Interpreting the Dynamics of Binding Interactions of snRNA and U1A Using a Coarse-Grained Model

Zhongjie Han,[1] Qi Shao,[1] Weikang Gong,[1] Shihao Wang,[1] Jiguo Su,[3] Chunhua Li,[1,2,*] and Yang Zhang[2,*]
[1]College of Life Science and Bioengineering, Beijing University of Technology, Beijing, China; [2]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan; and [3]College of Science, Yanshan University, Qinhuangdao, China

ABSTRACT   The binding interactions of small nuclear RNAs (snRNA) and the associated protein factors are critical to the function of spliceosomes in alternatively splicing primary RNA transcripts. Although molecular dynamics simulations are a powerful tool to interpret the mechanism of biological processes, the atomic-level simulations are, however, too expensive and with limited accuracy for the large-size systems, such as snRNA-protein complexes. We extend the coarse-grained Gaussian network model, which models the RNA-protein complexes as a harmonic chain of Cα, P, and O4′ atoms, to investigating the impact of the snRNA-binding interaction on the dynamic stability of the human U1A protein, which is a major component of the spliceosomal U1 small nuclear ribonucleoprotein particle. The results reveal that the first and third loops and the C-terminal helix regions of the U1A domain undergo a significant loss of flexibility upon the RNA binding due to the forming of mostly electrostatic and hydrogen bond interactions with RNA 5′ stem and loop. By examining the residues whose mutations significantly change the binding free energy between U1A and snRNA, the Gaussian network model-based calculations show that not only the residues at the binding sites that are traditionally considered to play a major role in U1A-RNA association but also those residues that are far away from the RNA-binding interface can participate in the long-range allosteric signal transmission; these calculations are quantitatively consistent with the data observed in the recent snRNA binding experiments. The study demonstrates a useful avenue to utilize the simplified elastic network model to investigate the dynamics characteristics of the biologically important macromolecular interactions.

INTRODUCTION

Protein-RNA interactions are critical to many important cellular processes, such as the regulation of gene expression, protein synthesis, and ribosome assembly and virus replication (1–3). The human U1A protein is one of the major protein components of the spliceosomal U1 small nuclear ribonucleoprotein particle (4,5). Its N-terminal RNA-binding domain (RBD1) binds to stem/loop II (SL2) of U1 small nuclear RNA (snRNA), participating in pre-messenger RNA splicing (6). It has been found that RBD1 of U1A protein can undergo a large conformational change upon binding to its target RNA, and any perturbation to the structure or dynamics of RBD1 would be propagated through the entire binding surface, leading to significant changes in RNA-binding affinity and specificity (7). Hence, the identification of the key residues of U1A RBD1 associated with SL2 RNA-binding and conformational transitions can be of important help for our understanding of the recognition mechanism between U1A RBD1 and SL2 RNA (hereinafter referred to as U1A and RNA) and the underlying alternative splicing process (8–10).

The high resolution crystal structure of U1A bound to RNA has been determined at 1.92 Å resolution (Fig. 1) (11). Here, the U1A domain involves a single polypeptide chain of 97 residues, characterized by a β_1-α_A-β_2-β_3-α_B-β_4 secondary structure that adopts a four-stranded antiparallel β-sheet as the primary RNA-binding surface supported by two α-helices (12,13). Before binding, U1A folds into the closed form in which the C-terminal helix (Helix-C, consisting of residues 90–97) covers part of the RNA-binding surface. Upon binding to RNA, U1A folds into the open form, Helix-C moves away uncovering the buried area to permit RNA access, and loop3 protrudes through the RNA loop (4,14). These changes strengthen the binding affinity by generating a series of hydrogen-bonding and stacking interactions.

Many experimental and theoretical researches focus on the binding dynamics and key residues involved in

FIGURE 1 Complex structure of U1A with SL2 RNA (PDB: 1URN) (*a*) and the secondary structure of RNA (*b*). U1A forms a β1-αA-β2-β3-αB-β4 secondary structure with β1 (Thr11-Asn15), β2 (Ile40-Val45), β3 (Ala55-Phe59), β4 (Arg83-Tyr86), αA (Lys23-Phe37), and αB (Val62-Met72).

U1A-RNA complex formation. For instance, using energetic pairwise coupling and residue mutation experiments, Kranz et al. found that RNA binding can mediate the local thermodynamic coupling between U1A interface β-sheets and the Helix-C (15) and across the RNA-binding surface of U1A (9). Law et al. later used a surface plasmon resonance-based biosensor to gain mechanistic insight into the role of Helix-C in mediating the interaction with RNA (16). On the theoretical side, all-atomic molecular dynamics (MD) simulation has been a widely used computational tool to explore protein structure dynamic features in atomic details and functionally key sites (17,18). For example, Reyes and Kollman carried out MD simulations and computational mutagenesis on protein and RNA residues in the U1A-RNA-binding surface to examine the origin of the binding specificity (19). Pitici et al. utilized MD simulations to obtain predictions of the structures for the unbound forms of U1A in solution and to elucidate the dynamical aspects of the induced fit upon RNA binding (20). Law et al. performed MD simulations of the wild-type and mutant U1A proteins in the absence of RNA to examine whether structural rearrangements could explain differences in their association rates with RNA and identified several key residues in complex formation and stability (21). Most recently, Guzman et al. adopted a series of MD simulations to delve into whether U1A protein alone is capable of undergoing the conformational dynamics similar to structural rearrangements upon RNA binding (13). Despite the success, the atomic-level simulation is often found too expensive and time consuming to study the large-scale conformational motion, especially when the size of the molecular system increases (in which a snRNA-protein complex typically consists of thousands of atoms). Meanwhile, the limited accuracy of the physics-based force field in the MD simulation has been a bottleneck for describing the subtle atomic interactions between RNA and protein because any atomic interaction errors can dramatically impact the distribution of the MD trajectories (17,22).

To address these issues, several coarse-grained models have been recently proposed to study the question of large-scale allosteric transitions. Among them, the elastic network model (ENM) has been shown to be a particularly effective computational technique for investigating the function-relevant motions of proteins and even RNAs at a coarse-grained level (23–27). In the conventional ENM, a protein structure is modeled as an elastic network of Cα atoms, in which the residue pairs within a given cutoff distance are considered to have interactions and are connected by a set of Hookean springs with a uniform force constant (28). Generally, the low-frequency motion modes obtained by ENM represent the large-scale collective motions relevant to the functions of the protein, whereas the high-frequency motion modes reflect the geometric irregularity in the protein structure (29–31). The ENM model has been successfully used to investigate the large-scale functional motions of several protein systems, including GroEL, GroEL-GroES complex, and calcium ATPase (32,33).

In this work, we examine the feasibility to extend the ENM model to interpret the dynamics of the binding interactions between snRNA and the human spliceosomal U1A domain. For this purpose, we make the use of a specific case of the ENM method, the Gaussian network model (GNM) in which residue nodes are assumed to undergo Gaussian-distributed fluctuations about their mean positions, to analyze the effects of RNA binding on the dynamics of U1A protein based on the residue fluctuations and movement coupling between them. To deduce the detailed binding structure, we further extend the GNM-based thermodynamic method to the identification of the residue sites of the U1A domain, which are functionally indispensable for the RNA-binding and conformational changes.

