



Allosteric mechanism for SL RNA recognition by polypyrimidine tract binding protein RRM1: An atomistic MD simulation and network-based study

Zhongjie Han¹, Zhixiang Wu¹, Weikang Gong, Wenxue Zhou, Lei Chen, Chunhua Li^{*}

Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Keywords:

PTB
Allosteric
MD simulation

ABSTRACT

Polypyrimidine tract-binding protein (PTB), an RNA-binding protein, is involved in the regulation of diverse processes in mRNA metabolism. However, the allosteric modulation of its binding with RNA remains unclear. We explore the dynamic characteristics of PTB RNA recognition motif 1 (RRM1) in its RNA-free and wild-type/mutant RNA-bound states to understand the issues using molecular dynamics (MD) simulation, perturbation response scanning (PRS) and protein structure network (PSN) models. It is found that RNA binding strengthens RRM1 stability, while L151G mutation in $\alpha 3$ helix far away from the interface makes the complex unstable. The latter is caused by long-distance dynamic couplings, which makes intermolecular electrostatic and entropy energies unfavorable. The weakened couplings between interface β sheets and C-terminal parts upon mutation reveal RNA recognition is co-regulated by these regions. Interestingly, PRS analysis reveals the allosteric caused by the perturbation on $\alpha 3$ helix has already been pre-encoded in the equilibrium dynamics of the protein structure. PSN analysis shows the details of the allosteric signal transmission, revealing the necessity of strong couplings between $\alpha 3$ helix and interface for maintaining the high binding affinity. This study sheds light on the mechanisms of PTB allostery and RNA recognition and can provide important information for drug design.

1. Introduction

Protein-RNA interactions play essential roles in numerous cellular processes, including gene expression and regulation [1]. In mammalian cells, >1000 diverse proteins can interact with RNA using different types of motifs, such as RNA recognition motif (RRM), double-stranded RNA binding motif (dsRBM), arginine-rich motif, GXXG motif, tetra loops (GX [GA]A) and so on [2,3]. RRM is the most abundant RNA-binding domain and is present in approximately 0.5 %–1.0 % of human genes [4]. Exploring the interactions of RRMs with target RNAs can not only promote the understanding of their specific recognitions but also provide important information for the development of related drugs.

Polypyrimidine tract-binding protein (PTB), composed of four RRMs, also called heterogeneous nuclear ribonucleoprotein I (hnRNP I), is an abundant eukaryotic RNA binding protein that can recognize CU-rich RNA and is implicated in many aspects of mRNA metabolism including splicing regulation, translation initiation, 3' end processing and mRNA stability [5–8]. The first two N-terminal RRMs (RRM1 and

RRM2) of PTB are connected by a 42 amino acid linker and can tumble independently, while the two C-terminal RRMs (RRM3 and RRM4) interact extensively with each other [9]. The RRM2 was recently shown to display only local adaptations of the backbone upon binding to RNA stem-loop structure [10], the tandem RRM3/RRM4 prefers single-stranded RNA sites [11], but the behavior of RRM1 upon interacting with a structured RNA target is still unknown. To gain a better insight into the ability of PTB to recognize structured elements in RNA targets, in 2020 Maris et al. solved the nuclear magnetic resonance (NMR) structure of RRM1 in complex with a stem-loop (SL) RNA that contains a UCUUU apical loop [12].

The structure of PTB RRM1 domain in complex with SL RNA is composed of three α -helices and four β -strands (Fig. 1). Their residue numbers are $\alpha 1$ (71–78), $\alpha 2$ (103–115), $\alpha 3$ (144–154), $\beta 1$ (60–64), $\beta 2$ (85–90), $\beta 3$ (96–101) and $\beta 4$ (127–130). The contacts between the protein and RNA mainly involve the protein β -sheet and RNA apical loop UCUUU. The CUU triplet of the apical loop lies on the β -sheet surface and makes base-specific interactions with the extended C-terminal loop

^{*} Corresponding author.

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

¹ Zhongjie Han and Zhixiang Wu contribute equally to this work.

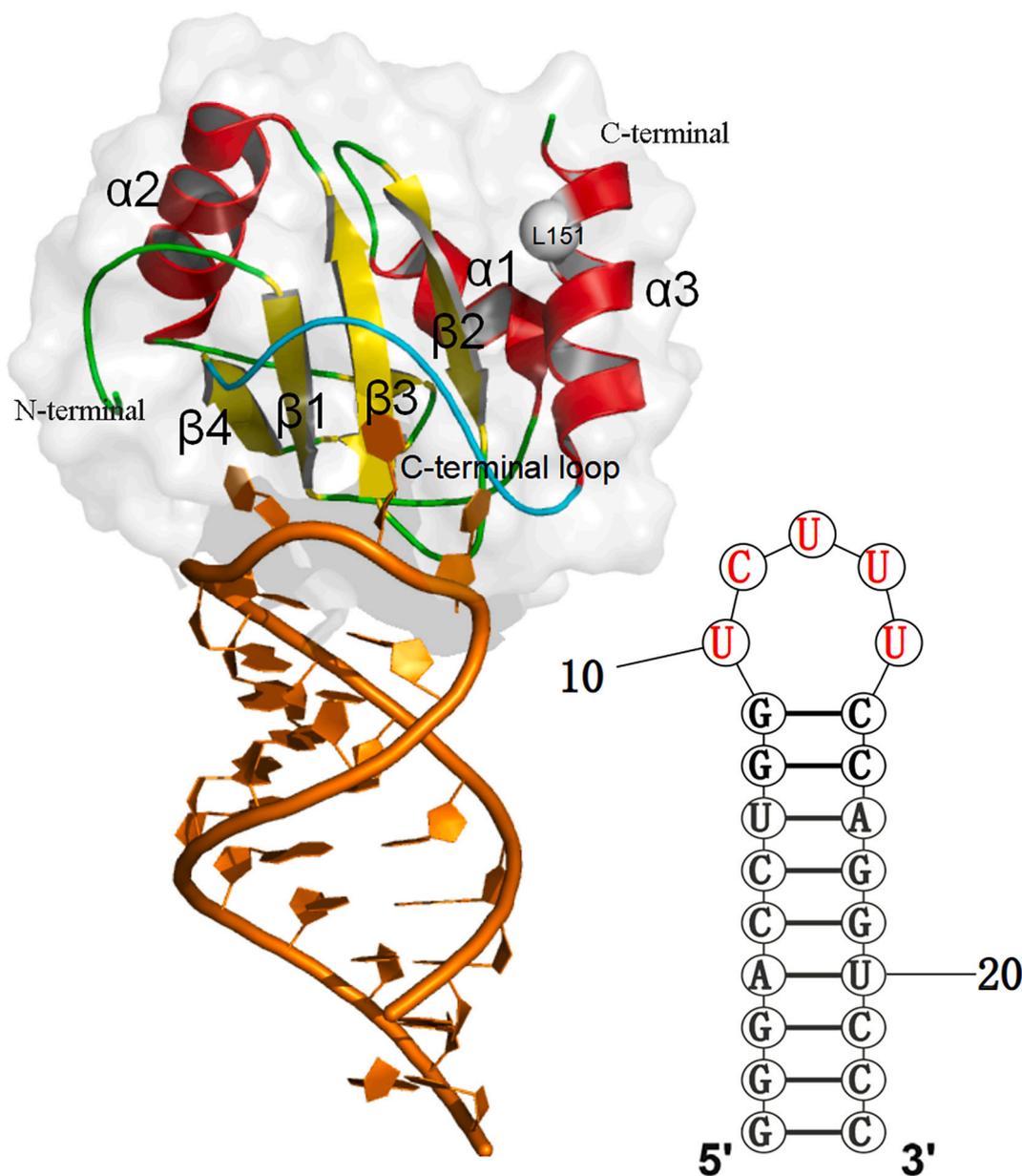


Fig. 1. Cartoon representation of PTB RRM1-RNA complex structure (PDB ID 2n3o) and RNA secondary structure. The protein helices, β sheets and C-terminal loop, and RNA are labeled respectively, and residue Leu151 is represented by a ball.

which also interacts with the β -sheet and forms an arch. Interestingly, the structure reveals a new C-terminal $\alpha 3$ helix that folds upon RRM1 binding RNA hairpin. Experiments have shown that the $\alpha 3$ helix acts as a sensor of RNA secondary structure [12], although it is far away from the interface and does not form any contact with RNA. Furthermore the residue mutations in the helix decrease the binding affinity of RNA to RRM1 [12], which implies that there exists allosteric phenomenon in the system. The recognition mechanism of PTB RRM1 with the target SL RNA and the long-distance allosteric regulation of the $\alpha 3$ helix for the binding affinity have attracted extensive attention from experimental and theoretical researchers.

