# Target Flexibility in RNA−Ligand Docking Modeled by Elastic Potential Grids

Dennis M. Krüger, Johannes Bergs,[†] Sina Kazemi,[†] and Holger Gohlke*

Department of Mathematics and Natural Sciences, Institute of Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-University, Universitätsstrasse 1, 40225 Düsseldorf, Germany

**S** *Supporting Information*

**ABSTRACT:** The highly flexible nature of RNA provides a formidable challenge for structure-based drug design approaches that target RNA. We introduce an approach for modeling target conformational changes in RNA−ligand docking based on potential grids that are represented as elastic bodies using Navier's equation. This representation provides an accurate and efficient description of RNA−ligand interactions even in the case of a moving RNA structure. When applied to a data set of 17 RNA−ligand complexes, filtered out of the largest validation data set used for RNA−ligand docking so far, the approach is twice as successful as docking into an apo structure and still half as successful as redocking to the holo structure. The approach allows considering RNA movements of up to 6 Å rmsd and is based on a uniform and robust parametrization of the properties of the elastic potential grids, so that the approach is applicable to different RNA−ligand complex classes.



**KEYWORDS:** RNA−ligand interactions, flexible docking, DrugScore, elasticity theory, structure-based drug design
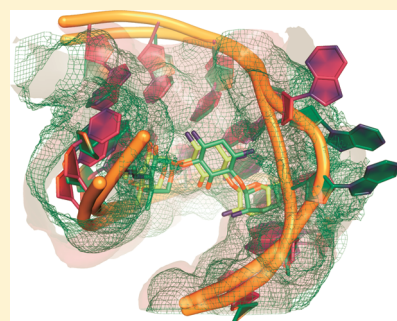
Structure-based drug design (SBDD) approaches that aim at developing novel antibacterial and antiviral drugs are fueled by a steep increase in functional and structural knowledge about RNA molecules. For these SBDDs to be successful, accurate and efficient docking and scoring methods must be available. While the past 30 years have seen great progress in the development of docking tools to predict protein−ligand interactions,[1,2] much less has been achieved in terms of efficiently and accurately modeling RNA−ligand interactions. The current approaches can be divided into three groups. First, computationally intense methods combine docking and molecular dynamics simulations.[3−6] Second, methods originally developed for protein-based drug design are subsequently applied to RNA. For example, Kuntz and co-workers used the DOCK program to identify small molecules with binding specificity to an RNA double helix.[7,8] Third, applications that have been newly developed for RNA−ligand complexes. These include regression-based scoring functions by Morley and Afshar[9] and James and co-workers[10] or a RNA-specific free energy function by Barbault et al.[11] For all of these functions, only a limited number of RNA−drug complexes were used for parametrization. Hence, the general applicability and predictive power of these functions remain elusive.[12] This provided the incentive for us to develop a knowledge-based scoring function to predict RNA−ligand interactions (DrugScore[RNA]).[13] DrugScore[RNA] is based on the formalism of DrugScore,[14] previously developed to score protein−ligand interactions. For DrugScore[RNA], distance-dependent pair potentials have been derived from 670 crystallographically determined nucleic acid−ligand and −protein complexes. When used as an objective function in connection with AutoDock for redocking of 31 RNA−ligand complexes, "good" binding geometries (rmsd < 2 Å) were identified in 42% of all cases on the first scoring rank.

Encouragingly, good docking results were also obtained for a subset of 20 NMR structures not contained in the knowledge base to derive the potentials.

The flexible nature of RNA[15] provides a formidable challenge to standard docking approaches and calls for new developments that consider target conformational changes upon ligand binding.[16] Analogous to protein−ligand docking,[17] three major classes of approaches are conceivable.[12] First, plasticity can be implicitly considered applying a soft-docking strategy with attenuated repulsive forces between target and ligand, but the range of possible movements that can be covered this way is rather limited. Second, only shifts of a few nucleotides are modeled, which assumes a rigid RNA backbone. This seems appropriate only in a limited number of cases, for example, the ribosomal A site.[18] Third, backbone motions are taken into account. This approach is necessary to deal with conformational changes, for example, observed upon binding to HIV-1 TAR RNA.[19,20] So far, only two approaches that fall into this class and are fast enough to allow for flexible RNA virtual screening have been introduced, a method by Guilbert and James,[21] which allows for the mutual conformational adaptation of flexible ligands and flexible targets by energy minimization, and a method by Moitessier et al.,[22] which combines multiple RNA conformations at the level of interaction grids.