MATERIALS AND METHODS

GNM

GNM is a special case of ENM, in which a biomolecule structure is modeled as a coarse-grained and elastic network by replacing one residue with several nodes (here, C_α atom for protein residue (32) and P and O4′ atoms for RNA nucleotide (34)) and imposing a harmonic potential with a uniform spring force constant between the nodes that are sufficiently close to lie within a given cutoff distance (35). By this simplification, the total internal potential energy of the network of N nodes can be written as

$$H = \frac{1}{2}\gamma\left[\Delta R^T(\Gamma \otimes E)\Delta R\right], \tag{1}$$

where γ is the harmonic force constant of the springs, the column vector ΔR represents the fluctuation of the N nodes, the superscript T denotes the transpose, E is the 3×3 identity matrix, $\otimes$ is the matrix direct product, and Γ is the $N \times N$ symmetric Kirchhoff matrix, the elements of which are described as

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum\limits_{i,j \neq i} \Gamma_{ij} & \text{if } i = j \end{cases}, \tag{2}$$

where r_c is the cutoff distance, and R_{ij} is the distance between the ith and jth nodes. It should be noted that the cutoff distances for nodes in a protein and an RNA are set to 8.0 and 9.0 Å respectively, and 9.0 Å for the protein-RNA complex interface node pairs (30).

The mean-square fluctuation of each node and the cross correlation fluctuations between different nodes are in proportion to the diagonal and off-diagonal elements of the pseudoinverse of the Kirchhoff matrix. The inverse of the Kirchhoff matrix can be decomposed as

$$\Gamma^{-1} = U \, \Lambda^{-1} U^{\mathrm{T}}, \tag{3}$$

where U is an orthogonal matrix whose columns u_i ($1 \leq i \leq$ N) are the eigenvectors of Γ, and Λ is the diagonal matrix of the eigenvalues λ_i of Γ. The fluctuation cross correlation between the ith and jth nodes and the mean-square fluctuation of the ith node can be written as

$$<\Delta R_i \bullet \Delta R_j> = \frac{3k_BT}{\gamma}\sum_{k=2}^{N}\lambda_k^{-1}[u_k]_i[u_k]_j, \tag{4}$$

$$<\Delta R_i \bullet \Delta R_i> = \frac{3k_BT}{\gamma}\sum_{k=2}^{N}\lambda_k^{-1}[u_k]_i^2, \tag{5}$$

where k_B is the Boltzmann constant, T is the absolute temperature, and the meaning of γ is the same as in Eq. 1. According to the Debye-Waller theory, the B-factor of the ith node can be calculated with the expression

$$B_i = 8\pi^2 <\Delta R_i \bullet \Delta R_i> /3. \tag{6}$$

The cross correlation is normalized as

$$C_{ij} = \frac{\left\langle \Delta R_i \bullet \Delta R_j \right\rangle}{\left[\left\langle \Delta R_i^2 \right\rangle \times \left\langle \Delta R_j^2 \right\rangle\right]^{\frac{1}{2}}}. \tag{7}$$

Generally, in the calculation of the normalized cross correlation, the contributions of the first few lowest-frequency normal modes (with the largest contributions to the residual displacements) are considered in Eq. 7 where the cross correlation between two nodes and their mean-square fluctuations are summed up, respectively, for the chosen normal modes as shown in Eqs. 4 and 5.

Additionally, based on the theory of statistical physics, the vibrational entropy of the system can be defined as (29)

$$S = \frac{\langle H \rangle - F}{T} = \frac{3}{2}(N-1)k_B + k_B \ln Z, \tag{8}$$

where $\langle H \rangle$ is the average vibrational Hamiltonian, $F = -k_BT \ln Z$ is the vibrational Helmholtz free energy, Z is the configurational integral part of the vibrational partition function given by $Z = \int \exp(- H/k_BT)d\{\Delta R\}$, N is the number of nodes, and the meaning of T, and k_B is the same as Eq. 4. When a perturbation is introduced to the force constant of the spring connecting the ith and jth nodes, the change of entropy can be written as

$$\Delta S = -\frac{\partial S}{\partial \gamma_{ij}}\Delta\gamma_{ij}, \tag{9}$$

where the negative sign represents the decrease of the force constant, and $\Delta\gamma_{ij}$ is the change of the force constant when a perturbation is introduced. Experimentally, when a residue is mutated to Ala, the interactions of the residue with other ones, and therefore the overall fold stability of the proteins, are reduced in most cases (36). Thus, to appropriately model such characteristics of mutations, the force constants of the springs connecting to the mutated residue are reduced in our method. According to Eq. 8, the derivation of the entropy S with respect to γ_{ij} is written by (25,37)

$$\frac{\partial S}{\partial r_{ij}} = -\frac{1}{2T}\left(\left\langle(\Delta R_i)^2\right\rangle + \left\langle(\Delta R_j)^2\right\rangle - 2\left\langle\Delta R_i \bullet \Delta R_j\right\rangle\right). \tag{10}$$

Thermodynamics cycle method for the identification of functional residues

The GNM-based thermodynamic method was previously proposed for identifying the functionally key residues in protein-small ligand interactions (37). Here, we extended it to protein-RNA interactions, in which the functional residues are identified as those whose mutations cause a significantly large change in the binding free energy between U1A and RNA.

As illustrated in Fig. 2, ΔG_1 denotes the binding free energy between the wild-type U1A and RNA, and ΔG_2 represents the binding free energy after a residue mutation is introduced to the receptor. We want to calculate the change of the binding free energy after residue mutation, i.e.,

$$\Delta\Delta G = \Delta G_2 - \Delta G_1. \tag{11}$$

Calculating the binding free energy ΔG_1 and ΔG_2 directly is hard because of the intricacy of the interactions between the protein and RNA. Therefore, we constructed a thermodynamic cycle to calculate the change of the binding free energy. We supposed two unphysical processes, protein → protein′ and complex → complex′, and the corresponding free energy changes were depicted as ΔG_3 and ΔG_4. Considering the free energy of the system is a state function, we can calculate its change between its initial and final states regardless of the path between them. Thus, Eq. 11 can be expressed as

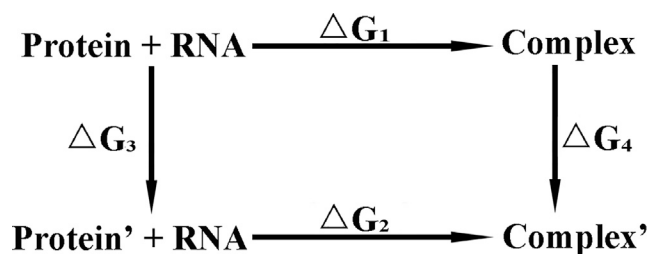$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3. \tag{12}$$

$$\text{Protein + RNA} \xrightarrow{\triangle G_1} \text{Complex}$$

$$\downarrow \triangle G_3 \qquad\qquad\qquad \downarrow \triangle G_4$$

$$\text{Protein' + RNA} \xrightarrow{\triangle G_2} \text{Complex'}$$

FIGURE 2 Diagram of thermodynamic cycle. $\varDelta G_1$ denotes the binding free energy between the wild-type U1A and RNA, and $\varDelta G_2$ represents the binding free energy after a residue mutation is introduced to the receptor, in which the mutation is marked by the apostrophe. protein → protein' and complex → complex' are the two nonphysical processes, and the corresponding free energy changes are depicted as $\varDelta G_3$ and $\varDelta G_4$.