For the recognition mechanism of PTB RRM1 with its target RNA, using NMR measurements, Simpson et al. explored the N-terminal RRM domains (RRM1 and RRM2) of PTB in 2004, giving the first insights into the structure and RNA binding properties of PTB [13]. Maris et al. performed site-directed mutagenesis on the PTB RRM1-RNA interface (H62A, K94A and P142G) and in the $\alpha 3$ helix (L151G) to evaluate their importance for intermolecular interactions. Combining NMR with

mutagenesis, binding and splicing assays, Joshi et al. presented the crystal structure of PTB RRM2 in complex with the Raver1 protein, revealing the molecular basis for the RNA binding [14]. Using molecular dynamics (MD) simulations and thermodynamic integration (TI) methods, Schmid et al. studied the stability of the tertiary fold of PTB RRM3/RRM4 and the sequence-specific RNA recognition by them [15]. For the phenomenon of long-distance allosteric regulation which is a common event in RRM domains mainly for strengthening intermolecular interactions [16–18], several computer simulation methods have been developed to explore the issue [19,20]. Our group extended the Gaussian network model (GNM)-based thermodynamic cycle method to investigate the impact of snRNA binding on the dynamics of the human protein U1A RRM (related to alternative splicing), revealing a significant loss of flexibility of the C-terminal helix upon snRNA binding, and the key sites for the binding-coupled opening of C-terminal helix [21]. Later, for the interaction of RNA with TDP-43 (an alternative splicing regulator containing two RRM domains), we utilized the equally weighted multiscale elastic network model (ewmENM) with the multiscale effect

of interactions considered, and MD simulations to characterize the large conformational arrangement between RRM1s upon RNA binding, and demonstrated that some mutations far away from the interface affect TDP-43-RNA binding affinity, revealing the long-distance allosteric regulation in the system [22]. These studies show that RRM1s can take a diverse allosteric way to strengthen their interactions with target RNAs. For PTB RRM1, experiments have revealed a binding-coupled change in the C-terminal $\alpha 3$ helix upon RNA binding, but now the allosteric mechanism is not completely clear.

Here, we perform all-atom MD simulations of PTB RRM1 in its RNA-free, wild-type and mutant RNA-bound states to investigate its dynamic properties and RNA recognition mechanism. In addition, we explore the protein allosteric properties using perturbation response scanning method and reveal the inner allosteric mechanism and signal communication using dynamic protein structure network model.

2. Materials and methods

2.1. Biomolecular systems and molecular dynamics simulation protocols

The complex of PTB RRM1 with the target RNA determined by NMR spectroscopy was obtained from the Protein Data Bank (PDB) [23] with PDB ID 2n3o. The nineteenth frame was selected as the representative structure which has the lowest average root mean square deviation (RMSD) from all the other frames. The free PTB RRM1 structure was built from the complex with RNA eliminated. The L151G mutant complex was constructed using CharmmGUI [24]. For convenience, the free PTB RRM1, wild-type and L151G mutant complexes are designated as PTB-WT, complex-WT and complex-L151G respectively. Three independent molecular dynamics simulations were performed using the GROMACS 2020 package [25] with the CHARMM36 all-atom force field [26,27]. In each simulation, the initial structure was solvated in a cubic periodic water box with the edges of the box at least 10 Å from any part of the solute. The three simulation systems contain 18,924, 17,533 and 17,540 TIP3P water molecules with a total of 59,814, 55,885 and 55,892 atoms, respectively. Sodium and chloride ions were added to each system to obtain a final ion concentration of 0.15 M. Each system was subjected to a rigorous energy minimization using the steepest descent and conjugated gradient algorithms with a tolerance of 100 kJ•mol⁻¹•nm⁻¹. Then, the system was equilibrated under NVT for 1 ns using position restraint which allows the solvent to equilibrate around the protein without disturbing the protein structure, and then under NPT for 1 ns unconstrained equilibrium simulation. Afterward, the systems were submitted for unbiased MD production runs of 1000 ns each. All simulations were conducted using the LINCS algorithm [28] to constrain bond lengths and angles involving the hydrogen atoms. The pressure was kept at 1 bar using the Parrinello-Rahman coupling algorithm by applying a semi-isotropic coupling constant of $\tau = 1$ ps. The temperature was kept at 303 K using Nose-Hoover with a coupling constant of $\tau = 0.1$ ps. A cutoff of 1.2 nm was used for van der Waals interactions, and long-range electrostatic interactions were computed using the PME method [29]. For each system, the MD integration step was set as 2 fs, and one snapshot was sampled every 5000 steps. A total of 100,000 conformations were collected for each trajectory for further analysis.

2.2. Dynamic cross-correlation map

Movement cross-correlation coefficient C_{ij} between two atoms i and j during the simulation is defined as follows:

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i \cdot \Delta r_i \rangle \langle \Delta r_j \cdot \Delta r_j \rangle}} \quad (1)$$

where Δr_i and Δr_j are the instantaneous displacements of the i th and j th atoms from their mean positions respectively, and $\langle \dots \rangle$ represents the trajectory average. A map describing the movement cross-correlation

coefficients between atoms in a molecular system is often called dynamic cross-correlation map (DCCM). The positively correlated residues move in the same direction, that is, $C_{ij} > 0$, while negatively correlated ones move in the opposite direction, that is, $C_{ij} < 0$. C_{ij} values are calculated for the C α atoms in PTB and P atoms in RNA.

2.3. Computation of binding free energy

The Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) [30] approach is used to calculate the binding free energy between RRM1 domain and RNA, which is carried out by gmx_MMPBSA [31]. The binding free energy (ΔG_{bind}) is calculated from the MD trajectory by the following equations:

$$\Delta G_{bind} = G_{pro+RNA} - (G_{pro} + G_{RNA}) = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{ele} \quad (3)$$

$$\Delta G_{sol} = \Delta G_{PB} + \Delta G_{SA} \quad (4)$$

In Eq. (2), $G_{pro+RNA}$, G_{pro} and G_{RNA} represent the free energies of the protein-RNA complex, protein and RNA in solvent, respectively. ΔG_{bind} can be decomposed into three terms: gas-phase energy ΔE_{MM} , solvation free energy ΔG_{sol} and entropy $-T\Delta S$ at temperature T . ΔE_{MM} is composed of van der Waals energy (ΔE_{vdW}) and electrostatic energy (ΔE_{ele}) as shown in Eq. (3). In Eq. (4), ΔG_{sol} includes nonpolar (ΔG_{SA}) and polar (ΔG_{PB}) terms, and the former is estimated using the solvent accessible surface area (SASA) and the latter calculated by the Poisson-Boltzmann (PB) model, where the solvent and solute dielectric constants are set to 78.54 and 2 respectively. The entropy ($-T\Delta S$) is calculated by the interaction entropy (IE) method proposed by Duan et al. [32]. The gmx_MMPBSA is also used to estimate the energy contribution per residue/nucleotide to the binding free energy. The last 200 ns in the MD simulations are used to calculate the binding free energy.

