J Comput Aided Mol Des (2014) 28:1205–1215
DOI 10.1007/s10822-014-9797-0

Structure–activity relationships of thiostrepton derivatives: implications for rational drug design

Antje Wolf · Sebastian Schoof · Sascha Baumann · Hans-Dieter Arndt · Karl N. Kirschner

Received: 17 April 2014 / Accepted: 15 September 2014 / Published online: 4 October 2014

Abstract The bacterial ribosome is a major target of naturally occurring thiopeptides antibiotics. Studying thiopeptide (e.g. thiostrepton) binding to the GAR's 23S·L11 ribosomal subunit using docking methods is challenging. Regarding the target, the binding site is composed of a flexible protein–RNA nonbonded interface whose available crystal structure is of medium resolution. Regarding the ligands, the thiopeptides are chemically complex, flexible, and contain macrocycles. In this study we developed a combined MD–docking–MD workflow that allows us to study thiopeptide–23S·L11 binding. It is shown that docking thiostrepton-like ligands to an MD-refined receptor structure instead of the medium resolution crystal leads to better convergence to the native-like docking pose and a better reproduction of experimental binding affinities. By applying an energy decomposition analysis, we identify key structural binding elements within GAR's rRNA–protein binding site and within the ligand structures.

Keywords Thiostrepton · GTPase-associated region · Docking · MM-PBSA · Antibiotics

Introduction

Over the last 60 years the pharmaceutical industry has experienced periods of both ample and sparse antibiotic development. Unfortunately, during the past few decades there has been little success in developing new antibiotics with novel molecular scaffolds [1–3]. During the same time, antibiotic resistance, particularly in gram-negative bacteria, has significantly increased [4, 5]. Antibiotic candidates that act at new targets or via distinct mechanisms have the greatest potential to overcome resistance [6]. However, scientific challenges associated with the discovery of novel antibiotics, and a poor return on investment, have led to many companies leaving the field [7–9]. Since the early 1960s only four new antibiotic classes have

A. Wolf
Department Bioinformatics, Fraunhofer-Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany

A. Wolf
Department of Life Science Informatics, Bonn-Aachen International Center for Information Technology (B-IT), Dahlmannstrasse 2, 53113 Bonn, Germany

S. Schoof
BASF SE, 67056 Ludwigshafen, Germany

S. Baumann
Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus C2.3, 66123 Saarbrücken, Germany

H.-D. Arndt
Institute of Organic Chemistry and Macromolecular Chemistry, Friedrich Schiller University, Humboldtstr. 10, 07743 Jena, Germany

K. N. Kirschner (✉)
Department of Simulation Engineering, Fraunhofer-Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany
e-mail: k.n.kirschner@gmail.com

K. N. Kirschner
Bonn-Rhein-Sieg University of Applied Sciences, Grantham-Allee 20, 53757 Sankt Augustin, Germany

been introduced, and none of these have yet made a major impact [10]. Thus, the search for new antibiotics and new classes thereof continues to be important, as highlighted by the Infectious Diseases Society of America's call for 10 new antibiotics by 2020 [11].

A major target for natural antibiotics is the bacterial ribosome [12]. Through binding, these antibiotic compounds inhibit necessary steps for protein synthesis. Unlike other ribosomal antibiotic binding sites, the GTPase-associated region (GAR) is not yet targeted for human antibacterial therapy. This makes it an attractive drug-design target as cross resistance with other antibiotics is unexpected. The only known class of molecules that bind at this site, the thiopeptide antibiotics [13], exhibit nanomolar affinities in vitro; unfortunately, their low aqueous solubility preclude their use in human antibacterial therapy. Understanding their binding mode and inhibitory action could open up new structural classes for novel lead compound design. A recent medium resolution crystal structure of thiostrepton (TS, Fig. 2) bound to the GAR [14] makes the system accessible for computer-aided structural investigations.

For "classical" drug targets, computational methods are routinely and productively used to complement experimental design of new active ligands [15, 16]. Often these drug targets, typically enzymes, have well defined binding pockets suitable for a "lock-and-key" interaction [17]. Ligand binding often competes with an endogenous small molecule or substrate [18]. Computationally investigating ligand binding to the GAR region is challenging for several reasons. The binding site of thiopeptides, the GAR's 23S·L11 ribosomal subunit, is a loose and large macromolecular interface that is composed of rRNA (i.e. 23S) and protein (i.e. L11). rRNA–protein interfaces are less well studied in comparison to binding sites consisting of a single biomolecular species.

When targeting a heterogeneous binding environment one needs to optimize the ligand's interaction with chemically distinct biomolecular species. These interfacial binding sites differ from "classical" enzyme binding pockets due to their larger surface area, having more solvent contact, and being more flexible. When investigating such regions, not only must one consider possible "lock-and-key" or "induced-fit" binding mechanisms, but also the conformational selection paradigm [19–22]. Conformational selections suggest that large biomolecules and their interfaces naturally sample an ensemble of conformational states, and ligands selectively bind to one of these states due to their compatible three-dimensional (3D) structure.

Another challenge is posed by the chemical nature of the thiopeptide ligands. Thiopeptides have a large molecular weight (TS: 1,664.83 $g \cdot mol^{-1}$) and their 3D structures resemble protein folds [23]. Currently, ribosomal binding modes have been experimentally determined for TS and nosiheptide (NS) [14]. Although in vitro and in vivo activities of semi-synthetic thiostrepton derivatives [24, 25] and synthetic fragments [26, 27] have been reported, their 3D structures and exact binding modes at the GAR remain unknown. In summary, the challenges in investigating the binding of thiopeptide ligands to the GAR include a large heterogeneous interfacial binding region, a likely ensemble of rRNA–protein conformations [14, 28–30], and large macrocyclic ligands that have unknown 3D structures.

