

# Dynamics of binding interactions of TDP-43 and RNA: An equally weighted multiscale elastic network model study

Xueqing Deng  | Shihao Wang | Zhongjie Han | Weikang Gong | Yang Liu  | Chunhua Li 

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

#### Correspondence

Chunhua Li, Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China.  
 Email: chunhuali@bjut.edu.cn

#### Funding information

National Natural Science Foundation of China,  
 Grant/Award Numbers: 31971180, 11474013

## Abstract

Transactive response DNA binding protein 43 (TDP-43), an alternative-splicing regulator, can specifically bind long UG-rich RNAs, associated with a range of neurodegenerative diseases. Upon binding RNA, TDP-43 undergoes a large conformational change with two RNA recognition motifs (RRMs) connected by a long linker rearranged, strengthening the binding affinity of TDP-43 with RNA. We extend the equally weighted multiscale elastic network model (ewmENM), including its Gaussian network model (ewmGNM) and Anisotropic network model (ewmANM), with the multiscale effect of interactions considered, to the characterization of the dynamics of binding interactions of TDP-43 and RNA. The results reveal upon RNA binding a loss of flexibility occurs to TDP-43's loop3 segments rich in positively charged residues and C-terminal of high flexibility, suggesting their anchoring RNA, induced fit and conformational adjustment roles in recognizing RNA. Additionally, based on movement coupling analyses, it is found that RNA binding strengthens the interactions among intra-RRM  $\beta$ -sheets and between RRMs partially through the linker's mediating role, which stabilizes RNA binding interface, facilitating RNA binding efficiency. In addition, utilizing our proposed thermodynamic cycle method combined with ewmGNM, we identify the key residues for RNA binding whose perturbations induce a large change in binding free energy. We identify not only the residues important for specific binding, but also the ones critical for the conformational rearrangement between RRMs. Furthermore, molecular dynamics simulations are also performed to validate and further interpret the ENM-based results. The study demonstrates a useful avenue to utilize ewmENM to investigate the protein-RNA interaction dynamics characteristics.

#### KEY WORDS

binding dynamics, ewmENM, key residues, TDP-43-RNA interactions

## 1 | INTRODUCTION

Transactive response DNA-binding protein 43 (TDP-43), an alternative-splicing regulator, plays multiple roles in pre-mRNA maturation, microRNA processing, and mRNA turnover, stabilization and transport.<sup>1</sup> TDP-43 can specifically recognize and bind long UG-rich

RNA sequences,<sup>2</sup> which is associated with a variety of neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD).<sup>3</sup>

TDP-43 was first found as a DNA-binding protein,<sup>4</sup> and later also as an RNA-binding protein and now was known to prefer to bind RNA.<sup>5</sup> TDP-43 can specifically recognize and bind UG-rich RNAs or

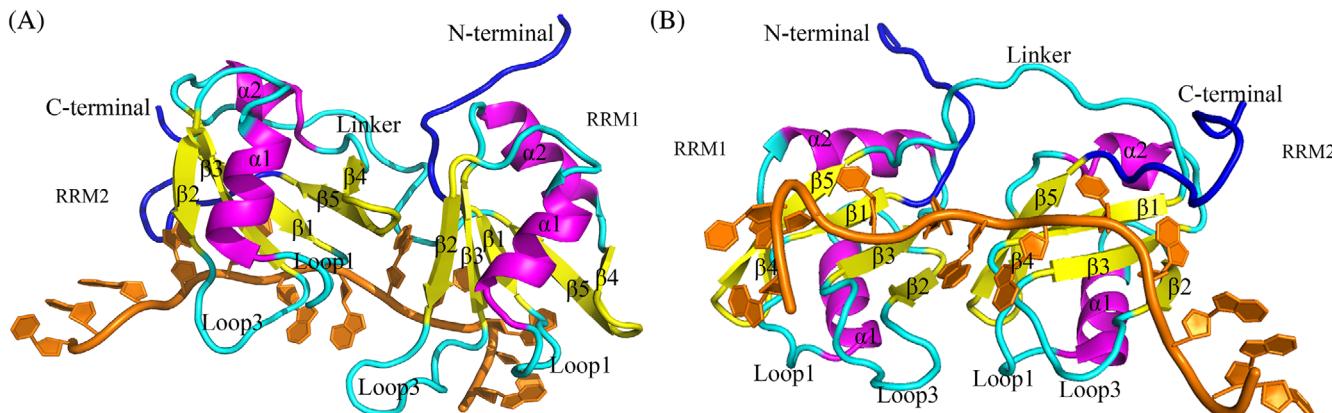
TG-rich DNAs by the RNA-binding domain (RBD).<sup>6</sup> Researchers have found that TDP-43 undergoes a large conformational change upon binding to its target RNA, and some perturbations to the structure or dynamics of TDP-43 would be propagated through the binding interface, leading to significant changes in RNA-binding affinity and specificity.<sup>7–9</sup> Therefore, exploring the dynamics of binding interactions of TDP-43 and RNA is helpful for understanding the mechanism of specific recognition and interactions between them and for developing therapeutic strategies targeting TDP-43 for neurodegenerative diseases.

