# UC Irvine
## UC Irvine Previously Published Works

**Title**

Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition

**Permalink**

**Journal**

**ISSN**

**Authors**

Frank, Aaron T.
Stelzer, Andrew C.
Al-Hashimi, Hashim M.
et al.

**Publication Date**

**Copyright Information**

Peer reviewed

# Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition

**Aaron T. Frank[1], Andrew C. Stelzer[2], Hashim M. Al-Hashimi[2],* and Ioan Andricioaei[1]**

[1]Department of Chemistry, University of California Irvine, 1102 Natural Sciences 2, Irvine, CA 92697 and
[2]Department of Chemistry & Biophysics, The University of Michigan, Ann Arbor, MI 48109, USA

## ABSTRACT

**We describe a strategy for constructing atomic resolution dynamical ensembles of RNA molecules, spanning up to millisecond timescales, that combines molecular dynamics (MD) simulations with NMR residual dipolar couplings (RDC) measured in elongated RNA. The ensembles are generated via a Monte Carlo procedure by selecting snap-shot from an MD trajectory that reproduce experimentally measured RDCs. Using this approach, we construct ensembles for two variants of the trans-activation response element (TAR) containing three (HIV-1) and two (HIV-2) nucleotide bulges. The HIV-1 TAR ensemble reveals significant mobility in bulge residues C24 and U25 and to a lesser extent U23 and neighboring helical residue A22 that give rise to large amplitude spatially correlated twisting and bending helical motions. Omission of bulge residue C24 in HIV-2 TAR leads to a significant reduction in both the local mobility in and around the bulge and amplitude of inter-helical bending motions. In contrast, twisting motions of the helices remain comparable in amplitude to HIV-1 TAR and spatial correlations between them increase significantly. Comparison of the HIV-1 TAR dynamical ensemble and ligand bound TAR conformations reveals that several features of the binding pocket and global conformation are dynamically preformed, providing support for adaptive recognition via a 'conformational selection' type mechanism.**

## INTRODUCTION

Many non-coding RNA molecules (ncRNAs) perform their biological functions by undergoing large conformational changes in response to specific cellular signals including the recognition of proteins, nucleic acids, metal ions, metabolites, vitamins, changes in temperature and even RNA biosynthesis itself (1–5). These conformational transitions guide RNA folding during co-transcriptional folding; provide the molecular basis for sensing and signaling transactions that allow riboswitches to regulate gene expression in response to changes in environmental conditions; allow ribozymes to dynamically meet the diverse structural requirements associated with their multi-step catalytic cycles; and enable complex ribonucleoproteins to assemble in a hierarchical and sequentially ordered manner.

Although it is clear that many ncRNAs undergo large changes in structure in order to carry out their function, the mechanism by which these conformational transitions occur remains poorly understood. A central question is whether cellular factors such as proteins and ligands act catalytically to induce the RNA conformational change via 'induced fit', or they select and bind distinct RNA conformers from a pre-existing dynamical ensemble via 'conformational selection' (6–8). Insights into such mechanistic questions have been impeded by lack of biophysical techniques that allow the 3D visualization of intrinsic RNA dynamics over biologically relevant timescales. The atomic resolution characterization of dynamics in complex biomolecules is currently a major challenge in structural biology and biophysics. NMR spectroscopy is one of the most powerful techniques for characterizing dynamics uniquely providing comprehensive information

---

regarding the amplitude, timescale and—in favorable cases—direction of motions with site-specific resolution (9–11). However, even with abundant measurements that can be made with the use of NMR, the total number of observables still pale in comparison to the total number of parameters needed to fully describe dynamics. Molecular dynamic (MD) simulations provide an all-atom description of dynamics; however, force fields remain to be thoroughly validated particularly for nucleic acids and simulation timescales remain limited to ∼100 ns (12,13).

Because they are complementary on the spatial and temporal scales, the limitations inherent to NMR and MD could in principle be overcome by combining the two techniques; MD can fill the shortage in NMR data and NMR can provide a means for validating and potentially correcting force fields and accelerate MD conformational sampling to millisecond timescales. Several studies have emerged in which MD and NMR are used in concert in studies of nucleic acid dynamics (5,14–17). The two techniques can also be integrated to yield a unified view of structural dynamics. While such combined NMR/MD approaches have successfully been used in studies of protein dynamics (18–21), extension to RNA can prove very difficult. This is because unlike globular proteins, overall motions in highly flexible RNAs can be strongly coupled to internal motions making it difficult, if not impossible, to predict NMR data from an MD trajectory (22–24).

Recently, a domain-elongation strategy was introduced for decoupling internal and overall motions in RNA (25,26). In this approach, a target helix is elongated so that the overall motion is slowed down relative to internal motions and rendered less sensitive to internal fluctuations in other parts of the molecule. The elongation also simplifies analysis of NMR spin relaxation (27) and residual dipolar coupling (RDC) (28,29) data because it predefines the overall diffusion or alignment tensor to be axially symmetric with principal axis oriented nearly parallel to the elongated helix axis. As we show in what follows, this makes elongated RNAs ideally suited for computing NMR observables from a given MD trajectory of a corresponding non-elongated RNA in which snapshots are aligned by superimposing the reference elongated helix.

Here, we report implementation of a general strategy that combines MD simulations and NMR RDCs measured in elongated RNAs for constructing atomic resolution dynamical ensembles with timescale sensitivity extending up to milliseconds. The dynamical ensemble of the HIV-1 and HIV-2 transactivation response element (TAR) (Figure 1a) constructed in this manner provides new insights into the bulge-length dependence of RNA dynamics and the mechanism of conformational adaptation upon target recognition.

## MATERIALS AND METHODS

### Molecular dynamics simulations

Simulations of wild-type HIV-1 and HIV-2 TAR were performed using the CHARMM package (30) with force field parameter set 27 (31). Model 3 of the unbound NMR structure of HIV-1 TAR (PDB 1ANR) (32) was used as starting coordinates for simulations of HIV-1 TAR. For HIV-2 TAR, starting coordinates were obtained by removing argininamide from model 1 of the argininamide-TAR NMR structure (PDB 1AKX) (33). The RNA was charge-neutralized using sodium counter ions and solvated in a 33–35 Å sphere of TIP3 water (34). A spherical boundary potential was applied to maintain the density of water around the RNA site (35). The system was minimized and heated to 300 K, while harmonically constraining the heavy atoms of the RNA with a force constant of 62 kcal/mol/Å for 100 ps, after which constraints were removed and the system equilibrated for 1 ns. A Nosé–Hoover thermostat (36,37) was used to maintain a constant temperature of 300 K throughout the simulation, with a 1 fs time-step and a coupling constant of 50 ps$^{-1}$. Fifty distinct trajectories were initiated from this equilibrated structure, by assigning different initial velocities. It has previously been demonstrated that this technique can be used to enhance conformational sampling relative to a single trajectory of the same total duration (38,39). The first 0.5 ns of each trajectory was discarded and next 1.6 ns used for analysis. Conformations from each of the 50, 1.6 ns, trajectories were pooled to give a total effective simulation time of 80 ns. These 80 000 conformations were used as a structural pool for the 'selection' phase of the select-and-sample strategy (SAS, see below). The same protocol used for HIV-1 TAR, was used to prepare, equilibrate and produce the MD trajectories for HIV-2 TAR, except that a 33 Å sphere of TIP3 water was used to solvate the system.