Thus, in the present study, we set out to evaluate whether our approach of elastic potential grids,[23] previously developed for protein−ligand docking, provides an accurate and efficient

means for representing intermolecular interactions in fully flexible RNA—ligand docking. The underlying idea is to adapt a 3D grid of potential field values, precalculated from an initial RNA conformation by DrugScore$^{RNA}$, to another conformation by moving grid intersection points in space, but keeping the potential field values constant. For this, RNA movements are translated into grid intersection displacements by coupling RNA atoms to nearby grid intersection points and modeling the 3D grid as a homogeneous linear elastic body applying elasticity theory.[24,25] More specifically, by solving Navier's equation (see eq 1 in ref 23), we compute the displacement $\vec{u}$ of a point in the body (in our case, a grid intersection point) due to forces $\vec{F}$ exerted at some other regions (in our case, due to protein movements). The relation between $\vec{u}$ and $\vec{F}$ is influenced by the modulus of rigidity $G$ and Lamé's constant $\lambda$, which determine the elastic properties of the body.[25] The elastic body-deforming forces $\vec{F} = k \cdot \vec{d}$ are computed according to Hooke's law, with $k$ being a force constant that describes the stiffness of harmonic springs that couple grid intersection points to nearby RNA atoms and $\vec{d}$ being movements of these atoms. An efficient interpolation scheme translates RNA—ligand interactions from the irregular, deformed 3D grid to a new lookup table. Interaction energies between ligand and RNA are then determined from this lookup table. In contrast to using static 3D grids, that way new RNA conformations can be accommodated during a docking run without the need to recalculate potential field values. A more technical description of the method is given in the Supporting Information of ref 23.

For parametrizing the elastic potential grid, the relation between the force constant $k$ and the modulus of rigidity $G$ was determined by a training procedure as described in ref 23. In agreement with ref 23, Lamé's constant $\lambda$, which is linked to the description of expansion or contraction of an elastic body perpendicular to an applied force, is set to zero. For the training, we used 27 RNA—ligand complex structures from nine different types of RNA, namely, aptamer and riboswitch RNA, 16S *E. coli* rRNA, 23S *H. marismortui* rRNA, 16S *T. thermophilus* rRNA, HIV-1 TAR RNA, Diels—Alder ribozyme RNA, dimerization initiation site of HIV-1 RNA, and the ribosomal decoding A site RNA (Table S1 in the Supporting Information). To some extent, these structures have served already in previous evaluations of fully flexible docking approaches[21,22] and represent RNA movements ranging from base flipping (16S *E. coli* rRNA) to backbone deformations (aptamer RNA). For 27 deformations of potential grids according to experimentally determined RNA movements across the nine different target classes, the ratio $k/G$ was found to be $29.45 \pm 0.19$ (Figure S1 in the Supporting Information). For comparison, for protein—ligand complexes, a ratio $k/G = 30.00 \pm 0.24$ has been determined.[23] This result is encouraging because it demonstrates that the parametrization of our approach is transferable between different RNA—ligand complex classes, irrespective of the kinds of conformational changes observed. At the same time, it reveals that for conformational changes due to ligand binding, which are modeled in terms of elastic deformations of binding site regions, almost identical parameters are obtained for proteins and RNA. This means that our approach should be applicable, too, if the target is a RNA—protein complex.

Docking into elastic potential grids was evaluated using a modified version of AutoDock4.1.6[26,27] as a docking engine and DrugScore$^{RNA13}$ as a scoring function. This combination has already proven reliable in a "redocking" evaluation.[13] Initially, the evaluation data set consisted of in total 60 *holo* structures as well

**Table 1. Docking Success Rates**[a]

| RNA type[b] | I (9) | II (1) | III (2) | IV (2) | V (2) | VI (1) |
|---|---|---|---|---|---|---|
| redocking[c] | 9 | 1 | 2 | 2 | 2 | 1 |
| *apo*-docking[d] | 2 | 1 | 0 | 0 | 1 | 0 |
| deformed grids[e] | 7 | 1 | 0 | 0 | 1 | 0 |