In this method, the residue mutations are introduced through reducing the force constant of all the springs connecting to the mutated residues. For a certain pair of nodes i and j, which form a contact both in the receptor and complex structures, the perturbation of the force constant γ_{ij} will lead to the same change of the potential energy (i.e., $\varDelta U_3 = \varDelta U_4$). Thus,

$$\varDelta \varDelta G = -T(\varDelta S_4 - \varDelta S_3). \tag{13}$$

The entropy change caused by the spring perturbation can be calculated via Eqs. 9 and 10, and then the value of $\varDelta \varDelta G$ can be obtained by Eq. 13. When a mutation is introduced to a residue, all the springs connecting to this residue should be perturbed simultaneously. In this way, the $\varDelta \varDelta G$ value caused by a residue mutation is the sum of the binding free energy changes for the perturbations of all the springs involved in this residue. In the process of identifying the key residues, every amino acid residue is mutated in the same way described above, and no special treatment is made to any amino acid residue. Additionally, the residue mutation may result in an increase or a decrease in binding free energy. The absolute value of $\varDelta \varDelta G$ is adopted here. The functionally key residues are those whose mutations cause a significantly large change in the binding free energy. It should be pointed out that the first 30 slowest normal modes were taken into account for the protein, and for the complex, the motion modes were projected onto these 30 slowest modes.

Receptor and complex systems

The crystal structure of U1A-RNA with Protein Data Bank (PDB): 1URN (Fig. 1) (11) was used to construct the GNM of the complex structure, and the GNM of the receptor was built based on the structure with RNA eliminated from the complex.

RESULTS

Theoretical B-factors of residue nodes in the GNM of complex structure

The complex structure was modeled as the GNM system with each amino acid residue in the protein simplified as a node (i.e., its C_α atom), each nucleotide residue two nodes (i.e., P and O4′ atoms), and the interaction network formed by the adjoining node contacts. B-factors of the complex residues can be acquired theoretically (see Materials and Methods). As the absolute value of the spring constant γ does not affect the relative size of residue fluctuations, it

has no influence on the correlation between the calculated and experimental B-factors, cross correlations between residue fluctuations, and the key residue identification (38). Therefore, $\gamma = 1$ was adopted here (30,39). The correlation coefficient between the calculated and experimental B-factors is 0.857, as shown in Fig. 3, the good agreement indicating that this simplified model is constructed reasonably and can be applied to the following analyses about cross correlations of residue fluctuations.

Role of RNA binding on the dynamics of U1A protein

The RNA binding can lead to a large conformational change of U1A, which is important for the specific protein-RNA recognition (15). To examine the specific role of RNA binding on the dynamics of the system, we computed the mean-squared fluctuation (MSF) of each residue of U1A in the GNM with and without RNA based on Eq. 5 (see Fig. 4).

It can be seen from Fig. 4 that the residues in U1A loop1 (residues 16–22) and loop3 (residues 46–54), both located at the RNA-binding surface, have a strikingly higher MSF in the RNA-free state than in the RNA-binding state. Additionally, loop3 undergoes a more flexible loss as it forms more interactions with RNA nucleotides (11). This agrees with the complex crystal structure where U1A loop3 protrudes into the RNA loop with the formation of many hydrogen-bonding and hydrophobic interactions, such as Ser48-C15, Leu49-A6, Gln54-G9, and Phe56-A11, whereas loop1 only contacts with nucleotides A2, U3, and C4 from 5′ stem through mostly side chain interactions, leaving the backbone of loop3 less conformationally constrained (40). Similar MSF changes have also been observed in previous
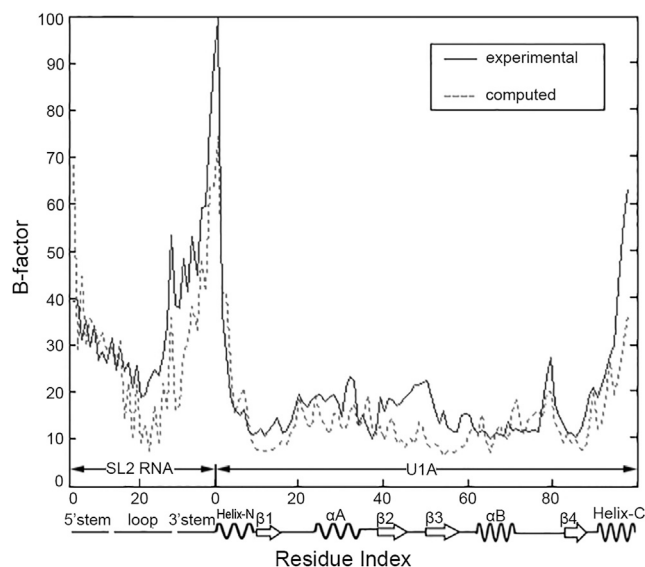

FIGURE 3 Comparison between the experimental (*solid line*) and computed (*dashed line*) B-factors of P, O4′, and C_α atoms of U1A-RNA complex.
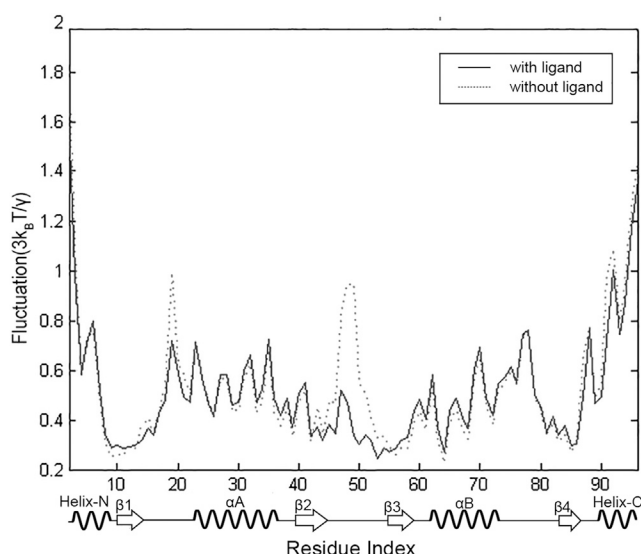
FIGURE 4 Fluctuations of residues of U1A in the RNA-bound state (*solid line*) and the RNA-free one (*dashed line*), respectively.

MD simulations and experimental studies (20), suggesting that the RNA binding indeed plays an important role in stabilizing the binding surface of U1A, especially for the loop1 and loop3 regions (41). This is consistent with the classic induced fit model that loop1 and loop3 play vital roles as U1A anchors to RNA bases (42).

Fig. 4 also shows that the N-terminal helix (Helix-N) and Helix-C residues have a relatively high flexibility in both RNA-binding and RNA-free states; this is consistent with the observation in structural and NMR dynamics studies (43,44). Interestingly, the residues in Helix-C have a small flexibility loss after RNA binding. This is understandable because in the absence of RNA, Helix-C is often partially or fully unfolded. With the recognition of RNA by the U1A protein, however, Helix-C reorients to permit RNA access. This accommodation can result in the formation of specific hydrogen-bonding interactions between U1A and RNA, such as Asp90-C12, Ser91-A11, and Asp92-C12, stabilizing the U1A-RNA complex (16,42,45).

In summary, the RNA binding leads to the large flexibility changes in U1A loop1 and loop3 as well as Helix-C regions because of the formation of a series of specific interactions with RNA target, which is important for the induced fit in U1A-RNA recognition and the high stability of the complex structure.