2.4. Perturbation response scanning approach

The perturbation response scanning (PRS) approach [33] based on linear response theory (LRT) [34] was designed to deduce protein allosteric properties. The method can be used to calculate the response of residue k to the perturbation at residue i . The 3 N -dimensional vector ΔR of node displacements in response to the exertion of a perturbation (3 N -dimensional force vector F) obeys Hooke's law $F = H \Delta R$, where H is the 3 $N \times 3 N$ Hessian matrix in the Anisotropy Network Model (ANM) theory. The force exerted on residue i is written as:

$$F^i = (000 \dots \Delta F_x^i \Delta F_y^i \Delta F_z^i \dots 000)^T \quad (5)$$

and the resulting response is:

$$\Delta R^i = H^{-1} F^i \quad (6)$$

where ΔR^i a 3 N -dimensional vector describes the displacements of all the residues away from their equilibrium positions in response to the exerted force F^i which are nonzero only for the three terms related to residue i .

Here, the average value of the squared residue k displacements ($\|\Delta R_k^i\|^2$) in respective response to multiple exerted perturbations on residue i is taken as the sensitivity of residue k to the perturbation at residue i [35]. The forces are along seven directions respectively, i.e., x-, y-, z-, both x- and y-, both x- and z-, both y- and z-, and all x, y, z directions. Repeating (scanning) this procedure for all sites yields a response matrix P (PRS heat map) with a size of $N \times N$, each column of which provides a measure of the sensitivities of all residues to the perturbation at the residue corresponding to the column. The average over the columns of the normalized P yields the sensitivity profile, and the residues corresponding to the peaks are often the functional residues

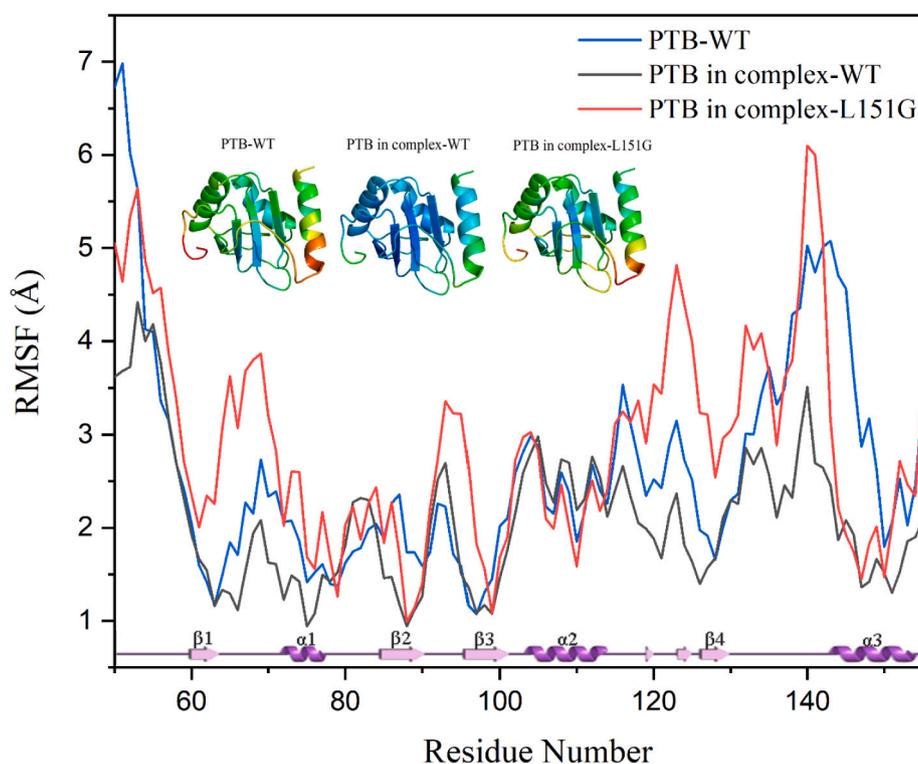


Fig. 2. Comparison of residue RMSFs for protein parts in PTB-WT, complex-WT and complex-L151G systems respectively, with RMSF values mapped on the corresponding structures also shown where blue and red colors represent rigid and flexible regions respectively.

as sensors involved in protein allostery [35].

2.5. Dynamic protein structure network model

Protein structure and function rely on the complex network of inter-residue interactions [36]. Based on the graph theory, a weighted molecular structure network model is constructed for the protein-RNA complex, where each residue is represented as a node (Ca for protein and P for RNA), and the node pairs within a cutoff distance (7, 13 and 10 Å for protein, RNA and interface regions respectively [21]) are connected by the edges of a weight from their fluctuation cross-correlations (C_{ij}) computed from the MD trajectory. The weight w_{ij} of the edge connecting nodes i and j is expressed by:

$$w_{ij} = -\log(|C_{ij}|) \quad (7)$$

For a network, the characteristic path length (CPL) is defined as the average length of the shortest paths between all pairs of nodes in the network:

$$CPL = \frac{1}{N_p} \sum_{j>i}^N d_{ij} \quad (8)$$

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 (ΔCPL_k) after removing node k from the network [37]. A Z-score analysis is used to measure the relative change of CPL:

$$Z\text{-score}_k = \left| \frac{\Delta CPL_k - \langle \Delta CPL_k \rangle}{\sigma} \right| \quad (9)$$

where ΔCPL_k is the change of CPL after removal of node k , $\langle \Delta CPL_k \rangle$ is the change ΔCPL_k averaged over all the nodes, and σ is the standard deviation.

3. Results and discussion

3.1. Comparative analyses of MD trajectories

Three 1000 ns MD simulations were carried out for PTB-WT, complex-WT and complex-L151G systems. Fig. S1 shows the time evolutions of RMSDs of PTB-WT, complex-WT and complex-L151G systems during MD simulations respectively. From Fig. S1, the RMSD values of the three systems are all stable after 100 ns, and thus the MD trajectories of the last 900 ns are chosen for further analyses. The complex-L151G system possesses a higher RMSD value (5.28 ± 0.90 Å) than complex-WT (5.05 ± 0.57 Å) (Fig. S1A), the main reason for which is from the protein part's higher RMSD in the former (Fig. S1B). The higher standard deviation indicates that the complex-L151G system is more unstable, consistent with the experimental result that L151G mutation reduces the binding affinity [12]. Additionally, it is found that RNA binding evidently decreases the RMSD of the protein part, especially for that in complex-WT (Fig. S1B), indicating that RNA binding is beneficial for the stability of PTB structure, while the mutation makes the protein unstable to some extent.