In a previous study we predicted the binding mode of TS and three derivatives by docking their NMR determined structures to a truncated GAR model, and correlated their molecular interactions with their biological activity [31]. However, the structural differences between the investigated compounds were limited. Herein, we extend this study to five further derivatives whose activities have been determined, but whose 3D structures are unknown. These compounds feature diverse chemical modifications at different positions of TS's scaffold, accompanied with measured GAR affinities that range over four orders of magnitude. Due to the size and complexity of the investigated compounds, 3D structure generators conventionally used in docking, such as CORINA [32], reach their limits. To predict reliable ternary complexes of the thiopeptides with the 23S·L11 subunit, we use an integrated docking and molecular dynamics (MD) approach. The MD simulations utilize a thiopeptide-tailored force field [33], providing good reliability and balance in the generated forces.

In our MD–Docking–MD workflow, (1) initial MD simulations are performed to generate refined input structures of (a) the receptor, and (b) the ligands, (2) ternary complexes are generated by docking, and (3) the resulting complexes are relaxed and rescored using a second MD simulation to account for flexibility and solvation effects that are not addressed in the docking. These MD simulations are used to calculate free energies of binding with the molecular mechanics generalized Born (Poisson Boltzmann) surface area (MM-GB(PB)SA) approach, and subsequently an energy decomposition analysis is also performed.

We show that using an MD-refined target structure, versus the use of a medium resolution crystal structure, can significantly improve the binding pose predictions and docking scores provided by molecular docking. By applying an energy decomposition analysis [34], we identify key structural binding elements within GAR's rRNA–protein binding site and within the ligand structures.

Results

The 3.3 Å crystal structure [14] of TS bound to the 50S ribosomal subunit has some uncertainty concerning the

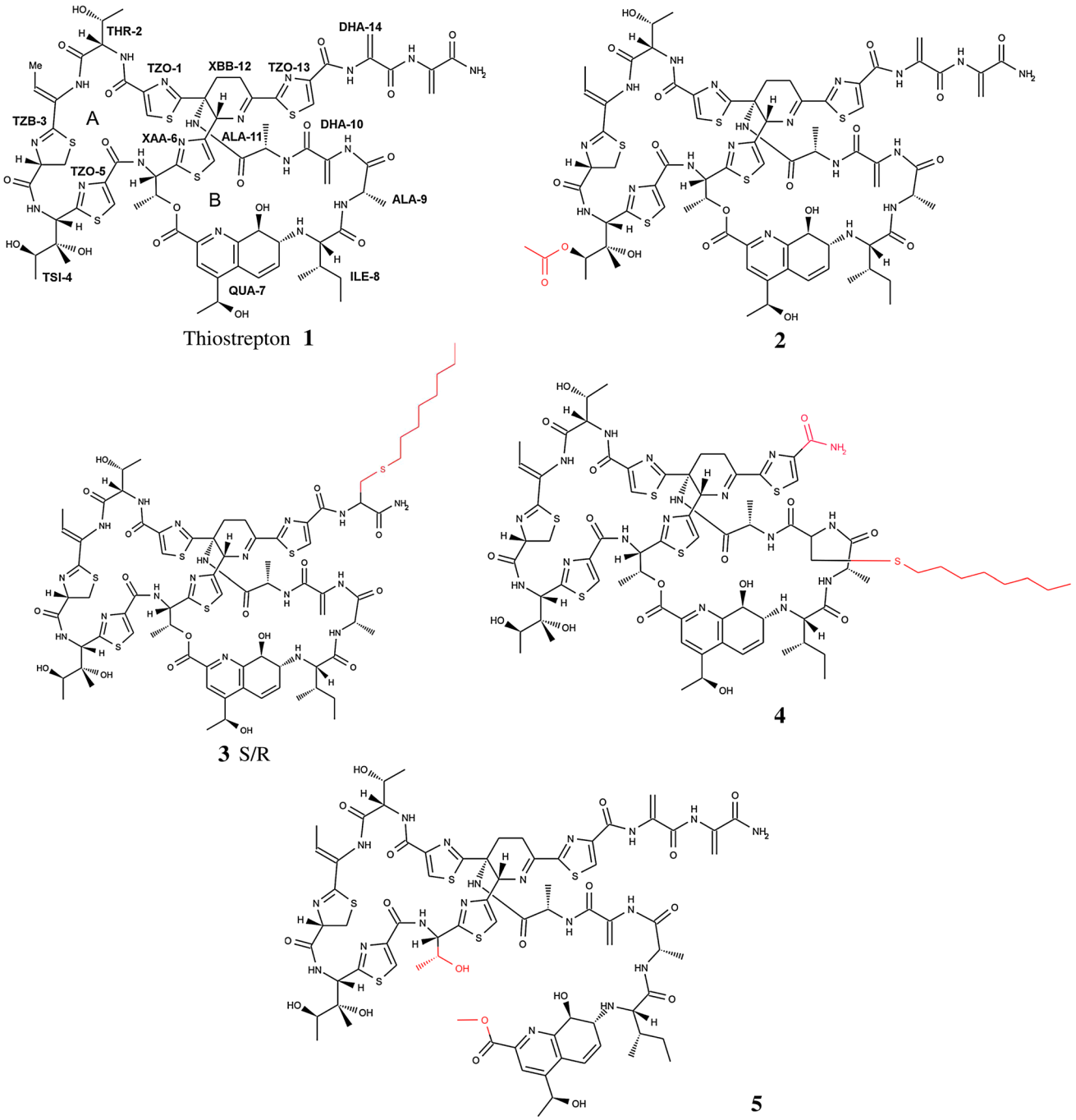

Fig. 1 Investigated TS derivatives

exact position of the residue side chains and of TS, making it a poor target structure for molecular docking studies. A model of the crystal structure's 23S·L11 ribosomal subunit was created by extracting the subunit, "capping" the resulting valencies, and then adding missing hydrogen atoms. Specifically, protein L11 and helices H43–H44 of the 23S rRNA where extracted, which will now be referred to as rRNA·L11. Subsequently, a molecular mechanics minimization was performed to relax the structure to its local potential energy minimum. Such a model building procedure results in a relaxed structure that closely resembles the original crystal structure.