Movement coupling between residues in complex and in the U1A protein

To detect the movement coupling between residues in U1A-RNA complex, we calculated the cross correlations between residue fluctuations in complex based on Eq. 7. Fig. 5 *a* gives the results obtained based on the first 15 slowest motional modes. The cross correlation between fluctuations can reveal
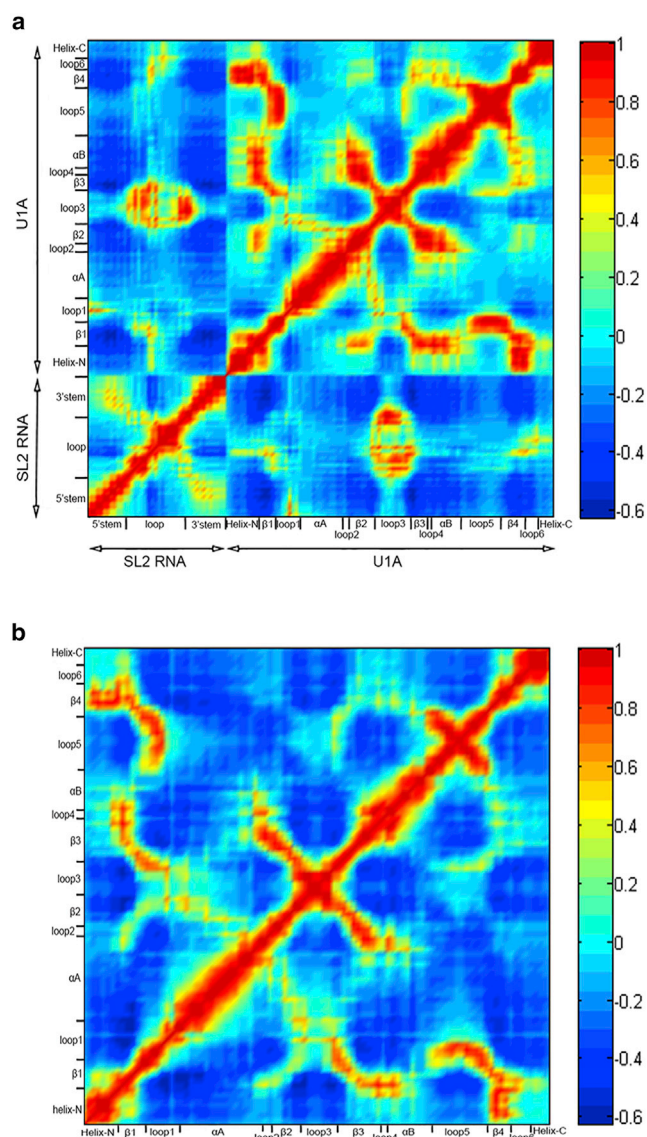

FIGURE 5 Fluctuation cross correlation calculated using the dominant 15 low motional modes contributing 56% and 52% to residue fluctuations for U1A-RNA complex (*a*) and protein U1A (*b*), respectively. As shown in the color bar, the blue regions in the figure indicate negative correlation and the green-yellow-red regions represent positive correlation. To see this figure in color, go online.

the information related to the spatial motion and interactions, with the values ranging from −1 to 1. The positive values indicate that the residues move in the same direction, and the negative ones indicate that they move in the opposite direction. The higher the absolute value is, the more the two residues are correlated. The value $C_{ij} = 0$ means that the motions of residues are completely uncorrelated.

For RNA in complex, the fluctuation of the single-stranded loop region is found to be anticorrelated with the stem region, yet the 5′ and 3′ stem nucleotides are medium positive correlated, which is consistent with the results of MD simulations (42,46,47).

For the two partners in complex, a strongly positive correlation is found between the RNA loop (nucleotides 6–15) and the U1A loop3 (residues 46–54) regions. The 5′ and 3′ stem parts near the loop have weak positive correlations with U1A loop3. These results can be attributed to the formed interactions between these regions in the x-ray crystal structure (11,46). For example, U1A Arg52 forms hydrogen bonds with RNA A6 and G16, U1A Ser46 and Leu49 form water-mediated interactions with G16, and U1A Ser48 forms hydrophobic interactions with C15 (42). Additionally, U1A loop1 (residues 16–22) shows a modest positive correlation with the RNA loop region (nucleotides 6–15) and 5′ stem (nucleotides 1–5). From the crystal structure, the interactions between them include the following: U1A Glu19 forms hydrogen bonds with RNA U7, and U1A Lys20 and Lys22 interact with the sugar-phosphate backbone of 5′ stem (46,47). The above analyses on the movement coupling between U1A loop1 and loop3 and the RNA loop region are consistent with the result that U1A loop1 and loop3 have a remarkable decrease in residue fluctuations when RNA binding occurs (see the Role of RNA binding on the dynamics of U1A protein), which is the reflection of some nonbonded interactions formed between them. Besides with U1A loop1 and loop3, the RNA loop (nucleotides 6–15) also forms more or less positive correlations with the complex interface part constructed by U1A β-sheets and their nearby regions, including loop5 (residues 73–82), loop6 (residues 87–89), and Helix-C (residues 90–97). From the complex crystal structure, the four β-sheets are in an antiparallel arrangement on the interface with hydrophobic residues in them packed on each other and forming a hydrophobic core that links the four β-sheets and is involved in extensive contacts with the RNA (42). For their nearby regions, U1A Ser91 and Asp92 (both in Helix-C) hydrogen bonds with C12, and Thr89 (in loop6) and Ser90 (in Helix-C) hydrogen bonds with A11 (46). These can explain the positive correlations between the RNA loop and U1A β-sheets and the nearby regions.

For U1A protein, to reveal the effect of RNA binding on movement coupling between U1A residues, the cross correlations between residue fluctuations in RNA-free U1A were also calculated (see Fig. 5 b) for comparison. From Fig. 5, a and b, for RNA-free U1A, β1 is partially positively correlated with β2 and positively correlated with β3, and β4 is anticorrelated with β2 and weakly positively correlated with β3, whereas the four β-sheets turn into positive correlations with each other upon RNA binding, indicating that RNA binding strengthens the interactions among β-sheets, which has been verified by Nagai et al (12). Additionally, Helix-C is anticorrelated with β1 and β2 and has weakly positive correlations with β4 and β3 when RNA is absent. However, upon binding with RNA, Helix-C forms more or less positive correlations with β-sheets, meaning that RNA binding mediates local motional cooperativity between Helix-C and the main binding surface β-sheets.

As a whole, RNA binding strengthens the interactions among β-sheets of U1A with positive correlations formed among U1A β-sheets and between them and RNA loop and mediates local motional cooperativity between Helix-C and the main binding surface β-sheets. Additionally, in the complex structure, U1A loop3 and loop1 are strongly and modestly positively correlated with the RNA loop, respectively, indicating their contribution to the induced fit in U1A-RNA recognition.

