Analyzing the Flexibility of RNA Structures by Constraint Counting

Simone Fulle and Holger Gohlke
Department of Biological Sciences, Molecular Bioinformatics Group, J. W. Goethe-University, Frankfurt, Germany

ABSTRACT RNA requires conformational dynamics to undergo its diverse functional roles. Here, a new topological network representation of RNA structures is presented that allows analyzing RNA flexibility/rigidity based on constraint counting. The method extends the FIRST approach, which identifies flexible and rigid regions in atomic detail in a single, static, three-dimensional molecular framework. Initially, the network rigidity of a canonical A-form RNA is analyzed by counting on constraints of network elements of increasing size. These considerations demonstrate that it is the inclusion of hydrophobic contacts into the RNA topological network that is crucial for an accurate flexibility prediction. The counting also explains why a protein-based parameterization results in overly rigid RNA structures. The new network representation is then validated on a $tRNA^{ASP}$ structure and all NMR-derived ensembles of RNA structures currently available in the Protein Data Bank (with chain length $\geq$40). The flexibility predictions demonstrate good agreement with experimental mobility data, and the results are superior compared to predictions based on two previously used network representations. Encouragingly, this holds for flexibility predictions as well as mobility predictions obtained by constrained geometric simulations on these networks. Potential applications of the approach to analyzing the flexibility of DNA and RNA/protein complexes are discussed.

INTRODUCTION

Understanding the flexibility characteristics of biomacromolecules is crucial to understanding their biological function. This holds true particularly for RNA molecules. Their intrinsic flexibility can be observed, e.g., during protein synthesis in the ribosomal complex (1,2). Other examples are the structural reorganization of riboswitches (3) or the hammerhead ribozyme (4). Conformational changes of RNA structures due to interactions with binding partners have been found as well (5–7).

RNA also has become a well-established target for drug design due to its key role in gene replication and expression (8–10). In this regard, considering flexibility is important because it enables the conformational adaptation of binding partners and influences binding thermodynamics (11–13).

However, knowledge about the dynamics of RNA is still limited because experimentally derived information about their flexibility characteristics, such as NMR relaxation measurements (14,15), are not yet as widely available as it is for proteins (16). Currently, the main source of dynamical information has been gained by the study of B-factors from x-ray crystallography or by atomic fluctuations derived from NMR structural ensembles, despite the limitations of these measures (17,18).

Alternatively, computational approaches such as MD simulations (19) or normal mode analyses (20) allow deeper insights into the dynamics of RNA structures in atomic detail. MD simulations are still too computationally expensive to investigate large macromolecules on a routine basis, however. Likewise, although all-atom normal mode analysis are reliable for investigating RNA structures, the much cheaper and, in the case of proteins, widely applied, elastic network models may not be best suited for more loosely packed systems such as RNA (21). Hence, there is still a need for efficient approaches that determine flexibility characteristics of RNA molecules, ideally on an atomic level.

In this study, we apply concepts firmly grounded in mathematics, solid state physics, and structural engineering that are promising in that sense. The FIRST (18) approach has been developed to identify flexible and rigid regions within biological macromolecules. Remarkably, a FIRST analysis of a molecule of several thousand atoms just takes a few seconds such that FIRST is also very efficiently applicable to large macromolecules (22,23) (S. Fulle, H. Gohlke, unpublished). For the analysis, a single, static 3D structure of the molecule is modeled as a so-called ''bond-bending network'' or ''molecular framework''. In these networks, vertices (joints) represent atoms, and edges (struts) represent covalent and noncovalent bond constraints (strong hydrogen bonds, salt bridges, and hydrophobic interactions) as well as angular constraints. A fast combinatorial algorithm, the pebble game (24,25), is then applied to determine the number and spatial distribution of bond-rotational DOF in the network and, hence, the local network rigidity.

The FIRST approach has been thoroughly validated to identify rigid clusters and collectively moving regions in proteins (18,26–29). The obtained rigid cluster decomposition also can serve as input for naturally coarse-grained

Submitted May 23, 2007, and accepted for publication December 20, 2007.

Address reprint requests to Holger Gohlke, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany. Tel.: 49-69-798-29411; Fax: 49-69-798-29527; E-mail: gohlke@bioinformatik.uni-frankfurt.de.

Abbreviations used: MD, molecular dynamics; FIRST, floppy inclusion and rigid substructure topography; 3D, three-dimensional; FRODA, framework rigidity optimized dynamics algorithm; DOF, degree(s) of freedom; dof, independent internal degree(s) of freedom; PDB, Protein Data Bank; GNM, Gaussian Network Model; NOE, nuclear Overhauser enhancement; RMSD, root mean-square deviation.

Editor: Kathleen B. Hall.

0006-3495/08/06/4202/18 $2.00

doi: 10.1529/biophysj.107.113415

simulations (30–32). Until now, however, FIRST has been applied to RNA structures only in the case of the ribosome (23). Although it is straightforward to investigate RNA structures based on the same flexibility and rigidity concepts applied to proteins, one needs to keep in mind that both systems have different structural features. Proteins are generally globular and more densely packed, whereas RNAs are elongated and more loosely packed (21). Likewise, the forces that lead to structure formation and stability are different in both cases—protein structures are dominated by the hydrophobic effect, whereas RNA structures are mainly stabilized by hydrogen bonds and base stacking interactions. Thus, a network representation used for the rigidity analysis that has been developed for proteins may not be appropriate for RNA systems.

Indeed, in this study, we show that the current protein-based parameterization of noncovalent bond constraints does not capture flexibility characteristics of RNA structures satisfyingly. Instead, it leads to too rigid RNA structures in general. This may be less severe in the case of canonical RNA double helices that are expected to be largely rigid (33). Yet, overconstrained RNA representations should be avoided for irregular RNA structures, where subtle differences in flexibility/rigidity may play an important role for function. Here, we present a new network representation that captures much better flexibility characteristics of RNA structures, based on modified geometrical and energetic criteria for including noncovalent constraints into the network.

THEORY

Rigidity analysis by constraint counting

The FIRST approach relies on a theorem by Laman (34) that precisely determines the dof within two-dimensional networks by applying constraint counting to all the subgraphs within the framework. In this way, rigid regions and flexible joints between them are identified in the network. According to the Molecular Framework Conjecture (35), such constraint counting can be generalized to a subtype of all 3D networks, bond-bending networks or molecular frameworks in which vertices are connected by edges, and bond-bending angles are modeled as additional constraints (18). Although the Molecular Framework Conjecture requires a rigorous proof, there are no known exceptions after years of exact testing (18,36).

The intrinsic flexibility within 3D generic (see below) bond-bending networks can be identified by determining the number and spatial distribution of bond-rotational DOF in the network (as implemented in the pebble game (24,25) algorithm). A constraint in the network is considered to be independent if breaking it affects the flexibility of the network. In contrast, a constraint is redundant if it can be removed without influencing the network rigidity. In the presence of dof (so-called ''floppy modes''), a region is underconstrained (flexible). In contrast, the corresponding region is overconstrained in the presence of redundant constraints. Finally, if there are as many internal DOF as there are constraints, the region is isostatically rigid. Note that the number of dof within a flexible region is usually much smaller than the number of rotatable bonds (''hinge joints''), because not all rotatable dihedral angles can be varied independently due to the interconnection of rings in the network (37).

Whereas the decomposition into rigid clusters and flexible regions only provides a qualitative picture, a continuous quantitative measure is given by a flexibility index f_i defined for each covalent bond i as follows:

$$f_i = \begin{cases} \dfrac{F_j}{H_j} & \text{in an underconstrained region} \\ 0 & \text{in an isostatically rigid cluster} \\ -\dfrac{R_k}{C_k} & \text{in an overconstrained region} \end{cases} \quad (1)$$

In underconstrained regions j, f_i relates the number of dof (F_j) to the number of potentially rotatable bonds (H_j). Conversely, in overconstrained regions k the number of redundant bonds (R_k) is related to the number of constraints (C_k). The flexibility index ranges from -1 to 1, with negative values in overconstrained regions and positive values in flexible ones. Further details about rigidity theory as well as the underlying algorithms, have been described elsewhere (18,24,36,38).

In contrast to interactions in force fields that allow for varying strengths, a constraint in a topological network is either present or it is not. Thus, given that the flexibility of biomacromolecules is largely determined by noncovalent interactions, the outcome of a flexibility analysis is mainly governed by the way hydrogen bonds, salt bridges, and hydrophobic interactions are modeled in the network (29).

MATERIALS AND METHODS

Topological network representation

As the modeling of the noncovalent constraints is crucial for a reliable flexibility prediction, we tested different criteria to include hydrophobic interactions and hydrogen bonds in the topological network representation of RNA structures.

Hydrogen bonds are included as distance constraints in the network depending on their geometry and interaction energy. A hydrogen bond is considered if 1), the donor-acceptor distance $\leq$3.6 Å; 2), the hydrogen-acceptor distance $\leq$2.6 Å; and 3), the donor-hydrogen-acceptor angle is $\geq$80°. Subsequently, the 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 (for details see Eq. S1 in Supplementary Material, Data S1 and Jacobs et al. (18)). By tuning the energy threshold E_{HB}, the number of hydrogen bonds included is varied, which influences the flexibility characteristics of the network. Choosing $E_{HB} = -0.6$ kcal/mol corresponds to the thermal energy at room temperature and so provides a natural choice (18). E_{HB}-values of -1.0 kcal/mol have also been reported in the literature, resulting in more flexible networks (31,32). Here, we tested the influence of E_{HB} on the flexibility prediction of RNA structures by setting E_{HB} to -0.6, -1.0, or -1.5 kcal/mol.

For the calculation of E_{HB}, the hybridization as well as donor or acceptor state for each nitrogen and oxygen atom in a RNA structure was defined. Both terminal oxygens of the phosphate group were considered to be neg-

atively charged in addition, allowing the formation of salt bridges with positively charged atoms (as may be the case in, for example, protein-RNA complexes). Salt bridges are considered to be stronger than hydrogen bonds and are treated by a different energy function (see Eq. S2 in Data S1).

Hydrophobic interactions between two carbon atoms are considered if the distance of the atoms D_{HC} is smaller than the sum of the van der Waals radii (1.7 Å for carbon) plus a threshold. Threshold values of 0.10 Å, 0.15 Å, and 0.20 Å were tested here, resulting in D_{HC} thresholds of 3.50, 3.55, and 3.60 Å, respectively. This threshold applies to all hydrophobic interactions, whether they occur between adjacent bases (''stacking interaction'') or not. Furthermore, according to the findings when counting constraints on canonical A-form RNA, the number of hydrophobic interactions between sequentially adjacent bases N_{HC} was limited to one or two.

Finally, the influence of fixing a priori the glycosidic bond was tested, as was the influence of modeling the ribose ring as flexible. For the latter case, the furanose ring was extended by two dummy atoms introduced between C4′ and O4′ as well as O4′ and C1′. This seven-membered ring system has one dof. The dummy atoms are otherwise ''inert,'' i.e., they do not form any noncovalent interactions with their molecular environment.

To evaluate the novel RNA parameterization proposed in this study, flexibility predictions of RNA structures were also performed applying 1), a topological network parameterization widely used to investigate proteins ($E_{HB} = -0.6$ kcal/mol and $D_{HC} = 3.65$ Å) (18,26) and 2), one used by Wang et al. (23) to analyze the flexibility characteristics of the ribosome ($E_{HB} = -1.5$ kcal/mol and $D_{HC} = 3.50$ Å). They are referred to as protein-based parameterization and the parameterization used by Wang et al. (23), respectively.

Constrained geometrical simulations with FRODA

FRODA uses a random walk strategy that does not take into account the outcome of previous moves to explore the conformational space of flexible parts of a macromolecule. The approach relies upon a decomposition of a macromolecule into rigid and flexible regions. Motions of a biomolecule are then guided by ''ghost templates'' that cover each rigid region. Atoms of a biomolecule are bound to the vertices of these rigid ghost templates. After a small random displacement of the atoms, bond and angle constraints (due to covalent and noncovalent bonds) are enforced by an iterative process in which ghost templates are fitted to the atomic positions, followed by fitting each atom to the position of its vertex to which it belongs. Overall, atoms in rigid regions are moved collectively. Dihedral angles are allowed to vary in that these angles are represented as ghost templates that overlap along the rotatable bond. Finally, inequality constraints associated with hard sphere van der Waals overlap are satisfied.

For the FRODA simulations, the decomposition of the RNA structures into rigid and flexible regions is used as input, as obtained from the respective topological network representation. For all other program parameters, default values were used. During the simulations, 10,000 conformers of each structure were produced, and every 100th conformation was saved for analysis. These simulations require between 1 and 2 h of computational time on a state-of-the-art single processor work station. The timings are comparable to all-atom vacuum normal mode analyses, given that the latter require extensive energy minimizations before diagonalizing the Hessian matrix.

