

# HIV-1 TAR RNA Spontaneously Undergoes Relevant Apo-to-Holo Conformational Transitions in Molecular Dynamics and Constrained Geometrical Simulations

Simone Fulle,<sup>†,‡,§</sup> Nina Alexandra Christ,<sup>†,‡</sup> Eva Kestner,<sup>‡</sup> and Holger Gohlke<sup>\*,‡,§</sup>

Department of Biological Sciences, Goethe-University, Frankfurt, Germany and Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Düsseldorf, Germany

Received March 13, 2010

We report all-atom molecular dynamics and replica exchange molecular dynamics simulations on the unbound human immunodeficiency virus type-1 (HIV-1) transactivation responsive region (TAR) RNA structure and three TAR RNA structures in bound conformations of, in total,  $\sim 250$  ns length. We compare the extent of observed conformational sampling with that of the conceptually simpler and computationally much cheaper constrained geometrical simulation approach framework rigidity optimized dynamic algorithm (FRODA). Atomic fluctuations obtained by replica-exchange molecular dynamics (REMD) simulations agree quantitatively with those obtained by molecular dynamics (MD) and FRODA simulations for the unbound TAR structure. Regarding the stereochemical quality of the generated conformations, backbone torsion angles and puckering modes of the sugar–phosphate backbone were reproduced equally well by MD and REMD simulations, but further improvement is needed in the case of FRODA simulations. Essential dynamics analysis reveals that all three simulation approaches show a tendency to sample bound conformations when starting from the unbound TAR structure, with MD and REMD simulations being superior with respect to FRODA. These results are consistent with the experimental view that bound TAR RNA conformations are transiently sampled in the free ensemble, following a conformation selection model. The simulation-generated TAR RNA conformations have been successfully used as receptor structures for docking. This finding has important implications for RNA–ligand docking in that docking into an ensemble of simulation-generated RNA structures is shown to be a valuable means to cope with large apo-to-holo conformational transitions of the receptor structure.

## INTRODUCTION

RNA structures show a remarkable ability to undergo large but controlled conformational changes to achieve their diverse functional roles.<sup>1</sup> The conformational changes span a wide range of amplitudes and time scales and are either self-induced or, more common, triggered upon binding to a range of cellular cofactors. To understand the functional mechanisms of RNA conformational transitions, it is highly desirable to have a detailed structural description of the underlying processes. In this regard, one of the best investigated RNA structures is the HIV-1 transactivation response element (TAR).<sup>2–6</sup> TAR RNA is a 59-nucleotide hairpin structure located at the 5'-end of all pre-mRNA in a HIV-1 transcript. Binding of the regulatory Tat protein to a bulge region within the TAR RNA structure is an essential step for HIV-1 viral replication. The bulge region links two A-form helical stem domains including a CUGGGA hexaloop. The secondary structure of this element is shown in Figure 1 and will be referred to as TAR RNA throughout the remainder of this text.

The TAR RNA is highly flexible and adopts rather different conformations upon binding to different ligands (Figure 1). Whereas the unbound structure has a pronounced

kink of  $45^\circ$  between the two stems, e.g., binding of the transactivator of transcription (Tat) protein results in a coaxial alignment of the two stems and, hence, an overall linear structure. This dramatic conformational change is required for the activation of the viral transcription in a stepwise manner.<sup>7</sup> In the search for therapeutic compounds, a promising strategy is to develop small molecules that either inhibit the Tat–TAR interaction or prevent the conformational change to the bound structure.<sup>8–10</sup> Currently, six different ligand-bound conformations are structurally known, comprising a conformation bound to a modification of the cognate Tat protein<sup>2,11</sup> and five conformations bound to small molecules that inhibit the TAR–Tat interaction (Figure 1).<sup>9,12–14</sup>

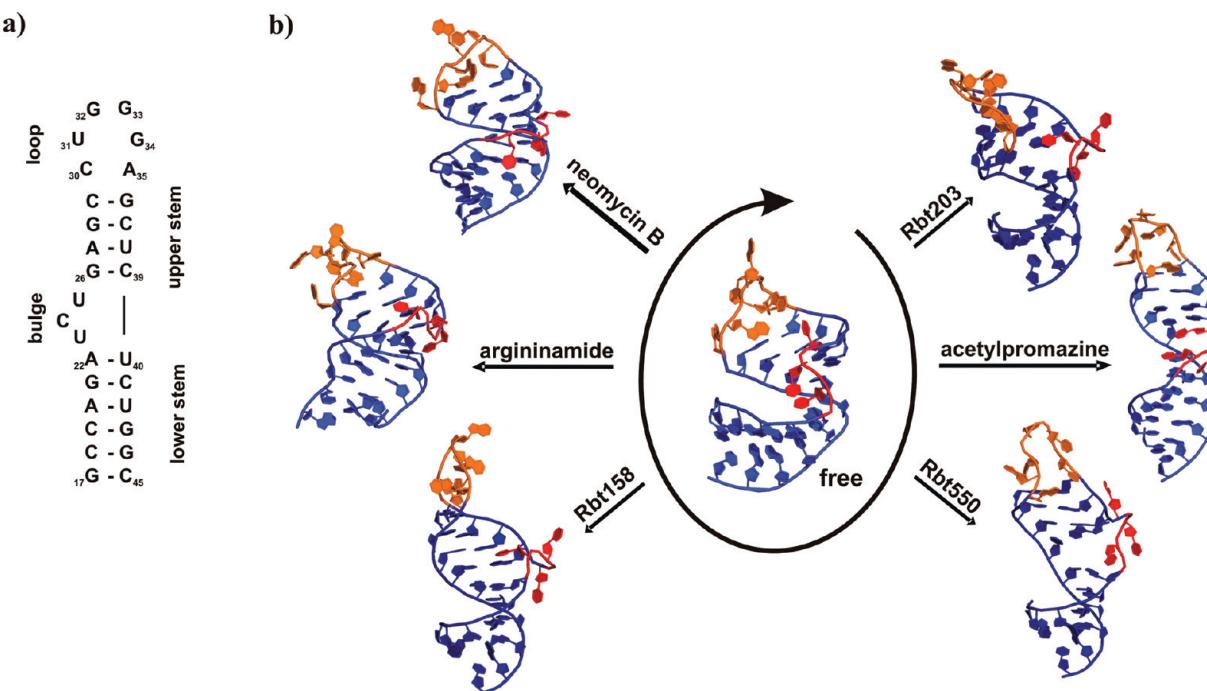
The flexible nature of TAR RNA calls for docking approaches that consider target conformational changes, including backbone motions, when it comes to structure-based drug design. So far, only two such approaches that are fast enough to allow for flexible-RNA virtual screening have been introduced.<sup>15,16</sup> Other strategies well-known in the field of protein–ligand docking, including parallel docking into multiple target conformations,<sup>17–20</sup> modeling target motions in reduced coordinates,<sup>21,22</sup> or deforming interaction grids according to target movements,<sup>23</sup> have not yet been described for RNA.<sup>24</sup> As a possible reason, all of these approaches require either one or multiple RNA structures close to bound conformations or knowledge about

\* Corresponding author. Telephone: (+49) 211-81-13662. Fax: (+49) 211 81-13847. E-mail: gohlke@uni-duesseldorf.de.

<sup>†</sup> These authors contributed equally to this work.

<sup>‡</sup> Goethe-University.

<sup>§</sup> Heinrich-Heine-University.



**Figure 1.** (a) Sequence and secondary structure of the 29-mer TAR RNA used for MD, REMD, and FRODA simulations. (b) Conformation of the HIV-1 TAR RNA in the free form and when bound to different ligands: argininamide, acetylpromazine, neomycin B, Rbt158, Rbt203, and Rbt550. In the structures, the bulge (red), loop (orange), and stem regions (blue) are marked. According to experimental evidence,<sup>4,5,25,26</sup> the unbound TAR RNA undergoes dynamical changes that sample the known ligand bound conformations along specific conformation pathways. Figure adapted from ref 86.

how a given RNA can move. What are the prospects to provide such information by computational means?

From an experimental point of view, TAR RNA can transiently sample the linear (bound) conformation, even in the absence of a ligand.<sup>4,5,25,26</sup> The linear conformation is then stabilized relative to other competing conformations by weak intermolecular interactions between the RNA and a ligand.<sup>27</sup> In such a conformation selection model,<sup>28,29</sup> a ligand captures a minor pre-existing population in the unbound state of the TAR RNA, while the population of the other conformers present in the unbound state are depleted.<sup>30</sup> Accordingly, if the TAR RNA can adopt bound conformers with low energy costs even in the absence of any ligands, then a strategy to find new therapeutically compounds is: (i) to sample the conformational space of the unbound TAR RNA structure and (ii) to carry out subsequently docking calculations on the transiently sampled RNA conformations. In total, this results in docking to an ensemble of different (fixed) receptor conformations, thereby implicitly taking into account conformational changes upon binding.<sup>31</sup>

Molecular dynamics (MD) simulation<sup>32</sup> is the state-of-the-art tool to generate conformational ensembles of biomacromolecules at an atomic level. Enhanced sampling methods, such as replica exchange molecular dynamics (REMD),<sup>33</sup> are increasingly used to overcome barriers between local minima. However, both simulation methods are very time-consuming and are only able to explore time scales on the tens to hundreds of ns scale. In contrast, binding to TAR RNA and related conformational transitions occur on a ns to ms time scale.<sup>5,25,26</sup> Thus, the likelihood to sample bound TAR RNA conformations by the above approaches is low. As an alternative, for enhancing sampling, constrained geometric simulation techniques<sup>34,35</sup> make use of a natural coarse graining of the biomacromolecule in terms of rigid

and flexible regions<sup>36</sup> and restrict the allowed conformational search space to an important subspace. These simplified simulation methods have been successfully applied to the protein world,<sup>34,35,37–40</sup> but experience is lacking for nucleic acid structures.<sup>41</sup>

Consequently, in the following the performance of MD and REMD simulations to sample the conformational space of TAR RNA will be compared to that of the constrained geometric simulation method framework rigidity optimized dynamics algorithm<sup>34</sup> (FRODA) for four HIV-1 TAR RNA structures. FRODA relies upon a decomposition of a macromolecule into rigid and flexible regions. The flexible parts are then moved through allowed regions of conformational space using random Brownian-type (Monte Carlo) dynamics, whereas atoms in rigid clusters are moved collectively. In particular, we will investigate which of the three simulation methods is able to sample known bound conformations starting from either the free (unbound) TAR RNA state or the bound conformations after removing the respective ligand. Furthermore, we will analyze the stereochemical quality of the generated TAR RNA conformations and test whether they are adequate for docking.