To detect the effect of RNA binding on PTB RRM1 flexibility, we calculated the root mean square fluctuations (RMSFs) of backbone C α atoms of RRM1 in PTB-WT, complex-WT and complex-L151G systems respectively, as shown in Fig. 2, with the corresponding results mapped on protein structures also shown. From Fig. 2, most of the protein regions in complex-WT system possess lower RMSF values than those in PTB-WT system, especially for C-terminal loop and $\alpha 3$ helix, indicating again that RNA binding increases protein stability. In contrast, upon mutation, the protein in complex-L151G has higher RMSF values than that in complex-WT, especially for the regions within and around $\beta 1$, $\beta 3$, $\beta 4$ and C-terminal loop which are mainly located at the protein-RNA binding interface, indicating that the L151G mutation is not conducive to the intermolecular interactions. Additionally, for RNA part (Fig. S2), its nucleotide RMSF values in complex-WT are evidently lower than those in complex-L151G, especially for nucleotides U10, C11 and U13

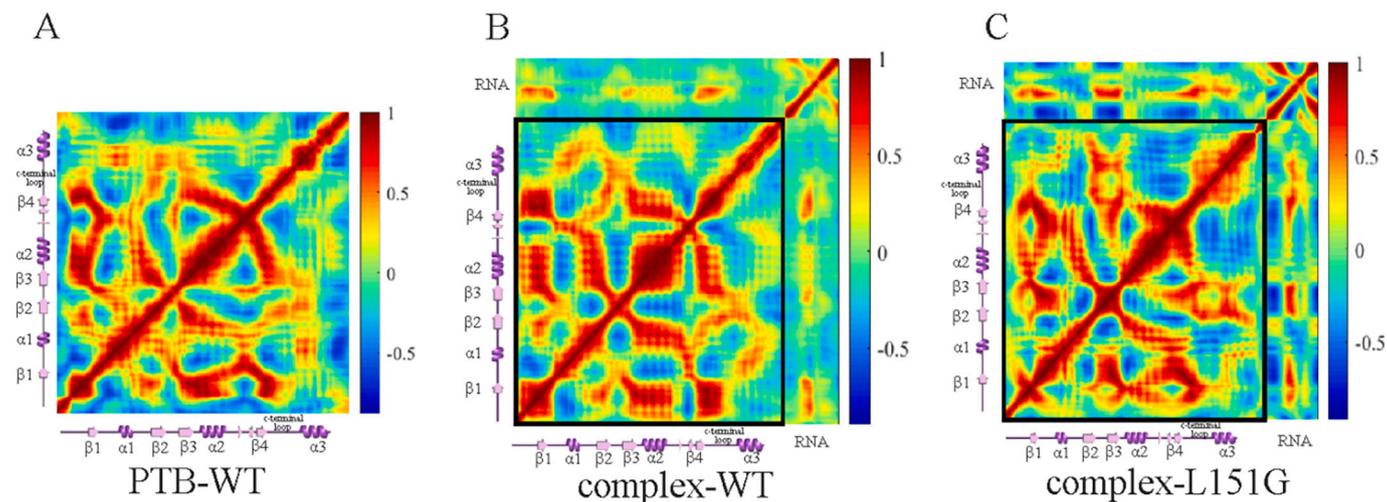


Fig. 3. Dynamic cross-correlation maps of PTB-WT (A), complex-WT (B) and complex-L151G (C) systems, respectively.

which are mainly responsible for interactions with protein, implying that the mutation disrupts the intermolecular interface.

In summary, RNA binding increases protein stability, and the mutation L151G in $\alpha 3$ helix far away from the interface makes the protein unstable to some extent and further disrupts the protein-RNA interface, revealing that there exists an interesting long-distance allosteric regulation.

3.2. Analyses of movement coupling

The dynamic coupling within RRM1 and between it and RNA can be detected by dynamic cross-correlation map (DCCM). According to Eq. 1, the DCCMs for PTB-WT, complex-WT and complex-L151G systems are presented in Fig. 3. For PTB-WT (Fig. 3A), some positive correlations are observed among the antiparallel β sheets, and meanwhile $\alpha 1$, $\beta 2$ and $\beta 3$ have certain positive correlations with $\alpha 3$ helix, indicating there exist motion couplings between $\alpha 3$ helix and the interface. Upon RNA binding (Fig. 3B), the correlated motions in complex-WT are enhanced and

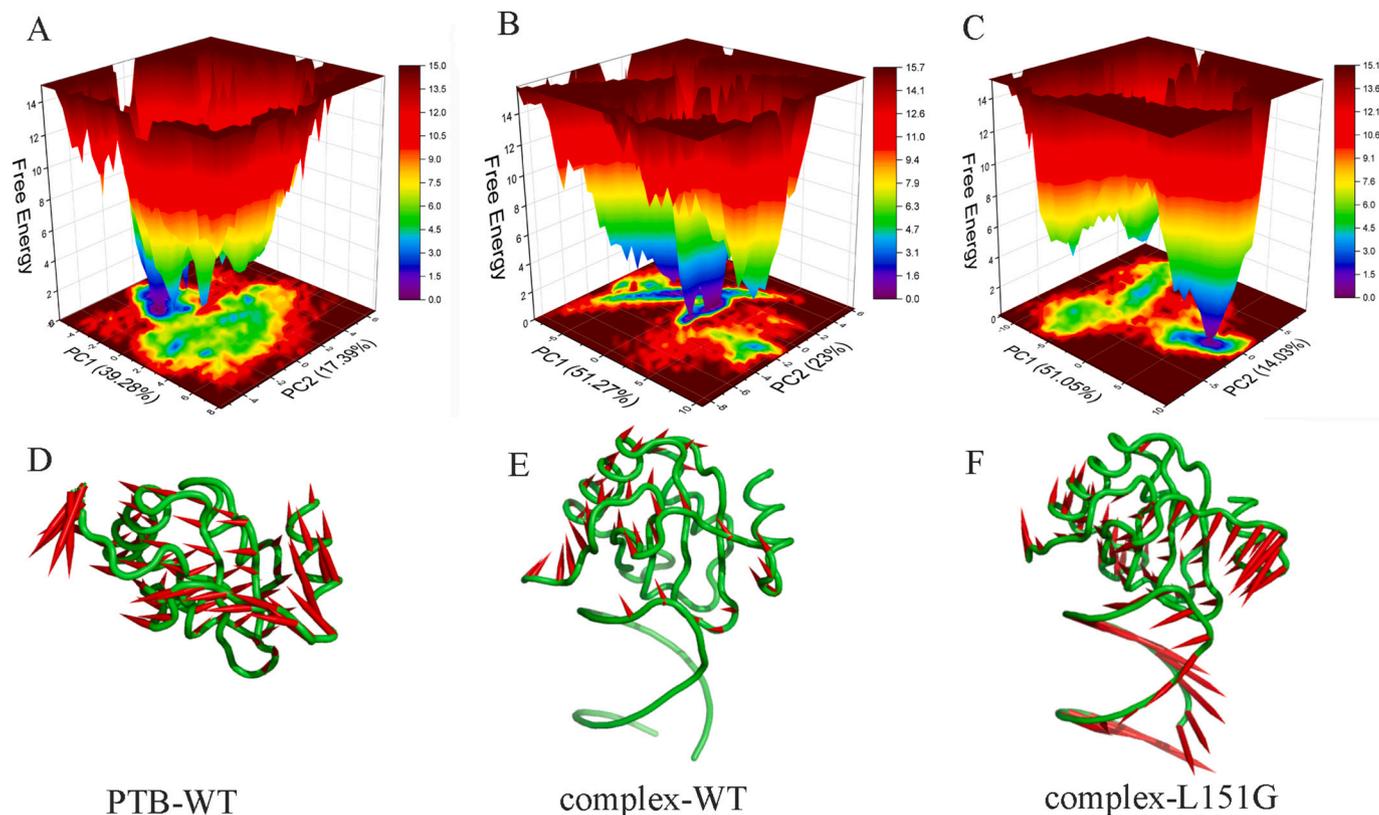


Fig. 4. Free energy contour maps depicted along the first two principle components PC1 and PC2 for PTB-WT (A), complex-WT (B) and complex-L151G (C) systems. The first principle component (PC1s) mapped on the corresponding average structures are displayed for PTB-WT (D), complex-WT (E) and complex-L151G (F) systems. The PC1s are shown as cone model with the length of cone proportional to the motion magnitude and the orientation of cone indicating the motion direction.

