation, its inhibition through thiostrepton and its stimulation through micrococcin. Details are described in the text. Picture taken from Ref. 32.

a central nitrogen heterocycle that connects one to two macrocycles and an acyclic tail. The best studied thiopeptide is *thiostrepton* (Fig. 1.5 **1**) which was discovered more than 50 years ago [34, 35]. It consists of a thiazole-rich macrocycle A, a second macrocycle B containing a quinaldic acid and a dehydroalanine-tail. The lesser known nosiheptide (Fig. 1.5 **2**) [36–38] also consists of two macrocycles, but ring B is much smaller than in TS and is not connected to the N-heterocycle. Although micrococcin (Fig. 1.5 **3**) was the first thiopeptide discovered [39], its absolute stereochemical assignment was described only recently [40, 41]. It possesses only one macrocycle which is structurally similar to TS's macrocycle A.

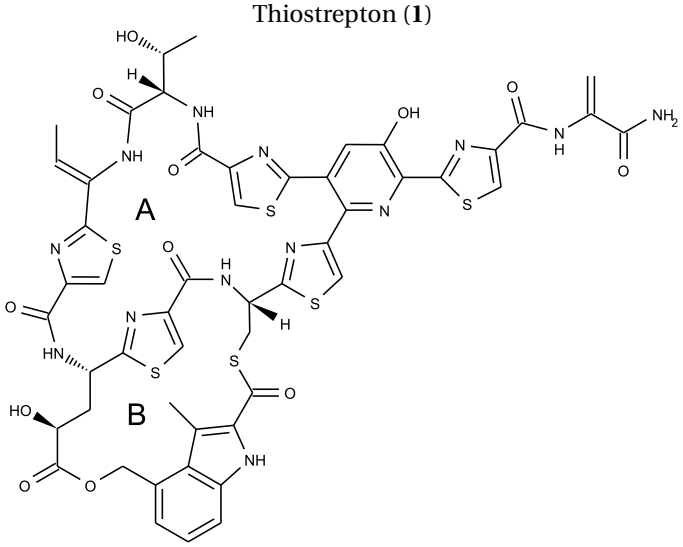
Interesting is their high activity (pico- to nanomolar range) against clinically relevant Gram-positive bacteria, for example MRSA and *Mycobacterium tuberculosis* which are resistant to most convential antibiotic treatments [17]. Because of their low aqueous solubility and poor bioavailability, thiopeptide antibiotics are only used in animal therapy as a topical food ingredient and not prescribed to humans.

1.4.2. Drug Design Attempts

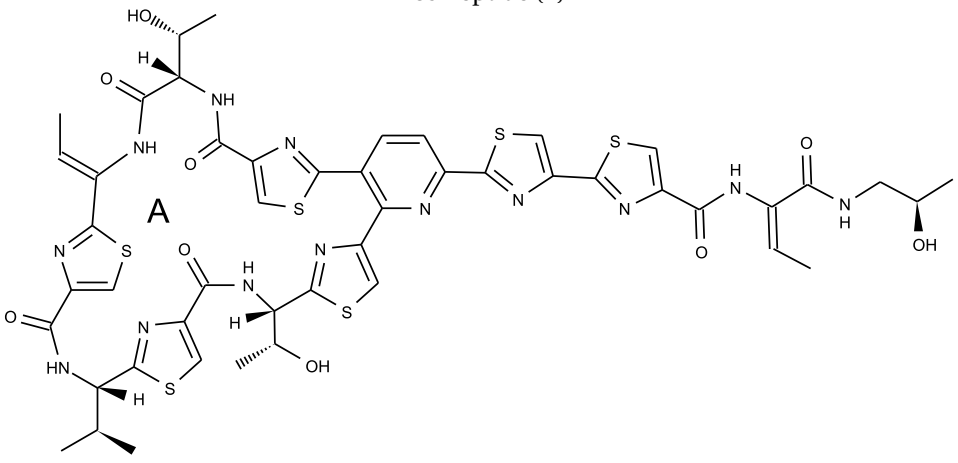
Although the thiopeptides are known for several decades now, the discovery of a biologically active TS fragment [42], and a complete description of its total synthesis [43] combined with the urgent pressure to find new antibiotics due to increasing resistance, has renewed interest in these compounds. Previous attempts to design new ligands that bind to the ribosomal stalk base have had moderate success so far. Bower et al. [44] identified TS fragments that bind selectively to the L11 binding domain of 23S rRNA but these compounds failed to show any antibacterial activity *in vivo*. This study, however,



Thiostrepton (**1**)



Nosiheptide (**2**)



Micrococcin (**3**)

Figure 1.5.: Chemical structures of important thiopeptide antibiotics.

only focussed on RNA binding, neglecting the interactions that thiostrepton makes to L11. Although TS can bind isolated 23S rRNA [45], the cooperative nature of the binding has been found in several studies [45–47], underlining the importance of taking both species at the binding site into account. A different route was recently pursued by Starosta et al. [48], who identified active thiopeptide precursor lead compounds, although with much reduced potency, using different high-throughput translation machinery assays.

Recent crystal structures of thiostrepton and nosiheptide bound to the GAR [32] now allow the precise location of these two compounds at the 23S rRNA–L11 interface, thus making the system accessible to structure-based ligand design. The work described in the current thesis represents the first computational structure-based drug design approach making use of these crystal structure data to complement experimental studies.

1.5. Modern Drug Discovery is 'Rationalized'

Computational methods are well established in modern drug discovery as a cost- and time-effective alternative or complement to experimental procedures [49, 50]. This kind of activities is usually summarized under the term *Computer-Aided Drug Design (CADD)*. The methods used range from bioinformatics sequence analysis tools in the target identification and validation process to chemoinformatics procedures in the thorough selection of screening candidates. Besides that, an ever increasing amount of data is stored in databases all over the world, providing valuable sources of information for scientists. The drug discovery process is a complex, cost-extensive and time-consuming procedure that usually includes the steps: target identification and validation - lead discovery and optimization - medical safety and evaluation - drug development - clinical trials - drug approval. For bringing a new drug to the market, timelines of 7–12 years and a cost of approximately 1.2 billion dollar are often cited [51]. *In silico* methods are primarily used in the early stages of the drug discovery process to significantly reduce the time and resource requirements for chemical synthesis and biological testing. An extended workflow diagram, showing the application of computational methods at different steps during the drug discovery pipeline, is depicted in Figure 1.6. The fast expansion of computational methods in this area has been made possible by advances in software and hardware computational power, sophistication in the identification of molecular targets, and an increasing database of publicly available target protein structures [51].

In the process of lead discovery, large databases of chemical compounds are screened. For virtual screenings two types of methods are used: ligand-based and structure-based methods [52]. Ligand-based methods take properties of known binders into account to find similar molecules. The underlying assumption is that structural similarity of compounds implies similar biological activity. Several ligand-based virtual screening techniques have been reviewed elsewhere [53, 54]. Structure-based methods make use of the 3D structure of the macromolecular target molecule. A wide range of different docking algorithms and scoring functions is used to predict the binding

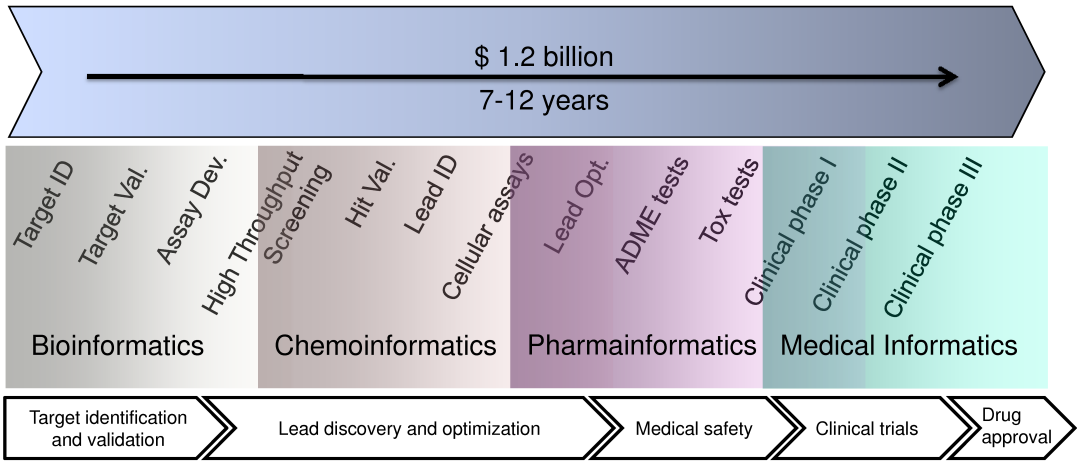


Figure 1.6.: Computational methods in the drug discovery pipeline.

geometry and affinity of a ligand with respect to a given target. An overview of different approaches and algorithms is given in the next section.

Especially structure-based methods are appealing because of their ability to identify new chemotypes and to give mechanistic insight into the binding mode between a compound and a target structure. The success of the method is evident from the drugs currently reaching the market, and it is clear that in many companies structure-guided approaches have become central to develop promising drug candidates [55]. An extensive overview of successful rational drug design examples is given in Chapter 5 of Klebe's book "Wirkstoffdesign" [56].

1.5.1. Targeting Macromolecular Complexes

Traditionally, drug targets are key proteins in a metabolic or signalling pathway associated with a certain disease or pathogen. Predominantly, these proteins are enzymes or receptors with well-defined binding pockets to which drugs bind in a "lock-and-key" manner, competing with an endogenous small molecule [57].

A completely different drug target paradigm was recently introduced by Pommier and Cherfils as the "interfacial inhibitor concept" [58]. The common mechanistic principle underlying interfacial inhibitors is characterized by a highly specific, reversible binding to interfaces between macromolecules (protein–protein, protein–nucleic acid or protein–lipid interfaces), often found in molecular machines [59]. Molecular machines, such as the ribosome, have a large number of fast-moving parts that are often synchronized with other molecular components – making them particularly vulnerable to small molecules [59]. The bound ligands transiently arrest the targeted molecular machines, and thus interfere with their function. As many identified interfacial inhibitors are natural products (among them thiostrepton), it was proposed that they are "one of nature's paradigms for generating inhibitors and toxins" [58]. In the meantime, also synthetic interfacial inhibitors have been designed,

for example inhibitors of HIV-1 integrase–DNA [60–62] and topoisomerase–DNA [63, 64] complexes, demonstrating that interfacial inhibitor design is a valid drug design concept.

It has been hypothesized [58] that interfacial inhibitors are very specific since they target both partners of a complex, and only in the context of their dynamic cellular pathways, like a molecular finger-print. This could reduce the risk of side-effects. Difficulties in molecular modelling are the huge size of the complexes (at least ternary complexes of up to three molecular species have to be modelled) and the dynamic nature of the biological interface [59]. Above all, taking into account the flexibility and plasticity of the interface is mandatory to successfully discover interfacial modulators [65].

1.5.2. RNA as Drug Target

Recently, there has been growing interest in RNA as a drug target due to the increased number of high-resolution RNA structures [66] and the rising awareness that RNA has more complex roles than just being a passive intermediate storage device for genetic information [67, 68]. Especially in the development of novel anti-infectives RNA plays a major role [69]. The wealth of high resolution crystal structures of complexes between antibiotics and their ribosomal binding sites provide the basis for structure-based approaches to be used in the design of appropriate modifications to existing antibiotics as well as in the discovery of completely new drug classes [70, 71]. Recent structure-based drug design successes at the bacterial ribosome include the oxazolidone class of compounds and the aminoglycosides [72]. A broad overview of structure-based design studies targeting the ribosome was recently given by Sherer [73].

Due to its specific physicochemical properties, however, further advances in understanding the chemistry and structure of RNA recognition are needed in order to design promising new drug candidates [74, 75]. Addressing protein–RNA interfaces by computational means poses, thus, even more challenges than protein–protein interfaces.

1.6. Aims and Challenges of this Thesis

This thesis represents a first step towards developing an understanding and computational strategies to investigate interactions of small molecules with “loose” biomacromolecular interfaces. These interfaces are of particular therapeutic interest as their functions help govern cellular biology. However, these non-traditional drug targets pose several challenges:

1. They usually have large, shallow, and solvent-exposed binding sites that possess inherent flexibility.

2. Their ligands often violate the classical concept of drug-likeness, established by Lipinsky more than 10 years ago [76]. For example the molecular weight of the thiopeptides is twice as much as that of drug-like compounds that are contained in classical virtual screening compound

libraries. Moreover, they contain macrocycles, making their use in docking even more difficult as these are not treated flexible during the docking process.

3. A third challenge is specifically associated with protein–RNA interfaces (in contrast to protein–protein interfaces): Classical scoring functions of docking programs are trained on protein–ligand complexes and it is unclear if they can be used in a straightforward way to score protein·RNA–ligand interactions.

This work focusses on a functionally important protein–RNA interface, the 23S·L11 subunit of the ribosomal GTPase-associated region. This system has been selected as a prototype due to its relevance in the development of novel antibacterials and the availability of experimental data through a collaroration with the Max Planck Institute for Molecular Physiology. The specific objective of this research project is to provide a detailed understanding of the influence that ligand binding has on the dynamics of the 23S·L11 interface and to elucidate the structural characteristics of this interaction, enabling the design and proposition of new chemical scaffolds targeting this binding site. The aforementioned difficulties do not allow for traditional application of state-of-the-art structure-based drug development techniques as in the case of "classical" targets (e.g. enzymes) [65]. Therefore, a substantional part of this thesis is to develop and validate new strategies to address these issues. Aspects of this are the integration of molecular dynamics techniques to account for the binding site's inherent flexibility, and to address thiopeptide–protein/RNA interactions by sophisticated scoring and free energy of binding calculation approaches. An MD–docking–MD workflow shall be created that can, in principal, be transferred to other protein–RNA interface target systems. The overall key aspect, however, is the steady feedback with experimental data to validate and improve the computational techniques.

After an introduction into state-of-the-art methods for computer-aided drug design in Chapter 2, the influence of thiostrepton binding on the dynamics of the GTPase-associated region is investigated by molecular dynamics simulations in Chapter 3. It is shown that thiostrepton restricts the conformational flexibility of the target system and that its mechanism of action is likely different from classical competitive inhibitors. The implications of these investigations for future drug design are derived. Chapter 4's focus is on the thiopeptide ligands. Several state-of-the-art docking and scoring functions are evaluated with respect to their performance at RNA·protein composite environments. The previously unknown binding modes of selected thiostrepton derivatives at the GAR are presented and the experimentally observed structure-activity relationships are rationalized. Key molecular motives at the binding site and within the ligand structures are identified. Finally, the gained knowledge is incorporated to develop the first pharmacophoric description of the 23S·L11·thiopeptide complex. In Chapter 5 a 4D-pharmacophore model is presented that additionally accounts for the dynamic stability of molecular interactions formed between TS and the ribosomal binding site as the fourth dimension. In Chapter 6, the results of this thesis are summarized and an outlook for future drug design at biomacromolecular interfaces is presented.

2. State-of-the-Art Methods for Computer-Aided Drug Design

2.1. Molecular Dynamics Simulations

2.1.1. Background and Theory

The following section gives a basic introduction into the theory underlying molecular dynamics simulations; it is mainly comprised from the textbooks by Frenkel and Smit [77] and by Andrew Leach [78].

Molecular dynamics (MD) simulations are used to obtain information about a system's time-dependent properties and/or to study the dynamic behavior of biological macromolecules at different timescales. Their execution is very similar to "real" experiments: at first, a sample has to be prepared, i. e. a model system with a certain number of particles has to be selected. Then, the simulation is run until the system is *equilibrated*, meaning the properties that we are interested in no longer change with time and are converged. Only then, the property of interest can be measured by computing a statistical ensemble average of a certain quantity.

Computational resources limit the complexity achievable in simulations. Therefore, trade-offs have to be made – either in the system's size or in the duration of the simulation. In practice, the time average is easier to obtain and usually the time average is computed instead of the ensemble average. Why can we calculate a time average if we are actually interested in a statistical ensemble average? This is based on one of the most fundamental theorems of statistical mechanics, the *ergodic hypothesis* [79], stating that, in infinity, the statistical ensemble average is equal to the time average of the system. If we simply wait long enough we can observe all possible *microscopic states* (i.e. the atomic positions and momenta) of the system in a certain *macroscopic state*, defined by, for example, its temperature T, pressure P and the number of particles N. An *ensemble* is a collection of all possible configurations which have different microscopic states but have an identical macroscopic or thermodynamic state. Three common types of statistical thermodynamics ensembles are usually distinguished in MD simulations: the microcanonical ensemble which is equivalent to a thermally isolated system (NVE; constant number of particles, volume and energy), the canonical ensemble, a system in contact with a heat bath (NVT, constant number of particles, volume and temperature), and

the isothermal-isobaric ensemble (NPT, constant number of particles, pressure and temperature).

The basic algorithms have hardly changed since the first MD study was done on hard spheres by Alder and Wainwright in the late 1950s [80, 81]. What has changed tremendously is the methodology of computer simulations. The adaptation to the fast evolving computer facilities, e.g. the parallelisation of codes to be able to run simulations on multiple cores and the most recent implementations of CUDA versions that can be run on graphics processing units (GPUs), have made MD simulations accessible to scientists without a supercomputer who use MD as an everyday tool to solve their research questions.

The many different MD programs available today (e.g. AMBER [82], CHARMM [83], Gromacs [84], NAMD [85]) are based on the same physical principle: The motions of a particle are simulated by numerically solving the differential equations embodied in Newton's second equation of motion:

$$M\ddot{r} = F \tag{2.1}$$

where F is the force exerted on a particle, M is the mass of that particle and $\ddot{r}$ is the acceleration of the particle (second derivative of its position r with respect to time). The force F can be calculated as the negative gradient of the potential energy:

$$F = -\nabla V(r) \tag{2.2}$$

Different integration schemes are used to solve Newton's equation of motion, such as the Verlet algorithm or a popular adaption of this algorithm, the Leap-Frog scheme. What is still needed to calculate the forces F of the particles, is the potential $V(r)$ which is described by a *force field*.

2.1.2. Force Fields

The potential $V(r)$ has most often the following functional form:

$$\begin{aligned} V(r) = &\sum_{bonds} \frac{k_b}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_\theta}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_n}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right) \\ &+ \sum_{i,j}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \end{aligned} \tag{2.3}$$

The first three terms are *intramolecular* or bonded terms, meaning that they are calculated for atoms in the same molecule involved in a bond (two atoms), angle (three atoms) or torsion (four atoms). The bond stretching and angle bending are modelled by a harmonic potential where the energy increases as the bond length l (the angle θ) deviates from its reference value l_0 (θ_0). The torsional potential models how the energy changes while rotating about a bond. The last term describes *intermolecular*, or non-bonded interactions, which are calculated between all pairs of

atoms residing in different molecules or between atoms within the same molecule but separated by at least three bonds. The van der Waals interactions are modelled by a 6-12 Lennard-Jones potential, and the electrostatic interactions are calculated using Coulomb's law.

The functional form, the parameters $k_b, k_\theta, V_n, \varepsilon_{ij}, \sigma_{ij}$ and the partial atomic charges (q) make up a *force field*. The parameters are specific for the atom types that are defined within the force field. An atom type is determined by the atomic number of the atom, its hybridisation state and its local environment. For example, most force fields differentiate sp^3-hybridised, sp^2-hybridised, and sp-hybridised carbon.

Force-Field Parametrization

Obtaining force-field parameters is a complicated and often time-consuming task. If available, experimental data can be used for the parametrisation, like it was done in the OPLS force field [86]. Jorgensen and Tirado-Reeves used computer simulations to derive non-bonded parameters that reproduced specific thermodynamic properties. Another way is to use data derived from quantum mechanics and validate them using experimental data, as done in deriving the GLYCAM06 force field [87]. A typical modern force-field parametrization workflow might be envisioned in the following way: First, a molecular fragment that represents the parameters of interest is chosen, often together with its local chemical environment. Several conformations of this fragment are produced and their potential energy calculated by quantum mechanics. Then, parameters are gradually refined in a 'trial and error' manner until they provide the best fit to the potential energy curve. Often, this process is heavily influenced by the experience, or chemical intuition, of the computational chemist. As it is almost impossible to simultaneously modify all parameters at once, the parameterisation is done in stages, for example first non-bonded parameters, then bonds, then angles, and finally the torsional parameters. It is usually necessary to modify the parameters in an iterative procedure because some degrees of freedom are coupled. Clearly, a force field obtained this way should be properly validated using experimental data. A key attribute of a generic force field is transferability, meaning that parameters should be transferable between different molecules and can be applied to a wider range of problems (e.g. different states and properties) than they were originally derived from.

Calculation of Partial Atomic Charges

Once the force field's functional form has been chosen and the parameters have been derived, one part of the force field is still missing: the partial atomic charges. Force fields either use a fixed charge for each atom type (e.g. OPLS) or assign charges dependent on the atom's local electrostatic environment. As the electrostatic properties of a molecule are the consequence of the distribution of the electrons and the nuclei, it is straightforward to use quantum mechanics to calculate partial atomic charges. Unfortunately, partial atomic charges are not an experimentally observable quantity. That is why there is no naturally 'best' method to calculate it but a variety of different approaches.

Table 2.1.: Examples of all-atom force fields and their applicability domain.

Name	Applicability domain	Reference
ff99	Proteins, nucleic acids	Wang et al. (2000) [88]
ff99SB	Improved backbone torsion parameters of ff99	Hornak et al. (2006) [89]
parmbsc0	Changes to ff99 for nucleic acids	Pérez et al. (2007) [90]
GLYCAM06	Carbohydrates	Kirschner et. al.(2008) [87]
GAFF	"General AMBER force field" for small organic molecules	Wang et al. (2004) [91]
OPLS	Protein, DNA, organics	Jorgensen et al. (1988) [86]
MMFF94	Organics	Halgren (1996) [92]

Fitting atomic charges to the molecular electrostatic potential (MEP) has proven to reliably reproduce condensed-phase properties, and to handle inter-molecular properties well [93]. The MEP at a given point p in the vicinity of a molecule is the force acting on a positive test charge (a proton) located at p through the electrical charge cloud generated through the molecules electrons and nuclei (see Fig. 2.1).

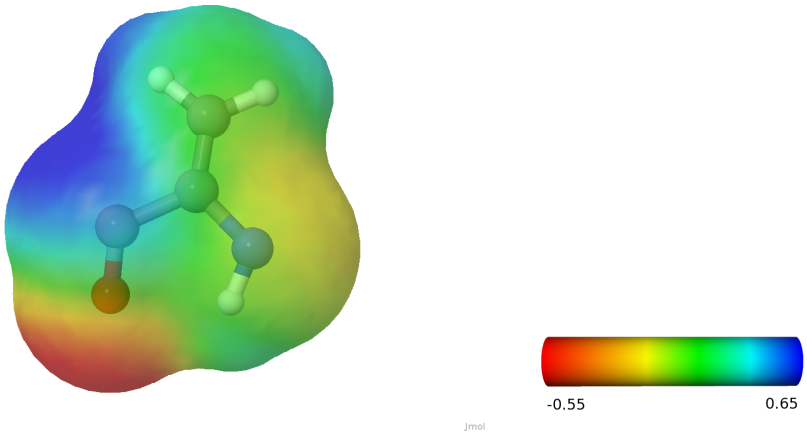


Figure 2.1.: Molecular electrostatic potential (MEP) mapped on the surface of the dehydroalanine (DHA) residue. Colored from red: negative to blue: positive.

The objective is to derive the set of partial charges that best reproduces the quantum mechanical electrostatic potential at a series of points surrounding the molecule. Different algorithms have been developed for this purpose. The CHELP method by Chirlian and Francl [94] uses spherical shells, centered on each atom with points symmetrically distributed on the surface, and a Lagrange multiplier method to determine the atomic charges. In the CHELPG algorithm of Breneman and Wiberg [95] the regular grid approach, introduced in 1981 by Cox and Williams [96], is combined with the Lagrange multiplier method by Chirlian and Francl. The algorithm of Singh and Kollman [97]

uses points on a series of molecular surfaces, constructed using gradually increasing van der Waals radii for the atoms. A modified version, the restrained electrostatic potential fit (RESP) [98], uses hyperbolic restraints on non-hydrogen atoms to hold down the electrostatic potential derived (ESP) charges with a minor impact on the fit. The RESP algorithm was used to generate partial charges for the 1994 AMBER force field [99].

A rapid method for calculating partial atomic charges is the Gasteiger-Marsili method [100] which uses information about the atoms in a molecule and their connections. In an iterative procedure less and less charge is transferred between bonded atoms at each step. The great advantage of this method is that it is much faster than the quantum mechanical approaches discussed before and is independent on the conformation or molecular orientation of the molecule. Therefore, it is preferentially used for docking applications, where charges have to be assigned to a large number of different molecules.

2.1.3. Modelling Solvent

In their natural environment proteins are surrounded by water. In general, there are two different ways to model solvent effects in an MD simulation: explicitly or implicitly. Explicit modeling means the solute (i.e. the protein or other molecule(s)) is solvated in a box (or octahedron) of water molecules. Modelling explicit water molecules allows specific interactions between the solute and solvent to be formed but also enormously increases the number of atoms of the simulation. To limit this number to a manageable value, usually periodic boundary conditions are applied, where the box is mirrored in each direction (see Fig. 2.2). This allows molecules to leave the box. If they leave the box at one side they “enter” it again at the opposite site. When determing the central box dimensions, it is important to choose a size preventing the solute molecules to “see” each other to prevent artifacts arising from their intermolecular interactions.

Despite its small size, water is one of the most challenging systems to model accurately. Consequently, many different water models have been proposed [101]. These models can be divided into three types. In the simple interaction-site water models a rigid water geometry and either three (TIP3P [102], SPC [103], SPC/E [104]), four (BF [105], TIP4P [102]), five (ST2 [106], TIP5P [107]) or six [108] interaction sites are used. Flexible models allow internal changes in conformation of the water molecule. An example is the model developed by Ferguson [109]. Finally, models have been proposed that include polarization effects, solution and interfacial properties of aqueous systems are particularly sensitive to [110].

If explicit modeling of solvent is too costly, or simply not necessary, a common technique is to represent it implicitly by an approximation of the mean-force potential for the solvation interactions. The solvent is modeled as a bulk with a specific dielectric constant. The electrostatic contribution of solvation is calculated by continuum methods like solving of the Poisson-Boltzmann (PB) equation [111, 112] or the simpler Generalized Born (GB) model [113]. The non-polar part of solvation is

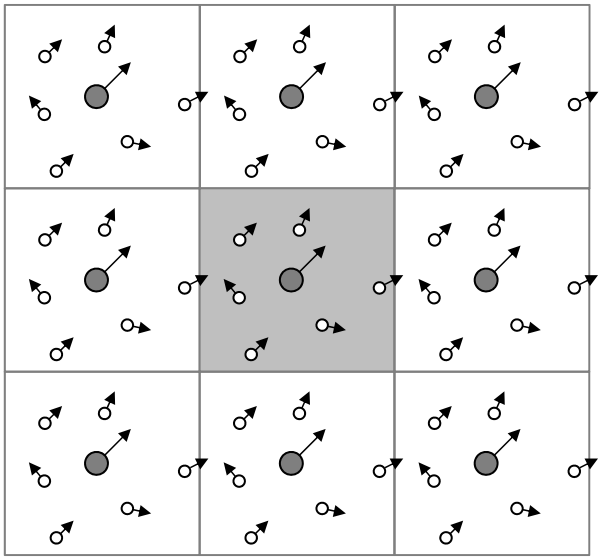


Figure 2.2.: Two-dimensional representation of periodic boundary conditions. The solute (big, dark-gray circle) and solvent (small, white circles) molecules in the central original cell (gray background) are replicated together with their velocities (arrows) in all dimensions.

most often approximated by an empirical term proportional to the solvent-accessible surface area (SASA). These combined approaches are then referred to as PBSA and GBSA, respectively.

2.1.4. Replica Exchange Molecular Dynamics

In a *replica exchange molecular dynamics (REMD)* simulation [114] multiple non-interacting copies (*replicas*) are simulated simultaneously. Each of the replicas samples a different thermodynamic state (cmp. Section 2.1.1), usually the temperature. In defined time intervals attempts are made to swap (neighboring) configurations (see Fig. 2.3). The transition probablities $W(X, X')$ between configuration X and X' are implemented by the Metropolis criterion to ensure detailed balance [115]:

$$W(X, X') = \min(1, \exp^{-\Delta}) \tag{2.4}$$

where $\Delta = (1/k_B T_j - 1/k_B T_i)(V_i - V_j)$. k_B is the Boltzmann constant, V_i and T_i are the potential energy and temperature of replica i. If the exchange is allowed between a pair of replicas, the target temperatures for the two replicas are swapped (i.e. system i changes to temperature T_j and system j changes to temperature T_i) and the velocities of each replica involved are adjusted by a scaling factor related to the previous and new target temperatures [116].

Metaphorically speaking, the high temperature simulations represent a configurational “card shuffling” to overcome high-energy barriers, the low temperature simulations serve to find local energy minima. The particular advantage of the replica-exchange method is that a replica trapped in a local energy minimum can escape via exchange to a different value of the independent variable

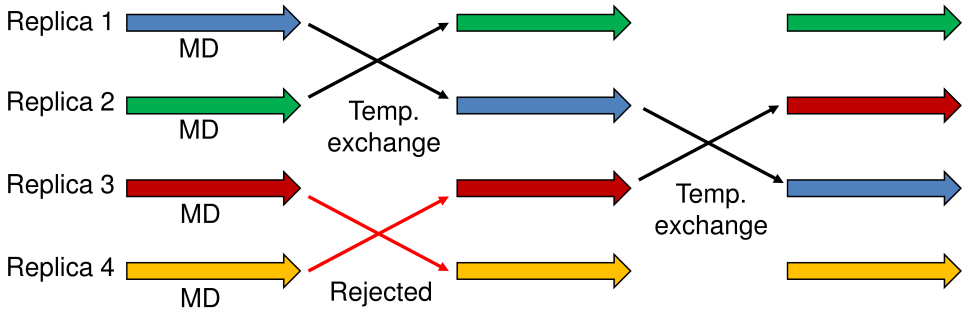


Figure 2.3.: Schematic for for the replica exchange molecular dynamics method (REMD). MD calculations are performed in parallel for individual replicas at different temperatures, which are exchanged during the calculations. Adapted from Ref. 117.

(temperature) [114]. Thus, a larger configurational space can be sampled.

2.1.5. Analysing Molecular Dynamics Trajectories

The output of a molecular dynamics run is a *trajectory*, typically containing a subset of the coordinates the investigated system passed through. It is up to the investigator how to analyse this trajectory to answer his or her research questions. There exist, however, several generic analysis techniques that can be applied, most of them are not exclusive to MD trajectories. The ones used throughout this thesis will be briefly explained in the following.

Root-Mean-Square Deviation

The *Root-Mean-Square Deviation (RMSD)* gives the average distance d between two position vectors p and q of N equivalent atoms in 3D space:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_i d_i^2} = \sqrt{\frac{1}{N}\|p_i - q_i\|^2} \tag{2.5}$$

It is a standard measure of structural distance between coordinate sets.

Root-Mean-Square Fluctuation

The *Root-Mean-Square Fluctuation (RMSF)* describes the atomic positional fluctuations within a considered time period. It is calculated by averaging over atom (or residue) i's deviations to its time-averaged position:

$$\mathrm{RMSF}(i) = \sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle} = \sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2} \tag{2.6}$$

The angle brackets $\langle\rangle$ denote a time average. This measure is similar to the RMSD, except that the averaging is over time instead of atoms. To remove system-wide translational and rotational

movements, the coordinates should be aligned on a common structure previously.

Radius of Gyration

The *radius of gyration* R_{gyr} is the distribution of the cross section about a central axis. In simple terms, it describes how much the molecule spreads out from its center. It is defined as the root-mean-square distance of the collection N of atoms i from their common center of gravity r_c:

$$R_{\text{gyr}} = \sqrt{\frac{1}{N}\sum_i (r_i - r_c)^2} \tag{2.7}$$

Dynamic Cross-Correlations

In some cases two residues move in concert with each other, either in the same direction (correlated) or in opposite directions (anti-correlated). The extent of the correlation (positive or negative) can be quantified by calculating the covariance between the fluctuations of two residues (equation 2.8). Elements of the cross-correlation matrix (C_{ij}) are calculated by their position vector r where i and j may correspond to any two atoms, residues or domains. The value of C_{ij} can vary from -1 (completely anti-correlated) to $+1$ (completely correlated).

$$C_{ij} = \frac{\langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle}{\sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle \langle (r_j - \langle r_j \rangle)^2 \rangle}} \tag{2.8}$$

The cross-correlation matrix can be visualized in form of a *Dynamic Cross-Correlation Map (DCCM)*, a symmetric contour plot (see for example Fig. 3.23 in Chapter 3). Typical characteristics of DCCMs include a line of strong cross-correlation along the diagonal, cross-correlations emanating from the diagonal, and off-diagonal cross-correlations. The high diagonal values occur where i = j, where C_{ij} is always equal to 1.00. In the case of proteins, positive correlations emanating from the diagonal indicate correlations between contiguous residues, typically within a secondary structure element or other tightly packed unit of structure. Typical secondary structure patterns include a triangular pattern for helices and a plume for strands. Off-diagonal positive and negative correlations may indicate potentially interesting correlations between domains of non-contiguous residues [118]. This information conveys aspects originating uniquely in the dynamical motions of the molecule and is not visual by inspecting the average structure [119].

Ransom-Wright and McCammon were among the firsts that used this type of cross-correlation maps from MD simulations to probe the spatial extent of the collective motions in cytochrome c [120].

Principal Component Analysis

Principal component analysis (PCA) is a statistical technique used for finding patterns in data of high dimension. It is often used as a tool in exploratory data analysis to reveal the internal data structure in a way that best explains its variance. Mathematically, this is achieved by diagonalization of the data covariance matrix C.

$$C = T\Lambda T^T \tag{2.9}$$

This results in the diagonal matrix Λ containing the eigenvalues λ as diagonal entries and the matrix T containing the corresponding eigenvectors. If the eigenvectors are sorted such that their eigenvalues are in decreasing order, the eigenvector with the largest eigenvalue (i.e. the first principal component) accounts for the highest proportion of variance within the data, the second component is orthogonal to the first one and accounts for the second highest proportion of variance, and so on. The last principal components can usually be neglected without loss of much information. Therefore, PCA is often used for data dimension reduction.

It is, however, important to note that PCA assumes linearity, which means PCA is limited to re-expressing the data as a linear combination of its basis vectors. Another assumption is a high signal-to-noise ratio, meaning that the largest variances contain the most important information.

When performed on a set of experimental structures or snapshots from a MD trajectory, the eigenvectors describe concerted atomic displacements and can highlight major conformational changes between the structures. Because it was shown that these motions are often essential for protein function [121] dynamics in this low-dimensional subspace – spanned by the first few eigenvectors – was termed “essential dynamics”[122].

Clustering

Clustering is an unsupervised learning/classification technique that groups similar objects into subgroups, called *cluster*. It is an optimization problem of minimizing intra-cluster and maximizing inter-cluster differences. Therefore, most clustering algorithms require a measure of similarity, or “distance”, of objects. When clustering is used on MD trajectory data, the RMSD between the MD snapshots can provide such a distance measure. Another possibility is to perform a PCA prior to clustering as this implicitly provides a native distance function for clustering: the Euclidean distance of the points in the PC (sub)space. In Section 3.2.2, the effects of different clustering algorithms on varying principle component subspaces and their influence on cluster quality are investigated.

There exists a series of different algorithms which can be divided into *partitional* and *hierarchical clustering* [123]. While partitional clustering is a division of the objects from the data set into non-overlapping subsets (clusters), hierarchical clustering allows nested clusters and results in a hierarchical tree (*dendrogram*). They are either *agglomerative* (“bottom-up” approach) or *divisive* (“top-down” approach). A partitional clustering can then be obtained by cutting the dendrogram at a

particular level. In the following, the two classes of clustering algorithms will be introduced.

Partitional Clustering Partitional techniques optimize a locally or globally defined criterion function to determine a pre-specified number of clusters, most often the squared error criterion [124]. Of these, the K-means algorithm is one of the oldest, fastest, and most widely used partitional clustering algorithms. The K-means algorithm attempts to find a pre-defined number (K) of clusters represented by their centroids (i.e. cluster centers). The user has to specify a number of desired clusters in the beginning. The algorithm places K initial centroids randomly. Repeatedly, K clusters are formed by assigning each point to its closest centroid. Then, the new centroid of each cluster is recomputed. These steps are repeated until the centroids do not change anymore.

K-means is simple and fast and can be applied to large data sets. The result depends, however, on the initially chosen centroids (i.e. not deterministic). It is also known that K-means fails when the "natural" clusters have non-spherical shapes or widely different sizes or densities, or when the data contains outliers [123]. When applied to MD trajectory data, K-means tends to produce "blocky" clusters of similar sizes [125].

Agglomerative Hierarchical Clustering Agglomerative hierarchical clustering is a collection of closely related clustering techniques that start with singleton clusters and then iteratively join the nearest clusters until all objects are grouped into one all-encompassing cluster. The different styles of this algorithm class differ by their definition of cluster proximity. *Single-linkage* defines cluster proximity as the proximity between the closest two points that are in different clusters, *complete-linkage* uses the farthest two points and *average-linkage* the average pairwise proximities of all pairs of points from different clusters. *Centroid* methods use the distance between the centroids of clusters. For *Ward*'s method, the proximity between two clusters is defined as the increase in the squared error that results when two clusters are merged.

The single-linkage method adopts a "friends of friends" clustering strategy. It is good at handling non-elliptical shapes, but is sensitive to noise and outliers. The complete-linkage method finds similar clusters. It is less susceptible to noise and outliers, but it can break large clusters and it favors globular shapes. The average-linkage and centroid methods can be regarded as aiming for clusters with characteristics somewhere between the single and complete link methods. Ward's method aims at finding compact, spherical clusters. Centroid-linkage and average-linkage were found to be the most useful methods for use with MD trajectories [125]. They can produce clusters of varying sizes and possibly concave shapes.

Agglomerative hierarchical clustering algorithms are deterministic and produce a dendrogram displaying the cluster-subcluster relationships and the order in which the clusters were merged. However, their high computational and storage requirements limit their use to smaller data sets.

Metrics for Cluster Validation Unsupervised methods are difficult to assess as they do not provide a native evaluation criterion. Although important, cluster evaluation is not commonly used in cluster analysis [123]. There exist several different metrics that at least give some general quantitative indication on cluster quality [123, 126, 127]. Each of them has their specific drawbacks and there is no consensus on which method should preferentially be used. In this thesis two metrics are used which have been shown useful in the context of MD simulations before [125].

A simple measure to determine the optimal number of clusters is to calculate the *SSR/SST ratio*, the quotient of the sum of squares regression (SSR or between sum of squares) and the total sum of squares (SST). The SSR is usually calculated via the sum of squares error (SSE or within sum of squares) that is the sum of the squared distances of all points x belonging to cluster C_i to its centroid c_i, summed together for all clusters K. The total sum of squares is the sum of squared distances for all N data points x_j to the overall mean c and is equivalent to the SSE if K is 1.

$$SSE = SST - SSR = \sum_{i=1}^{K} \sum_{x \in C_i} (x - c_i)^2 \tag{2.10}$$

$$SST = \sum_{j=1}^{N} (x_j - c)^2 \tag{2.11}$$

The SSR/SST ratio value lies between 0 and 1 and gives the percentage of explained variance by the data, equivalent to the R^2 value in regression analysis. As the ratio inherently rises with cluster count, one looks for an "elbow" in the curve where adding another cluster does not add much new information, as done in a scree test for principal component analysis [128].

A second metric is the *pseudo F-statistic* (equation 2.12, introduced by Caliński and Harabasz [129]). This metric is a measure for the "tightness" of clusters; high values usually indicate a better clustering.

$$pFS = \frac{SSR/K - 1}{SSE/(N - K)} \tag{2.12}$$

2.2. Molecular Docking

Molecular docking is used to predict the conformation and the binding affinity of a complex formed of two molecules. In the majority of applications a small molecule (ligand) is docked into a binding pocket of a protein (also often referred to as a receptor or target). Because of its pharmaceutical relevance to drug design, this overview is restricted to protein-ligand docking.

The docking challenge is two-fold: the first challenge is the conformational sampling of the ligand (and to some degree of the receptor), the second one is the prediction of the ligand's affinity to its target. The latter is solved by using a scoring function. Since the pioneering work of Kuntz and colleagues in the 1980s [130] more than 60 different docking programs and 30 scoring functions have been developed [131]. The most widely used docking programs for protein-ligand docking

are AutoDock [132] and AutoDock Vina [133], DOCK [134–136], FlexX [137–141], FRED [142, 143], Glide [144–146], GOLD [147], and ICM [148]. An exhaustive overview of current docking and scoring methods can be found in the review of Moitessier and coworkers [131].

The conformational sampling is accomplished by different search methods which can be broadly classified into a) systematic search methods, b) stochastic/random methods and c) simulation methods (see Table 2.2). To address the second challenge associated with molecular docking, the appropriate evaluation of the final complex, three different types of scoring functions have been developed. These are either a) force-field based, b) empirical or c) knowledge-based scoring functions. As each of these types has its short-comings (see Table 2.3), consensus scoring functions like X-Score [149] which combine several single functions have become very popular.

Table 2.2.: Overview of search methods used in docking. A short description and examples of docking tools implementing the methods are given.

Search method	Description	Examples
Systematic search		
Multiconformer docking	Rigid-body docking of multiconformer libraries, often with shape complementarity or interaction site matching algorithms	FRED, DOCK
Incremental construction	Ligand is built up on-the-fly in the active site, ligand is split into fragments, one anchor fragment is docked rigidly first with matching algorithm, others subsequently	FlexX, DOCK, Hammerhead [150]
Stochastic/random search		
Genetic algorithms	Ligand pose = chromosome; translation, rotation, torsion encoded in genes; evolved by genetic operators like mutation and crossover; selection by survival of the fittest	AutoDock, GOLD
Monte Carlo	Ligand pose modified through bond rotation, translation, new conformation with lower score is kept, a higher score is kept or rejected by e.g. Metropolis criterion	ICM, Glide, GlamDock [151]
Tabu search	Search is prevented from revisiting already explored areas of conformational space	PRO_LEADS [152]
Simulation methods		
	Molecular dynamics, simulated annealing, energy minimization (frequently in combination with other search methods)	ICM, DOCK, GlamDock

Over the years, many comparative studies investigating the performance of docking programs, with respect to docking accuracy and enrichment in virtual screening experiments, have been published [158–165]. Although the outcome of these studies varies, it is clear that the docking accuracy highly depends on the protein considered; none of the tools clearly outperforms all others. It is therefore highly advisable to test several programs on the system being investigated.

After an evaluation of different docking programs (see Section 4.2), AutoDock was selected for the docking studies throughout this thesis. Therefore, its search algorithm and scoring function will be described in more detail in the next section. Afterwards, the knowledge-based scoring function

Table 2.3.: Overview of scoring functions used in docking.

	Force-field based	**Empirical**	**Knowledge-based**
Description	Non-bonded terms (van der Waals and electrostatics) of established force fields like AMBER or Tripos FF	Decomposition into simpler, chemically intuitive interaction terms, weighted by scaling factors defined by regression to fit experimentally determined protein-ligand affinities and/or crystal structures	Statistical analysis of protein-ligand crystal structures to determine preferred binding geometries
Strengths	Already available	Many parameters; good predictions within the model; can be extended easily	Independent of experimental measurements; includes also entropic parts; computational simplicity
Weaknesses	Slow; only enthalpic contributions; arbitrarily chosen cut-off distances	Dependent on experimental data and training set	Dependent on data; some interactions (e.g. with metals) underrepresented in available crystal structures
Examples	AutoDock SF (FF/empirical) [153], DockScore, GoldScore, ICM SF	FlexX SF, ChemScore [154], GlideScore, PLP [155]	PMF [156], DrugScore [157]

Abbreviations: FF, force field; SF, scoring function.

DrugScore and its derivatives will be introduced because of their application in this thesis.

2.2.1. AutoDock

Search Algorithm

The AutoDock suite uses stochastic search algorithms to sample the conformational ligand space. The Lamarckian Genetic Algorithm (LGA), used by AutoDock, combines a genetic algorithm with a local search to achieve faster convergence [132]. Genetic algorithms adopt Darwin's theory of evolution [166] to optimization problems in other fields. In docking applications, the genome is made up by the translation, orientation and all torsional angles of the ligand. The genes are altered by genetic operators like crossover and mutation to create a new generation of ligand conformations. Afterwards, the fitness of the population's individuals is evaluated (by a scoring function) and only the fittest individuals are selected. The specialty of the LGA is that in regular intervals a local search is performed in coordinate space (phenotype) to find a local minimum that is used to update the genotype.

Scoring Function

The free energy of binding ΔG is estimated to be equal to the difference between (1) the energy of the ligand L and the protein P in a separated unbound state and (2) the energy of the ligand–protein complex [153]:

$$\Delta G = (V^L_{bound} - V^L_{unbound}) + (V^P_{bound} - V^P_{unbound}) + (V^{P-L}_{bound} - V^{P-L}_{unbound} + \Delta S_{conf}) \tag{2.13}$$

It is assumed that the two molecules are sufficiently distant from one another in the unbound state that $V^{P-L}_{unbound}$ is zero. As the protein is considered rigid, the bound state of the protein is identical to the protein unbound state, and the difference in their intramolecular energy is zero. Thus, Equation 2.13 further reduces to:

$$\Delta G = (V^L_{bound} - V^L_{unbound}) + V^{P-L}_{bound} + \Delta S_{conf} \tag{2.14}$$

ΔS_{conf} is an estimate of the conformational entropy lost upon binding that is proportional to the number of rotatable bonds in the ligand (N_{tors}):

$$\Delta S_{conf} = W_{conf} N_{tors} \tag{2.15}$$

The pair-wise potentials V are evaluated for all intermolecular protein–ligand atom pairs; intramolecular energies are calculated for all pairs of atoms within the ligand, excluding 1–2, 1–3 and 1–4 interactions. AutoDock uses a semi-empirical scoring function based on the non-bonded terms from the 1984 AMBER force field by Weiner et al. [167]. The scoring function includes terms for repulsion/dispersion, hydrogen bonding, electrostatics and desolvation:

$$\begin{aligned} V(r) = & W_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r^{12}_{ij}} - \frac{B_{ij}}{r^{6}_{ij}} \right) \\ & + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r^{12}_{ij}} - \frac{D_{ij}}{r^{10}_{ij}} \right) \\ & + W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon\left(r_{ij}\right) r_{ij}} \\ & + W_{sol} \sum_{i,j} \left(Sol_i Vol_j + Sol_j Vol_i \right) \exp^{\left(-r^2_{ij}/2\sigma^2\right)} \end{aligned} \tag{2.16}$$

The weighting constants W are optimized with respect to experimentally determined free energy values of 188 complexes [153]. Parameters A and B are taken from the 1984 AMBER force field [167]. Hydrogen bonds are accounted for with a special 12-10 potential and a directional term $E(t)$ that is dependent on the angle t between hydrogen-bond donor, hydrogen atom and hydrogen-bond acceptor [168]. The solvation potential is based on the volume (Vol) of the atoms surrounding a

given atom, weighted by an atom-type-dependent solvation parameter (Sol) and an exponential term based on the distance [169]. The distance weighting factor σ is set to 3.5 Å [153].

2.2.2. DrugScore

As already seen in the assembly in Table 2.3 DrugScore belongs to the knowledge-based scoring functions. The knowledge base is a set of experimentally determined protein-ligand complexes from the Protein Data Bank. In short, by a statistical analysis of structural data atom-pair potentials are derived, which are then used to score new complexes.

Theory

DrugScore was developed by Gohlke et al. in 2000 [157], but the formalism that it is based on was established earlier by Sippl [170–172]. The approach makes use of the inverse Boltzmann law. Interatomic contacts that occur more frequently than a mean distribution are considered energetically favorable. Rare contacts are considered energetically unfavorable.

The net statistical preferences ΔW_{ij} for ligand and protein atoms of type[1] i and j are obtained by comparing the mean statistical preferences of the subsystems i, j (W_{ij}) to the reference system (W). This difference is equivalent to the negative logarithm of the observed distribution for atoms i, j (g_{ij}) devided by the expected distribution (g):

$$\Delta W_{ij}(r) = W_{ij}(r) - W(r) = -\ln \frac{g_{ij}(r)}{g(r)} \tag{2.17}$$

$$g(r) = \frac{\sum_i \sum_j g_{ij}(r)}{i * j} \tag{2.18}$$

$g_{ij}(r)$ is the normalized radial pair distribution function for atoms of types i and j, separated by a distance in the interval of r and $r + dr$. The $g(r)$ term is the normalized mean radial distribution for any two atoms, corresponding to the reference state.

Solvent effects are modelled by incorporating the solvent-accessible surface area (SASA) of the protein and the ligand that becomes buried upon complex formation:

$$\Delta W_i(SASA, SASA_0) = W_i(SASA) - W_i(SASA_0) = -\ln \frac{g_i(SASA)}{g_i(SASA_0)} \tag{2.19}$$

g_i is the normalized distribution function of the surface area of an atom i in the buried state ($SASA$) (considering ligand and protein individually) compared to the solvated state ($SASA_0$). ΔW_i reflects the contribution that arises from differences in the solvent-accessible surface between the bound and fully solvated state.

[1] Sybyl atom types are used. For a full list, see Ref. 157.

The total preference ΔW for a particular binding geometry is then approximated by summing all individual contributions of k_i ligand atoms of type i and l_j protein atoms of type j:

$$\Delta W = \gamma \sum_{k_i} \sum_{l_j} \Delta W_{ij}(r) + (1-\gamma) \left[\sum_{k_i} \Delta W_i(SASA, SASA_0) + \sum_{l_j} \Delta W_j(SASA, SASA_0) \right] \tag{2.20}$$

γ is an adjustable parameter, optimized empirically to 0.5.

Compilation of potentials

For all atom types i and j present in the database, the radial pair distributions are calculated by counting the occurences of i and j at a distance between r and $r + dr(N_{ij}(r))$. They are normalized by dividing by the sum over all distances:

$$g_{ij}(r) = \frac{N_{ij}(r)/4\pi r^2}{\sum_r (N_{ij}(r)/4\pi r^2)} \tag{2.21}$$

Scaling to $4\pi r^2$ accounts for the volume of the spherical shell of radius r and thickness dr.

The solvent-accessible surface-dependent singlet potentials[2] are calculated by:

$$g_i(SASA) = \frac{N_i(SASA)}{\sum_{SASA}(N_i(SASA))} \tag{2.22}$$

$g_i(SASA)$ is the probability to find an atom of type i with an exposed solvent-accessible surface $SASA$ in a complexed state, while $g_i(SASA_0)$ is the probability to find the same atom with the same $SASA$ in a state totally separated from the complex.

DrugScore^{CSD}

The variant DrugScore^{CSD} [173] is based on the assumption that the physical nature of non-bonded interactions of protein-ligand complexes and small molecule crystal packings is identical. Therefore, the potentials are compiled from 28 642 entries of the Cambridge Structural Database (CSD) [174] providing relevant contact data in a more balanced distribution of atom types and higher resolution than the Protein Data Bank (PDB) [175]. It could be shown that DrugScore^{CSD} outperforms DrugScore with respect to the native pose identification [173].

DrugScore^{RNA}

In order to score ligand–RNA complexes Pfeffer and Gohlke developed DrugScore^{RNA} in 2007 [176]. It is again based on the same formalism like the other DrugScore functions, but derives the potentials from 670 nucleic acid–ligand and –protein complexes from the PDB. As RNA–ligand complexes alone

[2] in contrast to pair potentials

did not provide statistically significant potentials, DNA–ligand and nucleic acid–protein complexes were also included.

DSX

Very recently, a complete re-implementation of the DrugScore function, *DSX* (DrugScore eXtended), was introduced by Neudert and Klebe [177]. Beside from being parameterized on a larger and newer set of PDB and CSD complexes, the scoring function was extended by a much more specialized set of atom types.

2.3. Combining Docking and Molecular Dynamics Simulations

Docking and molecular dynamics simulations are highly complementary techniques concerning their specific strengths and weaknesses. The biggest advantage of docking is certainly its speed, allowing the screening of large compound libraries, on a scale of several hundred thousands, at reasonable costs. Conversely, its handling of receptor flexibility, solvent effects and binding entropy is rather poor due to the approximations made [131]. These weaknesses happen to be the particular strengths of MD simulations which simulate both ligand and receptor as flexible molecules in an explicit (or implicit, see Section 2.1.3) solvent environment. The drawbacks of MD simulations are that it is computationally expensive and that the conformational sampling of molecules is limited, i.e. molecular configurations can get trapped in local energy minima. It is thus not astonishing that approaches combining the two techniques in drug design gained much popularity over the last years. A comprehensive review of existing approaches is given by Alonso et al. [178].

Due to the high computational costs, the use of MD simulations is restricted to certain steps during the drug-design process. Usually, MD is used to complement docking 1) before docking, to optimize the target structure and account for its flexibility and/or 2) after docking, for the refinement of docked complexes, and to calculate binding free energies (see below). A third option is to dock via MD simulations but this approach is still very restricted and has only been applied to a limited number of case studies [178].

2.3.1. Using MD Simulations before Docking

If it is known that a receptor undergoes large conformational changes upon ligand binding, docking to a single receptor structures might be problematic because modern docking tools allow only very restricted receptor flexibility (usually limited to a few selected amino acid side chains within the binding pocket). Ideally, several high-resolution X-ray structures of the receptor, in complex with different ligands, would be used for an *ensemble* docking to include different conformational snapshots of the target [179]. If other experimental structures are not available, different conformations of the receptor can be generated computationally by taking representative structures from an MD trajectory.

As each additional conformation increases the chances of producing false positive solutions during the docking, it is better to use a few carefully selected conformations rather than every structure available [180].

2.3.2. Using MD Simulations after Docking

Applying MD to a docked complex is used to refine the complex by allowing the ligand and receptor to relax their conformations, to test its stability and to incorporate explicit solvent molecules. The most popular application, however, is to use MD-based calculations of binding free energies. The routine application of new approaches like the Molecular Mechanics Poisson-Boltzmann (or Generalized Born) Surface Area (MM-PBSA and MM-GBSA) method (cmp. Section 2.1.3), is becoming possible because of enhanced computational hardware [181].

MM-PBSA

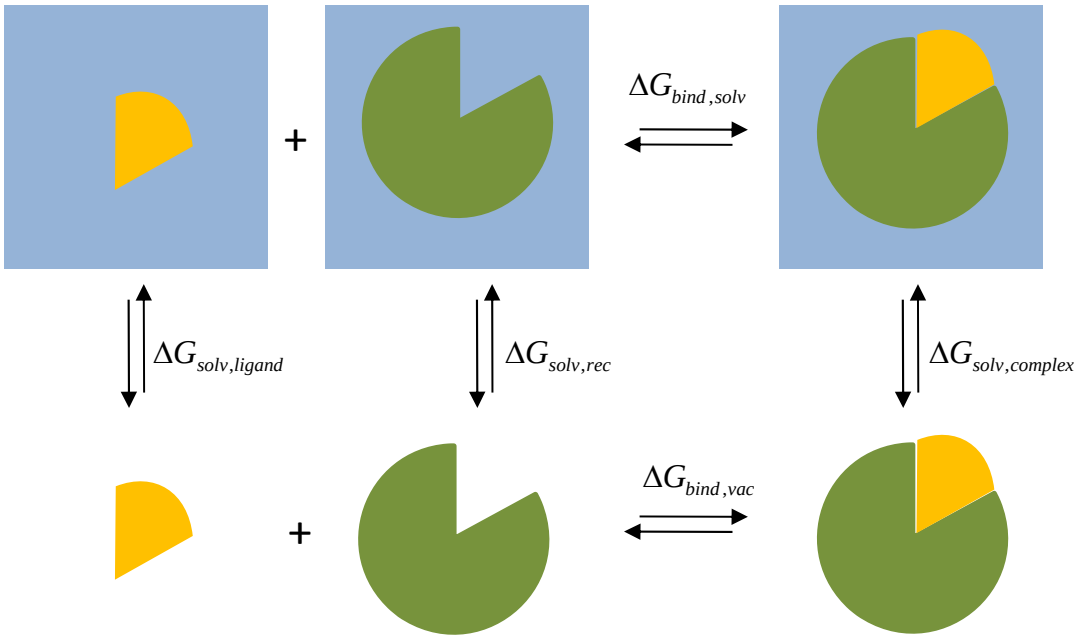


Figure 2.4.: Schematic representation of the solvation cycle. Adapted from [182].

The free energy of binding ΔG_{bind} is the energy difference between the bound state $G_{complex}$ and the unbound state, i.e. the energy of the receptor G_{rec} and the ligand G_{ligand}:

$$\Delta G_{bind} = G_{complex} - \left(G_{rec} + G_{ligand}\right) \tag{2.23}$$

By calculating ΔG_{bind} directly with equation 2.23 the fluctuations in total energy would be an order

of magnitude larger than the binding energy itself [182].

The MM-PBSA approach divides up the calculation according to the thermodynamic cycle (see Fig. 2.4) and takes advantage of multiple structures from a MD trajectory to get energy averages. According to the thermodynamic cycle, the binding free energy can be calculated as:

$$\Delta G_{bind,solv} = \Delta G_{bind,vac} + \Delta G_{solv,complex} - \left(\Delta G_{solv,ligand} + \Delta G_{solv,rec}\right) \tag{2.24}$$

The solvation free energies are calculated as described in Section 2.1.3, separately for the complex, ligand and receptor. Usually, the approximation is employed that the configurational space explored by the systems are very similar between the bound and unbound states; so every snapshot for each species is extracted from the same trajectory (single trajectory approach). For an efficient calculation, explicit water molecules and ions are stripped from the MD trajectory and implicit-solvent calculations are performed on the snapshots, either by numerically solving the Poisson-Boltzmann equation or by using the generalized Born approach. Electrostatic salt effects are incorporated by using Debye-Hückel theory [183].

$\Delta G_{bind,vac}$ is obtained by calculating the average interaction energy between receptor and ligand ΔE_{MM} via molecular mechanics and taking the entropy change ΔS upon binding into account (for example by a normal mode analysis):

$$\Delta G_{bind,vac} = \Delta E_{MM} - T\Delta S \tag{2.25}$$

The entropy is often neglected when ligands of equal size are compared as it is a time intensive calculation for very little added information.

2.4. Pharmacophores

A *pharmacophore* is the spatial arrangement of functional groups, shared by a number of biologically active molecules, forming the basis of their biological effect [184]. According to the official IUPAC-definition established by Wermuth in 1998 [185], a pharmacophore is:

- not a real molecule but an abstract concept accounting for the common interaction capacities of a group of compounds toward their target structure,
- the largest common denominator shared by a set of active molecules,
- defined by pharmacophoric descriptors, such as hydrogen bonds, hydrophobic and electrostatic interaction sites, which can be defined by atoms, ring centers or virtual sites.

An example of a pharmacophore is shown in Figure 2.5.

Deriving a pharmacophore can be approached in a ligand- or protein/structure-based manner [186]. A ligand-based pharmacophore is derived by flexibly overlaying a set of active analog

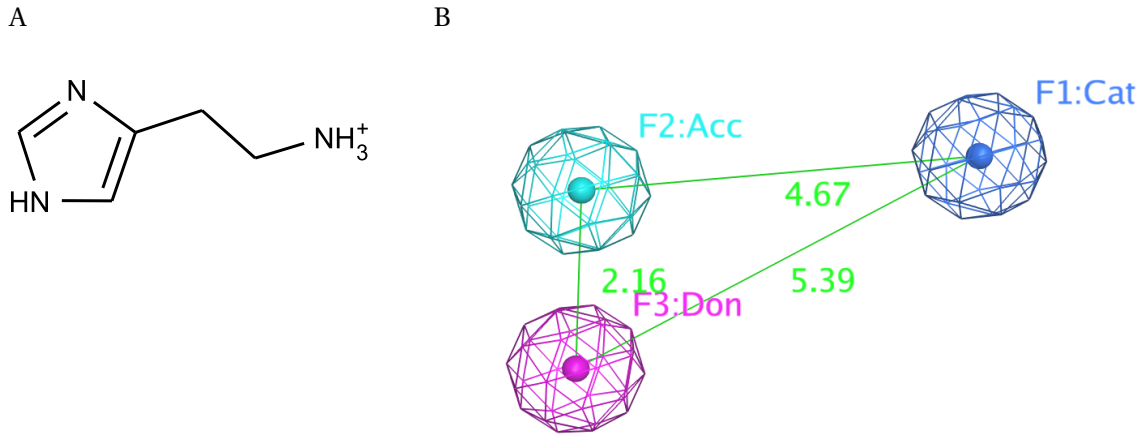


Figure 2.5.: 2D structure of histamine (A) and its derived pharmacophore (B) containing the features F1: Cation, F2: H-bond acceptor, F3: H-bond donor. Adapted from Ref. 184.

ligands; a protein-based pharmacophore can be defined by "hot spots", which are energetically favorable binding sites determined by the residue composition of the protein surface, as realized in the program *GRID* by Peter Goodford [168]. Recently, a third strategy has been developed by Wolber and Langer [187]. Their program *LigandScout* derives pharmacophores from protein–ligand complexes.

