



A study on allosteric communication in U1A-snRNA binding interactions: network analysis combined with molecular dynamics data



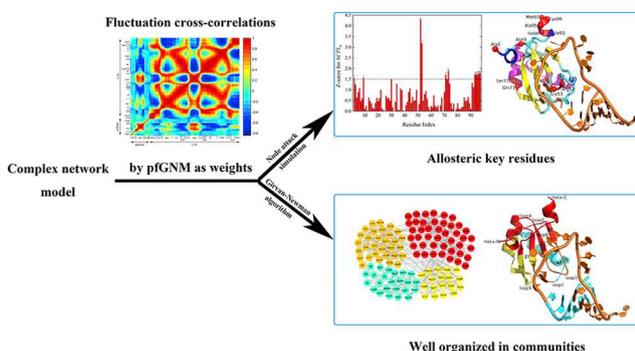
Qi Shao¹, Weikang Gong¹, Chunhua Li*

College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China

HIGHLIGHTS

- Effective combination of residual network with GNM models to investigate allosteric communication.
- Key residues were identified for U1A's allosteric signal transmission induced by snRNA binding.
- U1A is well organized in communities acting different roles for snRNA binding and allosteric regulation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

U1A-snRNA interactions
Gaussian network model
Complex network model
Allosteric communication
Key residues

ABSTRACT

The allosteric regulation during the binding interactions between small nuclear RNAs (snRNAs) and the associated protein factors is critical to the function of spliceosomes in alternative RNA splicing. Although network models combined with molecular dynamics simulations have shown to be powerful tools for the analysis of protein allostery, the atomic-level simulations are, however, too expensive and with limited accuracy for the large-size systems. In this work, we use a residual network model combined with a coarse-grained Gaussian network model (GNM) to investigate the binding interactions between the snRNA and the human U1A protein which is a major component of the spliceosomal U1 small nuclear ribonucleoprotein particle, and to identify the residues that play an important role in the allosteric communication in U1A during this process. We also utilize the Girvan-Newman method to detect the structural organization in U1A-snRNA recognition and interactions. Our results reveal that: (I) not only the residues at the binding sites that are traditionally considered to play a major role in U1A-snRNA association, but those residues that are far away from the RNA binding interface participate in the U1A's allosteric signal transmission induced by the RNA binding; (II) the structure of U1A protein is well organized with different communities acting different roles for its RNA binding and allosteric regulation. The study demonstrates that the combination of the residual network and elastic network models is an effective and efficient method which can be readily extended to the investigation of the allosteric communication for other macromolecular interaction systems.

* Corresponding author.

E-mail address: chunhuali@bjut.edu.cn (C. Li).

¹ Qi Shao and Weikang Gong contributed equally to this work.

1. Introduction

The RNA recognition motif (RRM), also known as RNA-binding domain (RBD), is one of the most abundant protein domains in eukaryotes that can undergo large dynamic and activity adaptability by allosteric regulation when binding to their RNA targets [1,2]. The most well characterized RRM is the N-terminal RNA-binding domain (RBD1) of the human U1A protein, one of the major protein components of the spliceosomal U1 small nuclear ribonucleoprotein (U1 snRNP) particle [3,4]. This domain can bind with sufficiently high affinity and specificity to stem-loop II (SL2) of U1 small nuclear RNA (hereinafter referred to as snRNA and the protein domain as U1A), participating in pre-mRNA splicing [5,6]. However, currently the effect of snRNA binding on U1A's dynamics and allosteric communication, and how U1A's structure is organized for its function are not completely clear.

The structure of U1A bound to snRNA has been determined by X-ray diffraction analysis at 1.92 Å resolution (Fig. 1 (a) and (b)) [7]. The U1A is characterized by a $\beta 1$ - αA - $\beta 2$ - $\beta 3$ - αB - $\beta 4$ sandwich fold that contains a four-stranded antiparallel β -sheet as the primary RNA-binding surface flanked by two α -helices on one side. During complex formation, the snRNA induces U1A conformational changes throughout the U1A-snRNA interface as well as the C-terminal helix (Helix-C) which is a little far away from the interface [8]. Before binding, U1A forms the closed state where the Helix-C covers part of the RNA-binding surface. Upon binding to RNA, U1A folds into the open form where the Helix-C is oriented away uncovering the buried area to permit RNA access (Fig. 1 (c)), and loop1 and loop3 also undergo a structural rearrangement with loop3 protruding through the snRNA loop [9].

In recent years, besides experimental studies, theoretical molecular dynamics (MD) simulations have focused on the relationship between the binding and conformational changes during their interactions. Reyes and Kollman carried out MD simulations and site-directed mutagenesis of U1A-snRNA interface residues to examine the origin of the binding specificity [10]. Utilizing MD simulations, Pitici *et al.* obtained the predictions for the structures of the unbound forms of U1A in solution in order to elucidate dynamical aspects of the induced fit upon RNA binding [11]. Most recently, Guzman *et al.* adopted a series of MD simulations to delve whether U1A protein alone is capable of undergoing the conformational dynamics similar to the structural rearrangements upon RNA binding [12]. On the experimental side, Law *et al.* used a surface plasmon resonance-based biosensor to gain mechanistic insight into the role of Helix-C in mediating the interaction of U1A with RNA [13].

Although the analyses above help to find the U1A's important structural elements involved in the binding and allosteric transition, the pathways of allosteric signal transduction favored by the network of inter-residue contacts and the key residues involved in the allostery remain poorly understood. MD simulation is a time-consuming method and it is difficult to investigate the large-scale functional motions of proteins. To address the issue, different levels of coarse-grained models have been developed [14,15]. Among them, the elastic network model (ENM), a harmonic potential-based and cost-effective computational method, has achieved great success in investigating the function-relevant motions in allosteric transitions [16–18] and the allosteric signaling caused by the ligand binding, mutations and their combinations [19–21]. Our group also utilized ENM-based methods to study the issues involved in the folding and allosteric processes of biomolecules [22–24]. In the conventional ENM model, the biomacromolecule structure is modelled as a coarse-grained elastic network, where the node pairs within a given cutoff distance are considered to have interactions which are modelled as a set of Hookean springs with a uniform force constant [25]. Later, Yang *et al.* proposed a parameter-free ENM (pfENM), in which all the node pairs are considered to be interacting with each other with the strength being inversely proportional to their square distance. Compared with the conventional cutoff-based ENM model, the pfENM model is not only simpler, i.e. without any parameters, but also can yield even better predictions for crystallographic B-factors and functionally correlated motions due to the consideration of long-range interactions [26].