[a] The number of complexes is given for which a ligand configuration with $\leq 2.5$ Å rmsd from the native structure was found on the first rank of the largest cluster of docking solutions. [b] The number of complexes for each RNA type is given in parentheses. I, 16S *E. coli* rRNA; II, Aptamer RNA; III, 23S *H. marismortui* rRNA; IV, 16S *T. thermophilus* rRNA; V, HIV-1 TAR RNA; and VI, Thi-box riboswitch RNA. [c] Docking the ligand back into its *holo* RNA structure. [d] Docking the ligand into an *apo* RNA structure. [e] Docking the ligand into a deformed grid whose potential values were calculated based on an *apo* RNA structure and which was subsequently deformed following RNA movements from the *apo* to the ligand's *holo* RNA structure.

as one *apo* structure for each of 11 RNA types (Table S2 in the Supporting Information). All RNA targets are characterized by pronounced movements upon binding, including backbone and base movements. To the best of our knowledge, this data set is the largest data set reported so far for the evaluation of an RNA—ligand docking study, with some of the structures having been used in related studies.[9,13,28] In a first redocking experiment, a success rate of 28% was obtained (Table S2 in the Supporting Information). Notably, successful dockings reported for RNA—ligand complexes in other studies could mostly be reproduced. Thus, the low success rate obtained for the present data set demonstrates a considerable data set dependence of RNA—ligand docking success and calls for the use of comprehensive evaluations data sets in general. For some RNA classes, no good docking solution was found at all (Table S2 in the Supporting Information). Possible reasons for failing to dock to RNA are an inappropriate treatment of electrostatic interactions or disregarding RNA—ligand contacts mediated by water.[12]

Consequently, we selected a subset of 17 *holo* structures for our evaluation data set where docking the ligands back into the bound RNA structures ("redocking") was successful in all cases. These structures comprise six different classes of RNA—ligand complexes (16S *E. coli* rRNA, aptamer RNA, 23S *H. marismortui* rRNA, 16S *T. thermophilus* rRNA, HIV-1 TAR RNA, and thi-box riboswitch RNA). Because of the limited number of cases where redocking was successful, 15 out of the 17 structures have already been used in the parametrization of the elastic potential grid. However, we do not expect a strong training effect when using these complexes for docking into deformed grids based on the facts that (i) almost no variation was found in the optimized parameter $k/G$ across the training set structures, and (ii) almost identical $k/G$ values are found for RNA—ligand and protein—ligand systems. This shows that $k/G$ and, hence, the elastic grid parametrization, is not system-specific. In contrast to redocking, docking to the *apo* structures largely failed in these cases (success rate 24%; Table 1; Table S3 in the Supporting Information). The "redocking" results for the subset thus illustrate that the combination of docking engine and scoring function used here is appropriate, whereas it is the conformational changes of the RNA structures that deteriorate the success rate in the case of "*apo* docking".