A pharmacophore model can be used to screen compound libraries in order to find molecules that share the pharmacophoric features, but have a different molecular scaffold or residues. This approach is used to find different compounds that are biologically active against the considered target, but either are not yet patented or exhibit superior physicochemical properties.

3. The Influence of Thiostrepton Binding on the Dynamics of the Ribosomal L11·23S Subunit

3.1. Introduction

The thiopeptide antibiotic thiostrepton [34, 35] inhibits bacterial protein biosynthesis by tightly binding to the cleft formed by the GTPase-associated region's L11 protein and H43/H44 of the 23S rRNA [46, 188–190] (see Section 1.3 for an introduction into the structure and biological role of the GAR). This ternary binding event was shown to be highly cooperative and necessitates the presence of both rRNA and protein components [45–47, 191]. Furthermore, recent data suggest that TS binding impedes stable binding of EF-G and EF-4 [192]. Interpretations of ribosomal crystal structures in the 3.2–4.2 Å resolution range have strengthened the hypothesis that TS restricts the protein's N-terminal domain movement, and thereby prevents proper binding and processing of translation factors [28, 32]. However, the resistance data for thiopeptide antibiotics cannot easily be aligned with a static lock-and-key model for inhibitor binding [191]. Obtaining atomic resolution data of the L11·23S binding site dynamics and the effects of inhibitor binding would improve the understanding of the multifaceted interaction between ligand, protein and RNA – and should guide further drug discovery.

One possible mean for acquiring such data are molecular dynamics simulations. Thereby insights into the structure and dynamics of large biomolecular complexes can be provided *in silico* at the atomic level that cannot yet be generated by other methods. When implemented and results interpreted carefully, MD simulations provide valuable qualitative and semi-quantitative information about the structure and dynamics of folded RNA molecules and protein-RNA complexes [193–195], especially when calibrated by experimental observation. For example, simulations have revealed functionally significant stochastic fluctuations of RNA building blocks, where ribosomal kink-turns can act as flexible molecular elbows during translocation [196, 197]. The intrinsic flexibility of H38 [198] and the H42–H44 segment of the 23S rRNA have been described by simulations [199]. More recent simulations on three ribosomal RNA three-way junctions demonstrated significant anisotropic hinge-like flexibility between stacked stems [200].

In the last decade, MD simulations of protein–RNA complexes have been enriched by the growing number of high-resolution crystal structures [194]. These works include coarse-grained [201] and all-atom [202, 203] simulations of the entire ribosome, as well as simulations of diverse rRNA–protein interactions [204–207], including the ligand-free GAR [206]. By using an MD simulation, Frank and coworkers identified two L11 conformations that matched cryo-EM maps and were interpreted as bound and unbound states to the aa-tRNA·EF-Tu·GTP ternary complex [206]. Thus, despite the ribosome's challenging size and molecular complexity, MD studies have contributed valuable insights into its functional properties.

In this chapter, results from all-atom MD simulations of the binary L11·23S complex and the ternary L11·23S·TS complex are compared to investigate the ligand's binding and its influence on the dynamics of the ribosomal protein-RNA complex. It is demonstrated that TS binding restricts the conformational flexibility of the nearby NTD, inluences the stability of its specific secondary elements, and exerts weak but detectable dynamic coupling to the distant CTD. Furthermore, distinct conformations of the far more flexible "apo" form of the N-terminal domain are identified that may reflect distinct interaction states with translation factors.

3.2. Materials & Methods

3.2.1. Molecular Dynamics Simulations

To construct the computational models, the crystal coordinates (PDB code 3CF5) [32] of the TS-bound 50S subunit from *Deinococcus radiodurans* were used. Only the atoms associated with the ribosomal protein L11, with H43–H44 of the 23S rRNA, and TS were extracted. For the simulation without ligand, the resulting vacancy was filled with water. All simulations were performed using AMBER9 and 11 [82]. The Parmbsc0 [90] and Parm99SB [89] force fields were used for the RNA and protein residues, respectively.

The TS model preparation (Dr. Karl N. Kirschner) required special consideration due to its constitution and large size. TS was separated into its constituent residues as defined by Bond et al. [208]. Parameters for the standard amino acid residues were taken from the Parm99SB force field [89]. For the non-standard residues, intramolecular and Lennard-Jones parameters were taken from the General Amber Force Field (GAFF) [91]. Each non-standard residue was capped using the nearest functional group present in the structure, and terminated using a methyl group – as exemplified in Figure 3.1. This was done to include any local electron delocalization present in the macrocycle during the determination of a residue's partial atomic charges. These residue analogues where optimized at the HF/6-31G(d) level of theory, and verified as minima using frequency analysis. Molecular electrostatic potentials were subsequently constructed using the CHELPG methodology [95]. A restrained electrostatic potential fit (i.e. RESP) was performed using a 0.01 weighting factor to obtain partial atomic charges that are compatible with Parm99SB [88]. The capping groups were

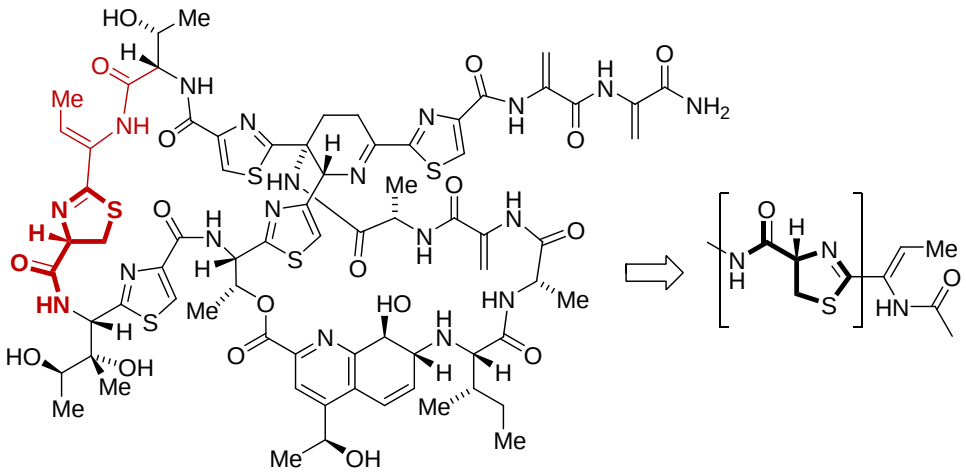


Figure 3.1.: Substructure extraction from thiostrepton (TS) and illustration of capping non-natural amino acid residues for partial charge determination. In the case shown, the thiazoline-4-carboxyl residue (brackets) is capped using the nearest functional groups of TS (dehydrobutyrine and thiostreptine).

subsequently removed from each residue, and the remaining residual charge was neutralized. All quantum mechanics calculations were performed using Gaussian03 [209].

The RNA–protein complex was neutralized using 50 sodium ions and solvated in a truncated octahedron of TIP3P [102] waters with a 15 Å padding in all directions. A series of minimizations were performed on the model to an energy gradient convergence of 0.6 kcal·mol^{-1}. Afterwards, three sequential MD simulations were performed where the portion of allowed motion included a) just the water molecules and sodium ions, b) water, sodium ions, added hydrogens, and capping residues, and finally c) the entire system. Each simulation was heated from 5 to 310 K in 100 ps and cooled back to 5 K in another 100 ps; for temperatures below 155 K the simulations were performed at a constant volume (NVT), while over 155 K the simulations were at a constant pressure of 1 bar (NPT). Finally, a constant pressure production run was conducted at 310 K without restraints. A chart describing this simulation protocol can be seen in Figure 3.2. Density, volume, and energy were monitored, as appropriate for each stage, to ensure simulation stability. A non-bonded cutoff of 9 Å was used, while the Particle Mesh Ewald method [210] was employed to capture the non-bonded interactions at longer distances. Electrostatic and vdW interactions between atoms separated by three bonds were reduced by factors of 1.2 and 2.0. Temperature regulation was controlled using Langevin dynamics with a collision frequency of 1 ps^{-1} and a unique random-prime-number seed was used at each restart point [82]. The SHAKE algorithm [211] was used to constrain bonds involving hydrogen atoms, and subsequently these bonds were excluded from the force evaluation. A time step of 1 fs was used during equilibration and increased to 2 fs during the production run. Coordinates were recorded every 1 ps. To verify observations, each of the two simulations were repeated for another 40 ns using the same protocol, but starting with a different random number seed.

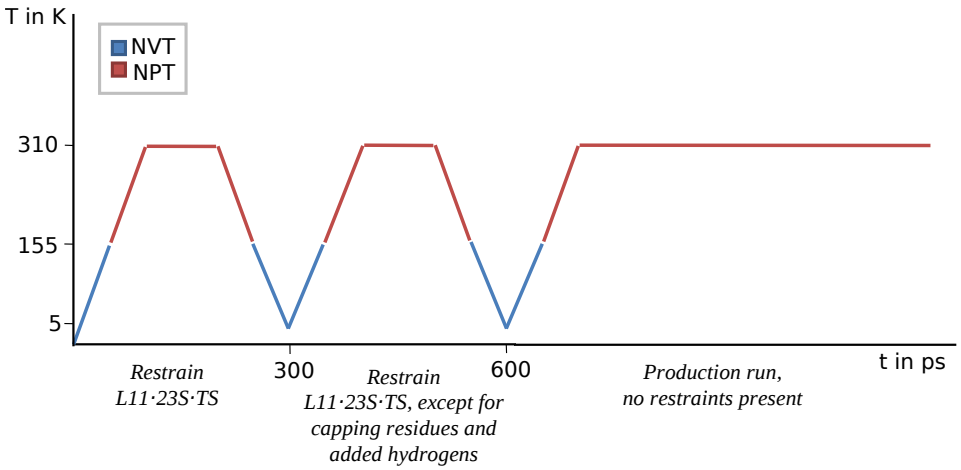


Figure 3.2.: The molecular dynamics method used for all models studied here. Blue segments correspond to constant volume, orange segments to constant pressure simulations.

3.2.2. Analysis of MD Trajectories

All analyses were performed on the trajectory data generated from the production simulation runs for both models. The underlying theory and equations of the used analysis methods can be found in Section 2.1.5. Prior to each analysis, the structures within each trajectory were aligned using the backbone atoms (i.e. Cα, C, N for the protein and P, C3', C4', C5', O3', O5' for the RNA) to the X-ray structure, except were stated otherwise. Such pre-analysis alignment removes system-wide translational and rotational movements.

RMSD and RMSF

The root-mean-square deviation (RMSD) and fluctuations (RMSF) were calculated for the protein and RNA backbone atoms using the PTRAJ module of AMBER [82]. The RMSF profile was generated residue-wise by averaging over the backbone atoms of each residue.

Cross-Correlation Analysis

To identify potentially interesting dynamic correlations between or within the RNA and protein structure, a dynamic cross-correlation analysis was performed. The necessary covariance matrices were computed using PTRAJ. The corresponding dynamic cross-correlation maps (DCCMs) were plotted using the statistics package R (V2.10) [212].

As usually only internal fluctuations are of interest, it is necessary to remove translational and rotational rigid-body motions from the simulation. As mentioned above, this is usually done by aligning the snapshots of the trajectory onto specific atoms of the first coordinate set or an average structure. However, the procedure for selecting these atoms is not unique and has even been a topic of debate [213, 214]. The heterogeneity of the investigated system (protein·RNA·ligand) further

complicates this decision. Consequently, several alignment strategies were tested for their influence on the resulting DCCMs for the simulation with TS (see Fig. 3.3). The alignment was done using the Cα atoms of the protein (Fig. 3.3A), the phosporus atoms of the RNA (Fig. 3.3B), a combination of them (Fig. 3.3C), and using the backbone atoms of the protein and RNA (Fig. 3.3D).

Figure 3.3 shows how misleading an alignment on either the protein or RNA part alone can be. When the trajectory is aligned on the Cα atoms only (Fig. 3.3A), the secondary structure of the protein is captured nicely but almost all of the RNA seems to be highly correlated. In Figure 3.3B, showing the alignment on phosphorus atoms, however, correlations are only visible within the C- and N-terminal domain of the protein, while the rest shows no correlation at all. Regions of correlation and anti-correlation became more differentiated, when both the RNA and protein part were included in the alignment procedure. No significant differences were observed when either Cα and phosphorus atoms, backbone atoms, or all heavy atoms of the residues were used, except that with increasing number of atoms included in the alignment the cross-correlations became less clear. Therefore, it was decided to align the trajectory on the RNA and protein backbone atoms.

Accordingly, different scenarios with an increasing number of atoms used for calculating the cross-correlations for the protein and RNA residues were tested while keeping the RMSD criterium constant (backbone atoms). Using either Cα+P, backbone, or all heavy atoms showed negliable effects in the resulting DCCMs (data not shown). Very local fluctuations, as from the side chains, seem to make relatively minor contributions, as was already noted by Ichiye et al. [215]. Therefore, a mass-weighted average over the backbone atoms was used for calculating the cross-correlations between residues.

Another aspect of analysing dynamic cross-correlations is the start and length of the simulation time window to be included in the calculation. In previous studies usually the complete simulation time was considered [119, 120, 213, 215–218]. Occasionally, maps of the first and second half of the simulation were compared to verify equilibration. To analyze the effect of considering different simulation time windows in more detail, dynamic cross-correlations averaged over a 10 ns-window shifted by 1 ns over the simulation with TS were calculated. Figure 3.4 shows selected DCCMs for a time-span of 10 ns shifted over the simulation from 0.5 to 30.5 ns. At the beginning of the simulation, large areas of (anti)correlations are visible, decreasing with simulation progression. For the last time windows (starting at 6.5 ns) the overall (anti)correlations are significantly reduced and stay more or less constant. The DCCMs up to this simulation time likely contain equilibration artifacts and thus show in general higher and broader areas of (anti)correlation. Analyzing DCCMs for different time-spans of the simulation can thus be used as an additional convergence criterion. Moreover, they can also reveal conformational transitions (see for example Fig. 3.23).

Principal Component Analysis and Clustering

Principal component analysis [121, 122] was done using R and the bio3D package [219], which was developed specifically for analyzing biomolecular data. Since initially the large-scale motion of the

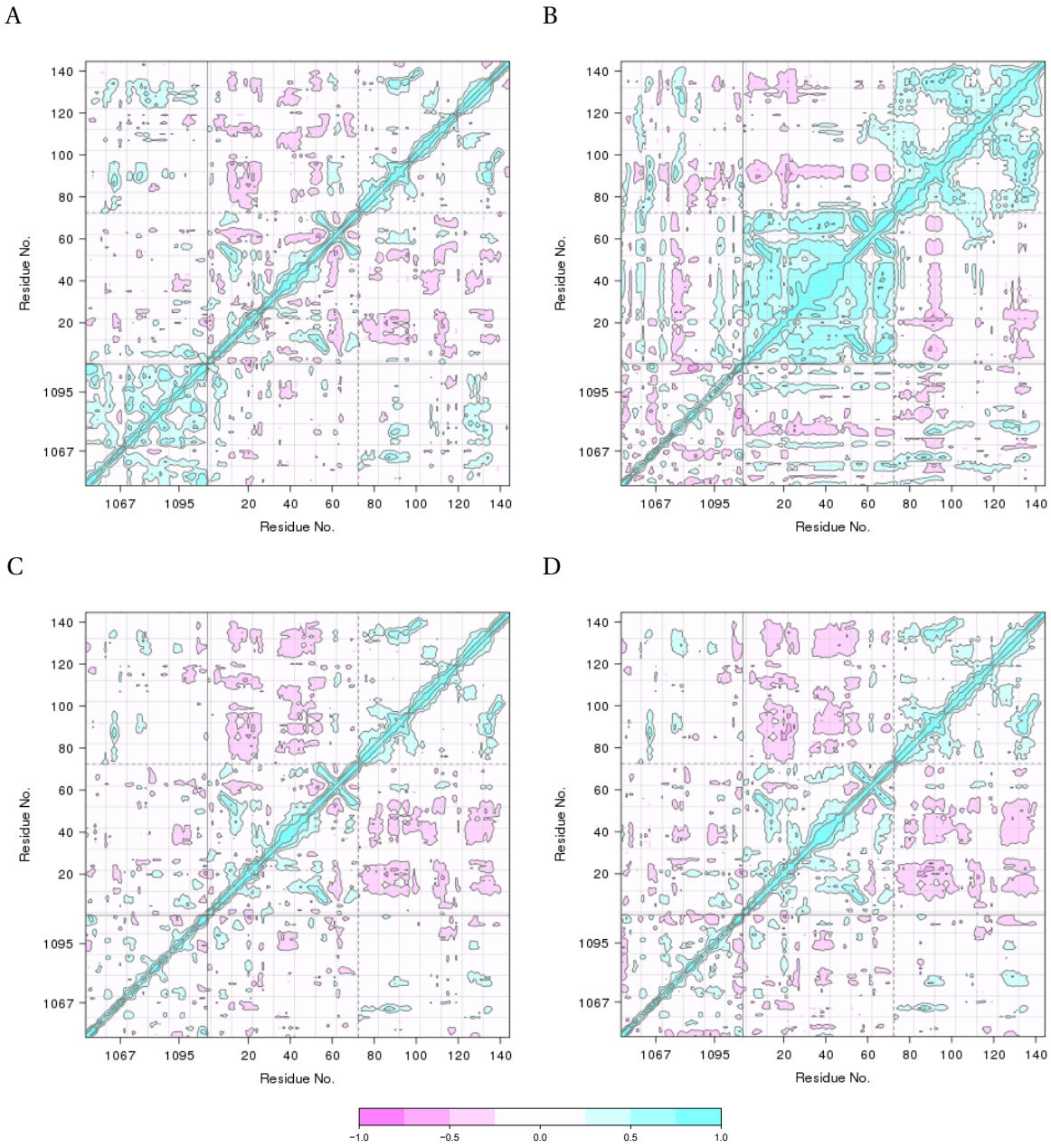


Figure 3.3.: Dynamic cross-correlation maps (DCCMs) of the L11·23S subunit using different alignment criteria. The last 33 ns of the simulation with TS were considered. RNA and protein residues are divided within the plots using solid black lines, the two protein domains are further divided by dashed gray lines. Cyan colors represent positive correlations, magenta anti-correlations (see color bar at the bottom). A) Alignment on Cα atoms of protein. B) Alignment on phosporus atoms of RNA. C) Alignment on Cα and phosphorus atoms. D) Alignment on RNA and protein backbone atoms.

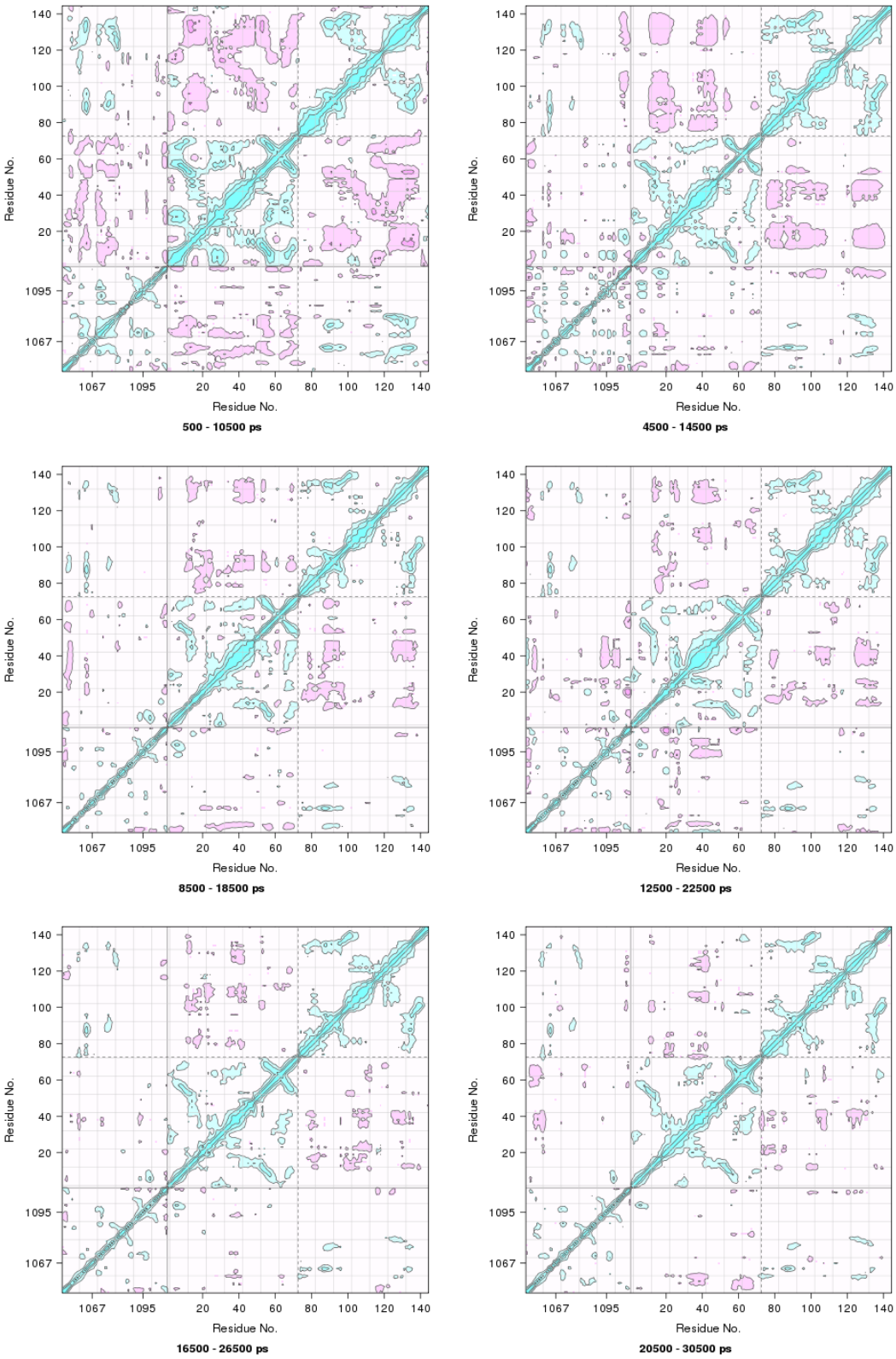


Figure 3.4.: Dynamic cross-correlation maps (DCCMs) for different time-spans of the simulation with TS.

L11·23S subdomain are of interest, side-chain local motion was excluded; this was achieved by using only the coordinates of the protein's Cα and RNA's phosphorus atoms. This approach does two things: 1) reduces the raw data analyzed into an amount that is more computationally tractable, and 2) is an initial filter that removes the higher frequency side-chain motion from the analysis. As done above, the trajectory snapshots were aligned to the X-ray structure using these atoms.

While PCA reveals the main motions contained in the MD trajectory it does not provide a partition of the snapshots into distinct conformational classes. This can be achieved by clustering the data. It will be investigated in the following if clustering PCA data improves cluster quality versus clustering the "raw" data set and how two different clustering algorithms perform. Two widely used clustering methods, each representing a different clustering approach: K-means (partitional) and average-linkage (hierarchical) were compared (see Section 2.1.5 for an introduction into clustering). For this investigation, the data from the simulation without TS after equilibration is used (2.5 to 40 ns, see Section 3.3.1 for how convergence of the simulation was determined).

K-means (algorithm of Hartigan and Wong [220]) and average-linkage were used, as implemented in R. Due to the high storage demand of the average-linkage algorithm not all 37,500 snapshots could be included into the clustering. Therefore, every 5th snapshot was considered when clustering with both algorithms (7,500 snapshots). A test with K-means on the complete snapshot set showed that this did not influence the clustering outcome (cmp. Figs. 3.9C and 3.17A).

Beside removing high-frequency side-chain motion from the cluster analysis by considering only backbone or the Cα+P trace, clustering can also be perfomed on a further reduced subset, such as the data contained within the first few PCs, where the high-frequency motion and "noise" has been removed. Then the clustering is performed on the data projected into the PC space. The assumption present in this combined technique is that the clustering focuses on data that is more relevant to large-scale motion within the interested molecular species. Consequently the question arises, "How many components should be included into the cluster analysis?" Common subsets of PCs, typically analysed in PCA, are the first two or three PCs since most of the overall variance is often captured within the first two or three components (e.g. 51.7 % and 60 %, respectively, as seen in Fig. 3.16). Moreover, these subspaces can be visually interpreted by humans and often also in a biologically meaningful way. Additionally, the fifth-dimensional PC subspace was included based on the scree test [128]; the fifth PC data point is where the percentage of variance becomes relatively horizontal as shown in Fig. 3.16, a criterion known as the "elbow-criterion." In summary, clustering was performed on two-dimensional (2D: PC one to two), three-dimensional (3D: PC one to three), and five-dimensional (5D: PC one to five) subspace, and the complete data set.

Another important question is, "How many significant clusters are there?" Determining the number of clusters in the data is a frequent problem in cluster analysis, especially for algorithms such as K-means which requires this value as an input parameter. The SSR/SST ratio (see Eqs. 2.10 and 2.11 on page 23) was calculated as a statistical measure to determine the optimal number of clusters for both algorithms (Fig. 3.5). Both algorithms show a SSR/SST ratio of ~0.5 at a cluster count of two

which means that the data's variance is already halved by chosing two subsets. Thus, the data clearly supports at least two clusters. Another increase is found at a cluster count of 4 although a clear "kink" cannot be identified, leaving a choice of 4 somehow arbitrary. However, the same conclusions can be drawn from a visual inspection of the average-linkage's cluster dendrogram (Fig. 3.6). Overall, both SSR/SST curves indicate a cluster count of 4 is reasonable since the steepness decreases thereafter.

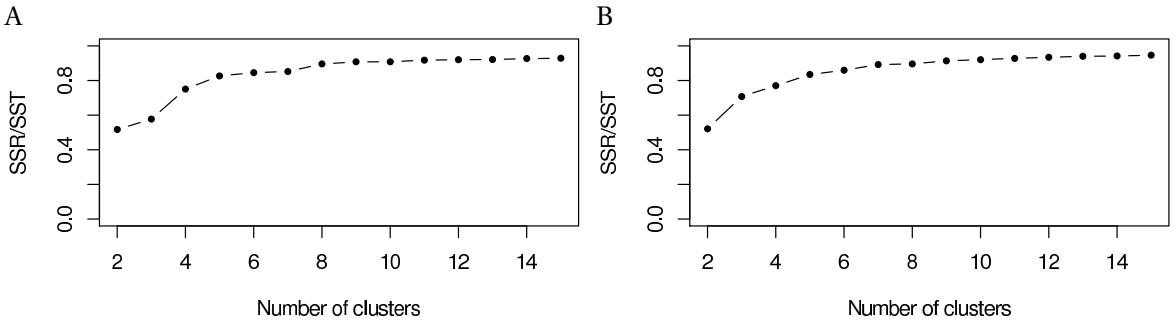


Figure 3.5.: SSR/SST ratio over cluster count for the entire MD trajectory data. A) Average-linkage; B) K-means algorithm.

In addition to analyzing the cluster dendrogram of the average-linkage algorithm (Fig. 3.6), the separation of data points along the axes of the first three PCs was considered as a second visual criterion (see Supplementary Animation 2 of Ref. 221). This visualization confirmed the choice of four clusters, showing that cluster 1 and 3 are clearly separated by the first PC. The second PC additionally separates cluster 2 and 3 from 1 and 4. From the SSR/SST ratio, and the visual analysis it is concluded that the trajectory contains a maximum of four distinct conformational classes and therefore a cluster count of 4 is fixed for further analysis.

In a similar fashion, by combining an objective statistical measure with different visual analytics, the clustering in the different PC subspaces was evaluated. Figure 3.7 shows the pseudo F-statistic (see Eq. 2.12 on page 23) for the two clustering algorithms. To compare the absolute pFS values, the number of clusters was fixed (i.e. 4) and the pFS was calculated on the complete data set, defining the clusters as obtained from clustering in the different subspaces. For K-means the values are very similar. For the 2D and 5D subspace and the complete data the pFS value is essentially the same. Only for the 3-dimensional subspace a small rise in the pFS value (i.e. a better clustering) can be seen. For clustering with the average-linkage algorithm, the pseudo F-statistic is consistently lower than for K-means and shows higher variation. Only in the 2D case the pFS is slightly higher than for the K-means algorithm but still lower than K-means clustering in the 3D subspace. In summary, the K-means clustering in the subspace defined by the first three PCs achieves a better separation of data points than in any other subspace and also the complete data set. These statistical results become clearer in Figures 3.8 and 3.9, which visualize the trajectory data points (i.e. structural snapshots) in the PC1–PC2 plane, colored according to their cluster membership.

Substantial differences in data-point assignments to different clusters are seen in using the average-

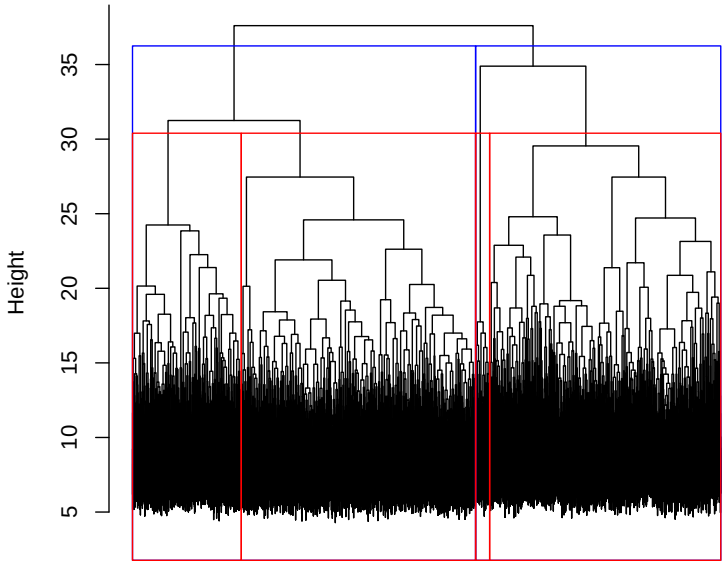


Figure 3.6.: Clustering dendrogram produced by the agglomerative average-linkage algorithm using the entire MD trajectory data. The blue rectangles denote the cut to obtain two different clusters, the red rectangles denote four different clusters.

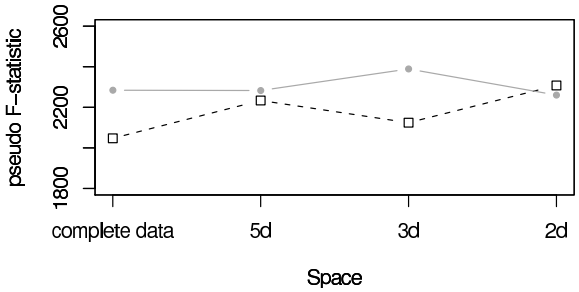


Figure 3.7.: Pseudo F-statistic for clustering in different subspaces. A fixed number of 4 clusters was used. Black squares: average-linkage, gray circles: K-means algorithm.

linkage algorithm when performed on different dimensions of PC subspaces (Fig. 3.8). Especially the size of cluster 2 (lighblue) changes. It is very small in favour of a large cluster 1 (blue) when clustering the complete data (Fig. 3.8A), medium-sized for 5D (Fig. 3.8B) and 3D (Fig. 3.8C) which differ in their boundary assignments between cluster 1 (blue) and 3 (red), and largest in the 2D case (Fig. 3.8D). In contrast, the K-means results are considerably more consistent in cluster size and shape, regardless if clustering was done using the complete data set or any number of PCs studied (Fig. 3.9A–D). Only a few data points are affected, corresponding to data within cluster boundary regions.

For both algorithms one can observe the largest difference going from two-dimensional (Figs. 3.8D and 3.9D) to three-dimensional subspace (Figs. 3.8C and 3.9C). Both clustering algorithms identified similiar core regions within the PC plane that belong to unique clusters. However, the algorithms differ in the assignment of data points within the boundry regions between clusters. For the average-

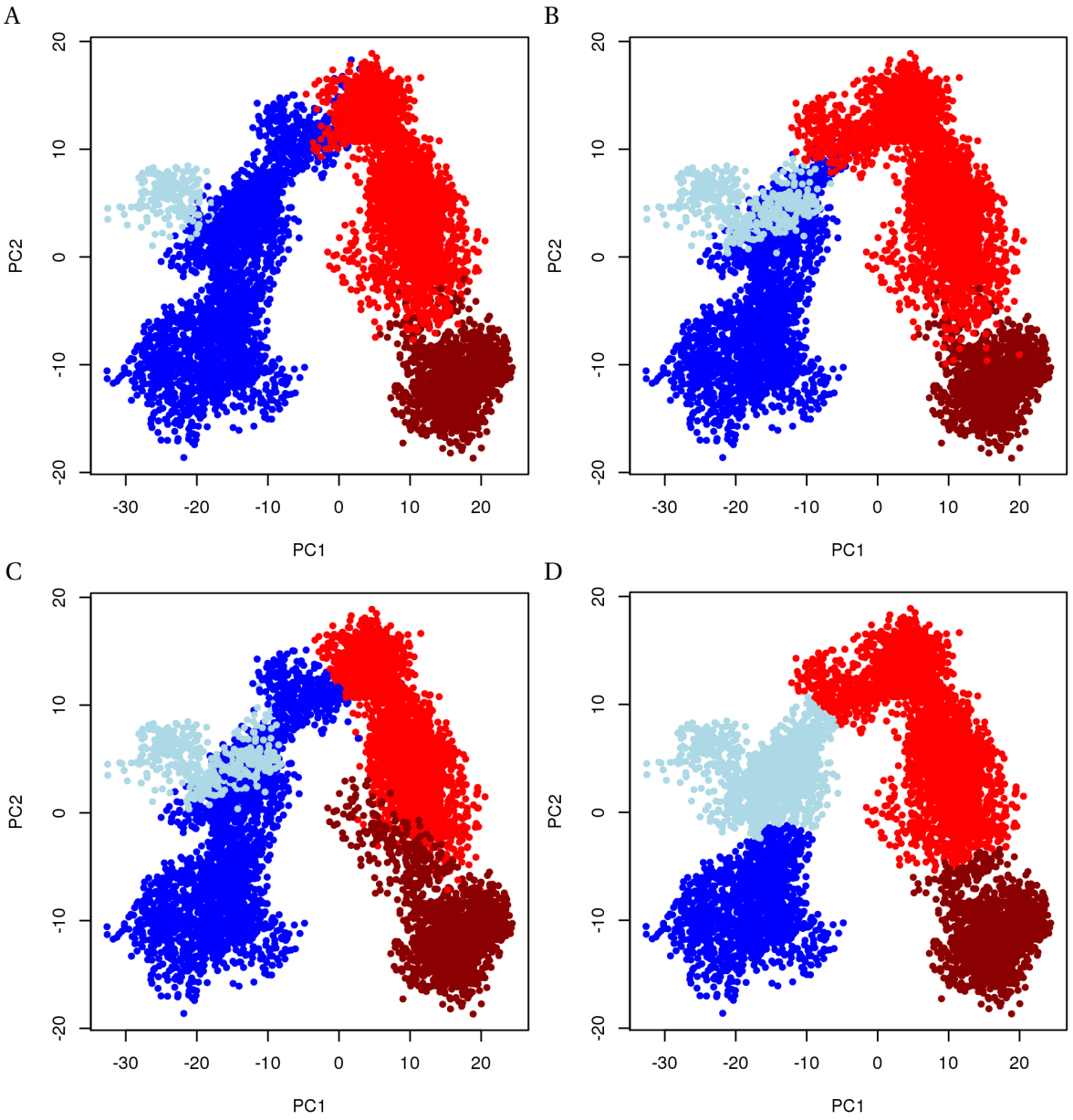


Figure 3.8.: Clustering results of average-linkage algorithm on different subspace dimensions projected on the 2D plane formed by the first two PCs. Clustering was performed on the entire MD trajectory data (A), and on the data from the first five (B), three (C), and two (D) principal components. Key: cluster 1 is blue, cluster 2 is lightblue, cluster 3 is red, and cluster 4 is darkred.

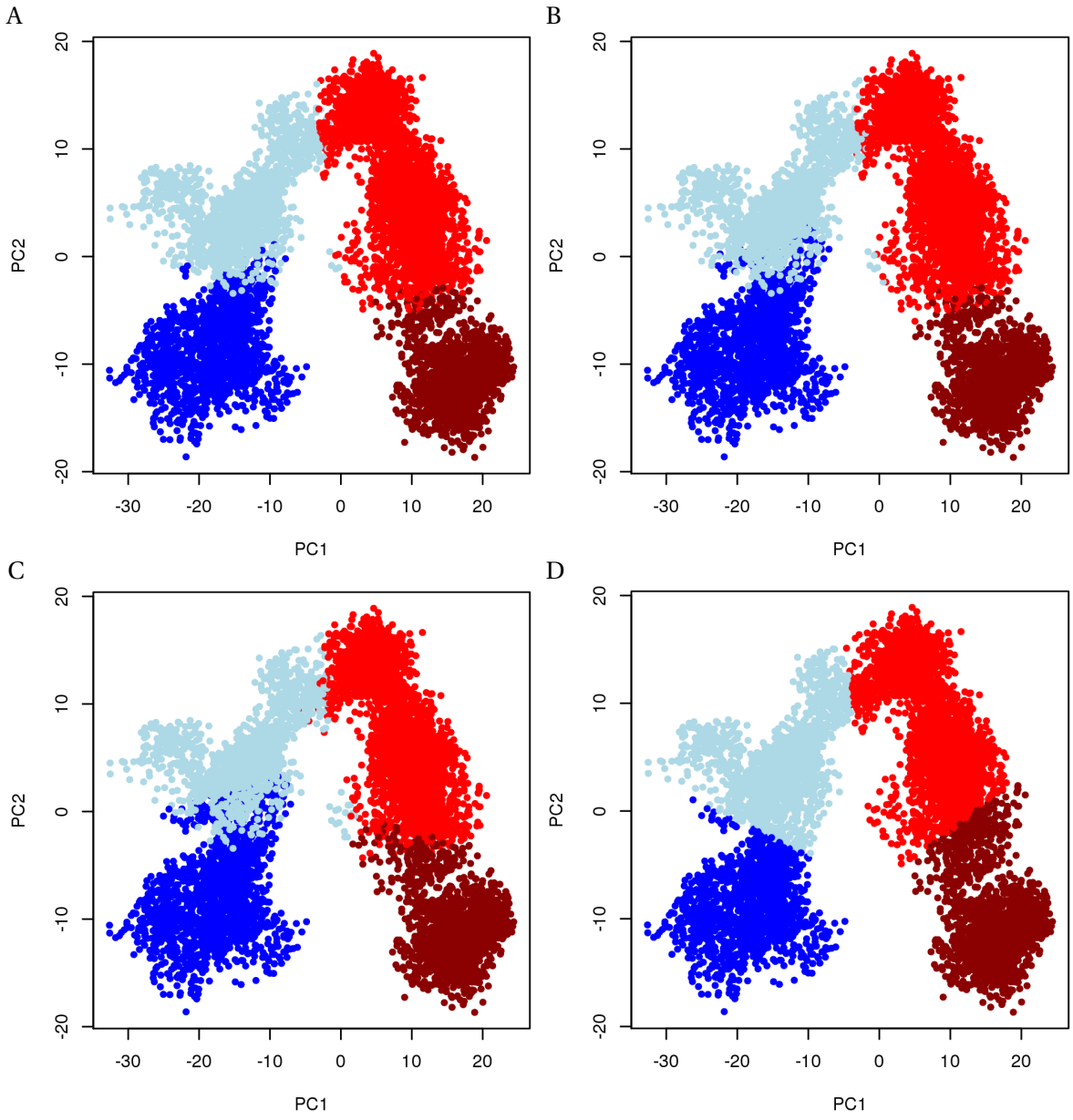


Figure 3.9.: Clustering results of K-means algorithm on different subspace dimensions projected on the 2D plane formed by the first two PCs. Four clusters were requested in the computation. A) Complete data; B) first five PCs; C) first three PCs; D) first two PCs.

linkage algorithm (Fig. 3.8) it is the region shapes and cluster size that changes significantly when varying the PC subspace. Such a dramatic change is not seen when using the K-means algorithm. Thus, the data-point membership to a specific cluster depend more on the selected subspace when using average-linkage clustering than for K-means. K-means appears to be more robust in this regard. Interestingly, selecting a subspace defined by the first two PCs (Figs. 3.8C and 3.9C) yield comparable clustering results for both algorithms and is the only clustering condition that leads to a slightly better clustering result for average-linkage. Whether or not this is a fortuitous agreement is currently unclear.

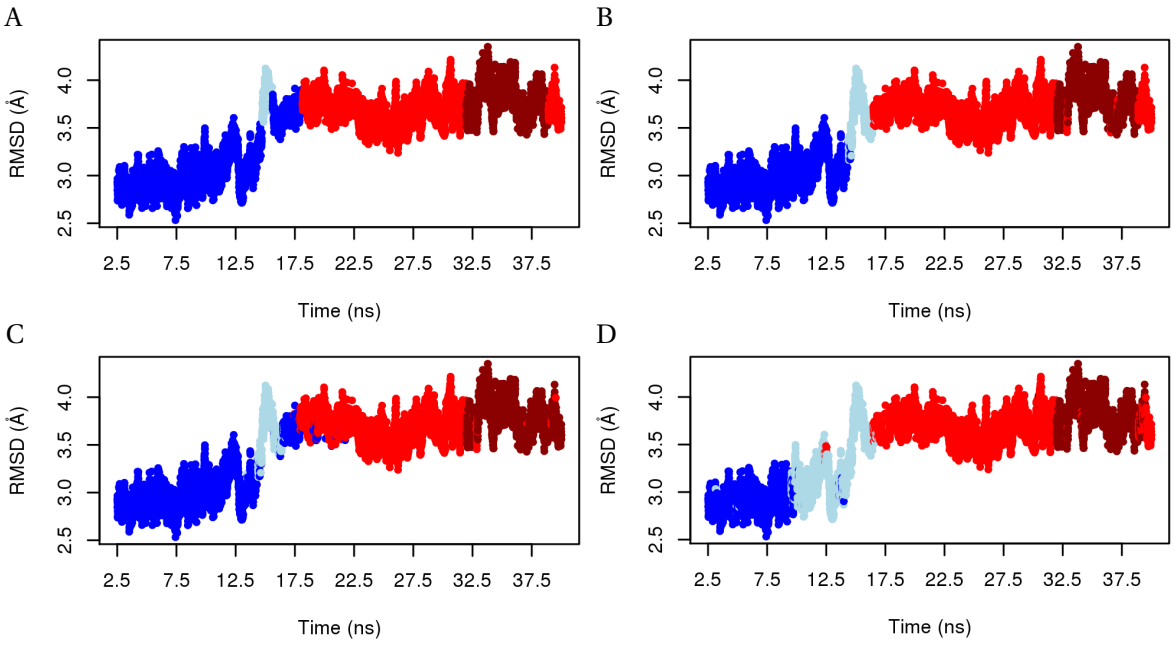


Figure 3.10.: Root-mean-square deviation (RMSD) over simulation time and color-coded by clusters obtained from average-linkage algorithm for the different PC subspaces and the complete data. The colors correspond to the cluster as defined in Fig. 3.8. A) Complete data; B) first five PCs; C) first three PCs; D) first two PCs.

Two-dimensional plots are particularly useful for deciding if a clustering result is reasonable and does somehow reflect the data. A projection into the PC1–PC2 plane is a good representation for seeing if a set of clusters "looks sensible" [222], as shown in Figures 3.8 and 3.9. As an additional way to verify and interpret the clustering results the RMSD over simulation time was plotted, and the data points color-coded by cluster membership as determined using the average-linkage (Fig. 3.10) and K-means (Fig. 3.11) algorithms on the different subspaces. Therefore, the resulting clusters were mapped back onto the RMSD versus time plot to provide trajectory time-frames for cluster existence. In contrast to Figures 3.8 and 3.9 these plots visualize the clustered snapshots over time along with their distance to a defined reference structure, which is the 3CF5 X-ray structure. If a dynamic event occurs during the simulation, which changes the structure's conformation, then

the clustering algorithms should be able to distinguish these conformations. At a minimum, if this conformational change is clearly seen in the RMSD vs. time plot, a change in the cluster membership should be seen as well.

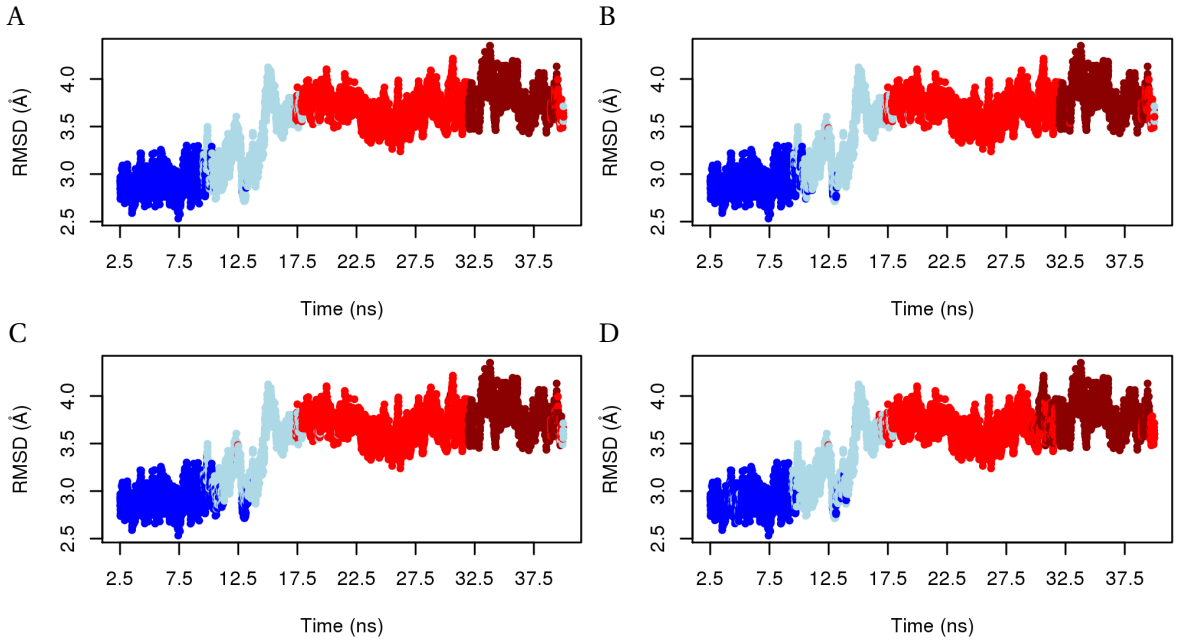


Figure 3.11.: Root-mean-square deviation (RMSD) over simulation time and color-coded by clusters obtained from K-means algorithm for the different PC subspaces and the complete data. The colors correspond to the cluster as defined in Fig. 3.8. A) Complete data; B) first five PCs; C) first three PCs; D) first two PCs.

As expected from Figures 3.7 and 3.9, hardly any differences in snapshot assignment to a cluster are observable for the different subspaces for K-means (Fig. 3.11). For average-linkage, however, the distribution of clusters over simulation is divisive (Fig. 3.10). Clearly separated in all four cases are cluster 1 (blue) at the beginning of the simulation (after equilibration), cluster 3 (red) in the middle, and cluster 4 (darkred) at its end. Disagreement occurs in the their lifetime and in the conformational snapshots assigned to cluster 2 (lightblue). Depending on the data clustered, cluster 2's membership can be fragmented across a dynamic event (Fig. 3.10A and B). Thus, average-linkage clustering of different PC dimensions require careful analysis since the snapshots cannot be unambigiously assigned. In this study, the clustering in the subspace defined by the first two PCs (Fig. 3.10C) provides the most coherent picture. Moreover, this clustering is supported by a good pseudo F-statistic and an almost identical K-means result.

In this study, K-means slightly outperformed average-linkage. K-means results gave better clustering statistics and provided more consistent clustering results (i.e. data-point membership to specific clusters, cluster shape and size) using the different dimensions of PC subspace. Drawbacks to its use include its need for a predetermined cluster count and its tendency to form blocky, spherical

clusters [123]. If the underlying data do not support this cluster structure, K-means will not provide good results.

The clusters found by average-linkage can be of varying size and shape, and the cluster dendrogram that hierarchical algorithms naturally provide proved valuable for determining the optimal cluster count. It was found that the outcome of the average-linkage algorithm was strongly dependent on the selected PC subspace. In the L11·23S model, the choice of the 3D subspace defined by the first three PCs combined with the K-means algorithm provided the best clustering results, even in comparison to using the original data.

In conclusion, using both clustering algorithms to analyze different PC subspaces allowed to form a coherent conclusion concerning the number of clusters present in the L11·23S MD trajectory data. Mapping these clusters onto 2D and 3D plots of the first two and three PCs, and onto an RMSD versus time plot, allowed to understand the clusters' time and conformational space relationship better. Based on these results, the snapshots for the two simulation were clustered along the PC subspace spanned by the first three PCs using the K-means method. Subsequently, the conformational representatives of each cluster were obtained by selecting the conformational snapshot with the lowest RMSD value to the cluster centroid (i.e. the average conformation).

Solvent Accessible Surface Areas

Solvent-accessible surface areas (SASA) were computed with AMBER's CPPTRAJ module using the LCPO algorithm [223]. For the simulation containing TS, the ligand was removed from the trajectory data prior to this analysis. For computing the crystal structure's SASA, capping groups and hydrogens were added to the crystal prior to the calculation.

Secondary Structure

The protein secondary structure was computed throughout each simulation using the DSSP method [224], as implemented in PTRAJ.

RNA Conformations

The χ-angle and sugar pucker of all nucleic acids were monitored throughout both simulations. The χ-angle was defined by the O4'-C1'-N9-C4 torsion, while the sugar puckering was characterized via the pseudorotational phase angle P as defined by Altona and Sundaralingam [225] (for an explanation of the pseudorotation concept see Section B of the appendix). Two additional water solvated simulations were performed on the adenosine monomer (A) and the adenosine trimer (AAA), whose results were used to help validate the finding in the ribosome simulations.

Water and Ion Occupancy

Water and ion occupancies were monitored for the equilibrated simulations. Therefore, water and ion densities were calculated by counting the number of sodium ions and water oxygen atoms in a three-dimensional grid centered on the simulated system using PTRAJ. The grid dimensions were chosen to be 80x80x80 with a 1 Å spacing. Since periodic boundary conditions were used during the simulations, the ions and water molecules were 're-imaged' into their original truncated octahedron prior to the analysis. The resulting contour files were visualized at different isovalues along with an average structure of the macromolecule with PyMOL [226].

3.3. Results

3.3.1. MD Model, Simulation, and Convergence

Simulation equilibration, general stability, and occurrence of dynamic events were monitored by computing the coordinates' relative deviations from the X-ray input structure as a function of simulation time. Figure 3.12 shows the RMSD of the protein and RNA's backbone atoms for both trajectories. Both simulations (with and without TS) displayed a prompt, characteristic deviation up to ~3 Å from the input coordinates after removing the restraints. Such RMSD profiles are not unexpected when using a medium-resolution crystal structure as input (3CF5: 3.3 Å) [32], and also represent the structure's local reorganization from the packed solid to solution phase. After this initial relaxation, the simulation with TS maintains a stable trajectory with an RMSD value of 3.2±0.18 Å, and a reduced fluctuation after 7 ns (sd 0.16 Å) which increases again at the end of the simulation. The simulation without TS equilibrates at a lower RMSD value (2.8±0.18 Å) after 2.5 ns, but a dynamic event takes place at 14.5 ns that introduces a sudden RMSD increase to 3.4±0.13 Å. As seen in Figure 3.12, the RMSD remains stable at this higher value for the remainder of the simulation.

The second set of simuations show similiar RMSD behavior in their trajectores. A comparison of the backbone RMSDs between the first and second simulation runs can be seen in Figure C.1 of the appendix. Whereas, the RMSD distributions stay comparably stable for the simulations with TS (Fig. C.1B), the RMSD increase was even more pronounced in the second simulation without TS (up to 5 Å, Fig. C.1A). For clarity, from hereon the observations gained from the first set of simulation runs will be discussed in detail. In general, similiar results and conclusions from the second set of simulations were found – however, when notable differences do occur they are mentioned appropriately.

Considering the RNA's and protein domains' backbone atoms separately (Fig. 3.13) reveals that in both simulations the rise of the RMSD is dominated by the protein, especially by its N-terminal domain. However, the RMSD of the NTD stays relatively stable in the simulation with TS (Fig. 3.13B), whereas in the simulation without TS the RMSD of the NTD displays a rise of almost 2 Å (Fig. 3.13A) – still visible, although alleviated by averaging over all backbone atoms, in Figure 3.12A.

As another convergence criterion the radius of gyration (R_{gyr}, see section 2.1.5 for its definition)

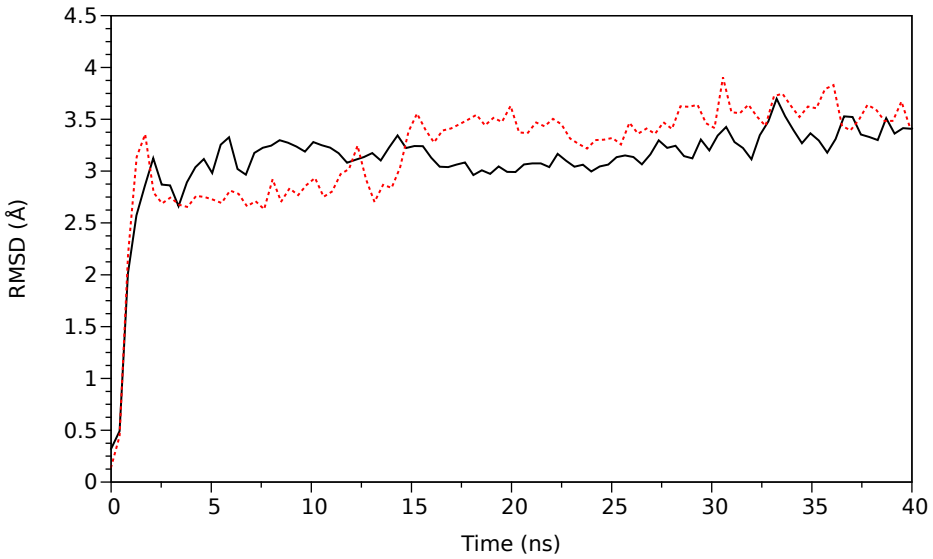


Figure 3.12.: Root-mean-square deviation (RMSD) from the X-ray structure over simulation time. Red dashed line: simulation without TS; black solid line: simulation with TS. Data points smoothed using Bezier curves.

of the protein·RNA complex was considered. At the beginning of both simulations, a large decrease of the radius of gyration can be seen (Fig. 3.14). In the simulation without TS, R_{gyr} reaches a first plateau after 2.5 ns before it drops again after ~14.5 ns. R_{gyr} of the simulation with TS equilibrates later, approximately after 7 ns. Interestingly, the complex exhibits a similar R_{gyr} of 19.8 Å in the first half of both simulations after equilibration (7–14.5 ns), during the last 25 ns the radius of gyration is consistently lower in the simulation without ligand. Overall, the radius span sampled during the simulation without TS is larger (19.16–20.22 Å) compared to the simulation with TS (19.44–20.16 Å).

Faster convergence can be seen in the protein's secondary structure, another indicator of the simulation's convergence. Figure 3.15 shows the percentages of L11's secondary structure elements over simulation time. Larger, short-time fluctuations are only visible within the very first nanoseconds. Afterwards, especially the percentage of sheets and turns becomes very stable. The percentages of helices and irregular structures (including loops) display the highest fluctuations and are also the largest differences between the simulations. In the simulation without TS (Fig. 3.15A), the helices' percentage fluctuates around the value observed in the X-ray structure (35.4 %). In the simulation with TS (Fig. 3.15B), the percentage of helices is approximately 5 % higher (38.9 % on average).

Based upon the investigations of the RMSD and the radius of gyration over simulation time (Fig. 3.12) it was decided that the production-run data begins at 2.5 and 7 ns for the simulations without and with TS.

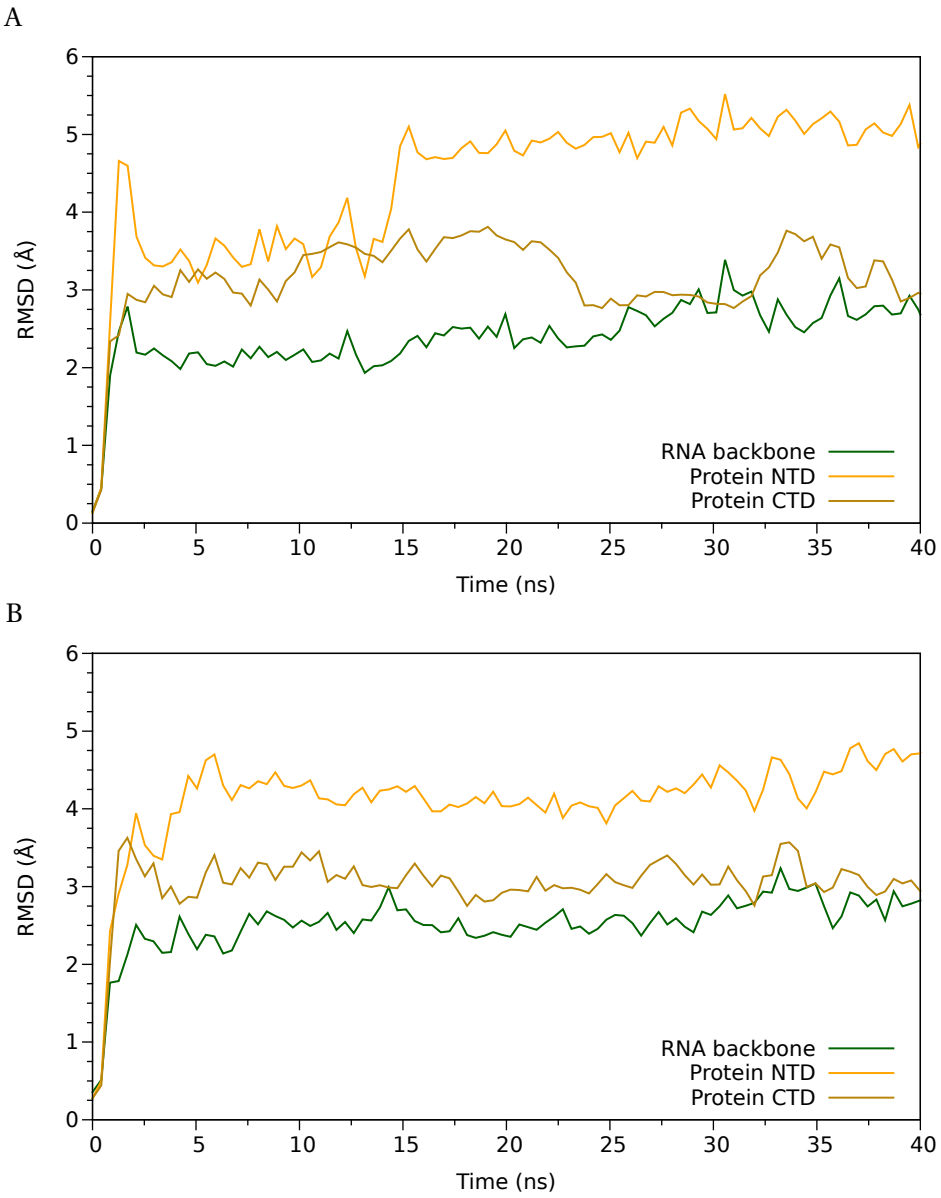


Figure 3.13.: Root-mean-square deviation (RMSD) from the X-ray structure over simulation time, separated for backbone atoms of RNA, L11 NTD, and CTD. A) Simulation without TS. B) Simulation with TS. Data points smoothed using Bezier curves.

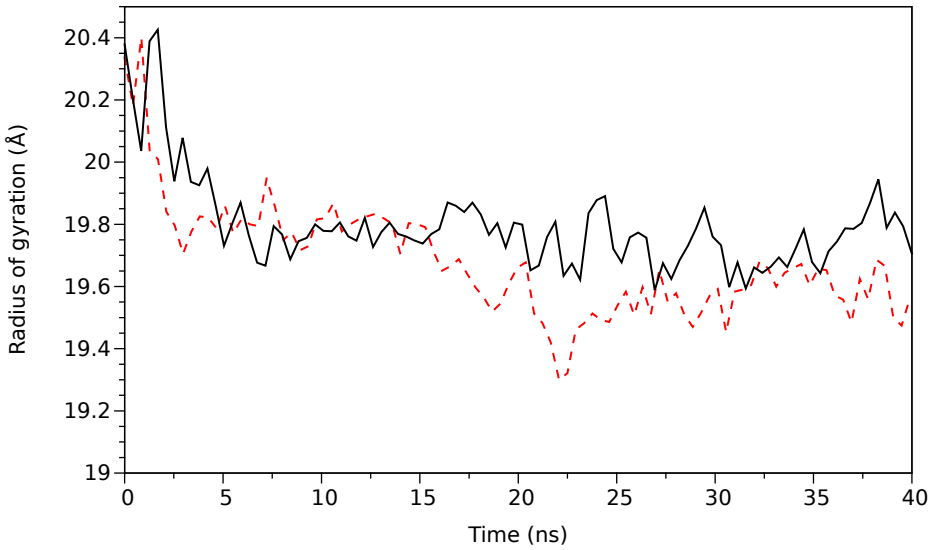


Figure 3.14.: Radius of gyration over simulation time. Red dashed line: simulation without TS; black solid line: simulation with TS. Data points smoothed using Bezier curves.