FRODA is similar to the CONCOORD approach by de Groot et al. (39) (which is based on distance geometry pioneered by Crippen (40)) in that it also generates random protein structures that fulfill a set of interatomic distance constraints. CONCOORD starts from random atomic coordinates for each step of structure generation. Corrections are then applied iteratively to the positions of those atoms that are involved in interatomic distances that violate the upper and lower distance bounds. This procedure ensures that bias in the results is minimal, that there is no correlation between any two structures generated, and, hence, that the accessible space defined by the distance bounds is efficiently sampled. In contrast, FRODA generates snapshots by applying small distortions to existing structures, leading to a correlation between subsequent conformers. In turn, application of ghost templates guarantees that only conformations with the correct stereochemistry are generated by FRODA, whereas CONCOORD can generate both images at a chiral center (as distance constraints do not contain chirality information) and, later, needs to be corrected for this. Furthermore, the FRODA algorithm allows adding directional biases to the atomic motions, so that they are not completely random. This allows exploring the conformational pathway between two conformers, as in steered MD simulations (41). Perhaps the foremost distinction between the two approaches is that FRODA operates on a coarse-grained representation of the biomacromolecule derived from an initial decomposition into rigid and flexible regions. CONCOORD instead uses an atomic biomacromolecule representation.

RNA structures used for validation

The flexibility predictions were tested on GNRA (PDB code: 1ZIF (GAAA), 1ZIG (GAGA), and 1ZIH (GCAA); (42)) and UUCG (PDB code: 1HLX; (43)) tetraloops, a $tRNA^{ASP}$ structure (PDB code: 2TRA; (44)) determined by x-ray crystallography, and all NMR-derived RNA structures available in the PDB that have a chain length of at least 40 nucleotides (Table 1).

In the case of disordered residues in the $tRNA^{ASP}$ structure, nucleotide atoms of the first alternative (marked ''A'' in the PDB file) were used. The crystallographically resolved Mg^{2+} ion together with six surrounding waters was included as part of the network—modeling interactions between the

TABLE 1 RNA structures of the validation set

PDB code	Reference	Description	Chain length	RMSD (Å)
1A51	(90)	Loop D/E arm of the 5S rRNA (*E. coli*)	41	0.68
1A60	(91)	Pseudoknot	44	3.52
1CQL	(92)	SRP RNA domain IV	43	0.65
1MNX	(93)	Loop E region of the 5S rRNA (*Spinach chloroplast*)	42	1.38
1P5M	(94)	HCV IRES domain IIa	55	1.19
1P5O	(94)	HCV IRES domain IIa	77	1.78
1S9S	(95)	MVL PSI site	101	3.22
1YMO	(96)	P2b-P3 pseudoknot from telomerase RNA (*H. sapiens*)	47	1.51
1Z2J	(97)	HIV-1 frameshift inducing element	45	2.12
1ZC5	(98)	Signal essential for translational frameshift in HIV-1	41	1.02
2ADT	(99)	GAAA tetraloop-receptor complex	86	3.08
2FEY	(100)	Stem loop IV (*Tetrahymena telomerase*)	43	2.27

The RMSD has been calculated for phosphorus atoms between average structures from FRODA-generated ensembles using the RNA parameterization and the respective NMR starting structure. In the case of 1A60 and 2ADT, RMSD values >3.0 Å result from highly mobile termini; in the case of 1S92, relative motions of two loosely coupled domains lead to RMSD values >3.0 Å.

metal and water as covalent bonds and between water and the RNA as hydrogen bonds. Nucleotides pairing with the anticodon triplet, a spermine molecule, and all other waters were omitted. Hydrogens were added to the $tRNA^{ASP}$ structure using the LEaP program from the Amber 8 package (http://amber.scripps.edu) (45). Topology files for modified nucleosides were taken from http://ozone3.chem.wayne.edu/. The residue numbering scheme of $tRNA^{ASP}$ follows the one given in the PDB file, i.e., residue 47 is skipped according to the residue numbering scheme of $tRNA^{PHE}$. In the case of the NMR-derived structures, the flexibility analysis was performed on the first structure of the conformational ensemble.

Comparing flexibility predictions with experimental data

Results of the flexibility analysis were compared with experimentally observed atomic fluctuations. In the case of the crystallographically determined $tRNA^{ASP}$ structure, atom-based flexibility indices were calculated as the average of f_i values of covalent bonds in which the atom is involved and compared to experimental B-values. Furthermore, the root mean-square amplitudes of motion determined by FRODA (Eq. 2) were compared with experimental root mean-square fluctuations about the mean position of atom i estimated from the crystallographic B-value B_i according to Eq. 3 as follows:

$$\langle r_i^2 \rangle^{1/2} = \left[\frac{1}{N} \sum_{j=1}^{N} \Delta x_i(j)^2 \right]^{1/2}, \quad (2)$$

where $\Delta x_i(j) = x_i(j) - \bar{x}_i$ with $x_i(j)$ is the coordinate vector of atom i in conformation j and $\bar{x}_i$ is the mean position of atom i during the sample period. N is the number of samples.

$$\langle r_i^2 \rangle^{1/2} = \left(\frac{3B_i}{8\pi^2} \right)^{1/2}. \quad (3)$$

Likewise, we determined conformational variabilities for phosphorus atoms from the NMR ensembles (Table 1) and compared them to root mean-square amplitudes of motion calculated by FRODA. It is important to note that FRODA results have not been scaled to best fit experimental data, in contrast to comparisons of calculated and experimental atomic fluctuations in the case of elastic network models (46).

For the correlation between computed and experimental atomic fluctuations, the square of the Pearson correlation coefficient (R^2) was calculated. The F-test was applied to determine whether a statistically significant linear relationship between the atomic fluctuation values exists. The null hypothesis H_0: $R^2 = 0$ is rejected if the p-value <0.05. In this case, it is assumed that the alternative hypothesis H_A: $R^2 \neq 0$ is valid.

As discussed below, multiple reasons may account for predicted fluctuations not perfectly agreeing with experimental measures of mobility. To test the impact of outliers on the calculated R^2-values, R^2-values were re-calculated omitting such outliers. For this, a data point i is regarded an outlier if the absolute value of its standardized residual $|e_i^+|$ exceeds 2, as described previously (47). The R statistics program (http://www.r-project.org) was used for these calculations.

GNM

The online server *o*GNM (48) (http://ignm.ccbb.pitt.edu/GNM_Online_Calculation.htm) was used to calculate normal modes of motion for the tested RNA structures based on a GNM representation. A Kirchhoff matrix was constructed using selected atoms of the nucleotides as network nodes within a chosen cutoff distance r_c. These nodes are connected by harmonic springs with a uniform spring constant γ. The magnitude of the spring constant is determined such that calculated squared atomic fluctuation values best fit experimental ones (46). In general, best results with respect to experimental mobility data were obtained by representing each nucleotide by a single sphere located at the position of the phosphorus atom and setting $r_c = 19$ Å (49). Slightly smaller correlation coefficients were obtained if an alternative GNM was applied (48). Here, each nucleotide was represented by three nodes centered on the position of the phosphorus, sugar C4′, and base C2 atoms, respectively, and setting $r_c = 7$ Å or $r_c = 10$ Å. GNM results reported in this study were thus calculated using the ''one-node/$r_c = 19$ Å'' parameterization.

Calculation of NOE intensities from ensembles generated by FRODA

Experimental NOE upper bounds were obtained from the BioMagResBank database (50) (http://www.bmrb.wisc.edu). NOE data was available for RNA structures 1A51 (mrblock_id: 38183), 1A60 (mrblock_id: 38211), 1P5M (mrblock_id: 47210), 1P5O (mrblock_id: 47233), 1S9S (mrblock_id: 244569), 1Z2J (mrblock_id: 52569), 2ADT (mrblock_id: 52123), and 2FEY (mrblock_id: 126693). For each pair of nuclei i and j for which an experimental NOE upper bound is reported, individual distances were averaged over the entire FRODA-generated ensemble according to $r_{i,j}^{avg} = \langle r_{i,j}^{-6} \rangle^{-1/6}$, assuming that the timescale of internal fluctuations is longer than the overall tumbling time in the case of the relatively small RNA molecules (51). A violation of the experimental NOE upper bound $nmr_{i,j}$ was determined by $v_{i,j} = \langle r_{i,j}^{-6} \rangle^{-1/6} - nmr_{i,j}$, with the violation $v_{i,j}$ being considered zero if $\langle r_{i,j}^{-6} \rangle^{-1/6} \leq nmr_{i,j}$. Finally, average violations $\langle v_{i,j} \rangle$ were calculated by averaging the individual violations over all N experimentally determined NOE upper bounds according to Eq. 4 (52) as follows:

$$\langle v_{i,j} \rangle = \frac{1}{N} \sum_{i,j} v_{i,j}. \quad (4)$$

RESULTS AND DISCUSSION

Here, we introduce a new parameterization for RNA that leads to better flexibility predictions than those that have been used previously. Motivated by the constraint counting on A-form RNA described in the following sections, we propose more stringent criteria for the inclusion of hydrophobic interactions in an RNA network. In addition, we investigated the criteria for the inclusion of hydrogen bonds. The general applicability of the novel parameterization is validated by comparison of rigidity analysis results to experimental measures of RNA mobility. These results are compared to GNM calculations (46).

Constraint counting on canonical A-form RNA

To derive a new network parameterization for RNA, it is instructive to analyze the flexibility properties of an RNA structure by direct counting on covalent and noncovalent constraints in the molecular framework. The explicit combinatorial analysis will reveal which internal DOF are available in the topological network of the RNA, and how noncovalent bonds influence the network rigidity. A canonical A-form RNA was exemplarily chosen for this because it is a major building block in many RNA structures. Similar to our analysis, Whiteley recently illustrated constraint counting on simple secondary structural elements of proteins (i.e., rings, α-helices, β-sheets, and β-barrels) (36).

General considerations

The analysis is based on the Molecular Framework Conjecture (35). The conjecture requires that only molecular frameworks in 3D-space are considered where the vertices are in a generic configuration (53). That is, these frameworks lack any special symmetry like collinear and parallel bonds. A visual inspection of several experimentally determined structures revealed that this condition is fulfilled for RNA in general. This holds even for substructural elements such as paired or stacked bases, which are coplanar (i.e., nongeneric or singular (53)) only in ideal canonical RNA geometries. Thus, constraint counting based on the Molecular Framework Conjecture can be applied to determine the rigidity of RNA networks.

For the analysis, the 3D RNA structure is modeled in atomic detail as bar-and-joint network. Distance constraints are inserted for covalent and noncovalent bonds as well as between next-nearest neighbors. The internal DOF result from torsional rotations around single bonds that are not locked in by other bonds. Dangling (or 1-valent) atoms (e.g., hydrogen atoms or atoms in hydroxyl-, amino-, and hydroxymethylen groups) do not affect the rigidity of the remaining network and are thus removed before the counting (54).

In the following sections, we will analyze the network rigidity of A-form RNA by successively considering substructural elements of increasing size. We will only present a summary of the constraint counting results here. Further details are provided in Data S1.

Sugar and base ring systems are rigid

According to constraint counting in 3D-space, the six-membered pyrimidine rings of cytosine and uracil are isostatically rigid, whereas the five-membered ring of ribose is overconstrained (36). Likewise, the purine ring systems of adenine and guanine are overconstrained. The corresponding networks are shown in Fig. 1 *a*).

Considering the five-membered furanose ring as rigid seems counterintuitive at first sight, because the ring is puckered rather than planar, leading to, in principle, 10 different envelope and twist conformations, respectively (55). However, the accessible conformational space of the sugar is restricted. Whereas desoxyribose in deoxyribonucleic acid (DNA) can transit between the two main conformations C3′-endo and C2′-endo, ribose in A-form RNA prefers the C3′-endo form (55). An ab initio conformational analysis also showed that the C3′-endo form of ribose is significantly more stable than the C2′-endo form (56). In addition, throughout a 30-ns long MD simulation of the ribosomal 16S helix 44, the sugar puckering remained consistently in the C3′-endo range (57). Thus, modeling the restricted motions of the ribose within RNA units by classifying it as rigid appears justified, although this precludes transitions between the C3′-endo and the C2′-endo conformations, which may be important for noncanonical elements. (For an alternative representation in which the ribose is modeled as a seven-membered ring with one dof see the Appendix.)