In addition to the network models based on the intra-protein interactions [27,28], the graph-theoretical approach (complex network model) has also been successfully used in the analyses of allosteric communications in protein systems [29–31] and the predictions of hub residues in allosteric signal transduction [32–34]. Some network parameters such as degree, closeness and betweenness centrality, can provide an estimate of the importance of the nodes in allosteric communication pathways [35,36]. The characteristic path length (CPL) is a very important parameter and Del Sol *et al.* found that the residues that greatly affect the CPL values upon removal are usually critical to allosteric signal transmission [37]. Furthermore, the behavior of the nodes that are highly correlated and within close physical proximity can be analyzed in terms of community structure [38]. The community analysis can identify relatively independent communities of residues that behave as semi-rigid bodies when propagating allosteric signals in biological systems [39]. Usually, the signaling strength between nodes is expressed as an edge weight, and thus much effort has been put into

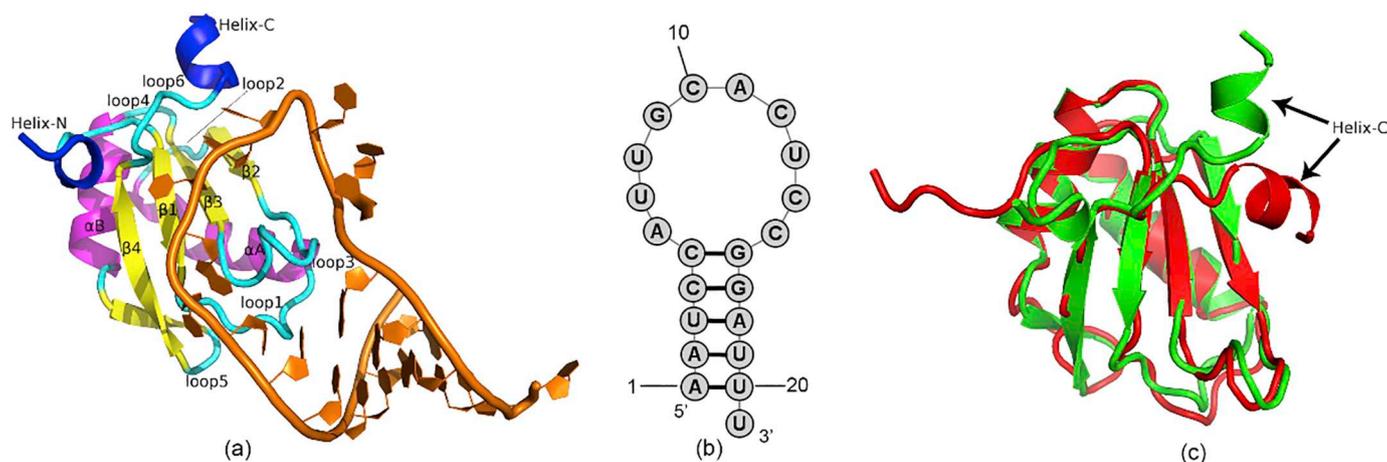


Fig. 1. Complex structure of U1A with snRNA (PDB code: 1URN) (a) and the secondary structure of snRNA (b). U1A forms a $\beta 1$ - αA - $\beta 2$ - $\beta 3$ - αB - $\beta 4$ sandwich fold with $\beta 1$ (Thr11-Asn15), $\beta 2$ (Ile40-Val45), $\beta 3$ (Ala55-Phe59), $\beta 4$ (Arg83-Tyr86) colored yellow, αA (Lys23-Phe37) and αB (Val62-Met72) colored pink. Helix-N and Helix-C are in blue, and loops (loop1-loop6) in cyan. (c). Superimposition between the open state of RNA-bound U1A and the closed state of apo-U1A structures.

the determination of the edge weight in order to explore the allosteric regulation effectively. Bhattacharyya *et al.* first used energy between nodes as a weight to examine the paths of allosteric communication in Pyrrolysyl-tRNA Synthetase [40]. VanWart *et al.* then utilized the motional correlation between nodes as a weight to explore the residue component contributions to dynamical network models of allostery [41]. Until recently, McClendon *et al.* computed the mutual information of Cartesian coordinates between nodes as a weight to identify inter-residue correlated motions in protein residue network [42]. Despite of the successes, the above approaches have to be performed on the basis of MD simulations to capture the conformational ensemble, which is often found too expensive and time-consuming.

Enlightened by the above approaches, in this work, we analyze the dynamic behavior of the binding interactions between the human U1A protein and snRNA using the parameter-free Gaussian network model pfGNM. To explore the key residues and the structural organization in the allosteric communication, we further construct the weighted residual network model with the weights coming from the inter-residue movement correlations computed based on the slowest motional modes from pfGNM, not from MD method. This work builds a new and effective avenue for investigating protein-RNA binding and allosteric dynamics.

2. Materials and methods

2.1. Parameter-free Gaussian network model

Different from the conventional cutoff-based Gaussian network model (GNM), the parameter-free GNM (pfGNM) model [26,43] adopts a distance-dependent spring constant set. In constructing the pfGNM model of protein-RNA complex, we model the molecular system as a coarse-grained and elastic network by replacing one residue with one node (C_α atom for protein residue, and P atom for RNA nucleotide [44]) and imposing a harmonic potential with the inverse-square distance dependent spring force constants between all pairs of nodes. Thus, the residue pairs that are far apart have weaker interactions than those pairs that are close to each other. By this simplification, the total internal potential energy of the network of N nodes can be written as

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

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

$$\Gamma_{ij} = \begin{cases} r_{ij}^{-2} & \text{if } i \neq j \\ -\sum_{i,j \neq i} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (2)$$

where r_{ij} is the distance between the i th and j th nodes.

The mean-square fluctuation of each node and the fluctuation cross-correlation 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

$$\Gamma^{-1} = U \Lambda^{-1} U^T \quad (3)$$

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 i th and j th nodes and the mean-square fluctuation of the i th node can be written as

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ij} \quad (4)$$

$$\langle \Delta R_i \cdot \Delta R_i \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ii} \quad (5)$$

where k_B is the Boltzmann constant, T is the absolute temperature, and the meaning of γ is the same as eq. (1). The cross-correlation and mean-square fluctuation associated with the k th mode can be given by

$$\langle \Delta R_i \cdot \Delta R_j \rangle_k = \frac{3k_B T}{\gamma} \lambda_k^{-1} [u_k]_i [u_k]_j \quad (6)$$

$$\langle \Delta R_i \cdot \Delta R_i \rangle_k = \frac{3k_B T}{\gamma} \lambda_k^{-1} [u_k]_i [u_k]_i \quad (7)$$

According to the Debye-Waller theory, the B-factor of the i th node can be calculated with the expression

$$B_i = 8\pi^2 \langle \Delta R_i \cdot \Delta R_i \rangle / 3 \quad (8)$$

The cross-correlation between residue fluctuations is normalized as

$$C_{ij} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{[\langle \Delta R_i^2 \rangle \times \langle \Delta R_j^2 \rangle]^{\frac{1}{2}}} \quad (9)$$

This value ranges from -1 to 1 . Positive values depict correlated motions occurring along the same direction and the negative represent correlated motions along 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.