## MATERIALS AND METHODS

**Data Set for Simulations.** The coordinates of the free TAR RNA (PDB code 1ANR)<sup>10</sup> and bound to argininamide (PDB code 1ARJ),<sup>11</sup> acetylpromazine (PDB code 1LVJ),<sup>13</sup> and neomycin B (PDB code 1QD3)<sup>12</sup> were used as starting structures for MD, REMD, and FRODA simulations. These structures were modified such that they all consist of the base sequence G<sub>17</sub>CCA<sub>20</sub>GAUCUGAGCC<sub>30</sub>UGGGAGCU<sub>40</sub>CUGGC<sub>45</sub>.

Furthermore, we used the TAR RNA structures bound to Rbt158 (PDB code 1UUI),<sup>14</sup> Rbt203 (PDB code 1UUD),<sup>14</sup>

and Rbt550 (PDB code 1UTS)<sup>9</sup> for our analyses. In the case of the NMR determined structures 1ANR, 1LVJ, 1QD3, and 1UTS, the first structure of the conformational ensemble was used. Only one model was reported in structures 1UUI and 1UUD. In the case of 1ANR, we followed a previous MD study and chose the second structure of the conformational ensemble.<sup>42</sup> For unbound simulations, the ligand molecules were removed from the respective RNA–ligand complexes. In the following, we will refer to the different *unbound* TAR RNA conformations as TAR-free (unbound), TAR-Arg (obtained from 1ARJ), TAR-AcP (obtained from 1LVJ), TAR-NeoB (obtained from 1QD3), TAR-Rbt158 (obtained from 1UUI), TAR-Rbt203 (obtained from 1UUD), and TAR-Rbt550 (obtained from 1UTS), respectively.

**MD and REMD Simulations.** The MD and REMD simulations were performed using the AMBER 8 suite of programs<sup>43</sup> together with the parm94 force field of Cornell et al.<sup>44</sup> For this, the system was neutralized by adding sodium counterions and solvated in a box of TIP3P water molecules,<sup>45</sup> forming a solvent shell of at least 11 Å between each face of the box and the solute. The particle mesh Ewald (PME) method was used with a direct-space nonbonded cutoff of 8 Å. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm, and the time step for all simulations was 2 fs. For the REMD simulations 16 replicas were used, and the temperatures were distributed exponentially between 300 and 350 K at 300, 303.0996, 306.2311, 309.3949, 312.5913, 315.8206, 319.0831, 322.3792, 325.7093, 329.0737, 332.4727, 335.9067, 339.3760, 342.8811, 346.4223, and 350 K. After equilibration, 50 ps of unrestrained MD in the canonical ensemble (NVT) were performed after which snapshots were sampled for analyses.

The production runs of the unbound MD simulations of TAR-free, TAR-Arg, TAR-AcP, and TAR-NeoP are of 50 ns length each and the unbound REMD simulations are of 13 ns length for the TAR-free structure, 14 ns for the TAR-Arg structure, 10 ns for the TAR-AcP structure, and 6.2 ns for the TAR-NeoB structure.

**Force Field Selection.** The validity of current MD simulations is influenced by the sampling of the conformational space and the approximate nature of the used force field.<sup>46</sup> Of the available force fields, the Amber force fields parm94<sup>44</sup> and parm99<sup>47</sup> are the most popular ones for RNA simulations given that they have provided stable simulations of numerous complex RNA structures for simulations of about 10 ns length<sup>46,48,49</sup> and allowed to reproduce experimental free energies of stability with a reasonable degree of accuracy.<sup>50,51</sup> A recent reparameterization of the  $\alpha$  and  $\gamma$  torsions in these force fields led to the parmbsc0 force field,<sup>52</sup> which improved the simulation of DNA duplexes on very long time scales (1  $\mu$ s range). However, further tests on A-RNA duplexes<sup>53</sup> and UAA/GAN internal loops of 23S rRNA H40<sup>54</sup> revealed that the penalization of the  $\alpha/\gamma$  transitions in the parmbsc0 force field may be too excessive for RNA structures and leads to a modest narrowing of the major groove due to the suppression of the  $\gamma$ -trans states.<sup>54</sup> In the present study, when investigating  $\alpha$  and  $\gamma$  torsions sampled during MD and REMD simulations using parm94, we found only one to two transitions toward  $\alpha/\gamma$  t/t substates, e.g., for nucleotides 20 and 28 in the two stem regions. This result agrees with the occasional finding of these substates in experimental structures<sup>54</sup> and, thus, supports the validity

of our simulations. Moreover, the force field variants parm94, parm99, and parmbsc0 all resulted in overall similar descriptions of RNA structures in simulations of up to 100 ns length in previous studies.<sup>54,55</sup> Thus, we expect that the main results of our simulations will not be sensitive to the used Amber force field variant.

**Constrained Geometrical Simulations.** The FRODA approach<sup>34</sup> consists of two steps. In the first step, a TAR RNA structure is decomposed into rigid clusters and flexible hinges in between. Rigid regions are those parts of a molecule that have a well-defined equilibrium structure and are expected to move as a rigid body with six degrees of freedom. Thus, no relative motion is allowed within rigid regions. In turn, flexible regions are hinge regions of the molecule where bond rotational motions can occur without a high cost of energy. For the analysis, a full atomic representation of the structure is modeled as a directed graph of covalent and noncovalent constraints. Noncovalent constraints are included according to a network parameterization that has been recently developed by us for RNA structures.<sup>41,56</sup> *Hydrophobic interactions* are considered between a pair of carbon and/or sulfur atoms if the distance between the atoms is smaller than the sum of the van der Waals radii (1.7 for carbon and 1.8 Å for sulfur) plus 0.15 Å. Hydrophobic interactions are modeled such that two degrees of freedom are removed from the network. This is supposed to model a less geometrically restrained interaction compared to a hydrogen bond. Furthermore, the number of hydrophobic interactions for base stacking is limited to one to prevent excessive hydrophobic contacts between sequentially adjacent bases. Hydrogen bonds and salt bridges are included as distance and angular constraints between hydrogen and acceptor atoms as well as two neighboring atoms, depending on their geometry and interaction energy. For this, potential hydrogen bonds are ranked according to an energy function that takes into account the hybridization state of donor and acceptor atoms as well as their mutual orientation.<sup>57</sup> By tuning the energy threshold  $E_{HB}$ , strong hydrogen bonds can be distinguished from weaker ones. Hydrogen bonds are included if their energy is  $\leq -1.0$  kcal/mol. Further details about the underlying rigidity theory and the network parameterization have been described elsewhere.<sup>41,57,58</sup>

Given a network representation of the RNA structure, the pebble game,<sup>59</sup> a fast combinatorial algorithm, is applied to exactly enumerate the number and spatial distribution of bond rotational degrees of freedom in the network. Based on the accessibility of rotational degrees of freedom, each bond is identified as part of either a rigid cluster or a flexible link in between. The resulting decomposition of the RNA structure into rigid and flexible regions is used in the second step of the approach as input for the FRODA simulation, which finally explores the molecule's mobility.<sup>34,60</sup> Default FRODA values were used for this second step. During the unbound FRODA simulations, 78 200 conformations of TAR-free, 27 200 conformations of TAR-AcP, 999 900 of TAR-Arg, and 999 900 of TAR-NeoB were produced.

**Analysis of the Trajectories.** To analyze conformational properties of the molecules, every 100th conformation of the FRODA simulation if less than 500 000 steps were simulated, every 200th conformation of the FRODA simulation if more than 500 000 steps were simulated, and snapshots saved at 10 ps intervals along the MD and REMD production runs

were used. The ‘ptraj’ module of the AMBER 9 suite of programs<sup>43</sup> was used for analyzing the root-mean-square deviation (rmsd) between structure pairs, the root-mean-square fluctuations about the mean position of atoms, and principal components of structural ensembles.

For rmsd calculations only the heavy atoms of those nucleotides that have at least one heavy atom within 5 Å of the respective bound ligand were considered, whereas loop (C30–A35) and terminal nucleotides (G17–C18 and G44–C45) were neglected. This resulted in the nucleotides A20–C29 and G36–U40 in the case of TAR-Arg, nucleotides G21–A27 and U38–C41 in the case of TAR-AcP, nucleotides C19–A27 and C39–G43 in the case of TAR-NeoB, nucleotides A20–A27 and C39–U40 in the case of TAR-Rbt550, nucleotides A22–U23, U25–C29, and G36–U40 in the case of TAR-Rbt203, and nucleotides A22–U23, G26–G28, and G36–U40 in the case of TAR-Rbt158.

For the principal component analysis (PCA),<sup>61</sup> a mass-weighted covariance matrix of all phosphorus atoms was determined for each simulation run, and the first ten eigenvectors and eigenvalues were calculated subsequently. Overall translational and rotational motions for the PCA were removed with respect to all phosphorus atoms of the TAR-free structure. Eigenvectors describe the directions of essential motions in conformational space, and the associated eigenvalues determine how much of the variance in the ensemble is explained by each eigenvector. Results of the PCA are compared with experimentally observed conformational changes in terms of directions of motions. For this, the first six eigenvectors were obtained by PCA analysis for an ensemble consisting of the experimentally determined TAR-free, TAR-AcP, TAR-Arg, TAR-Neo, TAR-Rbt158, TAR-Rbt203, and TAR-Rbt550 structures.

The maximal overlap value  $I_{\max}$  between the conformational changes observed in the experimentally determined structures  $\vec{u}_i$  and within a trajectory  $\vec{v}_j$  is calculated according to

$$I_{\max} = \max_{ij} \frac{|\vec{u}_i \cdot \vec{v}_j|}{(\vec{u}_i \cdot \vec{u}_i)^{1/2} (\vec{v}_j \cdot \vec{v}_j)^{1/2}} \quad (1)$$

An overlap of one indicates that the directions of both kinds of collective phosphorus displacements are identical.