Identification of functionally key residues in U1A protein

To identify the U1A key residues associated with RNA binding and conformational transition, each residue of the protein and complex systems was mutated, and the change of the RNA-binding free energy value $\Delta\Delta G$ in response was calculated utilizing the method described in the Materials and Methods. The residues with relatively high $\Delta\Delta G$ values were taken as key residues. The calculation results are shown in Fig. 6 a. From this figure, there are 14 clusters of residues (marked in Fig. 6 a) with high $\Delta\Delta G$ values, and the residue with the maximal $\Delta\Delta G$ value in each cluster is defined as the central residue of the corresponding cluster. These 14 residual clusters are, respectively, centered at
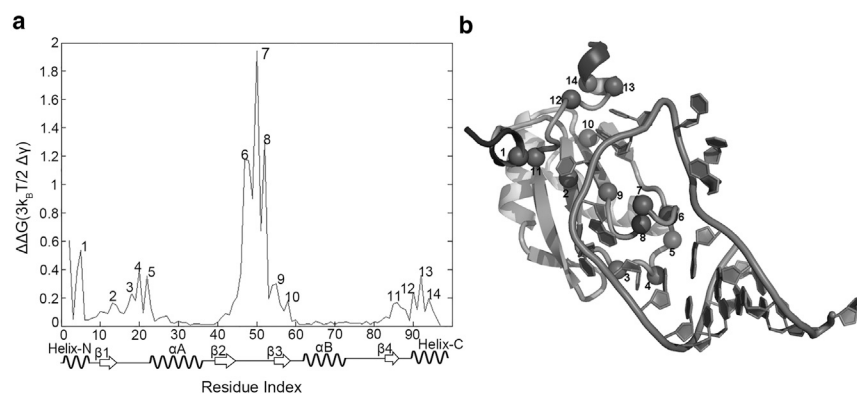

FIGURE 6 Identified functionally key residue clusters. (a) Binding free energy changes $\Delta\Delta G$ in response to residue mutations are shown. The cluster of key residues with relatively high $\Delta\Delta G$ values are marked by the numbers 1–14. (b) Locations of the central residues for the 14 clusters of key residues are shown.

Glu5, Tyr13, Asn18, Lys20, Lys22, Arg47, Lys50, Arg52, Ala55, Ile58, Tyr86, Asp90, Asp92, and Ile94. For clarity, the central residues are mapped on the tertiary structure of the complex system, as shown in Fig. 6 b. According to their positions, these 14 residue clusters are classified into three groups. The predicted key residues are compared with the available experimental and theoretical data, and the functional information of them will be discussed in detail below.

Residues at β-sheets

As shown in Fig. 6 b, there are four residual clusters 2, 9, 10, and 11 at the β-sheets (constructing the main interface), in which the former is located in β1, the middle two in β3, and the last one at the end of β4. Using the electrophoretic mobility shift method, Lutz-Freyermuth et al. have confirmed Phe59 in cluster 10 is essential for RNA binding and U1A-RNA interaction (48). Substitution of Tyr86 in cluster 11 with either Phe or Thr can result in a measurable reduction of complex stability (49). The x-ray crystallographic and NMR studies have identified Tyr13 in cluster 2 and Phe56 in cluster 9 as the two key aromatic residues that interact directly with RNA by forming aromatic stacking with the bases of C10 and A11, respectively (21). The site-directed mutagenesis study has shown that removal of the aromatic side chain of Tyr13 is very disruptive, leading to a dramatic decrease in its RNA-binding affinity (21). Additionally, the position of Tyr13 is stabilized by a strong hydrogen bond to Gln54 (stacks on RNA G9 (11,50,51)) in cluster 9, and these two residues have a strong cooperative role for the complex stability (21), normally helping position the β2-β3 loop in the RNA loop (52). Because of the loss of the hydrogen bond to Gln54, the mutation of Tyr13 to Phe causes a 40-fold reduction in RNA-binding affinity (21), supporting that Tyr13 and Gln54 play an important cooperative role for the complex stability (21,40,49). The mutation of Gln54 to Ala or Glu disrupts its stacking interaction with G9, resulting in a severe inhibition of RNA binding with the affinity weakened by three to five orders of magnitude, which indicates Gln54 plays a key role in complex association and stability (21). Mutation of Phe56 to Ala can lead to a large loss in binding affinity (25,000-fold), supporting that this is a critical role for this aromatic residue in the stabilization of the complex conformation (21,53,54).

In short, these results suggest that the identified key residues in these clusters can stabilize the structure of U1A-RNA complex and regulate the binding affinity and specificity by hydrogen bonding and stacking interactions with RNA.

Residues at loop regions

As shown in Fig. 6 b, the residues in clusters 3, 4, and 5 are located at U1A loop1, and those in clusters 6, 7, and 8 are situated at loop3. As for the key residue clusters at U1A

loop1, residue Glu19 in cluster 3 forms specific side chain-base hydrogen bonds with nucleotides A6, U7, and G9 (42,46). Using MD simulations, Tang and Nilsson suggested that these hydrogen-bonding interactions are important for the specific U1A-RNA recognition (42). Additionally, the mutations of Lys20Gln in cluster 4 and Lys22Gln in cluster 5, both near the RNA-binding pocket, result in a significant reduction of the electrostatic interactions with RNA, indicating that the positive charge in this region is important for the attraction of RNA to the binding pocket (55). As for U1A loop3, connecting strands β2 and β3 and protruding through RNA loop, it has been demonstrated to be the primary source of binding specificity (55,56). In the identified key residue clusters, Arg47 in cluster 6 forms electrostatic interactions with C4, thus stabilizing the double-helix stems (42). By site-directed mutagenesis, Katsamba et al. demonstrated that Lys50 in cluster 7 plays a role in the electrostatically mediated association between U1A and RNA (52). Arg52 in cluster 8 makes hydrogen bonds with the conserved CG basepair at the end of the RNA loop, and an MD study revealed that the hydrogen bonds are critical for U1A-RNA-specific binding (19,42). Also, the experimental study found that the specific U1A-RNA binding is abolished in Arg52Gln mutant (12).

Because U1A loop1 and loop3 regions undergo remarkable conformational changes in response to the binding of RNA, these identified key residue clusters in the flexible loop regions might contribute largely to the induced fit and further to the high binding affinity between U1A and RNA through forming specific electrostatic and hydrogen-bonding interactions.

Residues at Helix-C and Helix-N regions

Although Helix-C and Helix-N, both considerably flexible, are far away from the RNA-binding interface (43), some residue clusters located in them are still identified as important for RNA binding. As shown in Fig. 6 b, the predicted key residue clusters are 1 in Helix-N and 12, 13, and 14 in Helix-C. The mutation of Arg7Gln in cluster 1 can affect binding kinetics slightly and yet is deleterious to the stability of structure (55). As for Helix-C, the experimental and theoretical studies have demonstrated that it makes a significant contribution to RNA-binding affinity (16,45,57). The site-directed mutagenesis has proved that Ser91 in cluster 12 contributes a lot to RNA-binding specificity (57). Ile94 in cluster 14 makes a backbone hydrogen bond with Ser91, forming the first turn of the helix, and Ile94Pro mutant results in a 100-fold loss in complex stability, indicating its major role in stabilizing U1A-RNA complex structure (16). Asp92 in cluster 13 makes a hydrogen bond with C12, and the experiments on the mutations of Asp92Glu, Asp92Ala, and Asp92Phe showed that Asp92 plays a critical role in locking the U1A onto the RNA (16). Also, a

truncation, removal, or disruption of the Helix-C will result in a considerable loss of complex stability (16). The reason may be that although Helix-C residues do not make direct contacts with RNA, RNA binding can mediate local motional cooperativity between Helix-C and the main binding interface β-sheets, which has been found by the pairwise coupling free energy analysis (15) and our results about the effect of RNA binding on movement coupling between U1A residues (see the Movement coupling between residues in complex and in the U1A protein).