2.2. Complex network model

Based on the protein-RNA complex structure, a weighted complex network model is constructed with residues as nodes (C_α and P atoms stand for amino acid and nucleotide residues, respectively) and contacts as edges. Here, the contact is defined as the pairs of nodes within a cutoff distance. For node pairs in protein, RNA and their interface, the cutoff distances are set to 7.0, 13.0 and 10.0 Å respectively. The weight w_{ij} of an edge between nodes i and j is the probability of information transfer across that edge [45] as measured by their fluctuation cross-correlation C_{ij} computed based on the slowest motional modes from pfGNM model (in order to eliminate functionally unrelated high-frequency noise)

$$w_{ij} = -\log(|C_{ij}|) \quad (10)$$

The length d_{ij} of a path between distant nodes i and j is the sum of the edge weights between the consecutive nodes (k, l) along the path

$$d_{ij} = \sum_{kl} w_{kl} \quad (11)$$

The shortest path among all the paths between two nodes is found by the Floyd-Warshall algorithm that compares all possible path lengths between the two nodes. The characteristic path length (CPL) is defined as the average length of the shortest paths between all pairs of nodes in a network

$$CPL = \frac{1}{N_p} \sum_{j>i} d_{ij} \quad (12)$$

where N and N_p are the numbers of nodes and node pairs, respectively, and d_{ij} is the shortest path length between nodes i and j . The contribution of a node k to the information communication within a network can be measured with the change of the CPL after removing node k from the network. The change of CPL ($\Delta CPL_k = CPL_k - CPL$) was previously used to predict the important residues in allosteric communication within proteins [37]. A Z-score analysis provides a measure for the relative change in CPL

$$Z\text{-score}_k = \left| \frac{\Delta CPL_k - \overline{\Delta CPL_k}}{\sigma} \right| \quad (13)$$

where ΔCPL_k is the change of CPL after removal of node k , $\overline{\Delta CPL_k}$ is

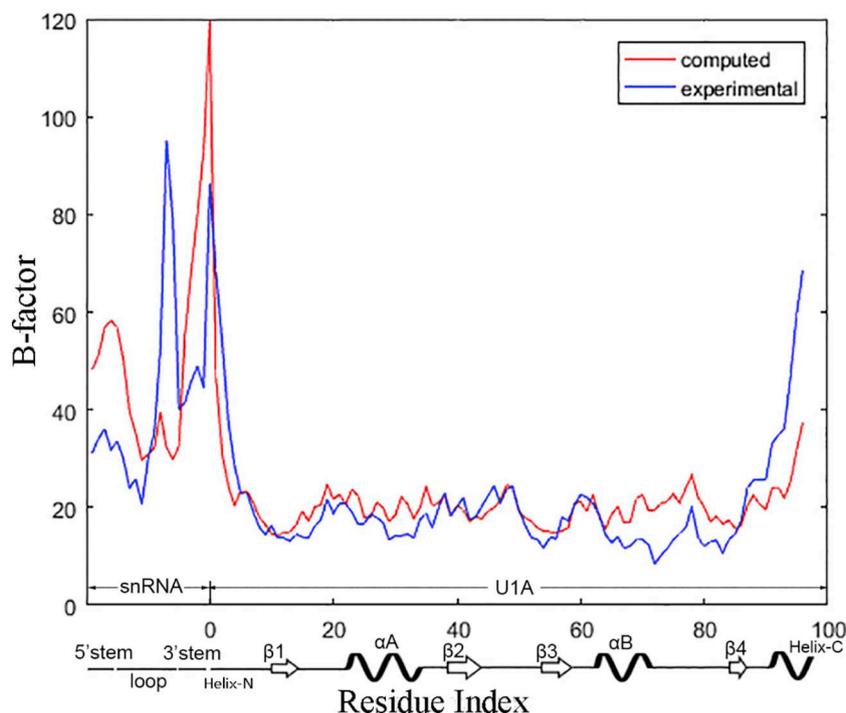


Fig. 2. Comparison between the experimental (blue line) and computed (red line) B-factors of P and Ca atoms of U1A-snRNA complex. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the change ΔCPL_k averaged over all the nodes, and σ is the corresponding standard deviation.

Additionally, the dynamic network of nodes and edges contains substructures or communities of nodes that are more densely interconnected to each other than to other nodes in the network. The community structure is identified by using the Girvan-Newman algorithm [38] which uses a top-down approach to iteratively remove the edge with the highest betweenness and recalculate the betweenness of all remaining edges until none of the edges remains. The betweenness

of a node i is defined to be the fraction of the shortest paths between pairs of nodes in a network that pass through the node i . The normalized betweenness is given as

$$b_i = \frac{2}{N(N-1)} \sum_{\substack{j < k \\ j \neq i \neq k}} \frac{g_{jk}(i)}{g_{jk}} \tag{14}$$

where g_{jk} is the number of shortest paths between nodes j and k , and $g_{jk}(i)$ is the number of shortest paths from node j to k that pass through