The analysis of the backbone torsion angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$ , and  $\zeta$ , the glycosidic torsion angle  $\chi$ , and the sugar puckering were carried out using the CURVES 5.2 algorithm.<sup>62</sup> Ranges of reference average backbone torsion angles of canonical A-form RNA structures are listed in Table 1.<sup>63</sup>

The conformational free energy difference  $\Delta G_{\text{free} \rightarrow \text{bound}}$  between the TAR-free structure and a bound conformation is determined from a standard thermodynamic relation<sup>64–66</sup> according to

$$\Delta G_{\text{free} \rightarrow \text{bound}} = -RT \ln \left( \frac{N_{\text{bound}}}{N_{\text{free}}} \right) \quad (2)$$

where  $R$  is the universal gas constant,  $T$  is the temperature and is set to 298 K, and  $N_{\text{bound}}$  ( $N_{\text{free}}$ ) is the number of sampled configurations close to the respective bound (TAR-free) conformation, as determined by rmsd calculations described above. Here, a sampled configuration is considered

**Table 1.** Average Backbone Torsion Angles of Canonical A-Form RNA<sup>63</sup>

torsion angle	atoms involved	range <sup>a</sup>
$\alpha$	O3' <sub>(n-1)</sub> –P–O5'–C5'	-95 to -50
$\beta$	P–O5'–C5'–C4'	165 to -150
$\gamma$	O5'–C5'–C4'–C3'	45 to 60
$\delta$	C5'–C4'–C3'–O3'	75 to 95
$\varepsilon$	C4'–C3'–O3'–P	170 to -150
$\zeta$	C3'–O3'–P–O5' <sub>(n+1)</sub>	-85 to -40
$\chi$	O4'–C1'–N1–C2 <sup>b</sup>	syn, 90 to 0
	O4'–C1'–N9–C4 <sup>c</sup>	anti, -120 to 180

<sup>a</sup> In degrees; ranges are given in clockwise rotation. <sup>b</sup> For pyrimidine bases. <sup>c</sup> For purine bases.

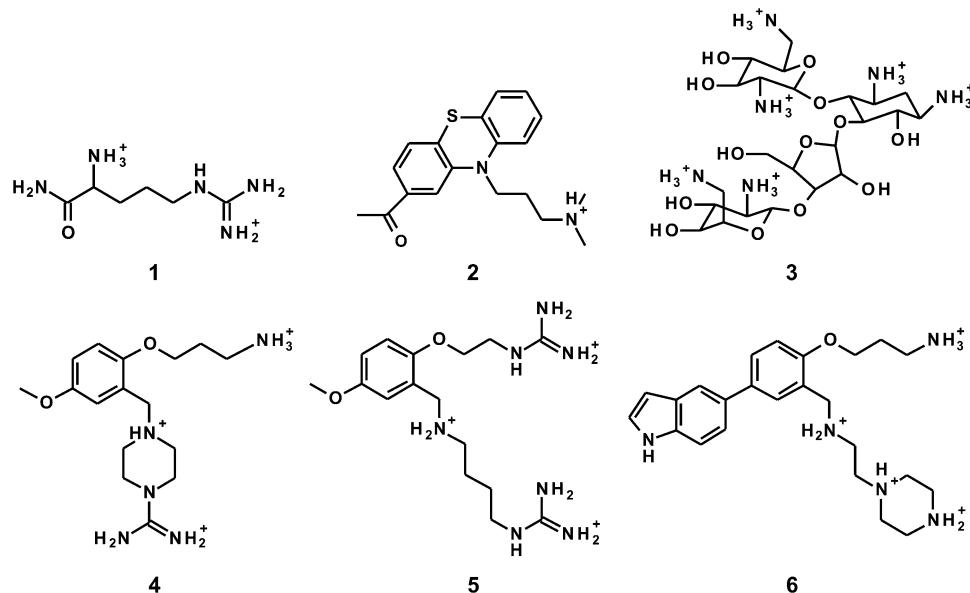
to be close to the TAR-free or bound conformation only if  $\text{rmsd} < 3.0$  Å with respect to the reference.

**Data Set for Docking.** RNA–ligand complexes were obtained from the PDB, as described in the Data Set for Simulations Section. After separating RNA and ligand molecules (Figure 2), the TAR RNA structures were converted to mol2 format using Moloc. DrugScore<sup>RNA</sup><sup>67</sup> grids were calculated, using a grid spacing of 0.375 Å, centered on the binding site of the respective ligand. The grids cover a volume that extends at least 7 Å beyond any ligand atom in the native bound conformation. Ligands were converted to mol2 format using PRODRG,<sup>68</sup> and atom types were assigned using Antechamber<sup>69</sup> together with manual corrections. Finally, rotatable bonds of the ligands were defined using the Autotors utility from the AutoDock suite of programs,<sup>70</sup> disallowing only torsion rotations in peptide bonds, ring systems, guanidinium groups, and terminal bonds between heavy atoms. In the case of the TAR-Arg structure, an arginine was present in the PDB file instead of an argininamide.<sup>11</sup> For the docking experiments, we modeled the argininamide by substituting the carboxylate group of arginine with an amide group.

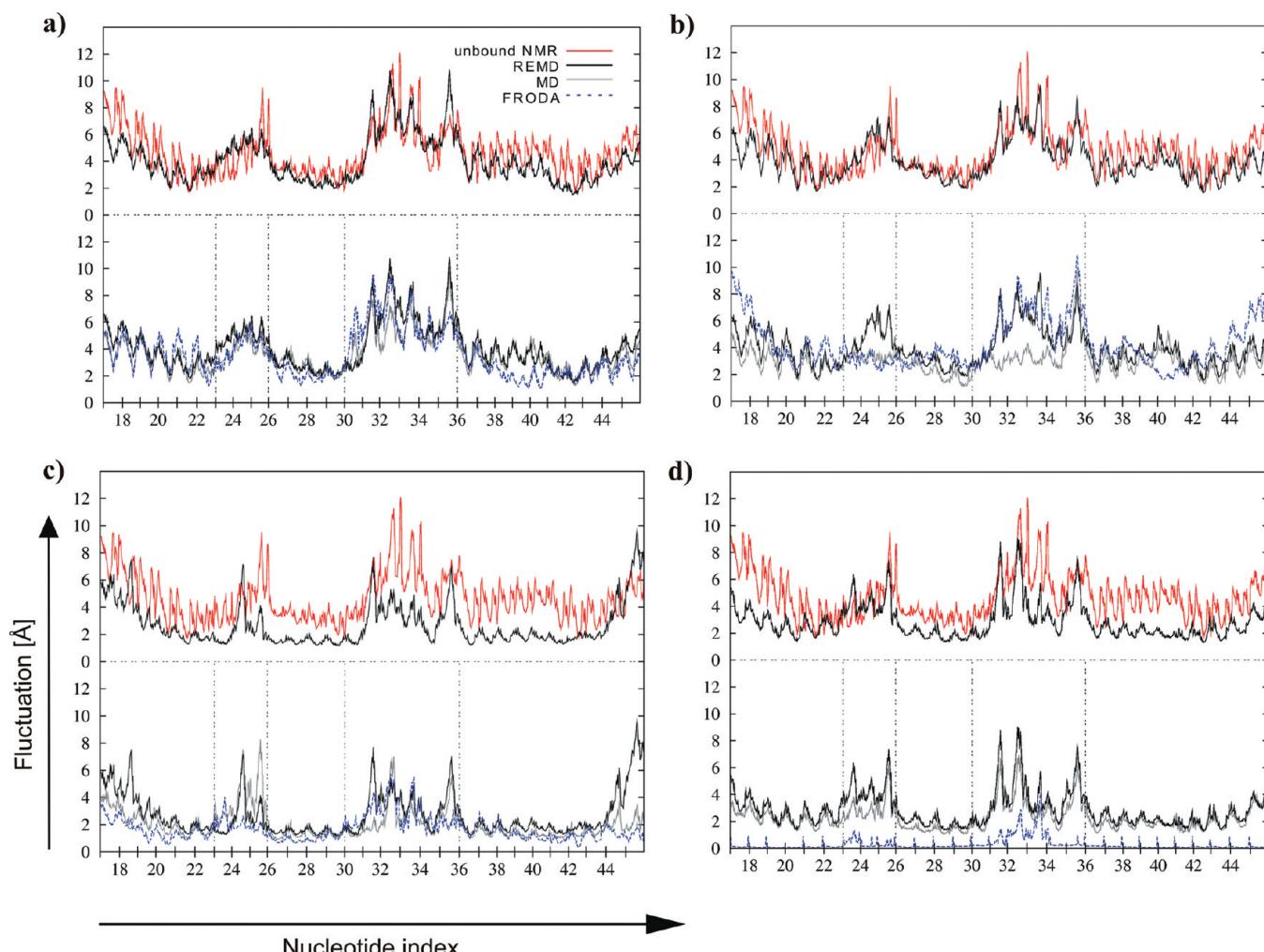
**Docking Parameters.** For each docking, 100 independent docking runs were performed with AutoDock 3.05 using the knowledge-based scoring function DrugScore<sup>RNA</sup><sup>67</sup> and holding the TAR RNA structures rigid. The Lamarckian genetic algorithm was applied using standard parameters provided by AutoDock, with an initial population size of 100, a termination criterion of  $3 \times 10^6$  energy evaluations, mutation and crossover rates of 0.8 and 0.02, respectively, a local search frequency of 0.06, and maximal 300 iterations. The final rank was determined only by intermolecular energies. At the end, the 100 generated ligand conformations were clustered using a rsmd tolerance of 1 Å. Solutions with  $\text{rmsd} < 2.5$  Å with respect to the experimentally determined structure are considered as native-like/successful.