### RDC-based SAS

In the original implementation, Chen and coworkers used $S^2$ NH spin relaxation order parameters to select structural ensembles for a series of protein systems (19). Here, we adapt this SAS approach to allow use of RDC data which probe the orientational dynamics of individual bond vectors over timescales extending up to milliseconds. In the SAS approach, one generates a set of conformations for the system of interest, in which we use MD to generate such structures. One then searches for an $N$-membered subset of structures that minimizes a cost function. Specifically, an $N$-membered subset of structures is randomly selected from a total pool of $M$ structures and an initial $\chi^2$-value is evaluated using Equation 1. Next, one of the $N$-membered structures is randomly chosen and replaced, by a random structure from the remaining $M–N$ conformational pool. The 'move' from step $k$ to $k + 1$ is then accepted if $\chi^2_{(k+1)} < \chi^2_{(k)}$. If $\chi^2_{(k+1)} > \chi^2_{(k)}$, the move is accepted with a probability $P = \exp((\chi^2_{(k)} - \chi^2_{(k+1)})/T_{eff})$, where $T_{eff}$ is an effective temperature that is linearly decreased in a simulated annealing scheme. The cost function is

$$\chi^2 = L^{-1} \sum_i^L (D_{ij}{}^{calc} - D_{ij}{}^{meas})^2 \qquad \textbf{1}$$
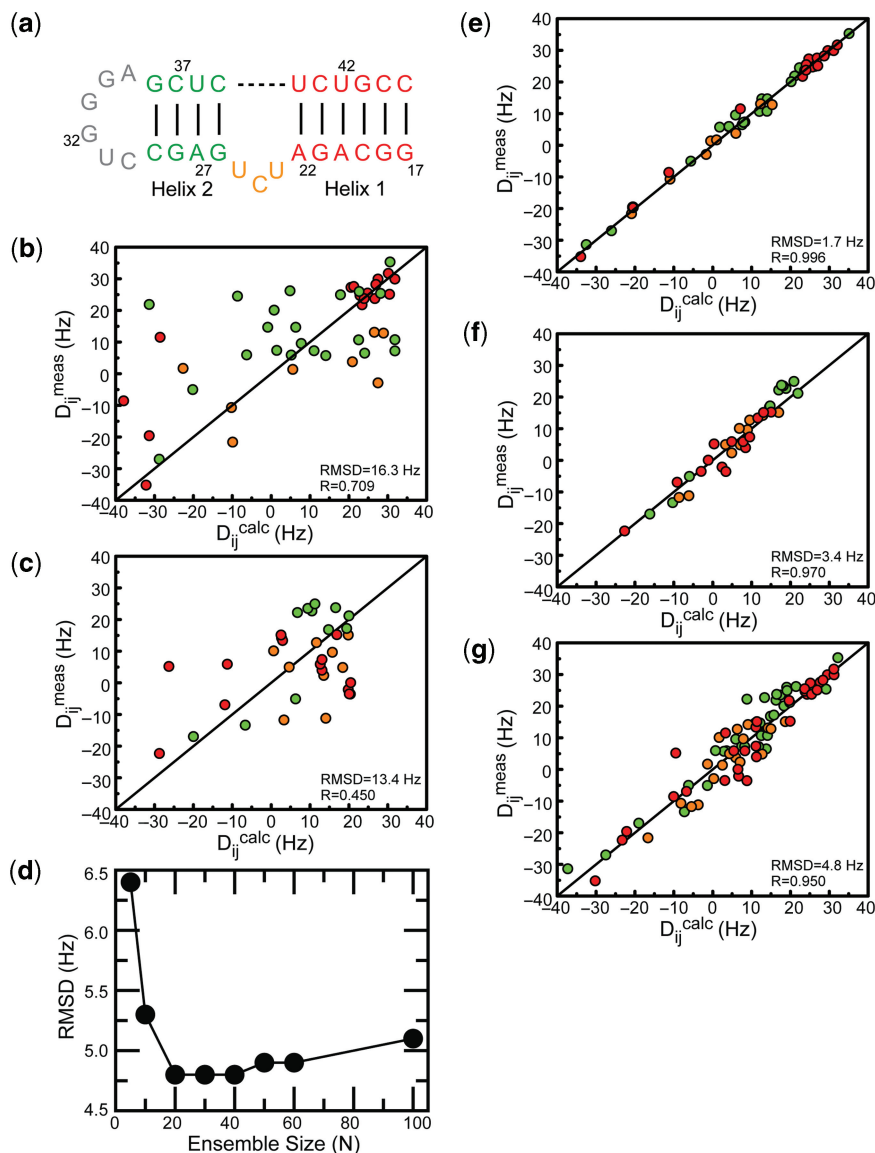
**Figure 1.** SAS analysis of HIV-1 E-TAR RDCs. (**a**) Secondary structure of HIV-1 TAR with helix I highlighted in red, helix II in green and tri-nucleotide bulge in orange. HIV-2 TAR lacks bulge residue C24. (**b** and **c**) Plots of experimental RDCs versus values computed from the 80 ns MD trajectory for (b) EI-TAR and (c) EII-TAR. Data for helix I, helix II, and bulge, are shown in red, green and orange, respectively. Also shown is the root-mean-square-deviation (rmsd) and correlation coefficient (*R*). (**d**) RMSD (Hz) between calculated and experimental RDCs as a function of *N*, following a SAS analysis using both EI and EII-TAR HIV-1 RDCs. (**e–g**) Plots of experimental RDCs versus values calculated from the *N* = 20 SAS ensemble using (e) EI-TAR, (f) EII-TAR and (g) EI-TAR + EII-TAR RDCs.

where $D_{ij}^{cal}$ and $D_{ij}^{exp}$ are the calculated and measured RDCs, respectively, $L$ is the total number bond vectors, and $D_{ij}^{cal}$ is calculated using,

$$D_{ij}^{cal} = \frac{\mu_0 \gamma_i \gamma_j h}{8\pi^3 \langle r_{ij}^3 \rangle} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \qquad 2$$

where $\gamma_i$ is the gyromagnetic ratio of the *i*-th nucleus, $r_{ij}$ is the bond length, $\theta$ is the angle between the inter-nuclear bond vector and the external magnetic field and the angular brackets denote a time-average over all sampled orientations. The time-averaged angular term can be expressed in terms of the time-independent orientation of an

internuclear vector relative to an arbitrary frame and the five order tensor elements ($S_{kl}$), (40,41).

$$\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle = \sum_{kl=xyz} S_{kl} \cos(\alpha_k) \cos(\alpha_l) \qquad 3$$

where $\alpha_n$ is the angle between the internuclear vector and the *n*-th axis of the arbitrary frame. Equation 3 assumes that internal motions do not affect overall alignment of the molecule. This assumption can break down in RNA because collective motions of helical domains can lead to large changes in the overall shape and thus overall alignment of the molecule (22,23,42). This can make it

impossible to accurately compute $D_{ij}{}^{cal}$ for a given ensemble and thus to use RDCs in selecting conformers form an MD trajectory. Domain-elongation allows one to overcome this problem by rendering the overall shape of the molecule far less sensitive to internal motions (25,26). Elongation also predefines the overall order tensor to be axially determined with principal direction oriented approximately parallel to axis of the elongated helix (26). This overall order tensor can be conveniently determined experimentally using RDCs measured in the elongated helix (26).