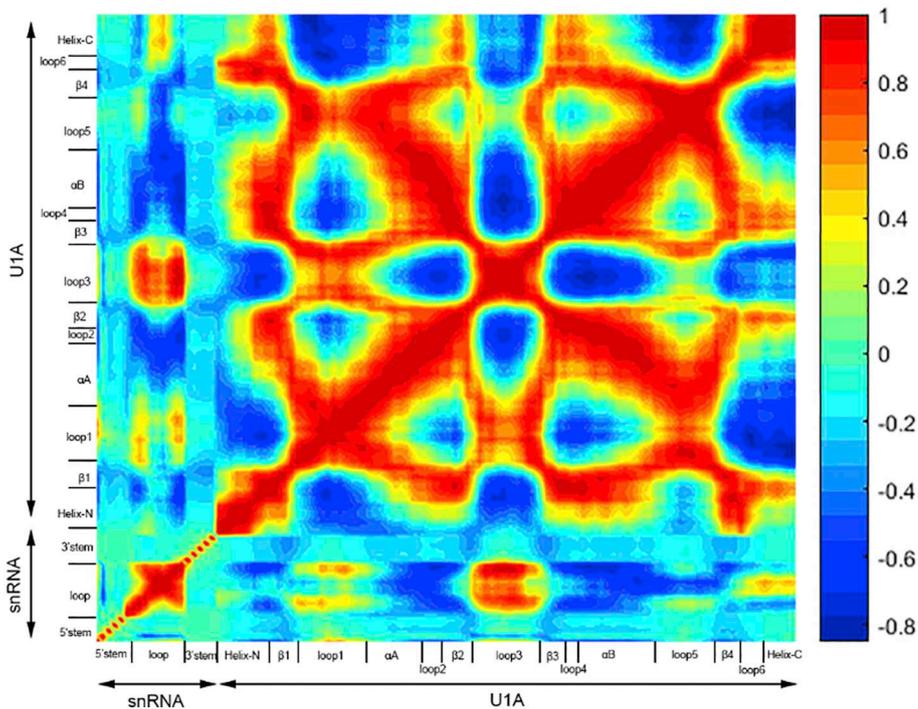


Fig. 3. Fluctuation cross-correlations calculated using the dominant ten lowest motional modes for U1A-snRNA complex. As shown in the color bar, the blue regions indicate negative correlations and the green-yellow-red regions present positive correlations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

node i . Betweenness is a measure of the centrality of a node in a network, and in some sense can be regarded as a measure of the influence that a node has over the spread of information through the network.

Another two basic parameters are the degree of a node and the clustering coefficient of a whole network. The former is defined as the number of its direct connections to other nodes, which is a centrality measure of the local connectivity in the interaction network. The latter, clustering coefficient CC is a measure of the probability for the neighbors of a node to be also neighbors of each other, which can be computed by

$$CC = \frac{1}{N} \sum_{i=1}^N \frac{2e_i}{k_i(k_i - 1)} \quad (15)$$

where k_i is the number of neighbors of node i , and e_i is the total number of edges actually connecting the neighbors of node i .

2.3. U1A-snRNA complex system

The crystal structure of U1A-snRNA with Protein Data Bank (PDB) code 1URN [7] is used to construct the pfGNM model, and the weighted residue complex network model of the complex structure.

3. Results and discussion

3.1. Theoretical B factors of residues in the pfGNM model of complex structure

The U1A-snRNA complex structure was modeled as the pfGNM model with each residue simplified as a node, and the interaction between any two nodes simulated as a spring of the inverse-square distance dependent force constant (see 2. MATERIALS AND METHODS). To test the feasibility of the model, we compared the theoretical and experimental B-factors of the residues. As the absolute value of the spring coefficient γ does not affect the relative size of residue fluctuations, it has no influence on the correlation between the calculated and experimental B-factors, and the cross-correlations between residue fluctuations [26]. Therefore, $\gamma = 1$ was adopted here. The correlation coefficient between the calculated and experimental B-factors is 0.651, as shown in Fig. 2, indicating that this simplified model is constructed reasonably and can be applied to the following analyses about cross-correlations between residue fluctuations.

3.2. Movement coupling between residues

Cross-correlations between residue fluctuations reflect the movement coupling and interactions between residues. For U1A-snRNA complex, we calculated the cross-correlations based on eq. (9). Fig. 3 gives the results obtained based on the first ten slowest motional modes accounting for more than 50% of the residue fluctuation. From Fig. 3, snRNA loop is found to be strongly positively correlated with U1A loop3, while modestly positively correlated with U1A loop1 and Helix C, which is in accordance with the crystal structure where many hydrogen bonding and hydrophobic interactions form for the former while only some contacts and several hydrogen-bonding interactions do through side chains for the latter [7,13,46–49]. The four β sheets

involved in the extensive contacts with RNA have relatively strong positive correlations with each other due to the formation of a hydrophobic core within them, and meanwhile have more or less positive correlations with Helix-C [47].

However, in the snRNA-free U1A, the correlations among the four β sheets and between them and Helix-C (see Fig. S1) are evidently weaker than the corresponding ones in the snRNA-bound U1A, and are even negative for the correlations of $\beta 1$ - $\beta 3$, $\beta 2$ - $\beta 3$, $\beta 1$ -Helix-C, $\beta 2$ -Helix-C. The results hint that RNA binding strengthens the interactions among β sheets and mediates local motional cooperativity between β -sheets and Helix-C, consistent with the observation made earlier [50–52].

3.3. Small-world characteristics of residue network and local hub residues

We employed a complex network analysis for U1A-snRNA system, which treats the complex as a network of interacting residue pairs with the edge weights coming from the cross-correlations between residue fluctuations obtained from pfGNM model. As the slow motional modes represent the large-scale collective motions associated with protein functions [53], the first ten slowest motional modes accounting for more than 50% of the residue fluctuation were utilized to produce edge weights. Then, the characteristic path length CPL (eq. (12)) and the clustering coefficient CC (eq. (15)) were calculated, as shown in Table 1. For comparison, the results for the regular and random networks with the same size as the U1A-snRNA residue network are also shown in Table 1. From Table 1, the CPL value of U1A-snRNA network is of the same order as that of the random one, and far less than that of the regular one. For the clustering coefficient CC , the former is seven times more than the middle one and less than the latter. Thus, based on the small value in the characteristic path length and the relatively large value in the clustering coefficient, we conclude that the U1A-snRNA residue network has small-world properties, in agreement with the amino acid network of proteins [54,55].

The residues with relatively large node degrees are often defined as hubs which are as well as hypothesized to act as a central backbone for information transduction, allowing for rapid integration and dissemination of the information [56–58]. To reveal the critical role played by the hub residues in information flow within the structure, we analyzed the position distributions of the hubs (~10% of all residues) with high degrees of connectivity exceeding the threshold of eleven (see Table S1). As shown in Fig. 4, these 12 hubs are mostly located at U1A four-stranded antiparallel β -sheets (Thr11, Ile12, Tyr13 and Asn15 located at $\beta 1$, Ile58 and Phe59 situated at $\beta 3$, Tyr86 in $\beta 4$), two in loop3 (Met51, Arg52), one in loop5 (Met82), one in loop6 (Ala87) and one in αB (Ala65). Most of these hubs are dispersed across the U1A-snRNA binding interface, forming a series of hydrogen bonds such as Arg52-A6 and Tyr86-C10, hydrophobic interactions such as Tyr13-C10 and Ala87-C10, as well as electrostatic interactions such as Arg52-G16 [47]. These hubs mostly appear on the interface, suggesting that they probably mediate the allosteric signal transduction between U1A and snRNA.