To create an appropriate solution-phase model, and to overcome potential high-energy states arising from the crystal structure's medium resolution, we performed explicit solvent all-atom MD simulation at 310 K on TS bound to the rRNA·L11 ribosomal subunit. For full details of this simulation, please refer to reference [35]. An

Table 1 Experimentally measured dissociation constants (K_D) and docking results using the X-ray and MD-refined rRNA·L11 conformations of investigated TS derivatives

Cmpd.[a]	Experiment	Docking X-ray		Docking MD-refined	
	K_D (nM)	Score[b]	#Poses	Score[b]	#Poses
1	0.20 ± 0.05	−8.55	35	−10.36	68
2	1.80 ± 0.21	−6.87	1	−8.16	26
3a[c]	3.59 ± 1.20	−6.85	1	−8.46	23
3b[c]	5.41 ± 2.03	−5.58	1	−6.91	25
4	24.70 ± 2.86	−7.12	5	−7.60	82
5	670.00 ± 46.00	−6.00	1	−6.95	1
R^2		0.43		0.63	
PI		0.55		0.71	

[a] See Fig. 1

[b] Unit kcal·mol^{-1}

[c] Different K_D values could be determined for the two diastereomers; however they could not be unambiguously assigned to the R or S stereoisomer. Therefore, the K_D constants were assigned according to the AutoDock ranking

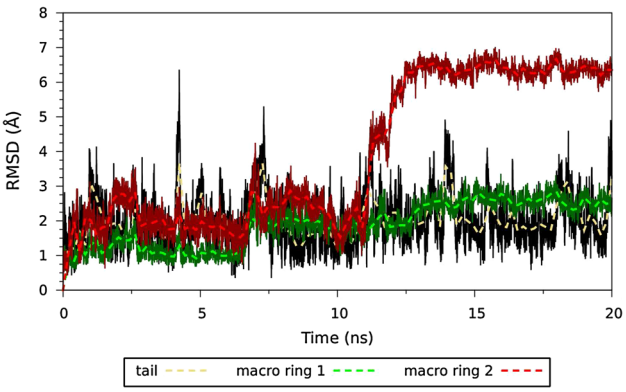

Fig. 2 Root-mean-square deviation (RMSD) from the input structure of **5** as a function of simulation time

rRNA·L11 conformation was extracted from the MD simulations that represents the average equilibrated complex. All docking studies were performed using (1) the minimized 3CF5 23S·L11 crystal structure and (2) the MD-refined structure as the targets.

Figure 1 displays the TS derivatives investigated in this study. They feature diverse modifications at different positions within the TS scaffold and four orders difference in the magnitude of measured affinities (see Table 1). To investigate a hypothesized structure–activity relationships (SAR), we performed molecular docking experiments using these derivatives. A prerequisite for docking are 3D structures for the receptor and the ligands. Since a 3D structure is only available for isolated TS in its crystalline state [23], an input structure for each derivative needed to be generated. Using TS structure as a 3D template, we modified the respective groups appropriately, performed a short minimization, and a subsequent 10 ns MD simulation in an organic solvent solution to relax them. Organic solvents were used to mimic potential NMR conditions [31].

To model the unique chemical functionalities (i.e. mixed organic and amino acid residues) present within the thiopeptides, a novel thiopeptide-tailored force field was developed that uses existing parm99SB parameters for the amino acids and optimized bond, angle and torsion parameters for the missing terms. These new parameters were developed to be balanced with parm99SB and transferable across small molecules with similar chemical functionalities [33]. The resulting tailored force field was used in the molecular mechanics and dynamics calculations.

The compounds' MD trajectory coordinates were clustered based on the macrocycles' RMSD between the snapshots, and the representative structure of the dominating clusters used for docking. One dominating cluster was seen in the compounds and, accordingly, only one conformation for compounds **1**–**5** was docked. Towards the end of the simulation of **5** a rise of the second macrocycle's RMSD was noticed (Fig. 2). Since **5** features the largest variation of the TS scaffold, a simulation length of 10 ns was obviously not long enough to fully stabilize its structure. Therefore, we extended the simulation to 20 ns and chose the representative structure of the last 5 ns as the docking input structure. While somewhat arbitrary, we chose the last 5 ns due to the reduced fluctuation and apparent improved stability in the last 10 ns relative to the conformations sampled in the first 10 ns. In future studies we will investigate more fully the structures that sample multiple conformations and how they effect the MD–Docking–MD workflow. Finally, the representative structure of compounds **1**–**5** was docked to the 23S·L11 subunit using AutoDock.

Choice of receptor conformation for docking

When the crystal coordinates of the 23S·L11 binding site were used as the target structure, the best ranked docking solution for all compounds, except for TS, was a inverted binding mode where the molecule's tail is inserted into the cleft formed between the rRNA and the protein. This binding mode is the result of a shape mismatch between the receptor and the ligand structures. When considering the largest cluster of docked conformations, a lower ranked TS-like binding mode is found for **2**, **4** and **3b** but not for **3a** (cluster 2:9 solutions) and **5** (cluster 5:1 solution). 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 1). In the majority of cases the first cluster was also

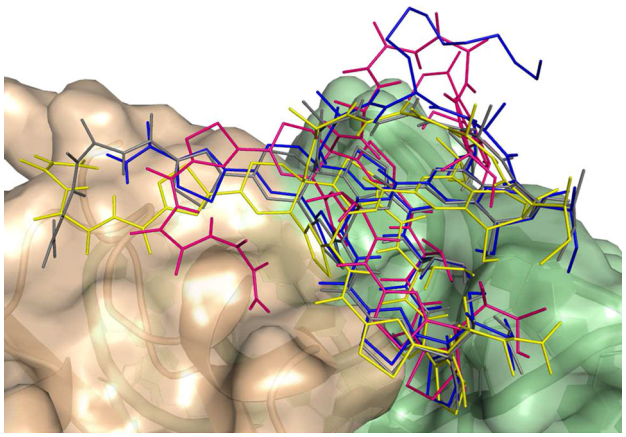

Fig. 3 Top-ranked docking poses of **1** (*gray*), **4** (*blue*), **2** (*yellow*), and **5** (*magenta*) using the MD-refined binding site conformation

the largest cluster (i.e. containing the most poses). The only exception is compound **3a** where the second cluster contains the largest number of poses (32). However, the observed binding mode and docking score (-8.42 kcal·mol^{-1}) are very similar to those of cluster 1.

