

## MD Simulations of the dsRBP DGCR8 Reveal Correlated Motions that May Aid pri-miRNA Binding

Christopher Wostenberg, W. G. Noid, and Scott A. Showalter\*

Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania

**ABSTRACT** Over the past decade, microRNAs (miRNAs) have been shown to affect gene regulation by basepairing with messenger RNA, and their misregulation has been directly linked with cancer. DGCR8, a protein that contains two dsRNA-binding domains (dsRBDs) in tandem, is vital for nuclear maturation of primary miRNAs (pri-miRNAs) in connection with the RNase III enzyme Drosha. The crystal structure of the DGCR8 Core (493–720) shows a unique, well-ordered structure of the linker region between the two dsRBDs that differs from the flexible linker connecting the two dsRBDs in the antiviral response protein, PKR. To better understand the interfacial interactions between the two dsRBDs, we ran extensive MD simulations of isolated dsRBDs (505–583 and 614–691) and the Core. The simulations reveal correlated reorientations of the two domains relative to one another, with the well-ordered linker and C-terminus serving as a pivot. The results demonstrate that motions at the domain interface dynamically impact the conformation of the RNA-binding surface and may provide an adaptive separation distance that is necessary to allow interactions with a variety of different pri-miRNAs with heterogeneous structures. These results thus provide an entry point for further in vitro studies of the potentially unique RNA-binding mode of DGCR8.

### INTRODUCTION

Mature microRNAs (miRNAs) are small, single-stranded RNAs (ssRNAs), 21–25 nucleotides in length, that affect gene translation by basepairing with messenger RNA (1). A majority of human genes (at least 60%) are regulated post-transcriptionally by one or more miRNAs (2). Recent studies have demonstrated that miRNAs contribute to the control of cellular homeostasis in multicellular organisms by regulating such biologically important processes as apoptosis, cell cycle progression, and cell-type differentiation (1,3), as well as disease states leading to cancer, Alzheimer's disease, and autoimmune diseases (1,4,5). Maturation of miRNA occurs in two independent and spatially separated steps. The first step occurs in the nucleus, where the single-stranded tail of primary microRNA (pri-miRNA) is removed by the microprocessor, which contains the RNase III enzyme Drosha and the double-stranded RNA (dsRNA)-binding protein (dsRBP) DGCR8 (known as Pasha in *Drosophila*) (6). The second maturation step occurs in the cytosol, where the terminal loop of pre-miRNA is cleaved by Dicer, another RNase III enzyme (6).

Originally, the dsRNA-binding domain (dsRBD) containing protein DGCR8 was identified in humans as a protein encoded in the region of chromosome 22 that is deleted in patients with DiGeorge syndrome (7). Since then, DGCR8 has been shown to play a critical role in processing pri-miRNA into pre-miRNA by binding to the double-stranded region of pri-miRNA (8). Throughout the maturation process, dsRBPs are required for recruitment and cleavage of the nascent miRNA. dsRBPs are seen in all forms of

life, from viral-encoded to prokaryotic and eukaryotic products, making them one of the most common RNA-binding motifs (9). Also, dsRBPs are found in both the cytoplasm and the nucleus, as evidenced by the compartmentalization of the first and second stages of miRNA processing in the nucleus and cytosol, respectively (6,10,11). The dsRBD, which is evolutionary conserved, consists of ~65–68 amino acids forming an  $\alpha\beta\beta\beta$  secondary structure (12–14). Some eukaryotic dsRBPs contain up to five dsRBDs, whereas others contain only one (15). As few as 11 basepairs of dsRNA (coinciding with one turn of canonical A-form helix) have been shown to interact with a single dsRBD (16). A-form double-helix RNA is the preferred binding partner of dsRBDs, which show little or no affinity for ssRNA, dsDNA, and DNA-RNA hybrids (17,18).

DGCR8 is a 773 amino acid residue protein that contains two dsRBDs at the C-terminal end of the protein, separated by a ~50 amino acid linker (Fig. 1 A). Except for a predicted WW domain in the middle of the sequence, the rest of the protein is intrinsically disordered and without known function. Even though DGCR8 does not contain known enzymatic activity, its role as a molecular anchor in directing Drosha to cleave the pri-miRNA ~11 basepairs from the ssRNA-dsRNA junction is vital for specificity in the miRNA maturation process (19).

Recently, Sohn et al. (20) determined the crystal structure of the Core region of DGCR8 (residues 493–720 of the human sequence), which contains both dsRBDs and the intervening ~50 amino acid residue linker. The crystallographic data revealed that the two dsRBDs are arranged in a pseudo twofold symmetry and packed against a well-defined secondary structure formed from the linker and the C-terminal tail of the construct (20). Notably, a key component of the interface is an  $\alpha$ -helix formed by the tail region at

Submitted November 6, 2009, and accepted for publication April 1, 2010.

\*Correspondence: sas76@psu.edu

Editor: Martin Blackledge.

© 2010 by the Biophysical Society  
0006-3495/10/07/0248/9 \$2.00

doi: 10.1016/j.bpj.2010.04.010

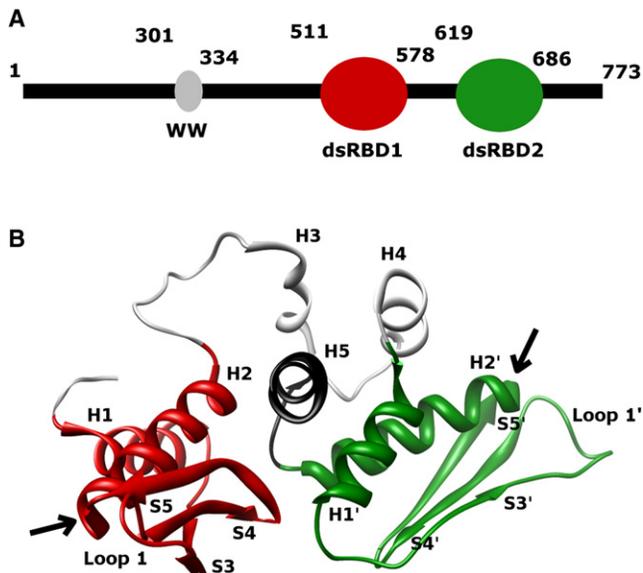


FIGURE 1 (A) Schematic representation of the primary sequence of DGCR8. (B) A ribbon diagram representing the crystal structure of DGCR8 Core (PDB 2YT4, residues 505–701) shows a well-structured linker (H3 and H4) and a C-terminal helix (H5, *black*). The nomenclature used is the same as the crystal structure nomenclature (20). dsRBD1 is shown in red and dsRBD2 is shown in green (in dsRBD2, helices and strands are denoted with primes to indicate equivalence with secondary structures to dsRBD1). The arrows indicate the proposed dsRNA-binding sites.