## RESULTS AND DISCUSSION

**Root-Mean-Square Atomic Fluctuations from MD, REMD, and FRODA Simulations.** Initially, the simulation results are compared to characteristics of atomic motions as obtained from conformational changes within the experimentally determined free NMR ensemble (Figure 3). Root-mean-square atomic fluctuations of all atoms as obtained from MD, REMD, and FRODA trajectories on the TAR-free, TAR-AcP, TAR-Arg, and TAR-NeoB structures are given in Figure 3.



**Figure 2.** Chemical structures of the ligands used in this study: argininamide (1), acetylpromazine (2), neomycin B (3), Rbt158 (4), Rbt203 (5), and Rbt550 (6).



**Figure 3.** Root-mean-square fluctuations of unbound HIV-1 TAR RNA as obtained from the NMR ensemble (red, continuous lines), MD (gray, continuous lines), REMD (black, continuous lines), and FRODA (blue, dashed lines) simulations. (a) TAR-free, (b) TAR-AcP, (c) TAR-Arg, and (d) TAR-NeoB structures were used as starting structures. Upper panels: The free NMR ensemble is compared with the REMD trajectories of the TAR-free, TAR-AcP, TAR-Arg, and TAR-NeoB structures, respectively. Lower panels: The MD, REMD, and FRODA trajectories of the TAR-free, TAR-AcP, TAR-Arg, and TAR-NeoB structures are compared. The bulge (nucleotides U23–U25) and the loop (nucleotides C30–A35) regions are indicated.

The overall trends of experimental fluctuations are reproduced very well by the REMD trajectories in the case of the TAR-free and TAR-AcP structures (Figure 3a and b), although details do not compare perfectly. Not surprisingly, the largest deviations between experimentally determined and computed fluctuations occur in the most mobile regions of TAR RNA, the bulge and the loop. As such, the pronounced experimental fluctuations of nucleotide U25, which is partially or transiently looped out of the helix in the free state,<sup>10</sup> are underestimated in both cases, whereas the fluctuations of nucleotide A35 are slightly overestimated in the REMD simulation of the TAR-free structure. Still, sampling is impressive given that fluctuations up to 10 Å are correctly reproduced by the REMD simulation of the TAR-free structure. For TAR-Arg and TAR-NeoB, locations of maxima of computed atomic fluctuations, as obtained from the REMD trajectory, again compare favorably with those determined from the free TAR RNA NMR ensemble (Figure 3c and d). However, in these two cases, the computed atomic fluctuations are generally too low compared to experiment.

When comparing atomic fluctuations obtained from the REMD trajectories with those obtained from MD and FRODA trajectories (lower panel in Figure 3), quantitatively similar patterns are observed in the case of the TAR-free structure. The experimental fluctuation pattern of nucleotide A35 is even better reproduced by MD and FRODA simulations than by REMD. In our opinion, it is striking that the FRODA approach, despite being conceptually much simpler, performs equally well compared to MD and REMD simulations in sampling large conformational movements in the bulge and loop regions of TAR RNA. A major reason is that the rigid-cluster decomposition of the TAR-free structure (Supporting Information, Figure S1) results in two larger rigid clusters located at the lower (comprising nucleotides G17–G21 and C41–C45) and upper (comprising nucleotides G26–G28 and C37–C39) stems, which are linked by flexible bulge and loop regions. This rigid-cluster decomposition reflects experimental findings that in the free conformation of TAR RNA the two stable helical stems collectively undergo a large-amplitude, hinge-like motion around the flexible bulge region.<sup>4,6</sup> Consequently, when used as input to the constrained geometrical simulation FRODA, TAR RNA motions are successfully predicted based on the identified bonds that limit motion within molecules and their coupling.

A different picture emerges in the case of TAR-AcP. First, while the REMD fluctuation pattern of the loop region is matched very well by the FRODA simulation, the fluctuation values obtained from the MD simulation are in general too low. Second, FRODA and MD simulations both underestimate fluctuations in the bulge region. This is not surprising in the case of the FRODA simulation, given that for this region a large rigid cluster was identified by the preceding flexibility analysis (Supporting Information, Figure S1).

**Analysis of the Stereochemical Quality of the Generated Structures.** The generated TAR RNA conformations were further analyzed by calculating the backbone torsion angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$ , and  $\zeta$ , the torsion angles of the glycosidic bond  $\chi$ , and the puckering mode of the ribose.<sup>62</sup> Sampled torsion angles of the TAR-free starting structure are plotted in Figure 4 for A20 of the lower stem, U23 of the bulge region, G28 of the upper stem, and G32 of the loop region.

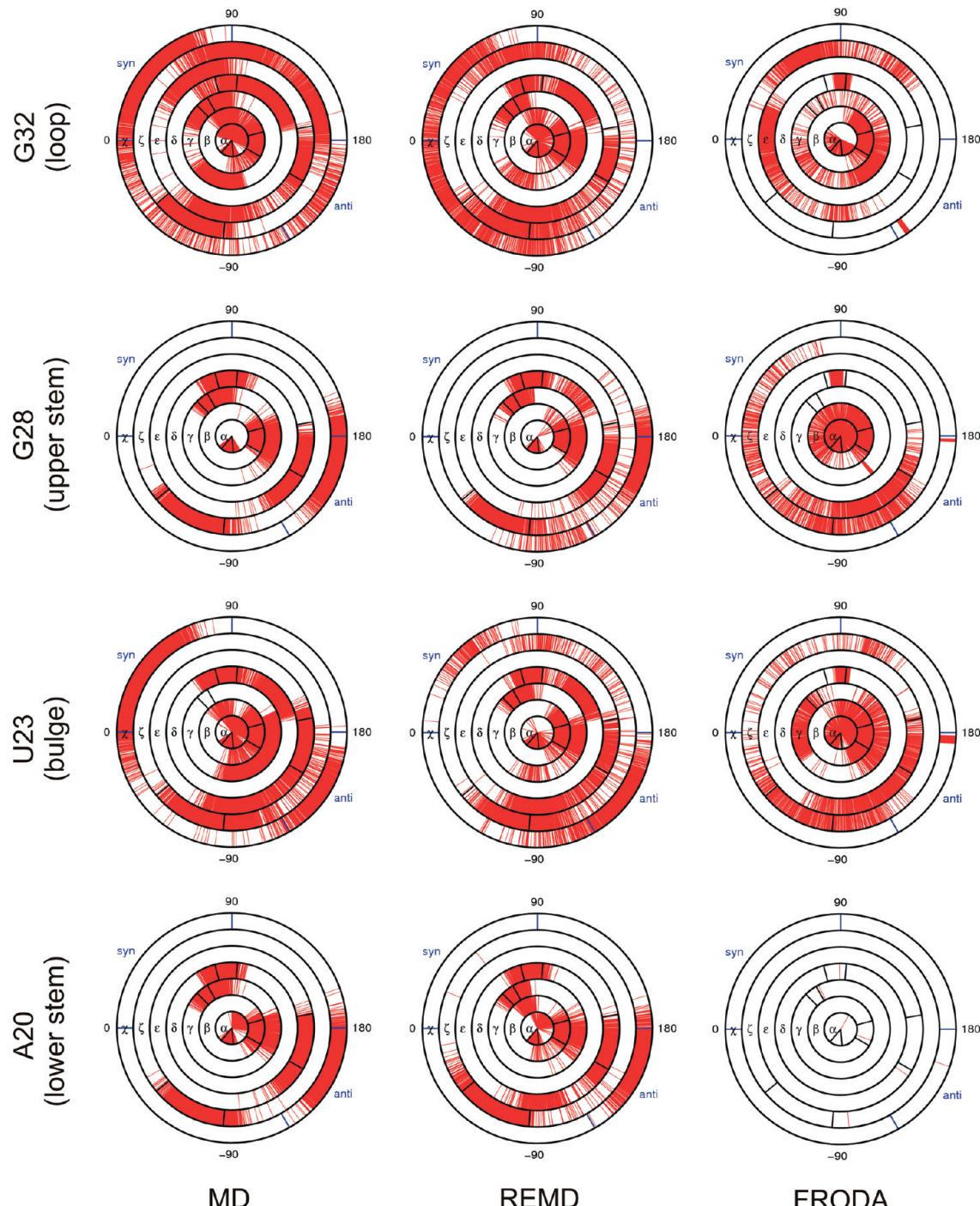
The range of average backbone torsion angles of canonical A-form RNA structures (Table 1) are indicated for comparison.<sup>63</sup> Corresponding plots for TAR-Arg, TAR-AcP, and TAR-NeoB structures are given in the Supporting Information (Figures S2–S4).

The sampled range of torsion angles of the MD and REMD simulations are strikingly similar, with larger deviations found for the U23 bulge and G32 loop nucleotides compared to that of the two stem nucleotides, A20 and G28. The glycosidic bonds of the stem nucleotides A20 and G28 stay preferably in the anti conformation, whereas the U23 bulge and G32 loop nucleotides sample much larger ranges of  $\chi$  values, varying between  $-180^\circ$  and  $90^\circ$ . Fluctuations about the glycosidic bond do not occur independently but lead to concerted torsion angle fluctuations in the sugar–phosphate backbone and the sugar puckering mode.<sup>71</sup> Along these lines, the distribution of the backbone torsion angles of the A20 and G28 stem nucleotides are close to those of standard A-form helices, whereas the backbone torsion angles of the G23 bulge and G32 loop nucleotides rotate more freely. The only exception is the backbone angle  $\gamma$ , which shows a bimodal distribution about  $60^\circ$  and  $180^\circ$  in the stem nucleotides A20 and G28. In general, these observations are also true for other nucleotides of the stem, bulge, and loop regions (data not shown), except terminal stem nucleotides. These are less ordered than central stem nucleotides, which is consistent with the pronounced atomic fluctuations found in these regions (Figure 3).

In contrast to the sampling of preferred torsion angles by MD and REMD simulations, in the case of FRODA simulations, torsion angles either are mostly freely sampled, as in the case of  $\alpha$  and  $\gamma$  of the U23 bulge nucleotide, or remain fixed at the starting value, as in the case of all torsion angles of the A20 stem nucleotide. Only in some cases, e.g., for torsion angles  $\varepsilon$  and  $\zeta$  of the U23 bulge and G32 loop nucleotides, respectively, a restricted torsion angle range is observed, which is probably due to steric hindrances of these nucleotides.