Furthermore, the experimental binding constants are better reproduced when using the MD-refined receptor conformation (see Table 1). This holds true not only for the coefficient of determination (R^2) and the predictive index (PI), but also for the absolute binding energies that are consistently lower and thus nearer to the experimental values. In summary, using the MD-refined structure as the docking target instead of the medium resolution crystal structure led to a better convergence to the native-like docking pose and a better reproduction of experimental binding affinities, thus validating the refinement approach.

Stability and decomposition analysis

The MM-PBSA and MM-GBSA methods are often used for calculating binding affinities of biomolecular complexes [36, 37], and we have applied these methods to our complexes. In the case of the crystal structure, a rescoring with MM-GBSA and with MM-PBSA did not lead to any observable correlation. Conversely, when using the MD-refined binding site structure, MM-PBSA [PI = 0.56 and (R^2) = 0.42] clearly outperformed MM-GBSA (PI =0.29 and R^2 = 0.02). However, its prediction accuracy is still not as good as the AutoDock scores.

To gain a better understanding of the individual residue's contributions to the binding enthalpy, MM-GBSA residue and pairwise energy decompositions [34] were performed. The observations for each ligand are discussed in the following paragraphs.

Table 2 Enthalpic MM-GBSA contributions of key residues in binding TS, separated as van der Waals (vdW), electrostatic (Elec.), polar solvation (Polar solv.) and non-polar solvation (Non-polar solv.) contributions in kcal·mol^{-1}

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

Thiostrepton

When **1** was docked to the MD-refined structure, the 3CF5 crystal binding mode was consistently reproduced (see Fig. 3). 86 % of all poses were contained in the first two clusters, exhibiting similar docking scores of -10.36 and -10.30 kcal·mol^{-1}. When docked to the crystal structure, the convergence of docking poses was worse but still acceptable, with 44 % of all poses contained in the first two clusters, which reproduces the crystal binding mode. During the subsequent two 2.1 ns MD simulations of the complex, the position of **1** showed only slight deviations relative to its docked conformation (Fig. 4a), underlining the complex's stability.

When decomposing the binding enthalpy on a per-residue basis, only a few residues contributed significantly (Figs. 4b, 5). These are A1067 and A1095 on the rRNA side, and Pro21, Pro25, and Gln29 of L11's α1 helix. The importance of the two nucleobases for binding of **1** is well-known [38, 39]. Interestingly, the energetic contributions of the flanking nucleobases, except for the direct neighbors G1068 and A1096, are slightly positive. Although the van der Waals (vdW) and the electrostatic contributions of A1067 are higher than those of the other residues, its interaction energy is significantly decreased by unfavorable polar solvation (Table 2). In contrast to A1067, the contributions of A1095 are dominated by vdW interactions, mediated by a stacking interaction with the TS's residue TZO-1. Hydrophobic interactions occur between TZB-3 and Pro21, and between the residues around the central dehydro-piperidine-core (XBB-12) and Pro25 (Fig. 4c). From the residue-wise decomposition of TS (Figs. 4d, 5) 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 rRNA and the protein, mediating contacts between A1067 and A1068 of the 23S and Pro21 and Pro22 of L11 (Fig. 5).

Compound 3

Compound **3** was obtained by a Thiol–Michael addition at the dehydro-alanine unit within the truncated tail. The

