

# Docking to RNA via Root-Mean-Square-Deviation-Driven Energy Minimization with Flexible Ligands and Flexible Targets

Christophe Guilbert and Thomas L. James\*

Department of Pharmaceutical Chemistry - MC 2280, 600 16th Street, University of California, San Francisco, California 94158-2517

Received January 23, 2008

Structure-based drug design is now well-established for proteins as a key first step in the lengthy process of developing new drugs. In many ways, RNA may be a better target to treat disease than a protein because it is upstream in the translation pathway, so inhibiting a single mRNA molecule could prevent the production of thousands of protein gene products. Virtual screening is often the starting point for structure-based drug design. However, computational docking of a small molecule to RNA seems to be more challenging than that to protein due to the higher intrinsic flexibility and highly charged structure of RNA. Previous attempts at docking to RNA showed the need for a new approach. We present here a novel algorithm using molecular simulation techniques to account for both nucleic acid and ligand flexibility. In this approach, with both the ligand and the receptor permitted some flexibility, they can bind one another via an induced fit, as the flexible ligand probes the surface of the receptor. A possible ligand can explore a low-energy path at the surface of the receptor by carrying out energy minimization with root-mean-square-distance constraints. Our procedure was tested on 57 RNA complexes (33 crystal and 24 NMR structures); this is the largest data set to date to reproduce experimental RNA binding poses. With our procedure, the lowest-energy conformations reproduced the experimental binding poses within an atomic root-mean-square deviation of 2.5 Å for 74% of tested complexes.

## INTRODUCTION

There have been successful efforts made to design ligands that bind to protein receptors using three-dimensional protein structures available via X-ray and NMR methods.<sup>1</sup> However, the increasing structural knowledge of RNA molecules and awareness of the essential role of RNA in development, gene expression, viral replication, blocking bacterial protein synthesis, and many other processes emphasizes the potential of certain RNA structural elements to be used as therapeutic targets for small molecule drug discovery. Similar to the active sites of proteins, RNA can form well-defined three-dimensional structures including bulges, hairpins, junctions, asymmetries in the deep and shallow grooves around noncanonical base pair clefts, and various binding pockets, all of which constitute attractive targets for specific and selective binding agents. RNA precedes proteins in the translation pathway; inhibiting the function of a single mRNA molecule can result in inhibition of multiple protein copies. Furthermore, there is precedence for RNA to serve as a target: the mechanism of action for many clinically important antibiotics entails binding to specific sites on rRNA.

Recent reviews discuss potential benefits of targeting RNA for the development of new antibacterial and antiviral drugs.<sup>2,3</sup> Some RNA-targeting examples include drugs such as aminoglycosides, macrolides, and other antibiotics, which are used heavily to fight bacterial infection by specifically targeting the RNA in the bacterial ribosome.<sup>4,5</sup> Their

widespread clinical use demonstrates that RNA is an important drug target. The development of novel antibacterials is the prime response to the threat presented by increasingly widespread resistance to existing antibiotics. Among the revealing X-ray structures of ribosomal subunits, one revealed the binding of aminoglycosides to the A site of the 16S rRNA (rRNA), their apparent functional target.<sup>6</sup> Such binding blocks the initiation of translation, elicits premature termination, and causes miscoding, ultimately leading to bacterial cell death. In addition to these well-known clinical applications of mostly natural product analogs and one FDA-approved, entirely synthetic compound, an oxazolidinone, as antibacterials, there has been some promising work targeting other RNA targets. Examples of other important targets include human immunodeficiency virus type 1 (HIV-1) RNA, specifically, the trans-activating response element (TAR),<sup>7–10</sup> and the Rev response element.<sup>11</sup> The functions of these regulatory RNA motifs of HIV-1 are impeded by ligands from interacting with their cognate viral target proteins Tat and Rev. Another HIV-1 RNA target is the dimerization initiation site, which binds aminoglycosides and may inhibit RNA genome packaging.<sup>12</sup> Aminoglycosides also act as inhibitors of catalytic RNAs such as the self-splicing group I introns,<sup>13,14</sup> the hepatitis delta virus ribozymes,<sup>15,16</sup> and the hammerhead ribozymes of some plant viroids.<sup>17</sup> More recently, in the development of anticancer therapies, the telomerase RNA/DNA heteroduplex formed during the catalytic cycle of telomerase was identified as a potential site for inhibition.<sup>18</sup>

There are typically two experimental approaches used to find low-molecular-weight compounds to bind to a new

\* Author to whom correspondence should be addressed. Phone: 415-476-1916. Fax: 415-502-8298. E-mail: james@picasso.ucsf.edu.

protein (or, rarely, RNA) target: (a) high-throughput screening of  $10^4$  to  $10^6$  compounds or (b) using the structure of a receptor with a ligand to suggest other compounds to synthesize and test. Both approaches are expensive but have certainly had some success. However, computational screening (also called virtual screening (VS) or docking) of a large database of small molecules for binding to the target structure can also be employed as the first stage in novel ligand discovery in the case that a structure of the receptor, either with or without the ligand bound, is available.

Kuntz et al. were the first to use the flexible-ligand docking program DOCK to identify compounds in the Available Chemicals Directory (ACD) that would bind to double-helical RNA.<sup>19</sup> We successfully screened the ACD to identify novel ligands binding to TAR RNA using rigid-ligand docking with DOCK followed by fully flexible-ligand docking via ICM.<sup>8,20</sup> ICM uses Monte Carlo minimization in torsion angle space with progressively more detailed conformational sampling on a progressively smaller list of top-ranking compounds.<sup>21</sup> Use of the MCSS method, which took advantage of the nucleic acid parameters developed in the CHARMM force field, was also proposed.<sup>22</sup> Studies were also carried out using a combination of docking methods and molecular dynamics simulations.<sup>23,24</sup> In the case of TAR RNA, docking and modeling RNA with bound ligands has resulted in characterization of and enhancement of ligand binding features.<sup>25–29</sup> Recently, automated docking methods specifically parametrized for docking to RNA were proposed by various authors: Morley and Afshar developed a new RNA-specific regression-based scoring function called RiboDock,<sup>26</sup> which was trained and validated on a limited set of 10 RNA–ligand complexes. Detering and Varani systematically tested automated docking tools developed for proteins using another set of experimental three-dimensional structures of RNA–small molecule complexes.<sup>30</sup> The results show that the native structures can generally be reproduced to within 2.5 Å about 50–60% of the time on a data set of 16 RNA–ligand complexes when using the original force field parameters developed for protein binders. Moitessier et al. recently reported docking to hydrated and flexible RNA.<sup>31</sup> However, this approach required extensive experimental information such as cocrystal structures of several ligands bound to the same target. It was tested for 11 aminoglycosides bound to the ribosomal A-site RNA. More recently, Pfeffer and Gohlke reported a new knowledge-based scoring function, Drug-Score<sup>RNA</sup>, to predict RNA–ligand interactions on the basis of a data set of 31 RNA–ligand complexes by deriving energetic information from the statistical analysis of structural parameters for known biomolecular complexes.<sup>32</sup> They used the Autodock program,<sup>33</sup> which holds the RNA receptor rigid during the docking process, and they obtained a success rate of 42% using a 2 Å cutoff.

Success rates of 70–90% are typically obtained on protein–ligand test sets using nearly any of the many docking programs. However, the few comparable studies on smaller RNA–ligand test sets yield success rates of 40–60%. Two main factors help explain the relative lack of VS accuracy for RNA compared to proteins: (a) the lack of a robust empirical scoring function, which can discriminate real binders from nonbinders during the docking process and which can estimate the free binding energy for ranking the best ligands, and (b) the necessity to include RNA flexibility

in ligand docking. Good scoring functions and flexibility are probably the two biggest issues for protein docking, but they pose a significant problem for RNA docking because there are relatively few experimental RNA–ligand structures for the development of scoring functions, and functional RNA targets can be notoriously flexible. In the study presented here, we are primarily concerned with addressing the issue of the malleability of RNA upon ligand binding.