The structure of TDP-43 in complex with RNA has been solved by nuclear magnetic resonance (NMR) spectroscopy,<sup>9</sup> as shown in Figure 1 for its representative structure with the average lowest root mean square deviation (RMSD) from the other models. TDP-43 consists of N-terminal, C-terminal and an RBD composed of two tandem RNA recognition motifs (RRM1 and RRM2) connected by a 14-amino acid linker. The linker confers TDP-43's adaptability to nucleic acids by allowing different orientations of the two RRMs.<sup>10,11</sup> Unlike the typical RRM containing four-stranded antiparallel  $\beta$ -sheets, such as hnRNP-A1<sup>12</sup> and Hud,<sup>13</sup> each RRM of TDP-43 has a  $\beta_1$ - $\alpha_1$ - $\beta_2$ - $\beta_3$ - $\alpha_2$ - $\beta_4$ - $\beta_5$  order of secondary structure elements with an additional  $\beta$ -sheet ( $\beta_4$ ), extending the  $\beta$ -sheet surface accessible for RNA binding. The intra-RRM  $\beta$ -sheets and inter-RRM interface  $\beta_2$  and  $\beta_4$  form hydrophobic cores, respectively.<sup>9,14</sup> TDP-43 adopts a positively charged groove composed of  $\beta$ -sheets, loop1s, loop3s, linker, and C-terminal (the positive electrostatic potential mainly provided by loop1 and loop3 in each RRM) to bind UG-rich repeats. RRM loop1 and loop3 protrude to RNA forming electrostatic and hydrogen-bonding interactions with RNA,  $\beta$ -sheets form the main RNA binding surface where the two highly conserved segments denoted as ribonucleoprotein domains (RNPs) are located (with RNP1 mainly situated at  $\beta_3$  and RNP2 at  $\beta_1$ ), and C-terminal is necessary for RNA specific recognition. For RRMs, they can move relatively independently in the RNA-free state, and upon RNA binding they adopt a fixed and suitable orientation to each other, of which RRM1 plays an important role in

binding with RNA compared with RRM2 partially due to the longer loop3 in RRM1.<sup>8</sup> The conformational rearrangement between RRMs strengthens the binding affinity of TDP-43 with RNA.<sup>15</sup>

Most of the experimental researches focus on the key residues involved in the complex formation of TDP-43 and RNA. Using electro-mobility shift assay (EMSA), Baralle et al. revealed that RRM1 plays a more important role in DNA/RNA binding compared with RRM2 that however is needed for the correct complex formation as a supporting role.<sup>16</sup> Later, they found that besides  $\beta$ -sheets, loop1s, and loop3s connecting  $\beta$ -sheets and helices are also crucial for nucleic acid recognition by providing specificity via cross-linking and immunoprecipitation assay.<sup>17</sup> Utilizing EMSA, Chen et al. observed that mutating some linker residues to glutamic acid creates the strong repulsive electrostatic interactions between TDP-43 and nucleic acid, completely abolishing the RNA binding capacity of TDP-43.<sup>18</sup> Theoretically, to our knowledge there are only several studies on TDP-43 itself. Through molecular dynamics (MD) simulations, Chiang et al. revealed that negatively charged Asp169Gly mutation located at  $\beta$ -turn connecting  $\beta_4$  and  $\beta_5$  in RRM1 induces a local conformational change, increasing the hydrophobic interactions in RRM1 core.<sup>19</sup> Still with MD method, Prakash et al. studied the unfolding process of TDP-43, observing that RRM1 has a different unfolding pathway from RRM2.<sup>14</sup>

MD simulation is a time-consuming method especially for the large-size systems. In order to address the issue, the coarse-grained elastic network models (ENMs) have been proposed for investigating the function-related motions of biomacromolecules.<sup>20</sup> Gaussian network model (GNM)<sup>21,22</sup> and anisotropic network model (ANM)<sup>23</sup> are the two mostly used ENM methods. Generally, low frequency motion modes calculated from ENM represent the large-scale collective motions usually relevant to protein functions.<sup>24</sup> Originally, Tirion et al. developed the conventional ENM method where a protein structure is modeled as a coarse-grained elastic network in which all the pairwise residues within a given cutoff distance are considered to interact with each other and are connected by harmonic springs with