On the one hand, fully fixed torsion angles are observed in FRODA simulations if a nucleotide is part of a rigid cluster, whose atoms are moved collectively per simulation step without any internal motions allowed. This is true per se for all ribose moieties, as five-membered rings come out as rigid bodies from the flexibility analysis. Modeling a biomacromolecule in such a coarse-grained fashion has been found to not impair the prediction of finite amplitude motions, however, as shown for proteins<sup>72,73</sup> and other RNA molecules,<sup>41</sup> provided that the decomposition in rigid and flexible regions is based on an appropriate network representation.<sup>41</sup> On the other hand, the free sampling of some of the torsion angles in bulge and loop nucleotides is a consequence of the random walk strategy used by FRODA and calls for a restriction of torsion angles to preferred ranges. As such, knowledge about discrete dihedral angle combinations could be integrated into FRODA. Such a modeling of rotameric states in terms of torsion angle constraints has been shown to significantly improve the stereochemistry of protein conformations sampled by the NMSim approach.<sup>73</sup>

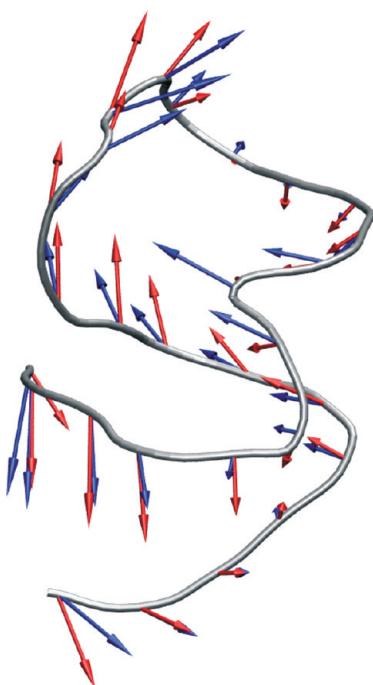
The average puckering mode of nucleotides of the MD (REMD) trajectories of the TAR-free structure amounts to (Supporting Information, Table S1) 70% (63%) in a C3'-endo conformation and only 5% (7%) in a C2'-endo



**Figure 4.** Conformation wheels depicting sampled backbone and glycosidic torsion angles of exemplary nucleotides in the stem, bulge, and loop regions during MD, REMD, and FRODA simulations of the TAR-free structure. Torsion angles are given in the range of  $-180^\circ$  to  $180^\circ$ . The occurrence of a particular torsion angle is denoted by a radial line at the appropriate angular location. Ranges of average backbone torsion angles of canonical A-form RNA structures (black lines) and syn and anti ranges of the glycosidic bond are indicated.<sup>63</sup>

conformation in the stem region, 30% (18%) in a C3'-*endo* conformation and 29% (50%) in a C2'-*endo* conformation in the bulge region, and 20% (26%) in a C3'-*endo* conformation and 39% (34%) in a C2'-*endo* conformation in the loop region. This is consistent with previous findings that the ribose in A-form RNA prefers the C3'-*endo* form but that outside regular helices, RNA riboses adopt C2'-*endo* conformations.<sup>63,74</sup> With respect to the average populated puckering mode, MD and REMD simulations show in general similar results for the stem and loop regions (Supporting Information, Table S1). However, compared to

the MD simulation, more C2'-*endo* conformations are sampled in the bulge region during the REMD simulation, which may be due to better sampling by the latter approach. The majority of the bulge nucleotides being in a C2'-*endo* conformation is consistent with previous MD<sup>75,76</sup> and NMR studies<sup>2</sup> on bulge systems. As already mentioned above, due to the rigid character of riboses in the FRODA approach, the puckering mode does not change from its starting structure value in FRODA simulations. Modeling the ribose ring as flexible instead yielded significantly worse mobility predictions for RNA structures in a previous study.<sup>41</sup>



**Figure 5.** Essential motions as given by the modes most involved in conformational changes from the experimentally determined unbound to the bound TAR RNA structures (red arrows) in comparison with those obtained from the REMD simulation on TAR-free (blue arrows). The amplitudes of the motions were scaled for best graphical representation.

**Essential Dynamics of Unbound and Bound TAR RNA Structures.** Unbound TAR RNA can transiently sample distinct bound conformations according to experimental evidence.<sup>4,5,25</sup> We thus analyzed all trajectories as to whether unbound TAR RNA also samples bound conformations during MD, REMD, and FRODA simulations. For this, essential motions were calculated from simulated TAR RNA ensembles and compared with essential motions derived from the ensemble of experimentally determined unbound and bound TAR RNA molecules. As an example, in Figure 5, essential motions derived from the experimental ensemble and from the REMD-generated ensemble of the TAR-free starting structure are mapped onto the TAR-free structure for those two modes that show the maximal overlap (eq 1). All maximal overlap values  $I_{\max}$  for modes of essential motions derived from the experimental ensemble and from simulation-generated ensembles for all starting structures are given in Table 2.

Using MD simulations,  $I_{\max} \geq 0.68$  is found for all four starting structures, with  $I_{\max} \approx 0.8$  found for TAR-Arg and TAR-NeoB starting structures. Very similar results are observed for REMD simulations, except for the simulation started from the TAR-Arg structure ( $I_{\max} = 0.56$  compared to  $I_{\max} = 0.78$  obtained for the MD simulation). Thus, given that at least twice as much total simulation time was applied for the REMD simulations in comparison to the MD simulations, it is discouraging that the enhanced sampling technique apparently does not provide a better sampling of known conformational transitions than that of standard MD. Comparing these results with those obtained for FRODA simulations, in general lower  $I_{\max} \approx 0.60$  is found for the latter. Only in the case of the TAR-AcP starting structure, the FRODA-generated ensemble yields  $I_{\max} = 0.73$ , slightly higher than found in MD and REMD simulations.

**Table 2.** Overlap of Essential Motions Observed in the Binding Process of HIV-1 TAR RNA Compared with Results Obtained from MD, REMD, and FRODA Simulations

starting structure	PCA modes <sup>a</sup>	overlap <sup>b</sup>
TAR-free	experimental (1)	MD (1) 0.72
	experimental (1)	REMD (1) 0.77
	experimental (1)	FRODA (1) 0.59
TAR-Arg	experimental (1)	MD (2) 0.78
	experimental (1)	REMD (2) 0.56
	experimental (1)	FRODA (1) 0.56
TAR-AcP	experimental (1)	MD (1) 0.68
	experimental (1)	REMD (1) 0.69
	experimental (1)	FRODA (2) 0.73
TAR-NeoB	experimental (1)	MD (2) 0.83
	experimental (1)	REMD (1) 0.80
	experimental (1)	FRODA (6) 0.56

<sup>a</sup> In parentheses, the numbers of the modes are given for which the overlap is maximal (eq 1). <sup>b</sup> Maximal overlap between the modes describing conformational changes in the experimentally determined structures and within a trajectory (eq 1) considering all phosphorus atoms of the TAR RNA structure, respectively.

The above results on the unbound TAR RNA structures reveal how much of the experimentally observed movement between unbound and bound conformations can be reproduced by the simulation approaches. Notably, the overlap between directions of collective movements reaches 77% in the case of the REMD simulation of the TAR-free structure. This indicates a high potential to sample bound conformations of the TAR RNA structure even when starting from a free form. However, it also shows that the sampling by MD and REMD simulations is still not sufficient as to completely reach a bound conformation. As shown in Figure 5, this is due to incomplete sampling of substructural regions, e.g., the bulge region in the case of TAR RNA, rather than a general deviation between essential motions determined from experiment and simulation. Similarly, starting from a complex structure, a tendency to sample known conformational transitions is revealed, with maximal overlap values between experimental and REMD- or MD-simulated directions of collective motions of  $\geq 80\%$  in the case of TAR-NeoB.

**Searching for Generated Bound Conformations.** Overall well reproduced trends of experimental fluctuations by the simulations and good agreement between experimentally observed and simulated directions of collective movements between unbound and bound conformations, at least for REMD and MD simulations, provided the incentive for us to search for bound conformations in the TAR RNA ensembles generated by simulations. More specifically, we were interested in bound conformations of the ligand binding regions of the known TAR RNA complexes in order to test whether these conformations would be suitable for docking (see below). Thus, rmsd for each binding region (see Materials and Methods Section for respective nucleotide numbers) of snapshots from MD, REMD, and FRODA trajectories started from the TAR-free, TAR-Arg, TAR-AcP, and TAR-NeoB structures were calculated with respect to all experimentally known bound TAR RNA conformations. Using an rmsd over the whole TAR RNA structure instead did not seem appropriate to us, because this rmsd would be mainly caused by interhelical domain motions and, thus, would be rather insensitive to structural details in the binding site regions.<sup>10,11,42,77</sup> Rmsd between experimental structures

**Table 3.** Rmsd of the Binding Regions of Experimentally Determined TAR RNA Conformations<sup>a</sup>

starting structure	reference structure						
	TAR-free	TAR-Arg	TAR-AcP	TAR-NeoB	TAR-Rbt550	TAR-Rbt203	TAR-Rbt158
TAR-free	—	5.1	4.7	5.6	4.2	4.4	2.7
TAR-Arg	4.9	—	5.1	3.9	5.4	1.8	2.1
TAR-AcP	4.7	6.1	—	4.6	5.2	4.0	3.4
TAR-NeoB	5.6	3.7	5.2	—	4.5	3.8	3.1
TAR-Rbt550	4.2	6.8	5.3	4.4	—	5.6	3.9
TAR-Rbt203	4.4	2.7	5.1	4.4	5.8	—	1.2
TAR-Rbt158	2.7	3.8	5.2	4.6	5.8	1.4	—

<sup>a</sup> In Å. For fitting and rmsd calculation, all heavy atoms of the binding region of the respective bound structure were considered. Cases with an rmsd < 3.0 Å are marked in italics.