3.4. Identification of key residues by characteristic path length change

It is conceivable that the residues that play an important role in receiving and propagating the allosteric signal should be central in the interaction network, lying on the shortest pathways between most residue pairs in the protein. Thus, we calculated the Z-score of the change in characteristic path length when one node and its links are removed from the network, which is a measure of its effect on communication within the entire network, as shown in Fig. 5 (a). From this figure, there are 11 residue clusters whose central residues (Ala2, Asn9, Arg52, Gly53, Ser71, Gln73, Ile93, Ile94, Ala95, Lys96 and Met97) are of higher Z-score values ($Z\text{-score}_k \geq 1.5$). For clarity, the central residues are mapped on the tertiary structure of the complex system, as shown in Fig. 5 (b). In accordance to the locations of these clusters in the

Table 1

Comparison in network parameters of U1A-snRNA residue network with the same-sized regular and random networks.

Network parameters	U1A-snRNA residue network	Regular network	Random network
CPL	3.82	7.41	2.47
CC	0.56	0.88	0.07

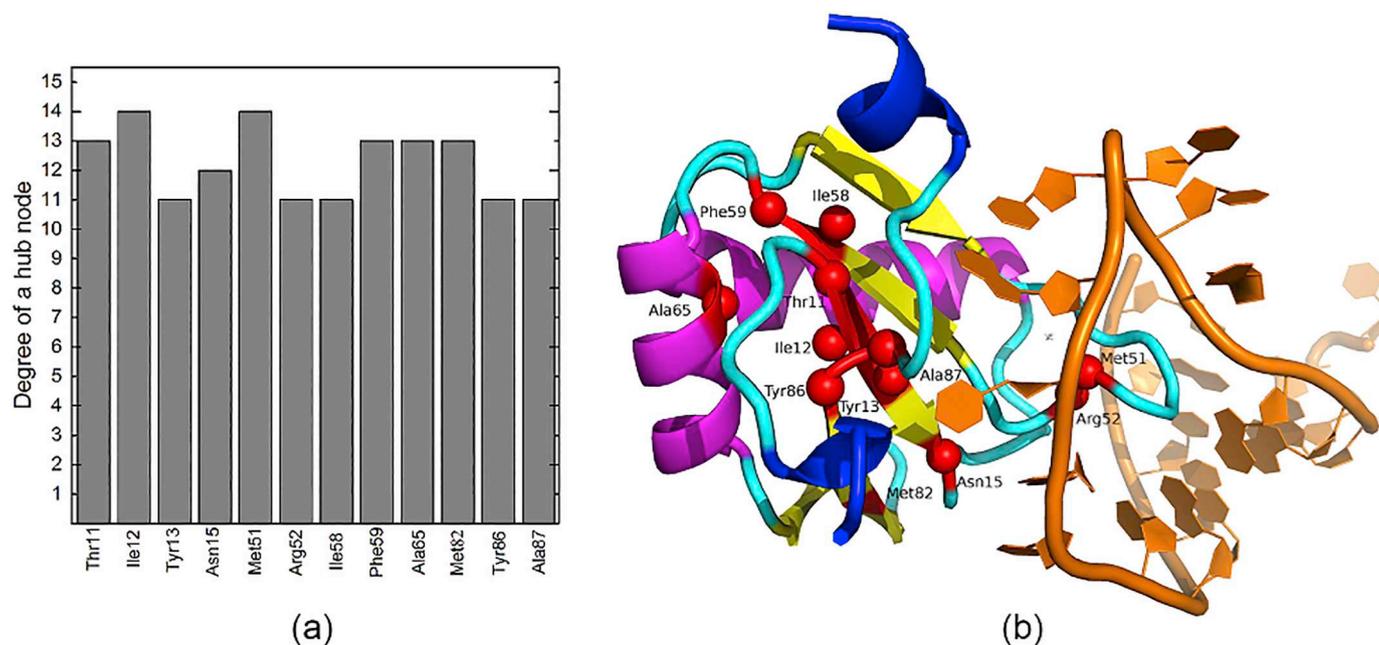


Fig. 4. Hub residues with high degrees of connectivity exceeding the threshold of eleven. (a) Degree values of these hub residues. (b) Positions of these hub nodes in U1A-snrRNA complex structure.

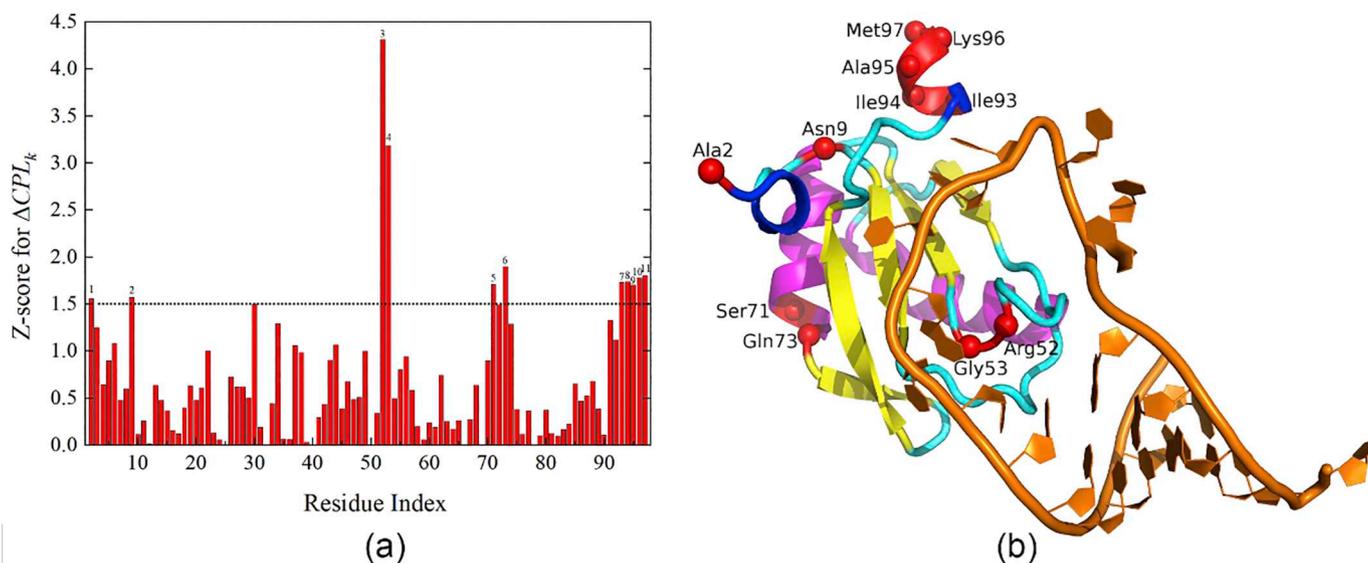


Fig. 5. Identified key residue clusters. (a) Z-score value of the change in the characteristic path length (ΔCPL_k) when one node and its links are removed from the U1A-snrRNA structure network. The clusters of key residues with relatively high Z-score values ($Z\text{-score}_k \geq 1.5$) are marked by the numbers 1–11. (b) Locations of the central residues for 11 clusters of key residues.

structure, they can be classified into two groups which are located respectively at Helix-N (clusters 1–2), Helix-C (clusters 7–11) regions, and loop regions (clusters 3–6). The functional information of them will be discussed in detail below by their comparison with the available experimental and theoretical data.