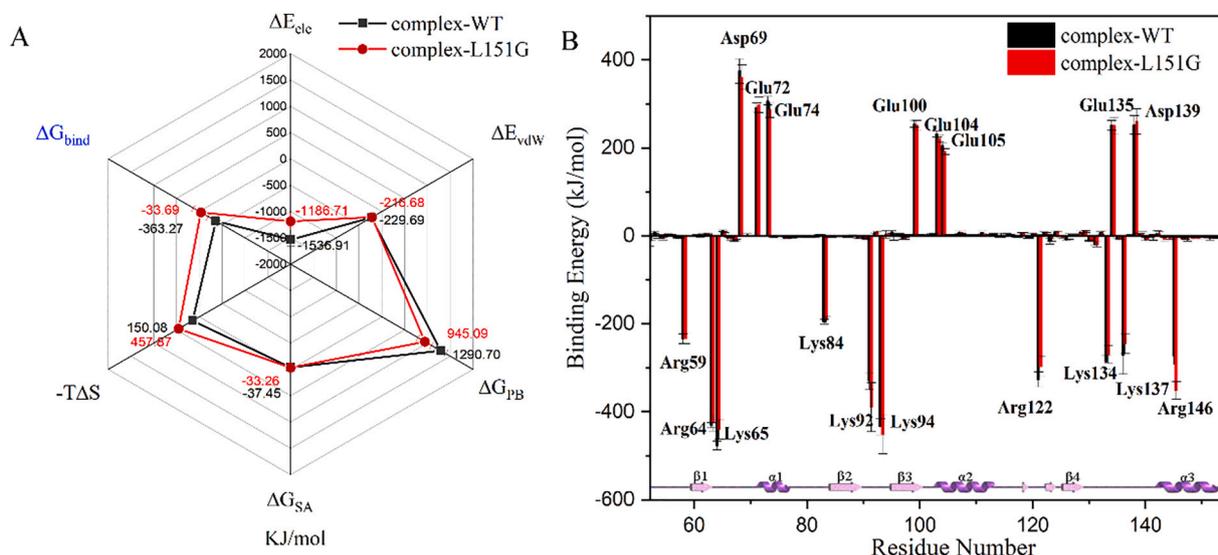


Fig. 5. Comparison of RRM1-RNA binding free energy components between complex-WT and complex-L151G systems (A) and the binding energy contribution per residue from the two complexes (B).

extended, presenting a more modular form. It is noted that the couplings among the β sheets at the interface are strengthened, and meanwhile the couplings between the C-terminal loop and interface, and between $\alpha 3$ helix and interface are strengthened to some degree, which is beneficial for the stable binding of RNA with protein. After mutation (Fig. 3C), the motion correlations within complex-L151G become less modular and more like those within PTB-WT, and there are more negatively correlated motions between $\beta 4$ sheet and $\alpha 3$ helix. Meanwhile, there are weakened couplings between $\beta 4$ and the other β strands and between the C-terminal parts and some β strands, which means that the mutation affects the couplings at the interface, between the interface and C-terminal loop, and between interface and $\alpha 3$ helix. For RNA in complex-WT, it shows positive correlations with protein antiparallel β sheets and C-terminal loop, which attributes to the formed interfacial interactions including hydrogen bonds and intermolecular stacking interactions [12]. Upon L151G mutation, the evident change is that the ends of RNA have a strong negative correlation with protein β sheets, which we think is not beneficial for the intermolecular interactions.

In summary, the binding of RNA strengthens the protein motion couplings, and L151G mutation in $\alpha 3$ helix weakens the couplings within and between local areas to some extent, especially for the couplings between the C-terminal parts and RNA binding interface. The results reveal that the RNA recognition by PTB is co-regulated by the C-terminal parts and β sheets, and there exists a long-range coupling between the $\alpha 3$ helix and the interface, which will be analyzed in the following section.

3.3. Motion mode analyses of the three systems

To further explore the principle motions related to the function exertion of PTB RRM1, both principal component analysis (PCA) and free energy landscape (FEL) analyses were carried out for the three MD trajectories. Fig. 4 shows the free energy contour maps depicted along the first two principal components (PCs) 1 and 2, with the first slowest motion modes PC1s mapped on the average structures for the three systems respectively. As shown in Fig. 4A, B and C, PC1 captures 39.28%, 51.27% and 51.05% of the system's variance for PTB-WT, complex-WT and complex-L151G systems respectively, suggesting that the cooperative motions of the whole system are enhanced by RNA binding. The energy barriers among the energy funnels in complex-WT are higher than those in the other two systems, which means that the complex-WT system has fewer conformation states for PTB's RNA recognition. From

Fig. 4E and F, compared with the smaller amplitude movement in complex-WT, complex-L151G has a much larger amplitude movement (consistent with the RMSF analysis), especially for the protein C-terminal loop and $\alpha 3$ helix, and RNA stem part, which we think is not beneficial for protein-RNA binding stability. The above results indicate the effect of RNA binding and mutation in $\alpha 3$ helix on the overall structural dynamics and stability of protein-RNA binding interface, highlighting the vital role of the $\alpha 3$ helix for PTB's RNA recognition.

Fig. S3 shows the superimposed low-energy conformations of complex-WT and complex-L151G systems. From Fig. S3, the main differences in PTB RRM1 are located at the C-terminal parts. The $\alpha 3$ helix has a partial unfolding, and the structural changes caused by the mutation give rise to the changes in the interface. Evidently, the C-terminal loop in complex-L151G is farther away from the β sheets than that in complex-WT, which causes a certain displacement of RNA apical loop in complex-L151G. We speculate that the position of the C-terminal loop is important for the stable binding of RNA to protein. In the following section, we will analyze the detailed interactions in the complex systems.

3.4. Comparison of interactions in the two complex systems

The binding free energies of the two complex systems were calculated by MM-PBSA method, with the results shown in Fig. 5A. From Fig. 5A, the complex-L151G system has a higher binding free energy, which is consistent with the experimental result that L151G mutation reduces the binding affinity [12]. The lower binding affinity attributes to the higher electrostatic energy change, and larger entropy reduction during the binding process of RNA with the mutant protein than with the wild-type protein. The unfavorable electrostatic energy change is partially due to the broken intermolecular hydrogen bonds (see analyses below), and the unfavorable entropy change comes from more unstable residue-nucleotide interactions due to the introduced mutation.

Next, we used the energy decomposition strategy to analyze the contributions of each residue and nucleotide to the binding free energy, with the results shown in Fig. 5B. From Fig. 5B, the main residues contributing favorable energies for RNA binding in the two complexes are the common 10 ones which are mainly located around the binding interface. Interestingly, we notice that the residue Leu151 far away from the binding interface contributes little to the energy, but its mutation L151G evidently decreases the contributions of most of the 10 interface residues to the binding free energy, suggesting that the L151G mutation

Table 1

Hydrogen bonds formed between RRM1 and RNA with occupancy over 40 % for complex-WT and complex-L151G systems.

Complex-WT			Complex-L151G		
RRM1	RNA	Occupancy	RRM1	RNA	Occupancy
PHE130-O	CYT11-N4	87.48 %	PHE130-O	CYT11-N4	82.67 %
ASN132-N	CYT11-O2	73.41 %	ASN132-N	CYT11-O2	69.67 %
SER131-OG	CYT11-O2	72.90 %	SER131-OG	CYT11-O2	66.09 %
SER131-CA	CYT11-N3	60.67 %	SER131-CA	CYT11-N3	52.61 %
ASN132-N	CYT11-N3	48.58 %	ASN132-N	CYT11-N3	49.70 %
ASN132-N	CYT11-C2	47.27 %	ASN132-N	CYT11-C2	40.00 %
ARG52-NH2	GUA8-O1P	59.45 %			
ARG64-NH2	GUA9-O2P	55.81 %			
ASN95-OD1	URA10-N3	53.67 %			
LYS65-NZ	GUA8-N7	53.54 %			

The common hydrogen bonds in the two systems are highlighted in bold.

affects the binding affinity through long-distance allosteric regulation. Additionally, we note that the residues evidently favorable or unfavorable for RNA binding are all the charged ones, reflecting the electrostatic interaction-dominated RNA binding in this system.