Several docking programs treat the receptor as a rigid object and cannot account for conformational changes upon binding. This simplification results in relatively poor cross-docking results and low enrichment factors in virtual screening studies.<sup>34</sup> Analyses of RNA have established that a complex and its free component molecules often differ in structural details. Examples for which an induced fit has been demonstrated upon small ligand binding include HIV-1 TAR RNA and bacterial A-site rRNA, which has been discussed.<sup>2</sup> This reflects the conformational plasticity of RNA often observed upon binding to a protein partner. A single structure, or even the weighted average provided by a crystal structure or NMR structure, may not adequately describe the conformations (substates) accessible to the target RNA. The “structure” determined for RNA structural elements other than double helix typically represents an average or possibly the lowest-energy conformer among several that differ only slightly in energy.<sup>35,36</sup> The populations of conformational states depend not only on conditions such as pH, temperature, and ionic strength but also, very importantly, on the introduction of a ligand into the system. Upon ligand binding, many systems undergo rearrangements, which range from local motions of side chains in proteins or base pairing in nucleic acids to large domain movements. We may envision that the initial interaction between a ligand and the receptor in solution would occur with preference for partly fitting receptor and ligand conformations among an ensemble of conformers followed by readjustments to the final stable complex. A combination of conformational selection and induced fit would seem to be the best description of the interaction between molecules that do not fit optimally via their unbound conformations. However, considering the kinetics and thermodynamics of a binding reaction, the induced fit is possible only if the match between the interacting sites is strong enough to provide the initial complex with enough strength and longevity that the subsequent induced fit can take place within a reasonable time.<sup>37</sup>

For protein-based drug design, the challenge of docking a flexible ligand to a receptor with varying degrees of flexibility has been addressed by a few groups. A plethora of recent good reviews,<sup>38–41</sup> and descriptions of methodology,<sup>21,31,42–44</sup> for flexible receptor docking have been published. However, in most of the actual docking programs used in these schemes dealing with receptor flexibility, the receptor is fixed for all of the calculations. For RNA-based drug design, there was one RNA docking study where some receptor flexibility was permitted for the best-scoring ligands resulting from rigid-receptor docking.<sup>8</sup>

One approach to accommodate small changes in conformation utilizes a “soft” scoring function, which permits some overlap between the ligand and the protein, allowing for a small, limited plasticity of the receptor.<sup>45</sup> Soft-docking also has the benefit of being fast. Significant improvement over single structure docking has been achieved through the use

of several receptor structures to mimic flexibility in the docking process; rigid receptor docking is simply carried out for several different receptor conformations. These different conformations could come from multiple experimental structures when available.<sup>38,46–49</sup> Alternatively, they could be simulated via molecular dynamics<sup>50–52</sup> or normal mode calculations.<sup>53–55</sup> However, docking many individual structures will be computationally limited with a large database to screen. For efficiency, multiple conformations would need to be combined and incorporated into the VS procedure, which is a nontrivial task. Autodock provides one means of doing so:<sup>33</sup> it uses grids that can be combined to approximate an ensemble of rigid conformations. One might also question whether or not the structures selected to represent the ensemble of receptor conformations truly represent those found in solution. Some docking programs use stochastic methods, such as Monte Carlo, which allow side-chain flexibility,<sup>21,43</sup> and even minor backbone rearrangements through molecular minimization.<sup>8,43</sup> Although incorporating side-chain flexibility was a big step forward, screening using induced fitting of both the ligand and the receptor simultaneously has still not been reported.

In the present paper, we describe a new procedure for docking ligands to RNA, which includes the flexibility of the receptor over the course of the binding search using molecular minimization techniques. This enables an induced fit of the ligand and receptor. The principle of the method is basically to add an atomic root-mean-square-deviation (rmsd) constraint energy term to the potential energy, which drives the ligand to probe the surface of the receptor using energy minimization. The use of a rmsd penalty in addition to the regular potential energy in molecular simulations is not new.<sup>56,57</sup> Such a rmsd-type force can be used to explore the conformational space of a protein, simulate large-scale molecular motions such as hinge bending in proteins, and simulate the transition pathway between two known structures.<sup>57</sup>

The program we have developed, MORDOR (Molecular Recognition with a Driven dynamics Optimizer), allows induced-fit binding of small molecule ligands with RNA targets via flexible-receptor, flexible-ligand docking. Rigorous molecular force fields are used: CHARMM-27 is used in the calculations for the receptor,<sup>58</sup> and the general AMBER force field is used for ligand molecules.<sup>59</sup> The protocol for using MORDOR is outlined. The importance of including the internal energy of the receptor in addition to ligand–receptor interaction energy in distinguishing valid ligand binding poses is also demonstrated.

To evaluate MORDOR, we have applied it to a data set of 57 RNA–ligand complex structures. The validities of our sampling procedure and our scoring function were assessed in terms of the ability to reconstruct the native RNA–ligand complex geometries. Our data set is the largest published so far and includes nearly all of the complexes used by other groups to validate their VS procedures; hence, our results can be compared to those studies.

## METHODS

**Data Set.** In order to test the performance of the MORDOR docking procedure, we compiled a list of complexes that were labeled as containing both an RNA

molecule and a small-molecule ligand from the protein data bank (PDB). We reviewed them all by hand and initially selected 40 of them (20 crystal and 20 NMR structures) for our study. Our selection criteria included (a) diversity of the receptors and ligands; (b) resolution of X-ray structures in cases where more than one structure was available for a complex; and (c) avoidance of structures with certain problems, such as structures solved with incorrect chiral centers, complexes with significant intermolecular contacts between ligands due to crystal packing, and structures with missing residues in the binding pocket due to missing electron density. For comparison purposes, we later added 17 more complexes used in previous studies.

RNA structures were prepared using the following principles. All RNA receptor–ligand complex structures were retrieved from the Protein Data Bank.<sup>60</sup> Since our data set uses three-dimensional RNA structures solved by NMR techniques, which unlike some crystal structures do not reveal the positions of metal ions and water, all counterions and water molecules in the PDB files were removed. For crystal structures solved with multiple occupancies of individual atoms, the highest occupancy was selected. For structures with multiple binding sites, we used the first site encountered in the PDB file. In cases where multiple experimental NMR structures were available or there were crystal structures with multiple independent molecules in the crystallographic unit cell, again, we took the first one encountered in the PDB file.

**Preparation of the Ligands.** The ligands were stripped from the receptor and “regularized” using OpenEye’s convert.py tool (OpenEye Scientific Software, <http://www.eyesopen.com/>), which protonates the ligand at pH 7.0. Next, the topology and parameters using the general GAFF force field and AM1-BCC partial charges were calculated using ANTECHAMBER.<sup>61</sup> The ligands were ready for subsequent docking after structures were minimized in vacuo and randomly rotated two times.

**Preparation of the Receptor.** Hydrogen atoms were added to the receptor and optimized in the presence of the ligand using the algorithm Hbuild in CHARMM.<sup>62</sup> The complexes were then minimized with CHARMM c34b1 using topology and parameters of CHARMM and GAFF for the receptors and ligands, respectively. The receptors were ready for docking after deleting the ligand. After minimization, all of the receptors had an average rmsd of 0.4 Å (ranging from 0.11 to 0.6 Å) from their experimental counterpart (data not shown).

The final data set had a total of 57 structures, which are listed in Table S1 of the Supporting Information. Our list includes 33 X-ray and 24 NMR structures. The resolution of the 33 experimental crystal structures ranges from 1.3 to 3.3 Å.

**MORDOR—Creation of Binding Site and Docking.** All calculations reported in this work were performed using a new software tool we developed in our laboratory called MORDOR. It is written in C++ and uses Python for script interface. MORDOR contains all of the tools for virtual screening. The software can perform the main docking using its own energetic and minimization routines, but it is also interfaced with ANTECHAMBER for building ligand databases,<sup>61</sup> which include ligand topology and parameters, and CHARMM for some state-of-the-art molecular simulation



features,<sup>62</sup> such as calculating the solvation energy. MORDOR can also find the binding sites at the surface of the receptor to initially place the ligands. Docking results from VS can be displayed in the VIEWDOCK module of the visualization software CHIMERA when both the ligand and the flexible receptor are displayed.<sup>63</sup> CHIMERA generates a table from MORDOR results, grouping various useful parameter columns, for example, binding energy, charge, and molecular weight. Each column in the table may be used for sorting, and the receptor–ligand complexes are visualized according to that order by selecting the row corresponding to the desired complexes. Inspection of the docking results is consequently very convenient.

**Sphere Generation.** In order to identify the regions on the receptor most likely to constitute a useful surface for the binding of small molecules, that is, a binding pocket, we generate spheres, which are used for the initial placement of the ligand to a “hot spot” before starting the docking. After the initial placement, the ligand is perfectly free to move away from the spheres during the docking protocol described below. The procedure of sphere generation consists of mapping the entire receptor with a probe that is represented by a dummy atom of varying radius (between 1 and 4.0 Å with a step of 0.2 Å) and varying formal charge (from −1 to +1 with a step of 0.1). The receptor is completely embedded in a three-dimensional grid with 1 Å spacing. The sphere with a given radius and charge is initially placed on one node of the grid, and its position is energy-minimized (van der Waals and electrostatic), so that the sphere can find the lowest-energy pocket in the vicinity at the surface of the receptor. This step is carried out repeatedly for all nodes of the grid. The receptor is held fixed throughout this process. This calculation is repeated for each combination of sphere radii and charge values.