A nucleotide adds six dof to the system

Linking a base to a sugar leads to a system with one dof, if the glycosidic bond is considered to be freely rotatable. (Alternatively, the glycosidic bond can be fixed a priori. See the Appendix for more details.) Similarly, a phosphodiester bond between two sequentially adjacent riboses adds five dof to a network, again corresponding to the five single bonds along the backbone.

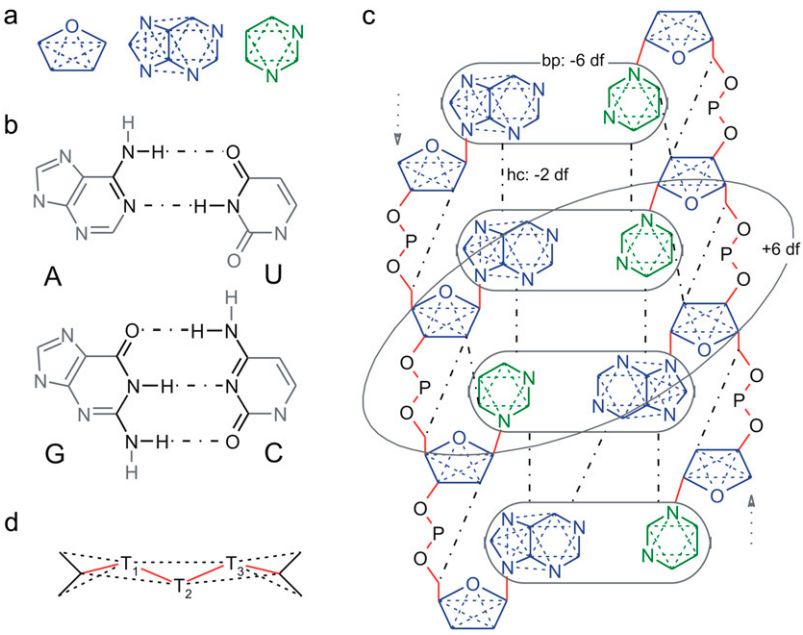

FIGURE 1 Topological network representation of RNA units. Constraints between nearest neighbors are indicated by straight lines, constraints between next nearest neighbors (angle constraints) by dashed lines. Flexible hinges are shown in red, rigid regions in green, and overconstrained regions in blue. (*a*) Scaffold representation of the ring systems of the ribose, purine, and pyrimidine bases. (*b*) Watson-Crick base pairing leads to the formation of one (in the case of AU) or two (in the case of GC) eight-membered ring systems in the network. The paired bases form a rigid cluster. (*c*) Network representation of a canonical A-form RNA applying the RNA parameterization. For reasons of clarity, angle constraints are only indicated in the sugar and base scaffolds, and hydrogen bonds between bases are omitted. Hydrophobic constraints are indicated by black dashed-dotted lines. (*d*) Hydrophobic interactions are modeled as bridges (''tethers'') of three pseudoatoms ($T_{1,2,3}$) between two hydrophobic atoms.

Base pairing reduces the number of dof by six

Watson-Crick base pairing leads to the formation of one (in the case of an adenine-uracil (AU) basepair) or two (in the case of a guanine-cytosine (GC) basepair) eight-membered ring systems in the network (Fig. 1 *b*), which are fused to other rings. Overall, this leads to the rigidification of the system of annelated five-, six-, and eight-membered rings in both the case of AU and GC. In total, the pairing of two bases removes six dof.

Each nucleotide pair adds six dof to a canonical A-form RNA

Each phosphodiester bond yields five dof, and each glycosidic bond adds another dof. Base pairing in turn removes six dof. Thus, adding a base-paired nucleotide pair to an RNA double strand in general adds six dof to the system.

One still needs to consider that, e.g., the 2′OH group of the ribose can be involved in hydrogen bonding, which may add additional constraints to the network. However, as revealed by MD simulations (58), the 2′OH group in RNA helices interacts preferentially with surrounding water molecules instead of, e.g., the O4′ atom on the 3′ side of the strand, as frequently suggested. Hence, we decided to not consider constraints involving the ribose 2′OH group in the A-form RNA analysis. Likewise, frequently occurring intrastrand C–H···O hydrogen bonds are not included because their interactions are usually weak.

In total, when only covalent and hydrogen bond constraints between base pairs are considered, a canonical A-form RNA with m base pairs has $8 + 6\,(m - 2)$ dof (see Data S1) and is thus highly flexible.

Hydrophobic interactions further rigidify the RNA

RNA helices intrinsically resist bend or twist deformations (33). Thus, they should be described as fairly rigid (at least locally), which is in contrast to the high flexibility found so far. For this, additional constraints due to hydrophobic (or stacking) interactions need to be included in the network to reduce the flexibility.

In the analysis of proteins, it has proven valuable to model hydrophobic interactions as bridges of three pseudoatoms between two hydrophobic atoms (Fig. 1 *d*). Each hydrophobic tether removes two DOF from the network (26). Taking into account that each nucleotide pair adds six dof to the network, the overall helix already becomes rigid if three independent hydrophobic tethers are inserted between sequentially adjacent nucleotide pairs.

Protein-based parameterization is not applicable for RNA structures

In a canonical A-form RNA, two or three hydrophobic tethers are inserted between neighboring pyrimidine or purine bases, respectively, if the protein-based parameterization (see Materials and Methods) is used. Between neighboring 3′-purine and 5′-pyrimidine bases, even seven hydrophobic tethers are included. Additional hydrophobic constraints are formed between bases and sugars and between the C2′ of residue i and the C5′ of the adjacent residue $i+1$. A protein-based parameterization thus results in a considerably overconstrained A-form RNA. As we show below, this is also true for noncanonical RNA conformations, which are predicted to be too rigid.

New criteria reduce the number of hydrophobic tethers in RNA units

To more realistically capture the flexibility properties of RNA structures, the number of constraints due to hydrophobic tethers in the network needs to be reduced. Initial tests showed that it is advantageous to address different types of hydrophobic interactions in the RNA network separately, which allows a better tuning of the spatial distribution of constraints.

First, we considered stacking interactions between sequentially adjacent bases. Because already three hydrophobic tethers between sequentially adjacent base pairs lead to the rigidification of an RNA double helix, we anticipated finding a better flexibility characterization if the number of hydrophobic interactions between such bases were limited. We tested limits of hydrophobic tethers N_{HC} of one and two. This criterion can be applied in general to all adjacent bases, irrespective of whether they are in a coplanar orientation (indicating stacking) or not.

Second, hydrophobic interactions between sugars or between sugars and bases were considered. Initially, the number of these interactions was reduced by applying a more stringent distance criterion D_{HC} for the identification of hydrophobic interactions. In the course of the study, it turned out to be advantageous to also apply this criterion for interactions between adjacent bases. Although it does not influence hydrophobic tether formation between stacked bases (as their distance is smaller than the cutoff value), it does so for bases that are not coplanar. In this regard, differences between hydrophobic and stacking interactions between adjacent bases in noncanonical RNA structures are taken into account. The resulting topological network of a canonical A-form RNA helix is shown in Fig. 1 *c*). In addition to interactions between bases, hydrophobic interactions are also included between C2′(i) and C5′($i + 1$) of the following residue and, in the case of adjacent pyrimidine bases, between C2′(i) and C6($i + 1$).

Taken together, this network representation results in a rigid canonical A-form RNA, i.e., canonical A-form RNA is modeled as a slightly overconstrained rod. Whereas this result is at variance with finite persistence lengths previously reported for double-stranded RNA (59,60), it is consistent with the view that RNA structures are formed from relatively rigid duplexes that are linked by flexible motifs (61). In fact, the asphericity especially found for medium-sized RNA has

been attributed to the formation of long helices due to coaxial stacking, which are expected to be rigid with large persistence lengths (61).

Applicability to other 3D motifs

It is reassuring that hydrophobic (stacking) interactions proved to be important in addition to Watson-Crick base pairing in the above analysis of A-form RNA. This raises the question, however, whether other 3D motifs that are (particularly) stabilized by non–Watson-Crick base pairs, base-phosphate, and base-sugar contacts would also come out of the analysis as stable substructures. As an example, we investigated RNA hairpins (62), in which specific sequence motives such as GNRA and UNCG tetraloops (where *N* is any nucleotide and *R* is a purine) are known to be unusually stable (63). Indeed, when applying the new network parameterization for RNA (see below), both structural motifs show great stability (Fig. S1 in Data S1).

In the case of the UUCG tetraloop (PDB code: 1HLX), a rigid stem comprising both the first and last loop nucleotides and a flexible loop tip are identified. This agrees well with order parameters derived from NMR experiments (14) and MD simulations (64). In particular, the looped out residue U7 (numbering according to Villa and Stock (64)) is known to exhibit enhanced conformational fluctuations (14,64). Accordingly, the backbone surrounding U7 is identified as flexible by our analysis. The unusual U·G base pair within the loop region involving both base-base and base-sugar hydrogen bonds sterically restricts the loop, leaving little flexibility (62). Our finding of a largely stable UUCG loop (except for U7) agrees with this finding (Fig. S1 *a* in Data S1).

In contrast, a higher degree of flexibility has been found in general for GNRA loops (62). All three GNRA hairpins investigated (GAAA, GAGA, and GCAA) show the same overall structural motif and are stabilized by networks of heterogeneous hydrogen bonds (42). Yet, different flexibility predictions of the GNRA loops result due to slightly different hydrogen bonding patterns and stacking interactions (42) (Fig. S1 *b* in Data S1). Whereas in the case of the GCAA loop, for example, the backbone of the third nucleotide is predicted to be flexible, it is rigid in the other two cases. This can be explained by the absence of stacking interactions between the second and third nucleotides in the first case (42). Conversely, only in the case of the GAGA loop do two hydrogen bonds occur between the first and the last loop nucleotides (42), leading to a rigidification of the two nucleotides and, thus, an overall stable loop backbone.

In summary, the flexibility predictions of RNA tetraloops are in agreement with experimental findings, particularly when considering that the motifs show considerable stability. This suggests that the new network parameterization should also be applicable in cases where favorable interactions other than Watson-Crick base-pairing or stacking occur.

Evaluating the flexibility predictions

The new network parameterization will now be evaluated by comparing results of rigidity analyses to experimental observations. In the case of proteins, crystallographic B-values and fluctuation data from MD simulations were used for this (18,29). We would like to note, however, that both measures report on the mobility of atoms, whereas analyzing a network by constraint counting provides information about flexibility/rigidity. Flexibility is a static property that describes the possibility of motion. Phrased differently, flexibility denotes the ability of a region to be deformed. From the study of flexibility alone, however, no information is available about the direction and magnitude of the possible motions (65). Thus, a perfect agreement between flexibility predictions and experiment cannot be expected in the case of, e.g., a mobile rigid body (such as a moving helix). Nevertheless, as detailed experimental information about RNA flexibility (as given by NMR relaxation data) is rarely available, we resorted to comparing flexibility predictions for RNA to crystallographic B-values as a first step.

In addition, we compared conformational variability (mobility) information derived from NMR ensembles with atomic fluctuations calculated from ensembles generated by constrained geometric simulations with FRODA (31). This is based on the following reasoning. FRODA relies upon a decomposition of a macromolecule into rigid and flexible regions. Flexible parts of the molecule are then moved through allowed regions of conformational space using random Brownian type (Monte Carlo) dynamics, whereas atoms in rigid clusters move collectively. The implicit assumption for the comparison is that only a physically realistic representation of the network will result in a proper decomposition into rigid and flexible regions of the RNA. Only then it can be expected that FRODA simulations will generate conformational ensembles whose fluctuation data will agree with that of NMR ensembles. Note, however, that because FRODA generates new conformers by satisfying existing constraints, only local motions consistent with the analyzed constraint network can be observed. Large conformational movements frequently observed in RNA (3–5,7) that require changes in the constraint network cannot be detected this way. To overcome this limitation, the approach must be extended in the future such as to allow for meaningful constraint breaking and formation during a constrained geometric simulation run.

Comparing flexibility predictions to crystallographic B-values

Qualitative comparison of flexibility indices with B-values

Initially, we compared the results from flexibility predictions with B-values of the $tRNA^{ASP}$ structure (Fig. 2; PDB code: 2TRA). For $tRNA^{ASP}$, detailed information about flexibility and mobility is also available from normal mode analysis (66), GNM calculations (49), and MD simulations (67), thus making it a well-suited test case.