### TAR ensembles by RDC-based SAS

The SAS RDC approach was implemented using in-house programs written in C++ available from the authors upon request. The previously reported (26) 47 and 35 one-bond base and sugar C-H RDCs measured in EI-TAR and EII-TAR, respectively were used independently or in combination to select an HIV-1 TAR structural ensemble from a pool of 80 000 conformers derived from the combined 80 ns MD trajectory. For HIV-2 TAR, 35 RDCs measured in EI-TAR were used. Note that although RDCs were measured on a TAR construct in which the apical loop was replaced by a UUCG loop, a detailed NMR study recently showed that this apical loop replacement does not affect inter-helical motions or local motions at the TAR bulge (43). When used in combination, both the EI-TAR and EII-TAR sets of RDCs were used ($L = 47 + 35 = 82$) in computing the $\chi^2$ in Equation 1. In all cases, the overall alignment tensor was assumed to be axially symmetric ($\eta = 0$) with principal direction ($S_{zz}$) oriented parallel to the elongated helix axis as computed using the program CURVES (44,45). The magnitude of $S_{zz}$ was obtained from an order tensor analysis as described previously (26).

Each RDC–SAS run was initiated from $N$ randomly selected conformers. A Monte Carlo (MC) simulated annealing scheme was then used to minimize the cost in Equation 1 as described above. Simulations were started at a high-effective temperature, where the MC acceptance probability was high ($\sim$0.99), and slowly decreased until the MC acceptance probability was $\sim$10$^{-5}$. At a given effective temperature $10^5$ MC steps were carried out. The effective temperature was then decreased, with $T_{i+1} = 0.9T_i$. The same protocol was used for HIV-2 TAR, except that only 35 RDCs measured in the helix I elongated sample were used in the SAS analysis. Here, there was less motivation to acquire an additional set of helix II elongated RDCs given the near linear alignment of the two helices which renders the two sets of data degenerate.

### Analysis of the SAS-derived TAR ensemble

The base angles, buckle ($\kappa$), opening ($\sigma$), propeller twist ($\omega$) and twist ($\Omega$), were calculated for eight non-terminal base pairs for each of the SAS selected TAR conformers using the program 3DNA (46). The inter-helical Euler angles $\alpha_h$, $\beta_h$ and $\gamma_h$ were also computed for each member of the ensemble, using the lower helix as a reference, as previously described (47). Note that degenerate sets of Euler angles reflect the same inter-helical orientation (47) and care had to be taken in selecting among degenerate angles to avoid non-realistic correlations.

## RESULTS

### SAS analysis of HIV-1 E-TAR RDCs

In Figure 1, we compare the RDCs measured previously (26) in EI-TAR and EII-TAR with those predicted based on the combined $50 \times 1.6 = 80$ ns MD trajectory. Though some correlation is observed between the measured and predicted RDCs for both EI-TAR (Figure 1b) and EII-TAR (Figure 1c), the deviations (RMSD ranging between 13 and 16 Hz) remains substantially larger than the estimated RDC measurement uncertainty ($\sim$4 Hz). The MD trajectory does not systematically under- or overestimate the RDCs measured throughout the RNA, indicating that it does not significantly over- or underestimate the amplitude of motions present. This is the case even though the RDC timescale sensitivity to motions (less than millisecond) is greater than that of MD ($\sim$80 ns). Thus, it appears that the motions in TAR saturate at the nanosecond timescales consistent with relaxation dispersion studies that provide no evidence for μs-ms motions in the stem-bulge-stem element of TAR (43).

We examined if SAS could be used to pull out subensembles from the MD trajectory that satisfy the measured RDCs. We first conducted a series of SAS runs with various ensemble sizes of $N = 5, 10, 20, 30, 40, 50$ and 100. The root-mean-square-derivation (RMSD) between measured and predicted RDCs when combining the EI-TAR and EII-TAR RDCs is shown in Figure 1d as a function of $N$. Increasing the ensemble size beyond $N = 20$ did not lead to significant improvements in the fit and in fact a deterioration was observed for $N > 20$. This is likely due to sampling problems during the MC simulating annealing minimization of the cost function as the number of possible combinations increases steeply with $N$ and the cost function exhibits, due to the frustration stemming from the underdetermined nature of the problem, a vast number local minima. An ensemble size of $N = 20$ was used in all subsequent SAS runs.

The RDCs calculated using a 20-member ensemble selected using the SAS approach exhibit a markedly improved fit to the EI-TAR (Figure 1e), EII-TAR (Figure 1f) and EI-TAR + EII-TAR (Figure 1g) RDCs as compared to those calculated from the entire MD trajectory. Importantly, sub-ensembles can be determined that simultaneously reproduce the EI-TAR and EII-TAR RDCs with an RMSD (4.8 Hz) that is comparable to the estimated experimental RDC uncertainty ($\sim$4 Hz) (Figure 1g). The SAS selected conformers also did not lead to any steric collisions with the elongated helices for both EI-TAR and EII-TAR despite the fact that the helices were not actually elongated in the MD simulations (data not shown).

To examine the uniqueness of the selected TAR ensemble, we carried out 100 independent $N = 20$ SAS runs. Here, a maximum of 2000 unique snapshots can be selected. However, repeated runs resulted in selection of

a narrow set of similar conformations. The total number of unique snapshots selected by the RDC-SAS optimization algorithm was 422, 48 and 121 for EI-TAR, EII-TAR and EI-TAR + EII-TAR RDCs, respectively. Thus, the RDC data favors selection of specific conformations from the available pool. As expected, the conformers selected did vary when changing the value $N$. However, the overall distribution of conformations remained similar as shown for example for the inter-helical orientation in Supplementary Figure S1.

## Local motions in the helices and bulge

In Figure 2a, we plot (in black) the average and standard deviation for various base angles computed for the 121 unique TAR conformers obtained from 100 $N = 20$ SAS runs. For comparison, the mean value and SD for idealized A-form helical geometry computed from a statistical comparison of high-resolution X-ray structures is shown in red (48). For the majority of the residues, very good agreement is observed between the SAS ensemble and canonical values. Large deviations are however observed for the junctional A22-U40 base-pair, which in the MD simulation frequently deviates from a hydrogen bonded alignment. This is in excellent agreement with previous NMR data showing that while the junctional G26-C39 base-pair forms a detectable hydrogen bond, the A22-U40 base-pair is flexible and does not form the expected base-pair in HIV-1 TAR (49–51).