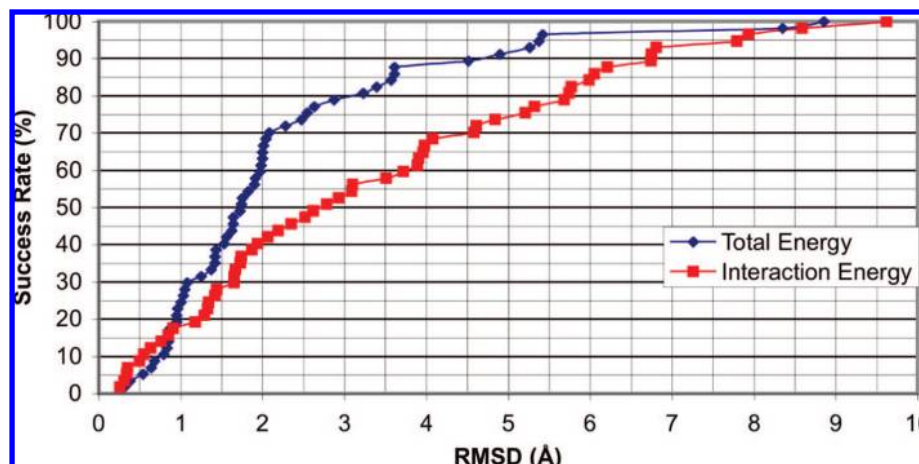
Finally, all of the resulting spheres are clustered and ranked according to their interaction energy with the receptor, assuring that they are at least 1.4 Å one from each other. We define the active site by selecting the 10 lowest-energy spheres found to be within a radius of 10 Å from any nonproton atom of the ligand in the experimental structure. We find that these spheres are remarkably able to find the experimental binding pocket but can sometimes identify another nearby “hot spot” as well (data not shown). The details of this approach to identify the hot spot regions will be more fully explained elsewhere.

**Docking via MORDOR.** The MORDOR method of docking a ligand to the receptor entails three steps. These are as follows:

The first step consists of placing the ligand at one of the “hot spots” on the receptor, which are defined by the spheres, by positioning one atom of the ligand in the center of one of the spheres. While the receptor is held rigid, the complex is very quickly minimized to place the ligand globally by removing any clashes between the receptor and the ligand. If no clashes are found, that is, no atom of the ligand is < 1.5 Å from any atom of the receptor, the complex goes to the second step. The same procedure was repeated 120 times with different orientations of the ligand. The initial orientation of the ligand was obtained by aligning the *x* axis of the ligand with one of the 120 vectors uniformly distributed on the trigonometric sphere. The receptor was held rigid during this procedure.

The second step entails energy minimization of the receptor–ligand complex; it allows the ligand to explore the receptor surface using a Path Exploration with Distance Constraints (PEDC) algorithm,<sup>57</sup> enabling an induced fit of both the ligand and the receptor. In this process, PEDC forces the ligand to probe the receptor, now permitted to have flexibility, thus exploring its surface. PEDC adds an atomic rmsd penalty to the potential energy. This rmsd penalty, or “distance constraint”,<sup>57</sup> is used as a driving force to constantly change the position of the ligand; the rmsd is calculated between the current and the reference ligand positions, where the reference is one of the previous ligand positions. The ligand movement is induced by setting a target distance (or target rmsd) value, which will “constrain” the ligand to move (using minimization) from its current position to a position where the rmsd from the reference structure will match the target rmsd value. The potential is applied in such a way that the ligand is being “pushed away”, beyond the target rmsd value, from the reference position. With a force constant of 500 kcal/(mol·Å<sup>2</sup>), the penalty decreases quadratically as the rmsd approaches the target value; that is, we pay a penalty when the rmsd is less than the target value; the penalty becomes zero when the rmsd is greater than the target value. The target rmsd is incrementally increased in small steps (0.1 Å); the receptor–ligand complex is minimized for each target rmsd. If the ligand rmsd reaches or exceeds the target rmsd, this minimum is stored as a binding pose for a subsequent rescoring in step 3; the reference is reset to the previous ligand position, and the target rmsd is reset to 0.1 Å. Otherwise, the target rmsd is incremented, and the complex is minimized again. The docking is terminated, and the whole procedure is repeated starting with the next orientation or sphere in step 1, when one of the following conditions is fulfilled: five minima have already been restored or the target rmsd exceeds 10 Å. In this way, the ligand explores many potential binding pockets, which could be as far as 20 Å from its initial position defined in step 1.

Because the energy minimization is performed in vacuo, the ligand while probing the surface of the receptor could favor electrostatic interactions and lead to unrealistic receptor conformations that, once established, are almost irreversible using minimization. To avoid this problem, the receptor was restrained during docking with another flat-well atomic rmsd potential, which allowed the receptor’s heavy atoms to change freely within an overall rmsd of 0.3 Å from the initial minimized structure but added a penalty with a force constant of 800 kcal/(mol·Å<sup>2</sup>) beyond 0.3 Å. The flat-well width of 0.3 Å and the force constants for the receptor and ligand rmsd potentials were chosen as a compromise to keep the receptor close enough to the minimized native structure and yet allow enough local flexibility to accommodate changes associated with the induced fit. Practically, with the protocol and force constant used in this study, the total receptor’s rmsd from the minimized experimental structures is in the range of 0.1–0.6 Å for the best-scoring poses but can be as high as 1.0 Å in some minima (stored poses). During the docking, the penalty driving movement of the ligand is strong enough to overcome the receptor restraint, so rmsd values for local conformational deviations may exceed the total rmsd of the receptor. For the binding interface (defined as all residues of the receptor within 5 Å of any ligand’s atom),



**Figure 1.** Success rate for MORDOR to predict the experimental binding pose for the data set of 57 RNA–ligand complexes as a function of the rmsd cutoff criterion chosen. The rmsd is calculated between the experimental structure and the best-scoring MORDOR prediction for all nonproton atoms of the ligand. Scoring was based on either the interaction energy (square) or total energy (diamond).

the local rmsd is within 0.1–0.8 Å for the best-scoring poses, and it can be as high as 2.1 Å in some minima, and even transiently reach 4.25 Å in the process of minimization. Allowing such local flexibility is especially important, if there is an induced fit of the receptor associated with ligand binding.

Steps 1 and 2 of the procedure are repeated for each heavy atom of the ligand for that initially chosen sphere. The whole procedure is then repeated for each of the spheres. (Typically, there are 10 spheres.) In this way, the ligand has been placed in various orientations at several different hot spots on the surface of the defined receptor (step 1), with numerous explorations of induced fitting of the ligand and receptor (step 2), thus enabling a broad sampling of induced-fit ligand binding to the receptor.

The third and last step of the procedure consists of rescoring the ensemble of docking poses for a given ligand with a more advanced scoring function (*vide infra*), which includes solvation energy, in an effort to improve the affinity prediction. The best total energy of the complex was used to rank and discriminate among the poses.

**Potential Energy and Scoring Function.** The docking simulations were conducted using the all-atom CHARMM27 force field for the RNA receptors<sup>58</sup> and the general AMBER force field for ligand molecules.<sup>59</sup> Both share the same classical functional form of potential energy. The AM1-BCC model charges were computed using ANTECHAMBER for the ligands.<sup>61</sup> Since the docking calculations are done in vacuo, for the sake of efficiency, we used a distance-dependent dielectric model with  $\epsilon(r) = r$ , where  $r$  is the distance between two atoms. We also used a nonbonded cutoff distance of 12 Å, a shifted function of the electrostatic interaction that began at 8 Å and ended at 10 Å, and a switched function for the Lennard-Jones potential.<sup>62</sup>

The scoring function, which helps to distinguish the correct docking poses from incorrect ones during the rescoring step, uses the total energy of the complexes. The components of the energy are in the classic forms found in CHARMM or AMBER and are composed of electrostatic, van der Waals, dihedral angle, torsion angle, bond, and Urey–Bradley terms in addition to the more sophisticated implicit solvent treatment: Generalized Born with simple switching (GMSW),<sup>64</sup> in CHARMM. The default GBSW parameters were used.