For the first group of residue clusters 1–2 and 7–11 at Helix-N and Helix-C regions, a little far away from the binding interface, they are of considerable flexibility (see Fig. 2), especially for the Helix-C which has a large conformational change or reorientation upon snRNA binding. In cluster 1, the residue Arg7 (adjacent to central residue Ala2 in space) at Helix-N is positively charged and its mutation leads to a loss of structure stability, as well as affects the binding kinetics slightly [59]. Previous research has showed that positively charged residues could have significant influence on RNA binding even if their distances from RNA reach

11 Å [60]. In cluster 2, the mutation of Thr11 at Helix-N to Ile could result in its hydrogen bond loss with Ser91 in Helix-C, which affects indirectly Ser91-A11 hydrogen bonding interaction [61], blocking signal communication among snRNA, Helix-C and Helix-N. As for Helix-C, the experimental and theoretical studies have demonstrated its significant contribution to RNA binding affinity [13,62]. Also, the position of Helix-C in U1A-snrRNA structure is determined mostly by hydrophobic interactions of its residues Ile93, Ile94 and Met97 (in clusters 7, 8 and 11) with His10 (in cluster 2) at Helix-N [61]. The truncation, removal or disruption of the Helix-C will result in a considerable loss (sometimes 100 fold) of complex stability [13], suggesting that its reorientation is probably for strengthening the interaction of U1A with snRNA, consistent with Law's point of view [13]. Many identified key residues at Helix-C combined with the experimental data indicate that Helix-C is a

critical part that bears the allosteric signal transmission, which facilitates achieving global conformational movements and ensuring normal progression of snRNA binding.

For the key residue clusters at loop regions, clusters 3 and 4 (centered at Arg52 and Gly53) are located at U1A loop3, cluster 5 (centered at Ser71) situated at α B very adjacent to loop5, and cluster 6 (centered at Gln73) at loop5. For U1A loop3, it protrudes through the RNA loop and plays vital roles for the induced fit as U1A anchors to RNA bases and for the high stability of the complex structure [47]. Thus, loop3 is important for the allosteric transition and the identified key residue clusters 3 and 4 are located in this part. Many experimental and theoretical studies have found that residue Arg52 (in cluster 3) plays a critical role for U1A-snRNA binding [10,47,52]. The mutations of the adjacent residue Gln54 account for a severe loss of RNA binding affinity [63]. The substitution of residue Gly53 (in cluster 4) with either Ala or Val can decompose the conformation of loop3, which is destructive for the induced fit in the allosteric communication between U1A loop3 and snRNA, leading to a dramatic decrease in RNA binding affinity and specificity [1]. Additionally, from the network point of view, Amitai *et al.* have shown that the key residues involved in the allosteric information communication tend to have high centrality in the structure [32]. Surprisingly, we also found that Arg52 and Gly53 have a higher betweenness centrality value (ranked 12th and 6th in all 96 residues, respectively, shown in Table S2), and the former is also with a high degree value (see Fig. 4), which suggests that these positions are crucial for the allosteric information transmission in the structure.

Finally, for the residue clusters 5 and 6 (centered at Ser71 and Gln73 respectively), we have not yet found the report on their importance for allosteric communications. However, they are strategically located between U1A β 3 and β 4 and connect two RNA binding regions like a bridge [7], indicating that they possibly belong to the allosteric information transfer station.

In fact, the residues with relatively high Z-scores commonly have a significant effect on the binding affinity or specificity between U1A and snRNA, which has been verified by experimental data. For instance, among these identified crucial residues, Arg52 and Gln53 have the largest Z-score values. Mutation of Arg52 completely abolishes the U1A-RNA binding [7]. Also, the RNA-binding assays show that the affinity between them is reduced by nearly 1.6×10^4 fold when Gly53 is replaced by Val [1]. For other key residues with similar Z-scores, mutational analyses reveal the magnitudes of the reduction in the binding affinity are relatively modest with about 10-fold magnitude [13,59].

3.5. Community analyses of U1A-snRNA system

To further reveal how U1A-snRNA structure is organized in the recognition and interaction, we applied the Girvan-Newman algorithm to partition U1A-snRNA complex into structurally contiguous communities. The algorithm splits the residue network of U1A-snRNA into four communities as shown in Fig. S2. The nodes in the same community are adjacent in structure but can be distant in sequence. Of the four communities, there are two communities including a combination of snRNA and U1A protein, and two communities containing only protein residues.

The red community with the most nodes is the largest one, and also has the most connections to other three ones. It consists of U1A Helix-N, Helix-C, part of α B, and most part of the four-stranded β -sheets. There are mainly hydrophobic interactions among residues in this community. For instance, the Helix-C is restricted by the interactions between its three hydrophobic core residues Ile93, Ile94 and Met97 and residues His10 (in Helix-N), Leu41 (in β 2), Ile58 (in β 3) and Val62 (in α B) in this community. Besides, there are important stacking interactions such as the interaction between Tyr13 (in β 1) and Phe56 (in β 3). The previous study has shown that the removal of the aromatic side chain of Tyr13 is very disruptive, leading to a dramatic decrease in structure stability [63]. The cyan community includes the whole α A, part of loop1 as well

as snRNA 5'stem, and this community contains the most positively charged residues (all positively charged residues Lys20 and Lys22 in loop1 and Lys23, Lys27, Lys28, His31 and Arg36 in α A). When RNA binding occurs, since the backbone phosphates are negatively charged, the positively charged residues are essential to attract snRNA and anchor it in correct position of U1A protein. The previous experiments have shown that the mutations of Lys20, Lys22 and Lys23 can result in a significant reduction of the electrostatic interactions with snRNA [59]. The orange community contains snRNA loop and 3'stem, and U1A loop3 and part of loop1, representing the primary U1A-snRNA recognition site. The importance of the orange community is emphasized by its incorporation of the critical U1A loop3 region which protrudes through the snRNA loop and locks the conformation of the complex with the hydrogen bonds formed between Arg52 (in loop3) with RNA A6 and G16 [47]. Tang and Nilsson have pointed out that U1A loop3 plays a critical role in the induced fit between U1A and snRNA [47]. The yellow community bridges β 4 and α B regions and mainly comprises of U1A loop5, the end part of α B and the beginning part of β 4. Our analyses on the characteristic pathway length found that the identified key residues (Ser71 in α B and Gln73 in loop5) in this community are of a higher betweenness centrality value (ranked 7th and 1st in all 96 residues respectively, see Table S2), which suggests that it may act as a transfer station for allosteric information flow within the complex structure.