The bulge residues exhibit different levels of motions. As shown in Figure 2b, the highly conserved U23 bulge stacks onto A22 in the majority of the conformations, consistent with observation of NOE connectivity between A22 and U23. The two residues undergo limited motions consistent with previous $^{13}$C relaxation studies of dynamics in elongated HIV-1 TAR (52). Interestingly, select conformations exist in which U23 adopts a looped out conformation as observed in several ligand bound TAR structures (PDB ID# 1QD3, 1UTS, and 397D). In contrast, residues C24 and U25 predominantly exist in a looped out conformation and are significantly more flexible, again in agreement with the previous carbon relaxation studies (52).

## Collective motions of helices

The relative orientation of two helices $i$ and $j$ can be defined using three inter-helical Euler angles that describe the twist angle $\alpha_h$ about helix $i$, the inter-helix bend angle $\beta_h$, and the twist angle $\gamma_h$ about helix $j$ (26,47). We computed these three inter-helical Euler angles for the 121 SAS selected TAR conformers and compared them to angles obtained for the entire MD trajectory. As shown in Figure 3a, the MD trajectory spans a large range of inter-helical angles. The SAS conformers are widely distributed across the MD trajectory (Figure 3a). For both the MD trajectory and SAS ensembles, significant correlation is observed between the three inter-helical angles, particularly between the twist angles $\alpha_h$ and $\gamma_h$. Such spatial correlations were reported previously based on an a three-state rigid-body refinement of the TAR inter-helical orientation (26). As shown in Figure 3b, the conformations obtained by SAS sample orientations that are in very good agreement with those obtained by a previous three-state analysis of the RDCs (26). The only significant deviations are observed in the $\alpha_h$-$\gamma_h$ plane for the near
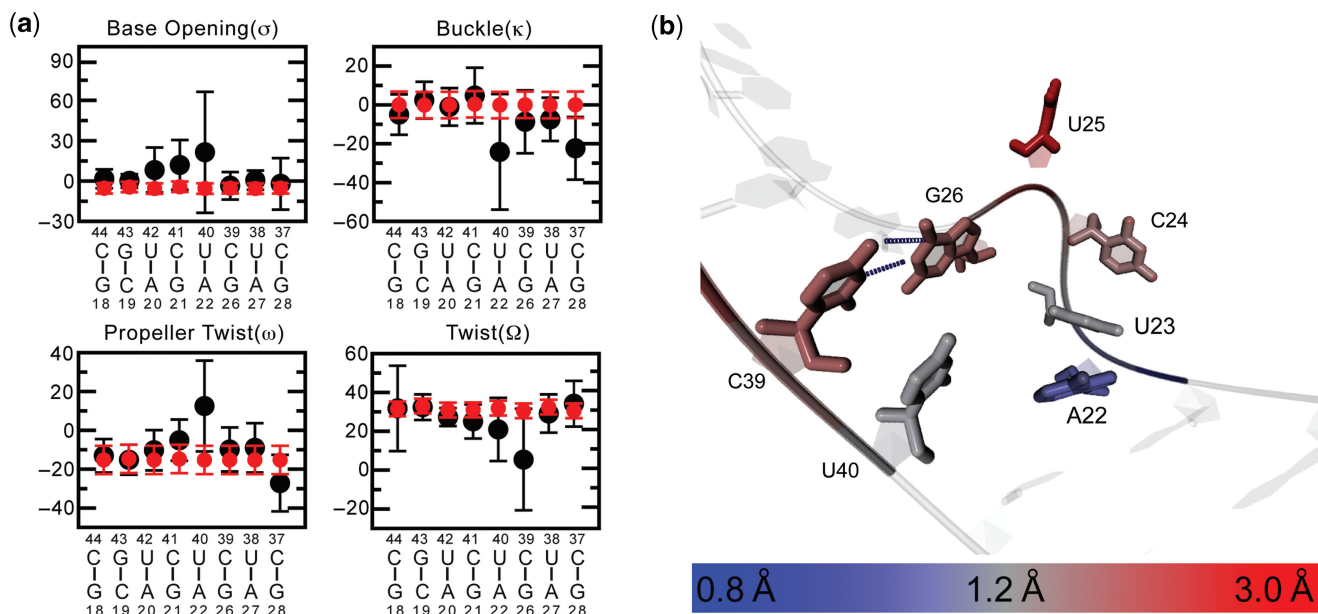


**Figure 2.** Local motions in the HIV-1 TAR dynamical ensemble. (**a**) Shown in black are the mean values for the base opening (σ), buckle (κ), propeller twist (ω) and twist (Ω) angles and their SD calculated over 121 HIV-1 TAR conformers obtained from multiple $N = 20$ SAS runs. For comparison, shown in red are corresponding values for an idealized A-form helix as obtained from a statistical survey of high resolution X-ray structures (48). (**b**) Average conformation of the HIV-1 TAR bulge and neighboring base-pairs calculated from the 121-membered SAS ensemble. The bases of the bulge and flanking base pairs are color-coded based on the root mean square fluctuations (r.m.s.f.) calculated over the ensemble.
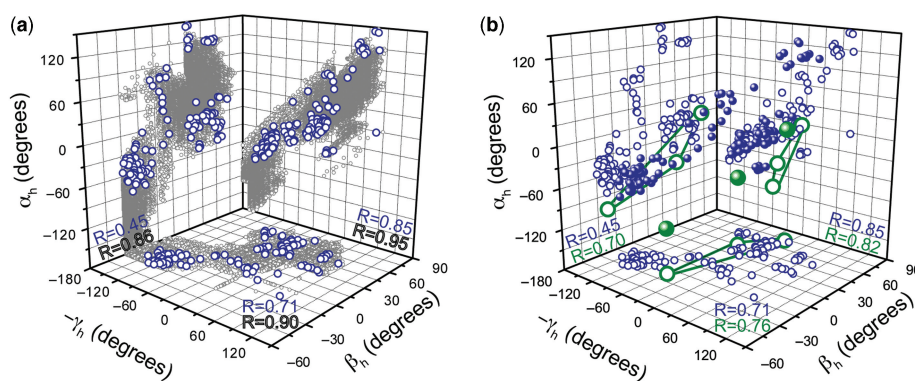
**Figure 3.** Global inter-helical dynamics in the HIV-1 TAR dynamical ensemble. (**a**) Shown in gray are the inter-helical twist ($\alpha_h$ and $\gamma_h$) and bend ($\beta_h$) angles for 80 000 HIV-1 TAR conformers derived from an 80 ns MD trajectory (conformer selected every 1 ps). In blue are the corresponding SAS selected conformers. The correlation coefficient ($R$) is shown on individual planes. (**b**) Comparison of the SAS selected inter-helical angles and those derived previously (26) based on a three-state rigid body ensemble analysis of E-TAR RDCs.

coaxial conformer ($\beta_h \sim -21°$). This is not surprising given that for this near coaxial conformer, the RDCs measured in both EI-TAR and EII-TAR are highly insensitive to the twist angles $\alpha_h$ and $\gamma_h$.