3.3.2. Essential Dynamics

In an attempt to rationalize the essential conformational changes due to the dynamic event seen in the simulation without TS, a principle component analysis on both trajectories was performed after their equilibration, whose results are presented in Figures 3.17 and 3.18. As outlined in Section 2.1.5, PCA explores the data's variance. In an MD trajectory, the data is the atomic coordinates over time and its variance their dynamic changes. Thus, a PCA can be used to explore the most prominent conformational changes present in an MD trajectory.

For the simulation without TS, the first principle component (PC1) explains more than one third of the overall variance (36.9 %, see Fig. 3.16A); the first three components explain more than 60 % of the complete variance. The residue-wise loadings for PC1 (see Fig. 3.17D for a graphical and Fig. 3.19A for a structural depiction) resemble the overall residue fluctuations (Fig. 3.21A), suggesting that the N-terminal domain contribution dominates this component. Within the N-terminal domain, the largest contributions map onto the terminus, the α1-helix (residues 14 to 27) and the loop connecting α2-β2 elements (residues 43 to 51). Apart from the terminal residues, the C-terminal domain and the RNA contribute little to the motion described by PC1.

The trajectory snapshots are clearly separated into two subunits along the first principal component (colored blue and red in Fig. 3.17A). These two subunits are further divided by the second PC, but these separations are less visually distinct. Therefore, the sampled conformations in the subspace spanned by the first three PCs were clustered into four clusters (Fig. 3.17A and B). The representative of the first cluster has a conformation that is similar to the crystal structure. The second cluster is best described as a conformational transition from cluster one to another stable conformation

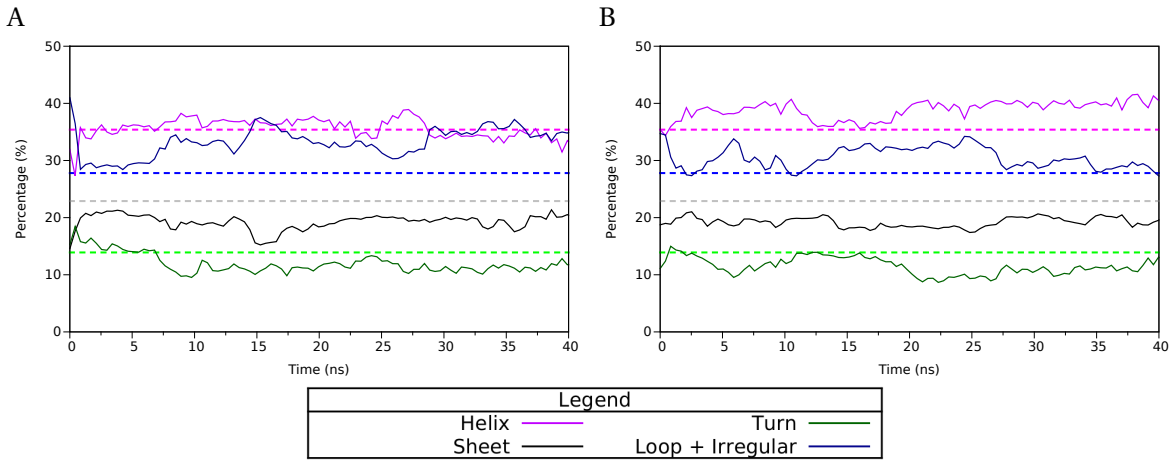


Figure 3.15.: Secondary structure of L11 over simulation time. The secondary structure percentages of the 3CF5 X-ray structure is shown in dashed lines. Data points smoothed using Bezier curves.

sampled by cluster three. Cluster three remains stable for ~15 ns before a closely related conformation (cluster four) is sampled, which lasts for the remainder of the simulation. Figure 3.17C displays the structural representatives of the four clusters, whose differences can be described as a rotational movement of the NTD with respect to the CTD (as indicated by the arrows). Their main differences are in the position of the N-terminus, the nearby loop connecting the β2-β3 strands, and the two loops connecting the helices with the β-strands on the far end (i.e. away from the binding site) of the N-terminal domain.

The PC analysis of the second run without TS created a similar picture (Fig. C.2, appendix). The first principal component, explaining 42.8 % of the overall variance (Fig. C.2C), clearly separates the snapshots into two distinct K-means determined clusters (Fig. C.2A,B). The representative structures of the two clusters identified in the second simulation aligned very well with cluster one and three identified in the first simulation (Fig. C.3, appendix). In both sets of clusters, one can clearly see a difference in the NTD orientation relative to the RNA, validating the observed conformational states. However, a significant difference is seen in the residue-wise loading for PC1. Apart from similar NTD contributions that were already seen in Figure 3.17D, the second simulation show that the artificially truncated RNA endings have a large contribution to the first PC (Fig. C.2D). As this is likely an artifact of the model creation, it was decided to focus the remaining analysis and conclusions concerning L11·23S behavior in the absence of TS on the first simulation since its RNA termini are dynamically more stable.

The first principal component for the trajectory containing TS explains a little less overall variance (29.8 %, see Fig. 3.16B). The residue contributions are distributed in a broader manner throughout the L11·23S complex (Figs. 3.18D and 3.19B). Notable are the missing large contributions from the first three N-terminal residues and those around residue 50 observed in Figure 3.17D when TS is not present. Instead, sharper peaks are seen around residues 20 and 62 of the N-terminal domain, which

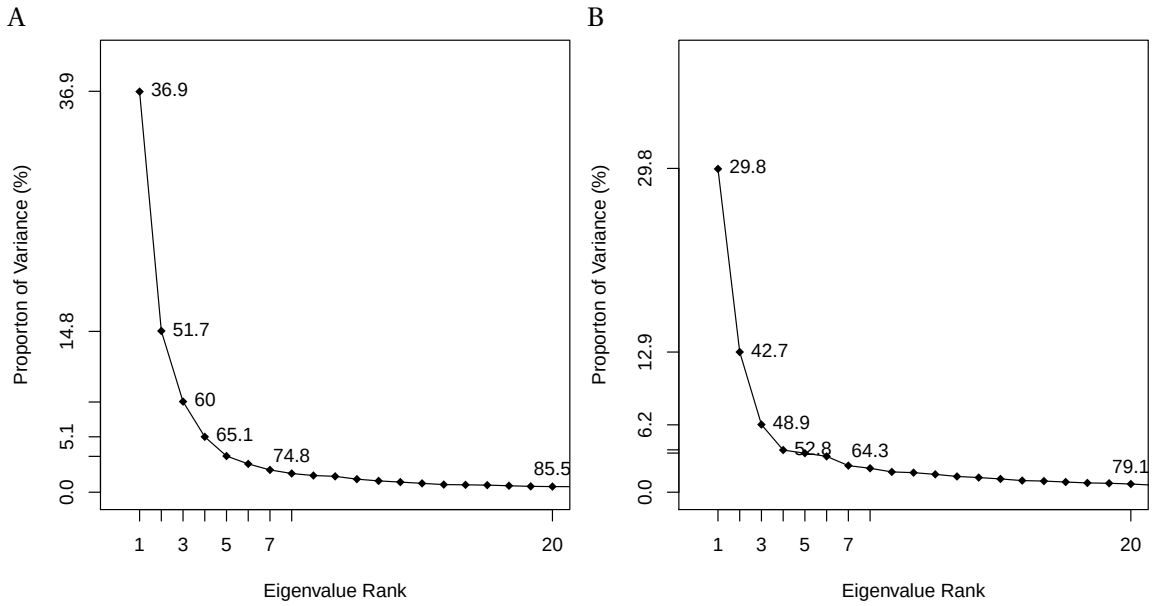


Figure 3.16.: Proportion of variance declared by the principal components (scree plot). A) Simulation with TS. B) Simulation without TS.

help form the loops connecting β1 to α1 and connecting β2 to β3. Clustering the conformations in the PC1 to PC3 subspace revealed that the cluster representatives were conformationally similar (Fig. 3.18C), more so than those observed in the simulation without TS. Furthermore, each cluster has similar RMSDs to the conformation described by cluster three found in the simulation without TS – using cluster three as a reference (i.e. the red structure in Figure 3.17C) backbone RMSDs of 2.0 to 2.16 Å (or 2.48 to 2.99 Å considering only the N-terminal domain) were found.

The differences between the N-terminal domain conformations described by PC1 are further underlined through an RMSD comparison of this domain to the X-ray structure for the different cluster representatives, whose values are seen in Table 3.1. Whereas the RMSD is nearly the same for all cluster representatives of the simulation with TS (ΔRMSD of 0.5 Å), the RMSD of the first three cluster representatives without TS *each* differ by ~1 Å (ΔRMSD of 2.0 Å).

Table 3.1.: Root-mean-square deviation (RMSD) in Å of the N-terminal domain's backbone atoms for the PC analysis cluster representatives, using the X-ray [32] conformation as a reference.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Simulation without TS	3.11	4.00	5.10	5.05
Simulation with TS	4.15	4.18	4.62	-

Global conformational changes of the L11·23S macromolecule were further explored by determining its solvent-accessible surface area, shown in Figure 3.20. For both simulations, there is a steep decrease in the SASA during the first part of the simulation, deviating from the value determined for

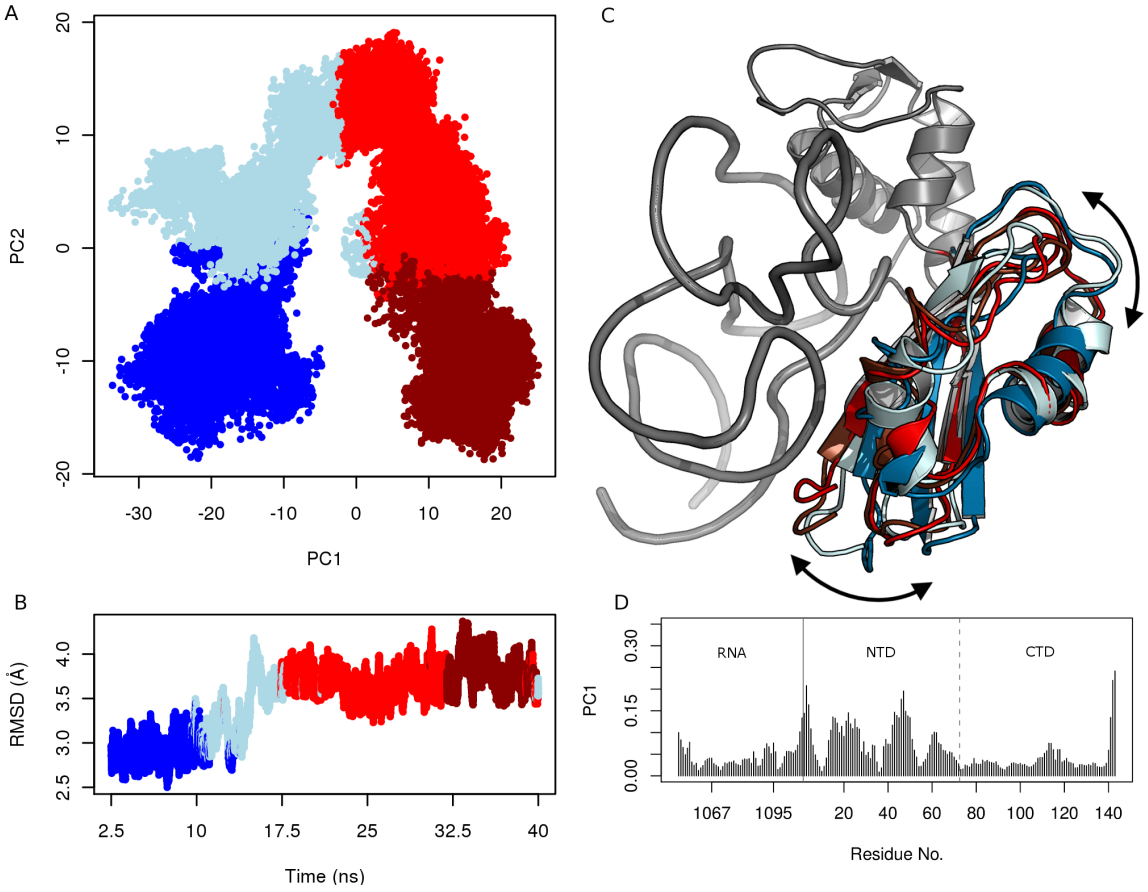


Figure 3.17.: Principal component analysis for the simulation without TS. A) Clustering in supspace spanned by first three PCs. Clusters one, two, three and four are colored as blue, light-blue, red and dark-red. B) RMSD plot color-coded according to clusters. C) Cluster representatives of the N-terminal domain, color-coded according to clusters. RNA and C-terminal domain are the ones from cluster representative three. The motion observed along the first PC is indicated by the arrows. D) Residue-wise loadings (contributions) in Å to the first principal component.

the crystal conformation (16,516 Å^2). The large SASA difference of almost 1,000 Å^2 between the crystal structure and its energy-minimized conformation (start of the simulations) can be likely attributed to 1) the medium resolution of the 3CF5 X-ray structure, 2) the structure's truncation to obtain only the L11·23S subunit model, and 3) the change of the hydrogen atoms' positions added to the crystal coordinates during minimization.

Considering an equal time frame for both trajectories (7–40 ns), the SASA for the simulation with TS stays relatively constant at 14,269±283 Å^2. The conformations without TS have a similar average SASA (14,083±426 Å^2), but sample a wider range as indicated by the higher standard deviation. Further inspection of the SASA computed for the simulation without TS reveals that the surface area corresponds well to the PC analysis clustering. The largest area (15,445 Å^2, L11 rendered in blue) is

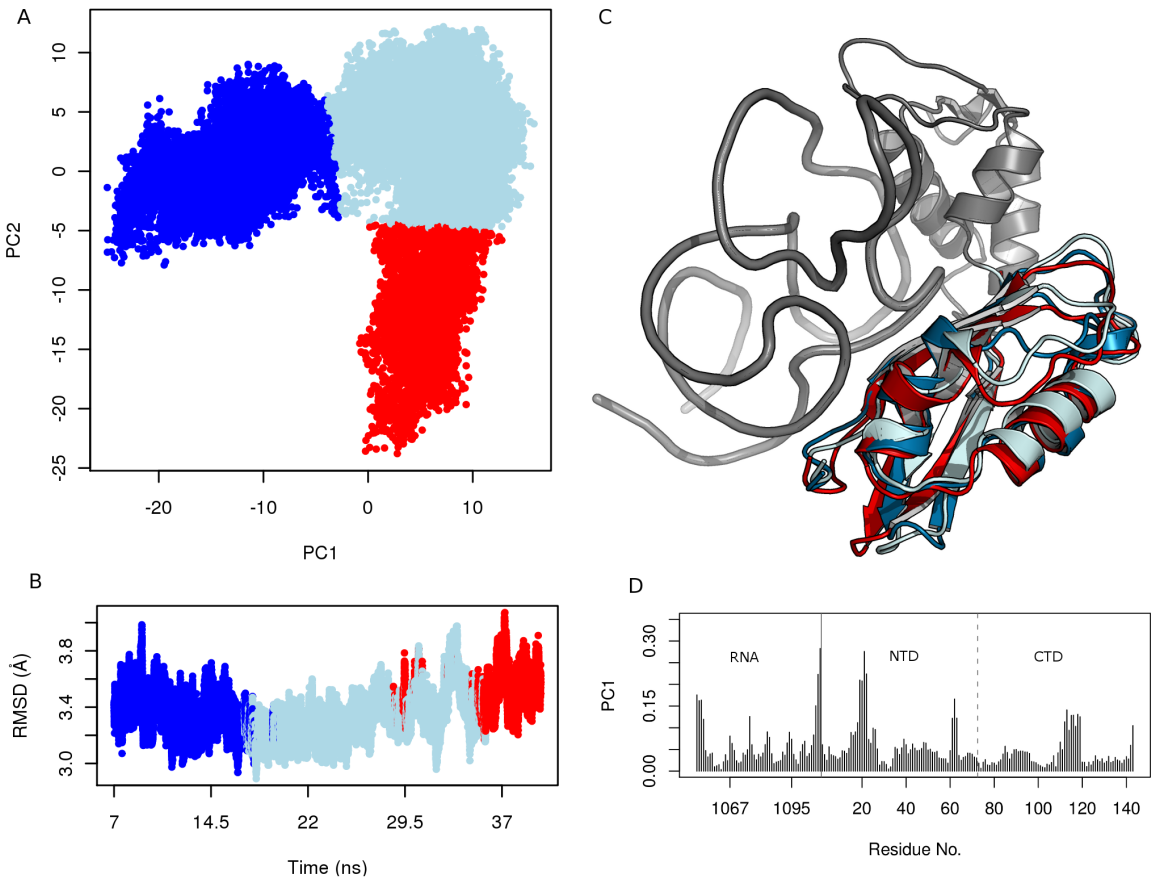


Figure 3.18.: Principal component analysis for the simulation with TS. A) Clustering in supspace spanned by first three PCs. Clusters one, two and three are colored as blue, light-blue and red. B) RMSD plot color-coded according to clusters. C) Cluster representatives of the N-terminal domain, color-coded according to clusters. RNA and C-terminal domain are the ones from cluster representative two. D) Residue-wise loadings (contributions) in Å to the first principal component.

seen in cluster one, the smallest (12,775 Å^2, L11 rendered in red) occurs in cluster three. The area difference is mainly caused by the conformational changes of the N-terminal domain. The SASA of the representative average structure[1] of the simulation with TS lies between the other two (14,290 Å^2, L11 rendered in gray).

3.3.3. Residue Fluctuations

The dynamic influence that TS has on the L11·23S complex can be investigated by computing each residue's Root-mean-square fluctuation. Due to the observations in the PC analysis, the obtained clusters were analyzed separately. Hereafter, four equal-length time frames for the simulation without

[1] The representative average structure of the simulation with TS was determined by selecting the structural snapshot most similar to the average structure of the last 33 ns of the simulation.

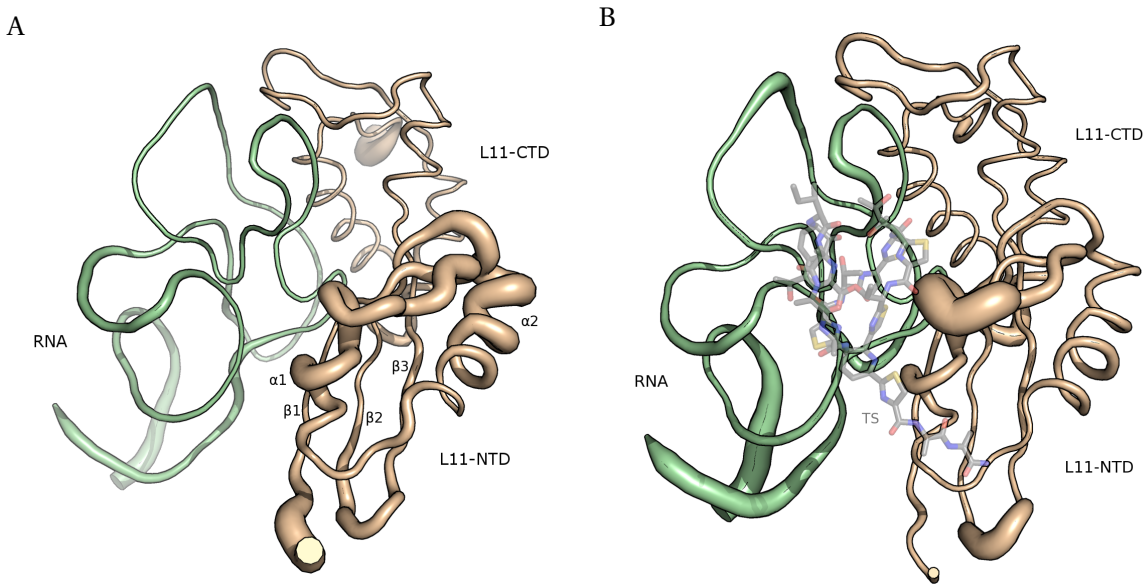


Figure 3.19.: Residue contributions to the first PC is shown by thickness amount as mapped onto the 3CF5 coordinates. The RNA is colored green, the L11 protein is colored wheat. A) Simulation without TS. B) Simulation with TS.

TS (2.5–10 ns, 10–17.5 ns, 17.5–25 ns, and 32.5–40 ns) and three time frames for the simulation with TS (7–16.5 ns, 16.5–25.5 ns, and 25.5–35 ns) will be considered.

When comparing the RMSF per residue between the two simulations (Fig. 3.21), the overall pattern is similar but differences are visible at specific and important positions. The fluctuations within the RNA and C-terminal domain are nearly identical between both simulations with one exception. Within the C-terminal domain, the loop (residues 112–122) connecting the α4-α5 helices shows a consistently high fluctuation when TS is present (Fig. 3.21B), whereas in the simulation without TS this loop's fluctuation progressively decreases with time (Fig. 3.21A). However, the high fluctuation of this loop could not be reproduced in the second simulation with TS (Fig. C.4B, appendix). The most notable difference in fluctuations between the two simulations occurs in the N-terminal domain, where TS's presence stabilizes its dynamics. Particularly noteworthy differences include the N-terminus (residues 1-6), α1, α2, and the turn connecting these two helices (residues 14–48). These differences are most pronounced early in the simulations (blue curves), but are still visible towards the end (red curves) when the fluctuations are drastically reduced in the simulation without TS. Exceptions to the general lower fluctuations observed in the simulation with TS are the consistently high fluctuations around residue 62 in the N-terminal domain and a fluctuation increase of the A1067 base, plus the already mentioned loop of the C-terminal domain. Interestingly, the RMSF for the loop (residues 39–49) connecting α2-β2 elements is ~2.0 Å at the beginning of the simulation without TS, then peaks at ~3.8 Å during the middle part, and reduces again to ~2.0 Å during the end of the trajectory.

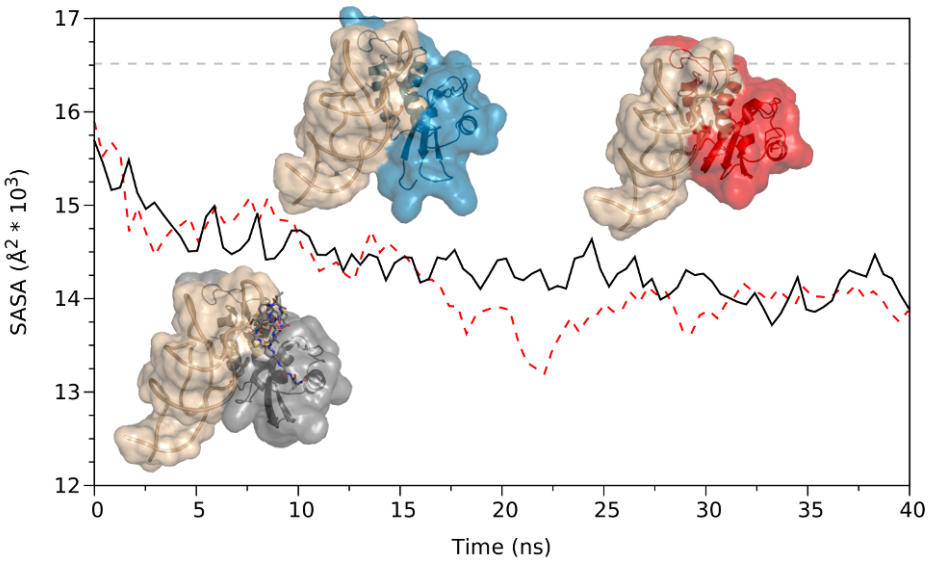


Figure 3.20.: Solvent-accessible surface area (SASA) over simulation time. Black line: simulation with TS; red dotted line: simulation without TS; gray line: SASA of X-ray structure. Data points smoothed using Bezier curves. The snapshots with the largest (blue) and the smallest SASA (red) are displayed for the simulation without TS. For the simulation with TS, the representative snapshot from 7 to 40 ns is shown (lower left corner and TS has been removed for graphical clarity, gray).

3.3.4. RNA Conformation

Overall, the RNA conformation remained fairly compact and immobile throughout both simulations, indicating a stable fold. With the exception of a few nucleic acids, all maintained an *anti* base conformation and a C3'-endo puckering was the preferred puckering sampled. The exceptions are listed in Table 3.2 and Table C.1 in the appendix, along with the percentage of base conformation and sugar puckering adopted during both simulations. Of these, the χ-angle of A1086 and A1088 both maintain their crystal conformation of *syn*, while their linking residue G1087 quickly switches from an *anti* to a stable *high-anti* conformation.

There are three loop apexes within the RNA that displayed different χ-angle conformations and dynamic motion that depend upon if TS is present or not. One apex is near the C-terminal domain and the loop connecting α4 and α5 elements. The residues U1082 and C1083 both maintain their crystallographic *anti* conformation throughout the simulation without TS, but sample a small percentage (12 % and 38 %) of *high-anti* conformation throughout the simulation when TS is present. Even larger changes are seen for the neighboring A1084 and A1085 residues, both of which are unpaired and have an *anti* conformation present in the crystal structure. In the simulation without TS, A1084 dynamically samples *high-anti* (34 %) and *syn* (62 %) conformations, while A1085 maintains the crystallographic *anti* conformation. However, in the simulation with TS, A1084 dynamically samples more *high-anti* (75 %) conformation at the expense of the *syn* (21 %) conformation. Likewise, A1085 adopts a stable *high-anti* conformation after 5 ns. Interestingly, the sugar puckering of A1085 also be-

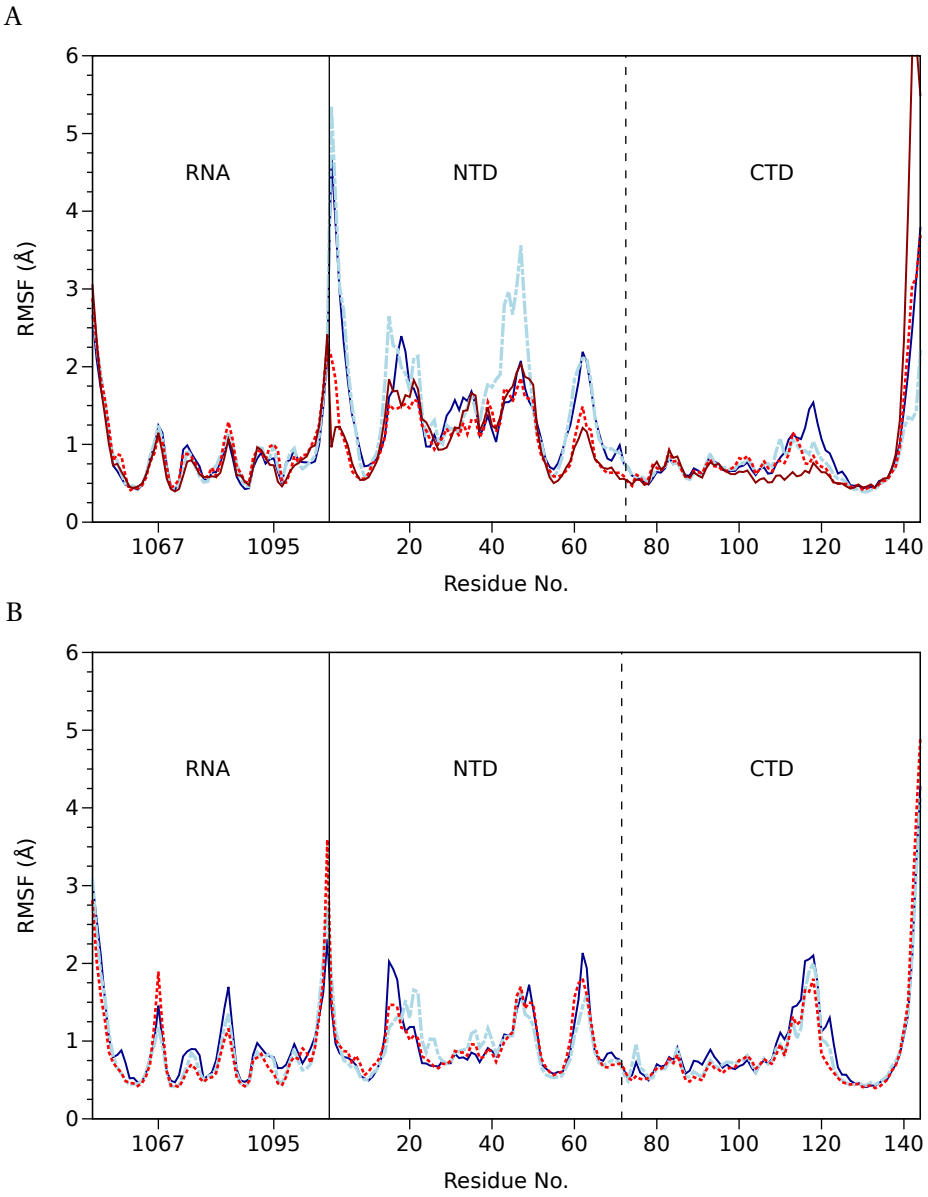


Figure 3.21.: Root-mean-square fluctuations (RMSF) of residue backbone atoms. A) Simulation without TS, evaluated at time frames 2.5–10 ns (blue), 10–17.5 ns (light-blue dashed), 17.5–25 ns (red dotted), and 32.5–40 ns (dark-red). B) Simulation with TS, evaluated at time frames 7–16.5 ns (blue), 16.5–25.5 ns (light-blue dashed), and 25.5–35 ns (red dotted).

Table 3.2.: The population percentage of χ-angle sampled during the simulations with and without thiostrepton bound. The residues listed are either involved in the binding of TS or showed interesting conformational sampling.

	Simulations (without TS | with TS)			X-ray
	anti	*high-anti*	*syn*	
A1067	4 | 70	3 | 2	93 | 28	anti
G1068	84 | 99	16 | 1	0 | 0	anti
U1082	100 | 88	0 | 12	0 | 0	anti
C1083	100 | 62	0 | 38	0 | 0	anti
A1084	4 | 4	34 | 75	62 | 21	anti
A1085	98 | 28	2 | 54	0 | 18	anti
A1086	2 | 0	0 | 0	98 | 100	syn
G1087	35 | 34	64 | 66	1 | 0	anti
A1088	2 | 0	0 | 0	98 | 100	syn
A1095	36 | 100	27 | 0	37 | 0	anti
A1096	67 | 100	33 | 0	0 | 0	anti
U1097	79 | 100	21 | 0	0 | 0	anti

comes more dynamic in the simulation with TS, sampling O4'-endo (44 %), C4'-exo (25 %), C3'-endo (16 %), and C1'-exo (12 %). This is the opposite behavior of A1084, which dynamically samples all ranges from C3'-endo to C2'-endo puckering in the simulation without TS, but preferentially adopts the C3'-endo puckering in the presence of TS.

Significant conformational changes were also observed at the apexes of the two helices that participate in TS binding, namely at bases A1067, G1068, A1095, and A1096. Both G1068 and A1096 predominately maintain an *anti* conformation during the simulation with TS, but adopt a small percentage of *high-anti* (16 % and 33 %) conformation during the simulation without TS. The largest changes are seen in A1067 and A1095, which are the two residues that interact directly with TS. The χ-angle of these two residues sampled during both simulations is shown in Figure 3.22. In the simulation with TS, A1067 maintains the crystallographic *anti* conformation for ~28 ns, and then samples the *syn* conformation for the remaining time, which corresponds to the higher fluctuation of this nucleotide at the simulation's end (cmp. Fig. 3.21). Likewise, A1095 maintains the crystallographic *anti* conformation throughout the simulation with TS. However, both residues behave differently in the simulation without TS. A1067 quickly adopts a dynamically stable *syn* conformation, while A1095 adopts both *anti* and *high-anti* conformation for ~24 ns and then adopts a *syn* conformation for the remaining ~16 ns. Note that in the second simulation without TS both nucleotides sample the same χ-angle conformations but with different lifetimes, with the exception that A1095 does not significantly sample any *anti* or *high-anti* conformation (Fig. C.5, appendix). For both residues a range of sugar puckers are sampled in the simulation without TS, while C3'-endo and C2'-exo are the predominate puckerings in the simulation with TS (Table C.1, appendix).

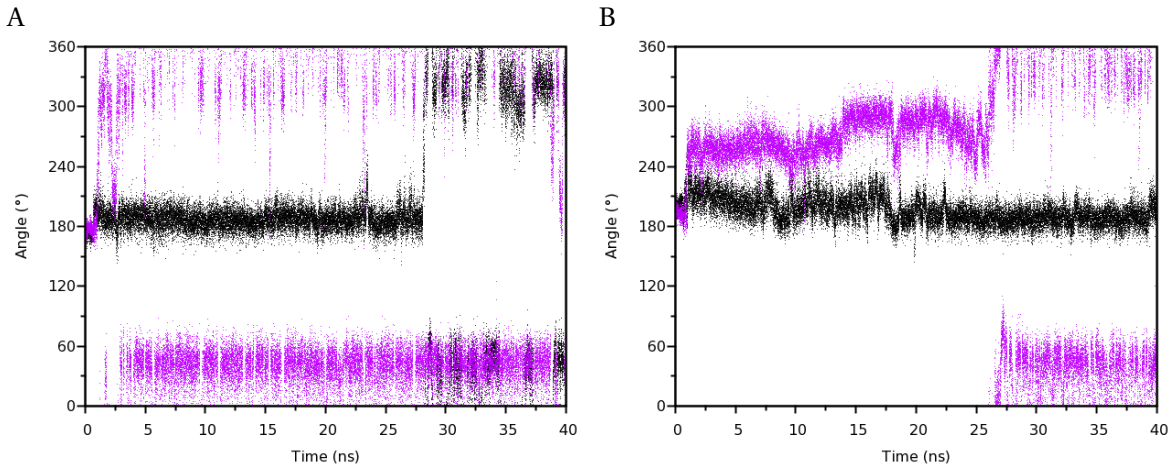


Figure 3.22.: χ-angle (O4'-C1'-N9-C4) for nucleic acids (A) A1067 and (B) A1095 as a function of simulation time. Black data points are from the simulation with TS, and purple data points are from the simulation without TS.

3.3.5. Dynamic Cross-Correlations and Protein Secondary Structure

Dynamic coupling between residues within the L11·23S complex was investigated in both simulations by computing covariance matrices [215] and visualized as dynamic cross-correlation maps (DCCMs). As done in the RMSF analysis, separate DCCMs were computed for each cluster, resulting in the maps shown in Figures 3.23 and 3.24. Each DCCM can be separated into three distinct areas, corresponding to RNA-RNA, RNA-protein, and protein-protein cross-correlation.

The percentages of the protein's secondary elements, as computed for each residue, are shown alongside each DCCM. In the simulation without TS, the protein's secondary structure appears to be very stable, with the exception of the N-terminal domain's α1 and α2-helices. These two secondary elements become more stable in the simulation with TS.

Protein·Protein coupling: In nearly all DCCMs shown, the secondary elements within the C-terminal domain are positively correlated (Figs. 3.23 and 3.24). Three minor exceptions are the negative correlation between the loops connecting α4-α5, α3-β4, and α4-β5 elements in Figures 3.23A and 3.24A. Another exception is the anti-correlation of the C-terminus with the domain's secondary structure elements in the last half of the simulation without TS (Figure 3.23C and D). Due to their infrequent appearance and small sizes, it is difficult to say if these are significant or an artifact of the sampling.

The motion within the N-terminal domain displays both positive and negative correlation when TS is not present, and almost exclusively positive-correlated motion when TS is present. Another general observation is that a coupled movement between the two protein domains exists almost exclusively as anti-correlated motion. The few small exception occur in the simulation when TS is not present (Figure 3.23A and B).

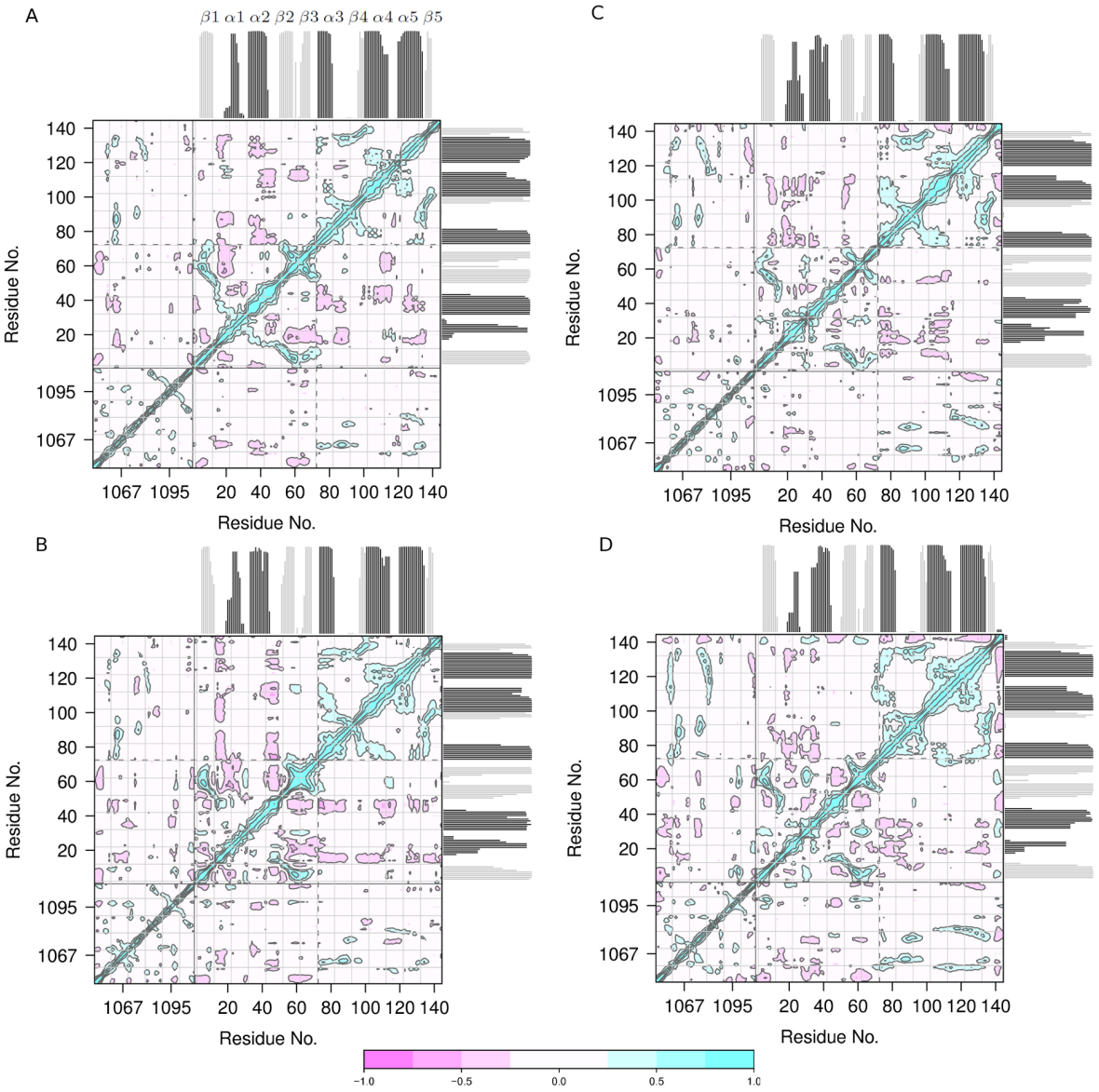


Figure 3.23.: Dynamic cross-correlation maps (DCCMs) for the simulation without TS. RNA and protein residues are divided within the plots using solid black lines, the two protein domains are further divided by dashed gray lines Cyan colors represent positive correlations, magenta anti-correlations (see color bar at the bottom). The percentage of protein secondary structure occupation during the simulation are displayed by light gray bars (strands) and in dark gray (helices). A) Analysis performed from 2.5 to 10 ns. B) Analysis performed from 10 to 17.5 ns. C) Analysis performed from 17.5 to 25 ns. D) Analysis performed from 32.5 to 40 ns.

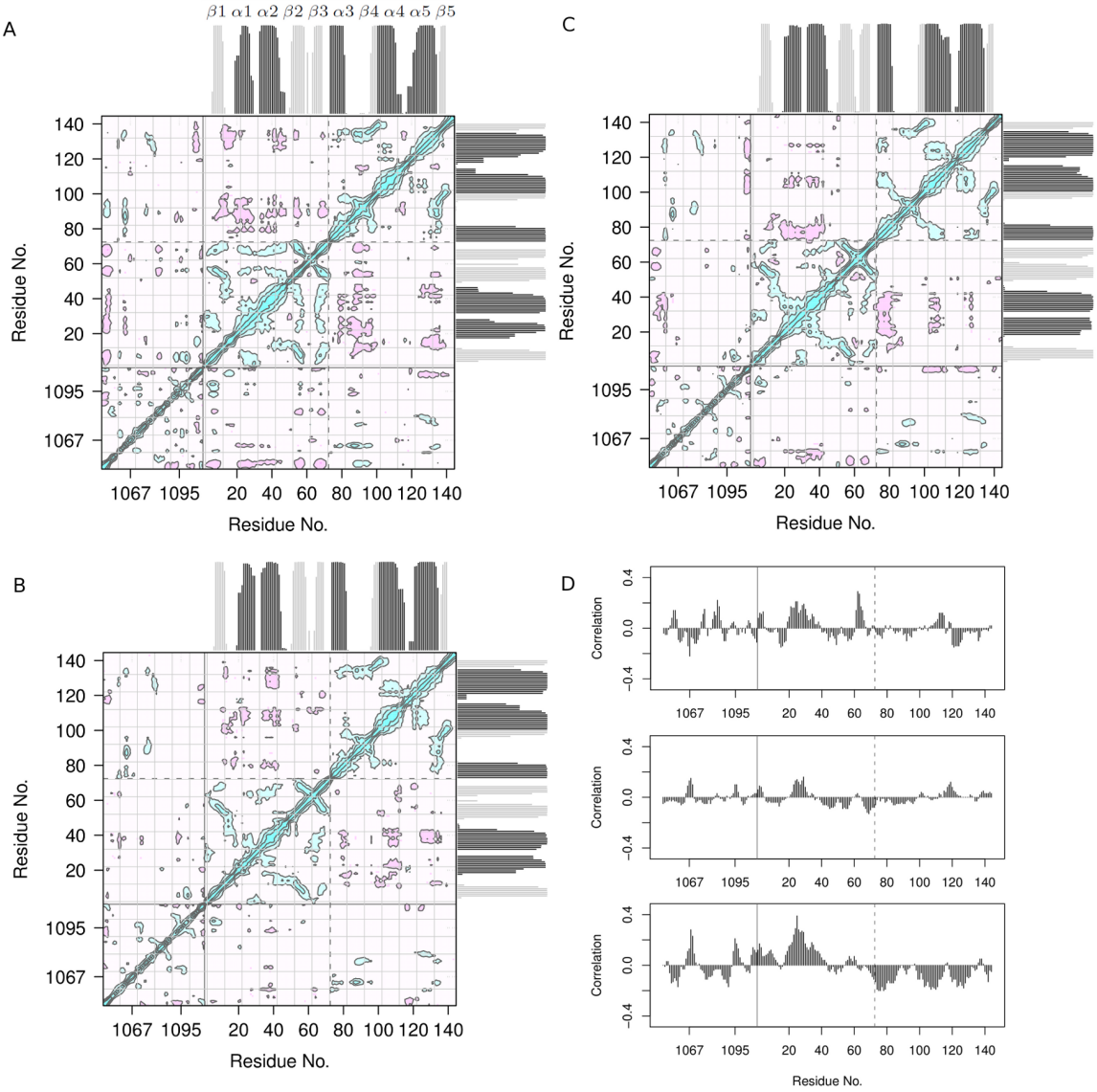


Figure 3.24.: Dynamic cross-correlation maps (DCCMs) for the simulation with TS. A) Analysis performed from 7 to 16.5 ns. B) Analysis performed from 16.5 to 25.5 ns. C) Analysis performed from 25.5 to 35 ns. D) Dynamic correlations of TS with 23S/L11 residues for the time frames analyzed in A, B and C, respectively.

Protein·RNA coupling: The largest correlations between the protein and RNA are found in the C-terminal loop connecting α3-β4 elements (residues 86–96) with the RNA residues that form the A1067 apex. This loop makes a significant number of salt bridges and hydrogen bonds to the RNA [30], and was shown to undergo an induced fit conformational change upon RNA binding as revealed by NMR studies [31]. Furthermore, this loop shows very little fluctuations (Figure 3.21) and its dynamics and conformation are unaffected by TS binding. In general, the C-terminal and RNA residues that interact through close-range non-bonded interactions are positively correlated, apart from a few anti-correlated residues associated with the protein and RNA termini. These negatively correlated residues associated with the RNA termini could be artifacts due to the truncation of the RNA during model building. Conversely, the N-terminal and RNA residues predominately have negative correlation, with the notable reoccurring exception of the positive correlation occurring between A1095 and the β1 structural element.

RNA·RNA coupling: Considering only the nucleic acid structure, positive correlation exists between RNA residues, with minor exceptions at the terminal residues where the RNA had been artificially truncated for model creation. Consistently occurring in both simulations are two small peaks arising from the interaction of residues near A1067 and A1095. These residues (A1070 and A1096) are part of the RNA helices that present A1067 and A1095 at their apexes.

L11·23S·thiostrepton coupling: Figure 3.24D contains the correlated motions of TS with the L11·23S biomacromolecule for the considered time frames. The correlation coefficients between L11·23S's residues and TS (Figure 3.24D) range between -0.2 and +0.4, and are thus significantly lower than those observed within the L11·23S biomacromolecule. The positive correlation of the N-terminal α1-helix (residues 25–35) is present in all time frames at varying magnitudes, and is probably due to its close spatial proximity to TS. The C-terminal domain is largely uncorrelated during the first and middle part of the simulation, but becomes anti-correlated during the last part of the simulation. The correlation between TS and the RNA fluctuates between the time frames, and shows little consistency. However, what is notable is that both A1067 and A1095 become positively correlated with TS at the end of the simulation, while the rest of the RNA becomes anti-correlated (apart from the termini).

3.3.6. Water and Ion Occupancy

Previous experimental studies clearly demonstrated the importance of a minimal ion concentration for the 23S rRNA fragment to fold correctly, leading to the hypothesis that the presence of ions is crucial for the stability of the RNA [31, 227, 228]. To determine if preferred regions of water and ion occurence existed – especially at the TS binding interface, their densities over simulation time were computed. Although the sodium ions moved throughout the periodic model during the simulations, strongly preferred occupancies were found within the RNA's interior. The same is true for the water

molecules whose densities were also highest within the RNA. Figure 3.25 shows the L11·23S·TS complex' average structure along with the sodium ion density at different isovalues and the most probable water oxgygen positions. No significant differences were found between the simulations with or without TS.

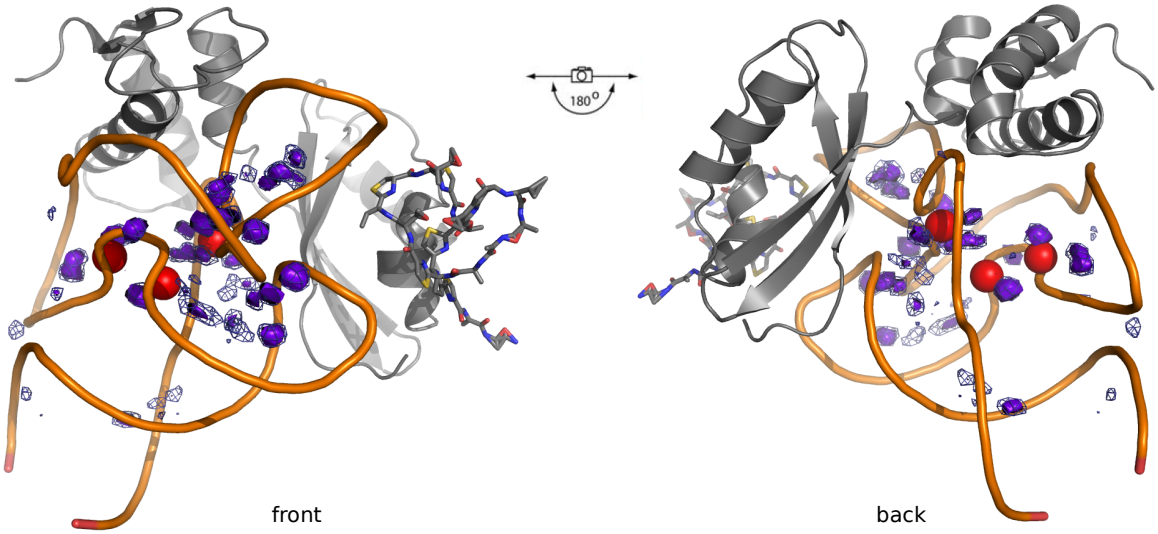


Figure 3.25.: Front and back view of average structure of last 33 ns simulation with TS. Sodium ion density grid is shown at isovalues representing 5 % (purple, solid) and 2.5 % (blue, mesh) of maximum density. The most probable (80 % of maximum density) water oxygen positions are represented as red spheres.

3.4. Discussion

MD simulations were initiated to provide atomic resolution information on how thiostrepton influences the L11·23S biomacromolecule. The ligand-bound state extracted from the X-ray crystal structure of the TS-bound 50S subunit of *D. radiodurans* [32] was chosen as the starting point. Other experimental data of the L11·S23 complex without a bound ligand [30], of the EF-G bound ribosome [28] or from NMR studies [31] have been reported, but were considered too remote from the ligand-bound state and/or to potentially display alternative conformations.

An all-atom MD study of the ribosomal L11·23S subunit, based on a crystal structure of the isolated *T. maritima* complex (PDB code 1MMS [30]), but without a ligand bound, has been reported before [206]. While the 1MMS and 3CF5 crystal structures originate from different organisms and were determined in different settings (RNA·protein complex vs. 50S subunit) and the 3CF5 crystal structure resolved ten additional amino acids in the protein termini, both similarities and characteristic differences are found between these data and the simulation with TS bound. Foremost, the overall RMSF patterns are comparable, with low and high fluctuations occurring within the C- and N-terminal domains. Especially good RMSF agreement is seen within the RNA, where five characteristic

peaks centered at nucleotides A1067, A1074, A1084, A1095, and C1110 were observed with nearly identical intensities. The L11·23S RMSD without TS shows similar trajectory values (2.5–3.1 Å) to the previous simulation (2-3 Å) for the initial ~15 ns, which is the extent the previous simulation was conducted. However, this agreement ended after the simulation underwent a dynamic event at 14.5 ns (see Figure 3.17B). Although starting from a different crystal structure and organism, the simulated conformations observed here show greater similarity to the 1MMS structure than to 3CF5 (see Table 3.3), both when TS is present (2.0–2.3 Å RMSD) and when it is absent (1.8–2.1 Å RMSD). This observation could be accounted for by the absence of the fully assembled 50S subunit in the 1MMS crystal structure and in the simulations here. Interestingly, both in the present and in Frank and coworker's simulation without TS [206], the A1067 residue predominately populated the *syn* conformation instead of the crystallographic *anti* conformation.

In the simulation without TS, two distinct and dynamically stable biomacromolecular conformations of the L11·23S complex were observed. These are represented by cluster one (dark blue) and clusters three and four (light and dark red), and separated by the first PC shown in Figure 3.17A. The conformation sampled by cluster one is structurally closest to the 3CF5 crystal structure, but is also relatively short lived. During the simulation this particular conformation reproducibly undergoes a dynamic change where it adopts another conformation (cluster three). This transition was verified in the second 40 ns simulation on the model without TS, which resulted in a very similar trajectory. Unexpectedly, the conformations sampled in cluster three are very similar to the conformations seen in the simulation with TS. This resemblance was surprising, since the TS-bound crystal structure with the ligand removed was used as the simulation's input. Of course, computational simulation and experimental structure determination feature distinctive differences. The 3CF5 X-ray structure itself was determined using preformed crystals of the complete 50S subunit soaked with a solution of thiopeptide ligand [32]. Given the apparent adaptability of the GAR, effects resulting from the crystallization of the apo form first, crystal packing, or from cryocooling could potentially lead to different conformations of the NTD relative to the CTD and RNA. On the other hand, the MD simulation is based on the crystal structure data, but cannot properly account for the low temperature of the electron density determination (100 K) and – most importantly – for the presence of the fully assembled 50S subunit.

Table 3.3.: Root-mean-square deviation (RMSD, in Å) between simulation cluster representatives and available crystal structures 3CF5 and 1MMS[a]. Backbone RMSD | NTD RMSD (with fitting to complete backbone).

		Simulation without TS				Simulation with TS		
	1MMS	1	2	3	4	1	2	3
3CF5	2.0 | 2.3	2.3 | 2.7	2.5 | 3.3	2.9 | 4.2	3.0 | 4.5	2.7 | 3.2	2.7 | 3.3	3.0 | 4.0
1MMS	-	1.8 | 2.4	2.0 | 3.1	2.1 | 3.0	2.1 | 3.1	2.0 | 2.2	2.0 | 2.2	2.3 | 2.9

[a] Only residues common to all structures were considered, i.e. nucleotides 1051 to 1108 for 23S rRNA, and residues 7 to 139 for L11.

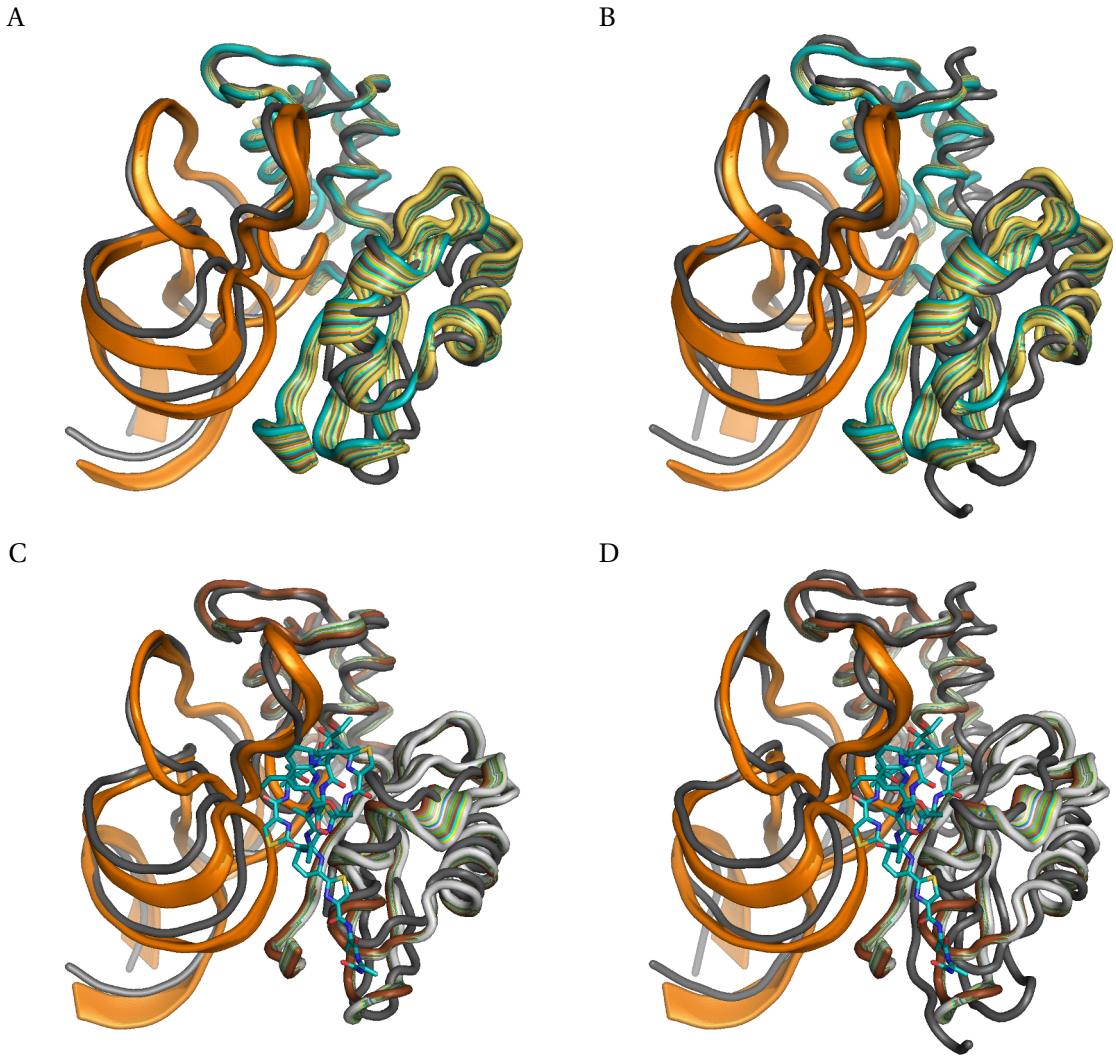


Figure 3.26.: Cα and phosphorus atoms trace of first principal component's interpolated trajectory. A) Simulation without TS and 1MMS (gray). B) Simulation without TS and 3CF5 (gray). C) Simulation with TS and 1MMS (gray). TS of the representative average structure is shown in cyan. D) Simulation with TS and 3CF5 (gray).

An overlay of the macromolecular conformations described by the first principal component with the available crystal structures can be seen in Figure 3.26, showing that the changes primarily occur within the protein's N-terminal domain and that the divergence of the structures in the apo form is higher than when TS is bound. The overall flexibility of the L11·23S complex, minus the ligand, can also be seen in the solvent accessible surface shown in Figure 3.20. Similar protein flexibility within L11·23S macromolecule has been previously found in simulations and structural studies, accompanied by the suggestion that the L11 protein can adopt "bound" and "unbound" conformations [31, 32, 206, 229]. The results suggest that a "TS-bound"-type conformation of the L11

protein inherently exists and does not require the ligand to induce its conformation into existence. Indicated by the simulations, but not fully proven due to inherently insufficient sampling, TS likely stabilizes this bound conformation thermodynamically, and might channel a partitioning of states into one distinct, "locked" conformation. These data are in line with the CD-spectroscopy results, which indicate the formation of a state with slightly less α-helical/A-form character in the presence of TS.

In the simulation with TS, the distribution of conformational snapshots in the PC subspace supported three clusters. However, these clusters are very similar to one another, as shown in the color-coded RMSD plot and the overlay of their representatives structures (Figure 3.18B and C). Most of the differences reside in the loop structure connecting α1 to α2. With regard to the binding cleft characterization, one macromolecular conformation is essentially sampled throughout the simulation. This conformation is structurally similar to but does not exactly match the "TS-bound"-like conformations sampled in the simulation without the ligand present, as suggested by the RMSD analysis.

Considering both simulations, three possible macromolecular conformational states can be seen within the 40 ns time-frame – a metastable "TS-bound"-like conformation, a TS-bound conformation, and a ligand-unbound conformation. It is very likely that additional conformations exist given the documented adaptability of the GAR, but to achieve statistically significant sampling of such large systems using all-atom MD simulations remains prohibited by current methods and computational means. Nevertheless, taking together the results from the PC analysis, clustering, SASA, and RMSD data, it is fair to conclude that the N-terminal domain is conformationally dynamic and is structurally sensitive to the presence of a bound ligand.

The dynamic effect that TS binding exerts on the system's separate subdomains (i.e. RNA, C-terminal, and N-terminal domains) was investigated by computing residue-wise RMSF and dynamic cross-correlation maps (DCCMs). In both simulations, the RNA residues had overall low fluctuations, with a few small prominent peaks (Figure 3.21). Likewise, the C-terminal domain displays low fluctuation apart from a prominent peak occurring in the loop (residues 112–122) connecting the α4-α5 helices, which decreases as the simulation progresses. For the simulation without TS, the N-terminal domain shows the greatest fluctuations of all the subdomains. Not unexpectedly, the highest fluctuations were observed during the dynamic event that takes place around 14.5 ns, which was also captured in clustering the PC space (light blue cluster 2 in Figure 3.17).

The binding of thiostrepton has a significant impact on the dynamics of the protein domains. The overall RMSF within the N-terminal domain is reduced, while leaving three loop structures with high fluctuations relative to the other amino acids. Surprisingly, there is an increase in fluctuations around the loop (residues 112–122) connecting the C-terminal's α4-α5 helices. This loop makes direct contacts to the RNA, and to the N-terminal domain of the ribosomal protein L10 within the assembled 50S subunit [27, 28]. Hence, in the presence of L10 this conformational transmission may become more pronounced and might productively contribute to a hinge-like motion of the GAR,

potentially coupling EF-G and TS binding to the L7/L12 stalk. Based on the observation that this loop's fluctuation decreases as the macromolecular conformation moves away from its "TS-bound" structure during the simulation without TS, it is likely that TS stabilizes a macromolecular protein conformation that results in the increased dynamic motion within this loop. Regarding the rRNA domain, TS causes an increase in the fluctuations in the A1067 residue, while showing overall little influence on the rest of the structure.