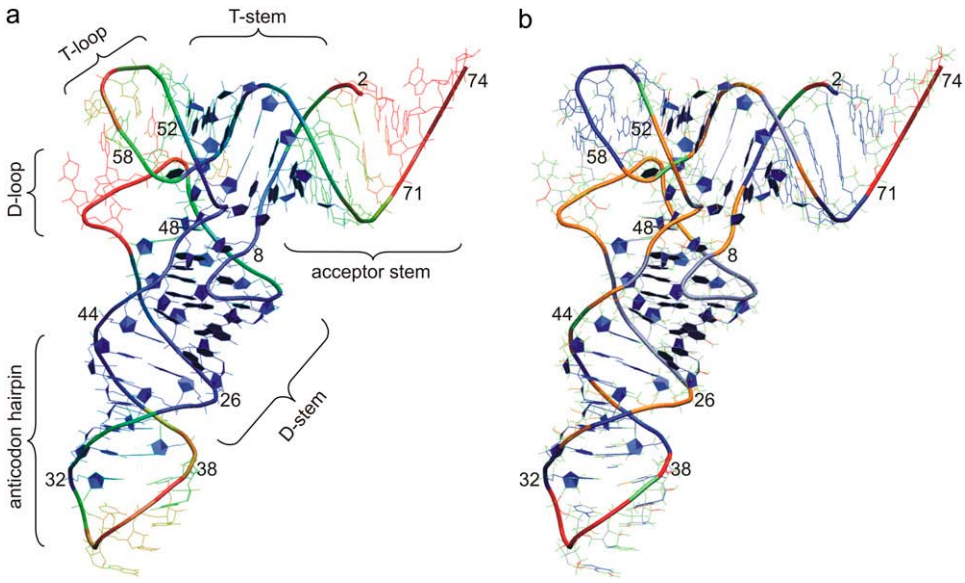

FIGURE 2 Experimental mobility information versus flexibility prediction for the $tRNA^{ASP}$ structure (PDB code: 2TRA). (*a*) Crystallographic B-values are mapped onto the $tRNA^{ASP}$ structure, using a color gradient ranging from blue (0 $Å^2$) over green to red (≥40 $Å^2$). B-values >40 $Å^2$ are truncated as described previously (46). The coloring of the backbone is according to the B-values of the phosphorus atoms. The core region (as defined by Auffinger et al. (67)) is highlighted with filled sugar and base scaffolds. (*b*) Color-coded representation of flexibility indices (Eq. 1) obtained by a flexibility analysis of the $tRNA^{ASP}$ structure using the RNA parameterization. Overconstrained regions are indicated by blue colors (*dark blue*, $f_i \leq -0.2$; *light blue*, $-0.2 < f_i < 0.0$), rigid regions are represented in green color ($f_i = 0.0$), and flexible regions are shown in red colors (*orange*, $0.0 < f_i < 0.2$; *red*, $f_i \geq 0.2$). The coloring of the backbone is according to the flexibility indices of the phosphorus atoms. The core region (as defined by Auffinger et al. (67)) is highlighted with filled sugar and base scaffolds.

In the tRNA structure, an Mg^{2+} ion bound to six water molecules and a spermine ligand have been resolved. They are located in the anticodon hairpin and in the acceptor stem, respectively. Because the position of the spermine does not allow the formation of any constraints to the RNA, the molecule was not included into the network representation. Similarly, in a normal mode analysis by Nakamura et al. (66), spermine was omitted for the same reason. In contrast, interactions between the Mg^{2+}-water complex and the tRNA were included as constraints, although performing the analysis without the Mg^{2+}-water complex did not change the flexibility prediction in this case, which is consistent with a GNM study on this structure (49). However, this finding cannot be expected in general, because Mg^{2+}-mediated bonding is a major source of stabilization for RNA structures (67), and we recommend including Mg^{2+}-mediated bonding as constraints in general.

Interactions mediated by other structural water molecules also may influence the flexibility of the tRNA. For example, MD simulations of the anticodon loop of $tRNA^{ASP}$ have demonstrated the importance of long-lived hydration patterns in the stabilization of loop structures (68). However, we decided not to include additional water molecules into the tRNA network based on previous findings that showed only a negligible difference in the flexibility characteristics of a protein-protein complex when structural waters were considered (29). Finally, interactions to crystal lattice neighbors can also impact the flexibility of the tRNA, particularly in the case of peripheral regions such as the acceptor and anticodon extremities (69). These intermolecular interactions were not considered in the flexibility analysis.

The best agreement between the flexibility prediction and the B-values was achieved by setting the energy threshold for a hydrogen bond (see Materials and Methods and Data S1) $E_{HB} = -1.0$ kcal/mol, modeling hydrophobic interactions between carbon atoms using $D_{HC} = 3.55$ Å, and restricting the number of hydrophobic interactions between sequentially adjacent bases N_{HC} to 1. These settings will be referred to as ''RNA parameterization'' in the following sections.

For validation, B-values and flexibility indices obtained by constraint counting (Eq. 1) are mapped color-coded onto the tRNA structure in Figs. 2, *a* and *b*, respectively. Green and blue colors indicate rigid regions, whereas red colors indicate flexible ones. We will focus on backbone atoms for the comparison because the sugar and base moieties are intrinsically rigid (see above).

As described by Auffinger et al. (67), we defined those residues as part of the core of the structure whose B-values are below 15 $Å^2$. Using the RNA parameterization, the flexibility prediction correctly identifies large parts of the core including the D-stem and parts of the acceptor and anticodon stem as rigid. Some core residues are classified as flexible, particularly in the variable loop (residues 44 to 48) and, to a lesser extent, in the T-stem. On the one hand, this finding may be attributed in part to missing Mg^{2+} ions, which play an essential role in the stabilization of the tRNA core. As such, for the structures of $tRNA^{GLY}$ (70) and yeast $tRNA^{PHE}$ (71), strong Mg^{2+} binding sites were detected near the D-stem, the base pair U8·A14, and the tertiary base pairs between the D- and TΨC-loops (67).

In contrast, identifying residues 8 and 48 in the core region to be flexible agrees with the hinge function of the regions comprising these residues as detected by normal mode analysis (66) and proposed by Olson et al. (72). These hinges provide the basis for twisting motions of the acceptor arm and the anticodon arm. Similarly, the nucleotides at position 26,

which separates the D-stem and the anticodon stem, and 44 and 45 of the variable loop have been suggested to form a hinge along which the relative angle between the two stems is changed (2,73). In agreement with this, residues 26 and 45 are identified to be flexible by our analysis.

Consistent with detected flexible regions, conformational changing of the tRNA structure is crucial for effective proofreading of codon-anticodon complexes during protein synthesis in the ribosome. For example, the D-loop shape changes during codon recognition (74), although the distance between that loop and the anticodon loop amounts to 45 Å. In agreement with this, the D-loop is predicted to be flexible by our approach. Furthermore, the anticodon loop (residues 32–38) and the outer acceptor stem (residues 71–74) are correctly predicted to be flexible (except residue 32 and 37, which are classified to be overconstrained and rigid, respectively). The analysis results agree with the high experimental B-values in these regions, as well as with MD simulation results of the tRNA during decoding, which demonstrate that the flexibility of the acceptor and the anticodon region is essential for the tRNA selection (75).

Recall that, for structures with moving rigid bodies, flexibility analysis results and crystallographic B-values may not necessarily agree well. This is the case for the elbow region (residues 52–58, containing the TΨC loop) of the tRNA. The region is predicted to be mostly overconstrained, which results from cross-strand stacking of purines belonging to the T- and D-loop (76). However, the B-values indicate a high mobility of the residues of the TΨC loop. The mobility must thus originate from flexible residues that flank the rigid body. Indeed, residues 52 and 59 are identified by the rigidity analysis as flexible, and the residues in between show large fluctuations in a constrained geometric simulation (see below and Fig. 3 *a*).

Quantitative comparison of calculated fluctuations with experimental mobility information

So far, we have compared flexibility predictions obtained with the RNA parameterization and crystallographic B-values qualitatively. For a quantitative comparison, we applied FRODA simulations to the $tRNA^{ASP}$ structure. From the generated conformational ensembles, the mobilities of phosphorus atoms were determined as the RMSD of each atom about its mean position (Eq. 2). These values were compared to atomic fluctuations calculated from the B-values according to Eq. 3. We note that the calculated fluctuations and the ones obtained from crystallographic B-values will not necessarily compare perfectly, because both include different contributions to the atomic motions (29,77–79).

When using the RNA parameterization, a fair correlation ($R^2 = 0.27$; if nucleotides 73 and 74 are omitted as outliers, $R^2 = 0.34$) is found between both measures of mobility, demonstrating that the rigid cluster decomposition before the simulations is appropriate. The agreement is better than in the case of the original protein-based parameterization ($R^2 = 0.20$; if nucleotides 17, 20, 21, 34, and 74 are omitted as outliers, $R^2 = 0.11$) (Fig. S2 in Data S1). A visual inspection reveals that, in the latter case, the tRNA structure is predicted to be too

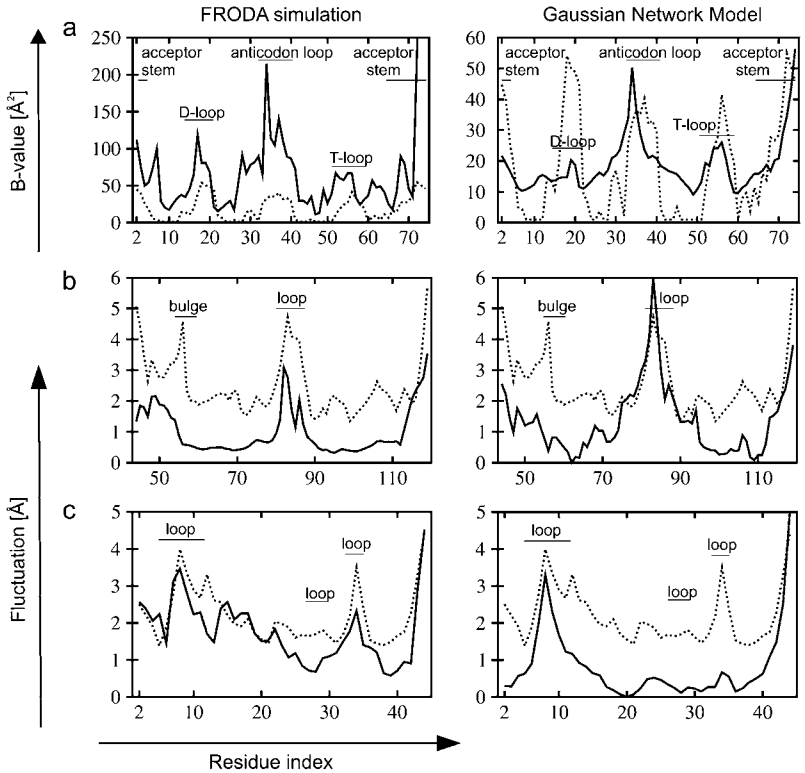

FIGURE 3 Mobility information of backbone phosphorous atoms predicted by FRODA simulations (*left panel*) and GNM (*right panel*) for the RNA structures (*a*) 2TRA, (*b*) 1P5O, and (*c*) 1A60 (*solid lines*). For comparison, crystallographic B-values (*a*) and conformational variabilities as measured in NMR are shown (*b* and *c*) (*dotted lines*). In the case of *a*, B-values predicted by FRODA of nucleotides 73 (575 $Å^2$) and 74 (1009 $Å^2$) are omitted for reasons of visualization.

rigid (Fig. S3 in Data S1), leading to insufficient mobility during the simulations. Doubling the simulation times did not change these results significantly, which indicates that the accessible conformational space of $tRNA^{ASP}$ has been appropriately sampled by the geometric simulations.

Overall, these findings demonstrate that the RNA parameterization reliably captures much of the conformational flexibility of the tRNA structure. In that sense, the new parameterization is superior to the original (protein-based) parameterization, which results in an overconstrained structure. It is particularly interesting to note that the approach allows identifying flexible residues that agree with previously determined hinge regions and, thus, provides insight into the flexibility characteristics of RNA structures on an atomic level.

Comparison to GNM results

We also compared our results with those obtained by a GNM (48). GNM calculations have been shown to provide a good description of the conformational dynamics of tRNA structures (49) and a data set consisting of 45 nucleotide-protein complexes and 19 RNA structures (including nine tRNA structures) (48). Here, we used the same GNM parameterization as in a previous study (49), which performed superior to alternative GNM representations (48) (see Materials and Methods for details).

Overall, equivalent results are obtained for the $tRNA^{ASP}$ structure (GNM: $R^2 = 0.32$; FRODA: $R^2 = 0.27$ (0.34)). For example, both methods correctly predict that very mobile nucleotides are located in the anticodon loop and outer parts of the acceptor stem (Fig. 3 *a*). Both methods predict nucleotide 34 in the anticodon loop to be overly mobile compared to experiment, which may be attributed to the fact that crystal-packing contacts involving this nucleotide are not considered in either approach (49). In the case of the D-loop, atomic fluctuations are considerably underestimated by the GNM, however. In contrast, a better agreement with experiment is found here by FRODA simulations based on the RNA parameterization.