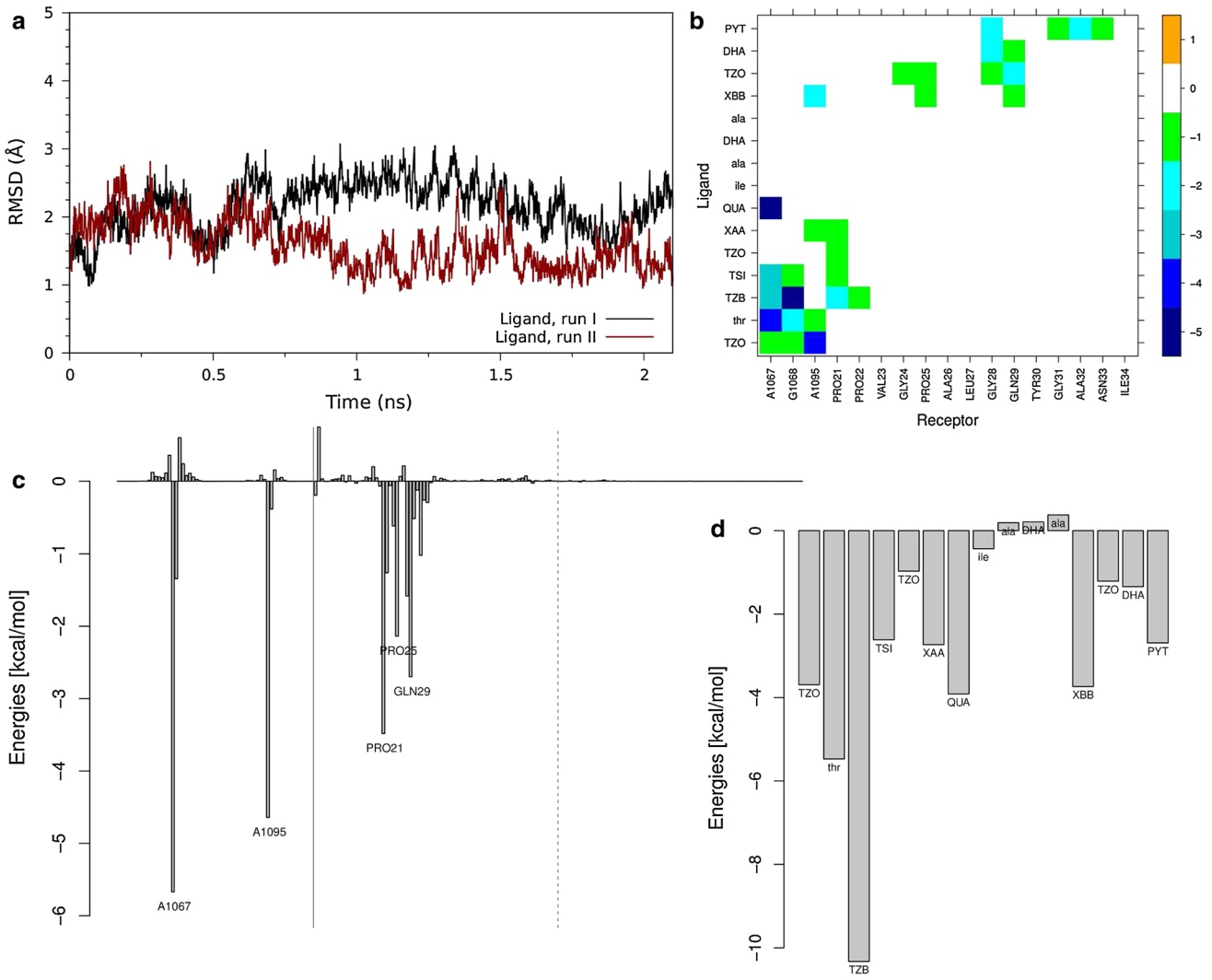

Fig. 4 MM-GBSA decomposition results for binding of TS. **a** RMSD of TS (without hydrogens) to the docked structure in the two rescoring MD simulations. **b** Pair-wise decomposition energies in kcal·mol^{-1}. **c** Residue-wise decomposition rRNA·L11. rRNA and protein residues are separated by a *black solid line*, the two protein domains are further divided by a *dashed line*. **d** Residue-wise decomposition ligand

stereochemistry of this group had a minor effect on the dissociation constant, but a 10-fold difference in activity was found in vivo [24]. Consequently, both diastereomers were modeled. As seen in Fig. 6, their binding modes are slightly divergent, especially regarding the tail. For both molecules 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 (−1.95 and −2.58 compared to −0.29 kcal·mol^{-1} for **1**, Table 3). On the other hand, the contribution of the polar Gln29 plays a smaller binding role. Due to steric restrictions, the interactions with Ile34 seem to be harder to accomplish for the S isomer whose interaction with A1095 is almost halved in comparison to TS. The MM-PBSA and MM-GBSA rescorings both preferred the largest cluster solution of **3a** and assigned it a very similar binding energy as that of the S stereoisomer (Table 3), consistent with the almost identical K_D values for the diastereomers (Table 1).

*Compound **4***

To obtain compound **4**, the octyl unit was added to macrocycle B's dehydro-alanine, and 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. Consequently, this residue was excluded from the MD trajectory's clustering, resulting in one large cluster whose representative structure was used as input for docking. After clustering the

Table 3 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

Cmpd.	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
3a	−6.17	−0.68	−6.03	−3.39	−2.32	−1.17	−1.95	−68.22	−45.19
3b	−5.17	−1.06	−2.47	−4.58	−1.95	−1.63	−2.58	−62.77	−46.98
4	−3.99	−1.23	−6.60	−1.71	−1.75	−2.05	−0.06	−47.33	−36.34
2	−1.37	−0.02	−2.88	−3.71	−1.42	−2.09	−0.60	−39.92	−35.21
5	−2.78	−3.14	−5.24	−4.07	−1.73	−1.43	−0.60	−53.62	−35.41

All values are in kcal·mol^{-1}

docking poses, the first cluster contained 82 poses, even more than TS, but obtained a slightly worse docking score (−7.60 kcal·mol^{-1}). The binding mode was very close to the one found for **1**, and had the octyl group sticking out into the solvent (Fig. 3). From the decomposition analysis (Table 3), the interactions with the key residues (A1067, A1095, Pro21, Pro25, and Gln29) are still present but considerably reduced, with the exception of A1095.

Compound 2

Compound **2** 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. 3) for **2**, a consistent shift of the docked position was noticed (RMSD ∼4 Å) in both MD simulations. The rRNA's energetic contributions are considerably reduced compared to TS (Table 3). For A1067 the lowest contribution among all derivatives was observed for **2**. Particularly, the electrostatic contribution is reduced (−1.29 kcal·mol^{-1}) compared to TS (−8.48 kcal·mol^{-1}). The added acetyl group leads to a steric clash and a loss of polar interactions with this nucleobase. Most likely, this is the cause for the observed shifting of the input docking pose that is seen during the MD simulations.

Compound 5

A transesterification of the macrocycle B lactone opened the macrocycle between residues XAA and QUA, leading to the derivative **5**. 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 relative to their position in the TS template (Fig. 2), and led to increased fluctuations of macrocycle B's residues. After ∼15 ns the RMSD became stable and the average conformation of the last 5 ns was selected as docking input. The docking of **5** did not converge, meaning that 100 single-pose-containing clusters were obtained. The molecule's many degrees of freedom (23 active torsions) are in the upper limit of AutoDock's search capabilities.