Determining the cross-correlation coefficients between the residue motions provided understanding into the collective motion within the L11·23S complex, revealing the dynamic relationships between residues, secondary structure elements, and domains. Without a bound ligand, the motions between the C- and N-terminal domains are predominately anti-correlated. Within the C-terminal domain, positive correlation dominates the motion, and the same is true within the RNA structure. The motion between the RNA and C-terminal domain are predominately positively correlated, while between the RNA and N-terminal domain the motion is nearly all anti-correlated. However, a reproducible complex coupling of both negatively and positively correlated motions characterizes the N-terminal domain when TS is absent.

In the presence of TS, the overall correlated motion of the L11·23S macromolecule is reduced (Figure 3.24). The concerted motion between the N- and C-terminal domains loose their weak positive correlation, and become entirely anti-correlated. There is also little effect on the correlated motion within the RNA. However, the most significant change occurs in the correlated motion within each protein domain where they become entirely positively correlated, which is especially pronounced within the N-terminal domain. The correlated motion between each of the protein domains with the RNA is also noticeably reduced. Furthermore, the motion between TS and A1067 and A1095 are slightly positively correlated (Figure 3.24D). Consistently throughout the simulation, TS's motion is positively correlated with the α1-helix, which itself shows correlation with various other secondary elements in both the C- and N-terminal domains.

On the local level TS does clearly effect the conformations of nearby residues, especially prominent is an altered A1067 and A1095 χ-angle preference and an increased α1-helix stability. The exposed nucleobase A1067 has been assigned multifunctional roles in tRNA and elongation factor recognition [28, 192, 206]. Both *syn* and *anti* conformations of its χ-angle were sampled during a simulation by Frank and coworkers in the absence of a ligand. The simulation without TS present conducted within this thesis confirms this property, but a stronger *syn* preference is found. In contrast to their finding, A1067's χ-angle fluctuations were observed throughout the trajectory (see Figure 3.22A), and no correlation between the χ-angle conformation and the N-terminal loop movements that connects the β2-β3 strands was found. Interestingly, in the simulation containing TS this χ-angle is strongly positioned in an *anti* conformation (187°±10°) for ~27 ns, and aligns well with the value observed by X-ray crystallography (174°). TS clearly stabilizes A1067's *anti* conformation, but does not entirely inhibit its nucleobase from adopting a *syn* conformation. This base-flipping event is preceded by a loss of the hydrogen bond formed between an amide group of TS's macrocycle A and N3 of A1067,

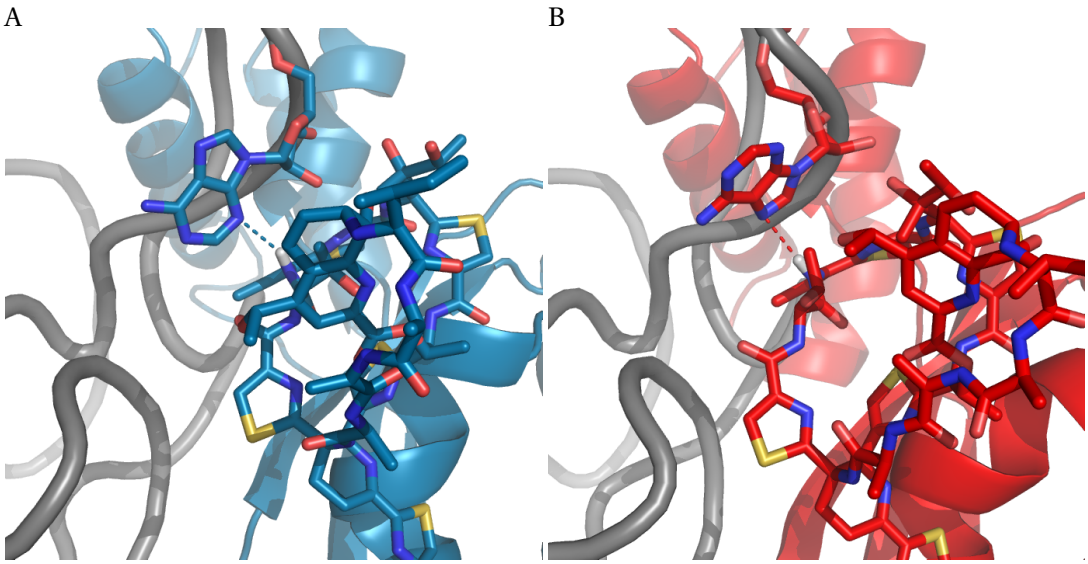


Figure 3.27.: Hydrogen bonds formed between TS and A1067's base. A) cluster representative 1, H-bond is formed with A1067:N3, A1067 is in *anti* conformation. B) cluster representative 3, H-bond is formed with A1067:N7, A1067 is in *syn* conformation.

whose presence probably fixed the base in the *anti* conformation (see Figure 3.27A). Instead, the formation of a hydrogen bond to N7 at the opposite side of the base is seen, stabilizing the *syn* conformation later in the simulation (Figure 3.27B). Furthermore, A1095 samples *anti*, *high-anti*, and *syn* conformations when TS is not present, but only the *anti* conformation when TS is bound. Due to the implicated importance of these residues to TS binding, two additional simulations were performed in the course of this work for validation – one on an adenosine monomer and the other on a trimer. In the adenosine monomer simulation, the *syn* preference was reproduced (see Fig. 3.28). However, the preference shifted to the *high-anti* conformation when flanked by two additional adenosines. This suggests that A1067's base is sterically unhindered by its surrounding environment in the absence of TS. To a lesser extent, TS appears to also influence the base orientations of U1082, C1083, A1084, and A1085, which are located near the protein's C-terminal domain. As a caveat, it cannot be completely ruled out that the preference for *high-anti* over *anti* conformations that was seen for some nucleobases is due to parameterization error within the Parm99SB and Parmbsc0 force fields [230, 231].

A computational study of allowed conformations in nucleic acid monomers indicated that the adenosine χ-angle preferentially adopts a *syn* conformation with a C2'-endo sugar pucker, and to a lesser extent when the sugar puckering is C3'-endo [232]. The *anti* conformation is populated when the sugar puckering is either C2'- or C3'-endo [233]. Consistent with this notion, the C2'-endo (50 %) and closely related C1'-exo (20 %) were found to be the predominant conformations for A1067's ribose in the simulation without ligand (Table C.1), and its base had a strong *syn* χ-angle preference (Table 3.2). However, when TS is bound, A1067's ribose preferentially adopts the C3'-endo (42 %) and

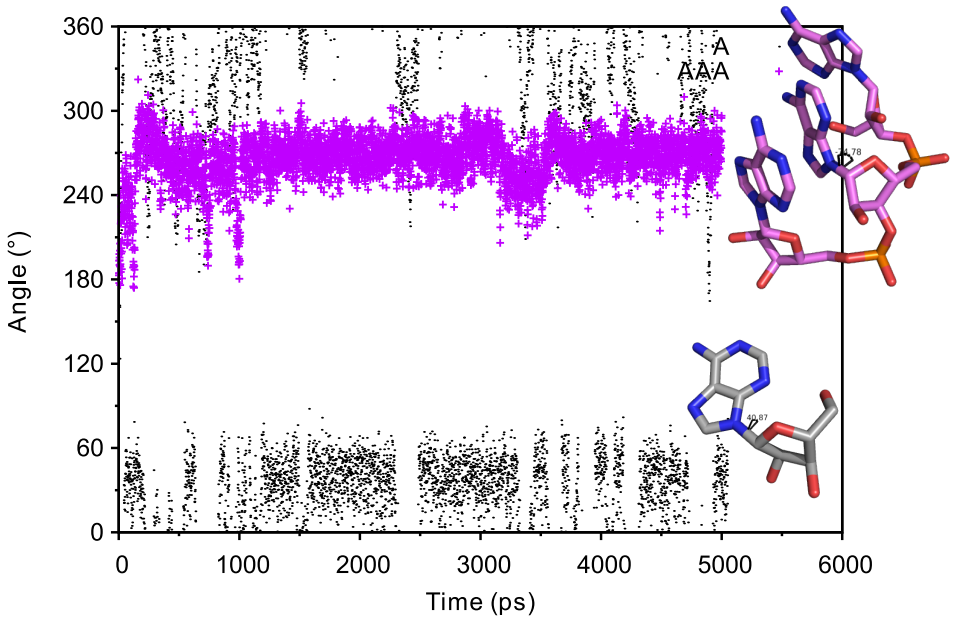


Figure 3.28.: χ-angle (O4'-C1'-N9-C4) for adenosine monomer (A, gray dots) and adonosine trimer (AAA, magenta crosses) as a function of simulation time. Structural representatives from the last simulation snapshot (5000th ps) are shown. Hydrogens are omitted for clarity.

C2'-exo (29 %) conformations, along with a high *anti* population of the χ-angle. The A1095 sugar populates several puckerings – C2'-exo (38 %), C1'-endo (17 %), C2'-endo (14 %), C1'-exo (12 %), and C3'-endo (8 %) – when TS is not present, coinciding with diverse sampling of its χ-angle. When TS is bound, the puckering predominately becomes C2'-exo (66 %) and C3'-endo (31 %), along with its *anti* χ-angle.

Within the ribosome a strong propensity for antibiotic binding sites to locate at non-paired bases, at *syn*-configured RNA bases, and at C2'-endo and C4'-exo sugar pucker sites was found [234]. A1067's conformational parameters conform to these general properties of ribosomal ligand binding sites. However, the binding site reorganizes when TS is bound. Since the actual binding event cannot be simulated for such a large system without inappropriate simplification, it remains unclear if this reorganization takes place during or after the ligand binding. This switch-like conformational reprogramming of A1067 and its surroundings might be an important part of the molecular activity of the thiopeptides. This property might also relate to the observed resistance pattern and specific mutations, which allow protein biosynthesis to proceed while TS is bound [191]. Furthermore, the natural mechanism of self-resistance features a crucial methylation of the 2'OH group at A1067. Beyond the breaking of apparent H-bond networks within the binding site [235], this modification could drive the conformational preferences of A1067 toward a state unreceptive for thiostrepton binding, potentially by indirectly modulating the sugar pucker and/or the *syn*/*anti* preference of the projected nucleobases [236, 237].

Preferred sodium ion and water molecule positions were identified within the rRNA, likely stabiliz-

ing its tertiary structure. This confirms the importance of ions and water molecules for the RNA fold, as indicated in previous experimental studies [31, 227, 228]. Importantly, no persistent ion or water molecules could be found within the TS binding site, making mediated contacts between TS and the 23S·L11 unlikely. This observation is crucial for the binding site's preparation in the subsequent docking studies (Chapter 4).

3.5. Conclusion and Implications for Drug Design

By embedding all components of the ternary L11·23 rRNA·TS complex in a fully solvated MD simulation, a computational model of the binding environment and the local dynamics of the GAR was created. Starting from an X-ray crystal structure of the ligand-bound state [32] an advanced and highly integrated target model was realized. It was found, that TS binding likely selects for a conformation of the GAR, which is also sampled in the ligand-free form. TS binding locally rigidifies the thiopeptide binding area on the L11 protein and characteristically reprograms the conformation of the apical nucleotide A1067, which mediates contacts to tRNA and incoming ribosomal factors. Otherwise, rRNA conformation is not detectably affected by TS binding. Removing TS from the L11·23S complex, immersing the ligand-free L11·23S complex in solvent, and heating it to room temperature resulted in a short-lived metastable conformation. A conformational transition during the simulation resulted in a state that was similar, but not identical, to the conformation adopted when TS is present. In the presence of TS, the overall dynamic motions and correlations thereof of the L11·23S complex became more stable. Principal component and dynamic cross correlation analyses of TS bound and unbound states located regions of fold change and of coupled motions. According to these data, conformational coupling to the RNA and the L10 protein could be mediated by the loop region connecting the α4 and α5 helices of L11.

These results align with CD spectroscopy data, which clearly showed a two-state transition of the RNA-protein complex upon ligand binding [238]. However, the ligand-free CD-spectrum of the L11·23S rRNA complex likely includes additional conformational states that have not been sampled by the MD trajectories. In the ribosomal X-ray crystallography work a shift in the distal portion of L11 (i.e. N-terminus and Glu62 turn) was noted when comparing the unbound and TS-bound form [32]. The unrestrained simulation performed in the context of this thesis suggests that this cleft can close in the absence of the ligand. In the average conformation of the simulation with and without TS (i.e. cluster three), the N-terminal's distal end is closer to the RNA than in the input crystal structure. Future research must clarify if these partially divergent aspects result from experimental conditions during crystal structure determination or from the absence of the full size ribosome in the NMR and MD simulation studies. In the simulation without TS, a small breathing motion is seen where the cleft closes and opens slightly.

Considering the energetics of ligand binding, a balance must exist between the binding enthalpy of the ligand and the entropy increase or decrease that occurs across the entire target and in the solvation

shell. On the one hand, the N-terminal domain gains an entropy decrease due to stabilization in the residue fluctuations, while on the other hand it has an entropy increase due to a reduction of overall correlated motions. The design and optimization of new inhibitors of the GAR will likely need to take both the binding enthalpy and the resulting entropy effects on the L11·23S macromolecule into account. The data presented herein should aid this process, by having identified crucial properties of the complex target, and by providing a validated model basis for ligand discovery strategies *in silico*.

4. Structure-Activity Relationships of Thiostrepton Derivatives

4.1. Introduction

In Chapter 3 the focus was on the target side, investigating the influence of thiostrepton binding on the dynamics of the 23S·L11 subunit. In the following chapters it will be investigated how molecular changes to the ligand structures influence binding to facilitate the optimization of future compounds.

Although structure-based drug design techniques, such as molecular docking, are now routinely used for protein targets, RNA has been recognized as a valuable drug target only recently [239]. Therefore, regarding RNA and DNA as a target, the development of docking and scoring tools has just begun. Most docking packages rely on scoring functions derived solely from protein data and it is unclear how this bias affects scoring involving RNA [75].

There are a number of studies indicating that protein-based tools with incorporated adaptations regarding the scoring function and receptor flexibilty can indeed be useful for structure-based drug design aimed at RNA [136, 240–246]. Moitessier et al. used a modified AutoDock protocol to dock aminoglycosides into conformational ensembles of the 16S ribosomal A-site [242]. Addressing the same target site, Barbault et al. used an RNA-adapted AutoDock 3 scoring function [247] followed by molecular dynamics simulations to establish structure-activity relationships of 16 newly derived aminoclycosides [243]. Baranger and colleages sequentially used DOCK and AutoDock for the identification of ligands binding to RNA tetraloops [244, 245]. Moreover, novel scoring functions have been developed that specifically address RNA, such as the empirical scoring function implemented within RiboDock [248] and the knowledge-based scoring function DrugScore^{RNA} [176]. However, these rare examples were derived from a limited data set.

In contrast to the ribosomal A-site mentioned above, which consist entirely of rRNA, the binding site of thiostrepton involves both rRNA and a protein [32, 249]. Protein–RNA interfaces have even been less well studied than binding sites consisting of only one species – protein or RNA. When targeting this kind of composite environments in a computational drug design project one is confronted with evaluating interactions of a ligand with two distinct molecular species. Moreover, the interface character makes this binding site also geometrically different from “classical” enzyme binding pockets as it has a larger surface and is much more solvent-exposed. It is, thus, completely uncertain how

reliable established protein scoring functions and the newer RNA scoring functions can predict molecular interactions between ligands and the L11·23S binding site.

Another challenge is posed by the nature of the considered thiopeptide ligands. Thiopeptides possess a considerable molecular weight (TS: 1664.83 g/mol) and their complex three-dimensional structures resemble protein folds [208]. Until now, ribosomal binding modes have only been determined for thiostrepton and nosiheptide [32]. Although *in vitro* and *in vivo* activities of semi-synthetic thiostrepton derivatives [250, 251] and synthetic fragments [42, 48] have been reported, their three-dimensional structures and exact binding modes at the GAR remain elusive.

Before reliable predictions of new thiopeptide compounds' binding modes can be made, a careful assessment of currently available docking programs and potentially necessary adjustments is needed. In the following section, it is presented how different existing docking and scoring packages, primarily developed for protein-ligand docking, performed in estimating the binding mode and affinity of different thiopeptides. Packages were included into the survey that have been shown to be able to successfully predict small molecule-protein and/or small molecule-RNA complexes and that were either free to academic institutions or were licensened by SCAI (FlexX). In particular, the docking tools FlexX (V3.1) [137], AutoDock (V4.2) [132], AutoDock Vina (V1.01) [133] and GlamDock (V0.4.9) [151] were tested. To address the scoring problem at a binding site composed of two molecular species, a combination of the existing scoring functions DrugScore [157] and DrugScore^{RNA} [176] was developed and tested as a re-scoring engine. Moreover, different approaches to predict three-dimensional structures of thiopeptide ligands were evaluated.

4.2. Evaluation of State-of-the-Art Scoring Functions at RNA-Protein Composite Environments

4.2.1. Materials & Methods

Molecule preparation The crystal structures from the nosiheptide-bound (PDB code 2ZJP [32]) and thiostrepton-bound (PDB code 3CF5 [32]) form of the 50S ribosomal subunit of *D. radiodurans* were downloaded from the Protein Data Bank [175]. Only relevant parts of the subunit were preserved, i.e. residues 1051–1108 (helices 43/44) of 23S rRNA and protein L11. The ligands' coordinates were saved as separate files. Hydrogens and Gasteiger partial atomic charges were added to the receptor by using MOE's [252] algorithm *protonate3d*. To prepare the ligands, hydogens were added, and correct atom types, bond types, and Gasteiger partial charges assigned in MOE. The same Gasteiger atomic charges were used in all docking programs, except for FlexX, which requires formal charges. Two different conformations of thiostrepton and nosiheptide with an increasing level of difficulty for the docking programs were created, i.e. 1) the crystal conformation but with translated (centered) coordinates and 2) an energy-minimized conformation (using the MMFF94x force field [92] to a gradient of 0.001 within MOE). All files were saved in mol2 format, which is recognized by all docking programs used.

FlexX FlexX uses an incremental construction algorithm that matches interaction types defined for the ligand and the receptor surface [137]. It uses an empirical scoring function based on the one defined by Böhm [253]. The receptor was prepared inside the FlexX GUI starting from the truncated PDB models. Hydrogens and formal charges were assigned using the FlexX built-in routines. The active site was defined by using a 9.5 Å cut-off radius around the reference ligand (TS and NS, respectively). The radius was chosen slightly larger than the default (6.5 Å) to accommodate the larger ligands used in this study. For the ligands, FlexX did not recognize the atom types correctly. Therefore, the automatic assignment was switched off and atom, bond types and hydrogens were used as found in the prepared mol2 files. A formal charge of zero was assigned for all ligands.

AutoDock A detailed description of AutoDock's docking algorithm and scoring function can be found in Section 2.2.1. Minor adjustments were made to the default parameters of AutoDock: The RMSD tolerance for clustering the resulting ligand conformations was set to 1.5 Å and 15 Lamarckian genetic algorithm (LGA) runs with $2.5 * 10^6$ energy evaluations were performed for each ligand. The grid box was centered on the bound TS conformation in the 3CF5 crystal structure with 70x70x70 grid points using the default grid spacing of 0.375 Å. The molecule input files (pdbqt) were prepared using the AutoDockTools scripts `prepare_receptor4.py` and `prepare_ligand4.py`, preserving the Gasteiger charges in the mol2 files calculated by MOE.

AutoDock Vina AutoDock Vina was recently released by the Scripps Research Institute. The developers claim a significant accuracy improvement of the binding mode predictions for small molecule–protein complexes, and an approximately two orders of magnitude speed-up in comparison with AutoDock 4 [133]. Vina uses an Iterated Local Search global optimizer similar to the one implemented in the docking program ICM [148] and an empirical scoring function inspired mainly by X-Score [149].

AutoDock's input pdbqt and grid files were used for docking with AutoDock Vina. The *exhaustiveness* parameter of the global search was increased from 8 (default) to 16, otherwise default parameters were used.

GlamDock GlamDock is based on a simple Monte Carlo sampling with subsequent minimization procedure and an empirical potential for scoring. Although GlamDock is not as well known as the other docking programs investigated, it was included in this evaluation as its *new high quality* scoring function has been reported to be very successful in reproducing protein–ligand and RNA–ligand complexes [254]. It was calibrated on a test set including partly RNA–small molecule complexes and contains an additional stacking term and an improved definition of lipophilic atoms to account for specific small molecule–nucleic acid interactions [254]. Two different GlamDock scoring scenarios were investigated: *high quality* (HQ) and *new high quality* (nHQ).

DrugScore DrugScore^{RNA} is the first knowledge-based scoring function to predict RNA–ligand interactions, available in the public domain. Based on the DrugScore formalism [157] (see Section 2.2.2), distance-dependent pair potentials were derived from 670 crystallographically determined nucleic acid–ligand and nucleic acid–protein complexes [176]. DrugScore^{RNA} showed superior predictive power when compared to DrugScore, DrugScore^{CSD}, and an RNA-adapted AutoDock 3 scoring function [241] on a test set of 15 RNA–ligand complexes [176].

To account for the specific thiopeptide–RNA and thiopeptide–protein interactions at the binding site, a combination of the scoring functions DrugScore and DrugScore^{RNA} was developed in the course of this study (referred to as DrugScore$^{Prot/RNA}$) and was used to re-score the docking poses produced by AutoDock and GlamDock. For comparison, DrugScore, and DrugScore^{RNA} alone, and a precurser[1] of the newest DrugScore version DrugScoreX (DSX) [177] were also used to re-score the complexes.

Evaluation measures To structurally evaluate the docking accuracy the RMSD (see Eq. 2.5 on page 19 for its definition) between the docking pose's atom coordinates and the ones in the "native" pose (i.e. the crystal or MD conformation) was determined. As the RMSD implementations slightly differ in the docking programs used, the external tool fconv (V0.98) [255] was used to calculate all RMSD values. Hydrogens were not included in the RMSD calculations.

For evaluating affinity predictions the predictive index (PI), introduced by Pearlman and Charifson [256], was used as an evaluation criterion. The PI measures how reliably a function can correctly choose the better binder for any pair of molecules. It is normalized to values from -1 (always wrong), 0 (random), to +1 (always correct). For a set of ligands i with experimentally determined binding energies E_i and corresponding predicted binding energies P_i, it is calculated as:

$$\mathrm{PI} = \frac{\sum_{j>i}\sum_i |E_j - E_i| C_{ij}}{\sum_{j>i}\sum_i |E_j - E_i|} \tag{4.1}$$

with

$$C_{ij} = \begin{cases} 1 & \text{if } [E_j - E_i] / [P_j - P_i] < 0 \\ -1 & \text{if } [E_j - E_i] / [P_j - P_i] > 0 \\ 0 & \text{if } [P_j - P_i] = 0 \end{cases} \tag{4.2}$$

4.2.2. Results

Re-docking accuracy

As an initial performance test, TS and NS were re-docked into their corressponding crystal structures (PDB codes 3CF5 and 2ZJP) using the docking tools FlexX, AutoDock, Vina, and GlamDock to evaluate how accurate the experimentally determined binding modes could be reproduced by the docking

[1] The DrugScoreX potentials were provided for testing by Gerd Neudert previous to the official release.

programs. These crystal structures have only a medium resolution (3.3–3.7 Å) and were obtained from soaking experiments [32] which implies uncertainty concerning the exact atomic positions of the protein's side chains, the nucleobases, and the ligands themselves. This low resolution renders these crystals imperfect target structures for docking experiments. Commonly, a re-docking RMSD below 2 Å is considered as a good result (see for example Ref. 161). Consequently, this value is often the upper limit of an X-ray structure's resolution to be included into test sets used in docking evaluation studies. During the MD simulation of the 23S·L11·TS complex based on 3CF5 a deviation up to ~3 Å from the input coordinates was observed after removing the restraints (see Chapter 3). The coordinates stayed approximately at this distance level for the rest of the simulation. Therefore, a representative average equilibrated 23S·L11 structure from the MD simulation, representing a refined complex structure in solution phase at 310 K, was used as an additional target for the docking studies (referred to as MD-refined structure hereafter). All ions and water molecules were removed from this structure as no preferred occupancy was detected at the TS binding interface (cmp. Chapter 3), rendering mediated contacts unlikely.

Table 4.1 contains the top-ranked (and minimal) RMSD values to the reference ligand structures found by the docking tools tested. As seen in Table 4.1, AutoDock and GlamDock clearly outperformed

Table 4.1.: Root-mean-square deviations[a] (RMSD, Å) of re-docking experiments for nosiheptide (NS, 2ZJP) and thiostrepton (TS, 3CF5).

Ligand	Input conf.[b]	FlexX	AutoDock	GlamDock	Vina
NS	c	4.30	4.87 (3.18)	1.48 (0.95)	27.90 (1.22)
NS	min	2.25 (1.92)	4.26 (3.09)	1.32 (0.87)	28.00 (2.11)
TS	c	13.27 (9.77)	2.42	3.78 (2.94)	31.97 (10.17)
TS	min	10.75 (10.55)	6.15 (4.06)	12.45 (6.18)	27.44 (14.00)
TS MD-refined	c	12.21 (7.46)	1.96 (1.64)	4.02 (1.00)	1.84 (1.84)
TS MD-refined	min	13.27 (9.26)	2.71 (1.89)	2.16 (1.56)	32.27 (2.63)

[a] Minimal RMSD found among the docking solutions is shown in brackets.
[b] Input conformation was either centered (c), or energy-minimized coordinates (min).

FlexX. FlexX was not able to place TS correctly into the binding pocket, as illustrated by the high RMSD values. The largest discrepancy between finding the correct pose (as expressed by the minimal RMSD in brackets) and ranking it correctly was seen for Vina. Vina is able to find a docking pose of NS close to the one observed in the crystal structure, but ranked a pose completely out of the binding pocket highest. For TS, however, its re-docking accuracy is similarly poor like FlexX's performance. GlamDock performed very well for NS (RMSD below 2 Å) and intermediate for TS in the crystal conformation but is sensitive to small conformational changes as sampled by the minimized TS conformation. Overall, AutoDock achieved the most stable results and the lowest RMSD for TS (2.42 Å). Thiostrepton's binding mode in the MD-refinded structure was consistently better reproduced than

in the crystal structure.

For nosiheptide, AutoDock's re-docking RMSD was higher than for the other tools, caused by a different binding mode where NS's tail was oriented towards L11's α1 helix instead of the protein-RNA cleft (Fig. 4.1). When the number of LGA runs was increased to 100, a solution was found at the top rank which was very close (RMSD 1.34 Å) to the crystal conformation (Fig. 4.1). Interestingly, the top solution found within the initial 15 runs was almost identical to the second-ranked cluster of AutoDock poses produced by the docking with 100 runs. This indicates that a second binding mode of NS might exist in addition to the one resolved in the crystal structure, a view also supported by biochemical studies that located NS's tail in this region of L11 [190].

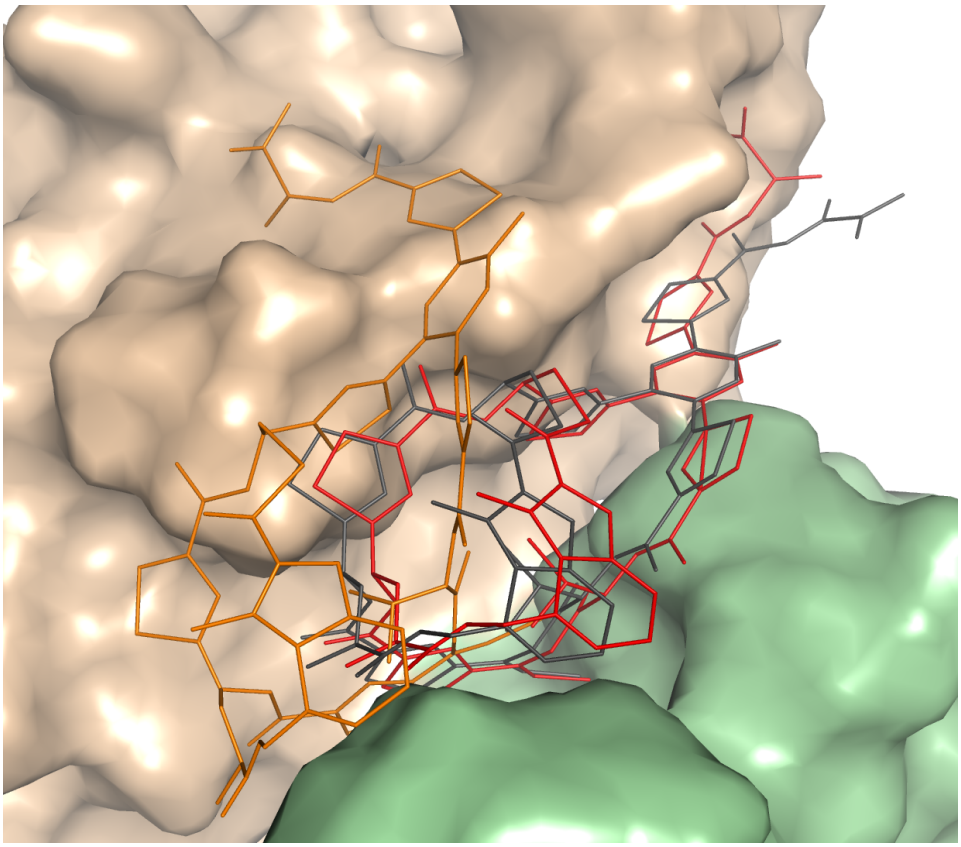


Figure 4.1.: AutoDock re-docking poses of nosiheptide to 2ZJP crystal structure. The L11 protein surface is colored wheat, the rRNA is colored green. NS binding mode as found in the crystal structural (gray) and top-ranked pose predicted by AutoDock when using 15 GA runs (orange) and when using 100 GA runs (red).

Due to FlexX's poor re-docking accuracy for TS and Vina's inapproriate ranking of docking poses, these tools were excluded from further evaluation at this step. Only AutoDock and GlamDock were tested for their affinity prediction of a larger set of thiopeptides.

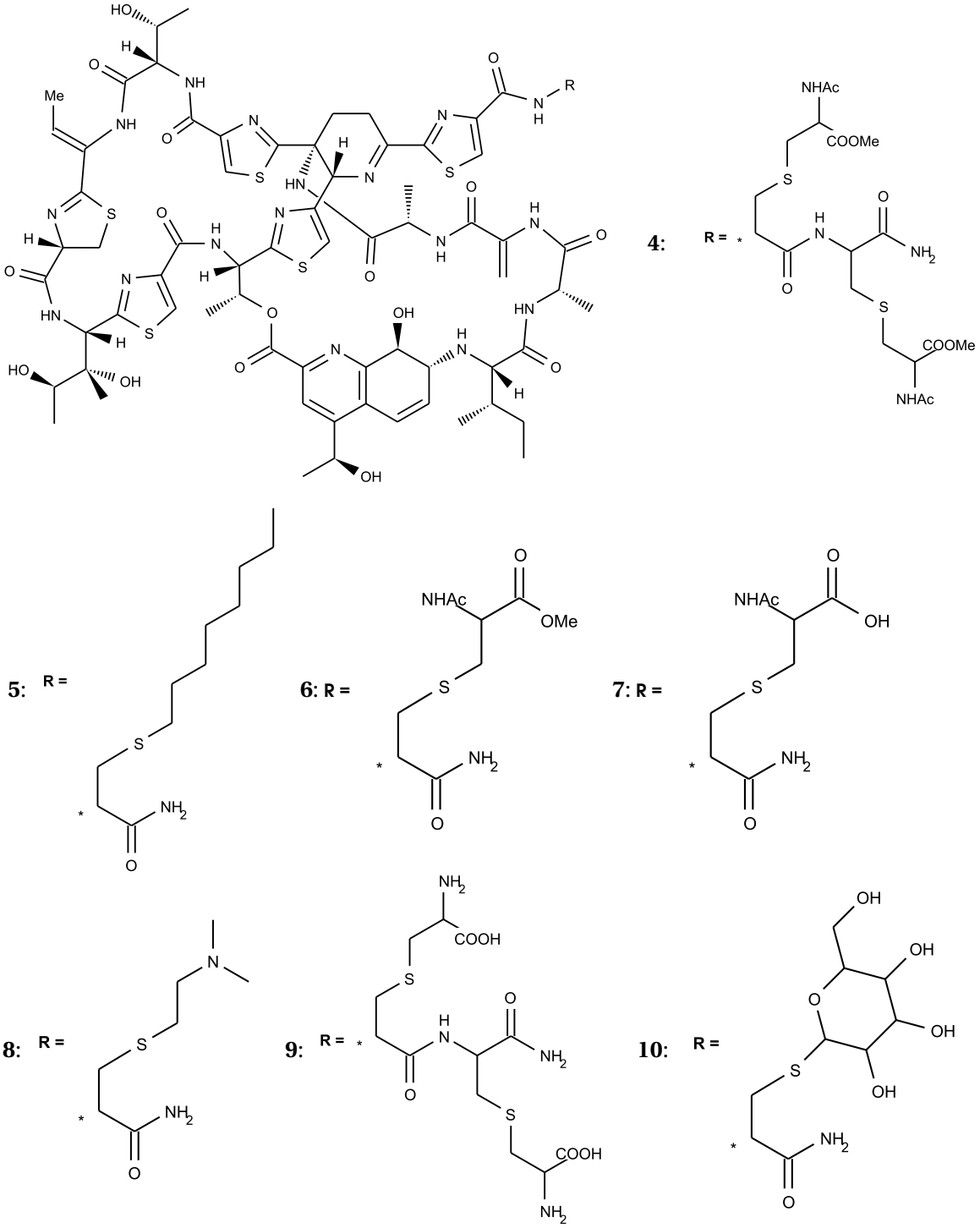


Figure 4.2.: Thiostrepton derivatives with modified tail.

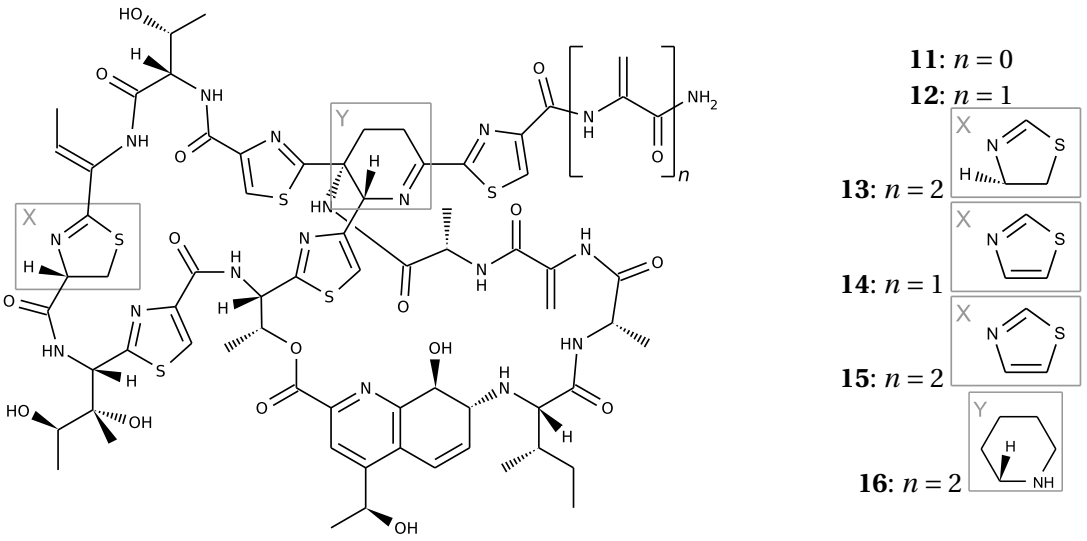


Figure 4.3.: Thiostrepton derivatives with modified rings and tail truncations.

Affinity prediction

The purpose of docking tools is twofold: they should 1) predict the correct binding modes of ligands, and 2) correctly predict their binding affinities (or at least rank different ligands correctly with respect to this property). Therefore, the docking tools that showed favorable re-docking accuracy in the initial test – AutoDock and GlamDock – were subsequently tested for their ranking ability. Experimentally available binding constants of TS, NS and 16 TS derivatives (Figs. 4.2–4.4, synthesis and testing by group of Dr. Hans-Dieter Arndt, particularly Dr. Sebastian Schoof and Dr. Sascha Baumann, Max Planck Institute for Molecular Physiology, Dortmund) were used as reference data. It was tested how well the scoring functions reproduced the experimentally determined binding constants by calculating their predictive indices (PIs). This test set was challenging since the ligands are structurally very similar and the affinities do not cover a large range (15 compounds 0.1–6.76 nM, **18**: 24.7 nM, **19**: 670 nM).

All docking tools investigated perform, in general, a flexible ligand docking. The ligand's translational and rotational degrees of freedom are searched for placing it in the binding site. Additionally, its torsional internal coordinates are searched to determine the ligand's bound conformation. However, this does not include macrocycles, as present in the TS scaffold. Thus, a macrocycle's backbone conformation is unchanged during docking; only side chains, if present, are modelled as being flexible. Therefore, a correct conformation of the two macrocycles in the input structure is critical for docking. An initial test to predict thiostrepton's three-dimensional structure was performed by using the 3D structure generation program CORINA [257], a common approach to predict 3D input conformations of small molecules in virtual screening applications. However, the resulting conformation deviated by almost 6 Å from the 3CF5 crystal conformation. Thus, a different route had to be found to generate reliable input conformations of the 16 TS derivatives.

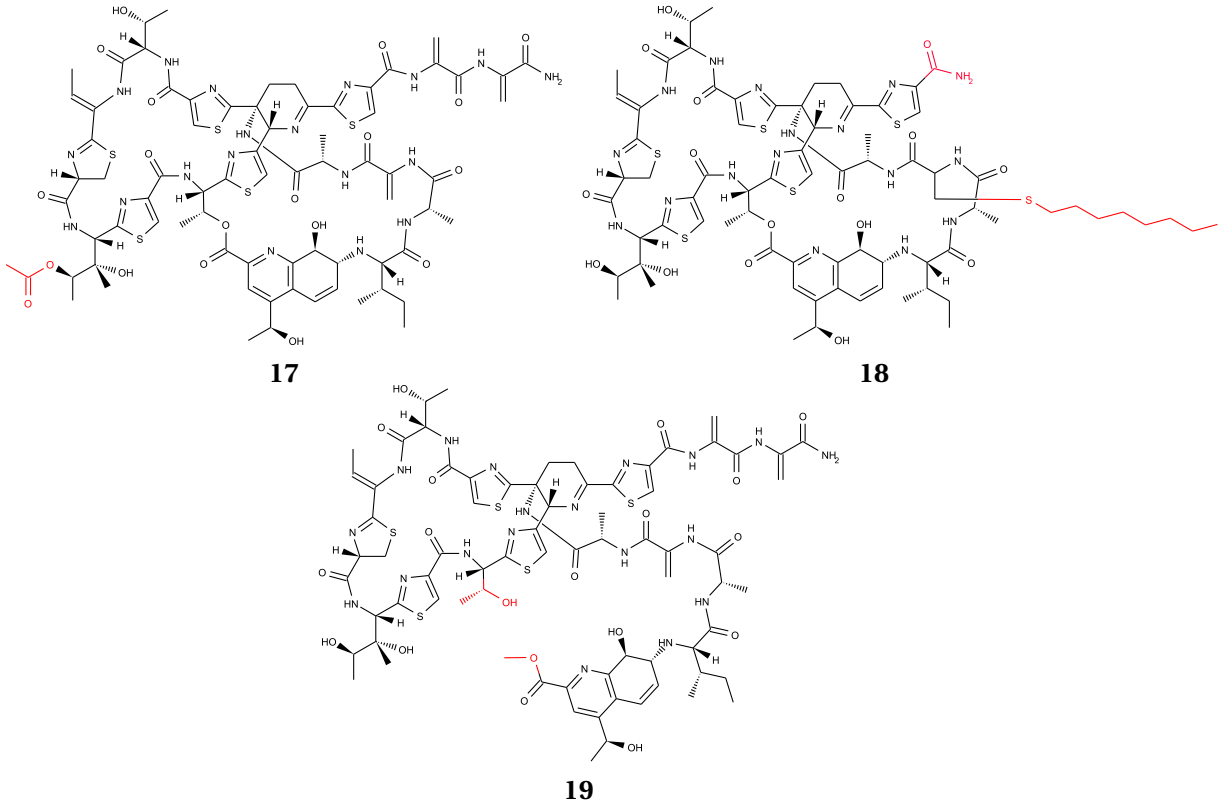


Figure 4.4.: Thiostrepton derivatives with modifications of side chains at macrocycle A or B.

To obtain three-dimensional structures of the TS derivatives, the 3D structure of TS was used as template, the respective changes to the molecules were made in MOE and the structures were subsequently minimized. All dockings were performed against the 3CF5 crystal structure and its MD-refined variant. Depending on the docking target, the TS's crystal or MD conformation was used as template for the derivatives in order to prevent a conformational bias. For GlamDock, two different scoring variants were investigated, HQ and nHQ (see Methods section). Additionally, the docking poses produced by AutoDock and GlamDock were re-evaluated (re-scored) with DrugScore, DrugScore^{RNA}, DSX, and the combined function DrugScore$^{Prot/RNA}$. As both AutoDock and GlamDock showed better performance when the MD-refined receptor structure was used, the DrugScore re-scoring was only performed for this target.

The predictive indices of AutoDock's and GlamDock's native scoring functions and a re-scoring with different DrugScore variants and the combined function DrugScore$^{Prot/RNA}$ for the MD-refined target are shown in Table 4.2. Except for the AutoDock score that ranked approximately half of the data set correctly with the MD-refined target structure (PI of 0.51), neither the GlamDock scoring functions nor any of the DrugScore re-scorings showed any predictive capabilities. Conversely, even negative PI values were often found, meaning that weaker binders were ranked higher than stronger ones. The new version of GlamDock's scoring function (nHQ) indeed improved the predictions but

Table 4.2.: Predictive indices (PIs) of docking scores and re-scoring of docking poses with different DrugScore variants.

	X-ray	**MD-refined**				
	Native Score		DrugScore	DrugScore^{RNA}	DSX	DrugScore$^{Prot/RNA}$
AutoDock	-0.29	**0.51**	-0.06	-0.19	-0.05	-0.22
GlamDock_HQ	-0.30	-0.28	-0.27	-0.52	n.d.	-0.50
GlamDock_nHQ	-0.46	0.07	-0.31	-0.36	-0.28	-0.43

only for the MD-refined target and in the way that it turned an incorrect prediction into a random one (i.e. -0.28 to 0.07, Table 4.2). An illustration of the improved fit can be seen in Fig. 4.5. However, the low PI values were often dominated by a few prominent outliers, for example removing only one compound (**19**) from the data set of 18 compounds increased the PI of the AutoDock+DSX scoring from -0.05 to 0.14.

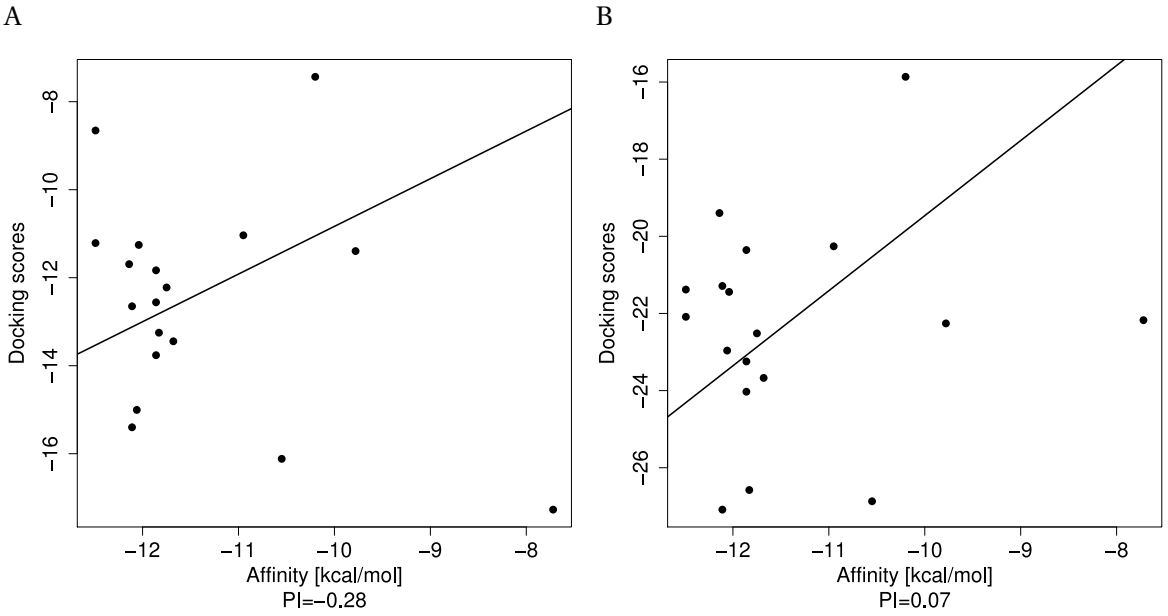


Figure 4.5.: GlamDock score versus experimentally determined binding affinity for 18 thiopeptides. A) *old high quality (HQ)* scoring function, and B) *new high quality (nHQ)* scoring function was used.

Since AutoDock was the only tool that displayed both, a convincing re-docking and reasonable affinity prediction accuracy, it was selected as the docking engine for all further docking studies throughout this thesis. Moreover, the fact that the AutoDock scoring function is based on the 1984 AMBER force field [167] (cmp. Section 2.2.1) was considered another advantage regarding the planned coupling with the AMBER MD package. In the following, adjustments are presented that were tested for a further improvement of AutoDock's performance.

AutoDock adjustments

An obvious parameter to tune when using stochastic algorithms is the number of runs (trials) that are performed. The AutoDock default number of genetic algorithm runs is 10 but it has been recommended to use a higher number (at least 100) to increase the exhaustiveness of the search, especially when docking very flexible ligands like peptides [258]. Docking the thiopeptides again with 100 runs instead of 15 improved the predictive index from 0.51 to 0.61 when using the MD-refined target structure. Like before, the PI for the crystal structure was considerably lower (0.28, see Fig. 4.6).

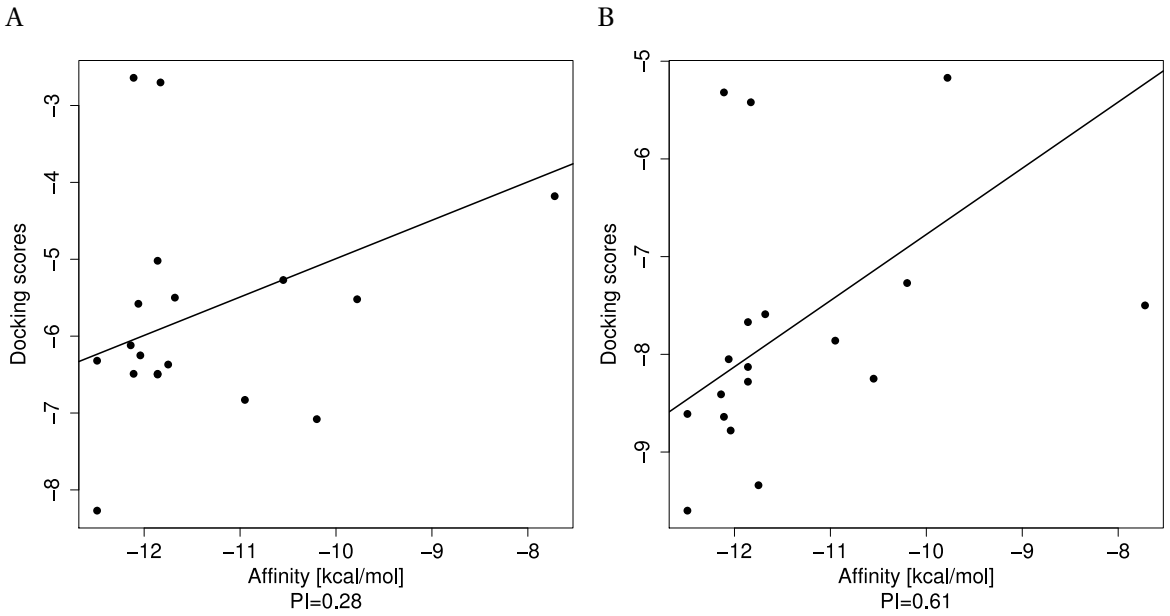


Figure 4.6.: AutoDock score versus experimentally determined binding affinity for 18 thiopeptides docked to A) the crystal structure and B) the MD-refinded structure. 100 docking runs.

Another setting investigated was the molecules' partial atomic charges. As described in Section 2.2.1, AutoDock uses Coulomb's law to describe electrostatic interactions between charged atoms. There exist several charge calculation methods (for an overview see Section 2.1.2) which lead to significant differences in the partial charges assigned to the different atoms [259]. Therefore, the method to calculate the partial atomic charges can largely influence the overall electrostatic contribution. There has been evidence [259–261] that more accurate charge calculation methods (quantum or semi-empirical) are inferiour to empirical charge calculation methods like the Gasteiger-Marsili method [100] in terms of docking accuracy as they can provide a more accurate modelling of electrostatic interactions.

For the MD simulation of the TS·23S·L11 complex (Chapter 3) RESP charges [98] compatible with the AMBER force fields [99] had already been calculated for the ligand, protein and RNA atoms. It was, therefore, straighforward to test if the RESP charge model would improve docking accuracy when docking TS back into its binding site. However, using RESP charges on the molecules' atoms

worsened the docking result. The previous docking energy of -13.37 kcal/mol changed to a positive value (3.12 kcal/mol) and the re-docking RMSD increased from 1.96 (1.64) Å to 13.58 (7.16) Å. Thus, Gasteiger charges were used in all further dockings with AutoDock. It was also found that the Gasteiger partial charges obtained with MOE and AutoDockTools (ADT) differed slightly[2]. These slight charge differences had a significant effect on the docking outcomes; the charges calculated by ADT produced consistently lower AutoDock scores. To obtain comparable results for different molecules, it was decided to use only ADT for charge calculations.

A second charge aspect was the negatively charged RNA backbone. In MD simulations of RNA this is accounted for by adding counterions to neutralize the simulated system. Docking simulations, however, are executed *in vacuo*, usually without considering any ions or water molecules. To solve the problem of the highly negatively charged RNA without the necessity to add explicit ions during the docking, Moitessier et al. [242] increased the partial atomic charge of the phosphorus atoms by +1. Analogously, this approach was tested with the 23S·L11 system, but did not lead to any observable changes in the docking outcome. One has to keep in mind, though, that all thiopeptides investigated are neutral. When dealing with positively charged molecules, such as the aminoglycosides investigated in the study by Moitessier et al., the proposed modification of the RNA charges might be worthwhile.

4.2.3. Discussion

As, until now, no assessment of available docking programs with respect to protein–RNA targets has been published, the performance of several protein-ligand programs when docking selected thiopeptides to the 23S·L11 subunit of the GTPase associated region was investigated. As noted earlier on other target systems [165, 262, 263], the performance among the different tools in terms of reproducing known binding modes and correctly ranking different ligands was quite diverse. FlexX and AutoDock Vina performed poorly and placed thiostrepton out of the binding pocket. Although in principal, these alternative binding modes cannot be ruled out, crystal and biochemical data in addition to inconsistent results between the two tools render them unlikely. FlexX's weakness at shallow binding pockets and very flexible ligands has already been observed, although with an older version [162, 264]. Moreover, its strategy to optimize for hydrogen bonds first might not be optimal for docking a compound possessing 18 hydrogen bond donors and 37 acceptors of which only a minority is actually involved in hydrogen bonds with the binding site. The complete lacking of modelling electrostatic interactions might be an explanation for Vina's poor performance.

The stochastic search algorithms implemented in GlamDock and AutoDock were obviously better suited to search the conformational space of the large and flexible ligands. As a drawback, using the latter tools accounts for a longer run time (approx. 75 min and 120 min per compound, respectively)

[2] The reason for this is currently unclear. A likely explanation is that the implementation of the Gasteiger-Marsili algorithm differs in the tools.

compared to FlexX and Vina (approx. 4 min and 13 min per compound, respectively). Using AutoDock with extensive search parameters (100 LGA runs) increases the runtime further to approximately 780 min and is therefore unsuited in a high-throughput context. Whereas GlamDock was slightly better than AutoDock with respect to re-docking accuracy, the AutoDock scoring function was the only one able to produce an acceptable ranking of the 18 investigated thiopeptides. Neither a re-scoring of the docking poses with a combination of the knowledge-based scoring functions DrugScore and DrugScore^{RNA} nor using a more accurate charge model could improve the results further. The failure of the DrugScore combination was astonishing as it was expected that specifically addressing ligand–protein and ligand–RNA interactions would be superior to AutoDock's scoring function that was solely trained on protein–ligand complexes. There are, however, two problems with the combined DrugScore$^{Prot/RNA}$ approach used here: 1) the data set that was used to derive the DrugScore^{RNA} potentials, and 2) the way the DrugScore and DrugScore^{RNA} potentials were combined. Redarding 1) The number of small molecule–RNA complexes in the PDB is much smaller than that of small molecule–protein complexes. This led to the necessity to include additionally small molecule–DNA complexes to obtain statistically significant potentials for DrugScore^{RNA} [176]. In contrast to DNA, which occurs almost always as double-stranded helices in the cell, RNA can form intricate three-dimensional folds, comparable in their complexity to protein structures [265]. Thus, although common motives exist, the interactions between small molecules/proteins and RNA are different to the ones observed for small molecule/protein–DNA complexes [266–268]. It is therefore questionable if a scoring function derived from both sets can model the interactions of small molecule-RNA complexes accurate enough. Regarding 2) DrugScore and DrugScore^{RNA} were derived from two independent data sets. Thus, their potentials might not be comparable and it might not be the correct way to simply add them for the protein and RNA part. Certainly, a reparameterization would be necessary to obtain statistically meaningful combined potentials. Since the results (also of the single DrugScore functions) were so discouraging, this route was not continued. Alternatively, it will be investigated in the following sections to what extend free energy computations (i.e. the MM-PB(GB)SA approach) can improve the affinity predictions.

An alternative explanation for the failure of the DrugScore^{RNA} and DrugScore$^{Prot/RNA}$ scoring is that the nature of the binding between thiopeptides and rRNA is much more "protein-like" than expected. The data suggest that despite the binding environment being dominated by the RNA [191], the binding site characteristics are more similar to a typical non-polar protein receptor than to isolated RNA. In fact, only surface-oriented base–ligand contacts and no charge–charge interactions are present in comparison to interactions of other ligands (for example aminoglycosides) with RNA.

The superiority of the simpler Gasteiger charge model compared to the RESP charges could be explained by the way the AutoDock scoring function was parameterized. The coefficients of the scoring terms (see Section 2.2.1) were determined by a regression analysis on a set of complexes with Gasteiger charges assigned [153]. Using a different charge set from the one used to derive the scoring function might lead to an imbalance between the electrostatics term and the other terms of

the scoring function.

Chosing a structural model representing the average equilibrated complex of an MD simulation of the TS·23S·L11 complex as target for the docking studies improved the docking results for all programs investigated. This demonstrates how important a high resolution structure is for structure-based investigations. If such a structure is not available, a structural refinement via MD might be a promising strategy. This will be further explored in Section 4.5. Further improvement of AutoDock's docking accuracy could be achieved by increasing the number of docking runs and thus the exhaustiveness of the conformational search. In summary, AutoDock showed acceptable re-docking and affinity prediction accuracy with the challenging system investigated in this thesis and will be used as docking engine in the following investigations.

4.3. Docking NMR Structures of Thiostrepton Derivatives for Characterization of the Ribosomal Binding Site

Through a collaboration with Dr. Henry Jonker (MPI Frankfurt/Main) it was possible to determine solution-phase structures of **1** and three derivatives (see Fig. 4.7) by NMR spectroscopy. This allowed to study the influence of selected molecular changes on the conformational space of TS. The determined solution structures were docked to the 23S·L11 subunit to determine their binding mode and correlate their molecular interactions at the ribosomal GAR with their bioactivity.

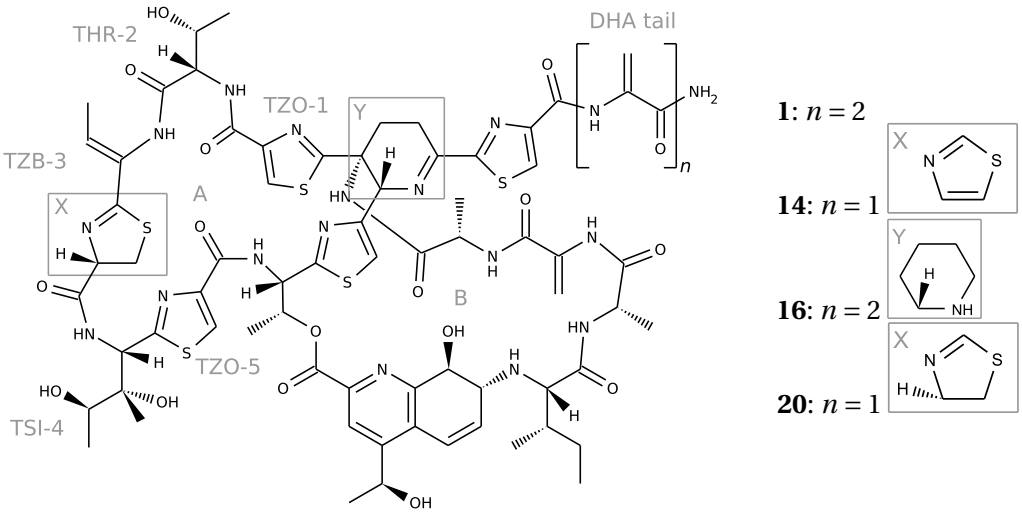


Figure 4.7.: Chemical structures of TS (**1**), its oxidized (**14**), reduced (**16**), and epimer (**20**) variant.

4.3.1. Materials & Methods

NMR Spectroscopy

As **1**, **14**, **16**, and **20** are barely soluble in biological buffer (low μM range) [250] and the binding site is largely hydrophobic, the moderately polar and adaptive [269] solvent mixture $CDCl_3$/[D_5]EtOH (5:1) was selected for NMR studies in solution. NMR spectra for TS and its derivatives (5–10 mM) were recorded at 298 K and 600 MHz. Further details are described in Ref. 235.

Biochemistry

Experimental ligand affinities to the target were determined by using a fluorescence-polarization-based assay, as described previously [191, 250].

Docking Studies

All docking studies were performed using AutoDock 4.2 with the optimal parameter settings determined previously (Section 4.2). The pose with the lowest score was chosen as representative of each cluster. Before docking, the L11 protein and 23S RNA portion of the ribosomal crystal structure (PDB 3CF5) were protonated and minimized using the Sander algorithm in AMBER 9. Two different docking scenarios were performed: 1) a rigid docking where all ligand torsions were inactivated in the input files, and thus inflexible during the docking, and the complete NMR bundle of 20 structures was docked in 15 independent GA runs. 2) For the semi-flexible docking, the target and the conformation of the macrocycles were kept rigid, the conformation closest to the average of the NMR bundle was selected as representative input conformation and 100 independent GA runs were performed.

4.3.2. Results

The solution structure of thiostrepton (**1**) that was determined by NMR spectroscopy compares very well with the X-ray crystal structure (PDB 1E9W) [208]. The structure of the quinaldic acid macrocyle (ring B) and the dehydrolanine tail are nearly identical for all derivatives. The main structural differences occur in the conformation of the A ring, specifically in residues 2–4 (Fig. 4.7). Their orientation significantly deviates in the oxidized compound **14** and, in particular, the epimer derivative **20** (Fig. 4.8). The X ring and the dihydroxyisoleucine side chain (TSI-4, cmp. Fig. 4.7) become displaced by 2.5–2.9 Å. Oxidation of the X ring flattens this subunit, but causes only a slight displacement of the adjacent dihydroxyisoleucine residue. The alteration in stereochemistry in epimer **20** leads to a much larger change in the structure, with a significant dislocation of ring X. Moreover, the adjacent thiazole ring (TZO-5) becomes tilted by 22°. The core structure of compound **16** is not strongly affected, but the orientation of the dehydroalanine tail is altered.

NMR structure ensembles generally cover a significant area of the energetically accessible conformational space, even though they may underestimate the truly accessible fluctuations [270].