Furthermore, the flexibility analysis based on the RNA parameterization appears to be more specific when defining hinge residues. As for GNM, hinge-bending regions are assumed to be highly restricted in motion (49). Based on this assumption, Bahar et al. (49) proposed hinge regions of $tRNA^{ASP}$ that comprise nucleotides 8–15 and 20–22 in the D-stem and 46–48 in the variable loop. Using the constrained counting instead, it is possible to specifically identify the involved hinge residues (8 and 48, 26 and 45) as flexible, in perfect agreement with previous results (2,73) (see above).

Comparing mobility predictions to NMR conformational variabilities

As an additional validation, we compared conformational variabilities of 12 RNA structures that have been determined by NMR and have a chain length of at least 40 nucleotides (Table 1) to atomic fluctuation data calculated by FRODA simulations. Each of the validation set structures consists of double helical regions and noncanonical structural elements. For example, in the pseudoknot structures of 1A60 and 1YMO single- and double-stranded regions are alternated. Furthermore, different types of loop structures are present, i.e., internal, stem, hairpin, tetra, and hepta loops. The data set structures thus comprise a broad range of irregular elements in RNA structures and should allow for a thorough validation of the flexibility predictions.

Influence of noncovalent constraints

We first tested the influence of including varying numbers of hydrogen bonds and hydrophobic interactions on the outcome of the simulation results. The squared correlation coefficients (R^2) between computed mobilities and experimental variabilities are listed in Table S1 in Data S1. Restricting the number of hydrophobic contacts between adjacent bases to $N_{HC} = 1$ has the largest impact on the results and is crucial for reliable flexibility predictions of RNA. This is demonstrated by the fact that, in this case, the best correlations between computed mobilities and experimental variabilities are found for most of the validation set structures (i.e., 10 out of 12 cases (Table S1 in Data S1)).

Next, adjusting the distance threshold D_{HC} up to which hydrophobic contacts are included is more important than choosing the cutoff value for the hydrogen bond energy function E_{HB}. For most of the investigated structures (i.e., 7 of those 10 cases for which $N_{HC} = 1$), the best correlations are obtained by setting $D_{HC} = 3.55$ Å between two carbon atoms.

Finally, we chose to set $E_{HB} = -1.0$ kcal/mol. First, none of the three tested E_{HB}-values (-0.6, -1.0, and -1.5 kcal/mol) resulted in significantly better results than the alternatives. The outcome of the approach is thus rather insensitive to this parameter, and choosing the median appeared natural. Second, an E_{HB}-value of -1.0 kcal/mol yields slightly more flexible structures than if an energy cutoff according to the thermal energy at 300 K (-0.6 kcal/mol) is chosen. This has been found to be advantageous in those cases where FIRST results are used as input to constrained geometric simulations (31) and rigid cluster normal mode analyses (32).

At this point, it is encouraging to note that the parameterization found to be optimal for predicting conformational variability by FRODA (i.e., $N_{HC} = 1$, $D_{HC} = 3.55$ Å, and $E_{HB} = -1.0$ kcal/mol) is identical to the one that works best when comparing flexibility indices to B-factors (see above). Both results thus mutually corroborate each other.

Comparison of fluctuation predictions based on different network parameterizations

We will now compare fluctuation predictions obtained by FRODA with the new RNA parameterization to those of the protein-based parameterization (18,26) and a parameteriza-

tion used by Wang et al. (23) for investigating the ribosome (see Materials and Methods). Correlation coefficients for all 12 investigated systems are provided in Table 2; backbone positional RMSD of the average structures of the simulated ensembles from the respective NMR starting structures are given in Table 1.

First, we tested whether these results are sensitive to the presence of ''outlier'' data points. For this, we repeated all analyses omitting such outliers (see Materials and Methods). The thus obtained correlation coefficients are provided in Table S2 in Data S1. Encouragingly, the same conclusions are found as if all data points were considered. Hence, we will only discuss results considering all data points below.

Second, we probed the robustness of the FRODA results by repeating simulations of the HCV IRES domain II (PDB code: 1P5O) nine times, always starting with a different random number seed. The resulting correlation coefficients are listed in Table S3 in Data S1. In 6 of 10 cases, the highest correlation coefficient was obtained with the RNA parameterization. The standard deviation of all 10 R^2-values calculated with the RNA parameterization is 0.1, demonstrating that the simulation results are reproducible. Based on this finding we decided to consider correlations to be significantly different if the difference of the R^2-values equaled at least 0.1.

Fig. 4 exemplarily shows plots of the correlations between conformational variabilities determined from NMR ensemble structures and calculated fluctuations by FRODA for the HCV IRES domain II (PDB code: 1P5O) and a pseudoknot RNA structure (PDB code: 1A60). Fig. 5 depicts the conformational ensembles from NMR and generated by FRODA simulations for both structures. Finally, Fig. 3, *b* and *c*, and Fig. S4 in Data S1 show atomic fluctuations predicted by FRODA compared to conformational variabilities from NMR on a per-residue basis for all 12 investigated systems.

TABLE 2 Coefficients of the correlation (R^2) between conformational variabilities derived from NMR ensembles and atomic fluctuations calculated by FRODA simulations or a GNM

PDB code	Protein-based*†	Wang et al.*‡	RNA-based*§	GNM¶
1A51	0.25	0.19	0.26	0.14
1A60	0.52	0.44	**0.62**	0.70
1CQL	*0.00*	**0.53**	0.21	–‖
1MNX	*0.00*	*0.03*	*0.09*	0.26
1P5M	*0.00*	*0.00*	*0.04*	0.08
1P5O	*0.04*	*0.04*	**0.52**	0.40
1S9S	0.07	*0.01*	0.18	0.32
1YMO	0.21	0.23	0.15	0.57
1Z2J	*0.01*	0.29	0.20	0.35
1ZC5	*0.00*	*0.03*	**0.38**	*0.01*
2ADT	*0.04*	0.36	0.27	0.46
2FEY	*0.04*	*0.07*	**0.39**	–‖

Cases for which no correlation was found at the $p = 0.05$ level are given in italics.

*R^2-values that are significantly the largest among all three parameterizations tested in the FRODA simulations are given in bold.

†D_{HC}: 3.65 Å, E_{HB}: −0.6 kcal/mol.

‡D_{HC}: 3.50 Å, E_{HB}: −1.5 kcal/mol.

§N_{HC}: 1, D_{HC}: 3.55 Å, E_{HB}: −1.0 kcal/mol.

¶Interaction sites are located at the P atoms; these sites are connected by harmonic springs with a uniform spring constant if the cutoff distance $r_c \leq 19$ Å.

‖Cases for which a negative correlation was found are not listed.

In both cases, the RNA parameterization (Fig. 4 *c*) leads to much better predictions of the fluctuation characteristics than the alternative parameterizations. Whereas both alternative parameterizations (Fig. 4, *a* and *b*) show no predictive power in the case of 1P5O ($R^2 \approx 0$), the RNA parameterization yields a fair correlation of $R^2 = 0.52$. Visual inspection of the rigid cluster decomposition of 1P5O shows that this is due to an overly constrained network representation in the first two cases, which results in no atom mobility or only a reduced atom mobility in the subsequent simulations. In contrast, the RNA is determined to be more flexible in the case of the new parameterization, which results in computed mobilities that agree much better with the experimental information: Only five residues show pronounced deviations from the correlation line. The mobility of residue U56 in the hinge bulge is underpredicted because the NMR ensemble contains two distinct conformations with the base moiety of U56 either stacked in or looped out, whereas only the looped out conformation is observed in our simulations. In turn, residues G82, C115, A116, and G117 are predicted to be too mobile, the first one residing in the hairpin loop region of domain IIb and the latter ones at the 3′ end of domain IIa.

In the case of the 1A60 structure, the alternative parameterizations already yield fair correlations ($R^2 = 0.52$ in the case of the protein-based parameterization and $R^2 = 0.44$ in the case of the parameterization used by Wang et al. (23)). Yet, the RNA parameterization results in a significantly better correlation of calculated and experimental fluctuation data, as indicated by $R^2 = 0.62$. Remarkably, the correlation line has a slope of 0.9 in this case, demonstrating that both small and large fluctuations are correctly predicted by the FRODA simulation. Only G5 in the last base pair of the hairpin stem of the T-arm is predicted to be too mobile.

This result is even more encouraging in view of the fact that the pseudoknot structure contains both single- and double-stranded regions. As fluctuations in both types of regions are equally well predicted by the simulations, this demonstrates that the underlying RNA network parameterization is equally well applicable in both cases.

Despite the success of the RNA parameterization, the predicted fluctuations do not perfectly correlate with the experimental conformational variabilities. Although experimental conformational variabilities <2 Å can usually be well predicted, predictions for those >3 Å show considerable deviations.

Because over- and underpredictions are observed, multiple reasons may account for this result. First, the RNA parameterization may still not be perfect, which can result in RNA regions being predicted too flexible or rigid. In fact, by system-specifically tuning the parameters that determine the inclusion of constraints into the network, better results may be obtained (data not shown). Nevertheless, the presented

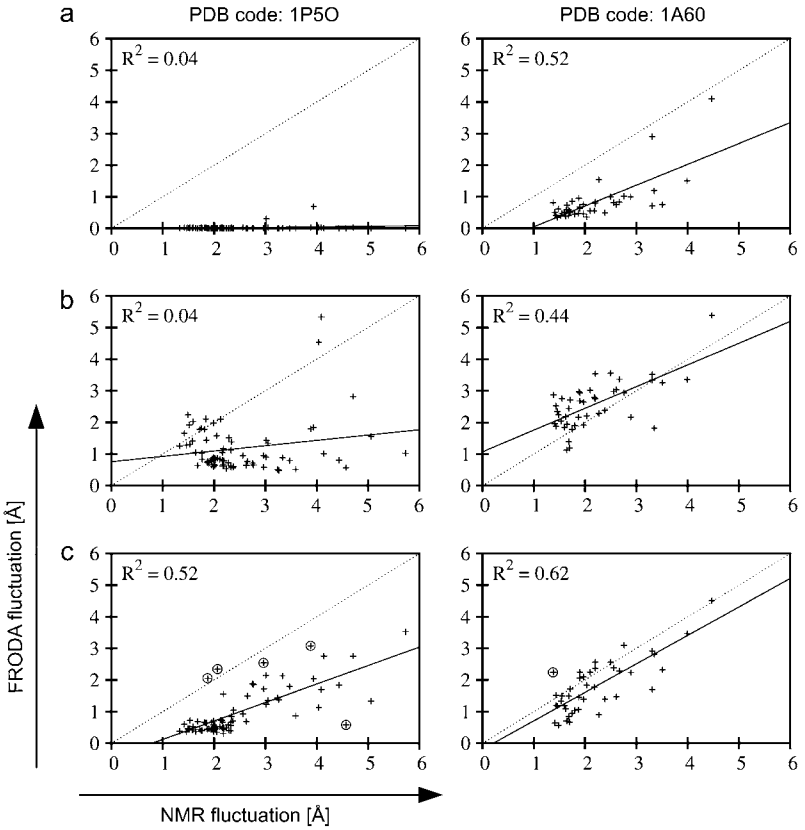

FIGURE 4 Atomic fluctuations predicted by FRODA simulations versus conformational variabilities as measured in NMR for RNA structures 1P5O (*left*) and 1A60 (*right*). For the FRODA simulations, a topological network representation according to (*a*) the protein-based parameterization, (*b*) parameterization used by Wang et al. (23), and (*c*) the RNA parameterization was used. In the latter case, residues that show pronounced deviations from the correlation line are marked by circles (see text for further discussion).

RNA parameterization shows the best general applicability, which is important for blind predictions in ''real-life'' scenarios. Second, overpredicted fluctuations may result if components that stabilize RNA structures such as Mg^{2+} ions (67) or structural waters (68) are not considered in the network representation. Such components were not included here, because no such information was available from the NMR structures. Finally, in the case of underpredicting fluctuations, sampling of flexible regions may not yet be sufficient. As demonstrated in the case of the tRNA, however, doubling the simulation time did not change the results qualitatively. In turn, fluctuations may seem to be underpredicted, although it is in fact the experimental value that overestimates the mobility. This may be true in particular as we interpret conformational variability from an NMR ensemble as atomic fluctuations. A region for which an insufficient amount of experimental restraints was available for structure determination (leading to a less well-defined structure) may then appear to be more ''flexible'' than it actually is. It can be expected that this is even more likely in regions that are mobile per se.