Based on the analyses above, we identified four community structures which play different roles within the complex structure. The red one is closely related to the stability of complex structure. The cyan one is important to the electronic attraction between U1A and snRNA. The orange is critical to the induced fit between the two molecules. And the yellow possibly acts as a transfer station of information communication within the complex structure.

4. Conclusions

We extended the parameter-free Gaussian network model (pfGNM) to the exploration of the effect of snRNA binding on the dynamics of U1A protein domain which is one of the major components of the spliceosomal U1 small nuclear ribonucleoprotein particle. The results reveal that snRNA binding strengthens the interactions among U1A β -sheets, and between the main binding surface β -sheets and Helix-C.

Furtherly, the complex network model with the edge weights coming from the cross-correlations between residue fluctuations obtained from pfGNM was utilized to examine the residues critical for the allosteric signal communication. The residue network model of U1A-snRNA complex structure is of small-world properties. The residues with high degrees of connectivity are mostly situated at the binding surface, implying their critical roles in information flow transmission within the structure. By removing one node one time, we simulated the attack on the complex network to identify the key residues in the allosteric communication process. The identified key residues are found in two groups according to their locations in the structure: [1] the residues at Helix-N and Helix-C with high flexibility and far away from the interface, which are thermodynamically coupled with the binding of RNA, and are important for the long-range allosteric signal transmission; [2] the residues at loop regions, which are mostly positively charged and highly flexible, and largely contribute to the induced fit and the high binding affinity between U1A and snRNA. Finally, we explored how the complex structure is organized using the community detecting algorithm. The structure is divided into four communities with different functional roles of the stability of complex structure, electronic attraction, induced fit and information transfer station in response to snRNA binding, which is found by combining previous experimental and theoretical data.

Considering the relatively good consistence of our results with experimental data, we believe that taking the cross-correlations between residue fluctuations in slow motional modes from pfGNM as the edge

weights in residual network model is a novel and efficient approach to analyze the allosteric signal communication and organization of the structure. The results can help reveal the relationship between molecular topological structures and their binding and allosteric dynamics.

Author statement

Shao Qi and Li Chunhua: designed the research, performed data analysis and wrote the draft. Shao Qi and Gong Weikang: developed the program. Gong Weikang: draw the figures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [31971180 and 11474013], and the Beijing Natural Science Foundation [4152011].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bpc.2020.106393>.