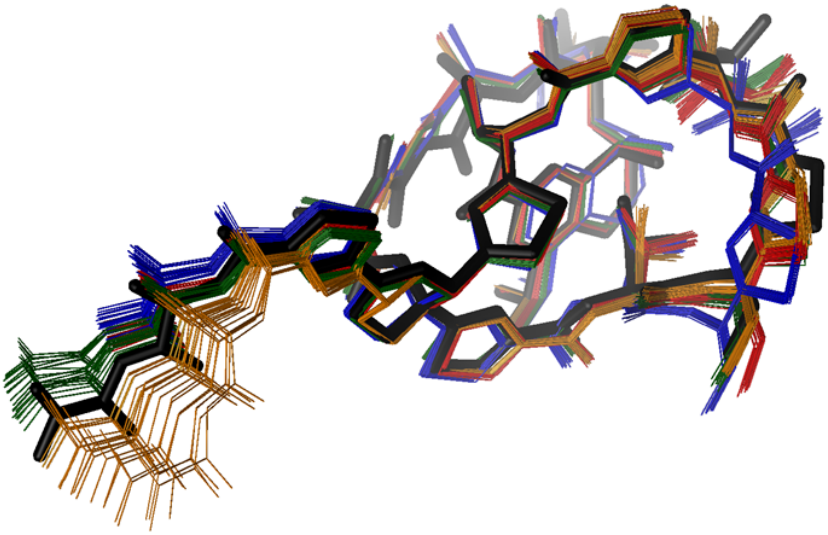


Figure 4.8.: Bundle of 20 minimum-energy NMR solution structures of compounds **1** (green, PDB code 2L2W), **14** (red, 2L2X), **20** (blue, 2L2Y), and **16** (orange, 2L2Z) superimposed on the previously known X-ray crystal structure of thiostrepton (black, 1E9W).

Therefore, all the individual members from the NMR-derived structure ensembles for **1**, its oxidized (**14**), reduced (**16**), and its epimer variant (**20**) were docked onto the crystal-structure coordinates (3CF5) of the L11 protein and 23S RNA binding site. By using these data for the completely rigid docking of **1**, consistently poses were found where the dehydroalanine (DHA) tail was inserted deeply into the cleft formed by L11 and 23S. Neither the recent structural data nor studies which localized the DHA tail biochemically support such a binding geometry [31, 32, 190, 191, 250].

Table 4.3.: Biological activities and docking scores.

Compound	K_D(L11/RNA) [nM][a]	Score[b]	Score[c]
1	0.20±0.05	-8.41 (4)	-8.35 (14)
14	0.23±0.14	-8.54 (23)	-8.54 (23)
16	0.10±0.05	-9.15 (2)	-8.71 (20)
20	5.60±1.10	-8.30 (6)	-7.92 (34)

[a] Apparent affinity to the reconstituted minimal L11·rRNA target complex from *T. thermophilus*.
[b] Best AutoDock score (number of conformations).
[c] Largest cluster docking score (number of conformations).

In a more refined semi-relaxed docking approach, the side chains of the ligand and the tail residues were treated as flexible[3]. For all four structures, poses with the best docking scores that resemble the overall binding mode in the crystal structure could be identified (Fig. 4.9A). These findings indicate that the flexibility of the side chains and the tail residues is important for proper binding to the target

[3]For a related technique, see for example Ref. 271

structure. A statistical evaluation of the binding conformations (Fig. 4.9B) consistently identified the crystal-like binding mode for **1**, **14**, and **16**[4], which exhibited comparable interactions with the binding site (see Fig. 4.10). In this analysis, the side-chain OH groups of THR-2 and TSI-4 form three hydrogen bonds to A1067, and TZO-1 displays a stacking interaction with A1095. The tail of **14** and **16** forms an additional hydrogen bond to the protein backbone. In contrast, most of these characteristic binding features were absent for epimer **20** (Fig. 4.10). Furthermore, the unlikely tail-in-cleft binding mode was frequently found. Taken together, these data and the lower score of the biochemically supported "best" solution strongly suggest that the epimer **20** is less well suited for binding to the target structure.

A B

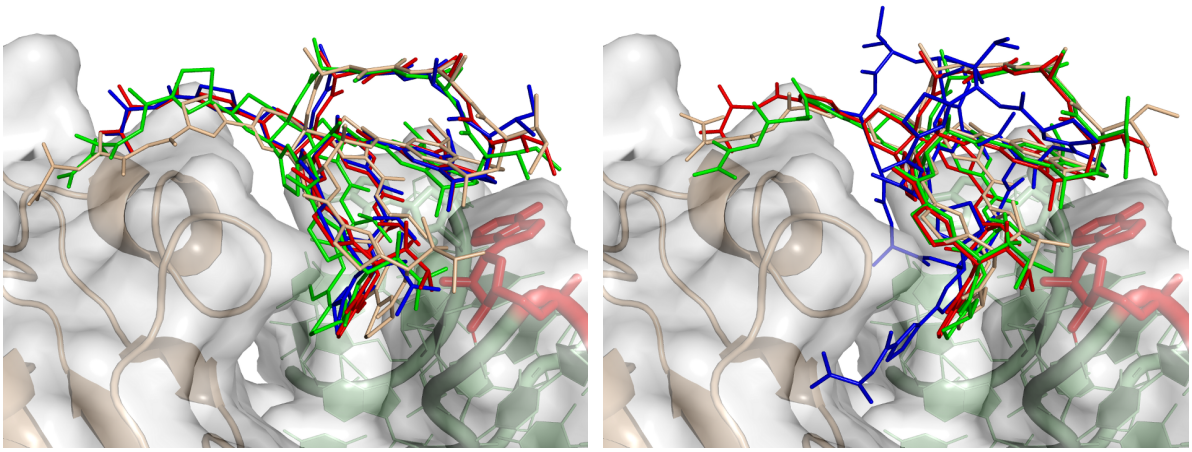
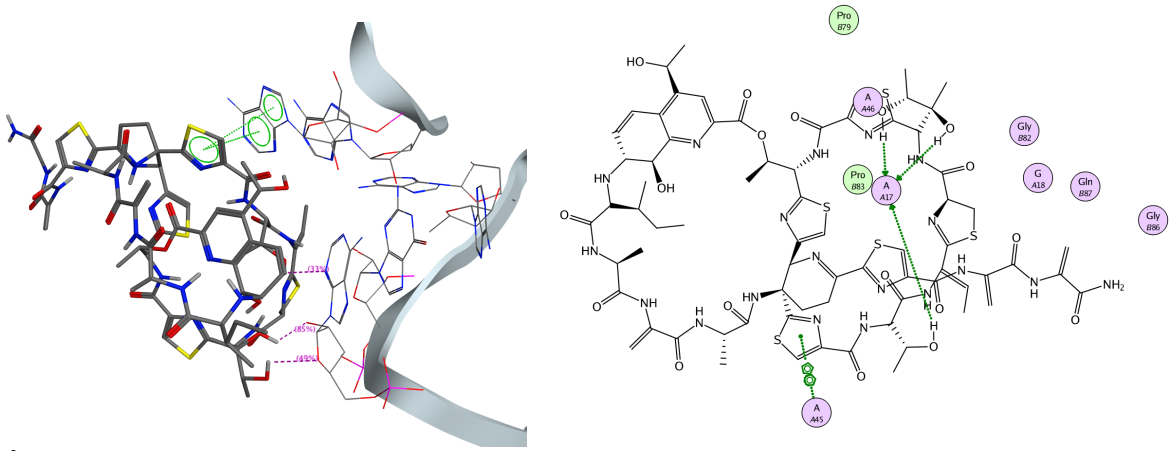


Figure 4.9.: Docking of **1** (green), **14** (red), **20** (blue), and **16** (light orange) to the L11 protein (wheat, only a backbone trace is shown) and 23S rRNA (pale green). The binding site's van der Waals surface is shaded gray, the crucial nucleobase A1067 is highlighted (red). A) Poses with the best docking scores. B) Representative poses from the largest clusters. Note the reverted binding pose being highly populated for **20**.

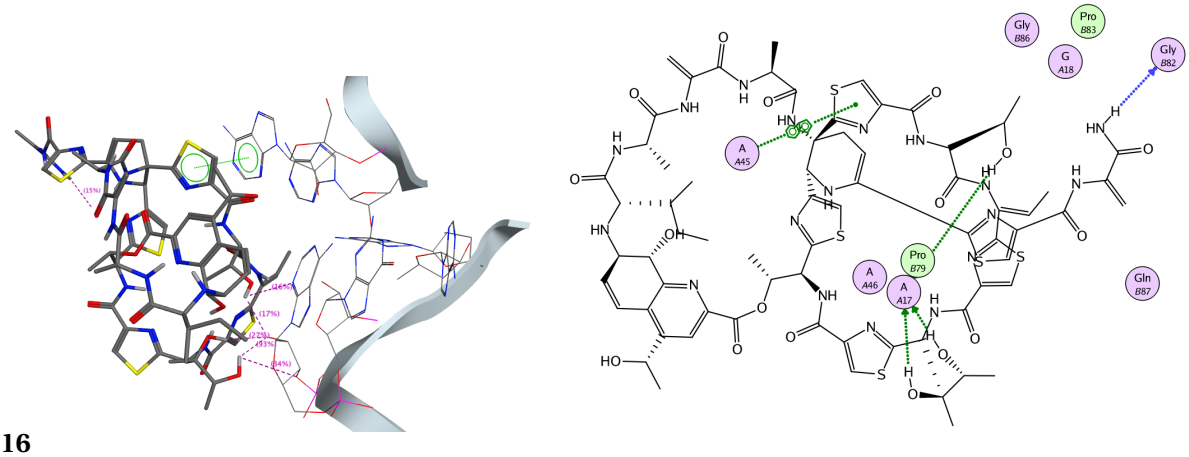
The docking studies ranked the reduced form **16** as the tightest binder, and placed thiostrepton (**1**) and the oxidized form **14** with nearly equal affinity in between **16** and **20** (Table 4.3). This computational trend was confirmed by biochemical measurements in which ligand affinities to the target were determined by using a fluorescence-polarization-based assay [191, 250]. The experimental conditions are slightly different because of the paucity of structural data (*D. radiodurans*) and reliable binding (*T. thermophilus*) and translation inhibition assays (*E. coli*). However, a high degree of conservation of the GAR structure and function among species [29] justifies the comparison. Thiostrepton (**1**), and the oxidized form **14** show very similar activity (Table 4.3). A significantly reduced affinity to the target (20–25 fold) was found for epimer **20**. As predicted by the docking studies, **16** had a significantly increased affinity to the isolated target – comparable to the highly active nosiheptide [17, 250, 273].

[4]The protonation state of the piperidine had no effect on the results.

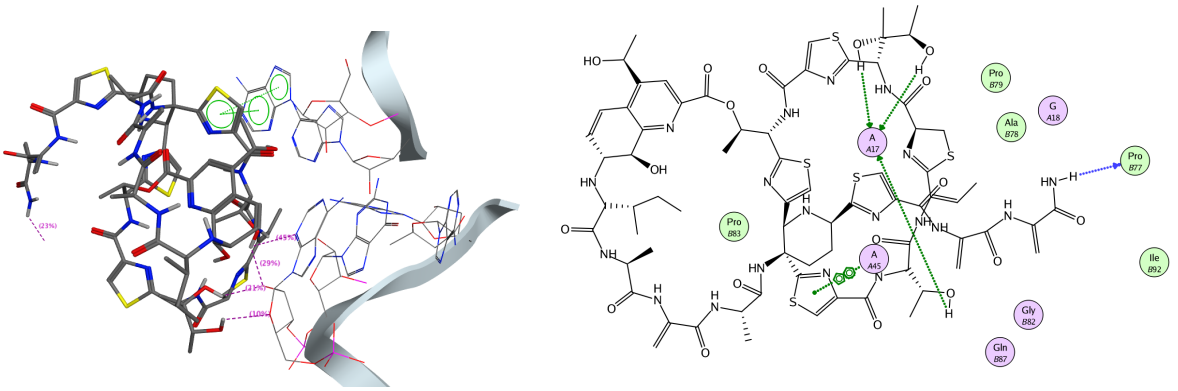
1



14



16



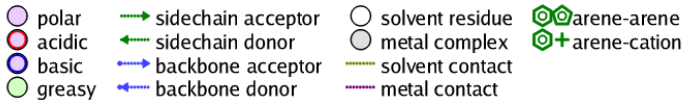


Figure 4.10.

20

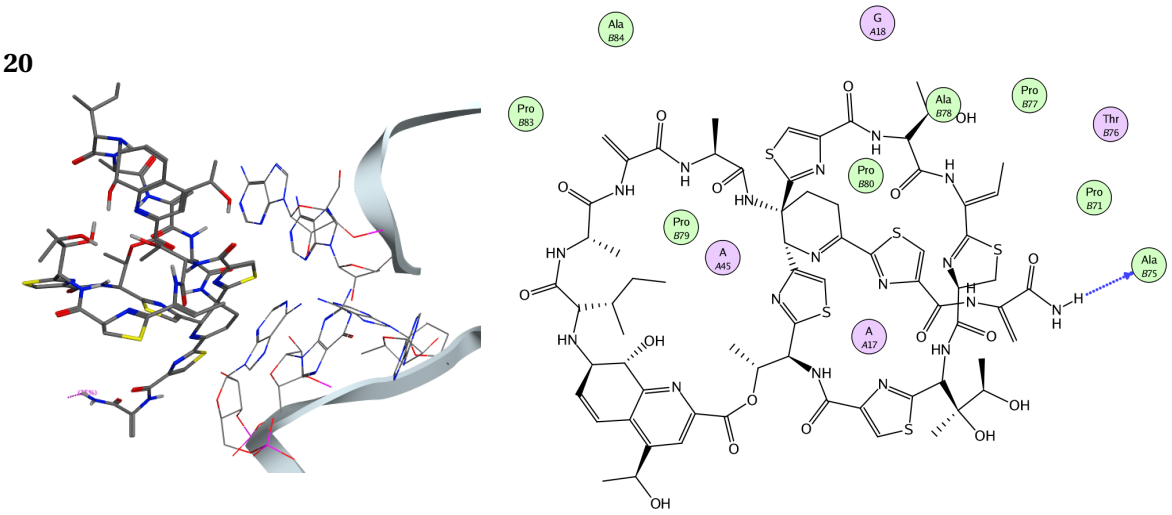


Figure 4.10.: Interactions of **1**, **14**, **16**, and **20** with the L11·23S binding site. Interactions are shown as a 3D (left) and 2D structural depiction (right) of the largest cluster docking solutions. For simplicity, only the RNA is displayed in the 3d depictions. The hydrogen bond's quality is expressed as percentage probability [272]. Note that the numbering of interacting residues is as found in the truncated model used for docking, for RNA residues (chain A) 1050 has to be added, for protein residues (chain B) 58 has to be subtracted.

4.3.3. Discussion

The epimer **20** showed a narrowing of the cleft between the A and B rings, which leads to a shape mismatch with the target structure. The insertion of the dehydrobutyrine residue (TZB-3) between the RNA and the L11 protein as well as the placement of TSI-4 at the A1067 residue then suffers. It is remarkable how well the experimental data was qualitatively captured by the integrated NMR-docking procedure, particularly given the moderate resolution of the target X-ray structure (3–4 Å). This trend strongly suggests that the inhibitor structures do not undergo major conformational reorganization on binding to the target. Thiopeptide inhibitors may have evolved to mimic this recognition motif, for example, to mimic domain V of the elongation factor EF-G in one of the states during the PRE to POST transition of the ribosome [28, 274, 275].

In summary, integrated semisynthesis, NMR spectroscopic solution-structure determination, computation, and biological evaluation studies have identified key conformational and structural parameters involved in thiostrepton targeting the ribosome's GAR. This ternary ligand·RNA·protein interaction seems to be driven by ligand shape and RNA recognition [191], but is moderately tolerant of structure variation. The NMR analysis of the structures of thiostrepton (**1**) and its derivatives revealed that the molecular scaffold of **1** may not be perfect for addressing the pharmacophore region of this highly complex RNA–protein target. The minor mismatch seems to be healed in part by the dihydroxyisoleucine side chain. Overall, these data define structural boundaries, wherein

an improvement of the overall pharmacological profile of these compounds or their analogues will be possible. These structural boundaries can be summarized as follows: 1) The truncation of the DHA tail by one unit had no detectable effect on the affinity, demonstrated by the almost identical K_D values determined for TS (**1**), and TS-1 [235] (data not shown). The limited number of contacts that were found by the docking studies between TS's tail and L11's N-terminal domain, and the considerable size of that protein region's surface, easily explain this observation. 2) Reduction of the dihydropiperidine imine function (Ring Y) led to a slight reorientation of the tail, allowing it to approach the protein surface more easily, and thus causing an increase of affinity. 3) The most pronounced effects were found for changes within macrocyle A, specifically ring X. The oxidation of ring X changes its conformation only slightly and still allows a similarly tight binding like TS. In contrast, the conformational change caused by ring X's epimerisation led to a shape incompatibility with the rRNA·protein cleft, illustrating the compounds' sensitivity to modifications of that molecular substructure.

4.4. Using Quantum Mechanics and Molecular Dynamics to Refine 3D Structures of Thiopeptides

In the previous section it was shown how the prediction of three-dimensional 23S·L11·thiopeptide complexes can aid the interpretation of structure-activity relationships and define boundaries for the modifications of the TS scaffold. The excellent agreement between predicted and biologically determined binding affinities for the compound set suggest that reliable predictions for newly designed compounds can be expected. However, only TS and three derivatives were investigated. It would be desirable to extend the SAR investigations to more diverse derivatives. The largest hurdle to overcome for achieving this goal is to obtain accurate 3D structures of the ligands for docking input. As the solution structure determination of thiopeptides by NMR involves significant effort and ressources, an *in silico* strategy would be a cost-effective alternative. The approach of using the TS structure as template, introduced in Section 4.2, turned out as a promising starting point, but it is questionable if a short minimization with a general force field can result in large structural changes. In the current section an alternative approach is presented that makes use of quantum mechanics and molecular dynamics simulation to refine 3D structures of thiopeptides.

Force-field optimization remains an important endeavor for modelling chemically diverse systems at atomistic resolution. Most MD simulations are on "traditional" systems involving standard amino and nucleic acids, where the force fields are well established. Modelling thiopeptides, however, is challenging using force field based methods. Like many natural products they have chemical functionalities and bonding that are not present in protein, DNA, and RNA macromolecules. Thus, an additional force field is needed whose parameters cover the unique functionalities within thiopeptides, and that is balanced with the protein/RNA force fields used. In this section the development

of a force field for thiopeptides is presented. The performance of this force field when used in a molecular dynamics simulation of thiostrepton (**1**) immersed in an organic solvent mixture is tested for exploratory and methodological development.

4.4.1. Materials & Methods

Force-field Parameter Optimization

TS was separated into its constituent residues as defined by Bond et al. [208] (Fig. 4.11). Each non-

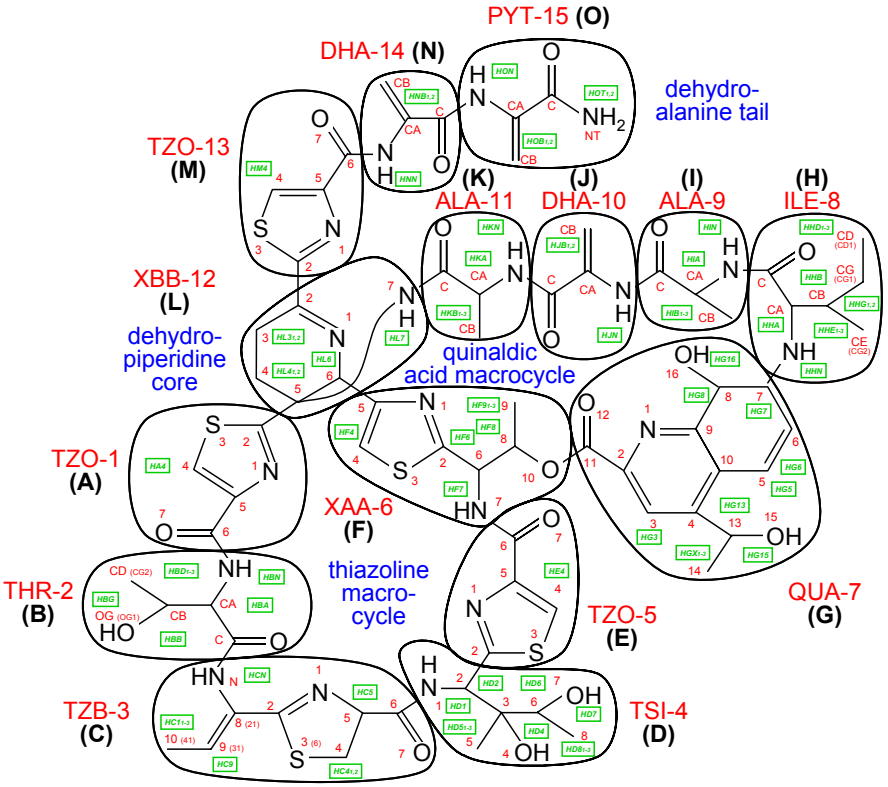


Figure 4.11.: Thiostrepton separated into its 15 constituent residues.

amino acid residue's initial conformation was taken as found in the TS crystal structure (3CF5 [32]), capped using the nearest functional group, and terminated using a methyl group (see Fig. 4.12). The standard residues ALA, ILE and THR were built in LEaP with ACE (CH_3CO) and NME ($NHCH_3$) capping groups. All residues were optimized using HF/6-31G(d) theory level, and electrostatic potentials were subsequently computed using CHELPG [95]. A RESP fit was performed via R.E.D. [276] using a 0.01 weighting factor to obtain charges compatible with Parm99SB [88]. The determined partial atomic charges can be found in Tables C.2 and C.3 in the appendix. Initial parameters for the standard amino acid residues were taken from Parm99SB [89], while the non-standard residues parameters were

taken from GAFF [91] (cmp. Chapter 3). A second intramolecular parameter set was optimized for modelling TS, chloroform, and methanol (Dr. Karl N. Kirschner). Since TS will be modelled bound to the ribosome, existing Parm99SB intermolecular Lennard-Jones parameters are used instead of optimizing them. Further methodological details are descibed in Ref. 277.

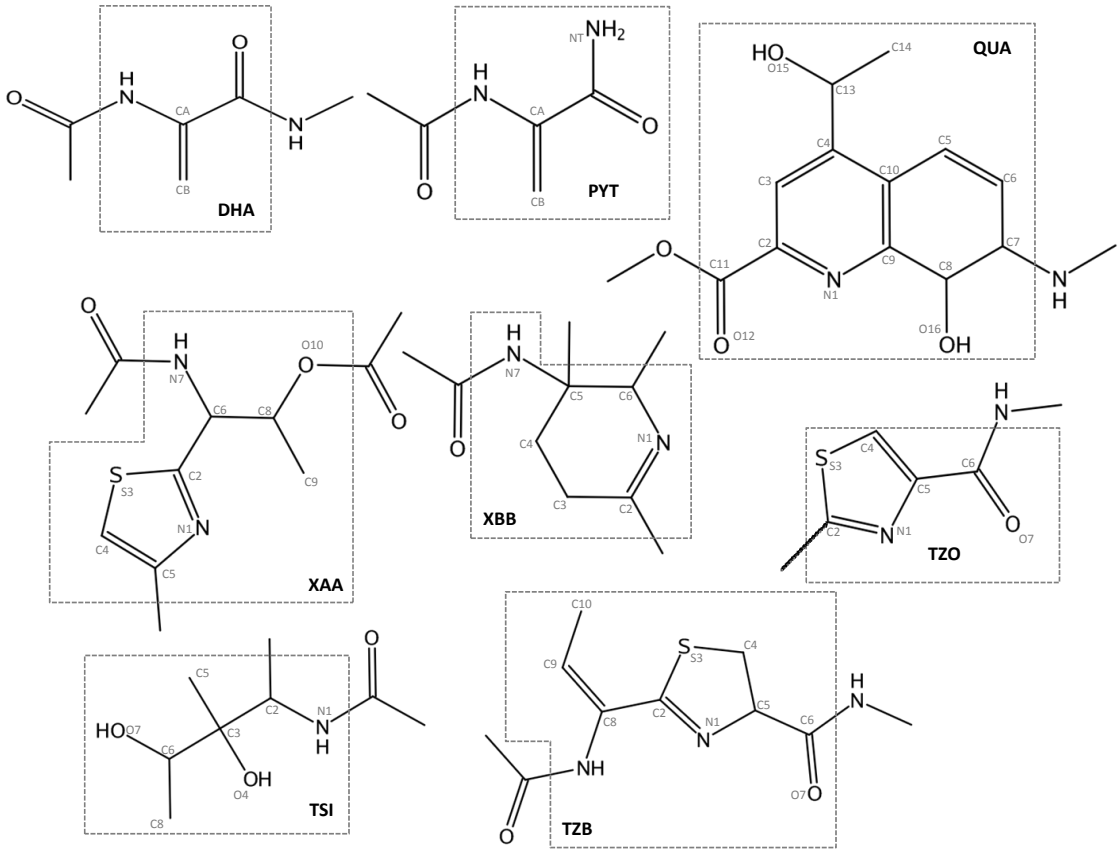


Figure 4.12.: Molecules used for deriving intramolecular parameters and RESP charges for non-standard thiostrepton residues. Molecular fragments are surrounded by dashed lines.

Molecular Dynamics Simulation for Parameter Validation

Both parameter sets were used in separate 20 ns isothermal-isobaric MD simulation performed at 300.15 K. 1-4 electrostatic and nonbonded scaling factors were set to one during the parametrization and the MD simulation using the novel force field, otherwise the default parameters were used (1.2 and 2.0, respectively). All MM-based and QM calculations were performed using AMBER [82] and GAMESS [278].

Replica-exchange Molecular Dynamics Simulation

In an attempt to predict TS's 3D structure from 2D input, replica-exchange molecular dynamics (REMD) simulations with AMBER 11 were set up. Since the current AMBER 11 REMD implementation does not allow explicit solvent simulations [116], TS was simulated in generalized Born (GB) implicit solvent. An initial three-dimensional TS model was generated with CORINA [257], and subsequently minmized using 500 steps steepest descent, followed by 500 steps conjugate gradient minmization. The OBC generalized Born model I [279]) with a nonbonded pairs truncation cutoff of 99.0 Å and a salt concentration of 0.0 M was used. To determine the number of replicas and their temperatures the "Temperature generator for REMD-simulations" [280] was used, requesting an initial temperature range from 300 to 600 K and an exchange acceptance ratio of 0.2. Each of the eight replicas was equilibrated individually to its target temperature using a Langevin thermostat in a 200 ps unrestrained simulation. A different random number seed was used for each replica. The REMD simulation was run for 40 ns with an exchange attempt frequency of once every 500 steps. All MD simulations were conducted using the SHAKE algorithm [211] to constrain hydrogen bonds and a step size of 2 fs. Coordinates were recorded every 2 ps. To overcome rotational barriers, a second REMD run was performed with ten replicas, covering a temperature range from 300 to 770 K.

4.4.2. Results & Discussion

Validation of the Optimized Parameters for Thiostrepton

An initial MD investigation of TS bound to the ribosome's GAR was done using the Parm99SB/GAFF parameter combination (Chapter 3). Using such a combined parameter set is a plausable route when confronted with an incomplete force field. However due to the developmental differences between force fields, such a "Frankenstein" approach can lead to untrustworthy results. Ultimately a force field is needed that is more reliable and whose errors can be tracked down and corrected more easily. Thus, TS's intramolecular degrees of freedom were optimized; in total, 35 bonds, 89 angles, and 153 torsions, using 18 atom types.

To partially validate the parameters, a model was created to reproduce the condition of the NMR experiment ([235] and Section 4.3) by immersing TS in a 5:1 chloroform:methanol mixture. As no such solvent mixture is readily available in AMBER, the desired mixture was created by 1) solvating TS in a chloroform box (CHCL3BOX with a padding of 28 Å in all directions), and 2) replacing every 6th chloroform residue by a methanol molecule. Two MD simulations were performed using the Parm99SB/GAFF combination and the novel optimized force field.

While the Parm99SB/GAFF parameter set performs adequately, as seen in Figure 4.13A, the optimized force field (Fig. 4.13B) reproduces the NMR structure better. However, a close examination of the torsion values and dynamics within TS's tail residues revealed some questionable behavior by the optimized force field. A re-examination of the critical torsion parameters showed that an alternative

solution could be found for the coupled parameters describing the rotation about C2-CP and C2-N bonds. After reoptimizing these parameters (Dr. Karl N. Kirschner), a third MD simulation was performed and provided improved structural agreement to the NMR data (Fig. 4.13C and Fig. 4.14). This led also to a better agreement in the quinaldic acid macrocycle (Ring 2) structure, which contains a residue that is dependent upon these parameters (Fig. 4.11).

Analysing Thiostrepton's Flexibility

After the optimized force field parameters had been finalized and validated, the data obtained from the MD simulation of TS in organic solvent allowed conclusions about the flexibility of the molecule by analysing the torsion angle distributions. Assessing molecular flexibility by MD approaches provides an advantage over experiments here since the structure of the dehydroalanine tail region could not be well defined by NMR spectroscopy [235]. Figure 4.15 depicts the distribution and location of the torsion angles exhibiting the highest varation during the simulation. Expectedly, and also visible from Figure 4.13, TS's tail is the most flexible part of the molecule. The standard deviation (sd) of the two C2-C2-N-CP torsion angles contained in the tail is ~48°. Beside from these two torsions, only the torsion angle in the isoleucin side chain (ILE-8) of Ring 2 showed comparable flexibility (sd of 46.7°). The rest of the torsions sample a very restricted angle range. Among those, several torsions contained in Ring 1 show at least a moderate flexibility (sd of 16.7–23.6°, Fig. 4.15B). Interestingly, they are located near the position where TS enters the cleft formed by the protein–RNA interface.

In summary, the MD simulation of thiostrepton in solution revealed that its two macrocycles possess a remarkable rigidity allowing only the tail and a few side chains to move. Thus, it is likely that only these portions can adjust to the 23S·L11 subunit upon binding. The inherent drawback of most docking algorithms to keep macrocycles rigid during docking is therefore not the worst approach for modelling the thiopeptide compounds. However, it again underlines the importance of a correct conformation of the macrocyclic core in the docking input structure. Otherwise, a correct placement at the binding site will not be found.

Predicting Thiostrepton's 3D structure from 2D input

Starting from an experimental input structure (3CF5 X-ray), the MD simulation using the TS optimized parameters gave an excellent geometric agreement with the NMR structure. However, the ultimate goal – valuable in a drug design project – would be to predict thiostrepton's three-dimensional structure *ab initio*. To test if starting from a 2D structure would lead to a similar agreement, an initial input conformation was generated with CORINA [257]. The resulting 3D conformation, deviating by 5.68 Å from TS's crystal structure, was subjected to a 20 ns MD simulation using the same protocol as before. However, a stable conformation hardly differing from the input conformation was obtained. Obviously, a conventional MD simulation was not able to sample the conformational space sufficiently.

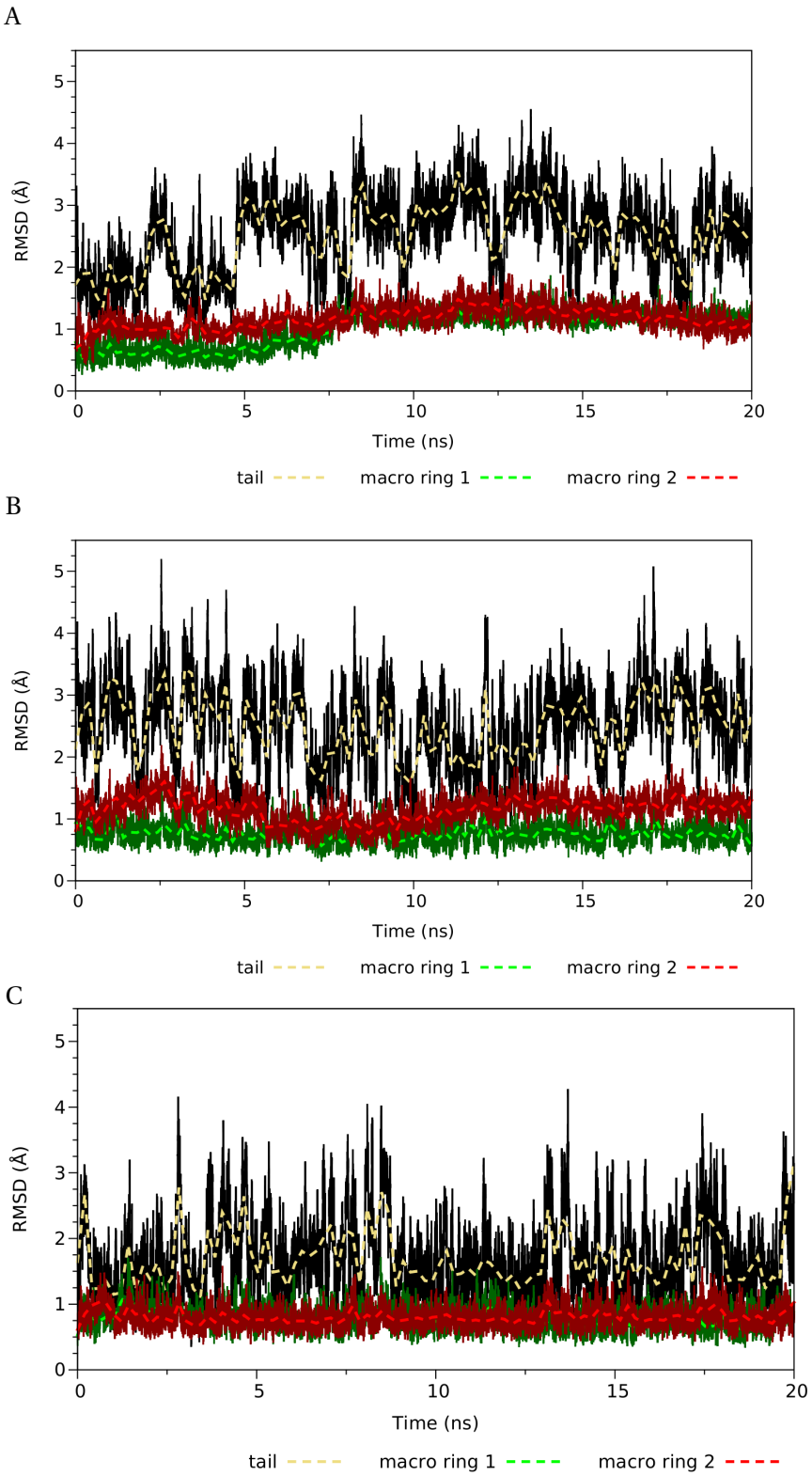


Figure 4.13.: Root-mean-square deviation (RMSD) to NMR reference structure for Parm99SB/GAFF force field simulation (A), simulations with initial optimized parameters (B), and reoptimized parameters (C).

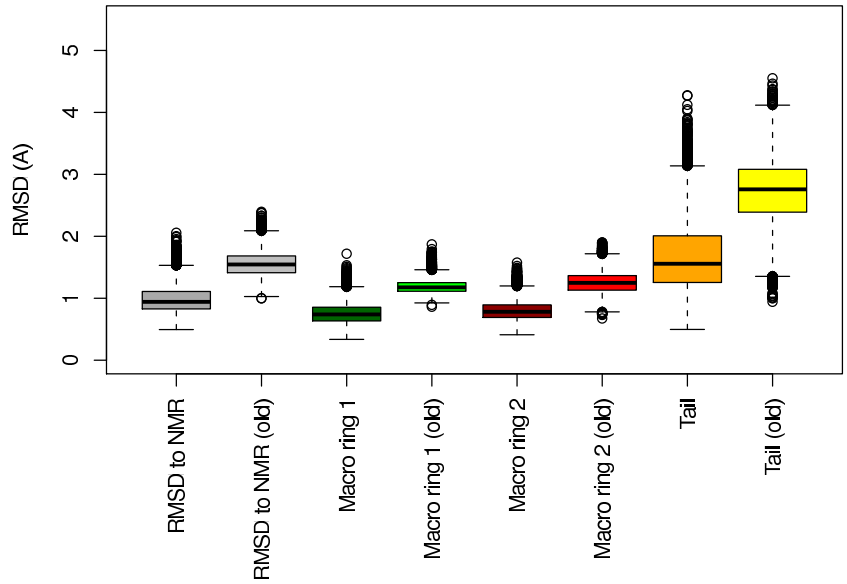


Figure 4.14.: Box plots of RMSD distribution for the overall RMSD to the NMR structure, and separated for the tail (yellow), macrocycle A (green), and macrocycle B (red). The darker colors represent the simulation results for the optimized force field, the lighter colors the results of the (old) Parm99SB/GAFF combination. Evaluated for the simulation time frame 8–20 ns.

To increase the sampling efficiency, a replica-exchange molecular dynamics simulation (REMD, see Section 2.1.4) [114] was set up (see Methods section for details of the REMD simulation conditions). Based on the number of atoms (199, excluding hydrogen atoms that were under SHAKE constraints), a temperature range from 300 to 600 K, and a requested exchange acceptance ratio of 0.2, an optimal number of eight replicas was determined by the "Temperature generator for REMD-simulations" [280]. Table 4.4 contains the temperatures for the replicas, along with their exchange acceptance ratios (EAR) for the first 1000 exchange attempts of the REMD simulation, validating that the required exchange ratio was reached.

Table 4.4.: Initial temperatures of replicas and exchange acceptance ratios (EAR) for 1000 exchange attempts.

Replica	1	2	3	4	5	6	7	8
Temp. (K)	300.00	336.27	375.67	418.61	465.28	516.07	571.35	631.59
EAR	0.16	0.19	0.18	0.22	0.20	0.19	0.19	0.24

When analysing the results of this simulation, it was found that TS's RMSD to the average NMR structure was still fluctuating around 5.5 Å. A conformation close to the NMR structure was not found – in any of the replicas, most likely because of unsufficient sampling due to too low temperatures and not enough replicas. A distribution analysis of selected torsion angles at the highest temperature (631.59 K) revealed that the temperature was not high enough to overcome barriers in peptide and

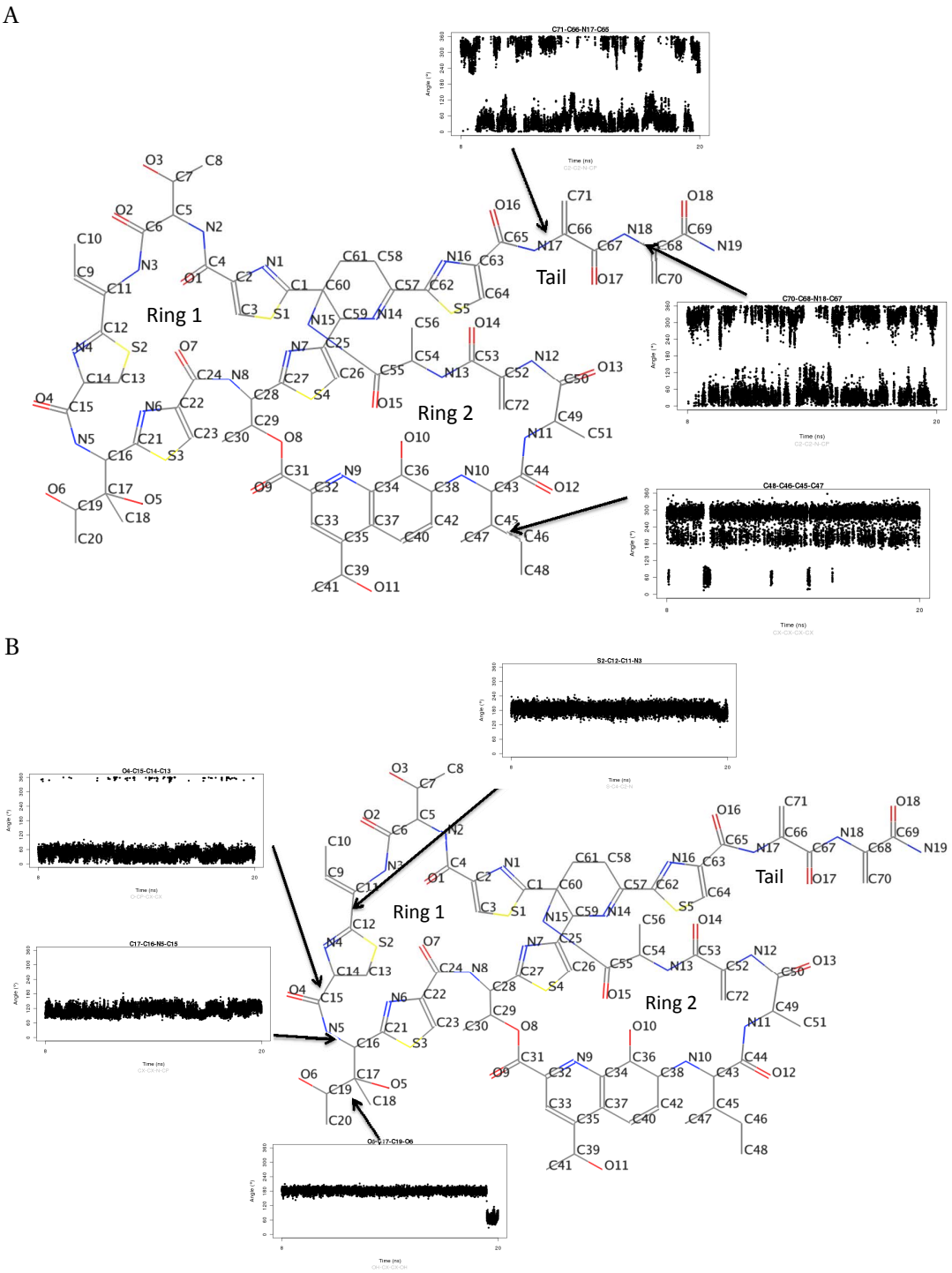


Figure 4.15.: Thiostrepton torsions showing high (A) and moderate (B) flexibility. High flexibility was determined based on sampling different conformational states, while moderate flexibility was determined based on a large standard deviation. The remaining torsion angles displayed very stable trajectories with low standard deviations.

double bond rotations. Therefore, a second REMD simulation with ten replicas up to a temperature of 768.68 K was performed. Figure 4.16 shows a section for the first 50 exchange attempts from this second simulation, ensuring that a good exchange between the ten replicas was reached during the simulation.

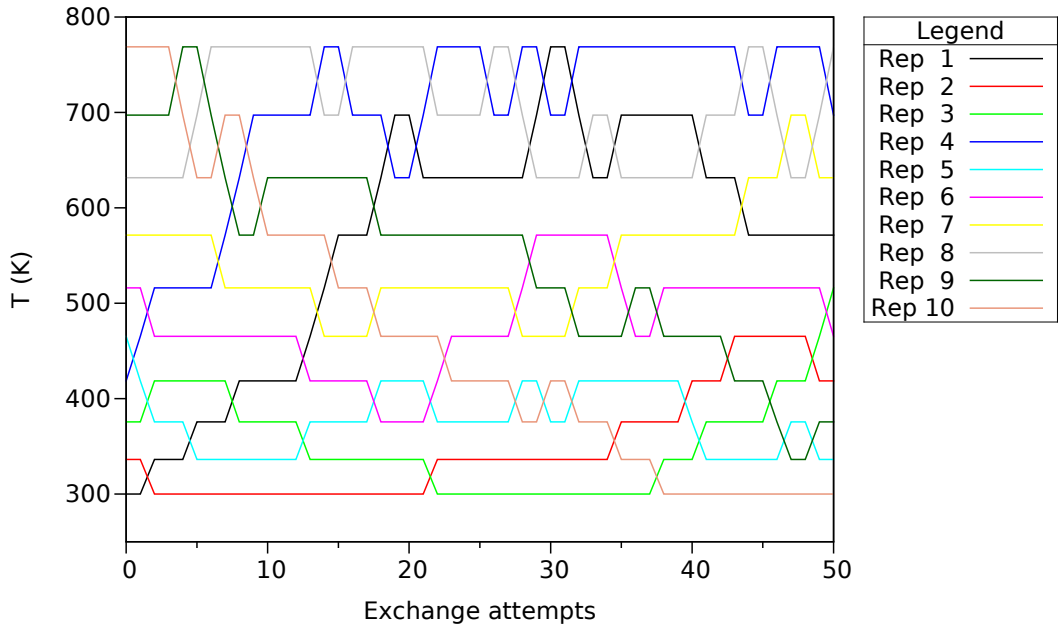


Figure 4.16.: Section from the replica exchange molecular dynamics (REMD) simulation process.

Even with these ten replicas, providing a temperature range from 300 to 768.68 K, it was not possible to sample a conformation below ~4 Å deviation from the NMR structure (Fig. 4.17). Assigning the RMSD to the tail, macroring 1, and macroring 2, separately (Fig. 4.17B), revealed that this arose from conformational restrictions in the two macrocyles. Although an initial RMSD decrease of ~1.5 Å can be seen and close conformations of the flexible tail part are found (closest distance 0.65 Å), the conformations of the two ring systems do not reach a deviation below 3.7 Å (Ring 1) and 3.4 Å (Ring 2). These results demonstrate that TS's structural complexity does not allow to predict its three-dimensional structure starting from 2D topology information. To obtain reliable 3D structures of this compound class' members, a knowledge-based strategy seems to be more promising. This approach will be explored in the next section.

4.5. Binding Mode Predictions and SAR Interpretations of Thiostrepton Derivatives

In Section 4.3 the binding mode prediction of TS and three derivatives, via docking their NMR determined structures to an X-ray structure of the GAR, was reported. Their molecular interactions were

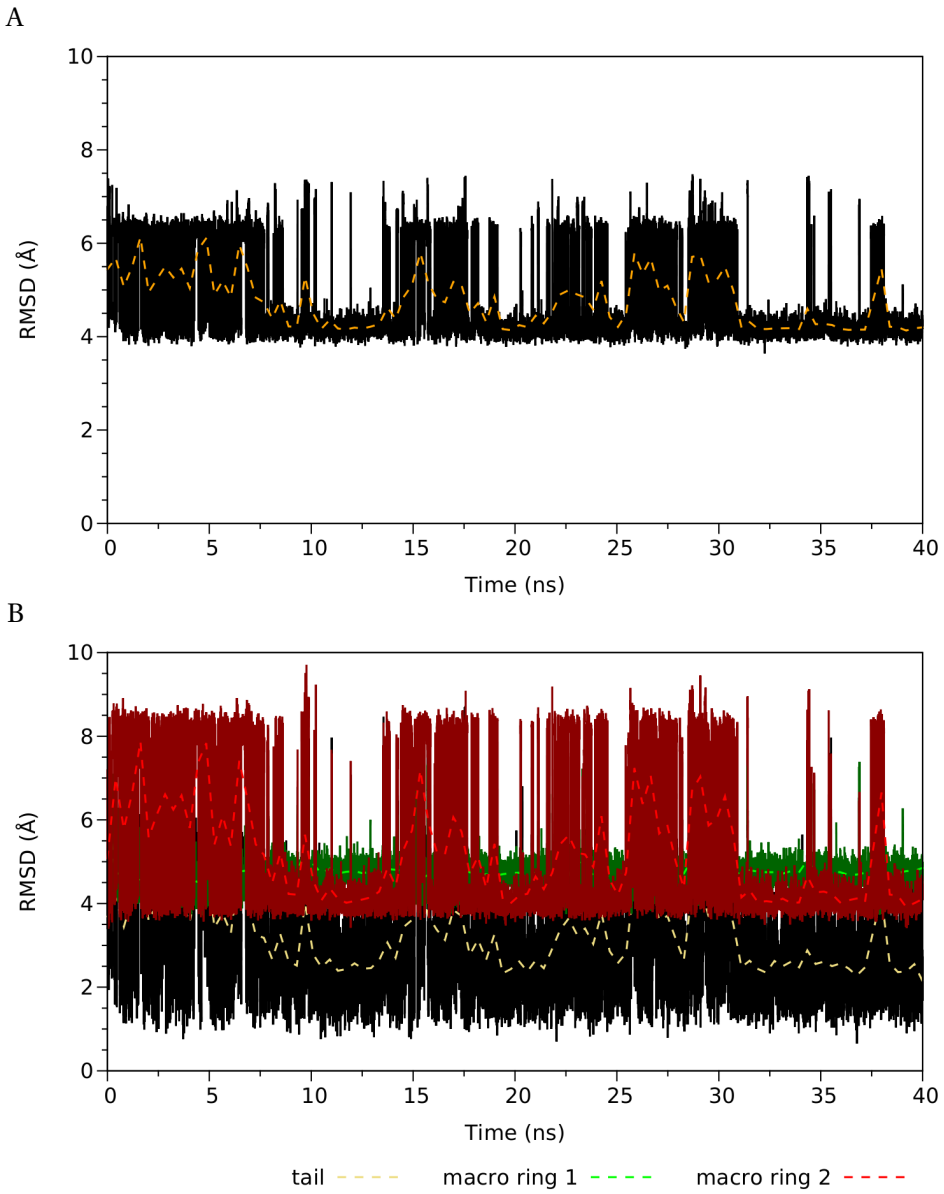


Figure 4.17.: Root-mean-square deviation (RMSD) to NMR reference structure for 300 K REMD trajectory. A) Overall RMSD. B) RMSD assigned separately for TS's tail (yellow), macrocycle A (green), and macrocycle B (red) components.

correlated with their biological activity to facilitate the optimization of future compounds. However, the structural differences between the investigated compounds were limited. In the following, this study is extended to five further derivatives whose activity was determined (group of Dr. H.-D. Arndt, MPI-DO), but whose three-dimensional structure is unknown. Figure 4.18 contains the structures of the investigated compounds, representing a subset of the TS derivates shown in Figures 4.2 and 4.4. These compounds feature diverse modifications at different positions of the TS scaffold and a four orders difference of magnitude measured affinities (see Table 4.5). Therefore, these molecules were selected for a closer analysis. Furthermore, a purely virtual fragment of **1** lacking the second macrocycle is investigated (compound **21** in Fig. 4.18).

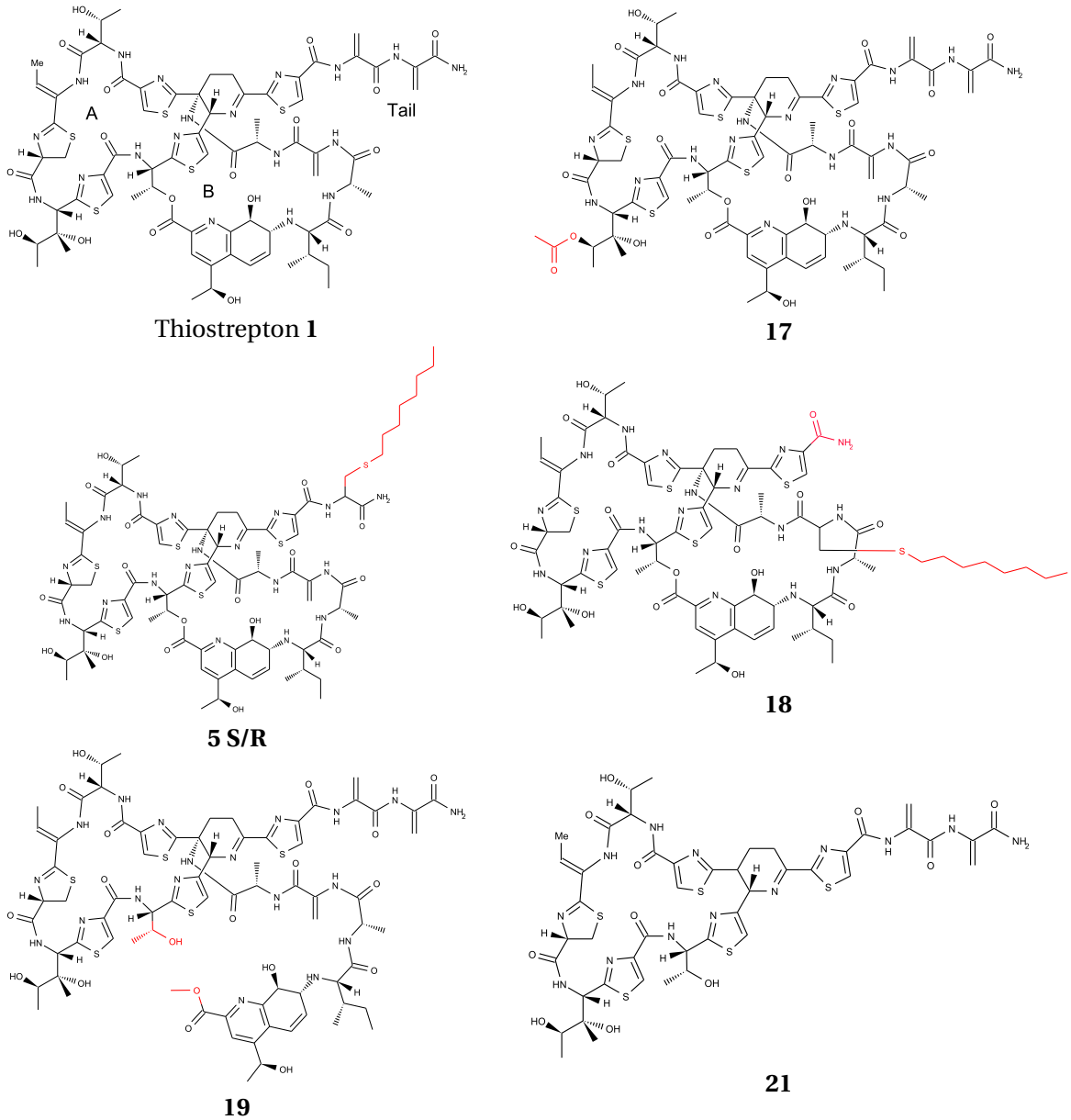
To predict reliable ternary complexes of the thiopeptides with the 23S·L11 subunit, an integrated MD–Docking–MD workflow is established (Fig. 4.19). In the proposed workflow, 1) initial MD simulations are used to generate refined input structures of a) the receptor, and b) the ligands, 2) complexes are generated via docking by AutoDock [132], and 3) these complexes are relaxed and re-scored via a second MD step to account for flexibility and solvation effects which are not addressed by AutoDock. Steps 1 and 2 of this workflow were already addressed partly in the previous sections. In Chapter 3 a refined structure of the medium-resolution 23S·L11·TS complex was obtained from a molecular dynamics simulation. The primary docking results of Section 4.2 indicated that this MD-refined structure might be a better suited target structure for docking of thiopeptides. In Section 4.4 a tailored thiopeptide force field was developed and used in an MD simulation to refine TS's three-dimensional structure. This approach is extended to the TS derivatives investigated here. Finally, a third workflow step is introduced, making use of MD simulations to relax the docked complexes. Moreover, these MD simulations are used to calculate free energies of binding using the MM-PB(GB)SA approach (see Section 4.5). A residue-wise decomposition of these energies is used to identify key structural elements for binding within the receptor binding site and within the ligand structures. These findings contribute to the pharmacophoric description of the 23S·L11·TS complex which is presented in Chapter 5.

4.5.1. Materials & Methods

Step 1: Generation of Input Structures

Receptor The crystal structure of TS bound to the 50S ribosomal subunit has a resolution of 3.3 Å [32], implying some uncertainty concerning the exact position of the side chains and TS. It is thus not an ideal target structure for molecular docking studies. Therefore, all docking studies were performed using 1) the (minimized) 3CF5 23S·L11 crystal structure (cmp. Section 4.3) and 2) the MD-refined structure obtained in Chapter 3 as target.

Ligands A pre-requisite for docking is the presence of three-dimensional structures for the receptor and the ligands. As a 3D structure is only available for TS [208], the input structure of the derivatives



Thiostrepton **1** | **17**

5 S/R | **18**

19 | **21**

Figure 4.18.: Investigated TS derivatives.

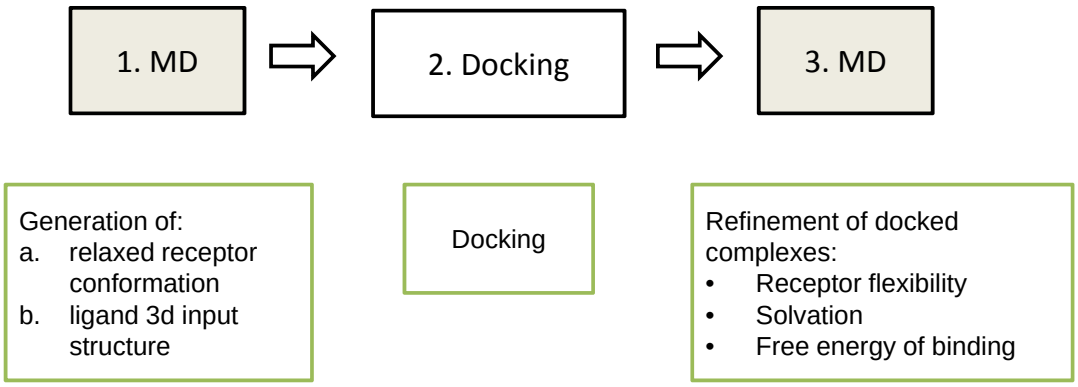


Figure 4.19.: Schematic representation of the proposed MD–Docking–MD Workflow.

had to be generated first. Due to the size and the complex structure of the investigated compounds, 3D generators conventionally used in docking, such as CORINA [257], reach their limits, as demonstrated in the previous sections. Therefore, the TS crystal structure (PDB code 1E9W [208]) was used as a template and the respective groups were modified using MOE, as done previously in Section 4.2. In contrast to the previous approach, the structures were subjected to 10 ns MD simulations in an organic solvent to relax them, using a thiopeptide-tailored force field with optimized bond, angle and torsion parameters. Most of the derivatives contained several parameters which were not contained in the force field derived for TS. Therefore, these additional parameters were optimized (Dr. Karl N. Kirschner) and the partial atomic charges for nine new residues (Fig. 4.20) were determined, as described in Section 4.4. The simulation procedure was similar to the one used to validate the optimized force-field parameters (see Section 4.4). The modified compounds were immersed in a 5:1 chloroform:methanol box with 28 Å padding in all directions. This moderately polar organic solvent was used to account for the thiopeptides' low solubility in water and the largely hydrophobic binding site at the 23S·L11 ribosomal subunit. After minimization, the system was heated to 310 K in a 100 ps constant volume simulation, followed by a 10 ns production run under constant pressure conditions. 1-4 electrostatic and vdW interactions were scaled by 1.0 each according to the parameterization of the force field.

Subsequently, the ligand MD trajectory was clustered based on the macrocycles' heavy atoms' RMSD between the snapshots, using the average-linkage algorithm as implemented in PTRAJ [125] with a cutoff distance of 0.8 Å and an initial trajectory sieve of 10. Representative structures of the dominating clusters were used as docking input. With exception of fragment **21**, for which two clusters were observed, one cluster dominated the other simulations, and thus only one conformation for the compounds was docked in each case. Towards the end of the simulation of compound **19** a rise of the second macrocycle's RMSD was noticed (Fig. 4.21). Therefore, the simulation of **19** was extended to 20 ns and the representative structure of the last 5 ns – after the RMSD had stabilized – was chosen as docking input.

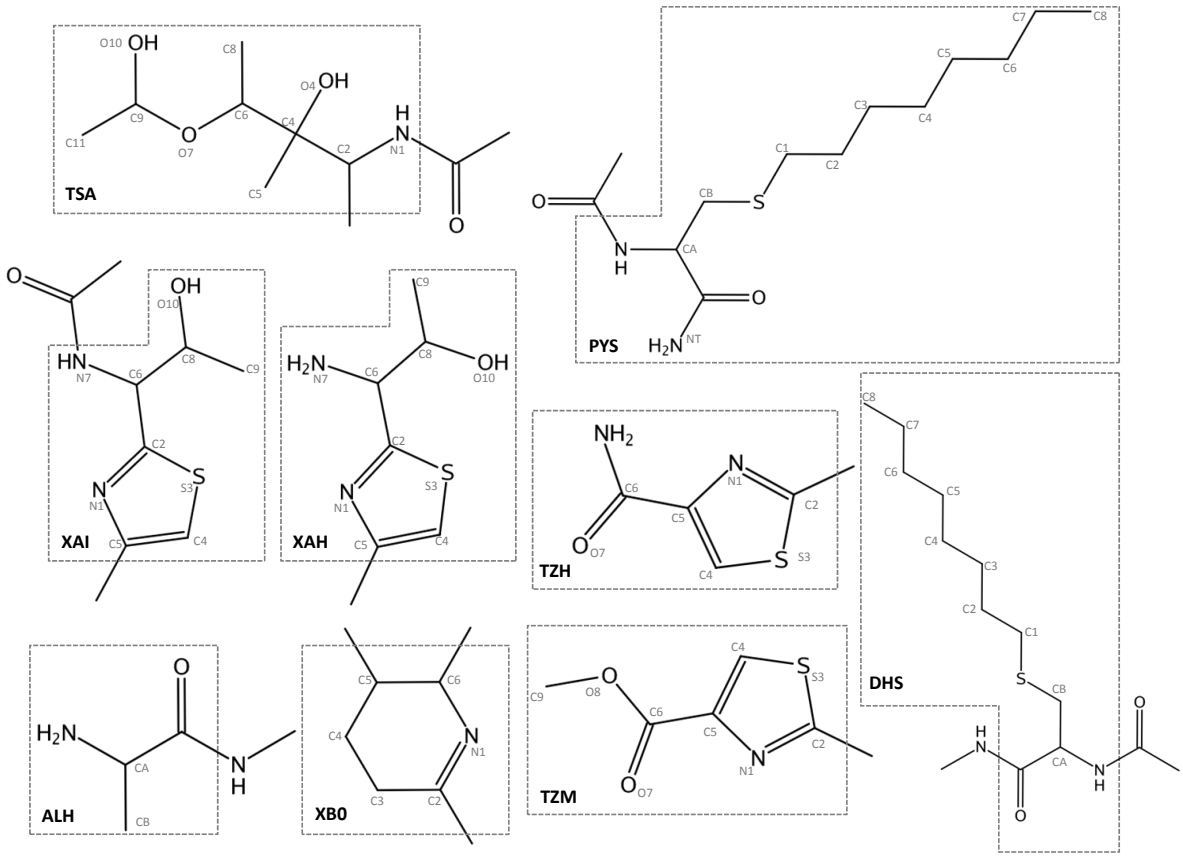


Figure 4.20.: Molecules used for deriving intramolecular parameters and RESP charges for additional non-standard residues contained in TS derivatives. Molecular fragments as found in the TS derivatives are surrounded by dashed lines.