Next, we evaluated the performance for docking into deformed potential grids. DrugScore$^{RNA}$ potential values on the grids were initially calculated based on the *apo* RNA structure.
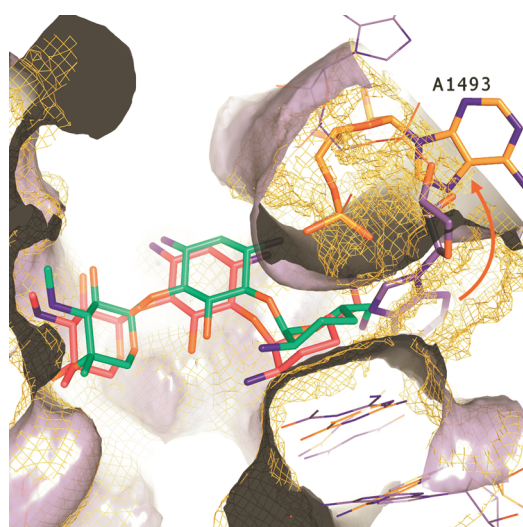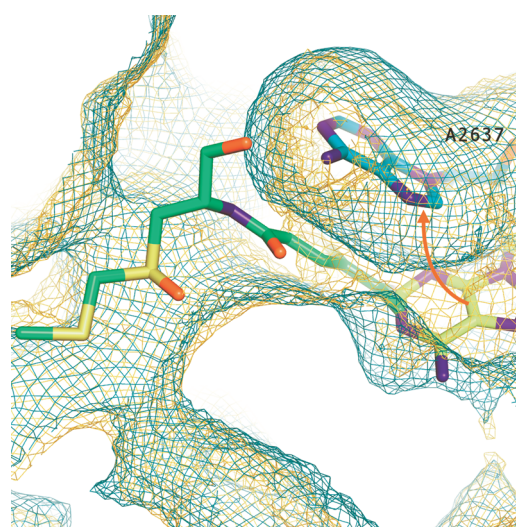
**Figure 1.** Docking of gentamicin C1a into potential fields generated from the 16S *E. coli* rRNA *apo* structure (PDB code: 2I2U) but deformed to the *holo* structure (PDB code: 2ET3). The overall binding site region of 2ET3 has moved by 1.43 Å rmsd, whereas the region containing the A1493 flip has moved by 6.22 Å rmsd with respect to that of 2I2U, respectively. Repulsive potential fields for aliphatic carbon and nucleotide conformations of 2I2U are depicted in blue. Deformed potential fields and nucleotide conformations of 2ET3 are depicted in orange. Aside from base movements, backbone motion can also be observed. Gentamicin C1a carbon atoms are displayed in green for the native structure and magenta for the solution found for docking into the deformed grids (rmsd to the native structure: 1.2 Å). Note how the most significant movement, an adenine flip, "drags along" the potential field.



**Figure 2.** Sparsomycin (green) binding to 23S rRNA (cyan; PDB code: 1VQ8). The apo structure of 23S rRNA is depicted in yellow (PDB code: 1FFK). Repulsive potential fields for aliphatic carbon are depicted in cyan in the case of the *holo* structure; repulsive potential fields for aliphatic carbon generated from the RNA *apo* structure but deformed to the *holo* structure are depicted in orange. Note that even after the deformation due to the rotational flip motion of A2637, repulsive potential field values remain at the location of sparsomycin's thymine moiety.

The grids were then deformed following RNA movements from the *apo* to a *holo* conformation, using $k/G = 30.00$ for each complex. As a result, the precalculated potential values from the *apo* conformation are shifted to new locations in the *holo* conformation. If the shifting is successful, we expect docking to the deformed grids to be as accurate as "redocking" into the *holo* structure. Convincingly, this is indeed true for 53% of the data set (Table 1). For three of the RNA classes, no improvement could be achieved, however. This did not change either if complex-specific $k/G$ values (Figure S1 in the Supporting Information) were used instead of $k/G = 30.00$, again demonstrating the general validity of this ratio (Table S3 in the Supporting Information). In the following, we will analyze in more detail these docking successes and failures.

An example of successful docking into deformed potential grids is depicted in Figure 1. For docking gentamicin C1a into the 16S *E. coli* rRNA, combined backbone and base motions must be modeled. Figure 1 shows that the RNA movements upon ligand binding are very well accommodated by the deformed potential grids, which were generated from the *apo* structure (PDB code: 2I2U) and deformed toward the bound structure (PDB code: 2ET3).The grid deformation results in grid intersection point displacements of maximal 1.31 Å (Figure S2 in the Supporting Information). Especially, the movement of the flipped-out adenine "drags along" the repulsive field for aliphatic carbon and opens a funnel-like space, which eventually allows docking of gentamicin C1a with 1.2 Å rmsd to the native structure.

The above data demonstrate as a proof-of-principle that, if the RNA movement is known, docking into deformed potential grids is about twice as successful as docking into an *apo* structure (9 vs