References

- [1] S.A. Showalter, K.B. Hall, Altering the RNA-binding mode of the U1A RBD1 protein, *J. Mol. Biol.* 335 (2004) 465–480.
- [2] C.G. Burd, G. Dreyfuss, Conserved structures and diversity of functions of RNA-binding proteins, *Science* 265 (1994) 615–621.
- [3] K.B. HALL, Interaction of RNA hairpins with the human U1A N-Terminal RNA binding domain, *Biochemistry-US* 33 (1994) 10076–10088.
- [4] W.T. Stump, K.B. Hall, Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain, *RNA* 1 (1995) 55–63.
- [5] D.B.W.V. Scherly, Identification of the RNA binding segment of human U1A protein and definition of its binding site on U1 snRNA, (1989).
- [6] K.B. Hall, W.T. Stump, Interaction of N-terminal domain of U1A protein with an RNA stem/loop, *Nucleic Acids Res.* 20 (1992) 4283–4290.
- [7] C. Oubridge, et al., Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature* 372 (1994) 432–438.
- [8] D. Anunciado, et al., Characterization of the dynamics of an essential helix in the U1A protein by time-resolved fluorescence measurements, *J. Phys. Chem. B* 112 (2008) 6122–6130.
- [9] D. Anunciado, A. Dhar, M. Gruebele, A.M. Baranger, Multistep kinetics of the U1A-SL2 RNA complex dissociation, *J. Mol. Biol.* 408 (2011) 896–908.
- [10] C.M. Reyes, P.A. Kollman, Investigating the binding specificity of U1A-RNA by computational mutagenesis, *J. Mol. Biol.* 295 (2000) 1–6.
- [11] F. Pitić, D.L. Beveridge, A.M. Baranger, Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: do molecular sub-states code for specificity? *Biopolymers* 65 (2002) 424–435.
- [12] I. Guzman, et al., Native conformational dynamics of the spliceosomal U1A protein, *J. Phys. Chem. B* 119 (2015) 3651–3661.
- [13] M.J. Law, et al., The role of the C-terminal helix of U1A protein in the interaction with U1hpII RNA, *Nucleic Acids Res.* 41 (2013) 7092–7100.
- [14] V. Tozzini, Coarse-grained models for proteins, *Curr. Opin. Struct. Biol.* 15 (2005) 144–150.
- [15] M.G. Saunders, G.A. Voth, Coarse-graining methods for computational biology, in: K.A. Dill (Ed.), *Annual Review of Biophysics*, 42 (2013), pp. 73–93.
- [16] J.G. Su, et al., Prediction of allosteric sites on protein surface with an elastic-network-model-based thermodynamic method, *Phys. Rev. E* 90 (2014).
- [17] T. Haliloglu, I. Bahar, B. Erman, Gaussian dynamics of folded proteins, *Phys. Rev. Lett.* 79 (1997) 3090–3093.
- [18] L.W. Yang, et al., oGNM: online computation of structural dynamics using the Gaussian network model, *Nucleic Acids Res.* 34 (2006) W24–W31.
- [19] E. Guarnera, Z.W. Tan, Z. Zheng, I.N. Berezovsky, AlloSigMA: allosteric signaling and mutation analysis server, *Bioinformatics* 33 (2017) 3996–3998.
- [20] E. Guarnera, I.N. Berezovsky, Structure-based statistical mechanical model accounts for the causality and energetics of allosteric communication, *PLoS Comput. Biol.* 12 (2016) e1004678.
- [21] E. Guarnera, I.N. Berezovsky, Toward comprehensive allosteric control over protein activity, *Structure* 27 (2019) 866–878.e1.
- [22] X.L. Xie, et al., Allosteric transitions of ATP-binding cassette transporter MsbA studied by the adaptive anisotropic network model, *Proteins* 83 (2015) 1643–1653.
- [23] C. Li, et al., 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 (2016) 014104.
- [24] Z. Han, et al., Interpreting the dynamics of binding interactions of snRNA and U1A using a coarse-grained model, *Biophys. J.* 116 (2019) 1625–1636.
- [25] A.R. Atilgan, et al., Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophys. J.* 80 (2001) 505–515.
- [26] L. Yang, G. Song, R.L. Jernigan, Protein elastic network models and the ranges of cooperativity, *P. Natl Acad. Sci. USA* 106 (2009) 12347–12352.
- [27] A.N. Naganathan, Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function, *Curr. Opin. Struct. Biol.* 54 (2019) 1–9.
- [28] Z. Liang, G.M. Verkhivker, G. Hu, Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications, *Brief. Bioinform.* 21 (2020) 815–835.
- [29] M.D. Daily, J., J. Gray, Allosteric communication occurs via networks of tertiary and quaternary motions in proteins, *PLoS Comput. Biol.* (2009) 5.
- [30] A.J. Rader, S.M. Brown, Correlating allostery with rigidity, *Mol. Biosyst.* 7 (2011) 464–471.
- [31] L. Di Paola, A. Giuliani, Protein contact network topology: a natural language for allostery, *Curr. Opin. Struct. Biol.* 31 (2015) 43–48.
- [32] G. Amitai, et al., Network analysis of protein structures identifies functional residues, *J. Mol. Biol.* 344 (2004) 1135–1146.
- [33] A.R. Atilgan, P. Akan, C. Baysal, Small-world communication of residues and significance for protein dynamics, *Biophys. J.* 86 (2004) 85–91.
- [34] A. Del Sol, H. Fujihashi, D. Amoros, R. Nussinov, Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families, *Protein Sci.* 15 (2006) 2120–2128.
- [35] O. Gacsi, A topological description of hubs in amino acid interaction networks, *Adv. Bioinforma.* 2010 (2010) 257512.
- [36] M.P. Joy, A. Brock, D.E. Ingber, S. Huang, High-betweenness proteins in the yeast protein interaction network, *J. Biomed. Biotechnol.* (2005) 96–103.
- [37] A. Del Sol, H. Fujihashi, D. Amoros, R. Nussinov, Residues crucial for maintaining short paths in network communication mediate signaling in proteins, *Mol. Syst. Biol.* 2 (2006).
- [38] M. Girvan, M. Newman, Community structure in social and biological networks, *P. Natl Acad. Sci. USA* 99 (2002) 7821–7826.
- [39] S. Bernhard, F. Noe, Optimal identification of semi-rigid domains in macromolecules from molecular dynamics simulation, *PLoS One* 5 (2010) e10491.
- [40] M. Bhattacharyya, S. Vishyeshwara, Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism, *Biochemistry-US* 50 (2011) 6225–6236.
- [41] A.T. VanWart, J. Eargle, Z. Luthey-Schulten, R.E. Amaro, Exploring residue component contributions to dynamical network models of allostery, *Abstr. Pap. Am. Chem. Soc.* 245 (2013).
- [42] C.L. McClendon, A.P. Kornev, M.K. Gilson, S.S. Taylor, Dynamic architecture of a protein kinase, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) E4623–E4631.
- [43] M.T. Zimmermann, R.L. Jernigan, Elastic network models capture the motions apparent within ensembles of RNA structures, *RNA-A Publ. RNA Soc.* 20 (2014) 792–804.
- [44] I. Bahar, R.L. Jernigan, Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms, *J. Mol. Biol.* 281 (1998) 871–884.
- [45] A. Sethi, J. Eargle, A.A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA: protein complexes, *P. Natl Acad. Sci. USA* 106 (2009) 6620–6625.
- [46] B.L. Kormos, A.M. Baranger, D.L. Beveridge, A study of collective atomic fluctuations and cooperativity in the U1A-RNA complex based on molecular dynamics simulations, *J. Struct. Biol.* 157 (2007) 500–513.
- [47] Y. Tang, L. Nilsson, 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 (1999) 1284–1305.
- [48] S.A. Showalter, K.B. Hall, Correlated motions in the U1 snRNA stem/loop 2: U1A RBD1 complex, *Biophys. J.* 89 (2005) 2046–2058.
- [49] I. Kurisaki, M. Takayanagi, M. Nagaoka, Combined mechanism of conformational selection and induced fit in U1A-RNA molecular recognition, *Biochemistry-US* 53 (2014) 3646–3657.
- [50] J.K. Kranz, K.B. Hall, RNA binding mediates the local cooperativity between the beta-sheet and the C-terminal tail of the human U1A RBD1 protein, *J. Mol. Biol.* 275 (1998) 465–481.
- [51] J.K. Kranz, K.B. Hall, 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 (1999) 215–231.
- [52] K. Nagai, et al., Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A, *Nature* 348 (1990) 515–520.
- [53] I. Bahar, A.R. Atilgan, M.C. Demirel, B. Erman, Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability, *Phys. Rev. Lett.* 80 (1998) 2733–2736.
- [54] M. Vendruscolo, N.V. Dokholyan, E. Paci, M. Karplus, Small-world view of the amino acids that play a key role in protein folding, *Phys. Rev. E* 65 (2002).
- [55] L.H. Greene, V.A. Higman, Uncovering network systems within protein structures, *J. Mol. Biol.* 334 (2003) 781–791.
- [56] V. Colizza, A. Flammini, M.A. Serrano, A. Vespignani, Detecting rich-club ordering in complex networks, *Nat. Phys.* 2 (2006) 110–115.
- [57] B. Misic, O. Sporns, A.R. McIntosh, Communication efficiency and congestion of signal traffic in large-scale brain networks, *PLoS Comput. Biol.* 10 (2014).

- [58] M.P. van den Heuvel, R.S. Kahn, J. Goni, O. Sporns, High-cost, high-capacity backbone for global brain communication, *P. Natl Acad. Sci. USA* 109 (2012) 11372–11377.
- [59] M.J. Law, et al., 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 (2006) 275–285.
- [60] C. Garcia-Garcia, D.E. Draper, Electrostatic interactions in a peptide-RNA complex, *J. Mol. Biol.* 331 (2003) 75–88.
- [61] F.H. Allain, P.W. Howe, D. Neuhaus, G. Varani, Structural basis of the RNA-binding specificity of human U1A protein, *EMBO J.* 16 (1997) 5764–5772.
- [62] Q.Y. Zeng, K.B. Hall, Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability, *RNA* 3 (1997) 303–314.
- [63] M.J. Law, et al., 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 (2005) 2917–2928.