Step 2: Docking

The docking studies were performed using AutoDock 4.2 [132] with docking parameters as described previously (Section 4.2). In short, the grid box was centered on the bound TS conformation with 70x70x70 grid points and the default spacing of 0.375 Å. The molecule input files (pdbqt) were prepared using AutoDockTools V1.5.4. 100 independent Lamarckian genetic algorithm runs, with $2.5 * 10^6$ energy evaluations, were performed for each ligand. Subsequently, the docking conformations were clustered using an RMSD threshold of 1.5 Å. The conformation with the lowest docking energy and the lowest energy conformation from the largest cluster (if different from the first) were subjected to MD simulations.

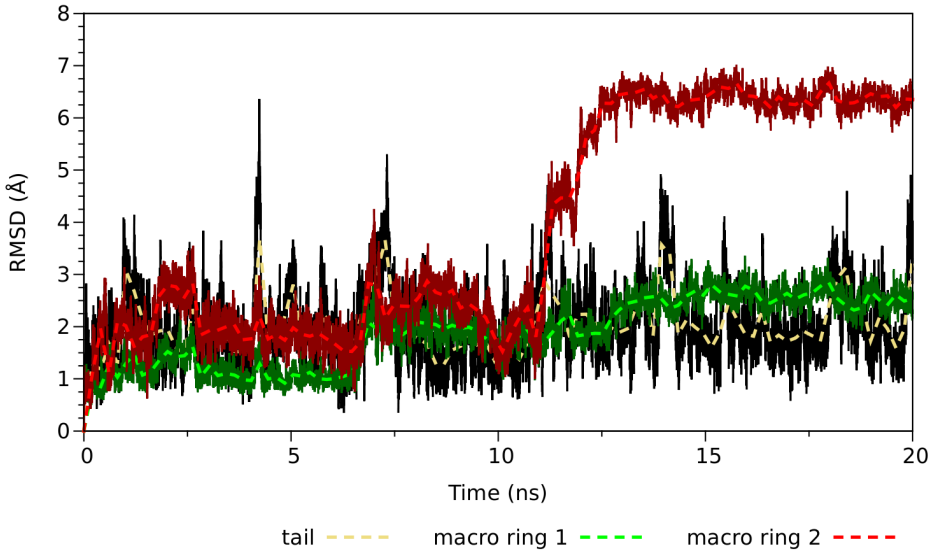


Figure 4.21.: Root-mean-square deviation (RMSD) from the input structure, assigned separately for compound **19**'s tail (yellow), macrocycle A (green), and macrocycle B (red) components, as a function of simulation time.

Step 3: Refinement of Docked Complexes and Free Energy Calculations

MD Simulations of Docked Complexes Each docked complex was neutralized using 50 sodium ions and solvated in a truncated octahedron of TIP3P waters with 15 Å padding in all directions. As the complexes consist of a relaxed receptor conformation and relaxed ligand structures form the previous simulations, the simulations were kept short. The system was first minimized with strong restraints on the solute followed by a second minimization without restraints. For heating the system to 300 K[5], a 20 ps constant volume simulation with weak restraints on the solute was performed, followed by a 50 ps constant pressure relaxation simulation without restraints. Finally, two 2100 ps production runs were conducted at constant pressure with different random number seeds. During the simulations mixed scaling factors for 1-4 interactions were used; 1.0 (vdW and electrostatic) for the ligand parameters from the thiopeptide force field and 1.2 (vdW)/2.0 (electrostatic) according to the Parm99SB [89] and the Parmbsc0 [90] force field, used for the protein and RNA, respectively.

MM-GB(PB)SA Calculations Molecular Mechanics Generalized Born (Poisson Boltzmann) Surface Area (MM-GB(PB)SA) computations were performed using the *MMPBSA.py* script of AmberTools 1.5. Explicit solvent and ions were stripped from the trajectory previous to the calculations, and equidistant snapshots were extracted for the ligand, the receptor and the complex from one trajectory. The electrostatic portion of the solvation free energy was calculated using the Poisson Boltzmann (PB)

[5]A temperature of 300 K instead of 310 K was chosen to have consistent conditions with the experimental affinity determinations for the MM-PB(GB)SA free energy calculations.

equation and by the generalized Born (GB) method. For the GB calculations, the method of Onufriev et al. [279] was used. Accordingly, the modified PBradii set 2 was used in LEaP when building the topolgy files. The PB equation was solved using the *pbsa* program of AmberTools. Internal and external dielectric constants were set to 1.0 and 80.0, respectively. The ionic strength was set according to experimental conditions (0.04 M). Non-polar solvation contributions were approximated by the *LCPO* method [223]. For the energy decomposition, 1-4 terms were added to internal potential terms. Entropic contributions were computed by a normal mode analysis in Generalized Born solvent, using every 200th snapshot of the first 2100 ps trajectory.

Evaluation Measures

In addition to the predictive index (PI), introduced in Section 4.2 (Eq. 4.1), the coefficient of determination R^2 was used to evaluate the performance of the native AutoDock score and the MM-PB(GB)SA calculations. For a set of ligands i with experimentally determined binding energies E_i and corresponding predicted binding energies P_i, R^2 is defined as:

$$R^2 = 1 - \frac{\sum_i (P_i - E_i)^2}{\sum_i (E_i - \overline{E_i})^2} \tag{4.3}$$

4.5.2. Results

Choice of Receptor Conformation for Docking

Docking to the MD-refined conformation gave an overall good convergence of docking poses: Out of the 100 docking poses generated for each ligand, a significant number was contained in the first (i.e. the lowest energy) cluster (see Table 4.5). In the majority of cases the first cluster was the largest cluster (i.e. containing the most poses) at the same time. The only exception is compound **5R** where the second cluster contains the largest number of poses (32). However, the observed binding mode and docking score (-8.42) are very similar to those of cluster 1. When the crystal coordinates of the 23S·L11 binding site were used as target structure, the best ranked docking solution for all compounds, except for TS, contained a reverted binding mode, where the molecule tail is inserted into the cleft formed by the ribosomal RNA and the protein. This binding mode was already observed when docking the NMR-determined structure of the epimer **20** (see Section 4.3), and is likely the result of a shape mismatch between the receptor and the ligand structure. When considering the largest cluster of docked conformations a TS-like binding mode is found for **17**, **18** and **5S** but not for **5R** (cluster 2: 9 solutions) and **19** (cluster 5: 1 solution).

Moreover, the experimental binding constants are better reproduced with the MD-refined receptor conformation (see Table 4.5). This holds not only for the coefficient of determination (R^2) and the predictive index (PI), but also for the absolute binding energies which are consistently lower and thus nearer to the experimental values. Using the MD-refined structure as docking target instead of the

Table 4.5.: Measured dissociation constants (K_D) and docking results of investigated TS derivatives, along with the predictive index (PI) and the coefficient of determination (R^2) of the AutoDock score for the two targets.

		X-ray		**MD-refined**	
Compound[a]	K_D [nM]	Score [kcal/mol]	#Poses	Score [kcal/mol]	#Poses
1	0.20 ± 0.05	-8.55	35	-10.36	68
17	1.80 ± 0.21	-6.87	1	-8.16	26
5R[b]	3.59 ± 1.20	-6.85	1	-8.46	23
5S[b]	5.41 ± 2.03	-5.58	1	-6.91	25
18	24.70 ± 2.86	-7.12	5	-7.60	82
19	670.00 ± 46.00	-6.00	1	-6.95	1
21	n.d.	-9.96	2	-11.00	28
PI		0.55		**0.71**	
R^2		0.43		**0.63**	

[a] See Figure 4.18.
[b] Different K_D values could be determined for the two diastereomers; however they could not be unamibiuosly assigned to the R or S stereoisomer. Therefore, the K_D constants were assigned according to the AutoDock ranking.

medium-resolution crystal structure led to a better convergence of native docking poses and a better reproduction of experimental binding affinities, thus validating the refinement approach.

Prediction of Binding Affinities by the MM-PB(GB)SA Method

The molecular mechanics Poisson-Boltzmann (or Generalized Born) surface area (MM-PBSA and MM-GBSA) method (cmp. Section 2.3.2) has become a popular method for calculating binding affinities of biomolecular complexes in recent years [181, 281, 282]. There is, however, no agreement on one protocol to perform this kind of computation. Studies on various systems have favored either the faster and more approximate generalized Born (GB) method [283], or the more rigorous Poisson-Boltzmann (PB) approach [284]. Approaches mainly differ in the number of MD snapshots used for the calculations. Using a single minimized complex [281, 284, 285] is as common as the use of several hundreds snapshots from nanoseconds MD simulations [182, 286, 287]. Divergent results were also observed when using different GB models and dielectric constants [283, 288].

Consequently, different protocols were tested for free energy computations with the PB and GB solvation model. Since for many dockings an obviously wrong binding mode was found when using the crystal structure target, these tests were performed using the MD-refined structure. Initially, two generalized Born parameter sets were tested: parameterization “I” and “II” of the model developed by Onufriev, Bashford and Case (GB^{OBC}) [279]. Overall, a slightly better performance was achieved using parameter set “II”, evidenced by an increase of the predictive index (PI) from -0.04 to 0.19 by changing the model from “I” to “II” (using the 2.1 ns trajectory for each complex). Therefore, this GB model was

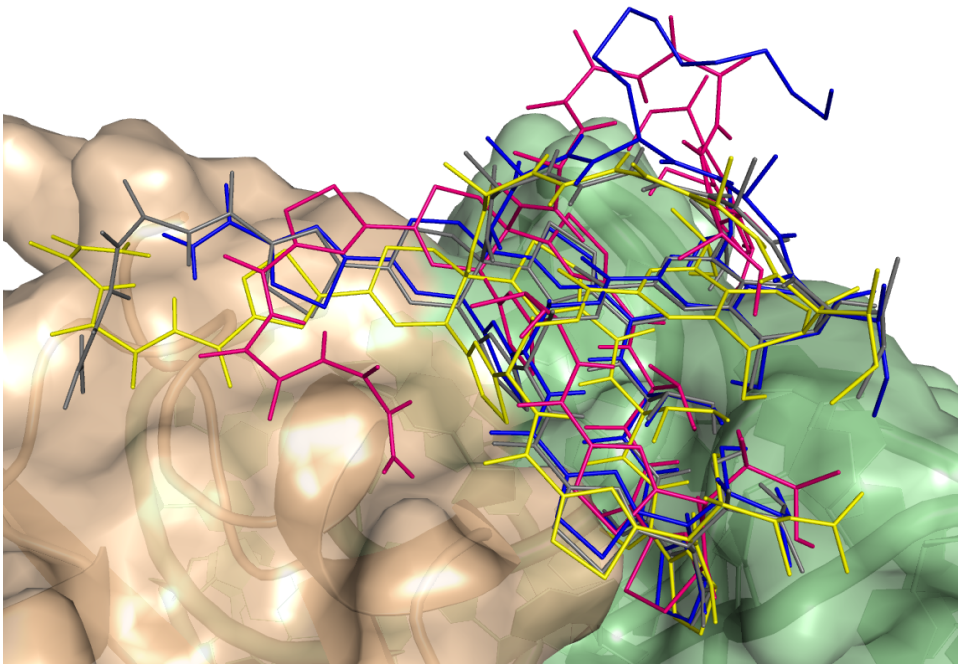


Figure 4.22.: Docking poses of **1** (gray), **18** (blue), **17** (yellow), and **19** (magenta).

selected for all further calculations. Secondly, computations were done using an increasing number of snapshots to test if this would level out inaccuracies: 1) one structure of the docked complex which was minimized in explicit solvent, 2) 100 snapshots taken from a short 200 ps MD simulation, 3) 105 snapshots taken from a 2.1 ns MD simulation, and 4) 210 snapshots from two separate 2.1 ns MD simulations. Table 4.6 lists the performance measures (PI and R^2) for different simulation lengths and number of snapshots for the GB and PB solvation model.

Table 4.6.: Predictive index and R^2 of MM-PB(GB)SA computations for different simulation lengths.

Simulation length (number of snapshots)	min (1)	0.2 ns (100)	2.1 ns (105)	2x 2.1 ns (210)
GB PI	-0.42	0.13	0.19	**0.29**
GB R^2	**0.23**	0.01	0.00	0.02
PB PI	0.57	0.66	**0.67**	0.56
PB R^2	0.34	0.21	0.26	**0.42**

Interestingly, the best values for the predictive index PI and the coefficient of determination R^2 do not always fall together, emphasizing the importance of not solely relying on one performance measure. Whereas the PI constantly rises with an increasing number of sapshots, the best correlation is found when using only the minimized complex. However, given the overall poor prediction accuracy of the GB implicit solvation model and the small number of ligands investigated, these numbers

should not be over-interpreted. The MM-PBSA protocol clearly outperformed MM-GBSA in all settings investigated. However, its best prediction accuracy (PI = 0.56 and R^2 = 0.42) is still below the performance of the AutoDock score (PI = 0.71 and R^2 = 0.63). Both implicit solvation methods showed the best balance between PI and R^2 for the largest number of snapshots (4th column in Table 4.6). However, given the only minor deviations to the performance when minimized complexes are used (first column in Table 4.6), it is questionable if the much longer computation time for running the MD simulations is worthwile.

Two additional MM-PB(GB)SA computation scenarios were tested. 1) Inclusion of entropy, calculated by a normal mode analysis using 11 snapshots of the first 2.1 ns MD trajectory, and 2) setting the solute's dielectric constant to 4 instead of 1 which was shown to be a better choice for ligand–nucleic acid complexes [289]. However, both scenarios rather worsened than improved the prediction accuracy (data not shown).

Stability and Decomposition Analysis

To gain a better understanding of the individual residue's contributions to the binding enthalpy, residue- and pairwise energy decompositions [290] were performed with the MM-GBSA approach. This protocol was chosen since a complete decomposition is currently only available with the GB model in AMBER. In the following the observations for each ligand are discussed.

Thiostrepton: When **1** was docked to the MD-refined structure, the binding mode found for TS in the 3CF5 crystal structure was consistently reproduced (see Fig. 4.22). 86% of all poses were contained in the first two clusters, exhibiting a similar docking score of -10.36 and -10.30 kcal/mol, respectively. When docked to the crystal structure target, the convergence of docking poses was worse but still acceptable: 44% of all poses were contained in the first two clusters, showing the X-ray binding mode. During the following 2.1 ns MD simulations of the complex, the position of **1** showed only slight deviations relative to its docked conformation (Fig. 4.23A, 2.14 and 1.00 Å final RMSD for the ligand), underlining the complex's stability.

When decomposing the binding enthalpy on a per-residue basis, only a few residues contributed significantly (Figs. 4.23B and 4.24). These are A1067 and A1095 on the RNA side, and Pro21, Pro25, and Gln29 of the α1 helix of L11. The importance of the two nucleobases for binding of **1** is well-known [47, 188]. Interestingly, the energetic contributions of the flanking nucleobases, except for the direct neighbours G1068 and A1096, are slightly positive. Although the vdW and especially the elecrostatic contributions of A1067 are higher than those of the other residues they are heavily decreased by the unvafourable polar solvation (Table 4.7). In contrast to A1067, the contributions of A1095 are dominated by vdW interactions, mediated by a stacking interaction with the TS residue TZO-1. Hydrophobic interactions occur between TZB-3 and Pro21 and the residues around the central dehydro-piperidine-core (XBB-12) with Pro25 (Fig. 4.23C). From the residue-wise decomposition of

TS (Figs. 4.23D and 4.24) the pivotal role of the TZB unit of macrocycle A is obvious. This part of **1** is deeply inserted into the cleft formed by the RNA and the protein and mediates contacts to both sides, A1067 and A1068 of the RNA as well as Pro21 and Pro22 of L11 (Fig. 4.24).

Table 4.7.: Enthalpic MM-GBSA contributions of key residues to binding of TS, separated as van der Waals (vdW), electrostatic (Elec.), polar solvation (Polar solv.) and non-polar solvation (Non-polar solv.) contributions in kcal/mol.

Residue	vdW	Elec.	Polar solv.	Non-polar solv.	**Total**
A1067	-7.60	-8.48	11.41	-0.99	-5.67
A1095	-6.07	-0.31	2.24	-0.50	-4.64
Pro21	-4.40	0.01	1.62	-0.71	-3.48
Pro25	-2.36	-0.11	0.55	-0.22	-2.14
Gln29	-3.69	-1.00	2.31	-0.32	-2.70

*Compound **5**:* Compound **5** was obtained by a thiol-Michael addition at the dehydro-alanine unit in the truncated tail. The stereochemistry of this group had a minor effect on the dissociation constant but a 10-fold difference in activity was found *in vivo* [250]. Consequently, both diastereomers were created. As can be seen in Figure 4.25, their binding modes are slightly divergent, especially regarding the tail. For both, the octyl unit is oriented towards the most hydrophobic area of the protein surface around Ile34. These hydrophobic interactions are visible in the increased energetic contribution of this residue in the decomposition analysis (-2.58 and -1.95 compared to -0.29 kcal/mol for **1**, Table 4.8). On the other hand, the contribution of the polar Gln29 plays a minor role. Due to steric restrictions, the hydrohobic interactions with Ile34 seem to be harder to accomplish for the S isomer whose interaction with A1095 is almost halved compared to TS. The MM-PBSA and MM-GBSA re-scorings both preferred the largest cluster solution of **5R** and assigned it a very similar binding energy like for the S stereomer (Table 4.8), consistent with the almost identical K_D values for the diastereomers (Table 4.5).

Table 4.8.: Enthalpic MM-GBSA contributions of key residues to binding, and overall binding enthalpy obtained by MM-GBSA and MM-PBSA computations for TS and its derivatives. All values are in kcal/mol.

Cmd.	A1067	G1068	A1095	Pro21	Pro25	Gln29	Ile34	GB	PB
1	-5.67	-1.34	-4.64	-3.48	-2.14	-2.70	-0.29	-60.41	-48.72
5R	-6.17	-0.68	-6.03	-3.39	-2.32	-1.17	-1.95	-68.22	-45.19
5S	-5.17	-1.06	-2.47	-4.58	-1.95	-1.63	-2.58	-62.77	-46.98
18	-3.99	-1.23	-6.60	-1.71	-1.75	-2.05	-0.06	-47.33	-36.34
17	-1.37	-0.02	-2.88	-3.71	-1.42	-2.09	-0.60	-39.92	-35.21
19	-2.78	-3.14	-5.24	-4.07	-1.73	-1.43	-0.60	-53.62	-35.41
21	-2.34	-1.59	-3.60	-4.43	-1.85	-1.45	-1.28	-48.49	-34.56

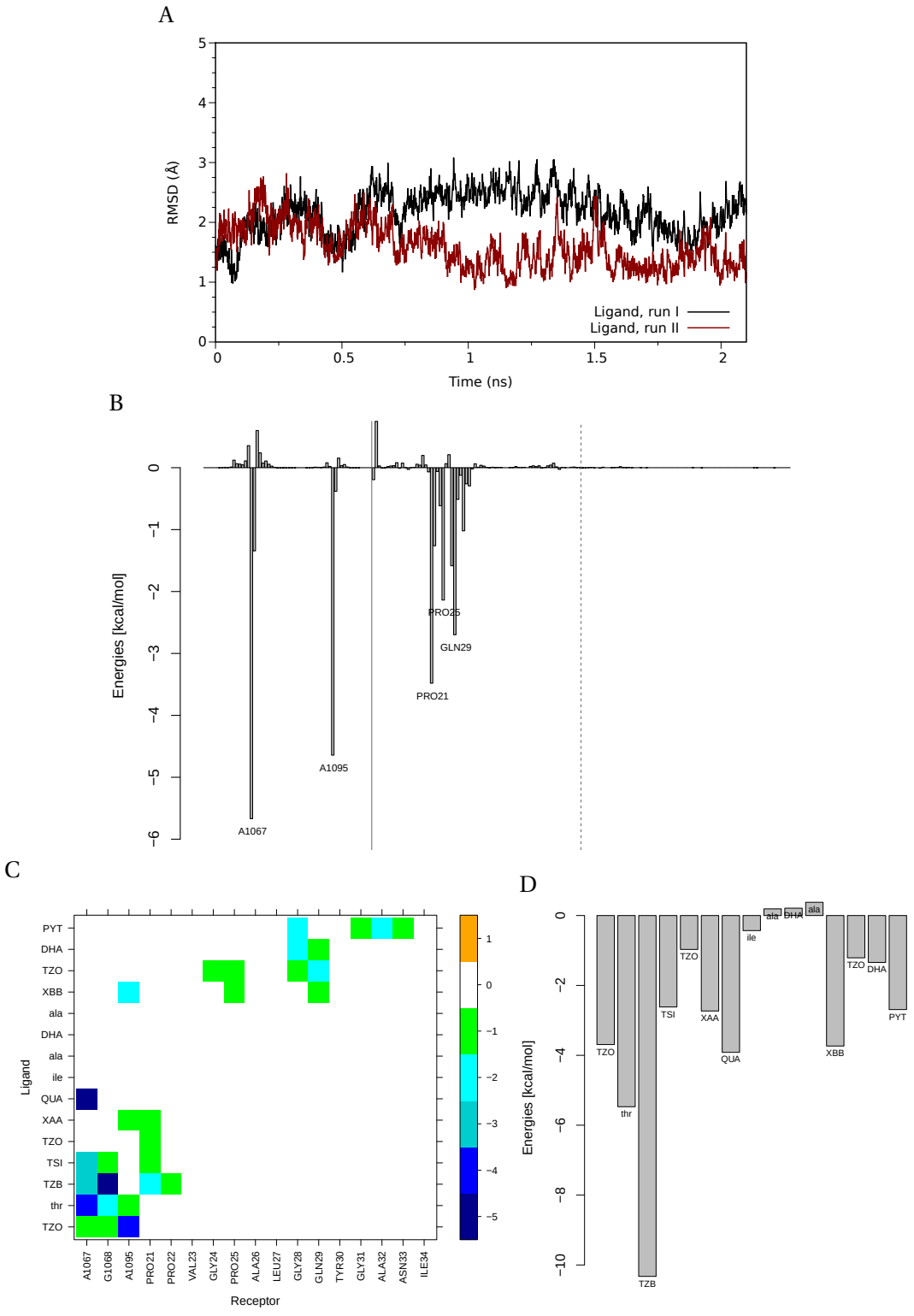


Figure 4.23.: MM-GBSA decomposition results for binding of TS. A) RMSD of TS (without hydrogens) to the docked structure in the two re-scoring MD simulations. B) Residue-wise decomposition 23S·L11. RNA and protein residues are separated by a black solid line, the two protein domains are further divided by a dashed line. C) Pair-wise decomposition energies in kcal/mol. D) Residue-wise decomposition ligand.

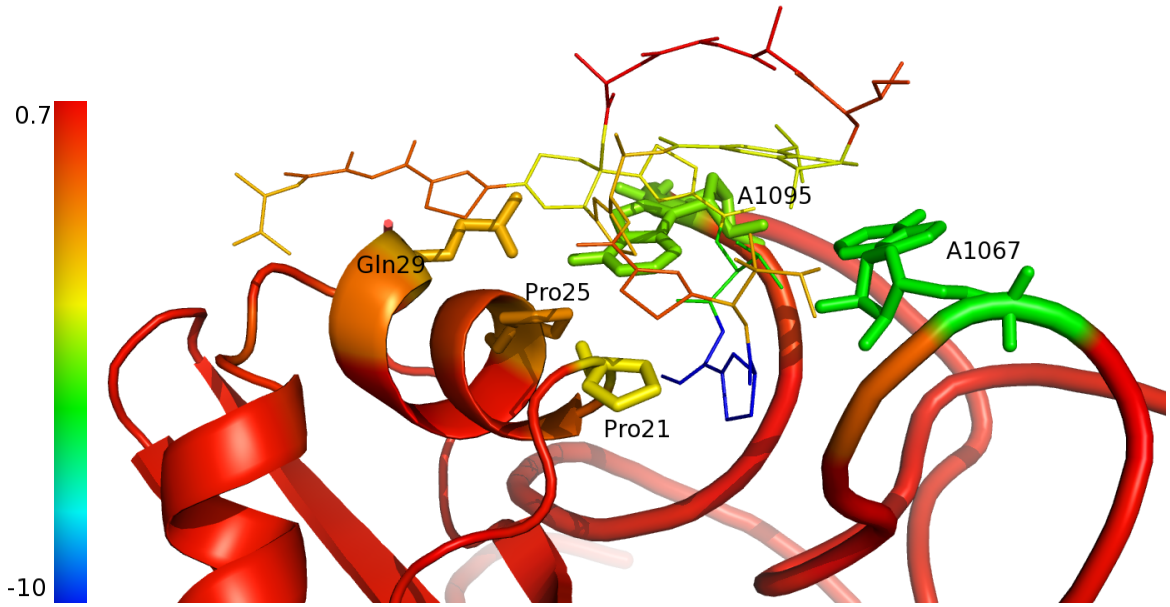


Figure 4.24.: MM-GBSA decomposition results mapped onto residues of the 23S·L11·TS docked complex. Significant binding site residues are shown in sticks representation. Color-coding according to enthalpies in kcal/mol (see color bar).

*Compound **18**:* To obtain compound **18**, the octyl unit was added to the dehydro-alanine contained in macrocycle B. Additionally, the tail was truncated by two units. These modifications led to a reduction of the tail's flexibility and to a flexibility increase of the residue containing the octyl unit (residue 10, Fig. 4.26). Consequently, this residue was excluded from the first MD trajectory's clustering, resulting in one large cluster whose representative structure was used as input conformation for docking. After clustering of docking poses, the first cluster contained 82 docking poses, even more than TS, but obtained a slightly worse docking score (-7.60 kcal/mol). The binding mode was very close to the one found for **1**, the octyl group sticking out into the solvent (Fig. 4.22). As suggested earlier [292], the unfavourable solvation of the hydrophobic octyl group is most likely the major reason for the worse binding affinity in comparison to TS. In the decomposition analysis (Table 4.8), the interactions with the key residues (A1067, A1095, Pro21, Pro25, and Gln29) are still present but, except for A1095, considerably reduced, explaining the lower predicted enthalpy for this compound.

*Compound **17**:* Compound **17** features a modification of macrocycle A's dihydroxyisoleucine side chain, obtained through an acetylation of the secondary OH group. Starting from a TS-like binding mode (Fig. 4.22) for **17**, a consistent shift of the docked position was noticed (RMSD ~4 Å) in both MD runs. The RNA's energetic contributions are considerably reduced compared to TS. For A1067 the lowest contribution among all derivatives was found. Particularly, the electrostatic contribution is reduced (-1.29 kcal/mol) compared to TS (-8.48 kcal/mol). The added acetyl group leads to a steric

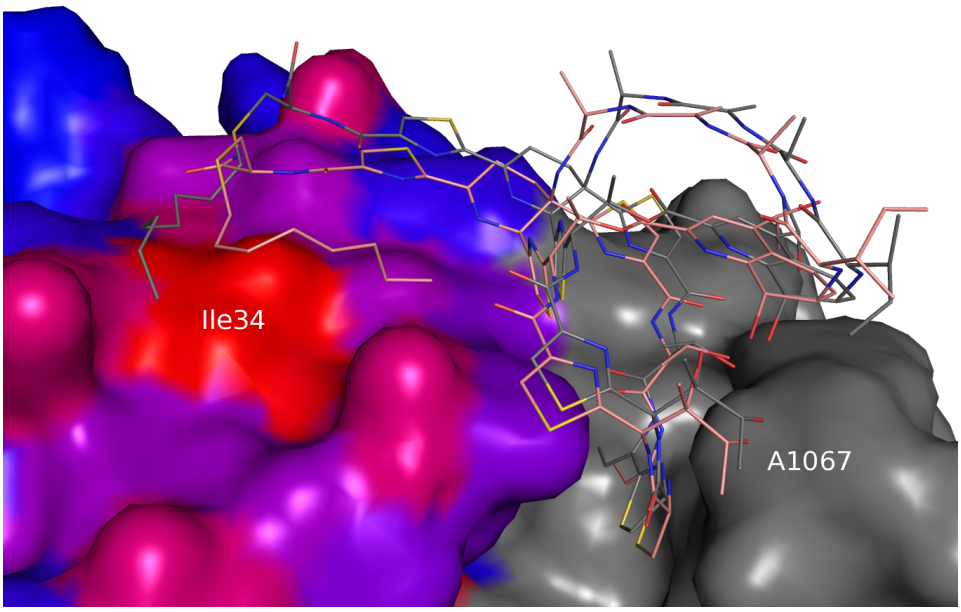


Figure 4.25.: Docking poses of stereoisomers of **5**. grey: **5S**, wheat: **5R**. RNA grey. Protein surface colored according to hydrophobicity [291] (blue: lowest hydrophobicity, red: highest hydrophobicity). Figure created using PyMOL with rTools plugin.

clash and a loss of polar interactions with this nucleobase. Most likely, this is the cause for the drift from the docking pose that was observed during the simulations.

*Compound **19**:* A hydrolysis of the macroring B lacton split the macrocycle between the residues XAA and QUA and led to derivative **19**. This modification represents the most extensive change of the TS scaffold. Not surprisingly, this was reflected by a significant RMSD increase of macrocycle B's atoms to their position in the TS template (Fig. 4.21), and led to increased fluctuations of macrocycle B's residues (Fig. 4.27). After ~15 ns the RMSD became stable and the average conformation of the last 5 ns was selected as docking input. The docking of **19** did not converge, i.e. 100 clusters containing one pose each were obtained. The molecule's many degrees of freedom (23 active torsions) are in the upper limit of AutoDock's search capabilities. Nevertheless, a binding mode similar to TS was found for the best scored pose (Fig. 4.22). Macroring A's position is similar to its position as found in TS, but a slight shift in residues THR-2 and TZB-3 is seen. The split macrocycle B's residues make additional contacts to the RNA. This slight shift of the binding pose led to a binding energy decrease (-2.78 kcal/mol) of A1067's contribution for the benefit of G1068's binding energy (-3.14 kcal/mol).

*TS fragment **21**:* The decomposition analysis of TS and its derivatives revealed that the second macrocycle hardly contributed to the binding enthalpy or even had a slight positive energy (see Fig. 4.23D, residues 7–11). It was, thus, investigated how a TS fragment, completely lacking the second macrocycle, would bind. To obtain the virtual fragment, macrocycle B was removed from the

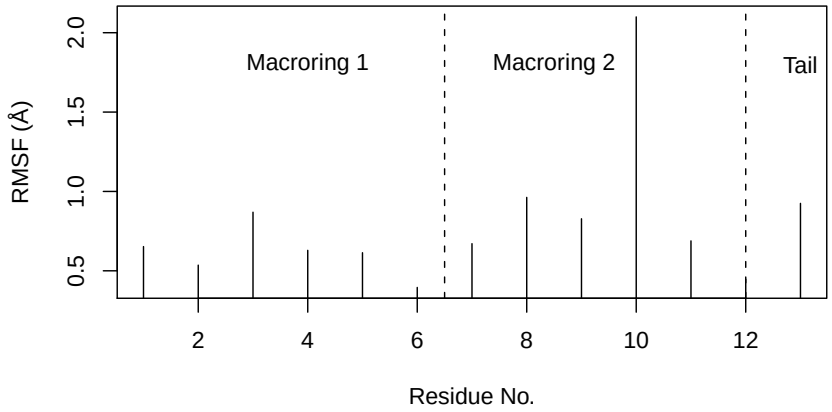


Figure 4.26.: Compound **18**'s residue-wise root-mean-square fluctuations as obtained from a 10 ns MD simulation in organic solvent.

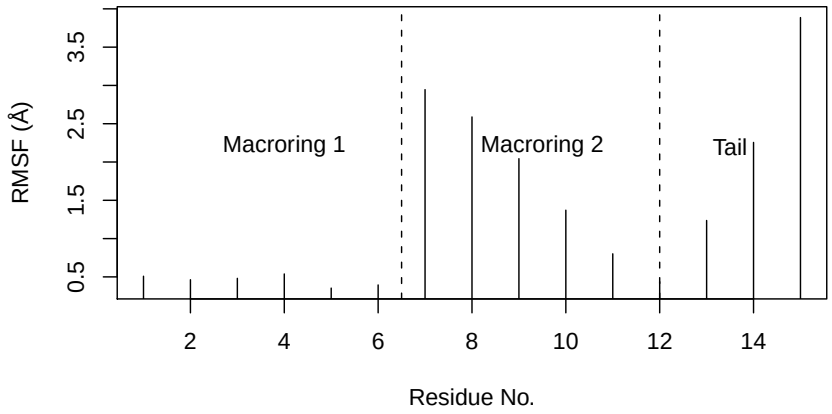


Figure 4.27.: Compound **19**'s residue-wise root-mean-square fluctuations as obtained from a 10 ns MD simulation in organic solvent.

TS structure and the resulting atomic vacancies were saturated with hydrogens (compound **21** in Fig. 4.18).

Although the simulation of **21** in organic solvent showed an increased flexibility of macrocycle A compared to TS, its overall conformation stayed similar to the TS crystal structure (~1 Å RMSD). Also in the docking experiments and subsequent MD simulations the binding mode was very similar to TS (Fig. 4.28). AutoDock even assigned a better score than for TS (Table 4.5), arising from a similar intermolecular energy but less penalty due to the lower number of rotatable bonds. The MM-PBSA re-scoring, missing such a penalty term, sets the binding enthalpy more in the region of the weaker binder **19**. Also on the enthalpic decomposition level, the contributions of the binding site's key residues were most similar to the ones found for compound **19** (Table 4.8).

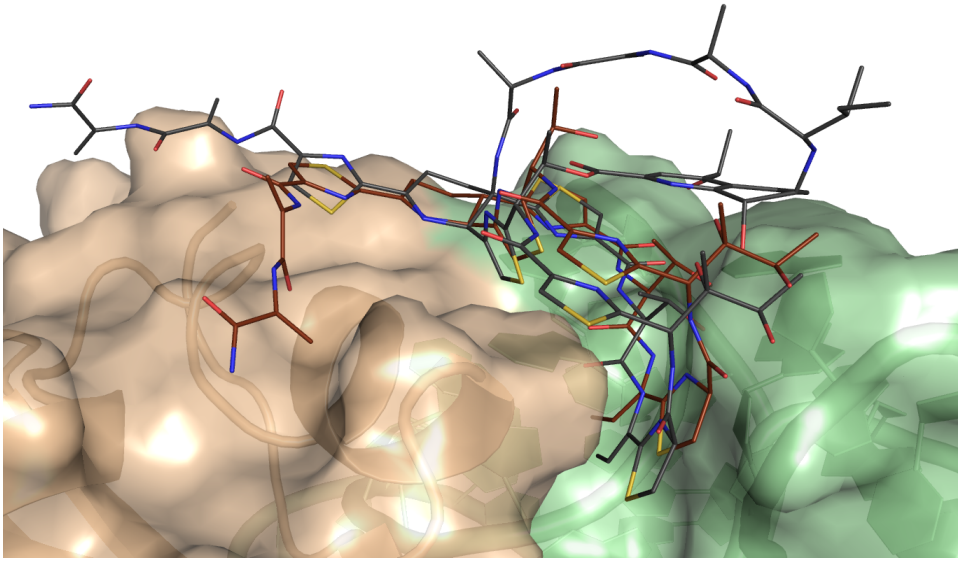


Figure 4.28.: Binding mode of TS (**1**, gray) and TS fragment (**21**, brown) after MD.

4.5.3. Discussion

A combined molecular dynamics and docking approach was used to describe the binding mode and molecular interactions of a number of thiostrepton derivatives with the 23S·L11 subunit of the bacterial ribosome. To generate input structures for docking, an approach was used that sequentially 1) uses the TS crystal structure as template, 2) refines the modified structures by an MD simulation using a thiopeptide-tailored force field, and 3) clusters the MD trajectory to obtain one (or several) representative conformation(s). Using the TS structure as a template is a valid approach as all investigated ligands are semi-synthetic derivatives from the same parent compound (TS). Therefore, it can be assumed that their 3D structures are similar. Considering the ligand structure refinement, the superiority of using an explicit-solvent MD simulation with a tailored force field compared to a short minimization *in vacuo* with a general force field (as done in Section 4.2) is shown by a lower prediction accuracy of the latter approach (PI = 0.65, R^2 = 0.23 compared to PI = 0.71, R^2 = 0.63 achieved here) under otherwise identical docking conditions (same target structure and docking parameters).

In contrast to previous studies [288, 293], a MM-PB(GB)SA re-scoring could not improve the native AutoDock score further. Almost no correlation was found between experimentally determined binding affinties and the binding energies calculated with the MM-GBSA approach. The more rigorous PBSA variant performed better (PI 0.67), coming close to AutoDock's prediction accuracy (PI 0.71). An explanation for AutoDock's better performance can be found in the high anti-correlation between the number of rotatable bonds and the binding affinity in the data set. This value directly contributes to the overall docking score as a rough measure for the loss of torsional entropy upon binding [153].

If only AutoDock's intermolecular energy (excluding the rotational term) is considered – which is a fairer comparison to the enthalpies computed by MM-PB(GB)SA – the predictive index becomes negative (-0.36). This indicates that entropic contributions are essential to the overall binding energy of the TS derivatives investigated. However, adding the entropic part of ligand binding – calculated by a normal mode analysis – led to a worse prediction (PI of 0.2 for MM-PBSA). It has been shown before that an inclusion of entropy does not always improve the prediction accuracy of the MM-PB(GB)SA approach [283]. Large fluctuations of the vibrational entropies are often observed with normal mode analysis [288, 293]. It can, thus, not be ruled out that the low number of snapshots used in the entropy calculations has led to inaccuracies.

Although the MM-PB(GB)SA re-scoring did not lead to a better overall affinity prediction, its energetic decomposition gave a clearer picture of the residues contributing to binding, both on the ligand and the receptor side. Among all derivatives, A1067, A1095 and Pro21 are the dominating residues of the rRNA·protein binding interface. For all ligands, by far the largest contribution to the overall binding enthalpy was seen for the TZB unit of macrocycle A which has contacts to the purines A1067 and G1068 on one side of the cleft and to prolines 21 and 22 on the protein side. In contrast, most of the second macrocycle's residues do not contribute or even show a positive contribution to the overall binding energy.

4.6. Conclusion

The thiopeptides represent a challenging compound class for docking programs, such as AutoDock, due to their molecular complexity, particularly the high number of rotatable bonds (14–31) and the contained macrocycles. Although the ligands are treated flexible in general, the conformations of large ring systems are not changed during the docking. A correct input structure is, therefore, crucial for obtaining correct placements. However, the three-dimensional structures of the derivates investigated here were largely unknown. It was shown that when either experimentally determined structures of thiopeptides are used or they are predicted by MD simulations with a tailored force field, binding mode and affinity predictions by the protein–ligand docking program AutoDock are possible with an acceptable reliability.

The conditions under which the 3CF5 X-ray structure was determined (preformed crystals of the complete 50S subunit soaked with a cryo-solution containing high concentrations of TS [32]) together with its medium resolution of 3.3 Å make it a non-ideal target structure for drug design purposes. Nevertheless, it is the only available crystal structure of TS bound to the GAR and provides valuable insights into the location and binding mode of the ligand. In this section, it was shown that docking TS-like ligands to a MD-refined receptor structure instead of the medium-resolution crystal leads to a better convergence of native docking poses and a better reproduction of experimental binding affinities. Similar observations were already made previously in this work (Section 4.2). However, all investigated ligands are structurally very similar to TS. Given the high adaptability of

the binding site, the suitability of a representative structure from an MD simulation with bound TS for docking structurally distant ligands remains currently unclear. Using an ensemble of different target conformations, for example the cluster representatives from the “apo” simulation, might be a promising alternative approach here.

A virtual TS fragment (**21**), lacking the second macrocycle, was found to exhibit a similar binding mode like TS’s macrocycle A, although being structurally more similar to the one-cyclic antibiotic micrococccin (**3**) for which a different binding mode was proposed [32]. Although apparently unimportant for binding, macrocycle B could be important for steric hindrance of incoming elongation factors, such as EF-G, and prevent them from proper binding to the GAR. Further experimental studies are necessary to verify this hypothesis. Nevertheless, fragment **21** could be a starting point of further structural exploitation for the design of novel, smaller antibiotics adressing the 23S·L11 ribosomal binding site.

5. A Pharmacophore Model for the 23S·L11·Thiostrepton Complex

5.1. Introduction

As already outlined in Chapter 1 thiopeptide antibiotics possess high potency but unwanted pharmacokinetic properties, such as their low solubility and bioavailability. For the design of new chemical scaffolds a pharmacophore model of TS, containing the essential features that determine the drug's biological activity, is therefore highly desired. The aforementioned low solubility of TS and the large size of the bacterial ribosome have long hampered obtaining high-resolution crystal structures of the complex and thus a precise understanding of TS's molecular interactions with the ribosome. Only recently in 2008, a crystal structure of the 50S ribosomal subunit with bound TS was published [32], but only in a medium-resolution range of 3–4 Å. Along with results from numerous biochemical studies [42, 45–47, 188, 189, 191, 208, 227, 229, 294–302], this structure enabled, for the first time, a clear positioning of TS into the ribosome's 23S·L11 subunit. Due to the limited electron density resolution, this structure still does not allow the study of atomic-resolution molecular interactions of the ternary complex. Therefore, a pharmacophore model is still missing.

In this chapter, the knowledge gained in the previous chapters is incorporated to design the first pharmacophore for thiostrepton's molecular interactions with the ribosomal GAR. The docking results from Chapter 4 have clearly shown the superiority of using an MD-refined complex instead of the medium-resolved crystal structure for drug design purposes. There is strong evidence that this will also hold for deriving a pharmacophore.

In its natural aquous environment, a macromolecular complex is characterized by an ensemble of thermodynamically accessible conformations. Therefore, crystal structure coordinates can give misleading static views of interactions [303]. One of the first approaches for generating a *dynamic* pharmacophore model was introduced by Carlson et al. for HIV-1 integrase [304]. In their work, Carlson at al. overlaid protein conformations from an MD simulation, and identified conserved binding regions for probe molecules to build a receptor-based pharmacophore for the integrase. In comparison with a static pharmacophore model and a pharmacophore derived from a set of active ligands, the dynamic pharmacophore model showed better performance in fitting known inhibitors, and was able to identify new inhibitors for HIV-1 integrase in a screening experiment.

Here, the idea of a dynamic pharmacophore is adopted, but implemented in a different manner. Similarly to Carlson et al., classical pharmacophore-development techniques are combined with molecular dynamics simulations. However, instead of using MD to generate an ensemble of binding pocket structures, an initial static 3D-pharmacophore model is built based on the MD-refined representative 23S·L11·TS complex. The stabilities of the suggested pharmacophoric features over a given MD simulation are then investigated to refine the initial pharmacophore model. Features that show high dynamic stability are more likely essential for binding, and are thus prioritized over features that are only visible in a limited number of snapshots. At the same time, new important interactions may appear that were not contained in the initial complex to build the pharmacophore. The complex-based (3D) pharmacophore approach is combined with observations from MD simulations for obtaining a time-resolved pharmacophore, termed "4D-pharmacophore".

5.2. Materials & Methods

Based on the unbound TS ligand, the MOE pharmacophore generator assigns 47 potential pharmacophoric features (PCH scheme, Fig. 5.1), far too many for a useful pharmacophore query. Most of these features are shared by the TS derivatives investigated in Chapter 4, so that the number of features can hardly be reduced by including them. Therefore, it was decided to derive a *complex-based* pharmacophore using LigandScout [187] that is currently the only available program for such an approach. Two initial complex-based pharmacophore models were built: 1) based on the energy-minmized 3CF5 crystal structure, and 2) based on the MD-refined structure. LigandScout supports the following pharmacophoric elements: hydrogen bond donor/acceptor, positive/negative ionizable area, hydrophobic interactions, aromatic ring, metal binding feature, and (optionally) excluded volume to define areas where ligand groups are sterically unfavorable.

LigandScout provided pharmacophoric observable features whose stability could subsequently be analyzed in the MD trajectory. The live time of the defined pharmacophoric features was analyzed using the 40 ns MD trajectory data of the 23S·L11·TS complex from Chapter 3. The dynamic stability of hydrogen bonds was investigated with the AMBER module PTRAJ, setting the heavy atoms distance cut-off to 3.5 Å and the angle cut-off between the acceptor, hydrogen, and donor atoms to 120°. The presence of stacking interactions was determined by monitoring the distance between the ring centers (centroids) and the angle between the axes perpendicular to the ring plane (surface normal vectors) over simulation time (see Fig. 5.2 for a geometrical description). A parallel displaced stacking interaction is characterized by a surface normal vectors angle of ~0/180°, and an interplanar separation of 3.4–3.6 Å [303, 305]. LigandScout defines a tolerance range of 2.8–4.5 Å for this distance.

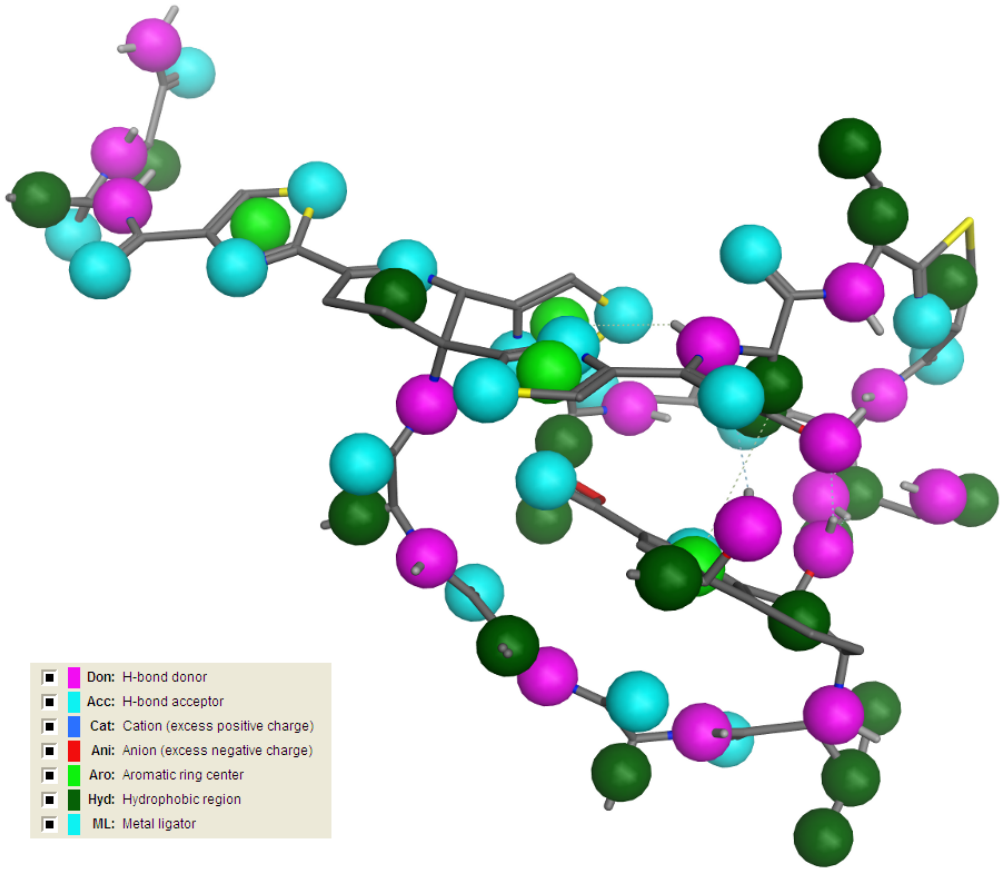


Figure 5.1.: Pharmacophoric features of unbound thiostrepton. Created using the MOE software with the PCH pharmacophore scheme.

5.3. Initial Pharmacophore based on the 23S·L11·Thiostrepton Complex

Figure 5.3 depicts the initially derived pharmacophore model for the 3CF5 crystal structure in a two- (A) and three-dimensional (B) representation. LigandScout assigned six pharmacophoric features to the TS structure: two hydrogen bond donors, and four hydrophobic centers. One hydrogen bond donor is assigned to the hydroxyl group of the quinaldic acid moiety in macrocycle B, interacting with nucleotide A1067's base (N3). The second H-bond donor is located in the tail, contacting L11's Ala32. Given the high number of possible hydrogen bond donors and acceptors (cmp. Fig. 5.1), the small number of actual hydrogen bonds formed with the 23S·L11 subunit is astonishing and cannot explain the strong binding. However, as a static view of the complex is considered, some hydrogen bonds might be missed. Two of the hydrophobic centers are located at macrocycle A's thiazoles (residues 5 and 6) that are in proximity of Pro21 and Pro25. The two remaining hydrophobic feature were found in TS's tail, at the thiazole and the second dehydroalanine unit.

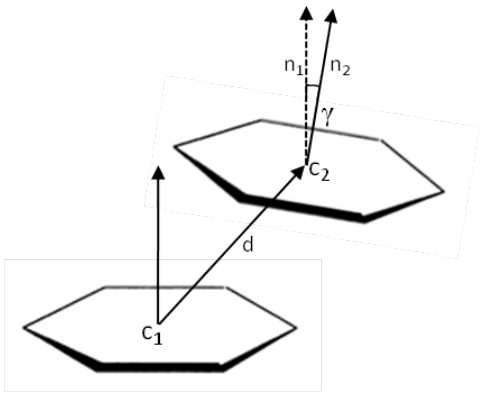


Figure 5.2.: Essential geometrical features used to describe stacking interactions between aromatic ring systems. d: distance between ring centroids c_1 and c_2; γ: angle between ring normals n_1 and n_2.

When the pharmacophore was built based on the MD-refined structure, eight features were assigned (Fig. 5.4). The four hydophobic features found in the first model are preserved, except for the thiazole unit in the tail. Instead, a fourth hydrophobic center is located at TS tail's first dehydroalanine. Considerable disagreement is seen in the assignment of hydrogen bonds. Two hydrogen bond donors are assigned to TZB-3's amide and to one hydroxyl group of the neighboring TSI-4. The hydroxyl group is marked as a double-feature by LigandScout, meaning that the oxygen can alternatively function as a hydrogen bond acceptor (green/red arrow in Fig. 5.4). Both features are found in TS's macrocycle A, interacting with nucleotide A1067's sugar moiety. Similarly to the first pharmacophore, another hydrogen bond donor is located in the tail, but here it is found between the first amide and L11's Gly28. Especially prominent is the stacking interaction between the thiazole of TZO-1 and base A1095 that was missed in the crystal structure pharmacophore. All other thiazoles present in macrocycle A are assigned hydrophobic centers, mediating interactions to the nucleotides G1068 and A1095 and the protein residues Pro21, Pro25 and Gly28.

In contrast to the crystal structure pharmacophore, macrocycle B completely lacks any pharmacophoric features. A considerable reorientation towards the solvent was observed early in the MD simulation, bringing the quinaldic acid moiety out of reach for an interaction with A1067's base. This reorientation of macrocylce B could be a result of the imperfect Parm99SB/GAFF force field that was used to model TS (cmp. Chapter 3). To test this hypothesis, the MD simulation with the optimized thiopeptide parameters of the docked TS·MD-refined complex (cmp. Section 4.5) was extended from 2.1 ns to 10 ns. No such reorientation was found during this simulation; the second macrocycle remained fairly stable in its initial position. The rest of the TS structure and also the binding region were almost exactly as found in the MD-refined representative structure.

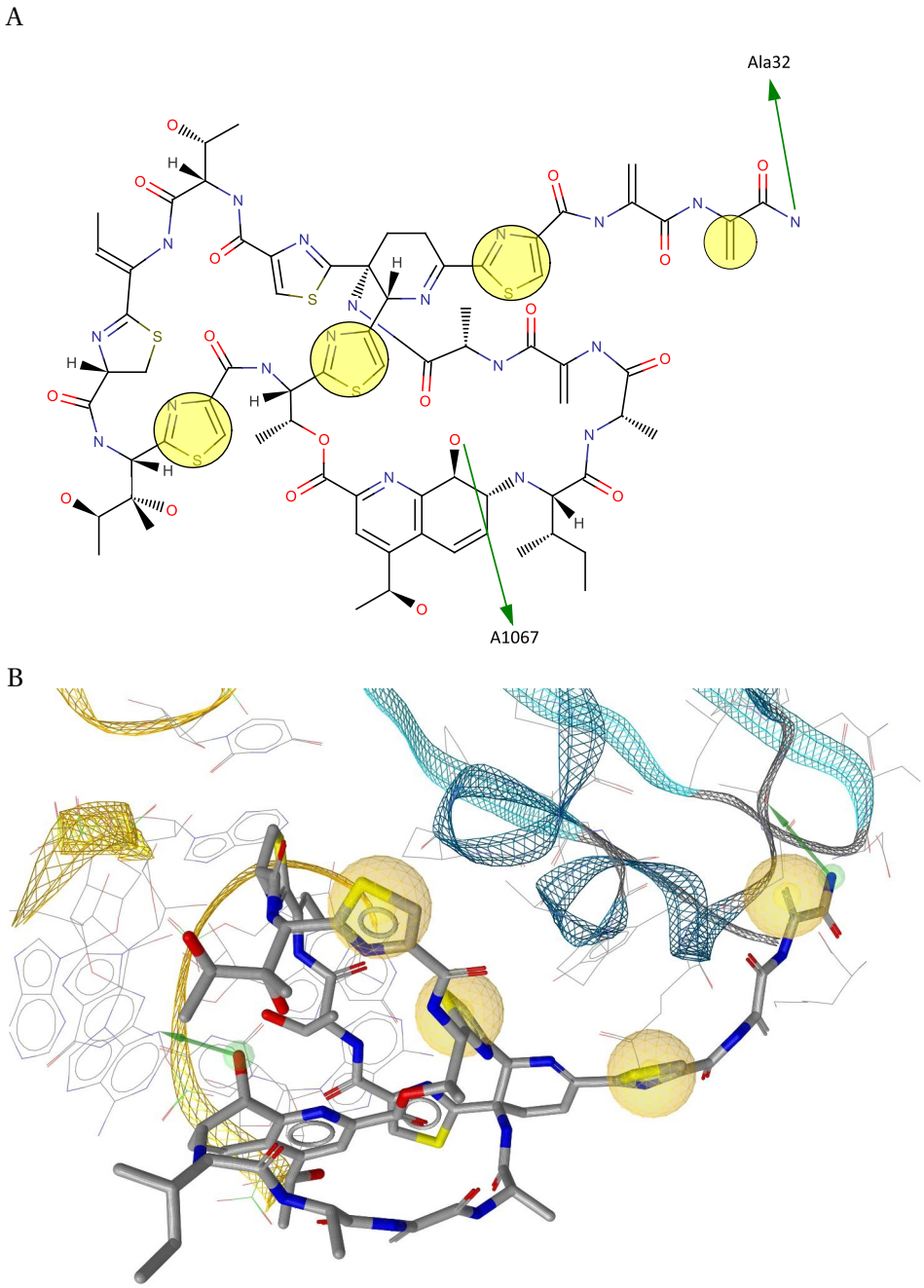


Figure 5.3.: Initial pharmacophore of the 23S·L11·TS complex based on the energy-minimized 3CF5 crystal structure built with LigandScout in a 2-dimensional (A) and 3-dimensional (B) depiction. Hydrophobic interactions are displayed as yellow circles/balls, and hydrogen bonds donors as green arrows.

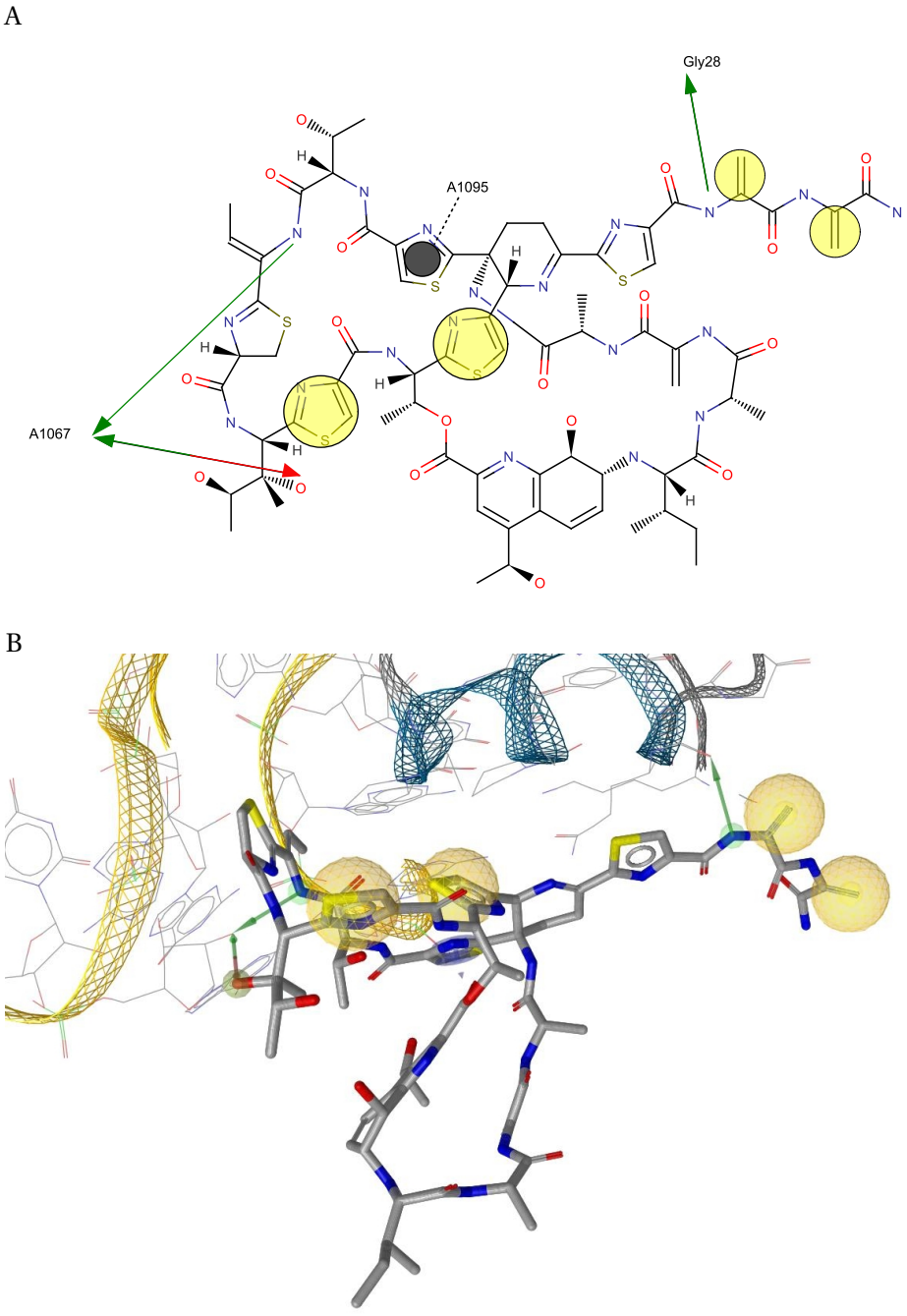


Figure 5.4.: Initial pharmacophore of the 23S·L11·TS complex based on the MD-refined structure built with LigandScout. A) 2D depiction; B) 3D depiction. Stacking interactions are displayed as grey circles, hydrophobic interactions as yellow circles/balls, hydrogen bonds donors and acceptors as green and red arrows, respectively.

5.4. Inclusion of Time – Development of a 4D-Pharmacophore

In a second step, the pharmacophoric features defined by LigandScout in both initial static pharmacophore models (3CF5 crystal coordinates and MD-refined complex) were investigated with respect to their live time in the MD simulation of the 23S·L11·TS complex (Chapter 3). Since only hydrogen bonds and stacking interactions possess clear geometrical constraints, analysis was focussed on these two interaction types.