4 successful cases, Table 1) and still about half as successful as docking to the respective *holo* structure (9 vs 17 successful cases, Table 1). We note that the deformed potential grids were generated from the *apo* structure, and "*apo* docking" is usually considered to be more difficult than docking to another *holo* structure or an average structure of the target.[12] In fact, the median rmsd of binding pocket regions of the respective *apo* and *holo* RNA structures in the evaluation data set is 2.60 Å. For comparison, a success rate of "good" binding geometries of 74% was found by the approach of Guilbert and James when tested on 57 RNA—ligand complexes.[21] When applied to 23 complexes in common with the present study, this rate drops to 67%. This finding again illustrates a data set dependence of such results. Furthermore, it is not clear to what extent the preparation of the docking data contributed to this success. Ligands were "stripped" from the receptor and, after minimization in vacuo, only randomly rotated twice. No randomization of torsion angles was reported. In the first step of the docking protocol, the ligand was then placed on "hot spots" of the receptor, essentially resembling a docking of a rigid ligand into a rigid receptor. As the ligand had not been purposefully distorted with respect to its bound conformation in the preparation step, it may thus not come as a surprise that good starting structures for the subsequent energy minimization step were generated that way.[16]

As already observed in the case of protein—ligand complexes,[23] it is the *type* of motion that influences the docking success in the case of elastic potential grids more strongly than the *magnitude* of motion. While our approach does allow one to consider backbone and base motions simultaneously, limitations of the approach become obvious if base movements are predominantly governed by rotational flip motions or if nucleotide movements lead to an exchange of interaction types, for example, H-bond donor vs acceptor. These types of movements cannot be modeled well by elastic deformations of potential grids: Although

grid intersection points can change their location, the grid topology (i.e., the ordering of grid intersection points) cannot be changed.

An example in that respect is given by sparsomycin binding to 23S *H. marismortui* rRNA upon which A2637 moves by 4.66 Å rmsd following a curvilinear path (Figure 2). Accordingly, while the deformed potential grids around the final A2637 position resemble very well those of the *holo* structure, grid points with repulsive potential field values still occupy the position of sparsomycin's thymine moiety even after the deformation. The fact that the maximal displacement of grid intersection points is 0.81 Å (Figure S3 in the Supporting Information) in this case, as compared to 1.31 Å for 16S rRNA, stresses that the type of motion is indeed the limiting factor here and not its magnitude. Instead of modeling the A2637 movement by elastic grids, excising grid regions close to the A2637 and carrying along these regions with the moving part should provide a more satisfying solution in this case.

Another limitation arises from the fact that our present approach does not allow one to model the creation of highly electronegative pockets for binding of positively charged groups, as often observed to occur in RNA by juxtaposing multiple phosphate groups in space. As we only move grid intersection points in space but keep the potential field values constant, no location in a deformed grid can have a more negative potential than the most negative value in the *apo* structure. Encouragingly, however, no correlation between the positively charged nature of a ligand and failure of docking into deformed grids was found (Table S3 in the Supporting Information): Considering only those cases where *apo* docking had failed, 5 out of 10 cases of docking highly ($\geq 3\ e$) positively charged ligands into deformed grids failed, too. However, also 2 out of 2 cases of neutral ligands failed.

At present, a key limitation of our approach is that the end states of deformations of the elastic potentials grids need to be known. However, we have demonstrated recently that molecular dynamics and constrained geometric simulations showed a tendency to sample bound HIV-1 TAR RNA conformations even when started from the unbound TAR RNA structure.[16] Notably, structural deviations could be reduced by up to 2.3 Å rmsd when compared to deviations between experimental structures. Furthermore, the simulated TAR RNA conformations were used successfully as receptor structures for docking.[16] With respect to the work presented here, receptor conformations generated by simulation techniques should thus provide attractive target conformations to which elastic potential grids generated from an *apo* conformation can be deformed to, without having to take the time to recompute the potential grids for each receptor conformation. In addition, our approach would also allow deforming grids to intermediate conformations in between the *apo* and a target structure as well as to conformations obtained from linear combinations of vectors of structural deviation between *apo* and multiple target structures. This is expected to be superior to parallel docking into an ensemble of fixed receptor structures. Research along these lines is underway in our laboratory.

## ASSOCIATED CONTENT

**ⓢ Supporting Information.** RNA−ligand complexes used for parameter training and evaluation of DrugScore[RNA] as well as detailed docking results. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author
*Fax: (+49)211-8113847. E-mail: gohlke@uni-duesseldorf.de.

### Present Addresses
†Department of Biological Sciences, Institute of Cell Biology and Neurosciences, Goethe-University, Max-von-Laue-Strasse 9, 60438 Frankfurt, Germany.