the C-terminal end of dsRBD2 (Fig. 1 B). This helix is tightly packed against the two dsRBDs and forms hydrophobic interactions with  $\alpha$ -helix H2 and  $\beta$ -strand S5 of both dsRBDs, with additional interactions from helix H1' of dsRBD2 (20). Lastly, Sohn et al. (20) noted several hydrogen bonds between both dsRBDs and the C-terminal  $\alpha$ -helix H5. This well-organized arrangement of the dsRBDs in the absence of RNA is in contrast to the NMR structure of the antiviral response protein PKR, which shows a flexible linker and structural independence between its tandem dsRBDs in the apo-state (21).

Both dsRBDs of DGCR8 are essential for high-affinity binding of dsRNA,  $K_D \sim 2.0 \mu\text{M}$  for pri-miR-16-1 (20,22). A mutational analysis showed that Lys-561, Lys-562, and Lys-565 in dsRBD1 (found in  $\alpha$ -helix H2) and Lys-669, Arg-670, and Lys-673 in dsRBD2 (found in  $\alpha$ -helix H2') contribute favorably to the energetics of pri-miRNA binding (20). To date, no RNA-bound structure of DGCR8 has been reported; however, Ryter and Schultz (13) and Sohn et al. (20) have proposed a binding mode based on mutational studies and homology modeling with the cocrystal structure of Xlrpba bound to dsRNA. This model suggests that the pri-miRNA would have to bend upon binding, or that DGCR8 goes through a conformational change upon binding, or both. Recent studies demonstrated that, under native conditions, proteins routinely sample conformations in the unbound state that are essential for bound-state function (23), indicating the likelihood that quantifying the apo-state

dynamics of DGCR8 will yield insight into the miRNA binding mechanism.

This motivated us to quantify the stability of the interactions at the interface and determine their impact on structure and function. In an initial analysis of structure-function relationships in DGCR8, Sohn et al. (20) performed a mutational analysis that focused on the proposed RNA binding site, leaving the novel and intriguing domain interface revealed by the structural work largely uninvestigated. In this study, we provide an integrative look at the interface between the two dsRBDs and the impact of its dynamics on the proposed interface for dsRNA binding using molecular-dynamics (MD) simulations. Despite the ubiquitous presence of dsRBDs in nature, and the availability of several atomic-resolution structures of dsRBD containing proteins (13,21,24–26), relatively few MD simulations of these domains have been calculated compared to other RNA-binding motifs and nucleic acid-protein complexes (27). In one notable example, Castrignanò et al. (28) simulated the *Drosophila* Staufen dsRBD3 free and bound to dsRNA, and showed a high degree of flexibility, even in the complex, of the RNA recognition loops (loops 2 and 4). To date, the RNA-binding protein that has been the most extensively studied by MD simulation methods is the splicing protein U1A from the U1 snRNP; however, this single-strand RNA-binding protein utilizes a fundamentally different binding mechanism mediated by the ssRNA sequence (29–31). Our *in silico* studies of DGCR8 provide a starting point for understanding what appears from known atomic structures of dsRBDs to be the unique binding mode of dsRNA by DGCR8, and lay a foundation for future *in vitro* and *in vivo* experiments.

## MATERIALS AND METHODS

A preliminary anharmonic normal mode (ANM) analysis on the crystal structure was done using the anisotropic network model web server (32). The suggested parameters from the web server were used to run the simulation, which included a 15 Å interaction cutoff and a distance weight of 2.5 for the interactions between C $\alpha$  atoms. Although the full set of modes reported by the server was analyzed, only the first mode is reported here.

MD trajectories were run in the AMBER 10.0 software package (33) using the ff99SB (34,35) force field. Simulations were carried out in explicit solvent represented by the SPC water model (36) under particle mesh Ewald periodic boundary conditions (37). Three initial protein lengths were generated from the crystal structure of RNA-free DGCR8 (PDB code 2YT4) (20): DGCR8-Core (505–701), DGCR8-dsRBD1 (505–583), and DGCR8-dsRBD2 (614–691). The crystal coordinates of DGCR8 are missing several residues that were built back into the starting conformations for the Core and dsRBD2 simulations. The loop formed by residues 643–648 was created in both starting conformations by restoring the residues VVPGKN using the Coot software package (38). For the Core simulation, residues 584–591 were also added by modeling the sequence SEEKPKD in Coot. A number of chloride counterions sufficient to neutralize the net positive charge on the proteins were added, and the resulting systems were solvated such that no solute atom was within 10 Å of a box edge. This required 21,936 water molecules for the Core, 8777 for dsRBD1, and 9232 for dsRBD2. The starting configurations were energy-minimized and equilibrated as previously reported (39). After the initial equilibration period, 250 ns of dynamics were

run in an isothermal-isobaric (NPT) simulation for each construct. Snapshots from each trajectory were stored to disk every 1.0 ps. The analysis of the trajectories was done in AMBER using the ptraj program (33). Molecular graphics images were created using the UCSF Chimera package (40). Additional analysis and visualization were accomplished in MATLAB (The MathWorks, Natick, MA).

## RESULTS AND DISCUSSION

The preorganization of DGCR8's dsRBDs seen crystallographically in the absence of RNA is in contrast to the only other previously determined structure of a protein with tandem dsRBDs, PKR, which shows a flexible linker (21). The well-defined structure in the linker and C-terminal end of DGCR8 produces a pseudo twofold symmetry in the Core that defines the relative orientation and spacing of the two dsRNA binding surfaces (Fig. 1 B). The novel interaction between the dsRBDs of DGCR8 seen in the crystal structure supports a new mechanism for dsRNA binding to dsRBDs and requires a quantitative understanding of the temporal and spatial dynamics in the linker, as they will directly impact the stability of the predefined spacing and orientation between the dsRNA-binding surfaces.