We also examined if there exists any correlation between the geometry of base-pairs at the junction and the inter-helical bend angles. The only significant observation was anti-correlation ($R \sim -0.70$) between the base-pair step angle ($\Omega$) at the G26-C39 junctional base-pair and $\alpha_h$ the twist angle about helix II. Thus, the unraveling of the G26-C39 base-pair may give rise to twisting motions around the axis of helix II.

### Comparison with HIV-2 TAR motions

We used the SAS approach to analyze RDCs previously measured in HIV-2 EI-TAR (26) in which the bulge residue C24 is omitted. Poor agreement (RMSD = 15.1 Hz) was again observed between the measured EI-TAR RDCs and values computed using entire MD trajectory (Figure 4a). By using SAS, we were able to find an $N = 20$ sub-ensemble that yields an RMSD of 1.7 Hz (Figure 4b and Supplementary Figure S2). Compared to HIV-1 TAR, repeated SAS runs resulted in selection of a larger number (276) of unique conformers for HIV-2 TAR. This could be attributed to a smaller RDC sensitivity to twisting motions both because the HIV-2 TAR structure is more linear and because only RDCs measured in EI-TAR was available for analysis.

Comparison of the SAS selected HIV-2 TAR conformers with those obtained for HIV-1 TAR (see Supplementary movies) revealed that reducing the length of the bulge in HIV-2 TAR led to a marked decrease in the local motions in the junctional A22-U40 base-pair (Figure 4c). In HIV-1, the SDs are approximately 30°, 45°, 13°, 16° for the base angles $\kappa$, $\sigma$, $\omega$ and $\Omega$, respectively. In HIV-2, they reduce to 10°, 6°, 8° and 6°, respectively. In contrast, we observe significant static deviations in the opening angle for the G26-C39 junctional base pair. Likewise, a significant reduction is observed in the local dynamics of bulge residues U23 and C25 (Figure 4d). U23 is less flexible and forms more stable stacking interactions

on an also less flexible A22. The root mean square fluctuations (r.m.s.f.) of the atomic positions of U23 and U25 decrease from 1.89 Å and 3.49 Å, respectively in HIV-1 TAR to 1.45 Å and 1.49 Å, respectively in HIV-2 TAR.

The reduction in the local motions in and around the bulge linker is, as expected, accompanied by a reduction in the inter-helical motional amplitudes, as shown in Figure 4e. Such a reduction is clearly observed for the inter-helical bending which decreases in standard deviation from $\sim 33°$ to $\sim 12°$. This is in agreement with an order tensor analysis of RDCs, which reported a reduction in the $\vartheta_{int}$ (which ranges between 0 and 1 for maximum and minimum inter-helical motional amplitudes, respectively) from $0.45 \pm 0.05$ to $0.77 \pm 0.04$ (26). As expected, the HIV-2 conformers cluster more tightly around more linear ($\beta_h \sim 0°$) conformations. While we do not observe a significant reduction in the amplitude of twisting motions about the two helices ($\alpha_h$ and $\gamma_h$) in HIV-2, the RDC sensitivity to these angles is diminished in HIV-2 TAR both because the structure is on average more linear and because RDCs were only measured on the domain I elongated construct. Nevertheless, we observe a stronger correlation between the twisting motions indicating that the $\alpha_h$ and $\gamma_h$ correlations originates in part from the steric drag one helix exerts on the other.

### Dynamics and adaptive recognition of ligands

Numerous studies have shown that HIV-1 TAR undergoes large conformational rearrangements that allow binding of diverse targets in and around the bulge, including peptide derivatives of its cognate protein Tat (53–55), divalent ions (56), and five chemically distinct small molecules (57–60). The three inter-helical conformers obtained by a three-state ensemble analysis of RDCs measured in unbound TAR revealed a global inter-helical motional trajectory that encapsulated many of the ligand bound conformations (26). As shown in Figure 5a, the SAS selected inter-helical conformers trace orientations similar to the ligand-bound TAR conformations, supporting the notion that unbound TAR can dynamically access its ligand-bound global conformations. This can also be
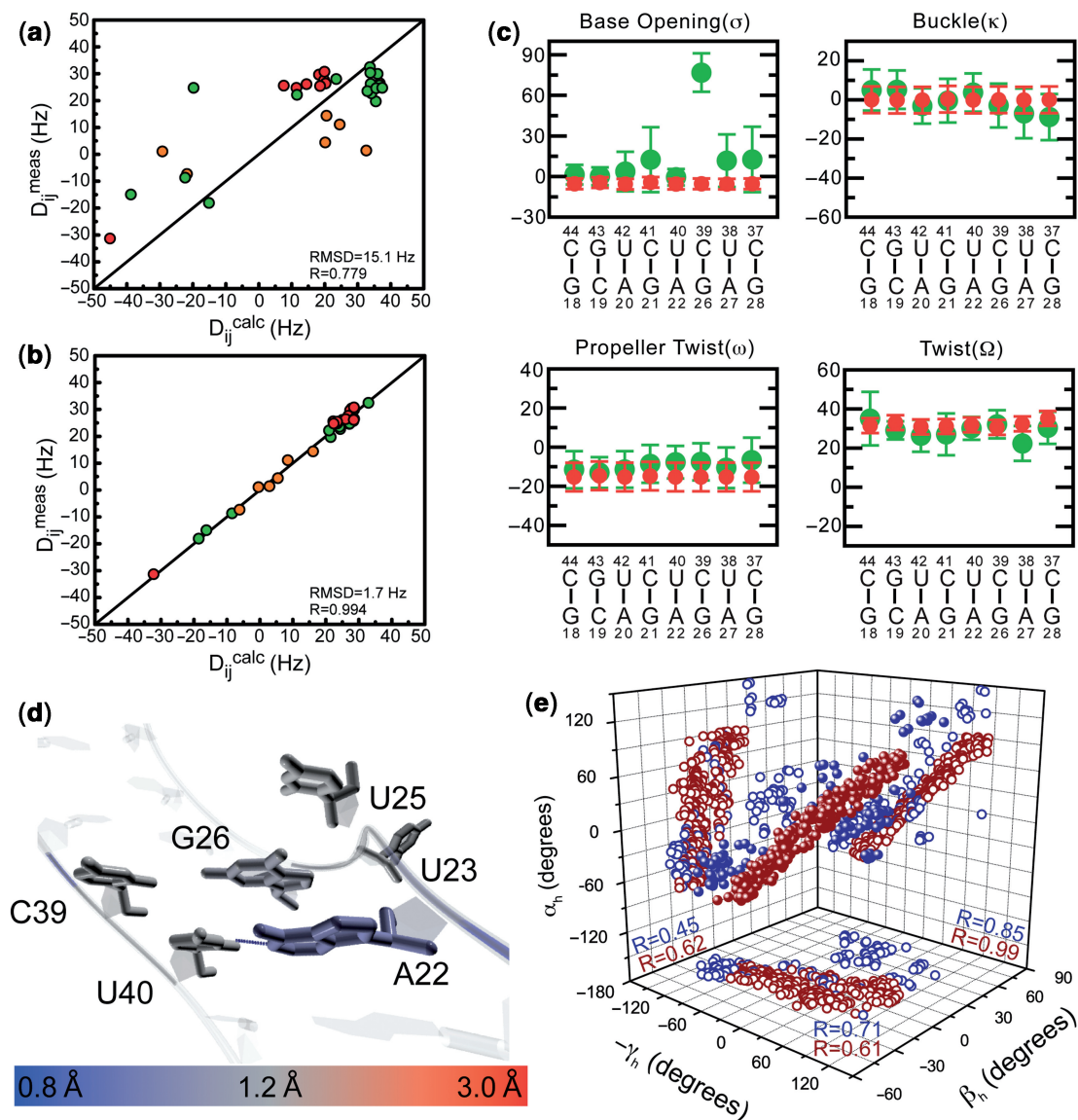
**Figure 4.** HIV-2 TAR dynamical ensemble. Plot of experimental HIV-2 EI-TAR RDCs versus values calculated from (**a**) an 80 ns MD trajectory of HIV-2 TAR and (**b**) following SAS selection of $N = 20$ conformers from the MD trajectory based on HIV2 EI-TAR RDCs. Coloring scheme is same as in Figure 1. (**c**) Shown in black are the mean base-pair angles and their SD calculated over a 279-membered HIV-2 TAR SAS ensemble. For comparison shown in green are corresponding angles for an idealized A-form helix as obtained from a statistical survey of high-resolution X-ray structures (48). (**d**) Average conformation of the bulge of HIV-2 TAR calculated from the 279-membered HIV-2 TAR SAS ensemble. Bases of the bulge and flanking base pairs are color-coded based on the root mean square fluctuations (r.m.s.f.) calculated over the ensemble. (**e**) Comparison of SAS selected HIV-1 (blue) and HIV-2 (red) inter-helical conformations. The correlation coefficient (*R*) is shown on individual planes.