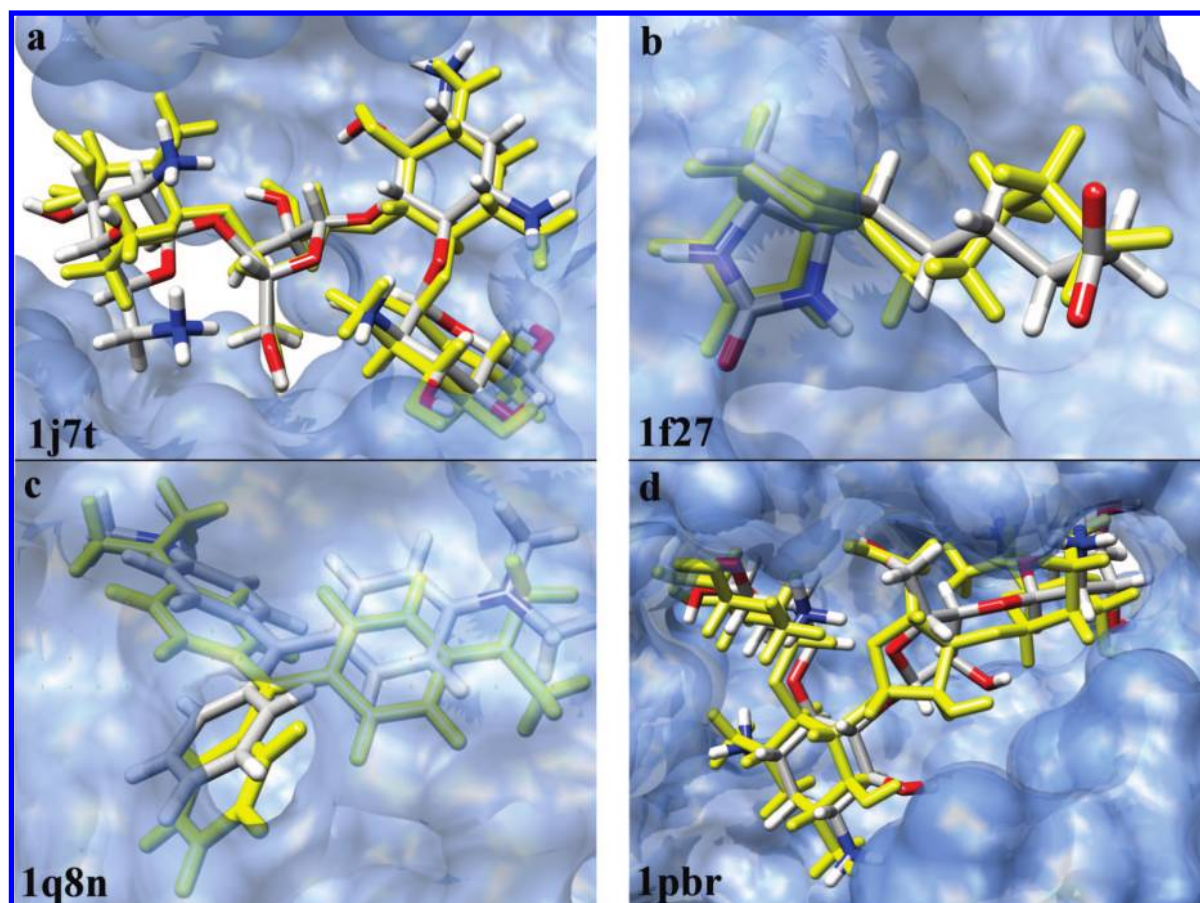
A small video illustrating the MORDOR procedure can be found at <http://mondale.ucsf.edu/jcim2008>. Typically, a docking run using MORDOR took from 0.5 to 3 h (depending on the receptor and ligand size) on a single AMD Athlon 64 3200 GHz CPU. This combination of software and hardware can typically provide virtual screening of a small database ligand library of 50 000 compounds within a few weeks using current technology clusters.

## RESULTS AND DISCUSSION

We explored the ability of MORDOR to reconstruct native RNA–ligand complex geometries in docking experiments. The data set used for docking consists of 57 RNA–ligand complexes. To assess how well MORDOR docking was able to reproduce the experimental RNA–ligand complex structures, rmsd values were calculated to characterize the difference between the ligand bound to the experimental structure and the conformation predicted by the docking program. For this purpose, the experimental and predicted complexes were first superimposed on the basis of the heavy atoms of the binding pockets of the receptors. The binding pocket was defined by all the residues that have at least one heavy atom within 5 Å of any heavy atom of the ligand. Then, the rmsd value was calculated using the heavy atoms of experimental and predicted positions of the ligand.

The individual rmsd values are reported for each tested complex in the Supporting Information (Table S1), and the data for all 57 complexes are summarized in Figure 1. These rmsd values can be used to assess the success of the docking. Many groups using rigid receptor docking consider a rmsd cutoff of 2 Å between the docking and experimental poses as the metric for success, although more recently for docking to RNA, a cutoff of 2.5 Å was used.<sup>30</sup> Although the choice of this cutoff is quite arbitrary, ideally, it should reflect the errors in determining the experimental structures (which are method-dependent) and real fluctuations in the ligand pose expected in solution at room temperature (which may be receptor- and ligand-specific). Typically, molecular structures determined by NMR are represented by an ensemble of structures, each of which is deemed to satisfy the experimental data. Analysis of such ensembles in the Protein Data Bank revealed that on average the position of





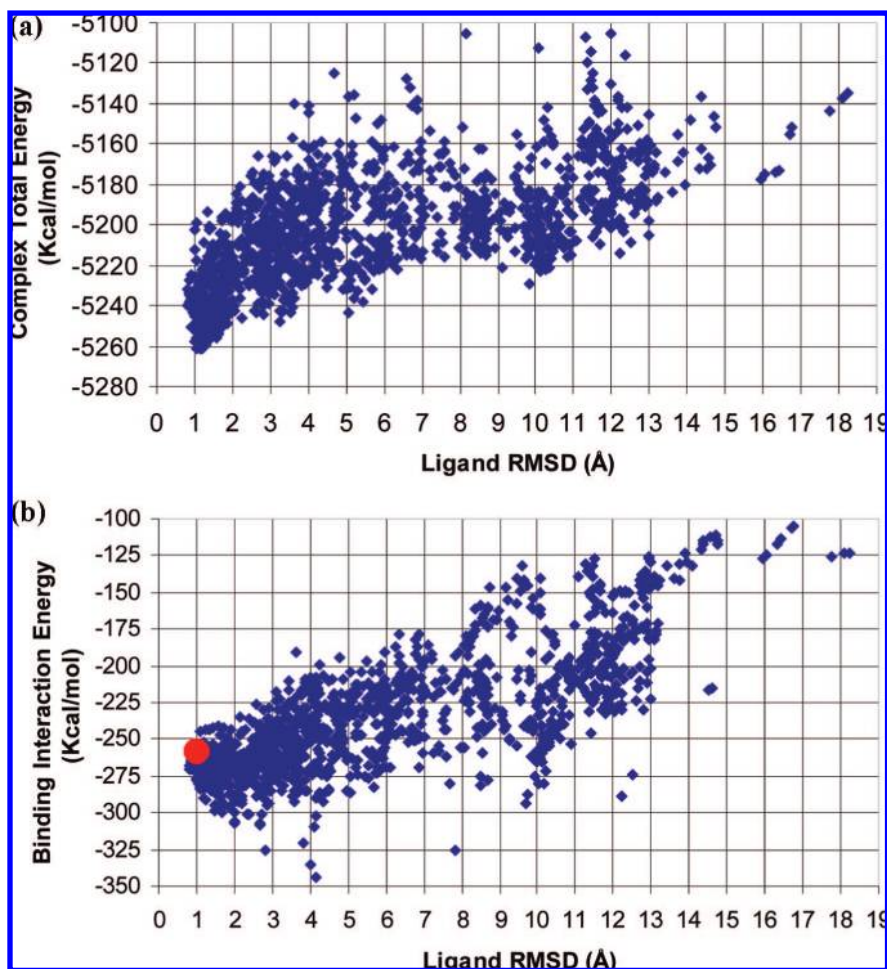
**Figure 2.** Examples of some correct poses predicted by MORDOR: the best pose with the highest score is shown in all-atom color, and the ligand in yellow is the experimental structure. The binding site of the experimental receptor is shown with the accessible surface colored blue (60% transparency). The Protein Data Bank accession number for the experimental structure is shown in the lower left corner. (a) X-ray structure of paromomycin bound to A-site rRNA. (b) X-ray structure of biotin bound to its aptamer. (c) NMR structure of malachite green A bound to its aptamer. (d) NMR structure of paromomycin bound to A-site rRNA.

a ligand is determined within a rmsd of 1.8 Å in NMR structures (Supporting Information, Table S2). Strictly speaking, this figure reflects the degree of indetermination of the structures by experimental NMR data (i.e., experimental errors), because each member of a NMR ensemble represents an independent structure determination attempt and not an individual solution conformer. However, the real flexibility of the complex in solution contributes indirectly to this figure, because increased flexibility leads to a paucity or even to internal conflicts in observed NMR parameters,<sup>65</sup> both of which would lead to an increase in the apparent imprecision of the structure. Importantly, rmsd in the ligand positions can be >3 Å in individual NMR structures (PDB codes 1UTS, 1QD3, 1FMN, 1AJU, 1NEM, 1FYF, and 1AM0) and as high as 8.5 Å (PDB 1EI2, Table S2). Most of these highly variable structures are complexes entailing aminoglycoside antibiotics, where the internal flexibility of the ligand contributes significantly to the overall rmsd (Table S2). In the case of crystal structures, the experimental variability in the ligand position was estimated by analyzing structures where the binding pocket–ligand pairs were independently determined more than once. They included symmetric RNA constructs with multiple binding sites, structures where crystallographic unit cells contained more than one independent molecule, and complexes solved in multiple crystal forms. Although generally lower than in NMR structures, the variability of the ligand position within the binding site

was also quite significant for crystal structures (Table S2), with an average rmsd of 0.7 Å and, in some cases, with rmsd as high as 2.8 Å (for the magnesium and barium crystal forms of a streptomycin complex with RNA aptamer;<sup>66</sup> PDB codes 1NTA and 1NTB). For simplicity and to facilitate comparison with published work, we will mostly use a cutoff of 2.5 Å as the criterion of a successful docking in the discussion.

As shown in Figure 1, 74% (42 out of 57) of the ligands in the 57 complexes are docked within 2.5 Å when using the total energy of the complex as the scoring function. The success rate drops to 65% (37 out of 57 structures) for the cutoff of 2.0 Å. Generally, at the cutoff of 2.5 Å, MORDOR reproduces the native X-ray structures better than NMR structures: 30 out of 33 and 12 out of 24, respectively (Table S1, Supporting Information), which correlates strongly with the greater variability of ligand positions in the experimental structures determined by NMR.