Together, these results hint that these key residues of clusters, thermodynamically coupled with the binding of RNA, participate in the long-range allosteric signal transmission.

DISCUSSION

Comparison with the parameter-free GNM model

In our approach, the GNM includes two parameters: a distance cutoff used to define the connections and placement of springs between residues, and a spring constant to represent the interaction strengths between residues. Recently, a parameter-free GNM (pfGNM) model (58,59) with distance-dependent springs has been proposed and showed a better B-factor prediction. Here, to examine the impact of the specific spring models on the results, we applied this model to the identification of U1A key residues associated with RNA-binding and conformational transition.

In the pfGNM model, the nodes representing amino acid and nucleotide residues are the same with those in the conventional GNM, and all pairs of residues are considered to interact with each other (no distance cutoff used) with interaction strengths weighted by the inverses of their square distances. Thus, residue pairs that are far apart have weaker interactions than those that are close in the pfGNM. As shown in Fig. S1, the correlation coefficient between theoretical and experimental B-factors of the residues is 0.902 for the pfGNM, slightly higher than that (0.857) when using the conventional GNM model, which is consistent with the former pfGNM results (58).

In Fig. S2, *a* and *b*, we present the binding free energy changes $\Delta\Delta G$ in response to residue mutations and the identified key residue clusters, respectively. Compared with Fig. 6, Fig. S2 shows similar regions of critical residues, although the differences in $\Delta\Delta G$ values among neighboring residues become smaller when using the pfGNM, which makes the discrimination of key residue clusters less explicit. This is understandable because the neighboring residues in the pfGNM are connected with all the other residues through long-range interacting springs. Thus, the network and free energy changes are less sensitive to the mutations than that in the conventional GNM, which defines the network connections through short-range springs in which a small mutation is more likely to trigger a network change. As shown in Fig. S2, five residual clusters can be identified

with relatively sharp peeks, which are, respectively, centered at Glu5, Glu19, Ser48, Met51, and Ser91. The number of key residues is lower than that (14) identified by the conventional GNM, probably because of the less sensitivity of the pfGNM model on the mutations. Nevertheless, the shape of the $\Delta\Delta G$ distribution is largely the same between the two models, and the five key residue clusters by the pfGNM are identical to a subset of the clusters of 1, 3, 6, 7, and 12 (Fig. 6 *a*) obtained from the conventional GNM-based thermodynamic method.

Comparison with the atomic-level GNM model

To examine the impact of the level of coarse graining on the results, we applied the atomic-level GNM model on the complex and protein U1A. Here, each heavy atom in protein and RNA is assigned a node, and the cutoff distance is shorter (than that in the conventional residue-level GNM) with the optimum value 4.0 Å for our system adopted, consistent with the set in a previous work (60). As shown in Fig. S3, the correlation coefficient between theoretical and experimental B-factors of the residues (C_α, P, and O4′ atoms considered for comparison) is 0.796, a little lower than that (0.857) obtained by the residue-level GNM. A previous study shows that the atomic-level ENMs can provide an improved representation for the collective motions of proteins compared with the coarse-grained models (60). Here, the reason for no improvement is probably because of the RNA molecule as when only the protein U1A considered an improvement from 0.674 to 0.677 can be observed. We also calculated the cross correlations between atomic fluctuations in complex and protein U1A using the first 93 modes, as shown in Fig. S4. There is a similar tendency with that (Fig. 5) obtained by the residue-level GNM, whereas the effect of the RNA binding on motional correlations between protein residues in the former is not as obvious as that in the latter.

Additionally, we present the binding free energy changes $\Delta\Delta G$ in response to residue mutations and the identified key residue clusters in Fig. S5. Compared with Fig. 6 *a* obtained by residue-level GNM, Fig. S5 also shows the similar regions of critical residues, and these regions are more remarkable, whereas the curve of $\Delta\Delta G$ is more rugged, which is probably caused by the atomic-level perturbation. There are 13 residual clusters that are identified with relatively sharp peaks, and they are, respectively, centered at Glu5, Thr11, Tyr13, Asn15, Glu19, Lys22, Leu44, Arg52, Gln54, Phe56, Gln85, Lys88, and Ser91. Among them, the nine key residue clusters (centered at Glu5, Tyr13, Glu19, Lys22, Arg52, Gln54, Phe56, Gln85, and Ser91) are identical to a subset of the clusters of 1, 2, 3, 5, 8, 9, 9, 11, and 12 (Fig. 6 *a*) obtained by the conventional residue-level GNM-based thermodynamic method. Although some clusters are not identified, the shape of the $\Delta\Delta G$ distribution is largely the same between the two GNM models.

Considering the remarkability of the identified critical residue regions, we think the atomic-level GNM-based method can be helpful for the residue-level GNM to identify the critical residues involved in RNA-binding and conformational transitions.

Comparison with the all-atom normal mode analysis

The conventional GNM is a coarse-grained model with the uniform force constant set, and we want to know the effects of using the all-atom normal mode analysis (NMA) with classical force field on the results. To detect the issue, the NMA method (61,62) was performed using all-atom models in Amber03 force field (63), which requires a very thorough energy minimization and the calculation of a Hessian matrix (See Supporting Methods for the detailed process). The result shows that the correlation coefficient between theoretical and experimental B-factors of the residues is 0.757, as shown in Fig. S6, lower than 0.796 and 0.857 obtained, respectively, by the atomic-level and conventional residue-level GNM models, which is consistent with the previous NMA result that the structural fluctuations predicted by NMA are often worse than those by GNM (64–66).

We calculated the cross correlations between atomic fluctuations in complex and protein U1A using the first 15 modes, as shown in Fig. S7. It is shown that the overall results are similar to those obtained from the GNM, but the former shows a not very distinct distribution, with some regions having correlations contrary to the corresponding ones in the GNM. From Fig. S7 *a*, for RNA in complex, the 5′ and 3′ stem nucleotides are anticorrelated, which is contrary to the results obtained by GNM. From the crystal structure, the 3′ and 5′ stems of RNA do form a series of basepairing interactions and should be of a positive motion correlation. For the two partners in complex, the 5′ and 3′ stem parts both have weak positive correlations with U1A loop3 in GNM, whereas the correlations are positive and negative, respectively, in NMA. Additionally, the four β-sheets turn into positive correlations with each other upon RNA binding in GNM, whereas there are still more or less anticorrelations between them in NMA.

In Fig. S8, we present the binding free energy changes $\Delta\Delta G$ in response to residue mutations and the identified key residue clusters. Compared with Fig. 6 *a* obtained by residue-level GNM-based methods, Fig. S8 also shows similar regions of functionally critical residues, although their extent of similarity is less than that between the results obtained by the two atomic- and residue-level GNM-based models, which is probably due to the common uniform force constant set in the latter two methods. A more rugged curve of $\Delta\Delta G$ is displayed in Fig. S8, like that in Fig. S5 obtained by the atomic-level GNM-based method, which may be caused by their atomic-level perturbation. As shown in Fig. S8, eight residual clusters can be identified with rela-

tively sharp peaks, which are respectively centered at Glu19, Lys23, Lys28, Lys50, Arg70, Phe77, Arg83, and Thr89. Among them, there are five key residue clusters (centered at Glu19, Lys23, Lys50, Arg83, and Thr89) which are identical to a subset of the clusters of 3, 5, 7, 11, and 12 (Fig. 6 *a*) obtained from the residue-level GNM-based method.