### Root mean-square deviation of DGCR8

The interfacial interactions of DGCR8 were studied using MD simulations of three different constructs derived from the crystal structure of the RNA-free Core (PDB 2YT4): DGCR8-dsRBD1 (505–583), DGCR8-dsRBD2 (614–691), and DGCR-Core (505–701). Two loops are absent in the crystal structure due to low electron density and were therefore modeled back into the structure before the simulations were run (see [Materials and Methods](#)). Analysis proceeded after the calculation of 250 ns isothermal-isobaric (NPT) trajectories of each construct.

Protein stability was checked by analyzing the backbone root mean-square deviation (RMSD) from the starting crystal structure over the course of the trajectory. Each construct is stable over the simulation timescale, as indicated by the plateau in the time trace of the RMSD (Fig. 2). The lowest RMSD is seen in dsRBD1, which indicates a higher stability for this domain and a smaller deviation from the starting crystal structure. A slightly higher RMSD is seen in dsRBD2, indicating that this domain is more dynamic than dsRBD1. Even excluding the loop that was added to the crystal structure in order to run the simulation, dsRBD2 still shows a higher RMSD than dsRBD1. dsRBD1 is more closely related to the canonical dsRBDs (Xlrpba, *Drosophila* Staufen, and RNase III) than dsRBD2 (20), which could also explain the difference in the dynamics of the two domains. This difference in the dynamics of the two isolated domains could lead to slightly different binding affinities. Early in the simulations, the Core reaches an RMSD of  $\sim 3.0$  Å, which then rises further to 4.5 Å after 60 ns and is maintained at that level for the rest of the simulation. As with dsRBD2, the high RMSD in

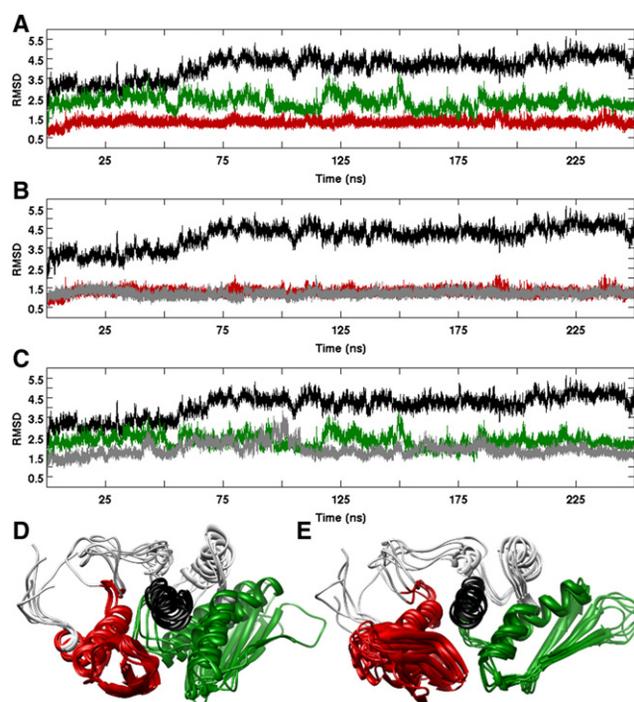


FIGURE 2 RMSD traces show the overall stability of the dsRBDs during the MD simulations and highlight the rearrangement of the domains relative to each other in the Core simulation. (A) RMSD traces of dsRBD1 (red), dsRBD2 (green), and Core (black) relative to the starting crystal structure. (B) RMSD trace of the isolated dsRBD1 (red) compared with the RMSD trace of dsRBD1 from the Core simulation (gray). (C) RMSD trace of the isolated dsRBD2 (green) compared with the RMSD trace of dsRBD2 from the Core simulation (gray). In both B and C the total Core RMSD (black) is given as a reference. (D) Ribbon bundle from the Core simulation in which dsRBD1 (red) is superimposed. (E) Ribbon bundle from the Core simulation in which dsRBD2 (green) is superimposed. Both bundles are created by taking structures from the simulation every 50 ns.

the Core can be partially attributed to an enhanced flexibility in the added loops that are not present in the crystal structure.

The high RMSD seen in the Core simulation is sufficient to cause concern in a single globular domain that the structure is unstable. However, DGCR8 is not a single global domain, and the somewhat large RMSD is reasonable for a multiple-domain protein if it can be attributed to the two domains reorienting themselves relative to each other while still retaining their overall structure. Mathematically, this would tend to inflate the RMSD because no single reference structure could serve well for the RMSD calculation over the entire time course. The RMSD trace shows that the Core simulation reaches a steady state of 4.5 Å deviation from the starting crystal conformation, indicating that the Core does reach a stable state, albeit one that deviates significantly from the starting structure. Rigid-body-type reorientations of the two domains with respect to one another are distinct from, for example, complete local unfolding of secondary elements in the linker. Therefore, to draw accurate conclusions from the RMSD traces, one must further explore the limiting example the Core most closely resembles.

If the two domains in the Core simulation are stable individually, then superposition of the residues corresponding to each dsRBD in the Core should yield low RMSD values comparable to those observed in the isolated-domain simulations. In dsRBD1 (Fig. 2 B), the RMSD trace from the isolated domain simulation and the RMSD trace from the individual domain in the Core simulation superimpose extensively onto each other. In dsRBD2 (Fig. 2 C), the isolated domain simulation has a slightly higher RMSD value throughout the simulation compared to the individual domain in the Core simulation, implying that features in the Core stabilize this domain.