**Table 4.** Minimal Rmsd of the Binding Regions of Structures Generated by MD, REMD, and FRODA Simulations with Respect to Experimentally Determined Conformations<sup>a</sup>

simulation method	starting structure	reference structure						
		TAR-free	TAR-Arg	TAR-AcP	TAR-NeoB	TAR-Rbt550	TAR-Rbt203	TAR-Rbt158
FRODA	TAR-free	—	4.5	4.8	4.4	3.5	3.9	2.3
	TAR-Arg	4.6	—	5.0	3.7	4.9	<b>1.8</b>	<b>2.0</b>
	TAR-AcP	4.5	4.4	—	4.6	5.0	3.9	3.3
	TAR-NeoB	5.6	3.9	5.1	—	4.5	3.8	3.0
MD	TAR-free	—	3.3	3.0	4.2	3.7	3.1	2.7
	TAR-Arg	3.0	—	4.6	3.4	4.1	<b>1.3</b>	<b>1.4</b>
	TAR-AcP	3.9	3.5	—	4.1	4.6	2.9	2.5
	TAR-NeoB	4.2	2.8	3.4	—	3.6	2.2	<b>1.6</b>
REMD	TAR-free	—	3.7	3.0	3.3	3.8	3.3	2.2
	TAR-Arg	4.4	—	4.6	3.7	5.2	<b>1.5</b>	<b>1.4</b>
	TAR-AcP	4.0	3.1	—	3.8	4.2	2.7	2.4
	TAR-NeoB	4.2	3.6	4.1	—	3.7	3.0	2.4

<sup>a</sup> In Å. For fitting and rmsd calculation, all heavy atoms of the binding region of the respective bound structure were considered. Cases with an rmsd < 3.0 Å are marked by italics, and cases with an rmsd ≤ 2.0 Å are given in bold and italics.

and minimal rmsd found for the MD, REMD, and FRODA-generated ensembles in comparison to the unbound and different bound experimental structures are reported in Tables 3 and 4, respectively.

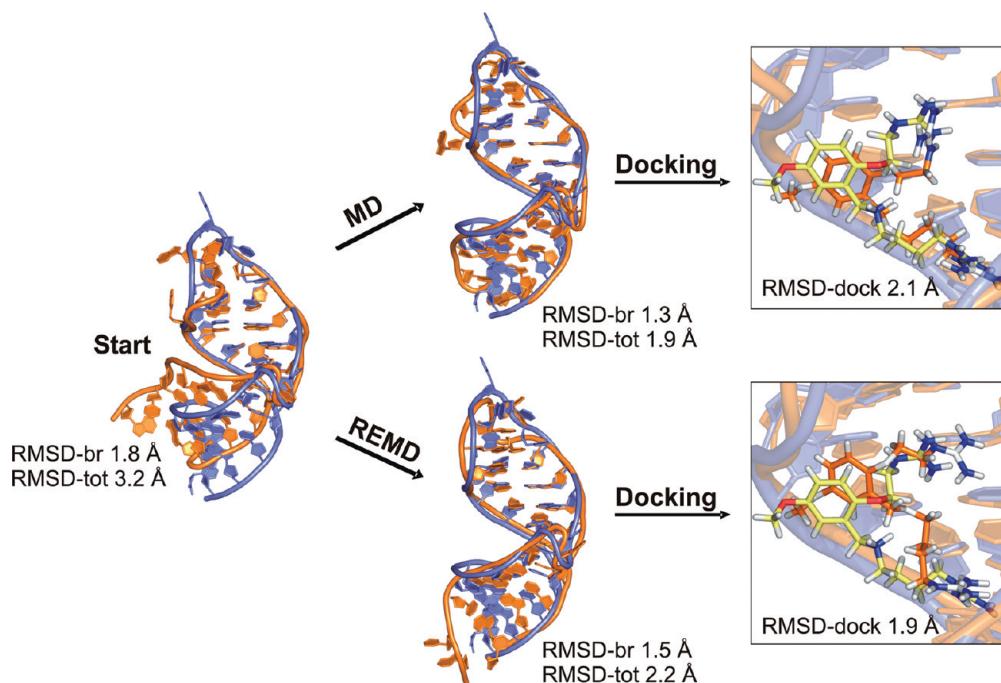
As for rmsd between experimental structures, bound TAR RNA conformations deviate by 2.7 to 5.6 Å (TAR-Rbt158 and TAR-NeoB, respectively) from TAR-free (Table 3). When compared to protein examples for which bound conformations have been successfully sampled starting from unbound structures, e.g., aldose reductase (rmsd between experimental structures: ~1.2 Å),<sup>78,79</sup> HIV-1 protease (3.3 Å for the flap residues),<sup>80</sup> and IL-2 (1.0 Å in the interface region),<sup>40,81</sup> these values already lead one to suspect that finding bound conformations starting from an unbound structure in the case of TAR RNA should be much more difficult. In that sense, it is reassuring that for all three simulations of TAR-free, conformations could be identified whose rmsd to a bound TAR RNA conformation is lower than that between the respective starting structures (Table 4). Also, when compared to the starting structures, structural deviations could be reduced by up to 2.3 Å rmsd (MD simulation of TAR-free vs TAR-NeoB: 5.6 → 3.3 Å). However, when considering minimal rmsd with respect to bound conformations for structures of the TAR-free simulation, values ~2.5 Å are only found with regard to TAR-Rbt158.

When determining the conformational free energy difference according to eq 2 in this case,  $\Delta G_{\text{free} \rightarrow \text{bound}} = 2.0 \text{ kcal mol}^{-1}$ , with the bound conformation being energetically disfavored compared to TAR-free, as expected. A similar

value of 2.2 kcal mol<sup>-1</sup> was found for the TAR-free → TAR-AcP transition, where the bound conformation was approached as close as 3.0 Å rmsd. Notably, these numbers are of the same order of magnitude as, but smaller than, the binding free energies of the respective ligands such that weak intermolecular interactions with the ligands should suffice to stabilize bound TAR RNA.<sup>27</sup> Furthermore, the small conformational free energy differences are consistent with the view that bound TAR conformations are transiently sampled in the free ensemble,<sup>4,5</sup> following the conformation selection model.

With respect to rmsd between bound experimental TAR RNA, all TAR conformations deviate by more than 3 Å from each other except for TAR-Arg vs TAR-Rbt158 and TAR-Rbt203 or TAR-Rbt158 vs TAR-Rbt203. Again, considerable sampling of conformational space can be observed during simulations, e.g., reducing structural deviations of the TAR-AcP structure with respect to the TAR-Arg structure by 3.0 Å rmsd in the case of the REMD simulation and by 2.6 Å in the case of the MD simulation. When looking at minimal rmsd values of simulated TAR RNA started from bound structures with respect to other experimental bound conformations, MD (REMD) show four (three) cases with rmsd < 3.0 Å and three (two) further cases with rmsd < 2.0 Å. In the case of FRODA simulations, only two examples with rmsd < 3.0 Å are found. These trends are also mirrored if average minimal rmsd values are considered (REMD: 3.4 Å; MD: 3.2 Å; and FRODA: 4.0 Å).

Overall, these results demonstrate that the two force field-based simulation methods outperform the constrained geo-



**Figure 6.** Alignment of TAR-Rbt203 (purple) and TAR-Arg (orange) used either as starting structures or as generated during MD and REMD simulations, respectively. The alignments with the lowest rmsd of the binding regions are given. For alignment only the nucleotides  $\leq 5 \text{ \AA}$  apart from the ligand Rbt203 were considered. The rmsd of the TAR RNA structures are given for the binding pocket region (rmsd-br) and for nucleotides C19–G43 (rmsd-tot). Docking solutions (orange) with minimal rmsd to the native ligand conformations (yellow) are shown.

metrical simulation method FRODA in sampling the TAR RNA conformational space. A possible reason is that the initial constraint network is not changed during FRODA simulations, which may result in too strict a limitation of movements. Instead, breaking and reformation of constraints during the simulation may be necessary to achieve conformational transitions as large as observed for TAR RNA. Furthermore, it is encouraging to note that rmsd deviations between TAR RNA binding sites could be reduced by  $\sim 3 \text{ \AA}$  with respect to the starting structures. However, an overall simulation time of  $>200 \text{ ns}$  in the case of the REMD simulation of the TAR-free structure is apparently still not sufficient to overcome structural differences of  $5.6 \text{ \AA}$  rmsd between TAR-free and TAR-NeoB structures.

**Docking into Sampled TAR RNA Structures.** Next, we investigated whether simulated TAR RNA conformations that have been identified as being close to bound conformations of other complex structures can be used as receptor structures for docking. For this, we extracted those generated conformations that have a minimal rmsd value  $<3.0 \text{ \AA}$  compared to another reference bound structures. This resulted in 17 TAR RNA conformations, into which the respective known TAR RNA ligand was docked using AutoDock together with DrugScore<sup>RNA</sup><sup>67</sup> as an objective function. The docking results are summarized in the Supporting Information, Table S2. In no case, a ‘good’ docking solution ( $\text{rmsd} \leq 2.5 \text{ \AA}$  to the native structure)<sup>15</sup> was found on the first scoring rank or in the largest cluster of generated ligand poses. However, for conformations extracted from MD and REMD simulations of the TAR-Arg structure in comparison with TAR-Rbt203 (rmsd of the binding pocket region  $\leq 1.5 \text{ \AA}$ ), docking poses with  $\text{rmsd} < 2.5 \text{ \AA}$  to the native structure had been generated among all 100 solutions, respectively (Figure 6). In the case of the FRODA simulations, no ‘good’ docking pose could be generated, which may result from less closely approaching

the bound TAR RNA conformations during the simulations as well as the lower stereochemical quality of the generated conformations.