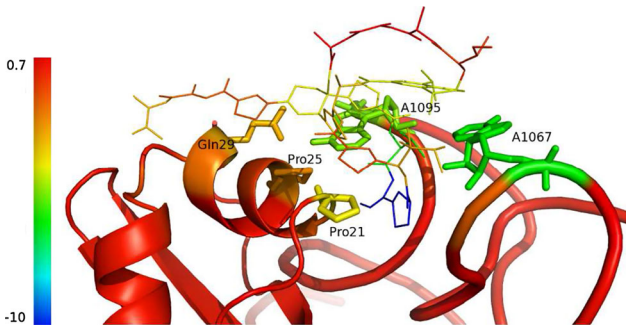

Fig. 5 MM-GBSA decomposition results mapped onto residues of the rRNA·L11·TS docked complex. Significant binding site residues are shown in sticks representation. Color-coding is done according residue contribution to the interaction enthalpies, measured in kcal·mol^{-1} (see *color bar*)

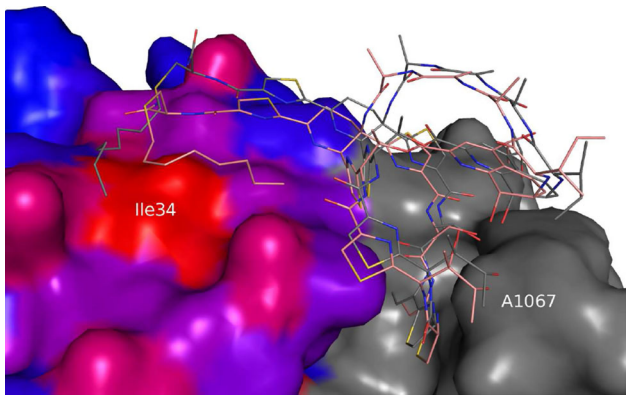

Fig. 6 Docking poses of stereoisomers of **3**. The ligands and rRNA are colored as **3b** = gray, **3a** = wheat, and rRNA = gray. The protein surface is colored according to hydrophobicity [40] (*blue* lowest hydrophobicity, *red* highest hydrophobicity). Figure created using PyMOL with rTools plugin

The shift in conformation seen within the MD simulation and the unconverged docking results inform us that

future studies involving significantly modified ligand (i.e. macroring openings) will require a more thorough MD simulation and docking study. Nevertheless, a binding mode similar to TS was found for the best scored pose (Fig. 3). 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 now opened macrocycle B's residues make additional contacts to the rRNA. This slight shift of the binding pose led to a binding energy decrease of A1067's contribution (-2.78 kcal·mol^{-1}) and a notable increase in G1068's binding energy (-3.14 kcal·mol^{-1}), the largest value seen in any of the compounds.

Discussion

The general goal of this study was to (a) develop a methodology for docking TS and derivatives thereof to the biomolecular heterogeneous L11·23S interfacial region, and (b) to gain insight into the important structural and residue features that are important for binding. The knowledge gained from this research will subsequently be used to guide the development of new computational-based predictions and organic synthesis towards the end goal of developing new active lead compounds. Due to the model's atomic size and the projection of performing the workflow a large number of times, we kept the MD simulation times low (i.e. 10–20 ns for the ligands, 2.1 ns for the complexes). Arguably these times are insufficient to obtain convergence in conformational sampling [41]; however, it appears to be generally sufficient to relax the local interactions due to ligand construction and those within the AutoDock generated complexes. These MD simulations times provide us with enough qualitative and quantitative (e.g. residue-wise energy decomposition analysis) data that allows an interpretation of the experimental binding data.

Using the crystal structure coordinates of L11·23S (i.e. 3CF5) as a docking target did not perform well for these macrocycle ligand studied, leading to posses that are inconsistent with experimental data. The conditions under which the X-ray structure was determined (i.e. preformed crystals of the complete 50S subunit soaked with a cryo-solution containing high concentrations of TS [14]) together with its medium resolution of 3.3 Å make it a less than ideal target structure for drug design purposes. Nevertheless, it is the only available atomic structural data of TS bound to the GAR, and provides valuable insights into the location and binding mode of the ligand. Immersing the rRNA·L11 fragment in explicit solvent, neutralizing it with ions, and performing a 40 ns MD simulation at body temperature [35] provided a target structure that lead to reasonable poses and stable trajectories.

In contrast to a previous study [42], an MM-PB(GB)SA rescoring could not further improve the native AutoDock score. Almost no correlation was found between experimentally determined binding affinities 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 [43]. 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 AutoDock-predicted binding of the TS derivatives investigated.

Conversely for MM-PBSA, adding the entropic component of ligand binding, as calculated by a normal mode analysis, led to a worse prediction (PI of 0.2 versus 0.67). It has been shown before that the inclusion of entropy is very difficult to properly accomplish [41, 44] and does not always improve the prediction accuracy of the MM-PB(GB)SA approach [45]. Large fluctuations of the vibrational entropies are often observed with normal mode analysis [42, 44]. It is likely that the low number of snapshots used in the normal mode analysis and the lack of statistical convergence in conformational sampling has led to inaccuracies in using this theory. A similar difficulty in using MM-PBSA for binding energy predictions was seen by Page and Bates [44] in their study of protein kinase inhibitors. However, they also concluded that energy decomposition is useful for gaining insight into ligand–receptor interactions.

Although the MM-PB(GB)SA rescoring did not lead to a better overall affinity prediction, the MM-GBSA energetic decomposition gave a clearer picture of the residues contributing to binding, both on the ligand and the receptor side. Among all derivatives, A1067 and A1095, and to a lesser extent Pro21, are the dominating residues of the rRNA–protein binding interface—consistent with a recent experimental study on several RNA and protein single-site mutations [46]. For all ligands, by far the largest contribution to the overall binding enthalpy was seen for the macrocycle A's TZB residue. TZB makes contacts to the purines A1067 and G1068 on the rRNA side and to prolines 21 and 22 on the protein side. In contrast, most of macrocycle B's residues do not contribute to the overall binding energy, with the exception of the QUA that has an ~ -4 kcal·mol^{-1} contribution.

Conclusion

Thiopeptides represent a challenging compound class for docking programs due to their molecular complexity, particularly the high number of rotatable bonds (i.e. 14–31) and the macrocycle scaffold. A correct input structure is crucial for obtaining reasonable binding configurations using AutoDock. Although the ligands are treated as flexible, the conformations of the macrocycle systems are not changed during the docking. Furthermore, the 3D structures of the derivatives investigated here were largely unknown. It was shown that generated thiopeptide models, using TS's experimentally known conformation as a template and 20 ns MD simulations with a tailored force field, were able to find the native-like binding mode and have their affinities reasonably predicted via docking.