The piecewise RMSD traces only provide information on the deviation of the individual domains being superimposed from their starting conformations in the crystal structure; therefore, to obtain information on the rest of the protein, ribbon bundles of the structure were analyzed (Fig. 2, D and E). As expected from the RMSD results, the domain being superimposed to generate each bundle does not deviate much, but the rest of the protein varies in its orientation with respect to the superimposed domain, consistent with the idea that the two domains may be reorienting relative to one another on the simulation timescale. Even though the linker and the nonsuperimposed domain fluctuate quite a bit in their relative orientation, the secondary structural elements of these regions are retained throughout the trajectories. These bundles rule out local unfolding of the secondary elements in the linker, as they are clearly retained. Additionally, from the ribbon bundles we find that dsRBD2 has a higher RMSD than dsRBD1 because loop 1' of dsRBD2 fluctuates more than loop 1 of dsRBD1 when superimposed. Thus, from the ribbon bundle structures, we have shown that the high RMSD from the Core is due not to instability of the structure, but to the rigid-body movement of the domains relative to each other.

### ANM analysis

We were concerned that the global reorientation of the dsRBDs seen in the simulations might have been an artifact, so we sought an independent measurement to verify the global reorientation before continuing with a more in-depth analysis of the correlations of the domain from the simulation. Elastic network models provide an inexpensive alternative to traditional atomistic normal mode analysis and have been shown to accurately reproduce anisotropic displacement factors in high-resolution crystals (41), as well as structural transitions in the functional cycle of the chaperonin GroEL (42). The ANM analysis, in particular, utilizes the same starting crystal structure used in our simulation with a fundamentally different description of the forces underlying atomic displacements, making it an effective verification tool for our study. We therefore performed an ANM analysis of the Core (residues 505–701) (32). The largest-amplitude eigenmode of the ANM calculation features correlated hinging motion in

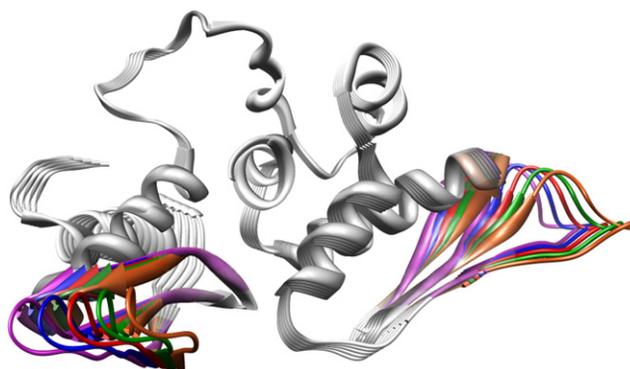


FIGURE 3 Ribbon diagrams spanning the range of motion in the largest-amplitude ANM illustrate the correlated movement of the dsRBD wings toward each other. Colorized regions highlight the sections of maximal displacement, with dsRBD1 and dsRBD2 colored identically in the given structure to guide the eye.

the dsRBDs, reminiscent of a butterfly flapping its wings (snapshots of structures spanned by the motion along this eigenmode are shown in Fig. 3). The motions of the domains are centered on the  $\alpha$ -helix formed by the C-terminal residues that pack between the dsRBDs. Keeping with the butterfly analogy, the two wings (formed by the two dsRBDs) flap in an anticorrelated movement that changes the distance between the two proposed RNA-binding surfaces, rather than in a correlated twisting motion that would leave the distance between the binding surfaces unchanged. The largest-amplitude motion is seen in loop 1 of both domains (residues 536–541 in dsRBD1 and 641–650 in dsRBD2), which is proximal to the proposed dsRNA-binding interface. The similarity between the ANM results and the features of the structural bundles from the MD simulation is striking. These results demonstrate that motions at the domain interface dynamically impact the conformation of the RNA-binding surface and may provide an adaptive separation distance that is necessary to allow optimal interactions with a variety of different pri-miRNAs with heterogeneous structures (14,18,19,43,44).

### Correlated dynamics in dsRBDs

Our initial motivation for calculating the MD trajectories was the unique prearrangement of the dsRBDs in the DGCR8 Core in the absence of RNA, which suggested an RNA-binding mechanism that depends on the collective arrangement of the two dsRNA-binding surfaces. As shown with the preliminary RMSD and ANM analyses, the function of these dynamics may be to adapt the separation distance and relative orientation of the two dsRNA-binding surfaces with respect to one another so that all of the heterogeneous pri-miRNA-binding targets can be recognized and bound with reasonably similar affinity.

One strength of MD simulations is their ability to reveal correlated dynamics, especially over long distances, through

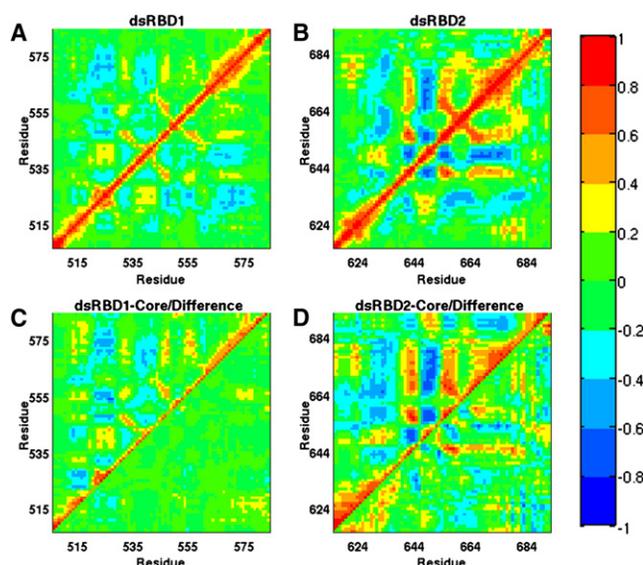


FIGURE 4  $C_{\alpha}$  correlation matrices reveal the collective backbone motions of isolated dsRBD1 (A), isolated dsRBD2 (B), dsRBD1 from the Core simulation (C), and dsRBD2 from the Core simulation (D). To highlight differences between the collective dynamics of the domains in isolation and in the Core, a difference matrix between panels A and C or panels B and D is shown below the diagonal in panels C and D, respectively. The color bar on the right shows the color scale indicating strong positive correlation (red), strong negative correlation (blue), and noncorrelated motion (green) used in the figure.

principal component analysis (45,46) or a variety of related techniques that have been developed to monitor the essential dynamics of selected degrees of freedom that are believed to be the most functionally relevant (47–49). The application of traditional Cartesian essential-dynamics methods requires that the snapshots of the trajectory must first be superimposed to remove the effects of translational and rotational diffusion, and therefore also requires the existence of a unique reference frame for superimposing the entire macromolecule. As was seen in the RMSD traces (Fig. 2), this is not possible for the overall Core, due to the reorientational dynamics of the dsRBDs with respect to each other. For the individual domains, however, removal of global motions is practical, and thus allows collective analysis of the individual site dynamics discussed above.