Examples of the successful performance of MORDOR on different sets of ligands are provided in Figure 2. Figure 2a–d show the superposition of the experimental structures and the best docking poses for four complexes. The examples from X-ray crystallography are paromomycin bound to A-site rRNA (Figure 2a, rmsd of 1.05 Å) and biotin bound to its aptamer (Figure 2b, 0.96 Å), and examples from NMR are malachite green A bound to its aptamer (Figure 2c, 1.90 Å) and paromomycin bound to A-site rRNA (Figure 2d, 1.64 Å). rmsd values between the experimental and best docking



**Figure 3.** Examination of docking details for one example: paromomycin bound to A-site rRNA (X-ray crystal structure with PDB code 1j7t). For the 1j7t docking run, there were 1531 docking poses generated. (a) The total energy of docking as a function of the rmsd value between experimental and predicted structures for the RNA–ligand complex. (b) Comparison of the binding interaction energy as a function of ligand rmsd for each of the docking poses; the red dot represents the docking pose with lowest total complex energy.

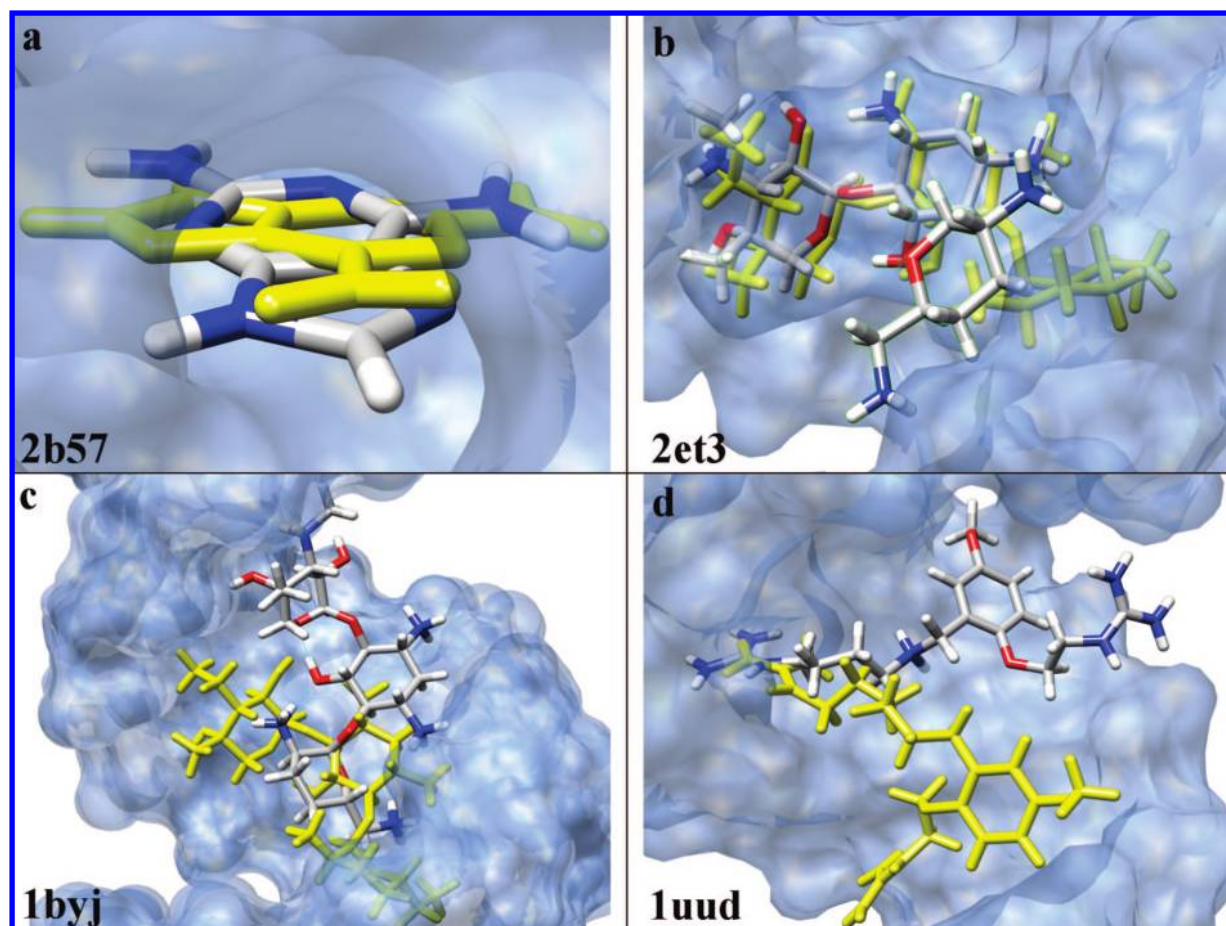
poses are shown in the Supporting Information (Table S1) for all of the complexes in the test set.

**Total Energy versus Interaction Energy as Scoring Criterion.** Very often, the internal energy of the receptor is neglected in the scoring function since most of the current programs keep the receptor fixed during docking. With a rigid receptor, it is possible to use the receptor–ligand interaction energy exclusively (usually, van der Waals energy and electrostatic energy including the energy of hydrogen bonds), disregarding any internal receptor energy change due to interactions with the ligand. To investigate the effect of the internal energy of the receptor, the same docking poses previously calculated were rescored using the interaction energy between the receptor and the ligand instead of the total energy of the complexes. MORDOR had a much lower success rate at reproducing the experimental conformations when the internal energy was omitted from scoring (Figure 1). Only 46% of the docked structures lie within the cutoff value of 2.5 Å from the experimental structures, compared to 74% when the internal energy of the receptor was added to the interaction energy; that success rate goes down to 40% for a cutoff of 2.0 Å. This shows the importance of taking into account the internal energy of the receptor during flexible docking, because of the ligand-induced change in the receptor's conformation. Overall, out of the 57 RNA–ligand complexes in our data set, the average rmsd between the

experimental and the best predicted pose is 2.19 Å using the total energy scoring function, compared to 3.34 Å using only the interaction energy. In all but a few cases, the rmsd value is equivalent or better when we use the total energy as the scoring criterion. However, using solely the interaction energy as the scoring criterion would have provided the correct docking poses in three cases where the total energy criterion failed: diaminopurine bound to a mutant G riboswitch (2b57), gentamicin C1A bound to A-site rRNA (1byj), and acetylpromazine bound to TAR RNA (1lvj). There is no obvious reason why these three cases would be better predicted by interaction energy alone, but they do not change the conclusion that the total energy was a much better predictor.

**Detailed Docking Example with One Structure.** As an example, we will show some details of the A site on 16S rRNA from *Escherichia coli* complexed with the aminoglycoside paromomycin (PDB code 1j7t). Figure 2a compares the structure of the experimental complex with that predicted by MORDOR, and Figure 3a shows the total energy of the complex as a function of the rmsd between the predicted and experimental ligand positions; the data result from the PEDC-driven docking procedure. As expected, the best total energies of the complex are observed when the docked ligand is about 1.1 Å from the experimental structure. The best energy is  $-5261.13$  kcal/mol, which corresponds to a rmsd





**Figure 4.** Examples of incorrect docking poses generated by docking with MORDOR. The best-scoring pose is depicted in all-atom color, and the ligand in yellow is the experimental structure. The binding site of the experimental receptor is shown with the accessible surface colored blue (60% transparency). The Protein Data Bank accession number for the experimental structure is shown in the lower left corner. (a) X-ray structure of diaminopurine bound to a mutant riboswitch. (b) X-ray structure of gentamicin C1A bound to A-site rRNA. (c) NMR structure of gentamicin C1A bound to A-site rRNA. (d) NMR structure of a substituted methoxybenzene with two charged sidechains (termed RBT203) bound to HIV-1 TAR RNA.

of 1.05 Å. However, among all of the docking poses, the closest ligand has a rmsd of 0.85 Å with an energy 40 kcal/mol higher (−5227.18 kcal/mol) than the best-ranked pose. This illustrates well the ruggedness of the potential energy surface when small changes in conformation can dramatically change the energy values. One can also see in Figure 3a how widely the algorithm allows the ligand to search the surface of the receptor; in the case of 1j7t, the ligand explored the receptor as far as 18 Å away from the experimental structure.

During docking, the rmsd fluctuation of the A-site rRNA receptor binding pocket was found to fluctuate between 0.37 and 1.07 Å from the experimental structure, with the average fluctuation of the 1531 docking poses being 0.62 Å. Initially before starting any docking calculation, the minimized receptor structure was 0.37 Å from the crystal structure.

In order to illustrate why MORDOR could not find the correct pose using the interaction energy as a scoring function, we report the binding interaction energy of the docking solution as a function of the ligand rmsd from the native ligand in Figure 3b. In general, if we compare Figure 3a and b, we can see that the correlation between low interaction energy and correct docking pose is not as sharp, which implies that the lowest interaction energy does not necessarily correspond to the best ligand pose. Moreover, if

we look in detail at the Figure 3b plot, the highest score would have picked a ligand 4.2 Å away from the experimental structure, well above the cutoff of 2.5 Å. The orange dot in Figure 3c represents the best docking pose using the best total complex energy as the scoring function; this pose is about 86 kcal/mol higher than the best score when considering only the interaction energy. In other words, use of the interaction energy as the criterion for selecting the best pose would not only yield a pose at substantial variance with the experimental structure but would have distorted the receptor in the process at a cost to the internal energy of the receptor.