Furthermore, we analyzed the hydrogen bonds between RRM1 and RNA, which were calculated by VMD 1.9.2 [38]. Table 1 shows the

intermolecular hydrogen bonds with an occupancy over 40 % for the two complexes. In complex-WT system, there are 10 interface hydrogen bonds formed by residues Phe130, Ser131 and Asn132 at the C-terminal loop, Arg64 and Lys65 at/adjacent to β 1, as well as Asn95 adjacent to β 3 and RNA nucleotides C11, G9, G8 and U10, while upon L151G mutation only the above 6 hydrogen bonds involving the C-terminal loop and C11 are maintained with lower occupancy generally. The results can partially explain the reason for the higher flexibility of RRM1 C-terminal loop and β sheets, and RNA G8, G9, U10 and C11, and the reason for the unfavorable intermolecular electrostatic energy change upon L151G mutation. The above result also reflects the important role of the C-terminal loop in stabilizing the RNA binding.

In summary, although L151G mutation is far away from protein-RNA binding interface, it decreases the binding affinity by altering the binding dynamics which makes the intermolecular electrostatic and entropy energies unfavorable. This long-distance dynamic coupling is the base for the allosteric modulation.

3.5. Allosteric characteristics of PTB RRM1 analyzed by PRS

The perturbation response scanning (PRS) approach combines the elastic network model (ENM) with LRT to assess the allosteric response of each protein residue to an external perturbation on the protein. We used PRS method to explore the allosteric characteristics and mechanism of PTB RRM1. First, the traditional ANM of PTB-WT was constructed where Ca atoms are taken as nodes and the node pairs within a cutoff distance are connected by springs of a uniform force constant. The cutoff distance was determined by maximizing the Pearson correlation coefficient (PCC) between the residue RMSFs obtained from the MD ensemble and the ANM model. Here, an optimal cutoff of 17 Å was

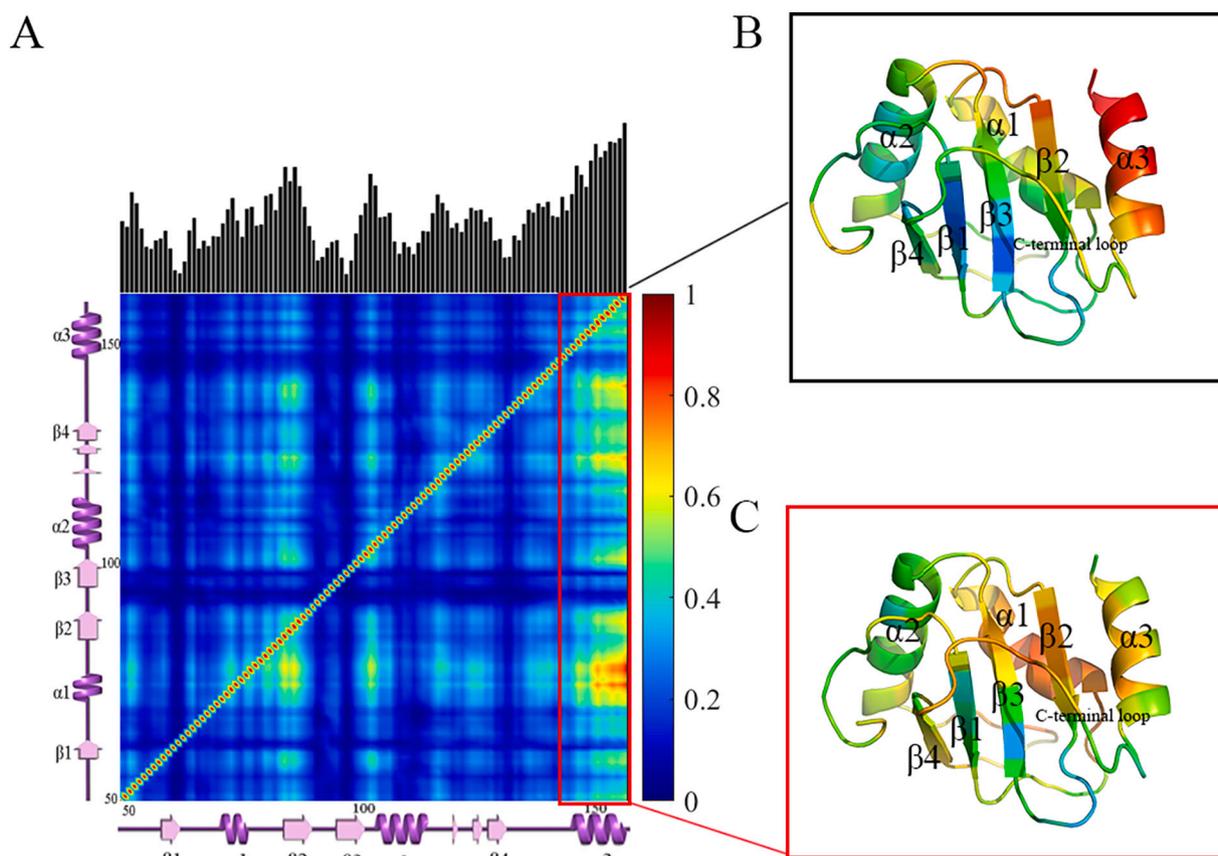


Fig. 6. PRS heat map of PTB-WT (A). The bar plot along the higher abscissa describes the potentials of residues to serve as sensors with the peaks indicating the residues having a high potential to serve as sensors. The potential is also mapped on the structure (B). The average responses of the other residues to the perturbations on the α 3 helix residues are mapped on the structure (C).

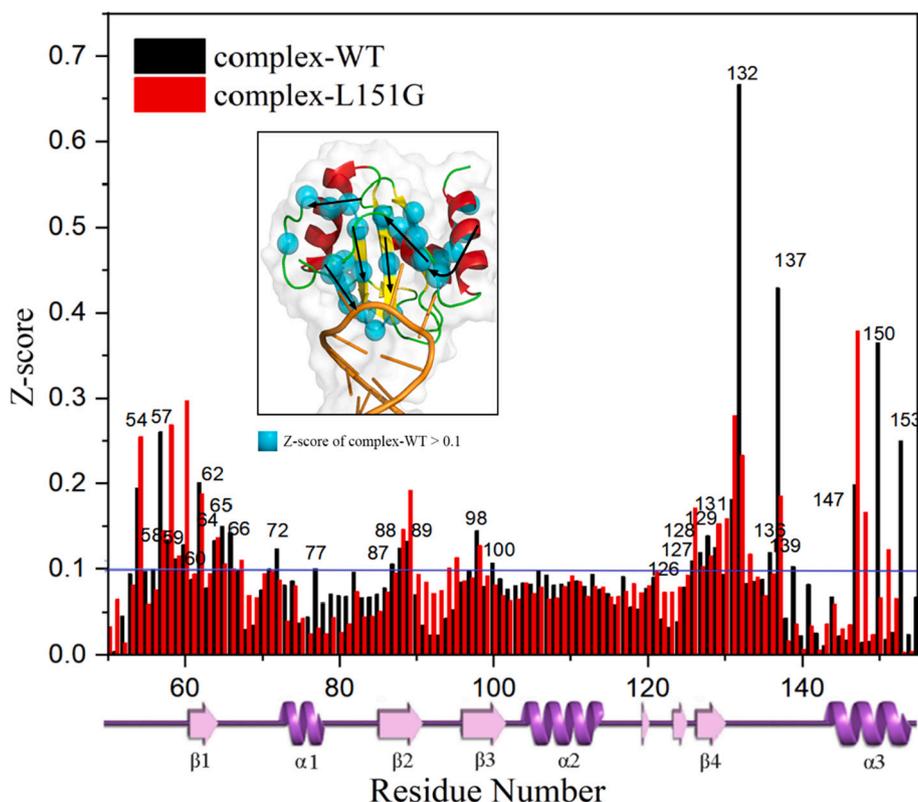


Fig. 7. Z-score value of the change in the characteristic path length (Δ CPL) when one node and its links are removed from the PTB-RNA structure network for Complex-WT and Complex-L151G systems. The residues with Z-score >0.1 in complex-WT are mapped on its structure. The black arrows represent the signaling pathway we speculate.

determined with $PCC=0.61$. Then, we calculated the PRS heat map and sensitivity profile according to the scheme mentioned in materials and methods, with the result presented in Fig. 6A and the sensitivity profile also mapped on the PTB structure (Fig. 6B).