An alternative to overcome this limitation appears to be the comparison of original NMR data in terms of atom-atom distances based on NOE intensities with those calculated from the FRODA-generated RNA ensembles. These comparisons are widely performed to judge the quality of MD simulations (52). Accordingly, we calculated the average violations of the experimentally determined NOE upper bounds (Eq. 4; see Materials and Methods), where the average goes over the entire generated ensembles of 8 of the 12

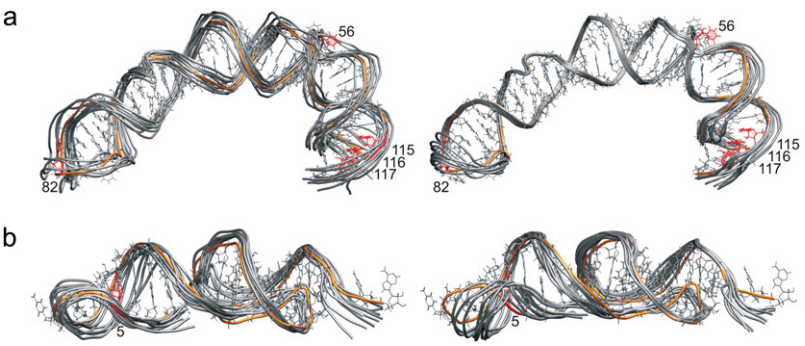

FIGURE 5 Conformational space of the RNA structures (*a*) 1P5O and (*b*) 1A60 spanned by the NMR ensemble (*left*) and conformations generated by a FRODA simulation using the RNA parameterization (*right*). From the NMR ensemble, the first 10 models of the PDB entry, and from the FRODA simulation, the starting structure (*orange*) as well as conformations generated every 1000 steps are shown. The first model of each NMR ensemble is highlighted in orange. Residues that are marked by circles in Fig. 4 are highlighted in red.

investigated RNA structures (in 4 cases, no experimental NOE upper bounds were available). In 7 of the 8 cases, more than 99% of the experimental restraints are matched with 0 Å violation (i.e., the computed atom-atom distance is smaller than the NOE upper bound). Only in the case of 1S9S, pronounced violations >0 Å occur in that 99.7 % (0.3%) of the experimental restraints are violated by <2 Å (>3 Å). When averaged over all experimentally determined NOE upper bounds, the simulated ensembles match the experiment with an average deviation of close to 0 Å again in 7 of 8 cases; the average deviation climbs to 1.73 Å only for 1S9S. Both the vanishing average deviation of upper bounds and the likewise vanishing number of individual NOE deviations in 7 of 8 cases can be understood by considering that FRODA generates only local motions consistent with the analyzed constraint network (see above). This is also reflected in backbone positional RMSD values between ensemble average structures and respective NMR starting structures of < or ~2 Å in 9 of 12 cases (Table 1). New conformers satisfy existing (short-range) constraints derived from the starting structure and, hence, very likely also satisfy experimental NOE upper bounds on which the structures are based. Thus, average deviations of NOE upper bounds (as well as individual NOE deviations) are only of limited value to judge the actual quality of the simulated FRODA ensemble. Only with recent developments in NMR methodology such as residual dipolar couplings can these limitations be addressed (16).

Table 2 shows R^2-values obtained for all tested RNA structures based on the protein-based parameterization, the parameterization used by Wang et al. (23), and the RNA parameterization. When considering differences between the three parameterizations as significant if the R^2-values differ by >0.1 (see above), the RNA parameterization outperforms the other ones in 4 cases, whereas the parameterization used by Wang et al. (23) does so only in 1 case. The protein-based parameterization is in no case superior to any of the other parameterizations. Thus, the RNA parameterization shows the overall best performance.

Comparison to GNM results

We also compared conformational variabilities obtained from the NMR ensembles to atomic fluctuations calculated by GNM (48) (Table 2), using the same GNM parameterization as for the $tRNA^{ASP}$ structure (49). This parameterization also provided the best agreement with the experiment in the case of the NMR structures compared to alternative ones (48). Nevertheless, in 3 of 12 cases, no correlation or even an anticorrelation between computed and experimental mobility values was obtained by GNM. Considering correlations to be significantly different if the difference of the R^2-values amounts to at least 0.1, either method (FRODA based on the RNA parameterization versus GNM) outperformed the other one in 5 of 12 cases. Only in the case of 1A60 and 1P5M did both methods perform equally well.

The complementary performance of FRODA and GNM on different RNAs is marked and calls for an explanation at the structural level. So far, it has been suggested that elastic network models (including GNM) are valid for many globular proteins systems but may not best suited for loosely packed systems such as small RNA structures (21). In fact, it was found that the quality of the GNM depends on the compactness of the RNA system (21). Further factors influencing the performance of the approaches could be the system size (48), the magnitude and type of the observed movements, or the degree of the rigid cluster decomposition used as input to the FRODA simulations. As for the 12 systems investigated, however, we were unable to identify a clear connection between the performance of FRODA or GNM and these factors.

Atomic fluctuations predicted by FRODA and GNM are compared to conformational variabilities from NMR on a per-residue basis for 1P5O and 1A60 in Fig. 3, *b* and *c*. Both methods correctly identify mobile nucleotides located in loop regions of 1P5O. In contrast, the GNM calculations underestimate the mobility of loop residues 33–35 in the case of 1A60, as do both methods for the bulge residues in the case of 1P5O. Overall, for both structures considered, the GNM appears to overestimate the mobility of one loop region at the expense of mobility predictions for the remaining structure. This ''tip effect'' occurs in systems with structural components that protrude out of the main body due to an imbalance of elastic forces among neighboring harmonic oscillators because of lighter packing around the tip region (80).

CONCLUSION

We have presented a new topological network representation of RNA structures that allows analyzing RNA flexibility/rigidity based on constraint counting. As constraints in the bond-bending network are ''all-or-nothing'' and the flexibility of biomacromolecules is largely determined by noncovalent interactions, sufficiently strong forces (which are included in the network) need to be distinguished from weaker ones (which are excluded).

As a first step, we analyzed the network rigidity of a canonical A-form RNA by counting on covalent and noncovalent constraints of network elements of increasing size, thereby determining the internal DOF that are available in the network. These theoretical considerations already show that it is the inclusion of hydrophobic contacts into the RNA topological network that is crucial for an accurate flexibility prediction and that the number of contacts between adjacent bases needs to be limited to capture the flexibility characteristics of RNA reliably. In particular, the counting explained why a protein-based parameterization results in overly rigid RNA structures.

The new network representation was validated on a $tRNA^{ASP}$ structure and all NMR-derived ensembles of RNA structures available that have a chain length of at least

40 nucleotides. From a methodological point of view, it is encouraging that the flexibility predictions demonstrate good agreement both qualitatively and quantitatively with the experimental mobility data and that the results are superior compared to predictions based on the parameterization used by Wang et al. (23) or a protein-based parameterization (18,26). This is not only true for information on infinitesimal (also called snap-shot or instantaneous (36)) motions as obtained directly by a flexibility analysis, but it also holds for finite amplitude motions determined by constrained geometric simulations using FRODA. Compared to GNM calculations, a comparable performance is found, with the FRODA-generated ensembles providing more detail than the coarse-grained GNM.

At first sight, one may be concerned that the somewhat disappointing performance of the protein-based parameterization (18,26) when applied to RNA results from overfitting of the parameters to protein. Hence, one may also worry about overfitting in the case of the RNA parameterization. In our opinion, this is not a cause for concern, for the following reasons: 1), As outlined in the Introduction, RNA and proteins have different structural features, and different forces lead to structure formation and stability in both cases. Hence, one should not expect a priori that a parameterization derived for one type of biomacromolecule is applicable to the other one, too. 2), The validation set of structures comprises a broad range of canonical and noncanonical structural elements. No bias of the flexibility prediction accuracy was found across the data set. 3), Most NMR RNA structures used for the validation are small, which may lead to the concern that the RNA parameterization will perform well on small RNA molecules but not large ones. When analyzing the protein-RNA composite network representation of the large subunit of the ribosome (see below), however, flexibility predictions for the ribosomal exit tunnel are in good agreement with experimental findings about the tunnel's role in cotranslational processes (S. Fulle, H. Gohlke, unpublished). Overall, we believe that these findings demonstrate the generality of the RNA parameterization.

Importantly, the flexibility analyses provide information on a local scale, i.e., on an atomic level, which is expected to provide more detailed insights than, e.g., results based on coarse-grained methods such as elastic network models. This has been successfully demonstrated for the identification of hinge residues in the case of $tRNA^{ASP}$. It is also particularly noteworthy that only a single structure is used as input, yet the results agree with experimental measures (crystallographic B-values or atomic fluctuations from NMR ensembles) obtained by averaging over ensemble and timescales that are out of reach for current state-of-the-art MD simulations. We note, however, that flexibility analyses can also be performed on ensembles of structures, e.g., derived by MD simulations, which becomes advantageous in the case of metastable systems that show more sensitivity of the flexibility analysis results to the underlying network representation (29).

With respect to the robustness, we have found that the analysis results only weakly depend on the choice of the energy threshold up to which hydrogen bonds are included into the network. This is particularly encouraging given that the energy function that is used for ranking the hydrogen bonds has been derived for proteins (81) and so may be less accurate for nucleic acids. In contrast, hydrophobic interactions have turned out to be critical for a successful characterization of the flexibility of RNA structures, as already found in the theoretical analysis of the A-form RNA. Still, several improvements of the network representation can be anticipated: 1), The strong repulsion between negatively charged phosphate groups is assumed to contribute to structural stability in nucleic acids (82). Modeling repulsion is difficult within the combinatorial approach followed in the pebble game, however, because this leads to one-way inequalities (the distance cannot become shorter) compared to two-way equalities (distance constraints) used so far (36). 2), Base stacking interactions are known to be both dependent on the type of the bases and the sequential context: Stacking interactions in general increase in the order pyrimidine-pyrimidine $<$ purine-pyrimidine $<$ purine-purine bases (55). In addition, stacking interactions are larger for sequences rich in G-C rather than A-U base pairs (83,84). Differences in base stacking interactions could be modeled by using varying numbers of pseudoatoms for the hydrophobic tethers. This approach has not been pursued so far.

In regard to using the RNA parameterization for analyzing DNA structures, one should notice that both types of molecules express different flexibility characteristics in response to the presence or absence of the 2′OH group (85). First, modeling the sugar as rigid may not be appropriate for DNA because the 2′-deoxyribose is known to transit between the two main conformations C3′-*endo* and C2′-*endo* (55). Second, a recent MD study revealed that the differences between flexibility and rigidity in both types of nucleic acids are much more complex than usually believed (86): RNA is very deformable along a small set of essential deformations, whereas DNA has a more degenerate pattern of deformability. Thus, more work clearly is needed in this area. Analyzing the flexibility of RNA in the presence of proteins is particularly interesting. First, RNA structural and dynamical changes associated with the recognition of cellular cofactors play an important role in biology (16). Second, the ribosome is a multi-complex composed of various RNAs and proteins, and understanding its flexibility characteristics is crucial to understanding its role as a molecular machine. In these cases, we propose to model the RNA components in the network according to the RNA parameterization presented here and the protein components according to the protein-based one. As the RNA/protein interface region is strongly stabilized by the formation of hydrogen bond networks and intermolecular hydrophobic cores (87), we suggest modeling these interactions according to the protein-based parameterization. We have analyzed such a composite network repre-

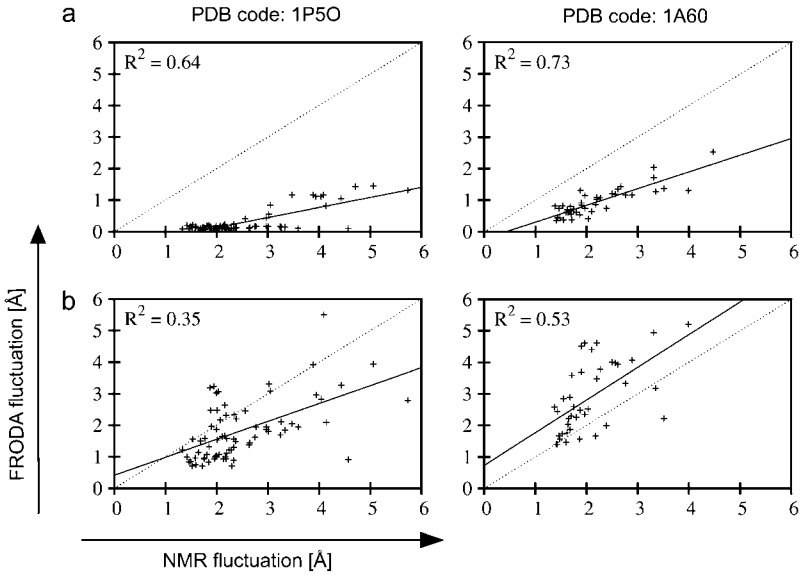

FIGURE 6 Atomic fluctuations predicted by FRODA simulations versus conformational variabilities as measured in NMR for RNA structures 1P5O (*left*) and 1A60 (*right*). For the FRODA simulations, a topological network representation according to (*a*) the RNA parameterization with fixed glycosidic bond dihedral angles and (*b*) the RNA parameterization with a flexible sugar ring was used.

sentation in terms of the large subunit of the ribosome (see above).