Analysis of the  $C_{\alpha}$  atomic fluctuations produces the covariance matrices shown in Fig. 4. From Fig. 4, A and B, it is evident that extensive correlations are observed within the individual domains, as assayed in the dsRBD1 and dsRBD2 simulations, respectively. Although the collective dynamics of both domains in the Core simulation are qualitatively similar to those of the isolated domain (upper triangles, Fig. 4, C and D), there are key differences. dsRBD1 shows few differences in the collective dynamics between the Core and isolated simulations (see the difference matrix in the lower triangle of Fig. 4 C). However, there are two notable exceptions: dsRBD1 makes contact with dsRBD2 and H5 in the C-terminus through the S4-S5 loop and the C-terminal

end of  $\alpha$ -helix H2 (Fig. 1 B). These two regions of the primary structure pack against one another in the domain, but their dynamics become significantly more correlated in the context of the Core (Fig. 4 C, lower triangle). Finally, constraint of the C-terminal residues of dsRBD1 by packing in the linker region of the Core strengthens the correlation between the residues at the C-terminus of H2 and those in the N-terminal region of  $\alpha$ -helix H1 (Fig. 4 C, lower triangle).

In contrast to dsRBD1, where the changes in correlated motion between the isolated and Core trajectories are limited to a small number of well-defined sites, the changes in dynamics of dsRBD2 are diffuse and encompass the entire domain (Fig. 4 D, lower triangle). Overall, the collective nature of the dynamics increases in the Core, with positive correlations becoming more positive and anticorrelations more negative, suggesting a global stiffening of the domain. Although these results are informative, they do not address the question of whether the collective dynamics in the Core as a whole produce functionally relevant changes in the character of the dsRNA-binding interface, because motions of the two domains relative to one another cannot be assessed by this method. To access the collective dynamics of the Core computationally, it is necessary to apply a method that analytically removes the effects of global rotational and translational motion without reference to a particular snapshot or average position.

### Isotropically distributed ensemble analysis

Several methods are available for assessing collective dynamics computationally without first removing global rotational motion, including isotropically distributed ensemble (IDE) analysis (50) and reorientational eigenmode dynamics (51,52). These methods have been applied successfully to describe the amplitude and timescale of internal dynamics in the single-domain proteins ubiquitin (50–52), calbindin (53), and RNA binding domain I (RBD1) from U1A (30), and have also provided insight into the collective dynamics of model RNA hairpins (54–56) and the complex between UI snRNA and U1A RBD1 (57).

Here, we choose to apply IDE analysis to the  $C_{\alpha}$  atoms of the DGCR8 Core because it allows us to quantify the collective dynamics in the MD simulation and test whether they are consistent with the ANM predictions. The covariance matrix constructed in IDE can be diagonalized, producing three eigenmodes corresponding to overall rotational motion and  $N-3$  ( $N = 197$ , the number of  $C_{\alpha}$  atoms) modes corresponding to internal motion. Plotting a parameter ( $\kappa$ ) that describes the number of atoms significantly affected by a given mode as a function of the eigenvalues ( $\lambda$ ) for the modes (Fig. 5 A) reveals the qualitative difference between the global and internal modes. The collectivity profile in Fig. 5 A clearly indicates that global and internal motions can be separated into discrete subsets of modes (notice the significant gap between the three largest eigenvalues, corresponding to the

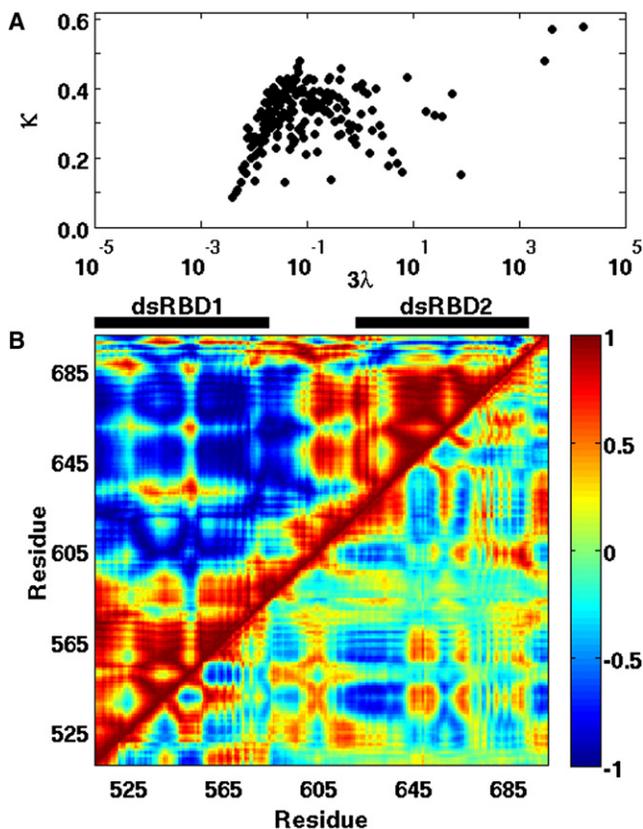


FIGURE 5 IDE analysis reveals motions correlating the conformations of dsRBD1 and dsRBD2 in the Core. (A) Mode collectivity plotted as a function of the eigenvalue for each mode reveals the separation of motion into three global reorientational modes (*upper right*) and the N-3 internal modes. (B) The IDE matrix from the Core simulation is represented by its cross-correlation coefficients (*upper triangle*), demonstrating the anticorrelated nature of the global reorientation of the two dsRBDs with respect to each other. Subtracting the contributions from the three global modes before constructing the cross-correlation coefficients preserves only the effects of internal dynamics (*lower triangle*). The color bar on the right shows the color scale indicating strong positive correlation (*red*), strong negative correlation (*blue*), and noncorrelated motion (*green*) used in the figure.

global modes, and the fourth), allowing reconstruction of a covariance matrix of internal motions from the N-3 internal modes and their eigenvalues. The overall IDE matrix from the Core simulation is shown in the upper triangle of Fig. 5 B, and the internal motion matrix is shown below the diagonal in the lower triangle. The features of the total IDE matrix reveal the presence of strongly anticorrelated motion of each RBD with respect to the other under the influence of the three global eigenmodes. This is consistent with the more-qualitative conclusions drawn from the bundles generated by superimposing each domain individually (Fig. 2, D and E), and confirms that the distance separating the two RNA-binding surfaces of DGCR8 is adapted dynamically on the timescale of these simulations.