**Docking Failures.** We have found it difficult to identify why all docking failures occur, because there are different kinds of docking failures with different causes. Figure 4 presents some examples of docking failures. Figure 4a and b show two examples of docking failures where the ligands, diaminopurine bound to a mutant riboswitch (2b57) and gentamicin C1A bound to A-site rRNA (2et3), are somewhat above the 2.5 Å cutoff, having rmsd values of 3.39 and 3.61 Å, respectively; both structures were determined by X-ray crystallography. In the case of 2b57, the small diaminopurine ligand is essentially flipped 180° such that stacking interactions with the purine rings and H-bonding capabilities inside the binding pocket are not so different compared to the



**Table 1.** Analysis of All Docking Failures Using MORDOR

PDB <sup>a</sup>	ligand	rmsd, <sup>b</sup> Å	rank <sup>c</sup>	rmsd, <sup>d</sup> Å	number of poses <sup>e</sup>	$\Delta E$ , <sup>f</sup> kcal/mol
<b>1B57</b>	diaminopurine	3.39	2	0.17	581	-4.23
<b>2FCX</b>	neomycin A	5.42	2	0.82	884	-4.74
<b>2ET3</b>	gentamicin C1A	3.61	2	0.89	885	-0.18
1TOB	tobramycin	2.63	13	2.46	1285	-18.00
1BYJ	gentamicin C1A	5.37	2	0.98	618	-1.11
1UTS	P13 RBT550	4.90	22	1.98	1473	-9.91
1LVJ	acetylpromazine	3.61	3	1.37	1158	-2.12
1EI2	neomycin	3.23	3	1.97	1456	-5.48
1UUD	P14 RBT203	8.85	3	1.79	922	-1.18
1FYP	paromomycin	2.54	2	0.97	1082	-2.99
1AJU	arginineamide	4.51	2	1.48	535	-0.52
1RAW	adenosine monophosphate	2.87	62	2.49	785	-7.64
<b>1XPF</b>	spermine	5.26	28	2.02	678	-15.24
1UUI	P12 RBT205	8.35	4	2.11	858	-1.46
1NBK	arginineamide	3.57	2	2.07	685	-1.64

<sup>a</sup> Protein Data Bank (PDB) code; crystal structures are shown in bold. <sup>b</sup> Deviation from the experimental ligand position of the best-scoring pose resulting from MORDOR; the rmsd was calculated as described in the Methods section. <sup>c</sup> Rank of the best-scoring pose among those that fall within the deviation cutoff of 2.5 Å. <sup>d</sup> Deviation from the experimental ligand position of the best-scoring pose among those that fall within the deviation cutoff of 2.5 Å. <sup>e</sup> Total number of poses stored during docking. <sup>f</sup> Energy difference between the best-scoring pose and the pose that scored best among those that fall within the deviation cutoff of 2.5 Å.

experimental structure. It might be noted that the crystal structure (2b57) contains nine Co(III) ions and numerous structural water molecules, which are not taken into account in our MORDOR docking (*vide infra*). For 2et3, the flexible gentamicin ligand placement is mostly correct, with the exception of the aminomethyl ring, which has flipped 180° compared to the experimental structure, preferring to form a strong electrostatic interaction with the RNA backbone.

There are also a few docked ligands displaying large rmsd values, such as the NMR structure of gentamicin C1A bound to A-site rRNA (1byj; Figure 4c) and a substituted methoxybenzene with two charged side chains termed RBT203 bound to HIV-1 TAR RNA (1uud; Figure 4d). Both structures were solved using experimental restraints from NMR. The deviations from the experimental ligand positions were 8.85 and 5.37 Å for 1uud and 1byj, respectively. The RBT203 ligand is one of three found by scientists at Ribotargets as good ligands for TAR RNA, and for which they published convincing model structures based on NMR restraint data.<sup>67,10</sup> Surprisingly, MORDOR docked all three comparatively poorly, on the basis of comparing the best-scoring pose from MORDOR with the reported structure models (see Table S1, Supporting Information). In the case of RBT203 (1uud; Figure 4d), a guanidino moiety is in approximately the correct position, but the rest of the ligand is reoriented deeper in the major groove than in the experimental ligand pose, which is more accessible to the solvent. In the case of gentamicin C1A (1byj; Figure 4c), the ligand simply found a better spot along the major groove. Given the large number of hydrogen-bonding possibilities with aminoglycosides binding to RNA, this is not too surprising.

In order to assess whether missed poses are due to insufficient conformational sampling during the docking, we reviewed all 15 failures with predicted poses deviating more than 2.5 Å from the experimental positions. Details are reported in Table 1. In particular, we list the rank of the best docking pose with deviation below 2.5 Å. In other words, the lowest-energy pose had a rmsd of more than 2.5 Å for these 15 ligands, but perhaps another pose with a

reasonably low score may be closer to the experimental pose. Table 1 also lists the energy difference between the docking pose with <2.5 Å rmsd and the best-scoring pose (lowest total energy). Of these 15 failures, a pose within the cutoff of 2.5 Å had a rank number of 2 in seven cases, a rank number of 3 in two cases, and a rank number of 4 in one case. The worst-ranking poses needed to achieve a rmsd < 2.5 Å were observed for a complex of spermine with RNA (PDB 1xpf), which ranked 28 among 678 docking poses stored for this complex, and for the complex of adenosine monophosphate with an RNA aptamer (PDB 1raw), which ranked 62 among 785 docking poses. Coincidentally or perhaps incidentally, these two are among the weakest binders in our data set of RNA complexes.

We find it very encouraging that, if we take the top four best-scoring poses for each docking run, MORDOR was able to reproduce the experimental docking pose in all but four, that is, >90%, of the cases. One of those four, tobramycin (1tob), even had a fairly decent rmsd of 2.63 Å for its best-scoring pose.

The last column of Table 1 lists the total energy difference between the docking pose and the top-ranking pose. An average of 5.1 kcal/mol is seen; however, the energy gap could be <1 kcal/mol (2et3 and 1aju), although it could be as much as -18 kcal/mol in an extreme case such as 1tob. This illustrates once more the ruggedness of the potential energy surface when small changes in conformation can dramatically change the energy values.

**Scoring Function Improvement.** Docking failures can also result from cumulative errors or approximations. The following comments suggest some possible strategies for reducing such scoring errors.

**Errors in Force Field Parametrization.** We used the best potential energy methods available at the time of this study without trying to optimize scoring functions. The scoring function still needs improvement to enhance the reliability of correctly discriminating docked from misdocked conformations. Many approaches could be used to improve the accuracy of the scoring functions: (a) Since we mixed topology and parameters from two force fields, GAFF

(AMBER) for the ligands and CHARMM for the receptors, charges and radius parameters may not be entirely compatible and should be recalibrated. (b) Radii for the ligands have not been optimized to work with the solvation energy, GBSW, when rescored the docking poses. (c) ANTECHAMBER sometimes produces an “unbalanced” topology and parameters for the ligand which could lead to an unrealistic conformation when interacting with the receptor.

**Errors in Electrostatics.** Given the highly localized charges of RNA molecules, which are counterbalanced by ions in solution, appropriate treatment of electrostatic contributions is crucial for docking some ligands, especially flexible ligands interacting with the large RNA major groove. These ligands can freely fluctuate, adopt different conformations in their binding site, and strongly interact with the highly charged RNA phosphate groups, leading to misdocked conformations. This is a recurrent problem with RNA-related calculations.<sup>30</sup> To improve our predictions, we tried a few simple things, such as screening the charge of the RNA phosphate group by trying different combinations of partial charges for the receptor and using a distance-dependent dielectric model with  $\epsilon(r) = 4r$ , where  $r$  is the distance between two atoms. Neither of these strategies improved the results globally; while some particular complexes could be “rescued”, the same protocol transformed some successful complexes into failures. It may very well be important to account for counterions to balance the highly localized charge in RNA and for water mediation at the interface of the biomolecular complexes. It may also be important to account for more specific structural cofactors such as Mg(II), which can be vital in establishing many RNA structures and consequently may be crucial for ligand binding. Any structural cofactors used in stabilizing crystals or crystal structures were not included with the receptor in docking. We assume that appropriately introducing water molecules and counterions would have improved our predictions in these cases. Not including them may allow ligands to have more conformational degrees of freedom; thus, one would expect more dispersion in the docking scores and rankings. Indeed, Moitessier et al. could greatly improve their accuracy of docking aminoglycosides to flexible RNA by including the first hydration shell water.<sup>31</sup>

**Wrong Way of Choosing the Correct Docking Pose.** In this procedure, we assume that the native complex structures reside in the deepest minima we were able to find by calculating the potential energy surface. We found that even small changes in the receptor conformation could lead to substantial changes in energy. This is because macromolecule receptors in an all-atom representation are characterized by a rough energy landscape with a large number of local minima separated by fairly high energy barriers. During a particular docking exercise, the system may become trapped in some local minimum; as docking is carried out for multiple different initial poses, we could generate many docking poses within the same energy range and consequently with less distinguishable multiple minima. Moreover, the force field could be “unbalanced” (*vide supra*), leading to unrealistic but energetically favorable conformations. This is particularly true when simulations are done in vacuo. The consequence is that the docking prediction should not be based solely on one single conformation energy calculation using mechanistic physics-based scoring functions and minimization techniques.