Only recently, we have begun to understand the role of conformational dynamics of RNA to achieve sufficient mechanistic and chemical complexity to undergo its diverse functions. We are convinced that analyzing RNA flexibility by constraint counting provides an efficient and reliable means to help understand functional motions and aid in structure-based efforts to find drugs targeting RNA.

APPENDIX

Influence of the glycosidic torsion angle

The nucleobase preferentially adopts an *anti*-conformation with respect to the ribose, in which the base is oriented away from the sugar. In the alternative *syn* conformation, the base is oriented over the furanose ring (55,88). Fluctuations about the glycosidic linkage do not occur independently (89). Rather, the glycosidic torsion angle varies collectively with the torsion angle fluctuations in the sugar-phosphate backbone and in the sugar rings (89). When compared to DNA, RNA structures have an enhanced rigid body behavior of the nucleoside units, which is associated with a decreased flexibility of the glycosyl linkage and the sugar moieties (85).

These findings led us to investigate the influence of the representation of the glycosidic bond in the RNA network on the outcome of the flexibility analysis. To address this question in comparison to the above results, we performed analyses on topological RNA networks in which glycosidic bonds are fixed a priori (using N_{HC}, D_{HC}, and E_{HB} as in the RNA parameterization). This resulted in rigid components that comprised at least the sugar ring and the base. After rigid cluster decomposition, FRODA simulations were performed, and atomic fluctuations were compared to conformational variabilities of NMR ensembles. The resulting R^2-values are listed in the Table S1 in Data S1. In 5 of 12 cases, R^2-values that are at least >0.1 are observed for FRODA simulations if the glycosidic bond is fixed. In contrast, in one case, worse results are obtained than if this bond is not restricted per se.

As depicted exemplarily in Fig. 6 *a* for the systems 1P5O and 1A60, this improvement, however, mainly results from less scatter in the data points describing conformational variabilities >3 Å. At the same time, the predicted mobilities are much too low, as demonstrated by correlation lines that are rather flat. Taken together, these findings suggest that RNA is predicted to be overly rigid if the glycosidic bond is fixed a priori.

Influence of the furanose ring flexibility

In the topological RNA networks, we have so far modeled the ribose as a five-membered overconstrained (i.e., rigid) ring (Eq. S3). This approach can be justified by findings that the ribose only shows restricted motions within RNA units (see above). Nevertheless, considering the ribose as rigid neglects the effect of sugar puckering (55).

To test the influence of the ribose ring modeling on the outcome of the flexibility predictions of RNA, we also analyzed topological RNA networks in which the ribose ring is considered flexible. This is achieved by extending the furanose ring by two dummy atoms, leading to a seven-membered ring with one dof (Eq. S4). As the dummy atoms do not form noncovalent interactions with their molecular environment, they do not influence the number of hydrogen bonds and hydrophobic interactions in the network (Fig. 7). Overall, this leads to a nucleotide adding seven dof to an RNA system (one due to the glycosidic bond, one due to the ribose, five due to the phosphodiester bonds).

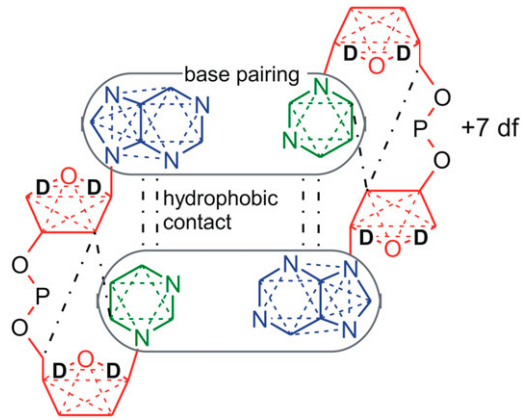

FIGURE 7 Topological network representation of a canonical A-form RNA where the ribose ring is considered flexible. This is achieved by extending the furanose ring by two dummy atoms *D*. Constraints between nearest neighbors are indicated by straight lines and constraints between next nearest neighbors (angle constraints) by dashed lines. For reasons of clarity, angle constraints are only indicated in the sugar and base scaffolds, and hydrogen bonds between bases are omitted. Hydrophobic constraints are indicated by black dashed-dotted lines. Flexible hinges are shown in red, rigid regions in green, and overconstrained regions in blue.

Not unexpectedly, if these topological RNA networks are analyzed using N_{HC}, D_{HC}, and E_{HB} as in the RNA parameterization presented above, the RNA structures are predicted to be too flexible (data not shown). We thus sought to compensate for the additional dof per nucleotide by increasing the number of stacking interactions between sequentially adjacent bases N_{HC} to two, which can result in up to four stacking interactions between adjacent base pairs in A-form RNA. (We note that, in the latter case, both base pairs form a rigid (overconstrained) cluster, because six dof of the two rigid objects face eight ($= 4 \times 2$) constraints due to the stacking interactions.) In fact, as shown in Fig. 8, part of the core of $tRNA^{ASP}$ and most of the acceptor stem is now predicted to be rigid again. However, a more detailed analysis reveals that the agreement with findings from experiment or other theoretical studies (72,75) is worse than in the RNA parameterization used above. For example, in the hinge region, only residue 8 is predicted to be flexible whereas residue 48 is predicted to be rigid. Furthermore, the anticodon loop is predicted to be largely rigid.

An explanation for these findings is given by the rigid cluster decomposition of the $tRNA^{ASP}$ structure (data not shown). Whereas rigid clusters comprise backbone regions in the case of the RNA parameterization introduced above, using seven-membered ribose moieties and $N_{HC} = 2$ results in (smaller) rigid clusters that are mostly confined to the base regions. Thus, modeling the ribose as flexible and compensating for this effect by additional stacking interactions ''shifts'' rigidity from the ''outside'' to the ''inside'' of RNA. In our opinion, such a shift is inappropriate, however. This view is also corroborated by FRODA simulation results for the structures 1P5O and 1A60 (Fig. 6 *b*). Using a seven-membered ribose ring and $N_{HC} = 2$ results for both structures in significantly worse correlations (1P5O: $R^2 = 0.35$; 1A60: $R^2 = 0.53$) than in the case of the RNA parameterization (1P5O: $R^2 = 0.52$; 1A60: $R^2 = 0.62$).

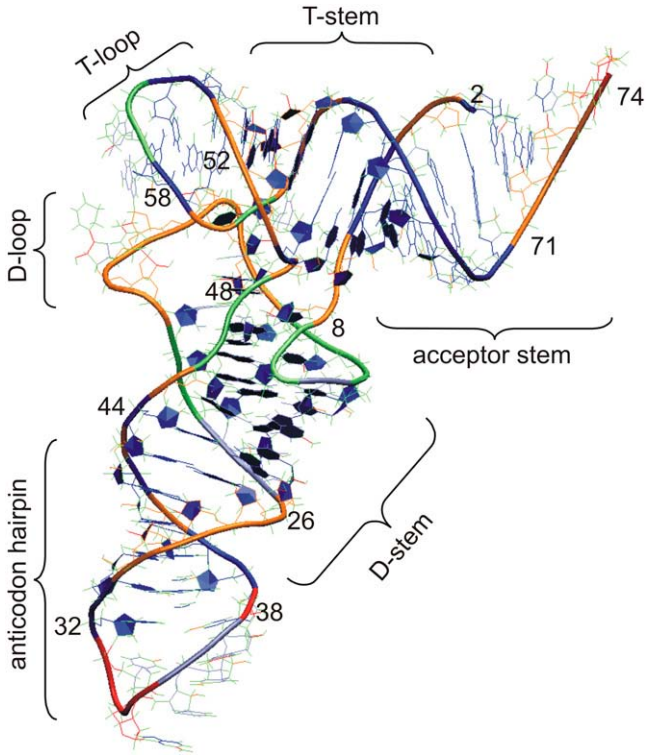

FIGURE 8 Flexibility prediction for the $tRNA^{ASP}$ structure (PDB code: 2TRA). Color-coded representation of flexibility indices (Eq. 1) obtained by a flexibility analysis of the $tRNA^{ASP}$ structure using the RNA parameterization, seven-membered ribose moieties, and $N_{HC} = 2$. Overconstrained regions are indicated by blue colors (*dark blue*, $f_i \leq -0.2$; *light blue*, $-0.2 < f_i < 0.0$), rigid regions are represented in green color ($f_i = 0.0$), and flexible regions are shown in red colors (*orange*, $0.0 < f_i < 0.2$; *red*, $f_i \geq 0.2$). The coloring of the backbone is according to the flexibility indices of the phosphorus atoms. The core region (as defined by Auffinger et al. (67)) is highlighted with filled sugar and base scaffolds.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

We are grateful to M. F. Thorpe, W. Whitely, A. J. Rader, and D. Metzler for insightful discussions.

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 579, ''RNA-ligand interactions'') and startup funds from J. W. Goethe-University, Frankfurt. Financial support was also provided by the Hessian Science Program and the Frankfurt International Graduate School for Science (S.F.). The RNA parameterization will be made available in the ambpdb program of the Amber 10 suite of packages (http://amber.scripps.edu) and the FIRST program (http://flexweb.asu.edu).

REFERENCES

1. Frank, J., and R. Agrawal. 2000. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature*. 406:318–322.
2. Valle, M., A. Zavialov, W. Li, S. Stagg, J. Sengupta, R. Nielsen, P. Nissen, S. Harvey, M. Ehrenberg, and J. Frank. 2003. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 10:899–906.
3. Mandal, M., and R. Breaker. 2004. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* 5:451–463.
4. Murray, J. B., D. P. Terwey, L. Maloney, A. Karpeisky, N. Usman, L. Beigelman, and W. G. Scott. 1998. The structural basis of hammerhead ribozyme self-cleavage. *Cell*. 92:665–673.
5. Puglisi, J. D., R. Tan, B. J. Calnan, A. D. Frankel, and J. R. Williamson. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*. 257:76–80.
6. Fourmy, D., S. Yoshizawa, and J. D. Puglisi. 1998. Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA. *J. Mol. Biol.* 277:333–345.
7. Nifosi, R., C. M. Reyes, and P. A. Kollman. 2000. Molecular dynamics studies of the HIV-1 TAR and its complex with argininamide. *Nucleic Acids Res.* 28:4944–4955.
8. Hermann, T., and E. Westhof. 1998. RNA as a drug target: chemical, modelling, and evolutionary tools. *Curr. Opin. Biotechnol.* 9:66–73.
9. Wilson, W. D., and K. Li. 2000. Targeting RNA with small molecules. *Curr. Med. Chem.* 7:73–98.
10. Hermann, T. 2005. Drugs targeting the ribosome. *Curr. Opin. Struct. Biol.* 15:355–366.
11. Hermann, T. 2000. Strategies for the design of drugs targeting RNA and RNA-protein complexes. *Angew. Chem. Int. Ed.* 39:1890–1904.
12. Faber, C., H. Sticht, K. Schweimer, and P. Rösch. 2000. Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* 275:20660–20666.
13. Teague, S. J. 2003. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* 2:527–541.
14. Duchardt, E., and H. Schwalbe. 2005. Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR 13C relaxation. *J. Biomol. NMR*. 32:295–308.
15. Shajani, Z., and G. Varani. 2005. 13C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.* 349:699–715.

16. Al-Hashimi, H. M. 2005. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *ChemBioChem*. 6:1506–1519.

17. Perez, A., A. Noy, F. Lankas, F. J. Luque, and M. Orozco. 2004. The relative flexibility of B-DNA and A-RNA duplexes: database analysis. *Nucleic Acids Res*. 32:6144–6151.

18. Jacobs, D. J., A. J. Rader, L. A. Kuhn, and M. F. Thorpe. 2001. Protein flexibility predictions using graph theory. *Proteins*. 44:150–165.

19. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol*. 9:646–652.

20. Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*. 13:373–380.

21. Van Wynsberghe, A. W., and Q. Cui. 2005. Comparison of mode analyses at different resolutions applied to nucleic acid systems. *Biophys. J*. 89:2939–2949.

22. Hespenheide, B. M., D. J. Jacobs, and M. F. Thorpe. 2004. Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus. *J. Phys. Condens. Matter*. 16:5055–5064.