seen in Figure 5b in which we compare the best matching ligand-bound TAR conformers and SAS-selected conformers as determined by superimposing all heavy atoms excluding the flexible terminal base-pairs (G17–C45) and the apical loop.

The SAS dynamical ensembles also allowed us to examine to what extent are local features of the ligand binding pocket dynamically preformed in the absence of ligands. In Figure 5c, we compare the best matching ligand-bound TAR conformers and SAS-selected conformers as determined by superimposing heavy atoms in the bulge and immediately neighboring base-pairs. In general, the largest deviations are seen for the highly flexible residues C24 and

U25, which are also known to be flexible in many of the ligand bound TAR conformations (26,61,62). Overall, these results suggest that local dynamics in an around the TAR bulge likely facilitate formation of the ligand binding pocket.

## DISCUSSION

RNA elongation provides a reliable approach for computing time-averaged RDCs in highly flexible RNAs providing a basis for integration with MD as we showed here using the SAS approach. That the measured RDCs in both HIV-1 and HIV-2 can be satisfied using conformers
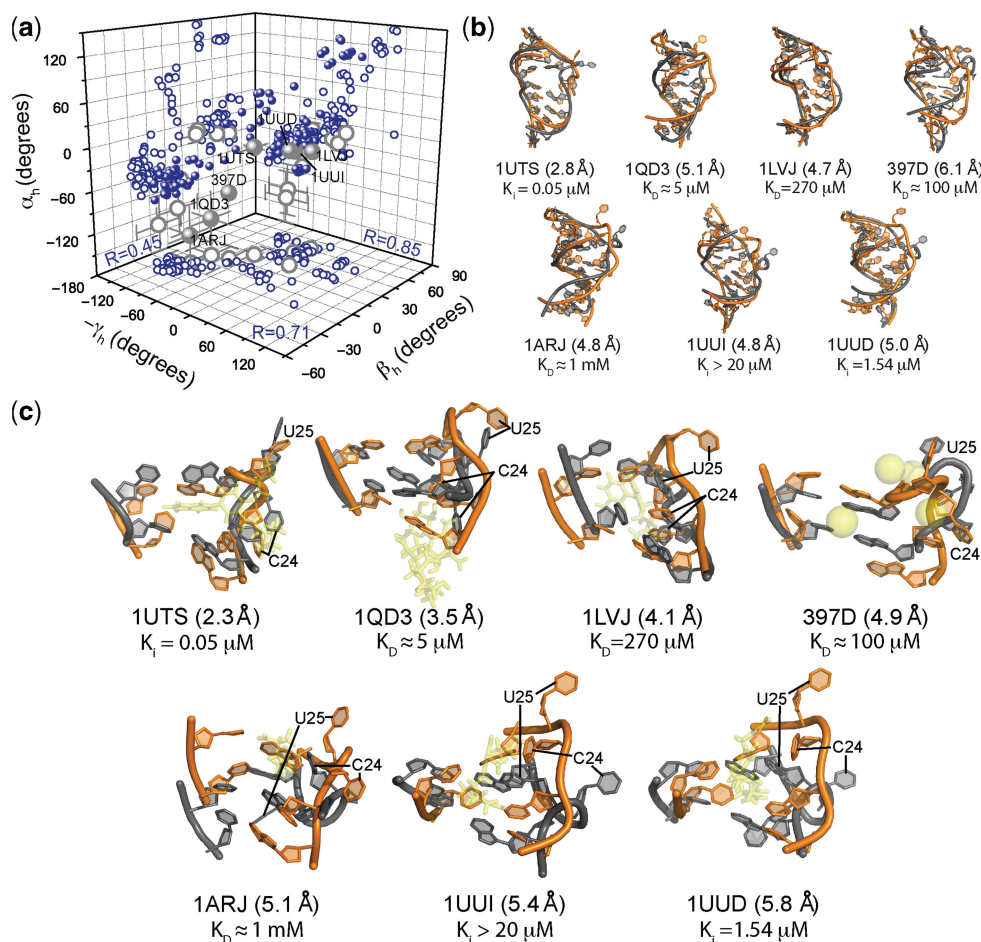
**Figure 5.** Comparison of the SAS derived HIV-1 TAR dynamical ensemble and ligand bound TAR conformations. (**a**) Comparison of the global inter-helical angles. Shown in blue are the SAS selected angles and in gray seven distinct ligand bound TAR structures. (**b** and **c**) Comparison of the b, global and c, local structure of SAS TAR conformers and seven distinct ligand bound TAR conformations (1QD3, 1UUI, 1UTS, 1UUD, 1ARJ, 1LVJ, and 397D). Shown are the pairs yielding the lowest RMSD fit when superimposing b, all heavy atoms excluding terminal base-pairs G17-C45 and the apical loop and c, all heavy atoms in the bulge and immediately adjacent base-pairs. Every model in the ligand bound NMR ensembles was used in the superposition. The corresponding ligand is colored yellow.

selected from the MD trajectory suggests that neither HIV-1 or HIV-2 TAR undergo significant μs-ms motions that are not sampled in the MD trajectory. This is consistent with relaxation dispersion NMR studies of HIV-1 TAR which provide no evidence for μs-ms motions in and around the bulge (43). In further support of this idea comes the observation that the first few low-frequency eigenmodes of quasiharmonic motion obtained by principal component analysis of the 80 000 snapshots capture the dynamics as described by the SAS ensemble (Supplementary Figure S3).