Hydrogen Bonds

Seven hydrogen bonds showed occupancies of at least 5 % during the MD simulation (see the PTRAJ output on page 153 in the appendix and Fig. 5.5 for a structural visualization of the five most occupied H-bonds). Six of these H-bonds involve A1067, and are formed either with its ribose (2'O) or its base (N3 and to a lesser extend N7). Predominantly, these groups within A1067 are found to be H-bond acceptors. For A1067's base, the corresponding donor within TS switches during the simulation: in the beginning it is the NH-group of TZB-3 (14.3 % occupancy), after approximately 13 ns the NH-group of TSI-4 (20.4 % occupancy). After A1067's base adopts a *syn* conformation late in the simulation (after ~27 ns) this H-bond is formed with N7 at the opposite side of the base instead of N3 (cmp. Fig. 3.27 on page 69). Hydrogen bonding involving the ribose is more pronounced early in the simulation – as an acceptor for TSI-4's NH- and OH-groups (21 % and 6.6 % occupancy). The close distance to TS's thiazoline also supports a hydrogen bond between the ribose's hydroxyl group and the thiazoline's nitrogen (TZB-3@N4, 12.7 % occupancy). The H-bond showing the highest stability is formed between the NH-group of TZO-13 in TS's tail and GLY28 in the proline-rich helix of L11-NTD. It is present over the entire simulation and seems to fix the position of the tail.

The hydrogen bonding features found by LigandScout for the MD-refined structure are represented by occupancies of 14.7 % (tail NH – Gly28), 6.6 % (TSI-4's OH – A1067 2'O), 3.4 % (A1067 2'OH – TSI-4's O), and 2.8 % (TZB-3's NH – A1067 2'O). The two hydrogen bonds observed in the cystal structure pharmacophore model (Fig. 5.3) did not emerge in the analysis of the 40 ns MD trajectory. During the MD simulation of the docked complex with the optimized parameters the hydrogen bond between QUA's C8-hydroxyl group and A1067's N3 was occasionally formed (17.7 % occupancy). A stronger H-bond partner was, however, found in the threonin's hydroxyl group (83.9 % occupancy). Therefore, an enthalpic contribution of ring B's quinaldic acid moiety to binding cannot be ruled out. Still, the presence of hydrogen bonds between A1067 and macrocyle A is much higher.

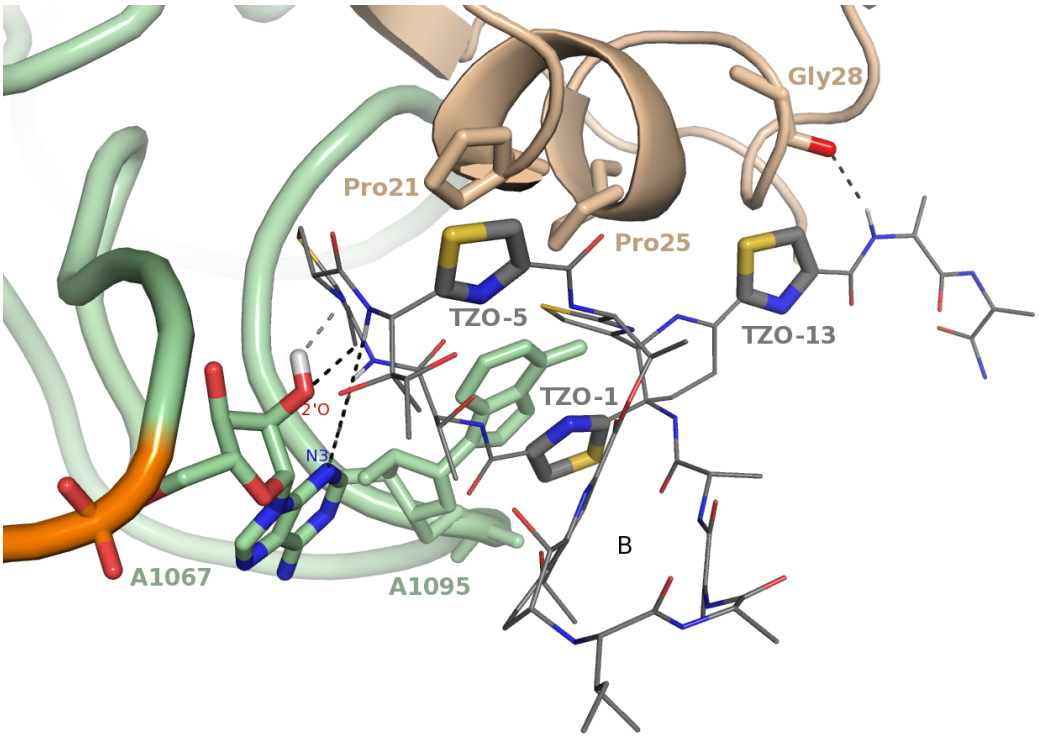


Figure 5.5.: Hydrogen bonds and possible stacking interactions between thiostrepton and the 23S·L11 binding site. Only hydrogens involved in H-bonds are shown. Important thiazole units and residues are shown in sticks representation.

Stacking Interactions

Harms et al. [32] defined "stacking" in a very generic sense and describe stacking interactions between 1) TZO-1 and the base of A1095, 2) TZO-5 and Pro21, and 3) TZO-13 and Pro25[1] (see Fig. 5.5). Therefore, these three potential stacking interactions were investigated in the MD simulation.

From the distance and angle distributions (Fig. 5.6) it becomes clear that a tight stacking interaction only occurred for TZO-1 – A1095(Fig. 5.6A). This stacking was also identified by LigandScout for the MD-refined structure (Fig. 5.4) but missed for the crystal coordinates. During the complete simulation the centroid distance was stable at approximately 3.8 Å, very close to the optimal distance (3.4–3.6 Å). Although this distance increased slightly to approximately 4 Å during the last half of the simulation, it did not loose its stability. The angle between the ring normals is between 0 and 20° in more than 80% of the simulation, meaning that the ring planes are always nearly parallel. This was not the case for TS's other thiazole interactions with the proline residues. Although the distance and angle between TZO-13 and Pro25 was also very stable through the whole simulation, the angle between the plane normals is too high (~50°) to stack upon each other. Moreover, the rings are moving apart shortly after

[1] Note that in the original publication [32] the proline residues are referred to as Pro22 and Pro26 as the *E. coli* sequence numbering is used.

the start of the simulation. The centroid distance equilibrates at a value of 5.6 Å, almost 2 Å further apart than in the crystal structure (see Fig. 5.6B). Considering Pro21, a large distance increase to TZO-5's ring cenroid can be seen (Figs. 5.6C). This distance increase was caused by a shift of helix α1's C-terminal end away from TS after ~16 ns simulation. This large distance between the rings definitely excludes a stacking interaction with this proline residue.

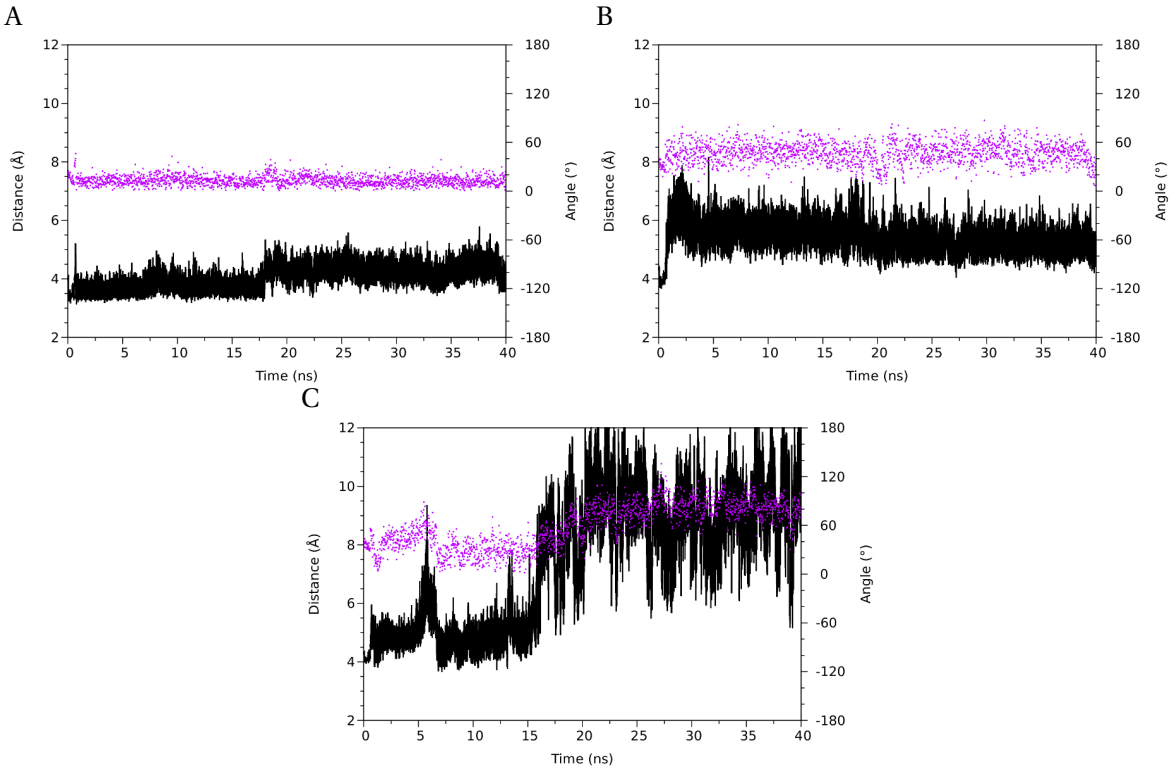


Figure 5.6.: Centroid–centroid distances (black, left axis) and angles between the ring normals (magenta, right axis) over simulation time for possible stacking interactions. A) TZO-1 – A1095. B) TZO-13 – Pro25. C) TZO-5 – Pro21.

5.5. Discussion

Complex-based pharmacophore models of the 23S·L11·TS complex were built whose features were investigated with respect to their dynamic stability in a molecular dynamics simulation. Three pharmacophoric features are especially noteworthy due to their dynamic stability: the extraordinarily stable stacking interaction between the TZO-1 thiazole and A1095's base, a hydrogen bond formed between TS tail's first DHA unit and L11' α1 helix backbone, and a complex hydrogen bond network between TS macrocycle A's residues 3/4 and A1067. These features were not found in the static view of the crystal structure, and their impact became only clear when their dynamics were analysed.

The stacking interaction with A1095 was consistently reproduced in the second MD simulation

(Fig. C.6, appendix) and the docking experiments, not only for thiostrepton but for most of its derivatives (cmp. Fig. 4.10 in Section 4.3). This fact, together with its enormous stability during the MD simulation and A1095's large energetic (particularly vdW) contribution to the overall binding affinity (see Section 4.5), render this interaction a crucial feature of TS's binding to the ribosome. The importance of this interaction is further underlined by the ~10-fold reduction of TS's binding affinity upon mutation of this nucleotide [191]. All other thiazole units are involved in hydrophobic interactions with the binding site (particularly Pro21 and Pro25) but their overall contribution is minor compared to TZO-1. Likely, they contribute as a summation of small hydrophobic effects.

The second crucial nucleotide is A1067 which is involved in several hydrogen bonds with TS's macrocycle A. TS has several functional groups that could form hydrogen bonds with A1067, located in the side chains and the amide backbone. Interestingly, the majority of them were observed to function as hydrogen bond donors. Hydrogen bonds were detected between the hydroxyl groups of THR-2 and TSI-4's side chains and A1067's sugar moiety (cmp. Fig. 4.10) in the docked complex of the NMR TS structure and the 23S·L11 crystal's coordinates. In the representative MD-refined structure hydrogen bonds were found between TZB-3's amide nitrogen and TSI-4's hydroxyl group and A1067's 2'OH. During the MD simulation, the nucleotide's 2'OH group and N3 are constantly involved in hydrogen bonds but with varying partners in the TS structure, mainly the amide nitrogens of TZB-3 and TSI-4, and TSI-4's first hydroxyl group. An acetylation of TSI-4 hydroxyl moiety (compound **17**) caused a large decrease of A1067's binding enthaply contribution in the MM-GBSA energetic decomposition, most likely as a consequence from the loss of one hydrogen bond donor plus steric hindrance of the larger side chain. This led to an overall reduction of the binding affinity and a loss of stability of the docking pose during the subsequent MD simulation (Section 4.5). The importance of TS's hydrogen bonding network with A1067 as a pharmacophoric featue is an explanation for the very effective natural resistance mechanism of its producer organim. *Streptomyces azureus* possesses a resistence gene, coding for a methyltransferase that catalyzes a selective 2'OH-methylation of A1067's ribose [306–308]. This resistence mechanism renders *S. azureus* highly insensitive to TS. Recently a ~3000-fold affinity decrease was reported [191]. Figure 5.7 depicts a structural model of A1067 2'O-methylation based on the MD-refined 23S·L11·TS complex. From this simple model it becomes obvious that the sugar methylation leads to 1) a steric clash with TS's binding position and 2) a disruption of the H-bonding network involving 2'OH. The high occupancy of A1067 N3's hydrogen bond is in compliance with the detected ~1000-fold affinity decreases for the transverse mutants A1067C and A1067U [191]. This high affinity loss was not seen for a transition within the purine family (A1067G), still featuring a nitrogen at this position. In summary, the hydrogen bonding network formed with A1067's base and sugar is an essential pharmacophoric feature but changing partners were observed during the MD simulation. As outlined in Chapter 3, the hydrogen bonds formed with A1067 might not only be necessary to fix TS in the binding site but might also be an important part of TS's mode of action by stabilizing A1067's base conformation and thus impede its interaction with EF-G.

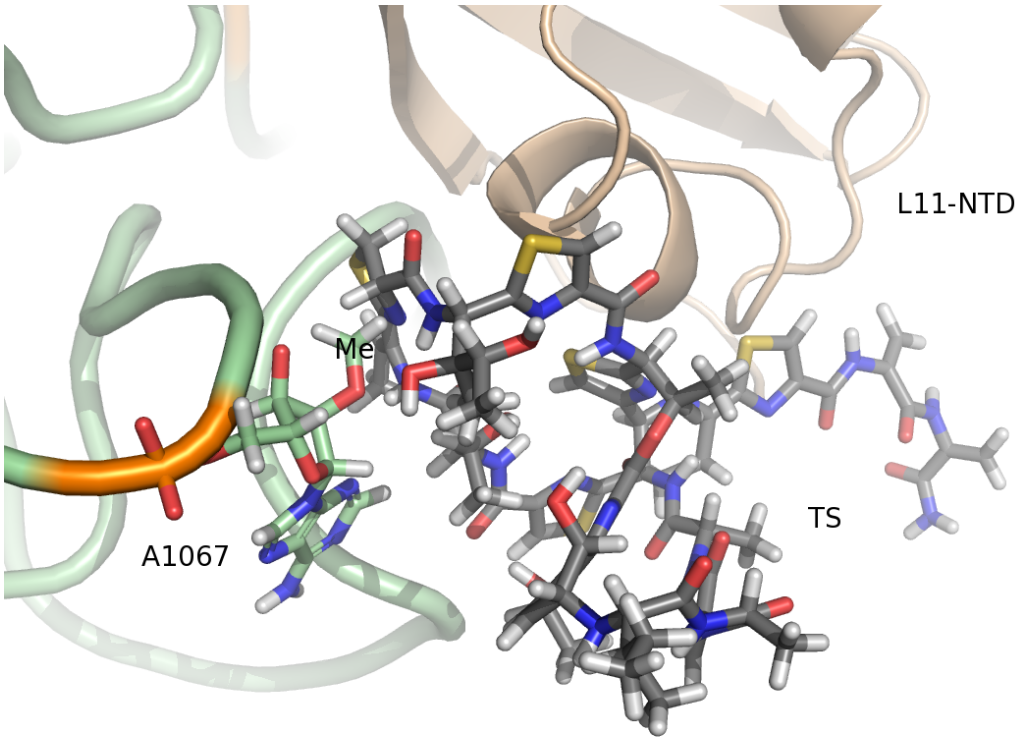


Figure 5.7.: Model of A1067 2'O-methylation based on the MD-refined 23S·L11·TS complex. L11 is colored in wheat, 23S rRNA is colored in palegreen. TS and A1067 are shown in sticks representation, the methyl group (Me) being labeled.

Another hydrogen bond donor was assigned to the first dehydroalanine unit in TS's tail, interacting with Gly28's backbone. This hydrogen bond showed enormous stability, being present throughout the whole 40 ns MD simulation. A hydrogen bond facility might therefore be another essential pharmacophoric feature in that region of the TS structure. Given the high flexibility and adaptability of the tail (see Section 4.4) and no significant enthalpic contributions from L11 residues in this area (see Section 4.5), it is not astonishing that TS was tolerant against truncation of its tail, and modestly tolerant against its modifications [250]. In the docking studies of the TS derivatives, a variety of different tail orientations along L11's proline-rich helix was found (cmp. Sections 4.3 and 4.5). All hydrogen bonds of the tail occured exclusively with L11's backbone. This could explain why different point mutations in this area hardly affect TS binding [191]. Interestingly, all TS derivatives still possessed the hydrogen bond donor in question. It seems likely that this hydrogen bond contributes to the stabilizing effect of TS on L11's N-terminal domain. An experimental test of a derivative lacking any hydrogen bonding feature in its tail could clarify if a hydrogen bond between TS's tail and L11 is essential.

The exact role of TS's macrocyle B for the antibiotic's inhibitory mechanism is still unclear. Except for the residue containing the quinaldic acid moiety (QUA-7), little enthalpic contributions of its

residues were computed (Section 4.5). A virtual TS fragment lacking macrocycle B (compound **21**) was found to have similar interactions like the complete TS molecule. Although the second macrocycle obviously has little contribution to binding, it can not be ruled out that it is necessary to sterically block incoming translation factors. In a study published 1999, Rodnina et al. concluded that TS does not interfere with factor binding, but rather inhibits the function of EF-G in subsequent steps [274]. A more recent study by Walter at al. contradicts these results [192]. They found that TS abrogates stable binding of EF-G and EF-4 to the ribosome, supported by the finding that an EF-G truncation variant lacking domains IV and V was insensitive to the presence of TS. The experimental testing of fragment **21** could help to clarify this issue. If it is found to have a similar binding affinity like TS but shows no effect on ribosomal translation, the steric blocking of translation factors by TS's macrocycle B will be confirmed as an important mechanimsm of TS's mode of action and has to be taken into account when new inhibitors are designed.

5.6. Conclusion and Outlook

In this chapter, the first pharmacophore model for TS binding to the ribosomal GAR was presented. Two initial three-dimensional static models were derived from the 3CF5 crystal structure and a refined 23S·L11·TS complex. The concept of a four-dimensional pharmacophore model was introduced that, in addition to a three-dimensional model based on a static complex, additionally accounts for the dynamic stability of molecular interactions as the fourth dimension. Three key pharmacophoric features within the TS structure were identified: 1) an aromatic ring to form a stacking interaction with A1095's base, 2) a hydrogen bond donor interacting with L11's α1 helix backbone, and 3) several functional groups to form hydrogen bonds with A1067. The significance of these features was revealed by monitoring their dynamic stability in an MD simulation and by discussing them in the light of recent mutation data.

The presented pharmacophore model provides a mean to screen for chemical fragments that fulfill the pharmacophoric restraints and thus potentially exhibit a similar mechanism like TS. Future synthesis and experimental testing of compounds will have to prove the usefulness of this model and will help to further refine it.

6. Summary and Outlook

6.1. Summary and Conclusion

The current work was initiated to obtain an atomistic understanding of the interactions of small molecules with biomacromolecular interfaces and for developing computer-based approaches to address these interfaces as non-traditional drug targets. Focussing on a protein–RNA interface of therapeutic importance, the specific aim of this thesis was to provide a detailed understanding of the influence that binding of thiopeptide antibiotics has on the dynamics of the 23S·L11 subunit and to elucidate the structural characteristics of this interaction to derive implications for the rational design of new compounds. The specific challenges associated with the L11·23S binding site, a protein–RNA interface at the bacterial ribosome, and the size and structural complexity of the thiopeptide ligands, did not allow for a straightforward application of established computer-aided drug design methods. These challenges were addressed by combining different computational chemistry approaches at different levels of granularity, ranging from quantum mechanics for deriving optimized intramolecular parameters and partial atomic charges for the thiopeptide compounds, to molecular dynamics simulations accounting for the binding site's flexibility, to molecular docking studies for predicting the binding modes of thiostrepton derivatives. Additionally, all computational results within this thesis are compared, validated and discussed with experimental studies.

In Chapter 3 the influence of thiostrepton, the best-studied thiopeptide representative, on the dynamics of the GAR was presented. Therefore, the results of two all-atom molecular dynamics simulations – one performed with the bound ligand and one without it – were compared. By creating an advanced and highly integrated target model, it could be shown that thiostrepton has a direct impact on the protein–RNA interface dynamics. Specifically, it restricts the conformational flexibility of the nearby N-terminal domain, and has a weak dynamic coupling to the distant C-terminal domain of the ribosomal protein L11. On the structural level, thiostrepton binding stabilizes the secondary structure elements of its binding area on the L11 protein (α1) and characteristically reprograms the conformation (χ-angle and sugar puckering) of nearby nucleotides, particularly the apical nucleotide A1067, which mediates contacts to tRNA and incoming ribosomal factors. A combined principal component and cluster analysis identified distinct conformations of the far more flexible "apo" form of the N-terminal domain that primarily differ in its orientation with respect to the 23S rRNA and the size of the cleft formed between the two. The different cluster identified may reflect distinct

interaction states with translation factors. The results of this study revealed a detailed understanding of the effect of thiostrepton binding at a previously unmet resolution, and provided a validated model basis for the structure-activity relationship and pharmacophore investigations addressed in Chapters 4 and 5. The observations made here, coupled with previous experimental knowledge [32], strongly support the hypothesis that restricting the binding site's conformational flexibility is an important component of thiostrepton's inhibitory mechanism. The observed extraordinary rigidity of thiostrepton's macrocyclic scaffold might be a structural asset to achieve this conformational restriction.

This thesis represents the first computational structure-based drug design attempt at a protein–RNA interface. It was, thus, not known beforehand if state-of-the-art structure-based design programs, developed for traditional targets, could be used for such an approach. Therefore, Chapter 4 starts with an evaluation of several established docking programs and an RNA scoring function with respect to their performance at protein–RNA composite environments. It could be shown that the protein–ligand docking program AutoDock achieved both, a convincing redocking and reasonable affinity prediction accuracy that could neither be outperformed by a combined RNA/protein scoring approach nor by molecular-mechanics-based free energy calculations. Around the docking engine AutoDock a MD-docking-MD workflow was established that allows a precise prediction of thiopeptide ligands' binding modes and affinities. Initial MD simulations are used to refine the coordinates of the binding site and to generate accurate input conformations of the ligands. Across all docking programs and ligands considered in this work, an MD-refined structural model of the binding site as target strongly improved the docking results. This implies that using molecular dynamics simulations is a promising strategy to overcome the inadequacies or limitations of experimentally determined threedimensional structures as drug targets.

Subsequent MD simulations of the docked complexes were used to test the stability of the ligands' binding poses and to decompose the enthalpic contributions of single residues to the overall binding affinity. With the help of this decomposition analysis crucial residues of the binding site and important moieties within the ligand structures could be identified that are in agreement with mutation studies. This knowledge was incorporated into a pharmacophore model, developed in Chapter 5. By investigating the dynamic stability of characteristic thiostrepton–GAR interactions, the complex thiostrepton scaffold could be reduced to a low number of significant pharmacophoric features.

The results of this thesis revealed, for the first time, the mode of action of thiopeptide antibiotics down to the details of their pharmacologically relevant parts and provide a framework for the *in silico* design of new ligands.

6.2. Future Work

The current thesis can be understood as a first detailed study into understanding and developing computer-based approaches for non-traditional drug targets, like the protein–RNA interface of the

bacterial GAR. Of course, open questions remain or came up in the course of the work on this thesis. As outlined in Chapter 5, the exact role of thiostrepton's second macrocycle is still unclear. Evidently, it has little enthalpic contribution to binding. Its opening or absence also did not lead to a significant destabilization of macrocycle A (compounds **19** and **21**). Based on the current state of knowledge, its most probable role is the steric blocking of translation factors as an enhancement of protein translation inhibition. The testing of this hypothesis is out of scope of the methods applied within this thesis. The chemical synthesis and experimental testing of the TS fragment **21** could enlighten this question.

With respect to the design of new chemical scaffolds targeting the L11·23S, the next step would be to use the derived pharmacophore model to screen for new chemical scaffolds. Due to the large spatial distance of the pharmacophoric features a fragment-based strategy seems promising. Here, different fragments that fulfill parts of the pharmacophore can be connected by linker fragments to design a larger compound matching the complete pharmacophore. By using the established MD-docking-MD workflow it can be subsequently tested if this compound meets the required properties, binding with high affinity and restricting the binding site's conformational flexibility.

This work can also be considered in a wider context: as a contribution to the design of modulators of biomacromolecular interfaces. For decades drug design has primarily focused on small molecules that bind to well-formed tight binding pockets, such as the catalytic centers of enzymes. Recently, there is increasing interest to design active compounds that disrupt or stabilize biomacromolecular interfaces (e.g. protein–protein, protein–DNA, protein–RNA, protein–lipid interfaces) to interfere with cellular pathways [59, 309]. However, investigating these systems using computational chemistry is still in its infancy. It is ancipitated that a more automated and streamlined use of molecular dynamics simulations in combination with established drug design approaches, and thinking beyond "drug-like" molecules will help to successfully address this research area.

6.3. Publications

1. Wolf, A., and Kirschner, K. N. (2012) Principal Component and Clustering Analysis on Molecular Dynamics Data of the Ribosomal L11·23S Subdomain. *J. Mol. Mod.* submitted. (Chapter 3)

   Content: This study investigates how clustering different principal component subspaces from molecular dynamics trajectory data effects the resulting clusters. The results demonstrate that by combining principal component analysis with subsequent data clustering researchers can obtain valuable dynamic and conformational information concerning their system.

   Significance: This is the first study that systematically explores the combination of principal component analysis with different clustering techniques for revealing conformational metastates in molecular dynamics simulation data. In times of improved hardware and software that allows simulations in the ns to μs time scale producing an enormous amount of data

such validations of quantitave analysis techniques become increasingly important. Moreover, the described methodology is established as a major analysis technique for investigating new modulators of the L11·23S subunit.

2. Wolf, A., Baumann, S., Arndt, H.-D., and Kirschner, K. N. (2012) The influence of thiostrepton binding on the dynamics of the ribosomal L11·23S subunit as revealed by molecular dynamics simulations. *PLoS ONE* submitted. (Chapter 3)

   Content: In this study it is explored how thiostrepton affects the dynamics of their ribosomal binding region at a previously unmet resolution. The findings of this work, coupled with previous experimental knowledge, strongly support the hypothesis that restricting the binding site's conformational flexibility is an important component of thiostrepton's mode of action which has to be taken into account when new inhibitors are designed.

   Significance: This is the foundation paper for the future work at designing new ligands for the L11·23S subunit. The data and analysis done within this work will serve as a guide and benchmark for comparison of future designed ligands.

3. Wolf, A., Reith, D., and Kirschner, K. N. (2011) Thiopeptide Antibiotics and the Ribosomal 23S·L11 Subunit – a Challenging Use Case for Semi-automatic Force-field Development. *From Computational Biophysics to Systems Biology*, 20–22 July, Jülich, Germany. (Chapter 4)

   Content: This paper describes the development of unique force-field parameters for thiopeptide antibiotics using quantum mechanical data. These parameters are validated by a molecular dynamics simulation in an organic solvent mixture against NMR structural data.

   Significance: The limited accuracy of force fields is seen as one of the major drawbacks in molecular mechanics applications. Although considered important, force-field parameterization is often not addressed as rigorously as in this work because of the huge effort that is necessary to obtain accurate parameters. The customized thiopeptide force field obtained from this work is compatible with existing protein/nucleic acid force fields and forms the basis to model future derivatives and their interactions with the L11·23S subunit.

4. Jonker, H. R. A., Baumann, S., Wolf, A., Schoof, S., Hiller, F., Schulte, K. W., Kirschner, K. N., Schwalbe, H., and Arndt, H.-D. (2011) NMR structures of thiostrepton derivatives for characterization of the ribosomal binding site. *Angew. Chem. Int. Ed. Engl.* **50(14)**, 3308–3312. (Chapter 4)

   Content: The activity of thiostrepton and derivatives with targeted shape changes was determined at their ribosomal binding site by using semisynthesis, NMR structure determination, docking, and biological evaluation in an integrated fashion. These data define structural boundaries, wherein an improvement of the overall pharmacological profile of these compounds or their analogues will be possible.

Significance: This paper demonstrates the sucessfull coupling of theory and experimental structure data, and validates the theoretical model.

Appendix

A. Acknowledgements

I would like to thank the people whose support has been crucial for the creation of this thesis. First of all, I would like to thank my doctoral advisor Prof. Dr. Martin Hofmann-Apitius for the opportunity to conduct my research within the Bioinformatics department of the Fraunhofer Institute SCAI. He made it as easy as possible for me to combine work with my family commitments. Furthermore, I would like to thank Prof. Dr. Michael Wiese for his willingness to be the co-referent of my thesis.

My special thanks go to my project's supervisor Dr. Karl N. Kirscher, one of the most enthusiastic scientists I have ever met. Karl, thank you for everything I could learn from you. Your support, guidance and critical feedback helped me a lot to become a more mature scientist. I would also like to thank Prof. Dr. Hans-Dieter Arndt from the Max Planck Institute for Molecular Physiology in Dortmund and his former Ph.D. students Dr. Sascha Baumann and Dr. Sebastian Schoof for the excellent collaboration within the ribosome project and many fruitful discussions.

My colleagues at the Bioinformatics department and the Computational Chemistry Engineering group at SCAI created a pleasant and creative working atmosphere. I would like to thank Dr. Astrid Maaß and Dr. Marc Zimmermann for proof-reading parts of this thesis. Special thanks go to Dr. Roman Klinger for being a companion and discussion partner along the long Ph.D. road. I would also like to thank Jasmin Zohren for keeping me free during the final stage of writing. Thanks for babysitting my daughter; it was a pleasure to share the office with you.

Without the support of my family this work would not have been possible. I am grateful to my parents who stood behind me in all my decisions and to my husband Stefan for his patience, support and love.

B. Structure and Dynamics of RNA

Nucleic acids are long, thread-like polymers, made up of a linear array of monomers called nucleotides. All nucleotides are constructed from three components: a nitrogen heterocyclic base, a pentose sugar, and a phosphate residue. The major bases are monocyclic pyrimidines (adenine (A) and guanine (G)) or bicyclic purines (cytosine (C), thymine (T, DNA only) and uracil (U, RNA only)). In ribonucleic acid (RNA), the pentose is D-ribose (in contrast to desoxyribonucleic acid (DNA) which is composed of 2-desoxy-D-ribose).

Sugar Pucker

The out of plane twisting of the sugar furanose ring is called 'puckering'. In the C2'-endo conformation the endo displacement of C2' is greater than the exo displacement of C3'. Of a C3'-endo conformation of the sugar pucker is spoken in the other case.

A common way to describe the sugar pucker is the concept of *pseudorotation*, that was first introduced by Kirkpatrick et al. for cyclopentane [310], and further developed by Altona and Sundaralingam for nucleotides [225]. The pseudorotational phase angle P is calculated by the five torsion angles of the furanose ring $\theta_0 - \theta_4$[1] (see Fig. B.1):

$$\tan P = \frac{(\theta_4 + \theta_1) - (\theta_3 + \theta_0)}{2\theta_2(\sin 36 + \sin 72)} \tag{B.1}$$

Values for P between 0 and 36° correspond to the C3'-endo, 144 to 180° to the C2'-endo conformation (Fig. B.2).

Conformation of the Nucleobase

There exist two distinct conformations of the bases with respect to the sugar, described by the torsion angle χ of the glycosylic bond. Nucleotides have a general preference for the *anti*-conformers of the glycosylic bond. Pyrimidines occupy a narrow range of *anti* conformations while purines are found in a wider range of *anti* conformations which can even extend into the high-*anti* range (see Fig. B.3).

RNA has greater structural diversity than DNA which is reflected in the functional diversity of RNA species. The structural elements that make up folded RNA are hairpin loops, interior loops, bulges, junctions, and other motifs [312]. The secondary structure elements of the 23S rRNA segment

[1] Note that the formula used here differs from the original one described by Altona and Sundaralingam due to their different numbering of the furanose torsion angles. [225, 311]

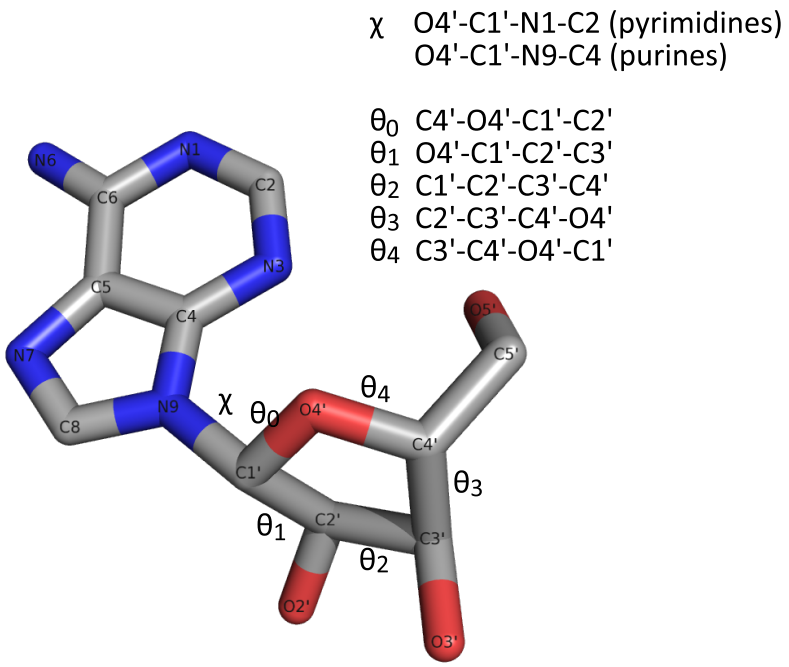


Figure B.1.: Definition of selected torsion angles found in nucleotides.

considered in this thesis is shown as a two-dimensional depiction in Figure B.4.

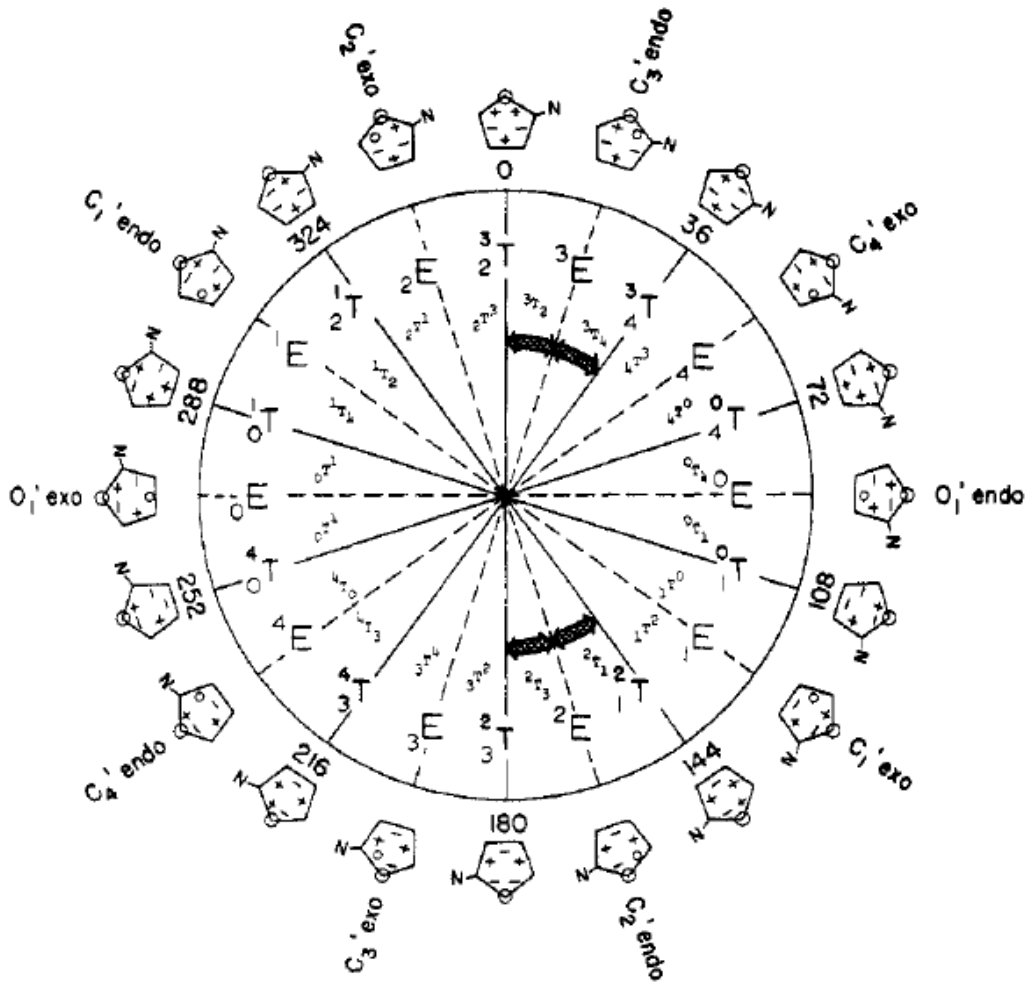


Figure B.2.: Pseudorotational cycle. The furanose ring with corresponding signs of the ring torsion angles is shown at the outer margin. Bold arrows indicate preferred pseudorotational regions for nucleotides. Taken from Ref. 225.

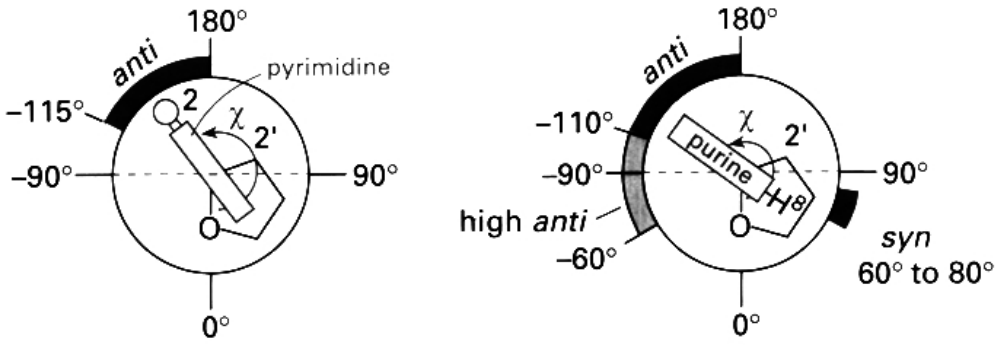


Figure B.3.: *Anti* and *syn* conformational ranges for glycosylic bonds in pyrimidines (left) and purines (right), taken from [233].

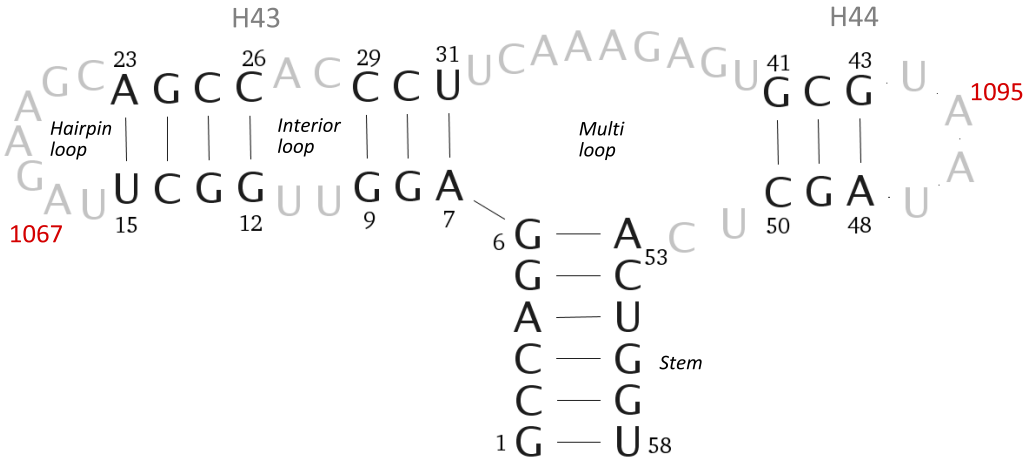


Figure B.4.: 2D representation of the H43/H44 23S rRNA segment's secondary structure (from *D. radiodurans*, PDB 3CF5 [32]). The crucial nucleotides A1067 and A1095 are assigned, otherwise the numbering corresponds to the nucleotides' positions in the PDB file (i.e. 1050 has to be added). Secondary structure generated with RNAView [313].

C. Additional Tables and Figures

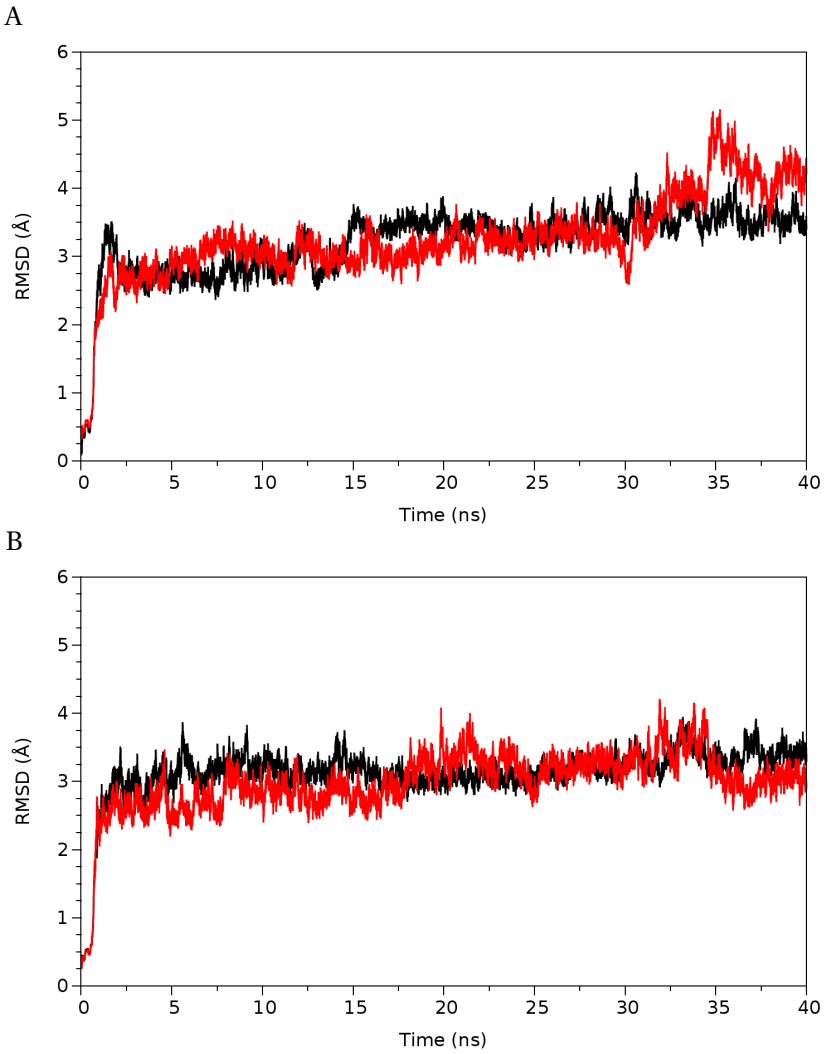


Figure C.1.: Backbone root-mean-squared deviations (RMSD) for simulations without TS (A) and with bound TS (B). Black line: run I; red line: run II.

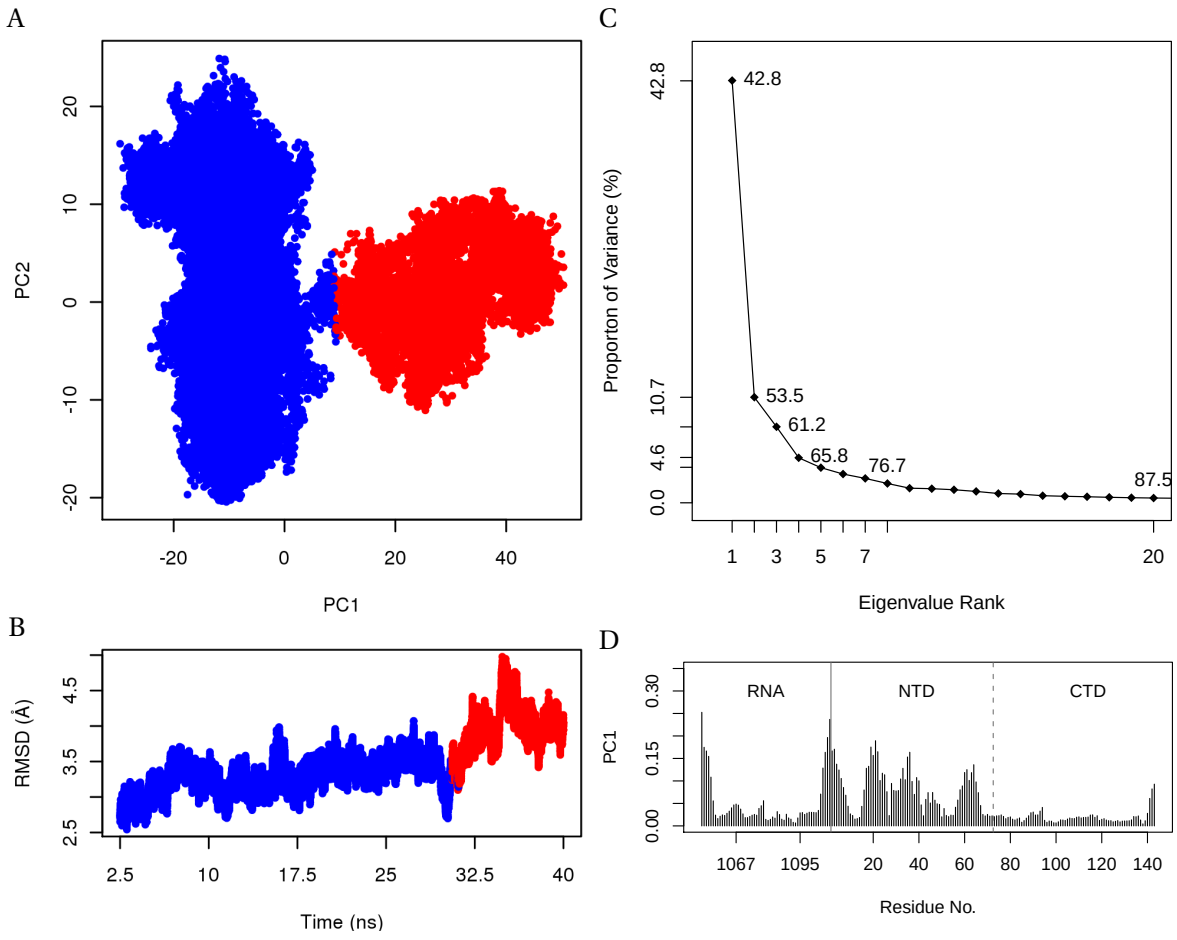


Figure C.2.: Principal component analysis (PCA) for the second simulation without TS. A) Clustering in supspace spanned by first three PCs projected onto 2D plane defined by PC1–2. Clusters 1, and 2 are colored as blue, and red. B) RMSD plot color-coded according to clusters. C) Scree plot. D) Residue-wise loadings (contributions) in Å to the first principal component.

A

B

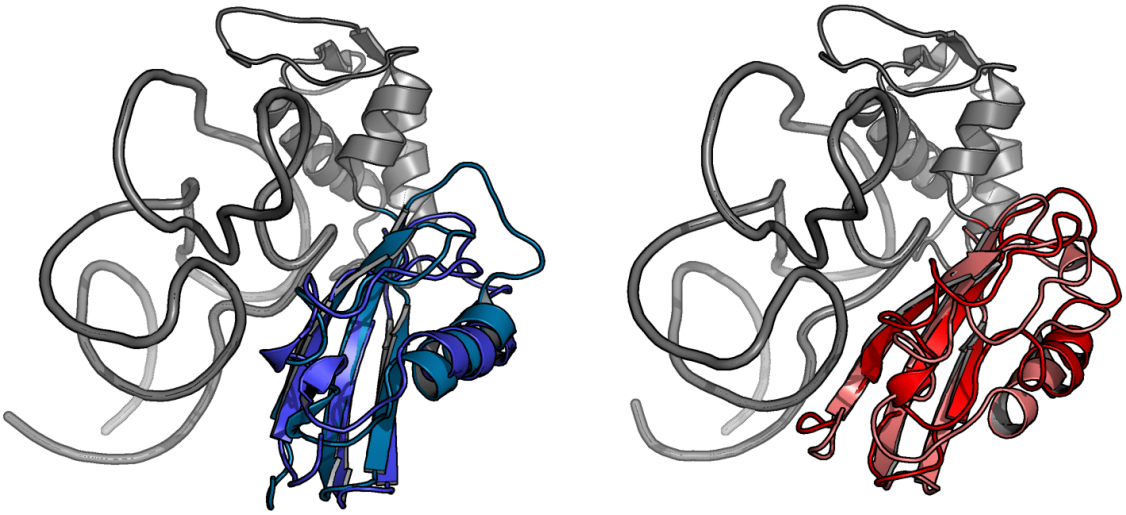


Figure C.3.: Cluster representatives for the two simulation runs without TS. A) Cluster one from the first run is displayed in marine, cluster one from the second run is displayed in darkblue. B) Cluster three from first run (red), cluster two from second run (light red).

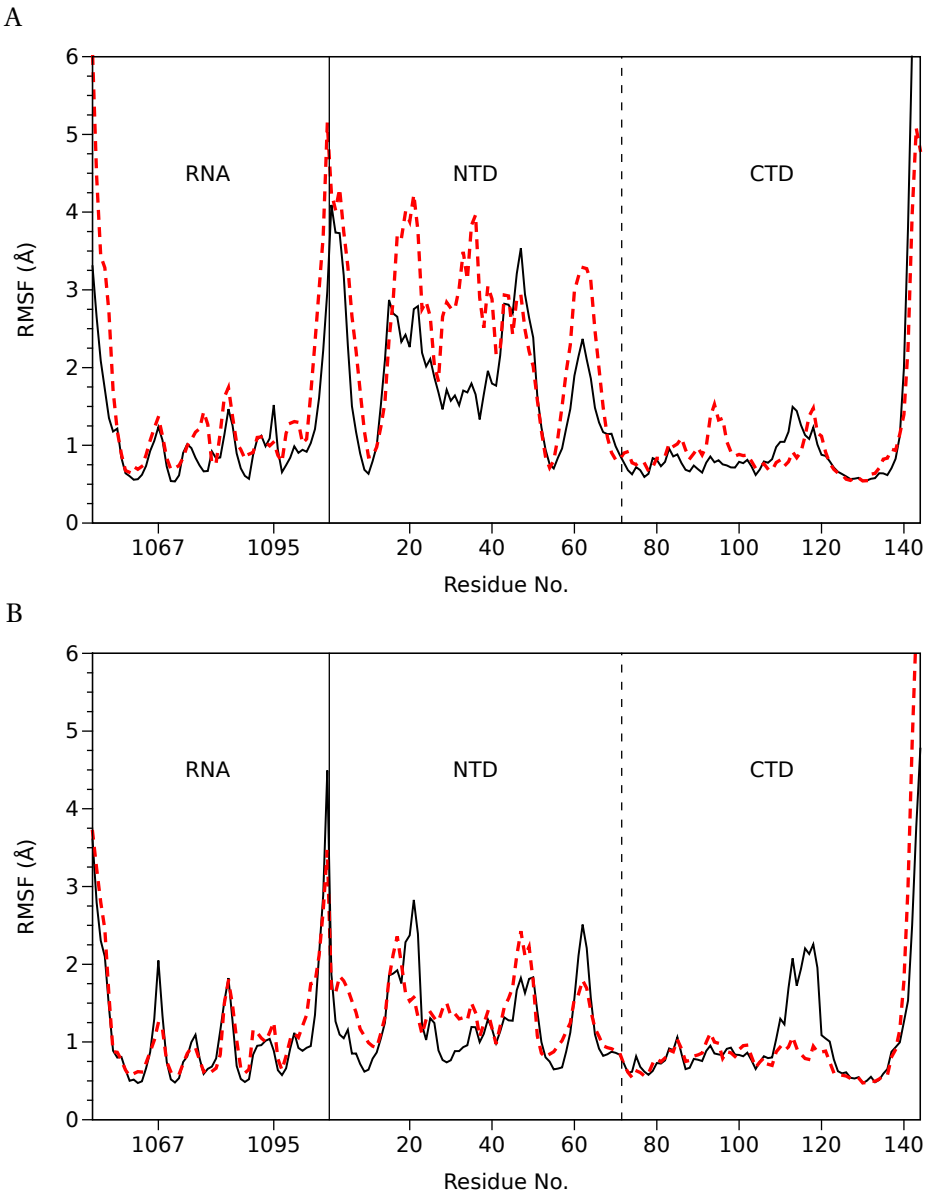


Figure C.4.: Root-mean-square fluctuations (RMSF) for simulations without TS (A) and with bound TS (B). Black line: run I; red line: run II. Evaluated at time frame 7–40 ns.

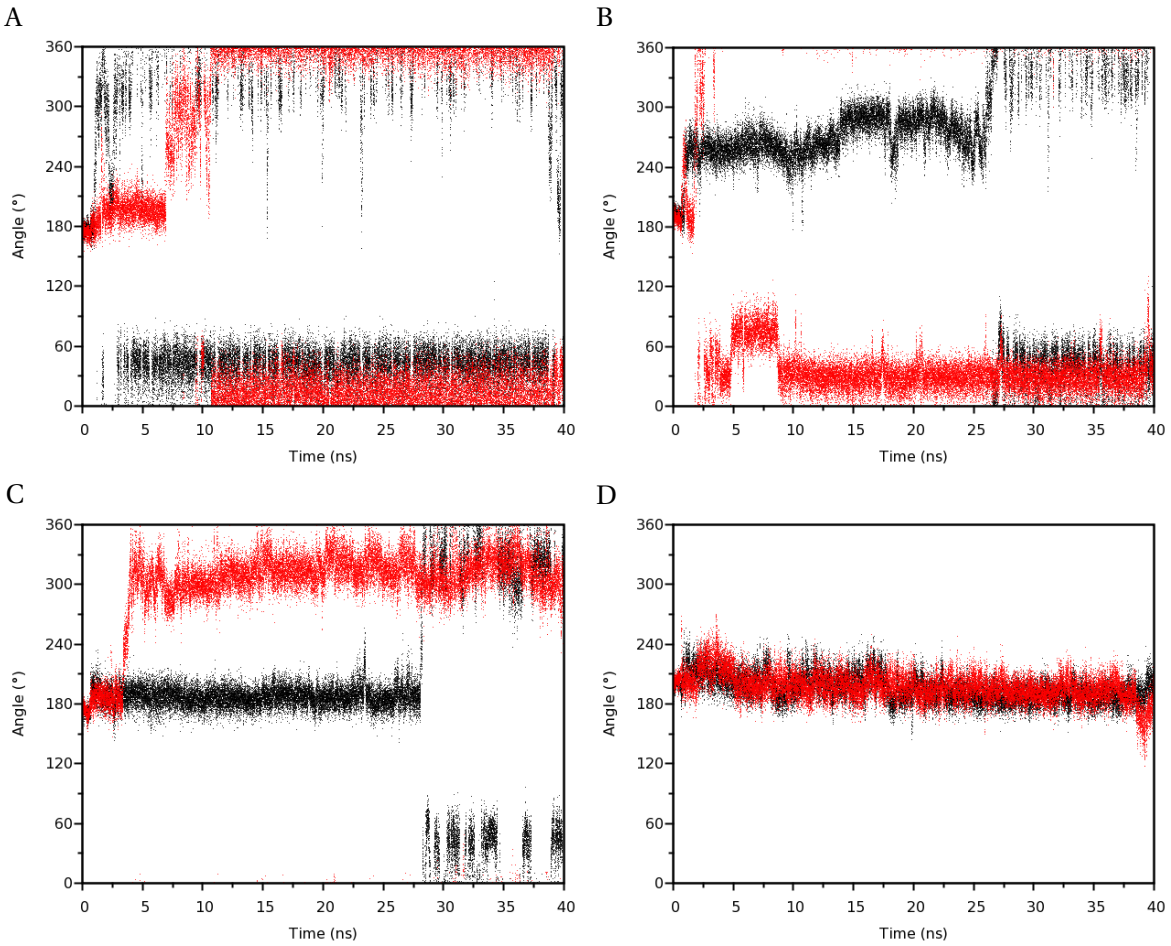


Figure C.5.: χ-angle (O4'-C1'-N9-C4) for nucleic acids (A, C) A1067 and (B, D) A1095 as a function of simulation time without TS (A, B) and with TS (C, D). Black data points are from the first simulation run, and red data points are from the second simulation run.

Table C.1.: The population percentage of different conformations of the selected riboses sampled during the simulations with and without TS. The residues listed are either involved in the binding of thiostrepton or showed interesting conformational sampling.

	Simulations (without TS | with TS)										X-ray
	C3'-endo	C4'-exo	O4'-endo	C1'-exo	C2'-endo	C3'-exo	C4'-endo	O4'-exo	C1'-endo	C2'-exo	
C1053	84| 79	12| 17								5| 4	C2'-exo
G1055	73| 87	1| 3								25| 11	C4'-exo
A1057	33| 40									64| 57	C2'-exo
G1059	79| 71	2| 14								19| 16	C2'-exo
U1060					39| 72	61| 28					C4'-exo
U1061				27| 48	72| 51	1| 1					C4'-exo
G1062	8| 8								2| 2	90| 90	C2'-exo
A1067	2| 42	3| 11	5| 2	29| 5	50| 7	8| 1	1| 0		1| 2	2| 29	C3'-endo
A1070					7| 4	85| 72	8| 24				C3'-endo
C1075	87| 87	1| 3								11| 10	C2'-exo
C1076	89| 62	3| 0								7| 38	C2'-exo
A1084	8| 69	9| 6	49| 5	19| 2	11| 0	2| 0			0| 1	1| 17	C3'-endo
A1085	82| 16	11| 25	0| 44	0| 12	0| 1					6| 2	C2'-exo
G1087	1| 0					2| 5	4| 6	18| 15	59| 60	17| 16	C3'-endo
A1088	0| 1	0| 32	1| 60	26| 4	66| 3						C3'-endo
G1089	0| 7				34| 24	64| 56	1| 1			0| 11	C3'-endo
A1095	8| 31	2| 0	4| 0	12| 0	14| 0	2| 0	1| 0	1| 0	17| 3	38| 66	C3'-endo
G1107	77| 74	2| 2								21| 24	C2'-exo
U1108	60| 43	9| 6	1| 1	4| 5	2| 31	2| 1				5| 4	C2'-exo

Table C.2.: RESP charges for non-standard thiostrepton residues and statistics of the fit.

QUA		TSI		TZB		XAA		XBB		PYT		DHA		TZO	
N1	-0.8100	N1	-0.4373	N	-0.2290	N1	-0.4430	N1	-0.6059	N	-0.2519	N	-0.2415	N1	-0.4841
C2	0.4835	H1	0.2651	H	0.1707	C2	0.2079	C2	0.3539	H	0.2191	H	0.2014	C2	0.4649
C3	-0.3358	C2	0.0158	N1	-0.4312	S3	-0.0692	C3	0.0006	CA	-0.0241	CA	-0.0115	S3	-0.0969
H3	0.1730	H2	0.1186	C2	0.3807	C4	-0.2901	H31	0.0060	C	0.7403	C	0.6365	C4	-0.1256
C4	-0.0294	C3	0.2731	S3	-0.3002	H4	0.2083	H32	0.0060	O	-0.5687	O	-0.5441	H4	0.2169
C5	-0.0059	O4	-0.6748	C4	0.1858	C5	0.3044	C4	0.0844	CB	-0.3563	CB	-0.3732	C5	-0.0523
H5	0.0744	H4	0.4327	H41	0.0481	C6	-0.0135	H41	-0.0108	HB1	0.1682	HB1	0.1662	C6	0.7028
C6	-0.2120	C5	-0.1825	H42	0.0481	H6	0.1423	H42	-0.0108	HB2	0.1682	HB2	0.1662	O7	-0.6256
H6	0.1699	H51	0.0605	C5	-0.0314	N7	-0.1824	C5	0.1312	NT	-0.8813				
C7	-0.4365	H52	0.0605	H5	0.0707	H7	0.0731	C6	0.0981	HT1	0.3884				
H7	0.2665	H53	0.0605	C6	0.6333	C8	0.1799	H6	0.1065	HT2	0.3884				
C8	0.1996	C6	0.3282	O7	-0.6232	H8	0.0255	N7	-0.4443						
H8	0.1019	H6	-0.0041	C8	-0.0554	C9	-0.0769	H7	0.2853						
C9	0.5638	O7	-0.7302	C9	-0.0826	H91	0.0232								
C10	-0.0678	H7	0.4431	H9	0.1267	H92	0.0232								
C11	0.5060	C8	-0.0803	C10	-0.0182	H93	0.0232								
O12	-0.5566	H81	0.0170	H101	0.0357	O10	-0.1361								
C13	0.4916	H82	0.0170	H102	0.0357										
H13	-0.0486	H83	0.0170	H103	0.0357										
C14	-0.0103														
H141	-0.0006														
H142	-0.0006														
H143	-0.0006														
O15	-0.7039														
H15	0.4088														
O16	-0.6176														
H16	0.3972														
RRMS[a]	0.21015		0.11422		0.22590		0.21973		0.15595		0.19047		0.22848		0.19686
RRMS[b]	0.17667		0.13408		0.21501		0.19863		0.17331		0.17948		0.19320		0.19564

[a] Relative root mean square (RRMS) for the molecular fragment.
[b] RRMS for the molecule (incl. capping groups).