23. Wang, Y., A. J. Rader, I. Bahar, and R. Jernigan. 2004. Global ribosome motions revealed with elastic network model. *J. Struct. Biol*. 147:302–314.

24. Jacobs, D. J., and M. F. Thorpe. 1995. Generic rigidity percolation: the pebble game. *Phys. Rev. Lett*. 75:4051–4054.

25. Jacobs, D., and B. Hendrickson. 1997. An algorithm for two-dimensional rigidity percolation: the pebble game. *J. Comput. Phys*. 137: 346–365.

26. Rader, A. J., B. M. Hespenheide, L. A. Kuhn, and M. F. Thorpe. 2002. Protein unfolding: rigidity lost. *Proc. Natl. Acad. Sci. USA*. 99:3540–3545.

27. Hespenheide, B. M., A. J. Rader, M. F. Thorpe, and L. A. Kuhn. 2002. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Graph. Model*. 21:195–207.

28. Rader, A. J., and I. Bahar. 2004. Folding core predictions from network models of proteins. *Polymer (Guildf.)*. 45:659–668.

29. Gohlke, H., L. A. Kuhn, and D. A. Case. 2004. Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins*. 56:322–337.

30. Lei, M., M. Zavodszky, L. Kuhn, and M. F. Thorpe. 2004. Sampling protein conformations and pathways. *J. Comput. Chem*. 25:1133–1148.

31. Wells, S., S. Menor, B. Hespenheide, and M. F. Thorpe. 2005. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol*. 2:127–136.

32. Ahmed, A., and H. Gohlke. 2006. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins*. 63:1038–1051.

33. Hagerman, P. J. 1997. Flexibility of RNA. *Annu. Rev. Biophys. Biomol. Struct*. 26:139–156.

34. Laman, G. 1979. On graphs and rigidity of plane skeletal structures. *J. Eng. Math*. 1970:331–340.

35. Tay, T.-S., and W. Whiteley. 1984. Recent advances in the generic rigidity of structures. *Structural Topoology*. 9:31–38.

36. Whiteley, W. 2005. Counting out to the flexibility of molecules. *Phys. Biol*. 2:116–126.

37. Thorpe, M. F., M. Lei, A. J. Rader, D. J. Jacobs, and L. A. Kuhn. 2001. Protein flexibility and dynamics using constraint theory. *J. Mol. Graph. Model*. 19:60–69.

38. Jacobs, D. J. 1998. Generic rigidity in three-dimensional bond-bending networks. *J. Phys. Math. Gen*. 31:6653–6668.

39. de Groot, B. L., D. M. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H. J. Berendsen. 1997. Prediction of protein conformational freedom from distance constraints. *Proteins*. 29:240–251.

40. Crippen, G. M. 1977. A novel approach to calculation of conformation: distance geometry. *J. Comput. Phys*. 24:96–107.

41. Isralewitz, B., J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten. 2001. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model*. 19:13–25.

42. Jucker, F. M., H. A. Heus, P. F. Yip, E. H. Moors, and A. Pardi. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol*. 264:968–980.

43. Allain, F. H., and G. Varani. 1995. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol*. 250:333–353.

44. Westhof, E., P. Dumas, and D. Moras. 1988. Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr. A*. 44:112–123.

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

46. Bahar, I., A. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des*. 2:173–181.

47. Tamhane, A. C., and D. D. Dunlop. 2000. Statistics and Data Analysis: From Elementary to Intermediate. Prentice Hall, Upper Saddle River, NJ.

48. Yang, L. W., A. J. Rader, X. Liu, C. J. Jursa, S. C. Chen, H. A. Karimi, and I. Bahar. 2006. oGNM: online computation of structural dynamics using the Gaussian Network Model. *Nucleic Acids Res*. 34:W24–W31.

49. Bahar, I., and R. L. Jernigan. 1998. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms. *J. Mol. Biol*. 281:871–884.

50. Ulrich, E., H. Akutsu, J. Doreleijers, Y. Harano, Y. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. Schulte, D. Tolmie, R. Kent Wenger, H. Yao, and J. Markley. 2008. BioMagResBank. *Nucleic Acids Res*. 36(Database issue):D401–D408.

51. Tropp, J. 1980. Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: the effect of fluctuating internuclear distances. *J. Chem. Phys*. 72:6035–6043.

52. Zagrovic, B., and W. van Gunsteren. 2006. Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us? *Proteins*. 63:210–218.

53. Whiteley, W. 1996. Some matroids from discrete applied geometry. *Contemp. Math*. 197:171–311.

54. Boolchand, P., and M. F. Thorpe. 1994. Glass-forming tendency, percolation of rigidity, and onefold-coordinated atoms in covalent networks. *Phys. Rev. B*. 50:10366–10368.

55. Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York.

56. Foloppe, N., L. Nilsson, and A. D. MacKerell. 2001. Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers*. 61:61–76.

57. Réblová, K., F. Lankaš, F. Rázga, M. V. Krasovska, J. Koča, and J. Šponer. 2006. Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers*. 82:504–520.

58. Auffinger, P., and E. Westhof. 1997. Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol*. 274:54–63.

59. Kebbekus, P., D. E. Draper, and P. Hagerman. 1995. Persistence length of RNA. *Biochemistry*. 34:4354–4357.

60. Abels, J. A., F. Moreno-Herrero, T. van der Heijden, C. Dekker, and N. H. Dekker. 2005. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J*. 88:2737–2744.

61. Hyeon, C., R. I. Dima, and D. Thirumalai. 2006. Size, shape, and flexibility of RNA structures. *J. Chem. Phys*. 125:194905.

62. Varani, G. 1995. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* 24:379–404.

63. Antao, V. P., S. Y. Lai, and I. Tinoco. 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* 19:5901–5905.

64. Villa, A., and G. Stock. 2006. What NMR relaxation can tell us about the internal motion of an RNA hairpin: a molecular dynamics simulation study. *J. Chem. Theory Comput.* 2:1228–1236.

65. Gohlke, H., and M. F. Thorpe. 2006. A Natural coarse graining for simulating large biomolecular motion. *Biophys. J.* 91:2115–2120.

66. Nakamura, S., and J. Doi. 1994. Dynamics of transfer RNAs analyzed by normal mode calculation. *Nucleic Acids Res.* 22:514–521.

67. Auffinger, P., S. Louise-May, and E. Westhof. 1999. Molecular dynamics simulations of solvated yeast tRNA(Asp). *Biophys. J.* 76:50–64.

68. Auffinger, P., and E. Westhof. 1997. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J. Mol. Biol.* 269:326–341.

69. Moras, D., A. C. Dock, P. Dumas, E. Westhof, P. Romby, J. P. Ebel, and R. Giege. 1986. Anticodon-anticodon interaction induces conformational changes in tRNA: yeast tRNAAsp, a model for tRNA-mRNA recognition. *Proc. Natl. Acad. Sci. USA.* 83:932–936.

70. Amano, M., and M. Kawakami. 1992. Assignment of the magnetic resonances of the imino protons and methyl protons of Bombyx mori tRNA(GlyGCC) and the effect of ion binding on its structure. *Eur. J. Biochem.* 210:671–681.

71. Jacka, A., J. E. Landnera, D. Rhodesa, R. S. Browna, and A. Kluga. 1977. A crystallographic study of metal-binding to yeast phenylalanine transfer RNA. *J. Mol. Biol.* 111:315–328.

72. Olson, T., M. J. Fournier, K. H. Langley, and N. C. Ford. 1976. Detection of a major conformational change in transfer ribonucleic acid by laser light scattering. *J. Mol. Biol.* 102:193–203.

73. Robertus, J. D., J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. C. Clark, and A. Klug. 1974. Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature.* 250:546–551.

74. Yarus, M., M. Valle, and J. Frank. 2003. A twisted tRNA intermediate sets the threshold for decoding. *RNA.* 9:384–385.

75. Sanbonmatsu, K. Y., S. Joseph, and C. S. Tung. 2005. Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA.* 102:15854–15859.

76. Hermann, T., and D. J. Patel. 1999. Stitching together RNA tertiary architectures. *J. Mol. Biol.* 294:829–849.

77. Sternberg, M. J. E., D. E. P. Grace, and D. C. Phillips. 1979. Dynamic information from protein crystallography. An analysis of temperature factors from refinement of the hen egg-white lysozyme structure. *J. Mol. Biol.* 130:231–253.

78. Simonson, T., and A. T. Brunger. 1992. Thermodynamics of protein-peptide interactions in the ribonuclease-S system studied by molecular dynamics and free energy calculations. *Biochemistry.* 31:8661–8674.

79. Garcia, A. E., J. A. Krumhansl, and H. Frauenfelder. 1997. Variations on a theme by Debye and Waller: from simple crystals to proteins. *Proteins.* 29:153–160.

80. Lu, M., B. Poon, and J. Ma. 2006. A new method for coarse-grained elastic normal-mode analysis. *J. Chem. Theory Comput.* 2:464–471.

81. Dahiyat, B. I., D. B. Gordon, and S. L. Mayo. 1997. Automated design of the surface positions of protein helices. *Protein Sci.* 6:1333–1337.

82. Mills, J. B., and P. J. Hagerman. 2004. Origin of the intrinsic rigidity of DNA. *Nucleic Acids Res.* 32:4055–4059.

83. Ornstein, R. L., R. Rein, D. L. Breen, and R. D. Macelroy. 1978. An optimized potential function for the calculation of nucleic acid interaction energies. I. Base stacking. *Biopolymers.* 17:2341–2360.

84. Gralla, J., and D. M. Crothers. 1973. Free energy of imperfect nucleic acid helices: III. Small internal loops resulting from mismatches. *J. Mol. Biol.* 78:301–319.

85. Pan, Y., and A. D. MacKerell. 2003. Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Res.* 31:7131–7140.

86. Noy, A., A. Pérez, F. Lankas, F. J. Luque, and M. Orozco. 2004. Relative flexibility of DNA and RNA: a molecular dynamics study. *J. Mol. Biol.* 343:627–638.

87. Perederina, A., N. Nevskaya, O. Nikonov, A. Nikulin, P. Dumas, M. Yao, I. Tanaka, M. Garber, G. Gongadze, and S. Nikonov. 2002. Detailed analysis of RNA-protein interactions within the bacterial ribosomal protein L5/5S rRNA complex. *RNA.* 8:1548–1557.

88. Foloppe, N., B. Hartmann, L. Nilsson, and A. D. MacKerell. 2002. Intrinsic conformational energetics associated with the glycosyl torsion in DNA: a quantum mechanical study. *Biophys. J.* 82:1554–1569.

89. Ravindranathan, S., C.-H. Kim, and G. Bodenhausen. 2003. Cross correlations between ^{13}C-^{1}H dipolar interactions and ^{15}N chemical shift anisotropy in nucleic acids. *J. Biomol. NMR.* 27:365–375.

90. Dallas, A., and P. B. Moore. 1997. The loop E-loop D region of *Escherichia coli* 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure.* 5:1639–1653.

91. Kolk, M. H., M. van der Graaf, S. S. Wijmenga, C. W. Pleij, H. A. Heus, and C. W. Hilbers. 1998. NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science.* 280:434–438.

92. Schmitz, U., S. Behrens, D. M. Freymann, R. J. Keenan, P. Lukavsky, P. Walter, and T. L. James. 1999. Structure of the phylogenetically most conserved domain of SRP RNA. *RNA.* 5:1419–1429.

93. Vallurupalli, P., and P. B. Moore. 2003. The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts. *J. Mol. Biol.* 325:843–856.

94. Lukavsky, P. J., I. Kim, G. A. Otto, and J. D. Puglisi. 2003. Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.* 10:1033–1038.

95. D'Souza, V., A. Dey, D. Habib, and M. F. Summers. 2004. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* 337:427–442.

96. Theimer, C. A., C. A. Blois, and J. Feigon. 2005. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell.* 17:671–682.

97. Staple, D. W., and S. E. Butcher. 2005. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J. Mol. Biol.* 349:1011–1023.

98. Gaudin, C., M. H. Mazauric, M. Traikia, E. Guittet, S. Yoshizawa, and D. Fourmy. 2005. Structure of the RNA signal essential for translational frameshifting in HIV-1. *J. Mol. Biol.* 349:1024–1035.

99. Davis, J. H., M. Tonelli, L. G. Scott, L. Jaeger, J. R. Williamson, and S. E. Butcher. 2005. RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex. *J. Mol. Biol.* 351:371–382.

100. Chen, Y., J. Fender, J. Legassie, M. Jarstfer, T. Bryan, and G. Varani. 2006. Structure of stem-loop IV of Tetrahymena telomerase RNA. *EMBO J.* 25:3156–3166.