As we know, in NMA, the system of motion equations is solved with a quadratic approximation to the potential energy function. This approximation is accurate enough only in a small neighborhood of the energy minimal state. In addition, NMA is conducted at the atomic level, which is subject to all the atomic-level errors induced in energy calculations because we currently do not have an efficient force field that can precisely describe the atomic interactions. As a result, the structural fluctuations predicted by NMA are often worse than those by coarse-grained GNM, which is built on a higher level of knowledge-based approximation (64,65); this has been confirmed again by Park et al. on the test of 104 proteins (66). Naturally, the fluctuation cross correlations predicted by NMA are often not as good as those by GNM. These are the case in our work. Here, the calculation of $\Delta\Delta G$ according to Eqs. 9 and 10 depends on the different values of the mean-square fluctuations of nodes and fluctuation cross correlations between nodes at the two states of protein U1A with and without RNA binding. Maybe, this can partially explain the reason why the predictive power of key residues by NMA is not as good as that by GNM on protein U1A.

With regard to computation cost, it will take a few minutes to identify key residues with a GNM-based method, whereas the time is about several hours when an NMA-based one is used on a four-core CPU. GNM is a simple model and easy to construct, and the dimension of the model is much smaller (proportional to the number of residues or atoms) than that of the all-atom NMA (three dimensions). In addition, GNM does not require the exact potential energy, Hessian matrix, and very thorough energy minimization even to the maximal force on atoms less than $\sim 10^{-9}$ kJ/(mol·nm), which can reduce not only computational cost but also possible errors in atomic energy calculations. Therefore, considering the efficiency and accuracy, our GNM-based thermodynamic method is better than the NMA-based one in identifying protein U1A key residues associated with RNA-binding and conformational transition.

Limitations of the method

In our GNM-based thermodynamic method, a protein (RNA) structure has been modeled as a homogeneous elastic network, which neglects the differences between 20 (4) kinds of amino acids (nucleotides) and the conformational changes induced by the removing of the RNA. This method aims to identify the topologically key residues important for

U1A's functionally conformational changes associated with RNA binding. As we know that the functional information of protein motions is largely encoded in the structural topology (29), we think that although this method has the above coarse-grained approximations, the topology-based model can help identify the key residues of U1A that are important for its functional motions involved in the process of RNA binding.

CONCLUSIONS

We have extended the GNM to examine the effects of RNA binding on the dynamics of the U1A domain, which is one of the major components of the spliceosomal U1 small nuclear ribonucleoprotein particle that is critical to the process of RNA alternative splicing. The analytical results of the GNM calculations show that the snRNA binding can lead to a large loss in the flexibility of U1A loop1, loop3, and Helix-C regions because of the formation of specific interactions with the RNA target, suggesting their importance for the induced fit in U1A-RNA recognition. The enhanced cross correlations between the conformational fluctuations reveal that RNA binding strengthens the interactions among the β-sheets of the U1A domain and between the main binding surface β-sheets and Helix-C. In the complex structure, the U1A loop3 and loop1 are found to be strongly and modestly positively correlated with the target RNA loop, respectively, consistent with the relative extent of the flexibility loss in U1A loop1 and loop3.

Further, we extended the GNM-based thermodynamic cycle method to the identification of the key residues of U1A whose mutations significantly change the binding free energy between U1A and snRNA. These identified key residues are found in three groups according to their locations in the protein structure: 1) residues at β-sheets, which provide a perfect binding surface for RNA and are directly involved in the binding of RNA with its receptor U1A; 2) residues at loop regions, which are positively charged and highly flexible and largely contribute to the induced fit and the high binding affinity between U1A and RNA through forming long-range electrostatic and specific hydrogen-bonding interactions; and 3) residues at Helix-N and Helix-C, which are highly mobile, thermodynamically coupled with the binding of RNA, and important for the long-range allosteric signal transmission. Because all the residues have been treated equally in the perturbation/mutation simulations, the results of identified critical residues should be of general implications. As a comparison, we also examined the pfGNM, atomic-level GNM, and classical all-atom NMA models, in which a similar shape of $\Delta\Delta G$ distribution was obtained despite less real key residues identified partially because the relatively lower sensitivity of the pfGNM networks on the mutation perturbations, the errors caused by the atomic-level perturbation, and the atomic-level errors induced in energy calculations and mu-

tation perturbations, respectively; however, the atomic-level GNM, with the ability to identify remarkably the critical residue regions, can be helpful for the residue-level GNM to identify the critical residues. These data with model-independent results help demonstrate the robustness of the approach and the generality of the key residue identifications.

Although based on an overly simplified ENM, the approach is featured with the analytical solution and time-saving virtues compared with atomic MD simulations, and its single parameter harmonic potential can reproduce in good detail the low-frequency, large-amplitude motions of proteins in their native states. Considering its simplicity and the consistence of the analytical results with experimental data, we believe that the Gaussian network modeling represents a promising approach, complementary to the atomic-level simulations, to investigating the binding and allosteric dynamics of the protein and RNA interactions, which can be readily extended for other macromolecular interaction systems.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.03.008.

AUTHOR CONTRIBUTIONS

Z.H. performed the NMA calculations. Q.S., C.L., and Y.Z. designed the research. Q.S. carried out the GNM calculations. Q.S., C.L., and J.S. performed data analyses. Q.S., C.L., Z.H., and Y.Z. wrote the manuscript. Q.S., W.G., and S.W. developed the program.

ACKNOWLEDGMENTS

REFERENCES

1. Allain, F. H., C. C. Gubser, ..., G. Varani. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature*. 380:646–650.

2. Maris, C., C. Dominguez, and F. H. Allain. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J*. 272:2118–2131.

3. Lukong, K. E., K. W. Chang, ..., S. Richard. 2008. RNA-binding proteins in human genetic disease. *Trends Genet*. 24:416–425.

4. Hall, K. B. 1994. Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry*. 33:10076–10088.

5. Stump, W. T., and K. B. Hall. 1995. Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA*. 1:55–63.

6. Pomeranz Krummel, D. A., C. Oubridge, ..., K. Nagai. 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature*. 458:475–480.

7. 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.

8. Krämer, A., W. Keller, …, R. Lührmann. 1984. The 5′ terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell.* 38:299–307.

9. Kranz, J. K., and K. B. Hall. 1999. RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface. *J. Mol. Biol.* 285:215–231.

10. Black, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72:291–336.

11. Oubridge, C., N. Ito, …, K. Nagai. 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature.* 372:432–438.

12. Nagai, K., C. Oubridge, …, P. R. Evans. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature.* 348:515–520.

13. Guzman, I., Z. Ghaemi, …, M. Gruebele. 2015. Native conformational dynamics of the spliceosomal U1A protein. *J. Phys. Chem. B.* 119:3651–3661.

14. Anunciado, D., M. Agumeh, …, A. M. Baranger. 2008. Characterization of the dynamics of an essential helix in the U1A protein by time-resolved fluorescence measurements. *J. Phys. Chem. B.* 112:6122–6130.