The PRS heat map describes the propensities of residues to sense perturbations and thus elicit cooperative responses, such as an allosteric conformational change induced upon ligand binding to a highly “sensitive” sensor [35]. From the bar plot (sensitivity profile) in Fig. 6A, some residues show high signals, suggesting their potential as a “sensitive” sensor. The residues with the strongest signals are mainly located at the C-terminal part, especially at the $\alpha 3$ helix (Fig. 6B), which hints that the $\alpha 3$ helix is a highly sensitive sensor in RRM1 allostery, consistent with the experimental result that $\alpha 3$ helix is responsible for RNA stem-loop recognition [12]. Additionally, it is noted that the columns corresponding to the $\alpha 3$ helix reflect that upon perturbation on $\alpha 3$ helix, the interfacial C-terminal loop, $\beta 2$, $\beta 3$, $\beta 4$ and $\beta 1$, as well as $\alpha 1$ helix have a significantly strong response, which again verifies that there exists a strong dynamic coupling between the $\alpha 3$ helix and the interface. Coincidentally, this strong coupling also occurs in the equilibrium dynamics of PTB RMM1 (Fig. 3A), suggesting that the dynamic allostery has already been pre-encoded in RRM1’s equilibrium dynamics, which is consistent with Hacisuleyman’s point of view on molecular allostery [39]. Furthermore, we mapped the average responses of the other residues to multiple perturbations on $\alpha 3$ helix residues on PTB RRM1 structure (Fig. 6C) to explore the allosteric response caused by the $\alpha 3$ helix. From Fig. 6C, it is likely that upon perturbation on $\alpha 3$ helix, $\alpha 1$ helix and $\beta 2$ have a relatively large response, and then together with the C-terminal loop affect $\beta 3$ and $\beta 1$ to a certain extent, and finally affect the remote $\beta 4$ through the C terminal loop to complete the regulation on the entire interface. The coupling between the $\alpha 3$ helix and the interface region is the main determinant for controlling the communication of allosteric signals.

In summary, perturbation response scanning is an effective approach to identify the key sensor and effector residues that function in molecular allostery, which reveals the important role of the $\alpha 3$ helix in PTB RRM1 allostery as a highly sensitive sensor.

3.6. Allosteric signal transmission in PTB RRM1 analyzed by PSN

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 protein. Thus we calculated the Z-score of the change of characteristic path length when one node and its links are removed from the residue network of complex-WT, which is a measure of its effect on communication within the entire network, the results (black bars) shown in Fig. 7. The residues with Z-scores >0.1 are mapped on the PTB RRM1 structure. From Fig. 7, the residues with a larger Z-score value are mainly distributed at the $\alpha 3$ helix, C-terminal loop and β sheets. The former two have almost the highest values, suggesting their important roles in allosteric communication, which is consistent with the PRS analysis results (Fig. 6C) that the coupling relationship between the two regions plays a pivotal role in the allosteric signal transmission and mediation; and the latter two participate in the direct interactions with RNA, indicating their key roles in mediating the signaling of protein-RNA interactions. Combined the results from the PRS analysis, here we speculate a possible signaling pathway connecting these hub residues, which starts from the $\alpha 3$ helix and passes through the C-terminal loop, transmitting signals to β sheets and thereby controlling the interactions of RRM1 with RNA (black arrows in Fig. 7). Furtherly, to see the changes in allosteric signal transmission upon L151G mutation, we also performed the corresponding calculations on complex-L151G, with the results (red bars) also shown in Fig. 7. From Fig. 7, generally the profile pattern of residue Z-score values does not change much. But

evidently the residues located at the C-terminal loop and $\alpha 3$ helix show an obvious decrease in Z-score, indicating their reduced role in signal mediation, which we think is partially due to the weakened dynamic coupling between the two regions and other parts including the interface (Fig. 3C). The weakened coupling affects the interactions between the protein and RNA. Thus, the maintenance of a high protein-RNA binding affinity requires the $\alpha 3$ helix to couple strongly to the interface, which is conducive to the tight protein-RNA interaction.

Here, based on the hub residues identified by PSN analysis, a possible allosteric signal transmission pathway is proposed. The important roles of the $\alpha 3$ helix, β sheets and C-terminal loop are revealed for signal communication. Also it is deduced that a strong coupling between the $\alpha 3$ helix and the interface is required to maintain a high protein-RNA binding affinity. The hub residues we identified can provide guidance for researchers to explore the functional sites of PTB and the design of the related drugs.

4. Conclusions

In this work, MD simulations are applied to investigate the allosteric properties and crucial interactions of PTB RRM1 domain with SL RNA. The comparative analyses of MD trajectories indicate that RNA binding increases the stability of RRM1 domain, while the mutation L151G in the $\alpha 3$ helix far away from the protein-RNA interface decreases the RRM1's stability, furtherly causing the whole system unstable. Through dynamic movement analyses, the obvious motion coupling between the C-terminal part and β sheets reveals that the RNA recognition by PTB is co-regulated by these segments. And the change of overall movement upon mutation indicates the existence of long-range couplings between the $\alpha 3$ helix and the interface. Although the L151G mutation is far away from the RNA binding interface, it changes the binding affinity by altering the binding dynamics, which makes the intermolecular electrostatic and entropy energies unfavorable, suggesting the important role of the long-distance dynamic coupling for the allosteric modulation on the intermolecular interactions. In addition, through the PRS analyses based on the ANM model, it is found that the allostery caused by the perturbation has already been pre-encoded in the equilibrium dynamics of the natural protein structure itself. Also it is found that the $\alpha 3$ helix plays a vital role in RRM1 allostery as a highly sensitive sensor, and the coupling between the $\alpha 3$ helix and interface region is the main determinant for controlling the communication of allosteric signals. Finally, through the analysis of the characteristic path length, we propose a possible allosteric signal transmission pathway. It is found that the coupling between the $\alpha 3$ helix and C-terminal loop plays a pivotal role in allosteric signal transmission and mediation. Meanwhile, the residues in the C-terminal loop and β sheets with a higher Z-score value mediate the signaling of protein-RNA interactions. Thus, the maintenance of a high protein-RNA affinity requires the $\alpha 3$ helix to couple strongly to the interface. This work sheds light on the allosteric mechanism in PTB, is helpful for the understanding of the RNA recognition by PTB, and can provide guidance for the design of the related drugs.

CRedit authorship contribution statement

Z.H. and C.L. conceived the study. Z.H. conducted MD simulations and protein structure network analysis. Z.H. and Z.W. conducted the MD analysis. W.Z. conducted the Molecular Mechanics Poisson Boltzmann Surface Area analysis. W.G. and L.C. conducted the perturbation response scanning analysis. C.L. validated the results. Z.H and C.L. wrote up the paper, and all the other authors reviewed the manuscript.

Declaration of competing interest

The authors confirm that there are no conflicts of interest.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Natural Science Foundation of China [32271294, 31971180].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2022.08.181>.