Table C.3.: RESP charges for standard thiostrepton residues and statistics of the fit.

	ILE		**THR**		**ALA**	
Atom	N	-0.4292	N	-0.3882	N	-0.3958
names	H	0.2737	H	0.2657	H	0.2311
and	CA	-0.0230	CA	0.0683	CA	0.0447
charges	HA	0.0638	HA	0.0614	HA	0.0491
	CB	0.0839	CB	0.1670	CB	-0.0392
	HB	0.0516	HB	0.0001	HB1	0.0221
	CG2	-0.1469	CG2	-0.0696	HB2	0.0221
	HG21	0.0340	HG21	0.0289	HB3	0.0221
	HG22	0.0340	HG22	0.0289	C	0.5858
	HG23	0.0340	HG23	0.0289	O	-0.5420
	CG1	0.0185	OG1	-0.6662		
	HG12	-0.0030	HG1	0.4420		
	HG13	-0.0030	C	0.5785		
	CD1	-0.0779	O	-0.5458		
	HD11	0.0188				
	HD12	0.0188				
	HD13	0.0188				
	C	0.5982				
	O	-0.5649				
RRMS[a]	0.12926		0.11236		0.15923	
RRMS[b]	0.12234		0.10933		0.14356	

[a] RRMS for the molecular fragment.
[b] RRMS for the molecule (incl. capping groups).

Hbond PTRAJ output for simulation with TS, run I:

```
HBOND SUMMARY:
    Data was saved to series hbt, output to file hbonds_lig_don_dist35.txt,
    data was sorted, intra-residue interactions are NOT included,
    Distance cutoff is 3.50 angstroms, angle cutoff is 120.00 degrees
    Hydrogen bond information dumped for occupancies > 0.00

Dumping schematic of time series after each h-bond, key follows:
 |          .       -       o      x      *      @    |
    0-5%   5-20%  20-40%  40-60% 60-80% 80-95% 95-100% occupancy

      DONOR         ACCEPTORH      ACCEPTOR
 atom# :res@atom   atom# :res@atom atom# :res@atom %occupied  distance       angle             lifetime maxocc
|  4067 :203@N4  |   547   :17@HO'2  546   :17@O2' |  12.68  2.906 ( 0.16)  32.54 ( 8.76)      2.2 (   2.0)     21 | -ooooooxooxo---.                                 |
|  4051 :203@O5  |   547   :17@HO'2  546   :17@O2' |   3.38  2.871 ( 0.21)  27.15 (10.59)      3.7 (   5.2)     40 |                                -o-       -       |

       DONOR         ACCEPTORH      ACCEPTOR
 atom# :res@atom   atom# :res@atom atom# :res@atom %occupied  distance       angle             lifetime maxocc
|   546  :17@O2' |  4063  :203@H74  4062  :203@N5  |  20.99  3.299 ( 0.23)  25.84 ( 9.93)      4.2 (   5.2)     41 |.-xxxx*xxx-oxx*o           .   .               . |
|   540  :17@N3  |  4063  :203@H74  4062  :203@N5  |  20.42  3.289 ( 0.20)  26.42 (10.17)      5.6 (   6.4)     56 |               .*xxx***oo*x -xx                   |
|  2263  :86@O   |  3999  :203@H79  3998  :203@N17 |  14.71  3.221 ( 0.26)  32.88 ( 9.21)      1.6 (   1.3)     17 |..-o-...-.....----.. .-. ..--.-.................  .|
|   540  :17@N3  |  4082  :203@H75  4081  :203@N3  |  14.34  3.398 ( 0.21)  22.65 (10.06)      4.4 (   5.8)     72 |.oxxxxx*xx   .o.                                  |
|   546  :17@O2' |  4052  :203@H38  4051  :203@O5  |   6.57  2.982 ( 0.22)  19.83 (10.54)      4.6 (   6.5)     64 |  .-...--.o-....              .       .           |
|   531  :17@N7  |  4082  :203@H75  4081  :203@N3  |   5.90  3.188 ( 0.20)  26.98 (10.11)      6.3 (   7.4)     47 |                                          .*o xx|
|   546  :17@O2' |  4082  :203@H75  4081  :203@N3  |   2.76  3.378 ( 0.25)  33.75 ( 8.55)      1.5 (   1.2)     14 |     ..... ..-.-                                 |
```

Hbond PTRAJ output for docked TS to MD-refined structure, optimized TS parameters, 10 ns simulation:

```
       DONOR         ACCEPTORH      ACCEPTOR
 atom# :res@atom   atom# :res@atom atom# :res@atom %occupied  distance       angle             lifetime maxocc
|  3999 :204@OG1 |   547   :17@HO'2  546   :17@O2' |  29.05  2.869 ( 0.16)  30.09 (14.08)      2.5 (   2.6)     29 |x. .           o-- .ooooooooo.  -ooo-oooxo--ooooo-.|
|  4002 :204@O   |  1438   :45@H61  1437   :45@N6  |  26.09  3.122 ( 0.19)  36.75 (12.10)      3.6 (   4.1)     31 |                       -*--oxoxxxxxxxx**o..o.-.-- |
|  4021 :205@N1  |   547   :17@HO'2  546   :17@O2' |  19.82  3.258 ( 0.16)  48.07 ( 9.03)      1.5 (   0.9)      9 |.  -    ....  .--- .o---------...---.-.------o-. |
|  4165 :216@O   |  2278   :87@HE22 2276   :87@NE2 |   8.55  3.021 ( 0.19)  23.43 (11.01)      8.2 (  13.8)    119 |                -o .x.---.             o-  --   |
|  4165 :216@O   |  2794  :121@HH12 2792  :121@NH1 |   5.72  2.998 ( 0.21)  33.96 (13.34)      6.3 (   9.3)     59 |     -xxx-                                        |

      DONOR         ACCEPTORH      ACCEPTOR
 atom# :res@atom   atom# :res@atom atom# :res@atom %occupied  distance       angle             lifetime maxocc
|   540  :17@N3  |  4000  :204@HG1  3999  :204@OG1 |  83.91  2.882 ( 0.16)  24.32 (14.06)      8.4 (  15.2)    266 |@ooox@@@*x***@@*@*@*@*******xxxxx****x****xx*****xo|
|  2318  :90@O   |  4178  :217@HT1  4177  :217@NT  |  31.50  2.919 ( 0.15)  21.19 (11.26)     41.9 (  50.6)    220 |@@@@@@*@@*@@@@@*.                                 |
|  2263  :86@O   |  4179  :217@HT2  4177  :217@NT  |  29.97  3.002 ( 0.17)  25.44 (11.94)     17.7 (  25.8)    216 |x@@@@@*@**@***@*.                                 |
|   546  :17@O2' |  4025  :206@H1   4024  :206@N1  |  23.27  3.261 ( 0.15)  45.95 (10.27)      1.8 (   1.4)     13 |- .-   ... .  .o-- .ooo----o...---------o--.oo--o. |
|  2263  :86@O   |  4162  :216@H    4161  :216@N   |  21.16  3.021 ( 0.16)  31.61 (13.65)      8.5 (  13.3)    117 |oox-o*@****o-xx                   .           .|
|   540  :17@N3  |  4078  :209@H16  4077  :209@O16 |  17.69  2.992 ( 0.18)  25.10 (11.67)     14.9 (  29.4)    177 |  o. .. *x*x      -        -@@@o         .x       |
|   546  :17@O2' |  4000  :204@HG1  3999  :204@OG1 |   9.81  2.970 ( 0.19)  45.15 (12.37)      1.6 (   1.6)     18 | ox-    -...  . . ..    .. ..--.......   ..       |
|   558  :18@O4' |  4011  :205@H    4010  :205@N   |   7.22  3.353 ( 0.12)  42.56 (11.31)      1.3 (   0.7)      6 |-              . . .-......   .  ..   ...-..-..-  |
|   546  :17@O2' |  4032  :206@H4   4031  :206@O4  |   4.34  3.059 ( 0.22)  20.17 (10.65)      4.7 (   8.6)     49 |-xx-               .                              |
|  2263  :86@O   |  4170  :217@H    4169  :217@N   |   4.22  3.208 ( 0.17)  41.50 (14.97)      1.5 (   1.1)      8 | ...-.. ...-...-.                                 |
```

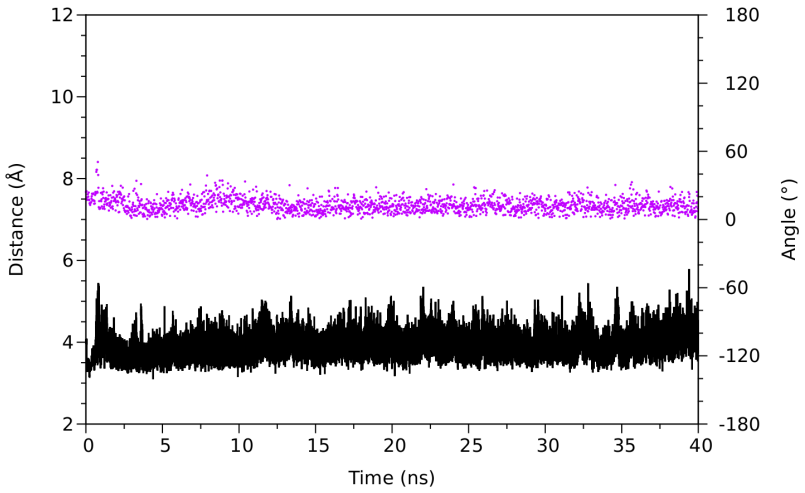


Figure C.6.: Centroid–centroid distances (black, left axis) and angles between the ring normals (magenta, right axis) over simulation time for TZO-1 – A1095 during the second MD simulation with TS.

Bibliography

[1] Coates, A. R. M. and Hu, Y. (2007) Novel approaches to developing new antibiotics for bacterial infections. *Br. J. Pharmacol.* **152(8)**, 1147–1154.

[2] Fischbach, M. A. and Walsh, C. T. (2009) Antibiotics for emerging pathogens. *Science* **325(5944)**, 1089–1093.

[3] Ray, A. and Rice, L. B. (2004) Wildcatters welcome: the need for new antimicrobial agents. *Therapy* **1(1)**, 1–5.

[4] Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* **6(1)**, 29–40.

[5] Jones, D. (2010) The antibacterial lead discovery challenge. *Nat. Rev. Drug Discov.* **9(10)**, 751–752.

[6] Braine, T. (2011) Race against time to develop new antibiotics. *Bull. World Health Organ.* **89(2)**, 88–89.

[7] Bush, K. (2004) Why is it important to continue antibacterial drug discovery? *ASM News* **70**, 282–287.

[8] Overbye, K. M. and Barrett, J. F. (2005) Antibiotics: where did we go wrong? *Drug Discov. Today* **10(1)**, 45–52.

[9] Organization., W. H. Antimicrobial resistance. Fact sheet N°194 (2011) `http://www.who.int/mediacentre/factsheets/fs194/en` (accessed 23 Jan 2012).

[10] Hancock, R. E. W. (2007) The end of an era? *Nat. Rev. Drug Discov.* **6(1)**, 28–28.

[11] Klevens, R. M. e. a. (2007) Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *JAMA* **298(15)**, 1763–1771.

[12] Dorman, S. E. and Chaisson, R. E. (2007) From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nat. Med.* **13(3)**, 295–298.

[13] Organization., W. H. 10 facts about tuberculosis. (2010) `http://www.who.int/features/factfiles/tuberculosis/en` (accessed 23 Jan 2012).

[14] Gwynn, M. N., Portnoy, A., Rittenhouse, S. F., and Payne, D. J. (2010) Challenges of antibacterial discovery revisited. *Ann. N. Y. Acad. Sci.* **1213**, 5–19.

[15] vonNussbaum, F., Brands, M., Hinzen, B., Weigand, S., and Häbich, D. (2006) Antibacterial natural products in medicinal chemistry–exodus or revival? *Angew. Chem. Int. Ed. Engl.* **45(31)**, 5072–5129.

[16] Silver, L. L. (2008) Are natural products still the best source for antibacterial discovery? The bacterial entry factor. *Expert Opin. Drug Discov.* **3(5)**, 487–500.

[17] Bagley, M. C., Dale, J. W., Merritt, E. A., and Xiong, X. (2005) Thiopeptide antibiotics. *Chem. Rev.* **105(2)**, 685–714.

[18] Wilson, D. N. and Nierhaus, K. H. (2004) Antibiotics and the inhibition of ribosome function. In Knud H. Nierhaus and Daniel N. Wilson, (ed.), Protein Synthesis and Ribosome Structure, pp. 449–527 Wiley-VHC Weinheim.

[19] Sohmen, D., Harms, J. M., Schlünzen, F., and Wilson, D. N. (2009) Enhanced SnapShot: Antibiotic inhibition of protein synthesis II. *Cell* **139(1)**, 212–212.e1.

[20] Munro, J. B., Altman, R. B., O'Connor, N., and Blanchard, S. C. (2007) Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell* **25(4)**, 505–517.

[21] Wilson, D. N. and Nierhaus, K. H. (2003) The ribosome through the looking glass. *Angew. Chem. Int. Ed. Engl.* **42(30)**, 3464–3486.

[22] Frank, J. and Agrawal, R. K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406(6793)**, 318–322.

[23] Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003) Locking and unlocking of ribosomal motions. *Cell* **114(1)**, 123–134.

[24] Frank, J. The positions of EF-Tu, EF-G, and tRNAs relative to the ribosome during the elongation cycle. (2008) `http://rpi.edu/dept/bcbp/molbiochem/MBWeb/mb2/part1/translate.htm` (accessed 7th August 2011).

[25] Sohmen, D., Harms, J. M., Schlünzen, F., and Wilson, D. N. (2009) SnapShot: Antibiotic inhibition of protein synthesis I. *Cell* **138(6)**, 1248.e1.

[26] Mohr, D., Wintermeyer, W., and Rodnina, M. V. (2002) GTPase activation of elongation factors Tu and G on the ribosome. *Biochemistry* **41(41)**, 12520–12528.

[27] Diaconu, M., Kothe, U., Schlünzen, F., Fischer, N., Harms, J. M., Tonevitsky, A. G., Stark, H., Rodnina, M. V., and Wahl, M. C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121(7)**, 991–1004.

[28] Gao, Y.-G., Selmer, M., Dunham, C. M., Weixlbaumer, A., Kelley, A. C., and Ramakrishnan, V. (2009) The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326(5953)**, 694–699.

[29] Wilson, D. N. and Nierhaus, K. H. (2005) Ribosomal proteins in the spotlight. *Crit. Rev. Biochem. Mol. Biol.* **40(5)**, 243–267.

[30] Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W., and Ramakrishnan, V. (1999) A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell* **97(4)**, 491–502.

[31] Jonker, H. R. A., Ilin, S., Grimm, S. K., Wöhnert, J., and Schwalbe, H. (2007) L11 domain rearrangement upon binding to RNA and thiostrepton studied by NMR spectroscopy. *Nucleic Acids Res.* **35(2)**, 441–454.

[32] Harms, J. M., Wilson, D. N., Schlünzen, F., Connell, S. R., Stachelhaus, T., Zaborowska, Z., Spahn, C. M. T., and Fucini, P. (2008) Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Mol. Cell* **30(1)**, 26–38.

[33] Connell, S. R., Takemoto, C., Wilson, D. N., Wang, H., Murayama, K., Terada, T., Shirouzu, M., Rost, M., Schüler, M., Giesebrecht, J., Dabrowski, M., Mielke, T., Fucini, P., Yokoyama, S., and Spahn, C. M. T. (2007) Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. *Mol. Cell* **25(5)**, 751–764.

[34] Donovick, R., Pagano, J. F., Stout, H. A., and Weinstein, M. J. (1955) Thiostrepton, a new antibiotic. I. In vitro studies. *Antibiot. Annu.* **3**, 554–559.

[35] Dutcher, J. D. and Vandeputte, J. (1955) Thiostrepton, a new antibiotic. II. Isolation and chemical characterization. *Antibiot. Annu.* **3**, 560–561.

[36] Prange, T., Ducruix, A., Pascard, C., and Lunel, J. (1977) Structure of nosipeptide, a polythiazole-containing antibiotic. *Nature* **265(5590)**, 189–190.

[37] Pascard, C., Ducruix, A., Lunel, J., and Prangé, T. (1977) Highly modified cysteine-containing antibiotics. Chemical structure and configuration of nosiheptide. *J. Am. Chem. Soc.* **99(19)**, 6418–6423.

[38] Benazet, F., Cartier, M., Florent, J., Godard, C., Jung, G., Lunel, J., Mancy, D., Pascal, C., Renaut, J., Tarridec, P., Theilleux, J., Tissier, R., Dubost, M., and Ninet, L. (1980) Nosiheptide, a sulfur-containing peptide antibiotic isolated from Streptomyces actuosus 40037. *Experientia* **36(4)**, 414–416.

[39] Su, T. L. (1948) Micrococcin, an antibacterial substance formed by a strain of Micrococcus. *Br. J. Exp. Pathol.* **29(5)**, 473–481.

[40] Lefranc, D. and Ciufolini, M. A. (2009) Total synthesis and stereochemical assignment of micrococcin P1. *Angew. Chem. Int. Ed. Engl.* **48(23)**, 4198–4201.

[41] Ciufolini, M. A. and Lefranc, D. (2010) Micrococcin P1: structure, biology and synthesis. *Nat. Prod. Rep.* **27(3)**, 330–342.

[42] Nicolaou, K. C., Zak, M., Rahimipour, S., Estrada, A. A., Lee, S. H., O'Brate, A., Giannakakou, P., and Ghadiri, M. R. (2005) Discovery of a biologically active thiostrepton fragment. *J. Am. Chem. Soc.* **127(43)**, 15042–15044.

[43] Nicolaou, K. C., Zak, M., Safina, B. S., Estrada, A. A., Lee, S. H., and Nevalainen, M. (2005) Total synthesis of thiostrepton. Assembly of key building blocks and completion of the synthesis. *J. Am. Chem. Soc.* **127(31)**, 11176–11183.

[44] Bower, J., Drysdale, M., Hebdon, R., Jordan, A., Lentzen, G., Matassova, N., Murchie, A., Powles, J., and Roughley, S. (2003) Structure-based design of agents targeting the bacterial ribosome. *Bioorg. Med. Chem. Lett.* **13(15)**, 2455–2458.

[45] Bausch, S. L., Poliakova, E., and Draper, D. E. (2005) Interactions of the N-terminal domain of ribosomal protein L11 with thiostrepton and rRNA. *J. Biol. Chem.* **280(33)**, 29956–29963.

[46] Porse, B. T., Leviev, I., Mankin, A. S., and Garrett, R. A. (1998) The antibiotic thiostrepton inhibits a functional transition within protein L11 at the ribosomal GTPase centre. *J. Mol. Biol.* **276(2)**, 391–404.

[47] Cameron, D. M., Thompson, J., Gregory, S. T., March, P. E., and Dahlberg, A. E. (2004) Thiostrepton-resistant mutants of Thermus thermophilus. *Nucleic Acids Res.* **32(10)**, 3220–3227.

[48] Starosta, A. L., Qin, H., Mikolajka, A., Leung, G. Y. C., Schwinghammer, K., Nicolaou, K. C., Chen, D. Y.-K., Cooperman, B. S., and Wilson, D. N. (2009) Identification of distinct thiopeptide-antibiotic precursor lead compounds using translation machinery assays. *Chem. Biol.* **16(10)**, 1087–1096.

[49] Jorgensen, W. L. (2004) The many roles of computation in drug discovery. *Science* **303(5665)**, 1813–1818.

[50] Taft, C. A., Da Silva, V. B., and Da Silva, C. H. (2008) Current topics in computer-aided drug design. *J. Pharm. Sci.* **97(3)**, 1089–1098.

[51] Kapetanovic, I. M. (2008) Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chem. Biol. Interact.* **171(2)**, 165–176.

[52] Rester, U. (2008) From virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. *Curr. Opin. Drug Discov. Devel.* **11(4)**, 559–568.

[53] Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., and Sastry, G. N. (2007) Virtual screening in drug discovery – a computational perspective. *Curr. Protein Pept. Sci.* **8(4)**, 329–351.

[54] Eckert, H. and Bajorath, J. (2007) Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today* **12(5-6)**, 225–233.

[55] Congreve, M., Murray, C. W., and Blundell, T. L. (2005) Keynote review: Structural biology and drug discovery. *Drug Discov. Today* **10(13)**, 895–907.

[56] Klebe, G. Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen chapter Erfolge beim rationalen Design von Wirkstoffen., pp. 333–605 Spektrum Akademischer Verlag (2009).

[57] Hopkins, A. L. and Groom, C. R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* **1(9)**, 727–730.

[58] Pommier, Y. and Cherfils, J. (2005) Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends Pharmacol. Sci.* **26(3)**, 138–145.

[59] Pommier, Y. and Marchand, C. (2012) Interfacial inhibitors: targeting macromolecular complexes. *Nat. Rev. Drug Discov.* **11**, 25–36.

[60] Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., and Cherepanov, P. (2010) Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc. Natl. Acad. Sci. USA* **107(46)**, 20057–20062.

[61] Hare, S., Gupta, S. S., Valkov, E., Engelman, A., and Cherepanov, P. (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464(7286)**, 232–236.

[62] Hare, S., Smith, S. J., Métifiot, M., Jaxa-Chamiec, A., Pommier, Y., Hughes, S. H., and Cherepanov, P. (2011) Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol. Pharmacol.* **80(4)**, 565–572.

[63] Ioanoviciu, A., Antony, S., Pommier, Y., Staker, B. L., Stewart, L., and Cushman, M. (2005) Synthesis and mechanism of action studies of a series of norindenoisoquinoline topoisomerase I poisons reveal an inhibitor with a flipped orientation in the ternary DNA-enzyme-inhibitor complex as determined by X-ray crystallographic analysis. *J. Med. Chem.* **48(15)**, 4803–4814.

[64] Marchand, C., Antony, S., Kohn, K. W., Cushman, M., Ioanoviciu, A., Staker, B. L., Burgin, A. B., Stewart, L., and Pommier, Y. (2006) A novel norindenoisoquinoline structure reveals a common interfacial inhibitor paradigm for ternary trapping of the topoisomerase I-DNA covalent complex. *Mol. Cancer Ther.* **5(2)**, 287–295.

[65] González-Ruiz, D. and Gohlke, H. (2006) Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* **13(22)**, 2607–2625.

[66] Drysdale, M. J., Lentzen, G., Matassova, N., Murchie, A. I. H., Aboul-Ela, F., and Afshar, M. (2002) RNA as a drug target. *Prog. Med. Chem.* **39**, 73–119.

[67] Sucheck, S. J. and Wong, C. H. (2000) RNA as a target for small molecules. *Curr. Opin. Chem. Biol.* **4(6)**, 678–686.

[68] Thomas, J. R. and Hergenrother, P. J. (2008) Targeting RNA with small molecules. *Chem. Rev.* **108(4)**, 1171–1224.

[69] Hermann, T. and Tor, Y. (2005) RNA as a target for small-molecule therapeutics. *Expert Opin. Ther. Pat.* **15(1)**, 49–62.

[70] Knowles, D. J. C., Foloppe, N., Matassova, N. B., and Murchie, A. I. H. (2002) The bacterial ribosome, a promising focus for structure-based drug design. *Curr. Opin. Pharmacol.* **2(5)**, 501–506.

[71] Franceschi, F. and Duffy, E. M. (2006) Structure-based drug design meets the ribosome. *Biochem. Pharmacol.* **71(7)**, 1016–1025.

[72] Wimberly, B. T. (2009) The use of ribosomal crystal structures in antibiotic drug design. *Curr. Opin. Investig. Drugs* **10(8)**, 750–765.

[73] Sherer, E. C. (2010) Chapter 9 - Antibiotics Targeting the Ribosome: Structure-Based Design and the Nobel Prize. In Ralph A. Wheeler, (ed.), Annual Reports in Computational Chemistry, pp. 139–166 Elsevier.

[74] Gallego, J. and Varani, G. (2001) Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges. *Acc. Chem. Res.* **34(10)**, 836–843.

[75] Foloppe, N., Matassova, N., and Aboul-Ela, F. (2006) Towards the discovery of drug-like RNA ligands? *Drug Discov. Today* **11(21-22)**, 1019–1027.

[76] Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46(1-3)**, 3–26.

[77] Frenkel, D. and Smit, B. (2001) Understanding Molecular Simulation: From Algorithms to Applications, Academic Press, 2nd edition.

[78] Leach, A. R. (2001) Molecular Modelling: Principles and Applications, Prentice Hall, 2nd edition.

[79] Ford, J. (1973) The Transition from Analytic Dynamics to Statistical Mechanics. *Adv. Chem. Phys.* **24**, 155–185.

[80] Alder, B. J. and Wainwright, T. E. (1957) Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **27**, 1208–1209.

[81] Alder, B. J. and Wainwright, T. E. (1959) Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **31**, 459–466.

[82] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26(16)**, 1668–1688.

[83] Brooks, B. R. et al. (2009) CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30(10, Sp. Iss. SI)**, 1545–1614.

[84] Hess, B., Kutzner, C., van derSpoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4(3)**, 435–447.

[85] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26(16)**, 1781–1802.

[86] Jorgensen, W. L. and Tirado-Rives, J. (1988) The OPLS Potential Functions for Proteins. Energy Minimization for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **110**, 1657–1666.

[87] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29(4)**, 622–655.

[88] Wang, J., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21(12)**, 1049–1074.

[89] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65(3)**, 712–725.

[90] Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **92(11)**, 3817–3829.

[91] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25(9)**, 1157–1174.

[92] Halgren, T. A. (1996) Merck molecular force field. *J. Comput. Chem.* **17(5-6)**, 490–641.

[93] Francl, M. M. and Chirlian, L. E. (2000) The Pluses and Minuses of Mapping Atomic Charges to Electrostatic Potentials. In Kenny B. Lipkowitz and Donald B. Boyd, (ed.), Reviews in Computational Chemistry, volume **14**, pp. 1–31 Wiley-VCH New York.

[94] Chirlian, L. E. and Francl, M. M. (1987) Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. *J. Comput. Chem.* **8**, 894–905.

[95] Breneman, C. M. and Wiberg, K. B. (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **11(3)**, 361–373.

[96] Cox, R. S. and Williams, D. E. (1981) Representation of the Molecular Electrostatic Potential by a New Atomic Charge Model. *J. Comput. Chem.* **2**, 304–323.

[97] Singh, U. C. and Kollman, P. A. (1984) An approach to computing electrostatic charges for molecules *J. Comput. Chem.* **5(2)**, 129–145.

[98] Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model *J. Phys. Chem.* **97(40)**, 10269–10280.

[99] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **117(19)**, 5179–5197.

[100] Gasteiger, J. and Marsili, M. (1980) Iterative Partial Equalization of Orbital Electronegativity – Rapid Access to Atomic Charges. *Tetrahedron* **36**, 3219–3288.

[101] Guillot, B. (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **101**, 219–260.

[102] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79(2)**, 926–935.

[103] Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., and Hermans, J. (1981) Interaction models for Water in Relation to Protein Hydration. In B Pullman, (ed.), Intermolecular Forces, pp. 331–342 Reidel Dordrecht, The Netherlands.

[104] Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.* **91(24)**, 6269–6271.

[105] Bernal, J. D. and Fowler, R. H. (1933) A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.* **1(8)**, 515–548.

[106] Stillinger, F. H. and Rahman, A. (1974) Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* **60(4)**, 1545–1557.

[107] Mahoney, M. W. and Jorgensen, W. L. (2000) A five-site model liquid water and the reproduction of the density anomaly by rigid, non-polarizable models. *J. Chem. Phys.* **112**, 8910–8922.

[108] Nada, H. and van derEerden, J. (2003) An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of H2O. *J. Chem. Phys.* **118(16)**, 7401–7413.

[109] Ferguson, D. M. (1995) Parameterization and evaluation of a flexible water model. *J. Comput. Chem.* **16(4)**, 501–511.

[110] Barnes, P., Finney, J. L., Nicholas, J. D., and Quinn, J. E. (1979) Cooperative effects in simulated water. *Nature* **282**, 459–464.

[111] Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **98(7)**, 1978–1988.

[112] Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science* **268(5214)**, 1144–1149.

[113] Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112(16)**, 6127–6129.

[114] Sugita, Y. and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding *Chem. Phys. Lett.* **314(1-2)**, 141–151.

[115] Garcia, A. E. (2008) Molecular dynamics simulations of protein folding. *Methods Mol. Biol.* **413**, 315–330.

[116] Case, D. A. e. a. Amber 11. Users' Manual. University of California, San Francisco (2010).

[117] Sugita, Y. Simulating the dynamics of proteins to understand protein functions. (2010) `http://www.rikenresearch.riken.jp/eng/frontline/6290` (accessed 31 Jan 2012).

[118] Grant, B. The bio3d Package. Users' Manual. University of California, San Diego (2011).

[119] Swaminathan, S., Harte, W. E., and Beveridge, D. L. (1991) Investigation of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer. *J. Am. Chem. Soc.* **113(7)**, 2717–2721.

[120] McCammon, J. A. (1984) Protein dynamics. *Rep. Prog. Phys.* **47(1)**, 1–46.

[121] Hayward, S. and deGroot, B. L. (2008) Normal modes and essential dynamics. *Methods Mol. Biol.* **443**, 89–106.

[122] Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins* **17(4)**, 412–425.

[123] Tan, P.-N., Steinbach, M., and Kumar, V. Introduction to Data Mining chapter Cluster Analysis: Basic Concepts and Algorithms, pp. 487–568 Addison-Wesley (2006).

[124] Jain, A. K., Murty, M. N., and Flynn, P. J. (1999) Data clustering: a review. *ACM Comput. Surv.* **31**, 264–323.

[125] Shao, J., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007) Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **3(6)**, 2312–2334.

[126] Milligan, G. and Cooper, M. (1985) An examination of procedures of determining the number of cluster in a data set. *Psychometrika* **50(2)**, 159–179.

[127] Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001) On Clustering Validation Techniques. *J. Intell. Inf. Syst.* **17**, 107–145 10.1023/A:1012801612483.

[128] Cattell, R. B. (1966) The scree test for the number of factors. *Multivar. Behav. Res.* **1**, 245–276.

[129] Caliński, T. and Harabasz, J. (1974) A dendrite method for cluster analysis. *Commun. Stat.* **3(1)**, 1–27.

[130] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Lange, R., and Ferrin, T. E. (1982) A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161(2)**, 269–288.

[131] Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., and Corbeil, C. R. (2007) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **153 Suppl 1**, 7–26.

[132] Morris, G. M., Goodsell, D. S., Huey, R., Hart, W. E., Halliday, R. S., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662.

[133] Trott, O. and Olson, A. J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31(2)**, 455–461.

[134] Ewing, T. J., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **15(5)**, 411–428.

[135] Moustakas, D. T., Lang, P. T., Pegg, S., Pettersen, E., Kuntz, I. D., Brooijmans, N., and Rizzo, R. C. (2006) Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput. Aided Mol. Des.* **20(10-11)**, 601–619.

[136] Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., Rizzo, R. C., Case, D. A., James, T. L., and Kuntz, I. D. (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **15(6)**, 1219–1230.

[137] Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261(3)**, 470–489.

[138] Rarey, M., Kramer, B., and Lengauer, T. (1999) Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics* **15(3)**, 243–250.

[139] Kramer, B., Rarey, M., and Lengauer, T. (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **37(2)**, 228–241.

[140] Rarey, M., Kramer, B., and Lengauer, T. (1997) Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. *J. Comput. Aided Mol. Des.* **11(4)**, 369–384.

[141] Rarey, M., Kramer, B., and Lengauer, T. (1999) The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **34(1)**, 17–28.

[142] McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A., and Brown, F. K. (2003) Gaussian docking functions. *Biopolymers* **68(1)**, 76–90.

[143] McGann, M. (2011) FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **51(3)**, 578–596.

[144] Friesner, R., Banks, J., Murphy, R., Halgren, T., Klicic, J., Mainz, D., Repasky, M., Knoll, E., Shelley, M., Perry, J., Shaw, D., Francis, P., and Shenkin, P. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47(7)**, 1739–1749.

[145] Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49(2)**, 534–553.

[146] Repasky, M. P., Shelley, M., and Friesner, R. A. (2007) Flexible ligand docking with Glide. *Curr. Protoc. Bioinformatics* **Chapter 8**, Unit 8.12.

[147] Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267(3)**, 727–748.

[148] Abagyan, R., Totrov, M., and Kuznetsov, D. (1994) ICM – a new method for protein modelling and design. Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15(5)**, 488–506.

[149] Wang, R., Lai, L., and Wang, S. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **16(1)**, 11–26.

[150] Welch, W., Ruppert, J., and Jain, A. N. (1996) Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **3(6)**, 449–462.

[151] Tietze, S. and Apostolakis, J. (2007) GlamDock: development and validation of a new docking tool on several thousand protein-ligand complexes. *J. Chem. Inf. Model.* **47(4)**, 1657–1672.

[152] Murray, C. W., Baxter, C. A., and Frenkel, A. D. (1999) The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput. Aided Mol. Des.* **13(6)**, 547–562.

[153] Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28(6)**, 1145–1152.

[154] Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **11(5)**, 425–445.

[155] Gehlhaar, D. K., Verkhivker, G. M., Rejto, P. A., Sherman, C. J., Fogel, D. B., Fogel, L. J., and Freer, S. T. (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **2(5)**, 317–324.

[156] Muegge, I. and Martin, Y. C. (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **42(5)**, 791–804.

[157] Gohlke, H., Hendlich, M., and Klebe, G. (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **295(2)**, 337–356.

[158] Bissantz, C., Folkers, G., and Rognan, D. (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **43(25)**, 4759–4767.

[159] Bursulaya, B. D., Totrov, M., Abagyan, R., and Brooks, III, C. L. (2003) Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided Mol. Des.* **17(11)**, 755–763.

[160] Stahl, M. and Rarey, M. (2001) Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **44(7)**, 1035–1042.

[161] Perola, E., Walters, W. P., and Charifson, P. S. (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56(2)**, 235–249.

[162] Kellenberger, E., Rodrigo, J., Muller, P., and Rognan, D. (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57(2)**, 225–242.

[163] Kontoyianni, M., Sokol, G. S., and McClellan, L. M. (2005) Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* **26(1)**, 11–22.

[164] Chen, H., Lyne, P. D., Giordanetto, F., Lovell, T., and Li, J. (2006) On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* **46(1)**, 401–415.

[165] Warren, G., Andrews, C., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M., Lindvall, M., Nevins, N., Semus, S., Senger, S., Tedesco, G., Wall, I., Woolven, J., Peishoff, C., and Head, M. (2006) A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **49(20)**, 5912–5931.

[166] Darwin, C. R. (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life, John Murray, London.

[167] Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106(3)**, 765–784.

[168] Goodford, P. J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–857.

[169] Stouten, P. F. W., Frommel, C., Nakamura, H., and Sander, C. (1993) An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Sim.* **10**, 97–120.

[170] Sippl, M. J. (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213(4)**, 859–883.

[171] Sippl, M. J. (1993) Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput. Aided Mol. Des.* **7(4)**, 473–501.

[172] Sippl, M. J. (1995) Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **5(2)**, 229–235.

[173] Velec, H. F. G., Gohlke, H., and Klebe, G. (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **48(20)**, 6296–6303.

[174] Allen, F. H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B* **58(Pt 3 Pt 1)**, 380–388.

[175] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28(1)**, 235–242.

[176] Pfeffer, P. and Gohlke, H. (2007) DrugScoreRNA – knowledge-based scoring function to predict RNA-ligand interactions. *J. Chem. Inf. Model.* **47(5)**, 1868–1876.

[177] Neudert, G. and Klebe, G. (2011) DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J. Chem. Inf. Model.* **51(10)**, 2731–2745.

[178] Alonso, H., Bliznyuk, A. A., and Gready, J. E. (2006) Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **26(5)**, 531–568.

[179] Rueda, M., Bottegoni, G., and Abagyan, R. (2010) Recipes for the selection of experimental protein conformations for virtual screening. *J. Chem. Inf. Model.* **50(1)**, 186–193.

[180] Totrov, M. and Abagyan, R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **18(2)**, 178–184.

[181] Foloppe, N. and Hubbard, R. (2006) Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **13(29)**, 3583–3608.

[182] Walker, R. and Steinbrecher, T. AMBER advanced tutorials. Tutorial 3. MM-PBSA. (2006) `http://ambermd.org/tutorials/advanced/tutorial3` (accessed 12 August 2011).

[183] Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999) Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **101**, 426–434.

[184] Klebe, G. (2009) Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen, Spektrum Akademischer Verlag, 2nd edition.

[185] Wermuth, C. G., Ganellin, C. R., Lindberg, P., and Mitscher, L. A. (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **70(5)**, 1129–1143.

[186] Wolber, G., Seidel, T., Bendix, F., and Langer, T. (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* **13(1-2)**, 23–29.

[187] Wolber, G. and Langer, T. (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **45(1)**, 160–169.

[188] Rosendahl, G. and Douthwaite, S. (1994) The antibiotics micrococcin and thiostrepton interact directly with 23S rRNA nucleotides 1067A and 1095A. *Nucleic Acids Res.* **22(3)**, 357–363.

[189] Bowen, W. S., Dyke, N. V., Murgola, E. J., Lodmell, J. S., and Hill, W. E. (2005) Interaction of thiostrepton and elongation factor-G with the ribosomal protein L11-binding domain. *J. Biol. Chem.* **280(4)**, 2934–2943.

[190] Baumann, S., Schoof, S., Harkal, S. D., and Arndt, H.-D. (2008) Mapping the binding site of thiopeptide antibiotics by proximity-induced covalent capture. *J. Am. Chem. Soc.* **130(17)**, 5664–5666.

[191] Baumann, S., Schoof, S., Bolten, M., Haering, C., Takagi, M., Shin-ya, K., and Arndt, H.-D. (2010) Molecular determinants of microbial resistance to thiopeptide antibiotics. *J. Am. Chem. Soc.* **132(20)**, 6973–6981.

[192] Walter, J. D., Hunter, M., Cobb, M., Traeger, G., and Spiegel, P. C. (2012) Thiostrepton inhibits stable 70S ribosome binding and ribosome-dependent GTPase activation of elongation factor G and elongation factor 4. *Nucleic Acids Res.* **40(1)**, 360–370.

[193] Jíri Sponer and Filip Lankas, (ed.) (2006) Computational studies of RNA and DNA, Springer, Dordrecht.

[194] McDowell, S. E., Spacková, N., Sponer, J., and Walters, N. G. (2007) Molecular dynamics simulations of RNA: an in silico single molecule approach. *Biopolymers* **85(2)**, 169–184.

[195] Ditzler, M. A., Otyepka, M., Sponer, J., and Walter, N. G. (2010) Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in. *Acc. Chem. Res.* **43(1)**, 40–47.

[196] Rázga, F., Koca, J., Sponer, J., and Leontis, N. B. (2005) Hinge-like motions in RNA kink-turns: the role of the second a-minor motif and nominally unpaired bases. *Biophys. J.* **88(5)**, 3466–3485.

[197] Rázga, F., Zacharias, M., Réblová, K., Koca, J., and Sponer, J. (2006) RNA kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure* **14(5)**, 825–835.

[198] Réblová, K., Rázga, F., Li, W., Gao, H., Frank, J., and Sponer, J. (2010) Dynamics of the base of ribosomal A-site finger revealed by molecular dynamics simulations and Cryo-EM. *Nucleic Acids Res.* **38(4)**, 1325–1340.

[199] Rázga, F., Koca, J., Mokdad, A., and Sponer, J. (2007) Elastic properties of ribosomal RNA building blocks: molecular dynamics of the GTPase-associated center rRNA. *Nucleic Acids Res.* **35(12)**, 4007–4017.

[200] Besseová, I., Réblová, K., Leontis, N. B., and Sponer, J. (2010) Molecular dynamics simulations suggest that RNA three-way junctions can act as flexible RNA structural elements in the ribosome. *Nucleic Acids Res.* **38(18)**, 6247–6264.

[201] Trylska, J., Tozzini, V., and McCammon, J. A. (2005) Exploring global motions and correlations in the ribosome. *Biophys. J.* **89(3)**, 1455–1463.

[202] Sanbonmatsu, K. Y. and Joseph, S. (2003) Understanding discrimination by the ribosome: stability testing and groove measurement of codon-anticodon pairs. *J. Mol. Biol.* **328(1)**, 33–47.

[203] Sanbonmatsu, K. Y., Joseph, S., and Tung, C.-S. (2005) Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA* **102(44)**, 15854–15859.

[204] Li, W., Ma, B., and Shapiro, B. (2001) Molecular dynamics simulations of the denaturation and refolding of an RNA tetraloop. *J. Biomol. Struct. Dyn.* **19(3)**, 381–396.

[205] Li, W., Ma, B., and Shapiro, B. A. (2003) Binding interactions between the core central domain of 16S rRNA and the ribosomal protein S15 determined by molecular dynamics simulations. *Nucleic Acids Res.* **31(2)**, 629–638.

[206] Li, W., Sengupta, J., Rath, B. K., and Frank, J. (2006) Functional conformations of the L11-ribosomal RNA complex revealed by correlative analysis of cryo-EM and molecular dynamics simulations. *RNA* **12(7)**, 1240–1253.

[207] Réblová, K., Spacková, N., Koca, J., Leontis, N. B., and Sponer, J. (2004) Long-residency hydration, cation binding, and dynamics of loop E/helix IV rRNA-L25 protein complex. *Biophys. J.* **87(5)**, 3397–3412.

[208] Bond, C. S., Shaw, M. P., Alphey, M. S., and Hunter, W. N. (2001) Structure of the macrocycle thiostrepton solved using the anomalous dispersion contribution of sulfur. *Acta Crystallogr. D* **57(Pt 5)**, 755–758.

[209] Frisch, M. J. e. a. Gaussian 03. (2004) Gaussian, Inc., Wallingford, CT.

[210] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103(19)**, 8577–8593.

[211] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.

[212] Team., R. D. C. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria (2009).

[213] Hünenberger, P. H., Mark, A. E., and vanGunsteren, W. F. (1995) Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* **252(4)**, 492–503.

[214] Karplus, M. and Ichiye, T. (1996) Comment on a "fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". *J. Mol. Biol.* **263(2)**, 120–122.

[215] Ichiye, T. and Karplus, M. (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11(3)**, 205–217.

[216] Karplus, M., Swaminathan, S., Ichiye, T., and vanGunsteren, W. F. (1983) Local and collective motions in protein dynamics. *Ciba Found. Symp.* **93**, 271–290.

[217] Ghosh, A. and Vishveshwara, S. (2007) A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA* **104(40)**, 15711–15716.

[218] Sethi, A., Eargle, J., Black, A. A., and Luthey-Schulten, Z. (2009) Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. USA* **106(16)**, 6620–6625.

[219] Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22(21)**, 2695–2696.

[220] Hartigan, J. A. and Wong, M. A. (1979) A K-means clustering algorithm. *Appl. Stat.* **28**, 100–108.

[221] Wolf, A. and Kirschner, K. N. (2012) Principal Component and Clustering Analysis on Molecular Dynamics Data of the Ribosomal L11·23S Subdomain. *J. Mol. Mod.* submitted.

[222] Jolliffe, I. T. (2002) Principal Component Analysis, Springer, New York.

[223] Weiser, J., Shenkin, P. S., and Still, W. C. (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **20(2)**, 217–230.

[224] Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22(12)**, 2577–2637.

[225] Altona, C. and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94(23)**, 8205–8212.

[226] DeLano, W. L. The PyMOL Molecular Graphics System. DeLano Scientific LLC (2006).

[227] Blyn, L. B., Risen, L. M., Griffey, R. H., and Draper, D. E. (2000) The RNA-binding domain of ribosomal protein L11 recognizes an rRNA tertiary structure stabilized by both thiostrepton and magnesium ion. *Nucleic Acids Res.* **28(8)**, 1778–1784.

[228] Ilin, S., Hoskins, A., Ohlenschläger, O., Jonker, H. R. A., Schwalbe, H., and Wöhnert, J. (2005) Domain reorientation and induced fit upon RNA binding: solution structure and dynamics of ribosomal protein L11 from Thermotoga maritima. *ChemBioChem.* **6(9)**, 1611–1618.

[229] Lee, D., Walsh, J. D., Yu, P., Markus, M. A., Choli-Papadopoulou, T., Schwieters, C. D., Krueger, S., Draper, D. E., and Wang, Y.-X. (2007) The structure of free L11 and functional dynamics of L11 in free, L11-rRNA(58 nt) binary and L11-rRNA(58 nt)-thiostrepton ternary complexes. *J. Mol. Biol.* **367(4)**, 1007–1022.

[230] Zgarbová, M., Otyepka, M., Sponer, J., Mládek, A., Banás, P., Cheatham, T. E., and Jurecka, P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **7(9)**, 2886–2902.

[231] Sklenovský, P., Florová, P., Banás, P., Réblová, K., Lankas, F., Otyepka, M., and Sponer, J. (2011) Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. *J. Chem. Theory Comput.* **7(9)**, 2963–2980.

[232] Yildirim, I., Stern, H. A., Kennedy, S. D., Tubbs, J. D., and Turner, D. H. (2010) Reparameterization of RNA chi Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* **6(5)**, 1520–1531.

[233] C. Michael Blackburn and Michael J. Gait, (ed.) (1996) Nucleic acids in chemistry and biology, Oxford University Press, Oxford and New York.

[234] David-Eden, H., Mankin, A. S., and Mandel-Gutfreund, Y. (2010) Structural signatures of antibiotic binding sites on the ribosome. *Nucleic Acids Res.* **38(18)**, 5982–5994.

[235] Jonker, H. R. A., Baumann, S., Wolf, A., Schoof, S., Hiller, F., Schulte, K. W., Kirschner, K. N., Schwalbe, H., and Arndt, H.-D. (2011) NMR structures of thiostrepton derivatives for characterization of the ribosomal binding site. *Angew. Chem. Int. Ed. Engl.* **50(14)**, 3308–3312.

[236] Kawai, G., Ue, H., Yasuda, M., Sakamoto, K., Hashizume, T., McCloskey, J. A., Miyazawa, T., and Yokoyama, S. (1991) Relation between functions and conformational characteristics of modified nucleosides found in tRNAs. *Nucleic Acids Symp. Ser.* **25**, 49–50.

[237] Kawai, G., Yamamoto, Y., Kamimura, T., Masegi, T., Sekine, M., Hata, T., Iimori, T., Watanabe, T., Miyazawa, T., and Yokoyama, S. (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* **31(4)**, 1040–1046.

[238] Wolf, A., Baumann, S., Arndt, H.-D., and Kirschner, K. N. (2012) The influence of thiostrepton binding on the dynamics of the ribosomal L11·23S subunit as revealed by molecular dynamics simulations. *PLoS ONE* submitted.

[239] Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* **11(13-14)**, 580–594.

[240] Filikov, A. V., Mohan, V., Vickers, T. A., Griffey, R. H., Cook, P. D., Abagyan, R. A., and James, T. L. (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided Mol. Des.* **14(6)**, 593–610.

[241] Detering, C. and Varani, G. (2004) Validation of automated docking programs for docking and database screening against RNA drug targets. *J. Med. Chem.* **47(17)**, 4188–4201.

[242] Moitessier, N., Westhof, E., and Hanessian, S. (2006) Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **49(3)**, 1023–1033.

[243] Barbault, F., Ren, B., Rebehmed, J., Teixeira, C., Luo, Y., Smila-Castro, O., Maurel, F., Fan, B., Zhang, L., and Zhang, L. (2008) Flexible computational docking studies of new aminoglycosides targeting RNA 16S bacterial ribosome site. *Eur. J. Med. Chem.* **43(8)**, 1648–1656.

[244] Yan, Z., Sikri, S., Beveridge, D. L., and Baranger, A. M. (2007) Identification of an aminoacridine derivative that binds to RNA tetraloops. *J. Med. Chem.* **50(17)**, 4096–4104.

[245] Warui, D. M. and Baranger, A. M. (2009) Identification of specific small molecule ligands for stem loop 3 ribonucleic acid of the packaging signal Psi of human immunodeficiency virus-1. *J. Med. Chem.* **52(17)**, 5462–5473.

[246] Li, Y., Shen, J., Sun, X., Li, W., Liu, G., and Tang, Y. (2010) Accuracy assessment of protein-based docking programs against RNA targets. *J. Chem. Inf. Model.* **50(6)**, 1134–1146.

[247] Barbault, F., Zhang, L., Zhang, L., and Fan, B. (2006) Parametrization of a specific free energy function for automated docking against RNA targets using neural networks. *Chemometr. Intell. Lab.* **82**, 269–275.

[248] Morley, S. D. and Afshar, M. (2004) Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *J. Comput. Aided Mol. Des.* **18(3)**, 189–208.

[249] Hermann, T. (2005) Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* **15(3)**, 355–366.

[250] Schoof, S., Baumann, S., Ellinger, B., and Arndt, H.-D. (2009) A fluorescent probe for the 70 S-ribosomal GTPase-associated center. *ChemBioChem.* **10(2)**, 242–245.

[251] Myers, C. L., Hang, P. C., Ng, G., Yuen, J., and Honek, J. F. (2010) Semi-synthetic analogues of thiostrepton delimit the critical nature of tail region modifications in the control of protein biosynthesis and antibacterial activity. *Bioorg. Med. Chem.* **18(12)**, 4231–4237.

[252] MOE (Molecular Operating Environment). Chemical Computing Group Inc., Montreal, Quebec, Canada `http://www.chemcomp.com/software.htm`, V2009.10.

[253] Böhm, H. J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **8(3)**, 243–256.

[254] Marialke, J., Tietze, S., Karasz, M., and Apostolakis, J. Robust Docking with GlamDock. Poster at 4. German Conference on Chemoinformatics, Goslar, Nov 9-11 (2008).

[255] Neudert, G. and Klebe, G. (2011) fconv: Format conversion, manipulation and feature computation of molecular data. *Bioinformatics* **27(7)**, 1021–1022.

[256] Pearlman, D. A. and Charifson, P. S. (2001) Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **44(21)**, 3417–3423.

[257] CORINA Molecular Networks GmbH `http://www.molecular-networks.com/products/corina`.

[258] Hetényi, C. and van derSpoel, D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **11(7)**, 1729–1737.

[259] Tsai, K.-C., Wang, S.-H., Hsiao, N.-W., Li, M., and Wang, B. (2008) The effect of different electrostatic potentials on docking accuracy: a case study using DOCK5.4. *Bioorg. Med. Chem. Lett.* **18(12)**, 3509–3512.

[260] Bikadi, Z. and Hazai, E. (2009) Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminform.* **1**, 15.

[261] Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., and Dill, K. A. (2007) Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **371(4)**, 1118–1134.

[262] Wolf, A., Zimmermann, M., and Hofmann-Apitius, M. (2007) Alternative to consensus scoring – a new approach toward the qualitative combination of docking algorithms. *J. Chem. Inf. Model.* **47(3)**, 1036–1044.

[263] Cole, J. C., Murray, C. W., Nissink, J. W. M., Taylor, R. D., and Taylor, R. (2005) Comparing protein-ligand docking programs is difficult. *Proteins* **60(3)**, 325–332.

[264] Schulz-Gasch, T. and Stahl, M. (2003) Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J. Mol. Model.* **9(1)**, 47–57.

[265] Hermann, T. (2000) Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes. *Angew. Chem. Int. Ed. Engl.* **39(11)**, 1890–1904.

[266] Allers, J. and Shamoo, Y. (2001) Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **311(1)**, 75–86.

[267] Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M., and Thornton, J. M. (2001) Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* **29(4)**, 943–954.

[268] Xie, Y., Tam, V. K., and Tor, Y. (2010) The Interactions of Small Molecules with DNA and RNA. In G. Mayer, (ed.), The Chemical Biology of Nucleic Acids, pp. 115–140 John Wiley & Sons, Ltd Chichester, UK.

[269] Gratias, R. and Kessler, H. (1998) Molecular Dynamics Study on Microheterogeneity and Preferential Solvation in Methanol/Chloroform Mixtures. *J. Phys. Chem. B* **102(11)**, 2027–2031.

[270] Spronk, C. A. E. M., Nabuurs, S. B., Bonvin, A. M. J. J., Krieger, E., Vuister, G. W., and Vriend, G. (2003) The precision of NMR structure ensembles revisited. *J. Biomol. NMR* **25(3)**, 225–234.

[271] Stauch, B., Simon, B., Basile, T., Schneider, G., Malek, N. P., Kalesse, M., and Carlomagno, T. (2010) Elucidation of the structure and intermolecular interactions of a reversible cyclic-peptide inhibitor of the proteasome by NMR spectroscopy and molecular modeling. *Angew. Chem. Int. Ed. Engl.* **49(23)**, 3934–3938.

[272] Labute, P. Probabilistic Receptor Potentials. *Journal of the Chemical Computing Group* (2001) `http://www.chemcomp.com/journals/cstat.htm` (accessed 23 Jan 2012).

[273] Hughes, R. A. and Moody, C. J. (2007) From amino acids to heteroaromatics–thiopeptide antibiotics, nature's heterocyclic peptides. *Angew. Chem. Int. Ed. Engl.* **46(42)**, 7930–7954.

[274] Rodnina, M. V., Savelsbergh, A., Matassova, N. B., Katunin, V. I., Semenkov, Y. P., and Wintermeyer, W. (1999) Thiostrepton inhibits the turnover but not the GTPase of elongation factor G on the ribosome. *Proc. Natl. Acad. Sci. USA* **96(17)**, 9586–9590.

[275] Stark, H., Rodnina, M. V., Wieden, H. J., vanHeel, M., and Wintermeyer, W. (2000) Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* **100(3)**, 301–309.

[276] Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W., and Cieplak, P. (2010) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **12(28)**, 7821–7839.

[277] Reith, D. and Kirschner, K. N. (2011) A modern workflow for force-field development – Bridging quantum mechanics and atomistic computational models. *Comput. Phys. Commun.* **182**, 2184–2191.

[278] Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M., and Jr, J. A. M. (1993) General atomic and molecular electronic structure system *J. Comput. Chem.* **14(11)**, 1347–1363.

[279] Onufriev, A., Bashford, D., and Case, D. A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55(2)**, 383–394.

[280] Patriksson, A. and van derSpoel, D. (2008) A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **10(15)**, 2073–2077.

[281] Kuhn, B., Gerber, P., Schulz-Gasch, T., and Stahl, M. (2005) Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **48(12)**, 4040–4048.

[282] Li, Y., Liu, Z., and Wang, R. (2010) Test MM-PB/SA on true conformational ensembles of protein-ligand complexes. *J. Chem. Inf. Model.* **50(9)**, 1682–1692.

[283] Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **32(5)**, 866–877.

[284] Ferrari, A. M., Degliesposti, G., Sgobba, M., and Rastelli, G. (2007) Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **15(24)**, 7865–7877.

[285] Degliesposti, G., Kasam, V., Costa, A. D., Kang, H.-K., Kim, N., Kim, D.-W., Breton, V., Kim, D., and Rastelli, G. (2009) Design and discovery of plasmepsin II inhibitors using an automated workflow on large-scale grids. *ChemMedChem.* **4(7)**, 1164–1173.

[286] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating

structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33(12)**, 889–897.

[287] Kieltyka, R., Englebienne, P., Moitessier, N., and Sleiman, H. (2010) Quantifying interactions between G-quadruplex DNA and transition-metal complexes. *Methods Mol. Biol.* **608**, 223–255.

[288] Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **51(1)**, 69–82.

[289] Misra, V. K. and Honig, B. (1995) On the magnitude of the electrostatic contribution to ligand-DNA interactions. *Proc. Natl. Acad. Sci. USA* **92(10)**, 4691–4695.

[290] Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* **330(4)**, 891–913.

[291] Kyte, J. and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157(1)**, 105–132.

[292] Baumann, S. Chemisch-Biologische Untersuchungen der Aktivität von Thiopeptidantibiotika an der 70S ribosomalen GTPase-assoziierten Region. PhD thesis Technische Universität Dortmund (2010).

[293] Page, C. S. and Bates, P. A. (2006) Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? *J. Comput. Chem.* **27(16)**, 1990–2007.

[294] Cameron, D. M., Thompson, J., March, P. E., and Dahlberg, A. E. (2002) Initiation factor IF2, thiostrepton and micrococcin prevent the binding of elongation factor G to the Escherichia coli ribosome. *J. Mol. Biol.* **319(1)**, 27–35.

[295] Conn, G. L., Gutell, R. R., and Draper, D. E. (1998) A functional ribosomal RNA tertiary structure involves a base triple interaction. *Biochemistry* **37(34)**, 11980–11988.

[296] Draper, D. E., Xing, Y., and Laing, L. G. (1995) Thermodynamics of RNA unfolding: stabilization of a ribosomal RNA tertiary structure by thiostrepton and ammonium ion. *J. Mol. Biol.* **249(2)**, 231–238.

[297] Lu, M. and Draper, D. E. (1995) On the role of rRNA tertiary structure in recognition of ribosomal protein L11 and thiostrepton. *Nucleic Acids Res.* **23(17)**, 3426–3433.

[298] Rogers, M. J., Bukhman, Y. V., McCutchan, T. F., and Draper, D. E. (1997) Interaction of thiostrepton with an RNA fragment derived from the plastid-encoded ribosomal RNA of the malaria parasite. *RNA* **3(8)**, 815–820.

[299] Rosendahl, G. and Douthwaite, S. (1993) Ribosomal proteins L11 and L10.(L12)4 and the antibiotic thiostrepton interact with overlapping regions of the 23 S rRNA backbone in the ribosomal GTPase centre. *J. Mol. Biol.* **234(4)**, 1013–1020.

[300] Ryan, P. C. and Draper, D. E. (1991) Detection of a key tertiary interaction in the highly conserved GTPase center of large subunit ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **88(14)**, 6308–6312.

[301] Ryan, P. C., Lu, M., and Draper, D. E. (1991) Recognition of the highly conserved GTPase center of 23 S ribosomal RNA by ribosomal protein L11 and the antibiotic thiostrepton. *J. Mol. Biol.* **221(4)**, 1257–1268.

[302] Xing, Y. and Draper, D. E. (1996) Cooperative interactions of RNA and thiostrepton antibiotic with two domains of ribosomal protein L11. *Biochemistry* **35(5)**, 1581–1588.

[303] Bissantz, C., Kuhn, B., and Stahl, M. (2010) A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **53(14)**, 5061–5084.

[304] Carlson, H. A., Masukawa, K. M., Rubins, K., Bushman, F. D., Jorgensen, W. L., Lins, R. D., Briggs, J. M., and McCammon, J. A. (2000) Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **43(11)**, 2100–2114.

[305] Hunter, C. A. and Sanders, J. K. M. (1990) The Nature of $\pi - \pi$ Interactions. *J. Am. Chem. Soc.* **112(14)**, 5525–5534.

[306] Cundliffe, E. (1978) Mechanism of resistance to thiostrepton in the producing-organism Streptomyces azureus. *Nature* **272(5656)**, 792–795.

[307] Bechthold, A. and Floss, H. G. (1994) Overexpression of the thiostrepton-resistance gene from Streptomyces azureus in Escherichia coli and characterization of recognition sites of the 23S rRNA A1067 2'-methyltransferase in the guanosine triphosphatase center of 23S ribosomal RNA. *Eur. J. Biochem.* **224(2)**, 431–437.

[308] Dunstan, M. S., Hang, P. C., Zelinskaya, N. V., Honek, J. F., and Conn, G. L. (2009) Structure of the thiostrepton resistance methyltransferase.S-adenosyl-L-methionine complex and its interaction with ribosomal RNA. *J. Biol. Chem.* **284(25)**, 17013–17020.

[309] Sperandio, O., Miteva, M. A., Segers, K., Nicolaes, G. A. F., and Villoutreix, B. O. (2008) Screening Outside the Catalytic Site: Inhibition of Macromolecular Inter-actions Through Structure-Based Virtual Ligand Screening Experiments. *Open. Biochem. J.* **2**, 29–37.

[310] Kilpatrick, J. E., Pitzer, K. S., and Spitzer, R. (1947) The Thermodynamics and Molecular Structure of Cyclopentane1. *J. Am. Chem. Soc.* **69(10)**, 2483–2488.

[311] Altona, C., Geise, H. J., and Romers, C. (1968) Conformation of non-aromatic ring Compounds–XXV : Geometry and conformation of ring D in some steroids from X-ray structure determinations. *Tetrahedron* **24(1)**, 13 – 32.

[312] Moore, P. B. (1999) Structural motifs in RNA. *Annu. Rev. Biochem.* **68**, 287–300.

[313] Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* **31(13)**, 3450–3460.