Analysis of the internal dynamics of the Core reveals further adjustments made to the relative orientation of the two RNA-binding surfaces with respect to one another.

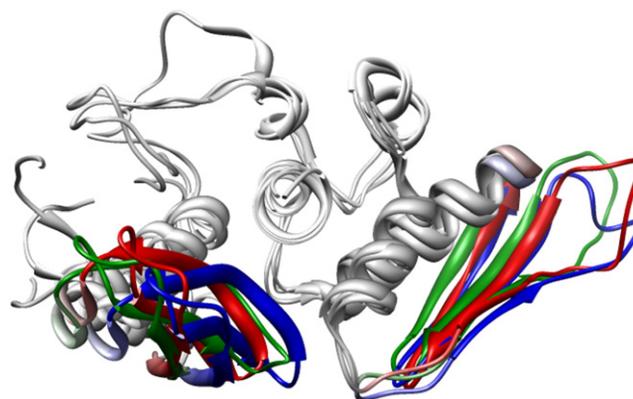


FIGURE 6 Ribbon diagrams spanning the range of motion sampled by projecting the limiting excursions along the largest-amplitude internal eigenmode of the IDE matrix onto the starting crystal structure. Colorized regions highlight the sections of maximal displacement, with dsRBD1 and dsRBD2 colored identically in the given structure to guide the eye.

The lower triangle of Fig. 5 B reveals a complex pattern of correlated and anticorrelated motions spanning the whole protein. Visualization of their effects can be achieved by projecting displacements along the largest-amplitude internal modes onto the starting structure from the simulation. In Fig. 6, the projected structures with the most extreme excursions along the first internal eigenmode of the IDE analysis are represented as ribbons, along with that of the starting conformation. Displacement of the RNA-binding loops is seen, along with a slight twist of the two domains relative to each other. The central  $\alpha$ -helix, on the other hand, remains largely unmoved under the influence of this motion, confirming its role as a pivot for the dynamics of the ensemble. Lastly, the portions of dsRBD1 and dsRBD2 in direct contact with one another are displaced in a correlated way, preserving contact, and further supporting the suggestion that direct domain-domain contacts must be maintained to facilitate the repositioning of the RNA binding surfaces.

### dsRBD domain interface

The results we have discussed so far present an intriguing picture of the impact global motions have on the relative positions of the two RNA-binding surfaces in DGCR8, but up to now we have avoided a detailed discussion of the atomic level interactions that underlie them. As a starting point for investigating the atomistic mechanism of RNA binding by DGCR8, we revisited the four hydrogen-bond interactions between the C-terminal  $\alpha$ -helix (H5) and the dsRBDs presented by Sohn et al. (20) (Fig. 7). Distance calculations indicate that throughout the simulations, the carbonyl oxygen of Ile-575 is an average of 4.5 Å away from the hydroxyl oxygen of Ser-693. A similar average is seen for distance of the amine nitrogen of Arg-630 to the carbonyl oxygen of Met-697 (4.2 Å). A higher average is seen for the amine nitrogen of Arg-630 with the carboxylate oxygen of Glu-701 and the

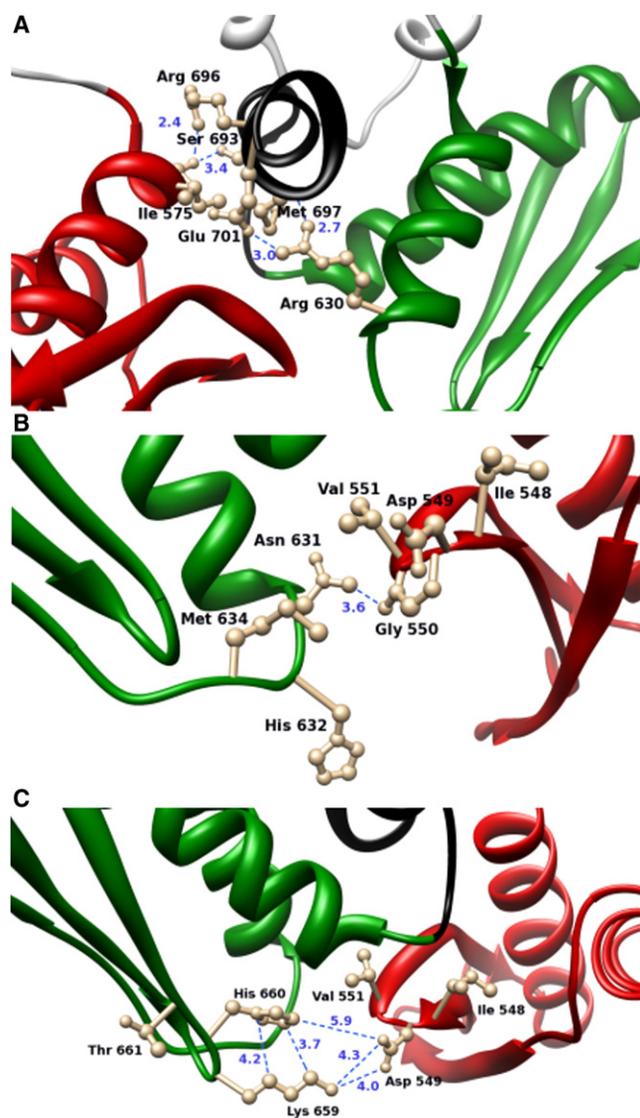


FIGURE 7 Expanded views of the Core crystal structure showing the interfacial interactions between dsRBD1 (red) and dsRBD2 (green). Stabilization of the interface comes from (A) multiple hydrogen bonds formed between the C-terminal helix and either dsRBD1 or dsRBD2; (B) a variety of packing interactions, including a hydrogen bond between Asn-631 and Gly-550; and (C) a salt bridge between Asp-549 and Lys-659. All distances shown are between the indicated pair of atoms in the crystal structure. Their variation in the MD simulations is discussed in the text.

carbonyl oxygen of Ile-575 with the amine nitrogen of Arg-696 (8.2 Å and 7.6 Å, respectively). Therefore, out of the four hydrogen bonds predicted from the crystal structure, only two seem likely to contribute significantly to the stability of the domain-linker interface formed by  $\alpha$ -helix H5. Based on our MD simulations, we conclude that the dynamically preserved hydrogen bonds found between Arg-630 and Met-697, and Ile-575 and Ser-693, are the ones responsible for preserving contact between  $\alpha$ -helix H5 and the dsRBDs.