References

- [1] K.J. Riley, J.A. Steitz, The "Observer Effect" in genome-wide surveys of protein-RNA interactions, *Mol. Cell* 49 (4) (2013) 601–604.
- [2] E. Jankowsky, M.E. Harris, Specificity and nonspecificity in RNA-protein interactions, *Nat. Rev. Mol. Cell Biol.* 16 (2015) 533–544.
- [3] R. Nagarajan, S.P. Chothani, C. Ramakrishnan, M. Sekijima, M.M. Gromiha, Structure based approach for understanding organism specific recognition of protein-RNA complexes, *Biol. Direct* 10 (2015) 8.
- [4] A. Clery, M. Blatter, F.H. Allain, RNA recognition motifs: boring? Not quite, *Curr. Opin. Struct. Biol.* 18 (3) (2008) 290–298.
- [5] F.C. Oberstrass, S.D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D.L. Black, F.H. Allain, Structure of PTB bound to RNA: specific binding and implications for splicing regulation, *Science* 309 (5743) (2005) 2054–2057.
- [6] S. Dai, C. Wang, C. Zhang, L. Feng, W. Zhang, X. Zhou, Y. He, X. Xia, B. Chen, W. Song, PTB: not just a polypyrimidine tract-binding protein, *J. Cell. Physiol.* 237 (5) (2022) 2357–2373.
- [7] J.M. Pina, L.A. Hernandez, N.M. Keppetipola, Polypyrimidine tract binding proteins PTBP1 and PTBP2 interact with distinct proteins under splicing conditions, *PLoS One* 17 (2) (2022), e263287.
- [8] J.A. Hensel, S.E. Nicholas, A.L. Kimble, A.S. Nagpal, O. Omar, J.D. Tyburski, E. R. Jellison, A. Menoret, M. Ozawa, A. Rodriguez-Oquendo, A.T. Vella, P. A. Murphy, Splice factor polypyrimidine tract-binding protein 1 (Ptbp1) primes endothelial inflammation in atherogenic disturbed flow conditions, *Proc. Natl. Acad. Sci. U. S. A.* 119 (30) (2022), e212227119.
- [9] F. Vitali, A. Henning, F.C. Oberstrass, Y. Hargous, S.D. Auweter, M. Erat, F. H. Allain, Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling, *EMBO J.* 25 (1) (2006) 150–162.
- [10] G. Dorn, A. Leitner, J. Boudet, S. Campagne, C. von Schroetter, A. Moursy, R. Aebersold, F.H. Allain, Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS, *Nat. Methods* 14 (5) (2017) 487–490.
- [11] C. Clerte, K.B. Hall, The domains of polypyrimidine tract binding protein have distinct RNA structural preferences, *Biochemistry-US* 48 (10) (2009) 2063–2074.
- [12] C. Maris, S. Jayne, F.F. Damberger, I. Beusch, G. Dorn, S. Ravindranathan, F. H. Allain, A transient alpha-helix in the N-terminal RNA recognition motif of polypyrimidine tract binding protein senses RNA secondary structure, *Nucleic Acids Res.* 48 (8) (2020) 4521–4537.
- [13] P.J. Simpson, T.P. Monie, A. Szendroi, N. Davydova, J.K. Tyzack, M.R. Conte, C. M. Read, P.D. Cary, D.I. Svergun, P.V. Konarev, S. Curry, S. Matthews, Structure and RNA interactions of the N-terminal RRM domains of PTB, *Structure* 12 (9) (2004) 1631–1643.
- [14] A. Joshi, M.B. Coelho, O. Kotik-Kogan, P.J. Simpson, S.J. Matthews, C.W. Smith, S. Curry, Crystallographic analysis of polypyrimidine tract-binding protein-Raver1 interactions involved in regulation of alternative splicing, *Structure* 19 (12) (2011) 1816–1825.
- [15] N. Schmid, B. Zagrovic, W.F. van Gunsteren, Mechanism and thermodynamics of binding of the polypyrimidine tract binding protein to RNA, *Biochemistry-US* 46 (22) (2007) 6500–6512.
- [16] J.M. Avis, F.H. Allain, P.W. Howe, G. Varani, K. Nagai, D. Neuhaus, 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 (2) (1996) 398–411.
- [17] C. Oubridge, N. Ito, P.R. Evans, C.H. Teo, K. Nagai, Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature* 372 (6505) (1994) 432–438.
- [18] I. Beusch, P. Barraud, A. Moursy, A. Clery, F.H. Allain, Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7, *elife* 6 (2017).
- [19] C. Maris, C. Dominguez, F.H. Allain, The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J.* 272 (9) (2005) 2118–2131.
- [20] Frederic H.T. Allain, Charles C. Gubser, Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation, *Nature* 380 (6575) (1996) 646.

- [21] Z. Han, Q. Shao, W. Gong, S. Wang, J. Su, C. Li, Y. Zhang, Interpreting the dynamics of binding interactions of snRNA and U1A using a coarse-grained model, *Biophys. J.* 116 (9) (2019) 1625–1636.
- [22] X. Deng, S. Wang, Z. Han, W. Gong, Y. Liu, C. Li, Dynamics of binding interactions of TDP-43 and RNA: an equally weighted multiscale elastic network model study, *Proteins* 90 (2) (2022) 589–600.
- [23] H.M. Berman, J. Westbrook, Z. Feng, G.L. Gilliland, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (Pt 6 No 1) (1999) 235–242.
- [24] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem.* 29 (11) (2008) 1859–1865.
- [25] A. Mja, D. Tm, C. Rsb, A. Sp, C. Jcsb, A. Bh, D. Ela, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers - ScienceDirect, *SoftwareX* s 1–2 (2015) 19–25.
- [26] A.J. Mackerell, M. Feig, C.R. Brooks, Improved treatment of the protein backbone in empirical force fields, *J. Am. Chem. Soc.* 126 (3) (2004) 698–699.
- [27] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.J. Mackerell, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, *J. Phys. Chem. B* 114 (23) (2010) 7830–7843.
- [28] B. Hess, P-LINCS: a parallel linear constraint solver for molecular simulation, *J. Chem. Theory Comput.* 4 (1) (2007).
- [29] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh ewald method, *J. Chem. Phys.* 103 (19) (1995) 8577–8593.
- [30] N. Homeyer, H. Gohlke, Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method, *Mol. Inform.* 31 (2) (2012) 114–122.
- [31] M.S. Valdes-Tresanco, M.E. Valdes-Tresanco, P.A. Valiente, E. Moreno, Gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS, *J. Chem. Theory Comput.* 17 (10) (2021) 6281–6291.
- [32] Li Yuchen, Cong Yalong, Feng Guoqiang, Zhong Susu, Z. John, The impact of interior dielectric constant and entropic change on HIV-1 complex binding free energy prediction, *Struct. Dyn.* 5 (2018), 064101.
- [33] C. Atilgan, A.R. Atilgan, Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein, *PLoS Comput. Biol.* 5 (10) (2009), e1000544.
- [34] M. Ikeguchi, J. Ueno, M. Sato, A. Kidera, Protein structural change upon ligand binding: linear response theory, *Phys. Rev. Lett.* 94 (7) (2005) 78102.
- [35] A. Dutta, J. Krieger, J. Lee, J. Garcia-Nafria, I. Greger, I. Bahar, Cooperative dynamics of intact AMPA and NMDA glutamate receptors: similarities and subfamily-specific differences, *Structure* 23 (9) (2015) 1692–1704.
- [36] L. Di Paola, M. De Ruvo, P. Paci, D. Santoni, A. Giuliani, Protein contact networks: an emerging paradigm in chemistry, *Chem. Rev.* 113 (3) (2013) 1598–1613.
- [37] A. Sethi, J. Eargle, A.A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA: protein complexes, *Proc. Natl. Acad. Sci.* 106 (16) (2009) 6620–6625.
- [38] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1) (1996) 33–38, 27–28.
- [39] A. Hacısuleyman, A. Erkip, B.Er Man, B.Er Man, Synchronous and asynchronous response in dynamically perturbed proteins, *J. Phys. Chem. B* 125 (3) (2021).