When assessing these results, it should be considered that docking to RNA structures is still more challenging than that to proteins.<sup>24</sup> In fact, the best success rates reported so far for docking solutions on the first rank are 42% (45%) for a data set of 31 RNA complexes using a  $2.0 \text{ \AA}$  ( $2.5 \text{ \AA}$ ) rmsd cutoff when docking with AutoDock/DrugScore<sup>RNA</sup><sup>67</sup> and 65% (74%) for a data set of 57 RNA complexes used in a study by Guilbert et al.<sup>15</sup> When tested on 28 complexes common to both studies, both methods perform equally (AutoDock/DrugScore<sup>RNA</sup>: 50%; and Guilbert et al.: 57%).<sup>15</sup> We note that this already appoints to a strong data set dependence.<sup>24</sup> In the case of the TAR RNA structures, both methods perform less satisfactorily, even in redocking experiments (Supporting Information, Table S3). Out of six cases, the method of Guilbert et al. identifies a good binding geometry on the first rank only for the TAR-NeoB complex ( $\text{rmsd} = 2.47 \text{ \AA}$ ), and AutoDock/Drug-Score<sup>RNA</sup> succeeds in the TAR-NeoB ( $\text{rmsd} = 0.88 \text{ \AA}$ ) and the TAR-Rbt203 ( $\text{rmsd} = 1.59 \text{ \AA}$ ) cases.

Overall, this demonstrates that some of the simulation-generated bound TAR RNA conformations are apparently similar enough to an experimentally determined complex TAR RNA structure such as to allow for successful pose generation. As a prerequisite for successful docking into the generated structures, a rmsd of respective binding pocket regions  $\sim 1.5 \text{ \AA}$  was identified. Even in the case of successful docking the Rbt203 ligand into TAR-Arg, sampling additional TAR-RNA conformations turned out to be essential. In this case, cross-docking the Rbt203 ligand into the experimentally determined TAR-Arg structure failed to produce ‘good’ ligand poses, although the experimentally determined TAR-Arg and TAR-Rbt203 structures differ by

only 1.8 Å rmsd. Not being able to identify these poses on the first rank or as members of the largest cluster must then be attributed to a weakness of the applied scoring function and, consequently, calls for further improvement here. When related to the protein–ligand docking field, for which much more experience is available, these results are encouraging in that a drop in docking accuracy compared to redocking was often found to be mirrored by the degree to which a protein moves upon ligand binding,<sup>82,83</sup> so that docking to an apo form usually shows the largest deterioration.<sup>84</sup> In this regard, being able to move closer to a bound conformation by up to 2.3 Å rmsd, as in the case of TAR-free versus TAR-NeoB, and to come as close as 2.2 Å rmsd to a bound conformation, as in the case of TAR-free versus TAR-Rbt158, already is a valuable achievement.

**Abbreviations.** TAR-free, conformation of HIV-1 TAR RNA in the free state; TAR-Arg, conformation of HIV-1 TAR RNA when bound to argininamide; TAR-AcP, conformation of HIV-1 TAR RNA when bound to acetylpro-mazine; TAR-NeoB, conformation of HIV-1 TAR RNA when bound to neomycin B; TAR-Rbt158, conformation of HIV-1 TAR RNA when bound to Rbt158; TAR-Rbt203, conformation of HIV-1 TAR RNA when bound to Rbt203; TAR-Rbt550, conformation of HIV-1 TAR RNA when bound to Rbt550.

## CONCLUSIONS

Several X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and molecular dynamics (MD) studies have provided insights into the structure, dynamics, and binding properties of HIV-1 TAR RNA, thus making it a valuable test case for simulation methods for RNA. Here, we have performed MD and replica-exchange molecular dynamics (REMD) simulations on the unbound human immunodeficiency virus type-1 (HIV-1) TAR RNA structure and on three TAR RNA structures in a bound conformation, totaling ~250 ns in length, in order to compare the extent of sampling with that of the constrained geometrical simulation approach FRODA and the quality of the generated conformations.

Root-mean-square atomic fluctuations obtained from unbound REMD simulations agreed very well with characteristics of atomic motions as obtained from conformational changes within the experimentally determined NMR ensemble of the free structure. When comparing atomic fluctuations obtained by REMD simulations with those obtained by MD and framework rigidity optimized dynamic algorithm (FRODA) simulations, quantitatively similar patterns are observed in the case of the TAR-free structure. This result is striking because FRODA, despite being conceptually simpler and computationally much cheaper, thus performed equally well in sampling local conformational movements in the bulge and loop regions of TAR RNA. An appropriate rigid-cluster decomposition of TAR RNA used as input to FRODA turned out to be detrimental for success, and results for FRODA-simulated bound TAR RNA conformations suggest that using a less coarse-grained representation will be advantageous in general. Regarding the stereochemical quality of the generated conformations, backbone torsion angles and puckering modes of the sugar–phosphate backbone were reproduced equally well by MD and REMD

simulations, but further improvement is needed in the case of FRODA simulations. In particular, a restriction of torsion angles to preferred ranges by including knowledge about discrete dihedral angle combinations from experimental structures should be integrated into FRODA.

Another main goal of this study was to investigate whether MD, REMD, or FRODA simulations are able to sample known bound conformations of HIV-1 TAR RNA when starting from an unbound structure. As indicated by essential dynamics analysis, all three simulation approaches showed a tendency to sample bound conformations, with MD and REMD simulations being superior with respect to FRODA, as shown by maximal overlap values  $I_{\max}$  for modes of essential motions derived from the experimental and simulation-generated ensembles. This can be explained in that FRODA generates new conformers by satisfying existing constraints, so that only local motions consistent with the analyzed constraint network can be observed. Instead, breaking and reformation of constraints during the simulation may be necessary to achieve conformational transitions as large as observed for TAR RNA. The finding was also corroborated when structural deviations of the binding pocket regions of sampled conformations, which were started from the TAR-free structure, were evaluated with respect to bound TAR conformations. Notably, structural deviations could be reduced by up to 2.3 Å rmsd when compared to deviations between experimental structures, although only in the TAR-Rbt158 case a minimal rmsd with respect to bound conformations of ~2.5 Å was found, accompanied by a conformational free energy change of 2.0 kcal mol<sup>-1</sup>. With respect to rmsd between bound experimental TAR RNA, again, considerable sampling of conformational space could be observed during simulations, which allowed reducing rmsd deviations between TAR RNA binding sites by 3.0 Å with respect to deviations in the starting structures. When looking at minimal rmsd values of simulated TAR RNA started from bound structures with respect to other experimental bound conformations, out of 14 cases, MD (REMD) showed four (three) instances with rmsd < 3.0 Å and three (two) cases with rmsd < 2.0 Å.

These above results are consistent with the experimental view that the bound TAR RNA is stabilized by weak intermolecular interactions with the ligand<sup>27</sup> and that bound conformations are transiently sampled in the free ensemble,<sup>4,5</sup> following the conformation selection model. The results are also encouraging from an application point of view in that those simulated TAR RNA conformations can be used as receptor structures for docking. In fact, simulation-generated bound TAR RNA conformations with an rmsd of respective binding pocket regions <1.5 Å were found to be suitable for successful pose generation.

At present, the question remains as to how one can identify suitable bound conformations from the generated ensembles without already knowing an experimentally determined bound conformation. An energetic evaluation appears to be difficult, because the conformational variability of a receptor leads to a disfavorable reorganization energy that can be large and vary strongly, even for relatively well preorganized binding sites.<sup>85</sup> Instead, we propose to cluster the generated conformations with respect to deviations in the binding regions and to dock preferentially into cluster representatives of well-populated clusters.

As long as large conformational changes of the receptor remain a challenge for current docking approaches, docking into an ensemble of receptor structures generated, e.g., by simulation techniques, provides an attractive alternative. This holds true especially for molecules like TAR RNA, whose conformational transitions upon binding follow a conformation selection model. The present study represents an attempt to fathom the scope and the limitations of this strategy in the field of RNA–ligand recognition.

## ACKNOWLEDGMENT

This work was supported by the DFG (SFB 579, “RNA-ligand interactions”), Goethe-University, Frankfurt, Germany and Heinrich-Heine-University, Düsseldorf, Germany. S.F. acknowledges financial support from the Hessian Science Program, the Frankfurt International Graduate School for Science (FIGSS), and the Otto-Stern-School in Frankfurt. H.G. acknowledges fruitful discussions at the workshop “Dynamics under constraints II”, McGill University’s Bellairs Research Institute, Barbados, 2007. We are grateful to Christopher Pfleger for assistance in generating wheel plots and to the “Zentrum fuer Informations und Medientechnologie” (ZIM) at the Heinrich Heine-University, Düsseldorf, for computational support.