These results led us to investigate the dynamic stability of other crystallographically observed interactions spanning the domain interface to clarify which are the most involved in

holding the interface together, and which simply appear to be preferred in the temporally and spatially averaged crystallographic model. In particular, we were drawn to the interface between the S4-S4 loop of dsRBD1 and H1' of dsRBD2, both of which undergo correlated displacements in the largest-amplitude internal mode of the IDE matrix. Recall that the S4-S5 loop and the end of helix H2 were also the portions of dsRBD1 that were found to experience the greatest difference in their  $C\alpha$  correlations by Cartesian principal component analysis of the isolated domain and Core simulations. In addition to an extensive van der Waals interface between the dsRBDs, including these regions of dsRBD1, further investigation of the crystal structure reveals two particularly important interactions spanning the dsRBD1-dsRBD2 interface: a hydrogen bond formed by Asn-631 and the backbone of Gly-550 (Fig. 7 B), and a salt bridge formed by Lys-659 and Asp-549 (Fig. 7 C). The hydrogen bond formed by the amide nitrogen of Asn-631 and the carbonyl in the backbone of Gly-550 is predicted from the crystal structure, with a distance of 3.6 Å. Throughout the simulation, an average distance of 5.4 Å is seen between the carbonyl oxygen of Gly-550 and the side-chain nitrogen atom of Asn-631, with a distance of <4.0 Å observed in 35% of the snapshots. This behavior indicates that over the timescale of the simulation, the hydrogen bond is retained, although dynamic excursions allow it to intermittently break. Additional van der Waals contacts between the side chain of Val-551 and various residues in dsRBD1 further strengthen this interface.

The interface between dsRBD1 and dsRBD2 is capped by a salt bridge between Lys-659 and Asp-549, as evidenced by the 4 Å approach of the charged moieties toward each other in the crystal structure (Fig. 7 C). The face of the imidazole ring of His-660 is oriented to provide packing interactions for the aliphatic chain of the Lys-659 side chain, further stabilizing its approach toward Asp-549. Additionally, the imidazole ring's NH moiety is 5.9 Å away from the carboxylate of Asp-549 in the crystal, reinforcing the stabilizing transdomain interactions involving this residue. An early event in the simulation reorients the S4-S5 hairpin, moving Lys-659 and Asp-549 more than 15 Å apart from each other, but they return to an average separation of 7.9 Å by the 60 ns mark, with an approach of <5 Å observed in 22% of the snapshots, and remain oriented to maintain an electrostatic interaction for the remainder of the trajectory. The salt bridge between Asp-549 and Lys-659 is likely to be vital for the protein's function and structure, given that the nearby His-660 provides stability and has a pKa that can be titrated near physiological conditions to enhance stability.

## CONCLUSIONS

We have presented the first MD simulations of the DGCR8 Core, a dsRNA-binding protein that contains two dsRBDs in tandem. Our results show that the dsRBDs are connected

through an extensive network of interactions in the dsRBD1-dsRBD2 interface and in the dynamically integrated interface of each domain with residues from the well-folded linker region and C-terminal  $\alpha$ -helix. These interactions, including key hydrogen bonds and salt bridges, are consistent with the initial configuration in the crystal structure, and are largely preserved in a 250 ns all-atom MD trajectory. Moreover, our results show that these interfacial connections are vital for maintaining the dynamics of the protein as a whole, which serves to fine-tune the distance separating the two dsRNA-binding surfaces and their orientations with respect to one another.

More importantly, our results demonstrate which residues might be effectively targeted in mutagenesis studies aimed at disrupting the collective dynamics of the Core under the hypothesis that this will reduce the functional benefits of pre-organizing the dsRNA-binding site. It is predicted that loosening of the transdomain interfaces will reduce the affinity of the Core for pri-miRNA because it would impede the cooperative function of the two domains. Additionally, such a disruption may render the protein less competent to bind pri-miRNA transcripts preferentially over other dsRNA molecules with different structures, providing insight into how DGCR8 selects pri-miRNA from the complex nuclear pool of partially dsRNAs.

We thank Prof. Phil Bevilacqua and Mr. Wayne Mullinax for helpful discussions.

This work was supported by startup funds from Pennsylvania State University.