**Table 2.** Comparison of Docking Success with Three Other Ligand–RNA Docking Studies<sup>a</sup>

pose cutoff	DrugScore <sup>RNA</sup>	Ribodock	Filikov et al.
<2 Å	13/28 ( <b>15/28</b> )	5/10 ( <b>8/10</b> )	0/5 ( <b>3/5</b> )
<2.5 Å	14/28 ( <b>16/28</b> )	5/10 ( <b>8/10</b> )	3/5 ( <b>3/5</b> )

<sup>a</sup> The fraction of complexes is given for which a docked pose with the best scoring rank had a rmsd < 2.0 Å or < 2.5 Å from the experimental ligand–RNA complex. Values in bold characters depict the success rate of MORDOR for compounds in common with those in the published studies. The DrugScore<sup>RNA</sup> data set contains 31 complexes.<sup>32</sup> The Ribodock data set contains 10 complexes.<sup>26</sup> The Filikov data set contains five complexes.<sup>20</sup> The Detering and Varani data set of 16 ligand–RNA complexes is not included in this table, because no rmsd value of the best docking pose was provided by the authors.<sup>30</sup> They measured the docking success rate by counting (in %) among 100 independent docking runs if the best docking pose (one per independent docking run) was below a cutoff of 2.5 Å.

One way to circumvent this problem is to treat a modest number of the top-ranked docking poses as an ensemble, because we clearly show that, despite the discrepancy in the force field and in the methodology used in the docking procedure, we nearly always found the correct docking pose among the few top-ranked poses. Ruvinsky recently addressed this issue: By ranking conformations with a standard energy term together with a rmsd-dependent term that favors conformations with many neighbors in conformational space, they were able to predict experimental ligand poses significantly better than simply using the best scoring pose.<sup>68</sup>

#### Comparison with Previous RNA-Docking Studies.

There are many literature reports comparing the docking performance of available programs. Any comparison must be done cautiously, because (a) we are not equally familiar with all programs and protocols, (b) the programs used were often trained to reproduce results known in advance, (c) the data sets were small or narrowly focused, which could easily favor one program over another in a nongeneral way, or (d) just because the docking protocols did not afford the same time for sampling. Docking has been considered a success according to different rmsd cutoff criteria and to different means of assessing ranking success (e.g., taking the first pose vs multiple poses).

Our data set comprises most of the RNA complexes that have been used in the studies of Pfeffer and Gohlke (31 complexes, 28 are common between both studies),<sup>32</sup> Detering and Varani (16 complexes),<sup>30</sup> Morley and Afshar (10 complexes),<sup>26</sup> Lind et al. (13 complexes),<sup>8</sup> and Filikov et al. (five complexes).<sup>20</sup> There are generally fewer members in the data sets as we go back in time, since more RNA–ligand complex structures have been reported over the past few years. Table 2 compares the docking accuracy from each of those studies as well as the present MORDOR results for the RNA complexes that are in common. We will discuss only results provided by the recent study by Pfeffer and Gohlke utilizing their DrugScore<sup>RNA</sup> scoring function, since that has the most extensive data set published to date (31 RNA complexes).<sup>32</sup> With DrugScore<sup>RNA</sup>, they successfully docked 42% of their data set using a docking criterion of <2 Å from the experimental complex using the first scoring rank. Using the same criterion, MORDOR gives a 65% success rate on all 57 RNA complexes.

If we take into account the 28 RNA complexes common to both studies and a criterion rmsd of 2.5 Å, Pfeffer and Gohlke identified 14 out of 28 native complexes against 16 out of 28 in our case. Their test set is predominantly composed of NMR structures. Interestingly, DrugScore<sup>RNA</sup>, unlike MORDOR, was not able to identify the correct docking poses for five X-ray RNA structures (1ft1, 1o9m, 1f27, 1mwl, and 1fuf). However, DrugScore<sup>RNA</sup> did perform better on docking NMR complexes than MORDOR (13/23 experimental ligand poses found for DrugScore<sup>RNA</sup> versus 11/23 for our scoring function).

## CONCLUSION

In this paper, we presented a new flexible docking procedure, MORDOR, suitable for virtual screening and structure-based drug design using biased minimization and classical molecular mechanics force fields as our scoring function. We applied this procedure to 57 cases, the most extensive docking test set of RNA–ligand complexes to date. We showed that a simple model based on physical interactions that makes use of the CHARMM force field for receptors, when combined with the ANTECHAMBER force field for ligands, can find the experimental binding pose within 2.5 Å with a 74% success rate. This brings RNA–ligand docking a step closer to the success found in protein–ligand docking, which typically ranges from 70% to 90%.

One of the most important aspects of our approach is that we permit an induced fit of a ligand to a flexible receptor target. We demonstrated that (a) rigorous force-field-based scoring functions using the classical molecular mechanics force fields are able to find the correct ligand conformation and discriminate the correct conformations from the mis-docked conformations solely on the basis of energy calculations 74% of the time when the top-ranking ligand pose is used and >90% of the time if we consider the four top-scoring poses and, (b) for accuracy of binding to flexible receptors, one needs to consider the internal energy of the receptor.

Our goal in this paper was to demonstrate the ability of MORDOR to dock known ligands to RNA and distinguish the correct docking pose. The next stage, essential for virtual screening, is to calculate the free energy of binding in order to compare different ligands with respect to their binding affinity toward a given RNA target. Our docking score is based on one single conformation by assuming that the native complexes reside in the deepest local minimum on the potential energy surface that was found by our algorithm. It is not possible to estimate the free binding energy and especially the entropy component on the basis of that single conformation. We are currently working on a method using statistical analysis which could exploit an ensemble of conformations generated during the docking procedure in order to select the correct docking pose and estimate the entropy of binding.

## ACKNOWLEDGMENT

We gratefully thank Dr. Nikolai Ulyanov for numerous insightful conversations regarding this work and for his help with building the RNA–ligand complex data set, Dr. Jay Stearns for comments on the manuscript, and Dr. Conrad

Huang for implementing MORDOR visualization in the VIEWDOCK module of CHIMERA. This work was made possible through Grant AI46967 from the National Institutes of Health. We thank OpenEye for generously providing the software suite that was used to regularize the ligand.

**Supporting Information Available:** The test set of ligand–RNA complexes and the variability of binding poses in NMR and crystal structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (1) Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **2003**, *10*, 787–97.
- (2) Foloppe, N.; Matassova, N.; Aboul-Ela, F. Towards the discovery of drug-like RNA ligands. *Drug Discovery Today* **2006**, *11*, 1019–27.
- (3) Hermann, T. Aminoglycoside antibiotics: Old drugs and new therapeutic approaches. *Cell. Mol. Life Sci.* **2007**, *64*, 1841–52.
- (4) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional insights from the structure of the 30s ribosomal subunit and its interactions with antibiotics. *Nature* **2000**, *407*, 340–8.
- (5) Hansen, J. L.; Ippolito, J. A.; Ban, N.; Nissen, P.; Moore, P. B.; Steitz, T. A. The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Mol. Cell* **2002**, *10*, 117–28.
- (6) Auerbach, T.; Bashan, A.; Harms, J.; Schlutzen, F.; Zarivach, R.; Bartels, H.; Agmon, I.; Kessler, M.; Pioletti, M.; Franceschi, F.; Yonath, A. Antibiotics targeting ribosomes: Crystallographic studies. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 169–86.
- (7) Litovchick, A.; Evdokimov, A. G.; Lapidot, A. Aminoglycoside-arginine conjugates that bind tar RNA: Synthesis, characterization, and antiviral activity. *Biochemistry* **2000**, *39*, 2838–52.
- (8) Lind, K. E.; Du, Z.; Fujinaga, K.; Peterlin, B. M.; James, T. L. Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target tar RNA. *Chem. Biol.* **2002**, *9*, 185–93.
- (9) Hwang, S.; Tamilarasu, N.; Kibler, K.; Cao, H.; Ali, A.; Ping, Y. H.; Jeang, K. T.; Rana, T. M. Discovery of a small molecule tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication. *J. Biol. Chem.* **2003**, *278*, 39092–103.
- (10) Murchie, A. I. H.; Davis, B.; Isel, C.; Afshar, M.; Drysdale, M. J.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T. M.; Mirza, S.; Prescott, C. D.; Vaglio, P.; Aboul-Ela, F.; Karn, J. Structure-based drug design targeting an inactive RNA conformation: Exploiting the flexibility of hiv-1 tar RNA. *J. Mol. Biol.* **2004**, *336*, 625–38.
- (11) Luedtke, N. W.; Tor, Y. Fluorescence-based methods for evaluating the RNA affinity and specificity of hiv-1 rev-rrt inhibitors. *Biopolymers* **2003**, *70*, 103–19.
- (12) Ennifar, E.; Paillart, J. C.; Bodlender, A.; Walter, P.; Weibel, J. M.; Aubertin, A. M.; Pale, P.; Dumas, P.; Marquet, R. Targeting the dimerization initiation site of hiv-1 RNA with aminoglycosides: From crystal to cell. *Nucleic Acids Res.* **2006**, *34*, 2328–39.
- (13) Hoch, I.; Berens, C.; Westhof, E.; Schroeder, R. Antibiotic inhibition of RNA catalysis: Neomycin b binds to the catalytic core of the td group i intron displacing essential metal ions. *J. Mol. Biol.* **1998**, *282*, 557–69.
- (14) Park, I. K.; Kim, J. Y.; Lim, E. H.; Shin, S. Spectinomycin inhibits the self-splicing of the group 1 intron RNA. *Biochem. Biophys. Res. Commun.* **2000**, *269*, 574–9.
- (15) Chen, P. J.; Wu, H. L.; Wang, C. J.; Chia, J. H.; Chen, D. S. Molecular biology of hepatitis d virus: Research and potential for application. *J. Gastroenterol. Hepatol.* **1997**, *12*, S188–92.
- (16) Chia, J. S.; Wu, H. L.; Wang, H. W.; Chen, D. S.; Chen, P. J. Inhibition of hepatitis delta virus genomic ribozyme self-cleavage by aminoglycosides. *J. Biomed. Sci.* **1997**, *4*, 208–16.
- (17) Borda, E. J.; Sigurdsson, S. T. Interactions of the antibiotics neomycin b and chlortetracycline with the hammerhead ribozyme as studied by zn<sup>2+</sup>-dependent RNA cleavage. *Bioorg. Med. Chem.* **2004**, *12*, 1023–8.
- (18) Rangarajan, S.; Friedman, S. H. Design, synthesis, and evaluation of phenanthridine derivatives targeting the telomerase RNA/DNA heteroduplex. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2267–73.
- (19) Chen, Q.; Shafer, R. H.; Kuntz, I. D. Structure-based discovery of ligands targeted to the RNA double helix. *Biochemistry* **1997**, *36*, 11402–7.
- (20) Filikov, A. V.; Mohan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. Identification of ligands for RNA targets