## REFERENCES

(1) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed. Engl.* **2002**, *41* (15), 2644–2676.

(2) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006**, *65* (1), 15–26.

(3) Srinivasan, J.; Leclerc, F.; Xu, W.; Ellington, A. D.; Cedergren, R. A docking and modelling strategy for peptide-RNA complexes: Applications to BIV Tat-TAR and HIV Rev-RBE. *Fold Des.* **1996**, *1* (6), 463–472.

(4) Leclerc, F.; Cedergren, R. Modeling RNA-ligand interactions: The Rev-binding element RNA-aminoglycoside complex. *J. Med. Chem.* **1998**, *41* (2), 175–182.

(5) Hermann, T.; Westhof, E. Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J. Med. Chem.* **1999**, *42* (7), 1250–1261.

(6) Mu, Y.; Stock, G. Conformational dynamics of RNA-peptide binding: A molecular dynamics simulation study. *Biophys. J.* **2006**, *90* (2), 391–399.

(7) Chen, Q.; Shafer, R. H.; Kuntz, I. D. Structure-based discovery of ligands targeted to the RNA double helix. *Biochemistry* **1997**, *36* (38), 11402–11407.

(8) Kang, X.; Shafer, R. H.; Kuntz, I. D. Calculation of ligand-nucleic acid binding free energies with the generalized-born model in DOCK. *Biopolymers* **2004**, *73* (2), 192–204.

(9) 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* (3), 189–208.

(10) 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* (6), 593–610.

(11) Barbault, F.; Ren, B.; Rebehmed, J.; Teixeira, C.; Luo, Y.; Smila-Castro, O.; Maurel, F.; Fan, B.; Zhang, L.; Zhang, L. Flexible computational docking studies of new aminoglycosides targeting RNA 16S bacterial ribosome site. *Eur. J. Med. Chem.* **2008**, *43* (8), 1648–1656.

(12) Fulle, S.; Gohlke, H. Molecular recognition of RNA: Challenges for modelling interactions and plasticity. *J. Mol. Recognit.* **2010**, *23* (2), 220–231.

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

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

(15) Fulle, S.; Gohlke, H. Analyzing the flexibility of RNA structures by constraint counting. *Biophys. J.* **2008**, *94* (11), 4202–4219.

(16) Fulle, S.; Christ, N. A.; Kestner, E.; Gohlke, H. HIV-1 TAR RNA Spontaneously Undergoes Relevant Apo-to-Holo Conformational Transitions in Molecular Dynamics and Constrained Geometrical Simulations. *J. Chem. Inf. Model.* **2010**, *51*, 6.

(17) Ahmed, A.; Kazemi, S.; Gohlke, H. *Front. Drug Des. Discovery* **2007**, *3*, 455.

(18) Tor, Y. The ribosomal A-site as an inspiration for the design of RNA binders. *Biochimie* **2006**, *88* (8), 1045–1051.

(19) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **1992**, *257* (5066), 76–80.

(20) Nifosi, R.; Reyes, C. M.; Kollman, P. A. Molecular dynamics studies of the HIV-1 TAR and its complex with argininamide. *Nucleic Acids Res.* **2000**, *28* (24), 4944–4955.

(21) Guilbert, C.; James, T. L. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model.* **2008**, *48* (6), 1257–1268.

(22) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, *49* (3), 1023–1033.

(23) Kazemi, S.; Krueger, D. M.; Sirockin, F.; Gohlke, H. Elastic potential grids: accurate and efficient representation of intermolecular interactions for fully flexible docking. *ChemMedChem* **2009**, *4* (8), 1264–1268.

(24) Chou, P. C.; Pagano, N. J., *Elasticity—Tensor, Dyadic, And Engineering Approaches*; Dover-Publications, Inc.: New York, 1992.

(25) Landau, L. D.; Lifshitz, E. M. *Theory of Elasticity*; Pergamon Press: London, 1959.

(26) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8* (3), 195–202.

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

(28) Zhao, X.; Liu, X.; Wang, Y.; Chen, Z.; Kang, L.; Zhang, H.; Luo, X.; Zhu, W.; Chen, K.; Li, H.; Wang, X.; Jiang, H. An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA, and RNA. *J. Chem. Inf. Model.* **2008**, *48* (7), 1438–1447.