15. Kranz, J. K., and K. B. Hall. 1998. RNA binding mediates the local cooperativity between the β-sheet and the C-terminal tail of the human U1A RBD1 protein. *J. Mol. Biol.* 275:465–481.

16. Law, M. J., D. S. Lee, …, I. A. Laird-Offringa. 2013. The role of the C-terminal helix of U1A protein in the interaction with U1hpII RNA. *Nucleic Acids Res.* 41:7092–7100.

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

18. Pezeshki, S., C. Chimerel, …, U. Kleinekathöfer. 2009. Understanding ion conductance on a molecular level: an all-atom modeling of the bacterial porin OmpF. *Biophys. J.* 97:1898–1906.

19. Reyes, C. M., and P. A. Kollman. 2000. Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol.* 295:1–6.

20. Pitici, F., D. L. Beveridge, and A. M. Baranger. 2002. Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: do molecular substates code for specificity? *Biopolymers.* 65:424–435.

21. Law, M. J., E. J. Chambers, …, I. A. Laird-Offringa. 2005. Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction. *Nucleic Acids Res.* 33:2917–2928.

22. Dror, R. O., R. M. Dirks, …, D. E. Shaw. 2012. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* 41:429–452.

23. Kurkcuoglu, O., R. L. Jernigan, and P. Doruker. 2010. Collective dynamics of large proteins from mixed coarse-grained elastic network model. *QSAR Comb. Sci.* 24:443–448.

24. Tuzmen, C., and B. Erman. 2011. Identification of ligand binding sites of proteins using the Gaussian network model. *PLoS One.* 6:e16474.

25. Su, J. G., H. J. Du, …, C. X. Wang. 2013. Identification of functionally key residues in AMPA receptor with a thermodynamic method. *J. Phys. Chem. B.* 117:8689–8696.

26. Su, J. G., X. M. Han, …, Y. D. Wu. 2016. Analysis of conformational motions and related key residue interactions responsible for a specific function of proteins with elastic network model. *J. Biomol. Struct. Dyn.* 34:560–571.

27. Pinamonti, G., S. Bottaro, …, G. Bussi. 2015. Elastic network models for RNA: a comparative assessment with molecular dynamics and SHAPE experiments. *Nucleic Acids Res.* 43:7260–7269.

28. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.

29. Bahar, I., A. R. Atilgan, …, B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–2736.

30. Wang, Y., A. J. Rader, …, R. L. Jernigan. 2004. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* 147:302–314.

31. Dutta, A., and I. Bahar. 2010. Metal-binding sites are designed to achieve optimal mechanical and signaling properties. *Structure.* 18:1140–1148.

32. Bahar, I., and A. J. Rader. 2005. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15:586–592.

33. Lopéz-Blanco, J. R., and P. Chacón. 2013. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* 184:261–270.

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

35. Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093.

36. Quan, L., Q. Lv, and Y. Zhang. 2016. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics.* 32:2936–2946.

37. Su, J. G., X. J. Xu, …, C. X. Wang. 2011. Identification of key residues for protein conformational transition using elastic network model. *J. Chem. Phys.* 135:174101.

38. Lv, D., C. Wang, …, X. Zhang. 2017. An efficient perturbation method to predict the functionally key sites of glutamine binding protein. *Comput. Biol. Chem.* 67:62–68.

39. Li, C., D. Lv, …, Y. Zhang. 2016. Approach to the unfolding and folding dynamics of add A-riboswitch upon adenine dissociation using a coarse-grained elastic network model. *J. Chem. Phys.* 145:014104.

40. 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.

41. Mittermaier, A., L. Varani, …, G. Varani. 1999. Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.* 294:967–979.

42. Tang, Y., and L. Nilsson. 1999. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284–1305.

43. Wittekind, M., M. Görlach, …, L. Mueller. 1992. 1H, 13C, and 15N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry.* 31:6254–6265.

44. Showalter, S. A., and K. B. Hall. 2004. Altering the RNA-binding mode of the U1A RBD1 protein. *J. Mol. Biol.* 335:465–480.

45. Kurisaki, I., M. Takayanagi, and M. Nagaoka. 2014. Combined mechanism of conformational selection and induced fit in U1A-RNA molecular recognition. *Biochemistry.* 53:3646–3657.

46. Kormos, B. L., A. M. Baranger, and D. L. Beveridge. 2007. A study of collective atomic fluctuations and cooperativity in the U1A-RNA complex based on molecular dynamics simulations. *J. Struct. Biol.* 157:500–513.

47. Qin, F., Y. Chen, …, H. F. Chen. 2010. Induced fit or conformational selection for RNA/U1A folding. *RNA.* 16:1053–1061.

48. Lutz-Freyermuth, C., C. C. Query, and J. D. Keene. 1990. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA.* 87:6393–6397.

49. Kranz, J. K., J. Lu, and K. B. Hall. 1996. Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. *Protein Sci.* 5:1567–1583.

50. Avis, J. M., F. H. Allain, …, D. Neuhaus. 1996. Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal

residues in structure stability and RNA binding. *J. Mol. Biol.* 257: 398–411.

51. Allain, F. H., P. W. Howe, ..., G. Varani. 1997. Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* 16:5764–5772.

52. Katsamba, P. S., M. Bayramyan, ..., I. A. Laird-Offringa. 2002. Complex role of the beta 2-beta 3 loop in the interaction of U1A with U1 hairpin II RNA. *J. Biol. Chem.* 277:33267–33274.

53. Zhao, Y., and A. M. Baranger. 2003. Design of an adenosine analogue that selectively improves the affinity of a mutant U1A protein for RNA. *J. Am. Chem. Soc.* 125:2480–2488.

54. Kormos, B. L., S. N. Pieniazek, ..., A. M. Baranger. 2011. U1A protein-stem loop 2 RNA recognition: prediction of structural differences from protein mutations. *Biopolymers.* 95:591–606.

55. Law, M. J., M. E. Linde, ..., I. A. Laird-Offringa. 2006. The role of positively charged amino acids and electrostatic interactions in the complex of U1A protein and U1 hairpin II RNA. *Nucleic Acids Res.* 34:275–285.

56. Katsamba, P. S., D. G. Myszka, and I. A. Laird-Offringa. 2001. Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J. Biol. Chem.* 276:21476–21481.

57. Zeng, Q., and K. B. Hall. 1997. Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA.* 3:303–314.

58. Yang, L., G. Song, and R. L. Jernigan. 2009. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA.* 106:12347–12352.

59. Zimmermann, M. T., and R. L. Jernigan. 2014. Elastic network models capture the motions apparent within ensembles of RNA structures. *RNA.* 20:792–804.

60. Sen, T. Z., Y. Feng, ..., R. L. Jernigan. 2006. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J. Chem. Theory Comput.* 2:696–704.

61. Go, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.* 80:3696–3700.

62. Brooks, B., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA.* 80:6571–6575.

63. Duan, Y., C. Wu, ..., P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.

64. Cui, Q., and I. Bahar. 2006. Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. *In* Chapman & Hall/CRC Mathematical and Computational Biology Series. Chapman & Hall/CRC Press, London/Boca Raton.

65. Micheletti, C., P. Carloni, and A. Maritan. 2004. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins.* 55:635–645.

66. Park, J. K., R. Jernigan, and Z. Wu. 2013. Coarse grained normal mode analysis vs. refined Gaussian network model for protein residue-level structural fluctuations. *Bull. Math. Biol.* 75:124–160.