However, it should also be noted that the SAS conformers represent a 'discrete' approximation to what is more likely a continuous complex distribution of many more conformations. The SAS conformers can be thought of as discrete points along the configuration space that may help define salient features of the motional trajectory (63). It should be kept in mind that averaging over such a discrete number of conformers leads to efficient averaging of the RDCs. A correspondingly larger continuous distribution of many more conformations will likely be required to

accomplish the same level of motional averaging. It is therefore not surprising that the MD trajectory evolves outside the envelope defined by the SAS conformers and that the SAS envelope increases with the size of the ensemble (Supplementary Figure S1). Thus, the most likely source of discrepancy between the measured RDCs and the MD simulation is the assignment of the relative weights to the various conformers. RDC studies of partially unfolded proteins have emphasized the exquisite sensitivity of RDCs to the underlying conformational distribution (64).

The comparison of the dynamical ensembles generated for HIV-1 and HIV-2 TAR provided fundamental new insights into the dependence of RNA dynamics on the bulge length. As would be expected, reducing the length of the HIV-1 trinucleotide bulge by a single nucleotide (HIV-2) led to a significant reduction in the local motions in and around the bulge, as well as global inter-helical motions (see Supplementary movies). In both HIV-1 and HIV-2, we observe spatial correlations between twisting motions about individual helices, and to a lesser extent

between twisting and bending as first reported based on the three-state ensemble analysis of the RDCs (26). Importantly, the spatial correlations between the twisting motions increase significantly in HIV-2 TAR (Figure 4e). This is expected if one were to consider limiting cases for the bulge length; at the limit of not having a bulge linker, the twisting dynamics becomes perfectly correlated, whereas for an infinitely long bulge, one would expect little correlation.

The HIV-1 TAR dynamical ensemble allowed us to directly examine if unbound TAR dynamically samples diverse ligand-bound conformations that have been reported to date. Our results suggest that many of the key features of the ligand bound TAR conformations, including the global inter-helical orientation and local aspects of the ligand-binding pocket appear to be dynamically pre-formed in unbound TAR. However, a detailed comparison of the SAS conformations and the ligand bound TAR structures remains complicated by a number of factors. As mentioned above, the SAS conformers only represent an approximate discrete state ensemble to what is likely a more continuous conformational distribution. One also has to consider the uncertainty in ligand bound TAR structures, both due to experimental imprecision and because the ligand bound states may be flexible themselves. Previous NMR studies reveal significant mobility in the TAR bulge, especially residues C24 and U25 when in complex with argininamide (49,65) and $Mg^{2+}$ (62,66). The ACP bound TAR structure also exhibits a large degree of inter-helical motions (61). Notwithstanding these complications, our results suggest that local dynamics in and around the bulge together with global motions of helices can drive much of the conformational adaptation required to bind to different ligand targets.

Our results also suggest that some of the TAR conformational changes require ligand binding to occur efficiently. For example, we do not observe the key U38-A27•U23 base-triple in either of the HIV-1 or HIV-2 SAS conformers that is known to form in the TAR-argininamide complex (53). In this regard, it is interesting to note that larger deviations between the SAS and ligand bound conformations are generally observed for the weaker binding ligands and *vice versa*. It is possible that the weaker binding ligands expend a greater fraction of the binding energy changing the TAR conformation. Additional studies are needed to shed light on these key energetic questions.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the TeraGrid and NERSC for their generous computing support.

## FUNDING

*Conflict of interest statement.* None declared.

## REFERENCES

1. Perez-Canadillas,J.M. and Varani,G. (2001) Recent advances in RNA-protein recognition. *Curr. Opin. Struct. Biol.*, **11**, 53–58.
2. Tucker,B.J. and Breaker,R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.*, **15**, 342–348.
3. Nudler,E. (2006) Flipping riboswitches. *Cell*, **126**, 19–22.
4. Al-Hashimi,H.M. and Walter,N.G. (2008) RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.*, **18**, 321–329.
5. Hall,K.B. (2008) RNA in motion. *Curr. Opin. Chem. Biol.*, **12**, 612–618.
6. Williamson,J.R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol.*, **7**, 834–837.
7. Leulliot,N. and Varani,G. (2001) Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, **40**, 7947–7956.
8. Al-Hashimi,H.M. (2005) Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chembiochem*, **6**, 1506–1519.
9. Mittermaier,A. and Kay,L.E. (2006) New tools provide new insights in NMR studies of protein dynamics. *Science*, **312**, 224–228.
10. Henzler-Wildman,K. and Kern,D. (2007) Dynamic personalities of proteins. *Nature*, **450**, 964–972.
11. Furtig,B., Buck,J., Manoharan,V., Bermel,W., Jaschke,A., Wenter,P., Pitsch,S. and Schwalbe,H. (2007) Time-resolved NMR studies of RNA folding. *Biopolymers*, **86**, 360–383.
12. Orozco,M., Noy,A. and Perez,A. (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.*, **18**, 185–193.
13. Mackerell,A.D. Jr and Nilsson,L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.*, **18**, 194–199.
14. Showalter,S.A. and Hall,K.B. (2005) Isotropic reorientational eigenmode dynamics complements NMR relaxation measurements for RNA. *Methods Enzymol.*, **394**, 465–480.
15. Duchardt,E., Nilsson,L. and Schleucher,J. (2008) Cytosine ribose flexibility in DNA: a combined NMR 13C spin relaxation and molecular dynamics simulation study. *Nucleic Acids Res.*, **36**, 4211–4219.
16. Trantirek,L., Caha,E., Kaderavek,P. and Fiala,R. (2007) NMR (13)C-relaxation study of base and sugar dynamics in GCAA RNA hairpin tetraloop. *J. Biomol. Struct. Dyn.*, **25**, 243–252.
17. Ferner,J., Villa,A., Duchardt,E., Widjajakusuma,E., Wohnert,J., Stock,G. and Schwalbe,H. (2008) NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.*, **36**, 1928–1940.
18. Lindorff-Larsen,K., Best,R.B., Depristo,M.A., Dobson,C.M. and Vendruscolo,M. (2005) Simultaneous determination of protein structure and dynamics. *Nature*, **433**, 128–132.
19. Chen,Y., Campbell,S.L. and Dokholyan,N.V. (2007) Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys. J.*, **93**, 2300–2306.
20. Markwick,P.R., Bouvignies,G. and Blackledge,M. (2007) Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. *J. Am. Chem. Soc.*, **129**, 4724–4730.
21. Showalter,S.A. and Bruschweiler,R. (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J. Am. Chem. Soc.*, **129**, 4158–4159.
22. Zhang,Q. and Al-Hashimi,H.M. (2008) Extending the NMR spatial resolution limit for RNA by motional couplings. *Nat. Methods*, **5**, 243–245.
23. Musselman,C., Al-Hashimi,H.M. and Andricioaei,I. (2007) iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys. J.*, **93**, 411–422.