- via structure-based virtual screening: Hiv-1 tar. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 593–610.
- (21) Totrov, M.; Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **1997**, *1*, 215–20.
  - (22) Leclerc, F.; Karplus, M. MCSS-based predictions of RNA binding sites. *J. Biomol. Struct. Dyn.* **1999**, *101*, 131–37.
  - (23) Leclerc, F.; Cedergren, R. Modeling RNA-ligand interactions: The rev-binding element RNA-aminoglycoside complex. *J. Med. Chem.* **1998**, *41*, 175–82.
  - (24) Hermann, T.; Westhof, E. Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J. Med. Chem.* **1999**, *42*, 1250–61.
  - (25) Mayer, M.; James, T. L. NMR-based characterization of phenothiazines as a RNA binding scaffold. *J. Am. Chem. Soc.* **2004**, *126*, 4453–60.
  - (26) Morley, S. D.; Afshar, M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using ribodock. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 189–208.
  - (27) Renner, S.; Ludwig, V.; Boden, O.; Scheffer, U.; Göbel, M.; Schneider, G. New inhibitors of the tat-tar RNA interaction found with a “Fuzzy” Pharmacophore model. *ChemBioChem* **2005**, *6*, 1119–25.
  - (28) Yu, X.; Lin, W.; Pang, R.; Yang, M. Design, synthesis and bioactivities of tar RNA targeting beta-carboline derivatives based on tat-tar interaction. *Eur. J. Med. Chem.* **2005**, *40*, 831–9.
  - (29) Mayer, M.; Lang, P. T.; Gerber, S.; Madrid, P. B.; Gómez Pinto, I.; Guy, R. K.; James, T. L. Synthesis and testing of a focused phenothiazine library for binding to hiv-1 tar RNA. *Chem. Biol.* **2006**, *13*, 993–1000.
  - (30) Detering, C.; Varani, G. Validation of automated docking programs for docking and database screening against RNA drug targets. *J. Med. Chem.* **2004**, *47*, 4188–201.
  - (31) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, *49*, 1023–33.
  - (32) Pfeffer, P.; Gohlke, H. Drugscore<sup>ma</sup>-knowledge-based scoring function to predict RNA-ligand interactions. *J. Chem. Inf. Model.* **2007**, *47*, 1868–76.
  - (33) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: Parallel applications of autodock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 293–304.
  - (34) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–25.
  - (35) Schmitz, U.; Donati, A.; James, T. L.; Ulyanov, N. B.; Yao, L. Small structural ensembles for a 17-nucleotide mimic of the trna t psi c-loop via fitting dipolar relaxation rates with the quadratic programming algorithm. *Biopolymers* **1998**, *46*, 329–42.
  - (36) Mujeeb, A.; Ulyanov, N. B.; Billici, T. M.; Farr-Jones, S.; James, T. L. Conformational ensemble calculations: Analysis of protein and nucleic acid NMR data. In *Biological magnetic resonance*; Krishna, N. R., Berliner, L. J., Eds.; Kluwer Academic: New York, 1999; Vol. 17, pp 201–20.
  - (37) Bosshard, H. R. Molecular recognition by induced fit: How fit is the concept. *News Physiol. Sci.* **2001**, *16*, 171–3.
  - (38) Damm, K. L.; Carlson, H. A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* **2007**, *129*, 8225–35.
  - (39) Kroemer, R. T. Structure-based drug design: Docking and scoring. *Curr. Protein Pept. Sci.* **2007**, *8*, 312–28.
  - (40) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. Virtual screening in drug discovery - a computational perspective. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–51.
  - (41) Zhou, Z.; Felts, A. K.; Friesner, R. A.; Levy, R. M. Comparative performance of several flexible docking programs and scoring functions: Enrichment studies for a diverse set of pharmaceutically relevant targets. *J. Chem. Inf. Model.* **2007**, *47*, 1599–608.
  - (42) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–7.
  - (43) Sherman, W.; Beard, H. S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–4.
  - (44) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of fitted 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435–49.
  - (45) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076–84.
  - (46) Knettel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424–40.
  - (47) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. Flexe: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–95.
  - (48) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins* **2002**, *46*, 34–40.
  - (49) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45–55.
  - (50) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632–3.
  - (51) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* **2003**, *68*, 47–62.
  - (52) Frembgen-Kesner, T.; Elcock, A. H. Computational sampling of a cryptic drug binding site in a protein receptor: Explicit solvent molecular dynamics and inhibitor docking to p38 map kinase. *J. Mol. Biol.* **2006**, *359*, 202–14.
  - (53) Keserü, G. M.; Kolossváry, I. Fully flexible low-mode docking: Application to induced fit in hiv integrase. *J. Am. Chem. Soc.* **2001**, *123*, 12708–9.
  - (54) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–40.
  - (55) Floquet, N.; Marechal, J. D.; Badet-Denisot, M. A.; Robert, C. H.; Dauchez, M.; Perahia, D. Normal mode analysis as a prerequisite for drug design: Application to matrix metalloproteinases inhibitors. *FEBS Lett.* **2006**, *580*, 5130–6.
  - (56) Engels, M.; Jacoby, E.; Krüger, P.; Schlitter, J.; Wollmer, A. The  $t \rightarrow r$  structural transition of insulin; pathways suggested by targeted energy minimization. *Protein Eng.* **1992**, *5*, 669–77.
  - (57) Guilbert, C.; Perahia, D.; Mouawad, L. A method to explore transition paths in macromolecules. Application to hemoglobin and phosphoglycerate kinase. *Comp. Phys. Com.* **1995**, *91*, 263–73.
  - (58) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and current status of the charmm force field for nucleic acids. *Biopolymers* **2000**, *56*, 257–65.
  - (59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–74.
  - (60) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–42.
  - (61) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
  - (62) Brooks, B. R.; Brucoleri, R. E.; Olafsen, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
  - (63) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Ucsf chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–12.
  - (64) Im, W.; Lee, M. S.; Brooks, C. L. 3rd Generalized born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–702.
  - (65) Görler, A.; Ulyanov, N. B.; James, T. L. Determination of the populations and structures of multiple conformers in an ensemble from NMR data: Multiple-copy refinement of nucleic acid structures using floating weights. *J. Biomol. NMR* **2000**, *16*, 147–64.
  - (66) Tereshko, V.; Skripkin, E.; Patel, D. J. Encapsulating streptomycin within a small 40-mer RNA. *Chem. Biol.* **2003**, *10*, 175–87.
  - (67) Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I. H.; Karn, J.; Lentzen, G.; Drysdale, M. J.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T. M.; Aboul-Elia, F. Rational design of inhibitors of hiv-1 tar RNA through the stabilisation of electrostatic “Hot spots”. *J. Mol. Biol.* **2004**, *336*, 343–56.
  - (68) Ruvinsky, A. M. Role of binding entropy in the refinement of protein-ligand docking predictions: Analysis based on the use of 11 scoring functions. *J. Comput. Chem.* **2007**, *28*, 1364–72.