It was shown that docking TS-like ligands to an MD-refined receptor structure instead of the medium resolution crystal leads to better convergence of native docking poses and a better reproduction of experimental binding affinities. However, all investigated ligands are structurally very similar to TS. Given the high adaptability of the binding site, the suitability of using a binding site conformation, derived solely from an MD simulation with a bound TS, to dock structurally distant ligands remains unclear. Using an ensemble of different target conformations, for example the cluster representatives from an "apo" simulation, might be a promising alternative approach for studying highly modified structures in the future.

In the future design of compounds, we will modify target residues within the TS macrocycle that have been indicated as important—either in stabilizing (e.g. TZB) or destabilizing the binding. We will also consider modification of the natural product scaffold that enables new or stronger interactions (e.g. dipole-dipole, hydrogen bonding) to occur with protein and rRNA residues. Prior to synthesis, we will computationally test these new molecules using the workflow described here, keeping in mind the observed strength and weakness of the applied methods. For example, if we modify a TS residue to be a stronger binder, then the MD–Docking portion of the workflow should provide us with a reasonable docking score and a good starting pose for the complex's MD simulation. Then an MM-GBSA residue-wise decomposition of the ligand—keeping in mind the difficulty of conformational sampling during the complex's MD simulation—should inform us if the proposed new molecule will provide experimental knowledge regarding specific binding interactions.

Materials and methods

Molecular dynamics simulation

All simulations were run with the AMBER 11 simulation package [47], using the SHAKE algorithm [48] on hydrogen atoms, a 2 fs time step and Langevin dynamics with a collision frequency of 1 ps^{-1} for temperature control. The structures were minimized in two steps: first with restraints on the solute, and second without restraints. Heating was performed with weak restraints on the solute in a constant volume simulation, the production runs were conducted at constant pressure without restraints. A cutoff of 9 Å was used, while the Particle Mesh Ewald method [49] was employed to capture the non-bonded interactions at longer distances. Coordinates were recorded every 1 ps.

Ligands

Intramolecular parameters for the thiopeptides' bonds, angles and torsions were taken from our optimized force field [33]. The modified compounds were immersed in a 5:1 chloroform:methanol box with 28 Å padding in all directions. We used this moderately polar organic solvent 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 100 ps, followed by a 10 ns production run. 1–4 electrostatic and vdW interactions were scaled by 1.0 each according to our parameterization scheme.

Complex

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 from the previous simulations, we kept the simulations rather 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, 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, snapshots for MM-PBSA computations were generated in two 2100 ps production runs 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 our force field and 1.2 (vdW)/2.0 (electrostatic) according to the

Parm99SB [50] and the Parmbsc0 [51] force field, used for the protein and rRNA, respectively.

Molecular docking

The docking studies were performed using AutoDock 4.2 [52]. The grid box was centered on the bound TS conformation and was chosen to be 70 × 70 × 70 which is slightly larger than the default (60 × 60 × 60) to accommodate the larger ligands used in this study. 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 MM-PBSA computations.

MM-GB(PB)SA computations

Molecular mechanics generalized Born (Poisson Boltzmann) surface area (MM-GB(PB)SA) computations were performed using the *MMPBSA.py* script of AMBER 11. Every 20th snapshot of each 2100 ps MD trajectory was used, 210 snapshots in summary. The ionic strength was set according to experimental conditions (0.04 M). For the GB calculations, the method of Onufriev et al. [53], model II, was used. Accordingly, the modified PBradii set 2 was used in LEaP when building the topology files. The PB equation was solved using the *pbsa* program of AmberTools. Nonpolar solvation contributions were approximated by the *LCPO* method [54]. 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.

Evaluation measures

We used two different metrics to evaluate the predictive capabilities of the AutoDock scoring function and the MM-GB(PB)SA approach for a set of ligands i with experimental E_i and corresponding predicted P_i binding energies. Beside the coefficient of determination R^2, defined as

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

we used the predictive index (PI) introduced by Pearlman and Charifson [55]. It measures how reliably a function can correctly choose the better binder for any pair of molecules:

$$\mathrm{PI} = \frac{\sum_{j>i} \sum_i w_{ij} C_{ij}}{\sum_{j>i} \sum_i w_{ij}} \tag{2}$$

with

$$w_{ij} = |E_j - E_i| \tag{3}$$

and

$$C_{ij} = \begin{cases} 1 & \text{if } [E_j - E_i]/[P_j - P_i] < 0 \\ -1 & \text{if } [E_j - E_i]/[P_j - P_i] > 0 \\ 0 & \text{if } [P_j - P_i] = 0 \end{cases} \tag{4}$$

It is normalized to −1 (always wrong), 0 (random), and +1 (always correct).