**Supporting Information Available:** Tables with the percent occupation of puckering modes and docking results as well as graphical representations of rigid-cluster decompositions and sampled backbone and glycosidic torsion angles of TAR RNA structures. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (1) Al-Hashimi, H.; Walter, N. RNA dynamics: It is about time. *Curr. Opin. Struct. Biol.* **2008**, *18*, 321–329.
- (2) 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*, 76–80.
- (3) Long, K. S.; Crothers, D. M. Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry* **1999**, *38*, 10059–10069.
- (4) Al-Hashimi, H. M.; Gosser, Y.; Gorin, A.; Hu, W.; Majumdar, A.; Patel, D. J. Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J. Mol. Biol.* **2002**, *315*, 95–102.
- (5) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **2006**, *311*, 653–656.
- (6) Musselman, C.; Al-Hashimi, H. M.; Andricioaei, I. iRED snalysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys. J.* **2007**, *93*, 411–422.
- (7) Cullen, B. R. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* **1986**, *46*, 973–982.
- (8) Gallego, J.; Varani, G. Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges. *Acc. Chem. Res.* **2001**, *34*, 836–843.
- (9) Murchie, A. I.; 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*, 625–638.
- (10) Aboul-ela, F.; Karn, J.; Varani, G. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* **1996**, *24*, 3974–3981.
- (11) Aboul-ela, F.; Karn, J.; Varani, G. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **1995**, *253*, 313–332.
- (12) Faber, C.; Sticht, H.; Schweimer, K.; Rösch, P. Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **2000**, *275*, 20660–20666.
- (13) Du, Z.; Lind, K. E.; James, T. L. Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.* **2002**, *9*, 707–712.
- (14) Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I.; Karn, J.; Lentzen, G.; Drysdale, M.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T.; Aboul-ela, F. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic “hot spots”. *J. Mol. Biol.* **2004**, *336*, 343–356.
- (15) Guilbert, C.; James, T. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model.* **2008**, *48*, 1257–1268.
- (16) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, *49*, 1023–1033.
- (17) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
- (18) Huang, S. Y.; Zou, X. Q. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 399–421.
- (19) Nabuurs, S. B.; Wagener, M.; De Vlieg, J. A flexible approach to induced fit docking. *J. Med. Chem.* **2007**, *50*, 6507–6518.
- (20) 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*, 9632–9640.
- (21) Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
- (22) Bahar, I.; Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.
- (23) Kazemi, S.; Krüger, D. M.; Sirockin, F.; Gohlke, H. Elastic potential grids: Accurate and efficient representation of intermolecular interactions for fully flexible docking. *ChemMedChem* **2009**, *4*, 1264–1268.
- (24) Fulle, S.; Gohlke, H. Molecular recognition of RNA. Challenges for modeling interactions and plasticity. *J. Mol. Recognit.* **2010**, *23*, 220–231.
- (25) Zhang, Q.; Stelzer, A.; Fisher, C.; Al-Hashimi, H. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **2007**, *450*, 1263–1267.
- (26) Olsen, G. L.; Bardaro, M. F., Jr.; Echodu, D. C.; Drobny, G. P.; Varani, G. Intermediate rate atomic trajectories of RNA by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **2010**, *132*, 303–308.
- (27) Al-Hashimi, H. M. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *ChemBioChem* **2005**, *6*, 1506–1519.
- (28) Tsai, C. J.; Ma, B. Y.; Nussinov, R. Folding and binding cascades: Shifts in energy landscapes. *Proc. Natl Acad. Sci. U.S.A.* **1999**, *96*, 9970–9972.
- (29) Tsai, C. J.; Kumar, S.; Ma, B. Y.; Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **1999**, *8*, 1181–1190.
- (30) Leulliot, N.; Varani, G. Current topics in RNA-protein recognition: Control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001**, *40*, 7947–7956.
- (31) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.
- (32) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (33) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (34) Wells, S.; Menor, S.; Hespeneheide, B.; Thorpe, M. F. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* **2005**, *2*, S127–136.
- (35) Lei, M.; Zavodszky, M. I.; Kuhn, L. A.; Thorpe, M. F. Sampling protein conformations and pathways. *J. Comput. Chem.* **2004**, *25*, 1133–1148.
- (36) Gohlke, H.; Thorpe, M. F. A natural coarse graining for simulating large biomolecular motion. *Biophys. J.* **2006**, *91*, 2115–2120.
- (37) Macchiarulo, A.; Nuti, R.; Bellocchi, D.; Camaiori, E.; Pellicciari, R. Molecular docking and spatial coarse graining simulations as tools to investigate substrate recognition, enhancer binding and conformational transitions in indoleamine-2,3-dioxygenase (IDO). *Biochim. Biophys. Acta* **2007**, *1774*, 1058–1068.
- (38) Jolley, C. C.; Wells, S. A.; Fronmre, P.; Thorpe, M. F. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. *Biophys. J.* **2008**, *94*, 1613–1621.
- (39) Macchiarulo, A.; Giacche, N.; Carotti, A.; Baroni, M.; Gabriele, C.; Pellicciari, R. Targeting the conformational transitions of MDM2 and MDMX: Insights into dissimilarities and similarities of p53 recognition. *J. Chem. Inf. Mod.* **2008**, *48*, 1999–2009.
- (40) Cozzini, P.; Kellogg, G. E.; Spyros, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.,

- Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sottriffer, C. A. Target flexibility: An emerging consideration in drug discovery and design. *J. Med. Chem.* **2008**, *51*, 6237–6255.
- (41) Fulle, S.; Gohlke, H. Analysing the flexibility of RNA structures by constraint counting. *Biophys. J.* **2008**, *94*, 4202–4219.
- (42) 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*, 4944–4955.
- (43) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (45) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (46) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. Molecular dynamics and quantum mechanics of RNA: Conformational and chemical change we can believe in. *Acc. Chem. Res.* **2010**, *43*, 40–47.
- (47) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
- (48) McDowell, E. S.; Spacková, N. a.; Sponer, J.; Walter, N. Molecular dynamics simulations of RNA: An in silico single molecule approach. *Biopolymers* **2007**, *85*, 169–184.
- (49) Orozco, M.; Noy, A.; Perez, A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185–193.
- (50) Freedman, H.; Huynh, L. P.; Le, L.; Cheatham, T. E.; Tuszyński, J. A.; Truong, T. N. Explicitly solvated ligand contribution to continuum solvation models for binding free energies: Selectivity of theophylline binding to an RNA aptamer. *J. Phys. Chem. B* **2010**, *114*, 2227–2237.
- (51) Kopitz, H.; Zivkovic, A.; Engels, J. W.; Gohlke, H. Determinants of the unexpected stability of RNA fluorobenzene self pairs. *ChemBioChem* **2008**, *9*, 2619–2622.
- (52) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T.; Laughton, C.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (53) Besseova, I.; Otyepka, M.; Reblova, K.; Sponer, J. Dependence of A-RNA simulations on the choice of the force field and salt strength. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701–10711.
- (54) Reblova, K.; Strelcova, Z.; Kulhanek, P.; Besseova, I.; Mathews, D. H.; Van Nostrand, K.; Yildirim, I.; Turner, D. H.; Sponer, J. An RNA molecular switch: Intrinsic flexibility of 23S rRNA helices 40 and 68 5'-UAA/5'-GAN internal loops studied by molecular dynamics methods. *J. Chem. Theory Comput.* **2010**, *6*, 910–929.
- (55) Zhang, Y. F.; Zhao, X.; Mu, Y. G. Conformational transition map of an RNA GCAA tetraloop explored by replica-exchange molecular dynamics simulation. *J. Chem. Theory Comput.* **2009**, *5*, 1146–1154.
- (56) Fulle, S.; Gohlke, H. Constraint counting on RNA structures: Linking flexibility and function. *Methods* **2009**, *49*, 181–188.
- (57) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins* **2001**, *44*, 150–165.
- (58) Gohlke, H.; Kuhn, L. A.; Case, D. A. Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* **2004**, *56*, 322–337.
- (59) Jacobs, D. J.; Thorpe, M. F. Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.* **1995**, *75*, 4051–4054.
- (60) Farrell, D. W.; Mamonova, T.; Kurnikova, M.; Thorpe, M. F. Generating pathways for free energy calculations in proteins using constraint-based conformational sampling. *Biophys. J.* **2009**, *96*, 407a.
- (61) Ichiye, T.; Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205–217.
- (62) Lavery, R.; Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic-acids. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.
- (63) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (64) Martinek, V.; Bren, U.; Goodman, M. F.; Warshel, A.; Florian, J. DNA polymerase beta catalytic efficiency mirrors the Asn279-dCTP H-bonding strength. *FEBS Lett.* **2007**, *581*, 775–780.
- (65) Udommaneethakanit, T.; Rungratmongkol, T.; Bren, U.; Frecer, V.; Stanislav, M. Dynamic behavior of avian influenza a virus neuraminidase subtype H5N1 in complex with oseltamivir, zanamivir, peramivir, and their phosphonate analogues. *J. Chem. Inf. Model.* **2009**, *49*, 2323–2332.
- (66) Bren, U.; Lah, J.; Bren, M.; Martinek, V.; Florian, J. DNA duplex stability: The role of preorganized electrostatics. *J. Phys. Chem. B* **2010**, *114*, 2876–2885.
- (67) Pfeffer, P.; Gohlke, H. DrugScore<sup>RNA</sup> - Knowledge-based scoring function to predict RNA-ligand interactions. *J. Chem. Inf. Model.* **2007**, *47*, 1868–1876.
- (68) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1355–1363.
- (69) 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.
- (70) Morris, G.; Goodsell, D.; Halliday, R.; Huey, R.; Hart, W.; Belew, R.; Olson, A. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1999**, *19*, 1639–1662.
- (71) Ravindranathan, S.; Kim, C.-H.; Bodenhausen, G. Cross correlations between <sup>13</sup>C-<sup>1</sup>H dipolar interactions and <sup>15</sup>N chemical shift anisotropy in nucleic acids. *J. Biomol. NMR* **2003**, *27*, 365–375.
- (72) Ahmed, A.; Gohlke, H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins* **2006**, *63*, 1038–1051.
- (73) Ahmed, A.; Gohlke, H. *Multiscale modeling of macromolecular conformational changes*. In *Proceedings of the 1st International Conference on Computational & Mathematical Biomedical Engineering - CMBE09*; Nithiarasu, P., Löhner, R., Eds.; Swansea: U.K., 2009; pp 219–222.
- (74) Auffinger, P.; Westhof, E. Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.* **1997**, *27*, 454–63.
- (75) Feig, M.; Zacharias, R.; Pettitt, B. M. Conformations of an adenine bulge in a DNA octamer and its influence on DNA structure from molecular dynamics simulations. *Biophys. J.* **2001**, *81*, 352–370.
- (76) Barthel, A.; Zacharias, M. Conformational transitions in RNA single uridine and adenosine bulge structures: A molecular dynamics free energy simulation study. *Biophys. J.* **2006**, *90*, 2450–2462.
- (77) Allain, F. H.; Varani, G. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **1995**, *250*, 333–353.
- (78) Steuber, H.; Zentgraf, M.; Gerlach, C.; Sottriffer, C. A.; Heine, A.; Klebe, G. Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. *J. Mol. Biol.* **2006**, *363*, 174–187.
- (79) Zentgraf, M. Characterization of binding pocket flexibility of aldose reductase., Ph.D. Dissertation, Philipps-University Marburg: Marburg, Germany, 2006.
- (80) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915–920.
- (81) Eyrisch, S.; Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **2007**, *50*, 3457–3464.
- (82) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.
- (83) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-ligand docking against non-native protein conformers. *J. Chem. Inf. Model.* **2008**, *48*, 2214–2225.
- (84) 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.
- (85) Ahmed, A.; Kazemi, S.; Gohlke, H. Protein flexibility and mobility in structure-based drug design. *Front. Drug Des. Discovery* **2007**, *3*, 455–476.
- (86) Casiano-Negroni, A.; Sun, X.; Al-Hashimi, H. M. Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: New insights into the role of counterions and electrostatic interactions in adaptive recognition. *Biochemistry* **2007**, *46*, 6525–6535.