24. Showalter,S.A., Baker,N.A., Tang,C.G. and Hall,K. (2005) Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *J. Biomol. NMR*, **32**, 179–193.

25. Zhang,Q., Sun,X., Watt,E.D. and Al-Hashimi,H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653–656.

26. Zhang,Q., Stelzer,A.C., Fisher,C.K. and Al-Hashimi,H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, **450**, 1263–1267.

27. Lipari,G. and Szabo,A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.*, **104**, 4546–4559.

28. Tjandra,N. and Bax,A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**, 1111–1114.

29. Tolman,J.R., Flanagan,J.M., Kennedy,M.A. and Prestegard,J.H. (1995) Nuclear magnetic dipole interactions in field-oriented proteins – information for structure determination in solution. *Proc. Natl Acad. Sci. USA*, **92**, 9279–9283.

30. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, **4**, 187–217.

31. MacKerell,A.D., Banavali,N. and Foloppe,N. (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, **56**, 257–265.

32. Aboul-ela,F., Karn,J. and Varani,G. (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.*, **24**, 3974–3981.

33. Brodsky,A.S. and Williamson,J.R. (1997) Solution structure of the HIV-2 TAR-argininamide complex. *J. Mol. Biol.*, **267**, 624–639.

34. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

35. Brooks III,C.L. and Karplus,M. (1983) Deformable stochastic boundary in molecular dynamics. *J. Chem. Phys.*, **79**, 6312–6325.

36. Nose,S. (1984) A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, **81**, 511–519.

37. Hoover,W.G. (1985) Canonical dynamics – equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695–1697.

38. Caves,L.S., Evanseck,J.D. and Karplus,M. (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.*, **7**, 649–666.

39. Auffinger,P. and Westhof,E. (1996) H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations. *Biophys. J.*, **71**, 940–954.

40. Saupe,A. (1968) Recent results in the field of liquid crystals. *Angew Chem. Int. Ed. Engl.*, **7**, 97–112.

41. Losonczi,J.A., Andrec,M., Fischer,M.W.F. and Prestegard,J.H. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.*, **138**, 334–342.

42. Zhang,Q., Throolin,R., Pitt,S.W., Serganov,A. and Al-Hashimi,H.M. (2003) Probing motions between equivalent RNA domains using magnetic field induced residual dipolar couplings: accounting for correlations between motions and alignment. *J. Am. Chem. Soc.*, **125**, 10530–10531.

43. Dethoff,E.A., Hansen,A.L., Musselman,C., Watt,E.D., Andricioaei,I. and Al-Hashimi,H.M. (2008) Characterizing complex dynamics in the TAR apical loop and motional correlations with the bulge by NMR, MD and mutagenesis. *Biophys. J.*, **95**, 3906–3915.

44. Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.

45. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.

46. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.

47. Bailor,M.H., Musselman,C., Hansen,A.L., Gulati,K., Patel,D.J. and Al-Hashimi,H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat. Protocols*, **2**, 1536–1546.

48. Musselman,C., Pitt,S.W., Gulati,K., Foster,L.L., Andricioaei,I. and Al-Hashimi,H.M. (2006) Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J. Biomol. NMR*, **36**, 235–249.

49. Pitt,S.W., Majumdar,A., Serganov,A., Patel,D.J. and Al-Hashimi,H.M. (2004) Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg2+-induced conformational stabilization. *J. Mol. Biol.*, **338**, 7–16.

50. Al-Hashimi,H.M., Gosser,Y., Gorin,A., Hu,W., Majumdar,A. and Patel,D.J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J. Mol. Biol.*, **315**, 95–102.

51. Al-Hashimi,H.M., Gorin,A., Majumdar,A., Gosser,Y. and Patel,D.J. (2002) Towards structural genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. *J. Mol. Biol.*, **318**, 637–649.

52. Hansen,A.L. and Al-Hashimi,H.M. (2007) Dynamics of large elongated RNA by NMR carbon relaxation. *J. Am. Chem. Soc.*, **129**, 16072–16082.

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

54. Aboul-ela,F., Karn,J. and Varani,G. (1995) The Structure of the Human-Immunodeficiency-virus type-1 Tar RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.*, **253**, 313–332.

55. Long,K.S. and Crothers,D.M. (1999) Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry*, **38**, 10059–10069.

56. Ippolito,J.A. and Steitz,T.A. (1998) A 1.3-angstrom resolution crystal structure of the HIV-1 trans- activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc. Natl Acad. Sci. USA*, **95**, 9819–9824.

57. Faber,C., Sticht,H., Schweimer,K. and Rosch,P. (2000) Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.*, **275**, 20660–20666.

58. Du,Z., Lind,K.E. and James,T.L. (2002) Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.*, **9**, 707–712.

59. Davis,B., Afshar,M., Varani,G., Murchie,A.I., Karn,J., Lentzen,G., Drysdale,M., Bower,J., Potter,A.J., Starkey,I.D. *et al.* (2004) Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.*, **336**, 343–356.

60. Murchie,A.I., Davis,B., Isel,C., Afshar,M., Drysdale,M.J., Bower,J., Potter,A.J., Starkey,I.D., Swarbrick,T.M., Mirza,S. *et al.* (2004) Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.*, **336**, 625–638.

61. Pitt,S.W., Zhang,Q., Patel,D.J. and Al-Hashimi,H.M. (2005) Evidence that electrostatic interactions dictate the ligand-induced arrest of RNA global flexibility. *Angew Chem. Int. Ed. Engl.*, **44**, 3412–3415.

62. Al-Hashimi,H.M., Pitt,S.W., Majumdar,A., Xu,W. and Patel,D.J. (2003) Mg2+-induced variations in the conformation and dynamics of HIV-1 TAR RNA probed using NMR residual dipolar couplings. *J. Mol. Biol.*, **329**, 867–873.

63. Fisher,C.K., Zhang,Q., Stelzer,A. and Al-Hashimi,H.M. (2007) Ultra-high resolution characterization of domain motions and correlations by multi-alignment and multi-reference RDC NMR. **450**, 1263–1267.

64. Meier,S., Blackledge,M. and Grzesiek,S. (2008) Conformational distributions of unfolded polypeptides from novel NMR techniques. *J. Chem. Phys.*, **128**, 052204.

65. Dayie,K.T., Brodsky,A.S. and Williamson,J.R. (2002) Base flexibility in HIV-2 TAR RNA mapped by solution (15)N, (13)C NMR relaxation. *J. Mol. Biol.*, **317**, 263–278.

66. Casiano-Negroni,A., Sun,X. and Al-Hashimi,H.M. (2007) Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. *Biochemistry*, **46**, 6525–6535.