References

1. Projan SJ (2003) Why is big pharma getting out of antibacterial drug discovery? Curr Opin Microbiol 6:427–430
2. Fabbretti A, Gualerzi CO, Brandi L (2011) How to cope with the quest for new antibiotics. FEBS Lett 585:1673–1681
3. Moellering RC Jr (2011) Discovering new antimicrobial agents. Int J Antimicrob Agents 37:2–9
4. Coates ARM, Hu Y (2007) Novel approaches to developing new antibiotics for bacterial infections. Br J Pharmacol 152:1147–1154
5. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D et al (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin Infect Dis 48:1–12
6. Gwynn MN, Portnoy A, Rittenhouse SF, Payne DJ (2010) Challenges of antibacterial discovery revisited. Ann N Y Acad Sci 1213:5–19
7. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov 6:29–40
8. Jones D (2010) The antibacterial lead discovery challenge. Nat Rev Drug Discov 9:751–752
9. Braine T (2011) Race against time to develop new antibiotics. Bull World Health Organ 89:88–89
10. Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science 325:1089–1093
11. Infectious Diseases Society of America (2010) The 10 × '20 initiative: pursuing a global commitment to develop 10 new antibacterial drugs by 2020. Clin Infect Dis 50:1081–1083
12. Wilson DN, Nierhaus KH (2004) Antibiotics and the inhibition of ribosome function. In: Nierhaus KH, Wilson DN (eds) Protein synthesis and ribosome structure. Wiley, Weinheim, pp 449–527
13. Bagley MC, Dale JW, Merritt EA, Xiong X (2005) Thiopeptide antibiotics. Chem Rev 105:685–714
14. Harms JM, Wilson DN, Schlünzen F, Connell SR, Stachelhaus T et al (2008) Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. Mol Cell 30:26–38
15. Jorgensen WL (2004) The many roles of computation in drug discovery. Science 303:1813–1818
16. Taft CA, Da Silva VB, Da Silva CH (2008) Current topics in computer-aided drug design. J Pharm Sci 97:1089–1098
17. Fischer E (1894) Einfluss der Configuration auf die Wirkung der enzyme. Berichte der deutschen chemischen Gesellschaft 27:2985–2993

18. Hopkins AL, Groom CR (2002) The druggable genome. Nat Rev Drug Discov 1:727–730
19. Ma BY, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12:713–720
20. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci USA 96:9970–9972
21. Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. Protein Sci 8:1181–1190
22. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5:789–796
23. Bond CS, Shaw MP, Alphey MS, Hunter WN (2001) Structure of the macrocycle thiostrepton solved using the anomalous dispersion contribution of sulfur. Acta Crystallogr D 57:755–758
24. Schoof S, Baumann S, Ellinger B, Arndt HD (2009) A fluorescent probe for the 70 S-ribosomal GTPase-associated center. ChemBioChem 10:242–245
25. Myers CL, Hang PC, Ng G, Yuen J, Honek JF (2010) Semisynthetic analogues of thiostrepton delimit the critical nature of tail region modifications in the control of protein biosynthesis and antibacterial activity. Bioorg Med Chem 18:4231–4237
26. Nicolaou KC, Zak M, Rahimipour S, Estrada AA, Lee SH et al (2005) Discovery of a biologically active thiostrepton fragment. J Am Chem Soc 127:15042–15044
27. Starosta AL, Qin H, Mikolajka A, Leung GYC, Schwinghammer K et al (2009) Identification of distinct thiopeptide-antibiotic precursor lead compounds using translation machinery assays. Chem Biol 16:1087–1096
28. Li W, Sengupta J, Rath BK, Frank J (2006) Functional conformations of the L11-ribosomal RNA complex revealed by correlative analysis of cryo-EM and molecular dynamics simulations. RNA 12:1240–1253
29. Lee D, Walsh JD, Yu P, Markus MA, Choli-Papadopoulou T et al (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 367:1007–1022
30. Jonker HRA, Ilin S, Grimm SK, Wöhnert J, Schwalbe H (2007) L11 domain rearrangement upon binding to RNA and thiostrepton studied by NMR spectroscopy. Nucleic Acids Res 35:441–454
31. Jonker HRA, Baumann S, Wolf A, Schoof S, Hiller F et al (2011) NMR structures of thiostrepton derivatives for characterization of the ribosomal binding site. Angew Chem Int Ed Engl 50:3308–3312
32. CORINA. Molecular Networks GmbH. http://www.molecular-networks.com/products/corina
33. Wolf A, Reith D, Kirschner KN (2011) Thiopeptide antibiotics and the ribosomal 23S–L11 subunit: a challenging use case for semi-automatic force-field development. In: Carloni P, Hansmann UH, Lippert T, Meinke JH, Mohanty S et al (eds) From computational biophysics to systems biology (CBSB11). Jülich, Germany, pp 65–69
34. Gohlke H, Kiel C, Case DA (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 330:891–913
35. Wolf A, Baumann S, Arndt HD, Kirschner KN (2012) Influence of thiostrepton binding on the ribosomal GTPase associated region characterized by molecular dynamics simulation. Bioorg Med Chem 20:7194–7205
36. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) Validation and use of the MM-PBSA approach for drug discovery. J Med Chem 48:4040–4048
37. Li Y, Liu Z, Wang R (2010) Test MM-PB/SA on true conformational ensembles of protein–ligand complexes. J Chem Inf Model 50:1682–1692
38. Cameron DM, Thompson J, Gregory ST, March PE, Dahlberg AE (2004) Thiostrepton-resistant mutants of *Thermus thermophilus*. Nucleic Acids Res 32:3220–3227
39. Rosendahl G, Douthwaite S (1994) The antibiotics micrococcin and thiostrepton interact directly with 23S rRNA nucleotides 1067A and 1095A. Nucleic Acids Res 22:357–363
40. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132
41. Genheden S, Ryde U (2012) Will molecular dynamics simulations of proteins ever reach equilibrium? Phys Chem Chem Phys 14:8662–8677
42. Hou T, Wang J, Li Y, 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 51:69–82
43. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. J Comput Chem 28:1145–1152
44. Page CS, Bates PA (2006) Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? J Comput Chem 27:1990–2007
45. Hou T, Wang J, Li Y, 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 32:866–877
46. Baumann S, Schoof S, Bolten M, Haering C, Takagi M et al (2010) Molecular determinants of microbial resistance to thiopeptide antibiotics. J Am Chem Soc 132:6973–6981
47. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
48. Ryckaert JP, Ciccotti G, 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
49. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H et al (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593
50. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A et al (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725
51. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE et al (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys J 92:3817–3829
52. Morris GM, Goodsell DS, Huey R, Hart WE, Halliday RS et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662
53. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 55:383–394
54. Weiser J, Shenkin PS, Still WC (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comput Chem 20:217–230
55. Pearlman DA, Charifson PS (2001) Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. J Med Chem 44:3417–3423