## REFERENCES

- Wang, W. X., B. W. Rajeev, ..., P. T. Nelson. 2008. The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of  $\beta$ -site amyloid precursor protein-cleaving enzyme 1. *J. Neurosci.* 28:1213–1223.
- Friedman, R. C., K. K. H. Farh, ..., D. P. Bartel. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19:92–105.
- Stanczyk, J., D. M. Pedrioli, ..., D. Kyburz. 2008. Altered expression of microRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. *Arthritis Rheum.* 58:1001–1009.
- Kennerdell, J. R., and R. W. Carthew. 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell.* 95:1017–1026.
- Timmons, L., D. L. Court, and A. Fire. 2001. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene.* 263:103–112.
- Jinek, M., and J. A. Doudna. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature.* 457:405–412.
- Shiohama, A., T. Sasaki, ..., N. Shimizu. 2003. Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochem. Biophys. Res. Commun.* 304:184–190.
- Gregory, R. I., K. P. Yan, ..., R. Shiekhattar. 2004. The microprocessor complex mediates the genesis of microRNAs. *Nature.* 432:235–240.
- Varani, G., and K. Nagai. 1998. RNA recognition by RNP proteins during RNA processing. *Ann. Rev. Biophys. Biomol. Struct.* 27:407–445.
- Nowotny, M., and W. Yang. 2009. Structural and functional modules in RNA interference. *Curr. Opin. Struct. Biol.* 19:286–293.
- Patel, R. C., and G. C. Sen. 1998. PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J.* 17:4379–4390.
- St Johnston, D., N. H. Brown, ..., M. Jantsch. 1992. A conserved double-stranded RNA-binding domain. *Proc. Natl. Acad. Sci. USA.* 89:10979–10983.
- Ryter, J. M., and S. C. Schultz. 1998. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* 17:7505–7513.
- Tian, B., P. C. Bevilacqua, ..., M. B. Mathews. 2004. The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.* 5:1013–1023.
- Fierro-Monti, I., and M. B. Mathews. 2000. Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem. Sci.* 25:241–246.
- Manche, L., S. R. Green, ..., M. B. Mathews. 1992. Interactions between double-stranded RNA regulators and the protein kinase DAI. *Mol. Cell. Biol.* 12:5238–5248.
- Saunders, L. R., and G. N. Barber. 2003. The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J.* 17:961–983.
- Bevilacqua, P. C., and T. R. Cech. 1996. Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry.* 35:9983–9994.
- Han, J. J., Y. Lee, ..., V. N. Kim. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell.* 125:887–901.
- Sohn, S. Y., W. J. Bae, ..., Y. Cho. 2007. Crystal structure of human DGCR8 core. *Nat. Struct. Mol. Biol.* 14:847–853.
- Nanduri, S., B. W. Carpick, ..., J. Qin. 1998. Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J.* 17:5458–5465.
- Yeom, K. H., Y. Lee, ..., V. N. Kim. 2006. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res.* 34:4622–4629.
- Bahar, I., C. Chennubhotla, and D. Tobi. 2007. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* 17:633–640.
- Ramos, A., P. Bayer, and G. Varani. 1999–2000. Determination of the structure of the RNA complex of a double-stranded RNA-binding domain from *Drosophila* Staufen protein. *Biopolymers.* 52:181–196.
- Gan, J. H., J. E. Tropea, ..., X. Ji. 2006. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell.* 124:355–366.
- Wu, H. H., A. Henras, ..., J. Feigon. 2004. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc. Natl. Acad. Sci. USA.* 101:8307–8312.
- Mackerell, Jr., A. D., and L. Nilsson. 2008. Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.* 18:194–199.
- Castrignano, T., G. Chillemi, ..., A. Desideri. 2002. Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals dynamic features of the intermolecular interface and its hydration. *Biophys. J.* 83:3542–3552.
- Reyes, C. M., and P. A. Kollman. 1999. Molecular dynamics studies of U1A-RNA complexes. *RNA.* 5:235–244.
- Showalter, S. A., and K. B. Hall. 2002. A functional role for correlated motion in the N-terminal RNA-binding domain of human U1A protein. *J. Mol. Biol.* 322:533–542.
- Hermann, T., and E. Westhof. 1999. Simulations of the dynamics at an RNA-protein interface. *Nat. Struct. Biol.* 6:540–544.

32. Eyal, E., L. W. Yang, and I. Bahar. 2006. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*. 22:2619–2627.
33. Case, D. A., T. E. Cheatham, 3rd, ..., R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
34. Hornak, V., A. Okur, ..., C. Simmerling. 2006. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*. 103:915–920.
35. Hornak, V., R. Abel, ..., C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 65:712–725.
36. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
37. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
38. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60:2126–2132.
39. Showalter, S. A., and R. Brüschweiler. 2007. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. *J. Chem. Theory Comput.* 3:961–975.
40. Pettersen, E. F. G., T. D. Goddard, ..., T. E. Ferrin. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.
41. Riccardi, D., Q. Cui, and G. N. Phillips. 2009. Application of elastic network models to proteins in the crystalline state. *Biophys. J.* 96:464–475.
42. Yang, Z., P. Majek, and I. Bahar. 2009. Allosteric transitions of supramolecular systems explored by network models: application to Chaperonin GroEL. *PLoS Comp. Biol.* 5:e10000360.
43. Berezikov, E., V. Guryev, ..., E. Cuppen. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*. 120:21–24.
44. Zeng, Y., and B. R. Cullen. 2003. Sequence requirements for micro RNA processing and function in human cells. *RNA*. 9:112–123.
45. Kitao, A., and N. Go. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* 9:164–169.
46. Koyama, Y. M., T. J. Kobayashi, ..., H. R. Ueda. 2008. Perturbational formulation of principal component analysis in molecular dynamics simulation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78:046702.
47. Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins*. 17:412–425.
48. Berendsen, H. J. C., and S. Hayward. 2000. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* 10:165–169.
49. Antoniou, D., J. Basner, ..., S. D. Schwartz. 2006. Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem. Rev.* 106:3170–3187.
50. Prompers, J. J., and R. Brüschweiler. 2002. Dynamic and structural analysis of isotropically distributed molecular ensembles. *Proteins*. 46:177–189.
51. Prompers, J. J., and R. Brüschweiler. 2001. Reorientational eigenmode dynamics: a combined MD/NMR relaxation analysis method for flexible parts in globular proteins. *J. Am. Chem. Soc.* 123:7305–7313.
52. Prompers, J. J., and R. Brüschweiler. 2002. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124:4522–4534.
53. Showalter, S. A., E. Johnson, ..., R. Brüschweiler. 2007. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J. Am. Chem. Soc.* 129:14146–14147.
54. Showalter, S. A., N. A. Baker, ..., K. B. Hall. 2005. Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics. *J. Biomol. NMR*. 32:179–193.
55. Musselman, C., S. W. Pitt, ..., H. M. Al-Hashimi. 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.
56. Musselman, C., H. M. Al-Hashimi, and I. Andricioaei. 2007. iRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys. J.* 93:411–422.
57. Showalter, S. A., and K. B. Hall. 2005. Correlated motions in the U1 snRNA stem/loop 2:U1A RBD1 complex. *Biophys. J.* 89:2046–2058.