**FIGURE 1** Reference structure of TDP-43-RNA complex determined by NMR spectroscopy (PDB ID: 4BS2). For clarity, the top-view and bottom-view of the structure are shown, with the linker (cyan) connecting RRM1 and RRM2, and five  $\beta$ -sheets (yellow), two  $\alpha$ -helices (magenta) and five loops (cyan) constituting each RRM structure, two ends being C-terminal and N-terminal (blue) and RNA (orange) lying on the  $\beta$ -sheet surface

a uniform force constant.<sup>20</sup> Later, considering the long-range effect of inter-residue interactions, Yang et al. proposed a parameter-free ENM (pfENM) where all the residues are considered to interact with each other with the force constant of the spring connecting two residues inversely proportional to the second power of inter-residue distance.<sup>25</sup> Hinsen adopted an exponential decay function dependent of the square of inter-residue distance to describe the force constant.<sup>26</sup> As we know, in biomolecules different kinds of interactions, such as covalent, Van der Waals, hydrophobic, and electrostatic interactions have their individual characteristic action ranges instead of a single one. Thus, with the multiscale feature of interactions considered, the multiscale ENM (mENM) was introduced by Wei et al. based on their previously proposed flexibility rigidity index (FRI) method,<sup>27–30</sup> where the force constant is represented as the sum of multiple weighted kernel functions.<sup>31</sup> The mENM is able to capture the multiscale behavior and gives rise to a significant improvement in B-factor prediction over the conventional ENM methods. Our extension of the mGNM to the dynamics analyses for a dataset containing 77 nonredundant RNAs achieves a remarkable improvement in residue flexibility prediction compared with the conventional GNM and pfGNM, especially for those loosely packed RNAs.<sup>32</sup> However, in mENM the weight optimization by the least squares method only considers the diagonal elements of Kirchhoff matrix or Hessian matrix, which means that the details of pairwise interactions are neglected in this process. When adopting equal weights in mENM (denoted as ewmENM), we achieve a better balance of performance in residue flexibility calculation and fluctuation cross-correlation calculation.<sup>33</sup> Additionally, based on the conventional GNM, our group developed a thermodynamic cycle method to identify the key residues whose mutations cause a large change in binding free energy.<sup>34</sup> The method is able to identify the key residues associated not only with the ligand binding, but also with the allosteric transition coupled with the binding.<sup>35,36</sup>

In this work, considering the advantage of ewmENM in reproducing function-related protein dynamics, we extend ewmENM, as an application, to interpret the effect of RNA binding on the dynamics of TDP-43. In addition, combined with ewmGNM, the thermodynamic cycle method is utilized to identify the key residue in TDP-43, which are functionally indispensable for RNA specific recognition and binding.

## 2 | METHODS AND MATERIALS

### 2.1 | Equally weighted multiscale ENM

Different from the conventional cutoff-based ENM<sup>20</sup> and the parameter-free ENM (pfENM) which adopts a distance-dependent spring constant set,<sup>25</sup> the equally weighted multiscale ENM (ewmENM) utilizes multiple kernel functions to represent inter-residue interactions, considering the multiscale effect of the interactions.<sup>33</sup>

### 2.1.1 | Equally weighted multiscale Gaussian Network Model (ewmGNM)

In ewmGNM<sup>33</sup> originated from mGNM,<sup>31</sup> a biomacromolecule is modeled as a coarse-grained elastic network where residues are replaced by several nodes (here  $C_\alpha$  atom for protein residue and P atom for RNA nucleotide) and the force constant of springs connecting two nodes is represented as the sum of multiple kernel functions with equal weights. Thus, in a nutshell, the B-factor prediction of the  $i$ th node of a biomolecule consisting of  $N$  nodes can be expressed as<sup>22,37</sup>

$$B_i \propto (\Gamma^{-1})_{ii}, \quad (1)$$

where  $(\Gamma^{-1})_{ii}$  is the  $i$ th diagonal element of the matrix inverse of the Kirchhoff matrix, which can be written as

$$\Gamma_{ij}(\Phi) = \begin{cases} -\Phi(R_{ij}^0; \eta), & \text{if } i \neq j \\ -\sum_{jj \neq i}^N \Gamma_{jj}(\Phi), & \text{if } i=j \end{cases}, \quad (2)$$

where the kernel function  $\Phi$  representing inter-residue interactions adopts an exponential form in the following with the interactions exponentially decreasing with increasing distances between nodes.

$$\Phi(R_{ij}^0; \eta, \kappa) = e^{-(R_{ij}^0/\eta)^\kappa}, \quad \kappa \geq 0 \quad (3)$$

where  $R_{ij}^0$  is the equilibrium distance between the  $i$ th and  $j$ th nodes, and the parameters  $\eta$  and  $\kappa$  control the decay extent. Thus, the corresponding Kirchhoff matrix of the  $n$ th kernel function can be described as

$$[\Gamma_n]_{ij} = \begin{cases} -\Phi_n(R_{ij}^0; \eta_n, \kappa_n), & \text{if } i \neq j \\ -\sum_{jj \neq i}^N [\Gamma_n]_{jj}, & \text{if } i=j \end{cases}. \quad (4)$$

Here, two kernel functions with equal weights are adopted. Thus, the total Kirchhoff matrix in Equation (2) can be written as

$$\Gamma = \sum_{n=1}^2 \Gamma_n. \quad (5)$$

The mean square fluctuation (MSF) of the  $i$ th node and the fluctuation cross-correlation between the  $i$ th and  $j$ th nodes are calculated, respectively by

$$\langle (\Delta R_i \cdot \Delta R_i) \rangle \propto [\Gamma^{-1}]_{ii} \quad (6)$$

$$\langle (\Delta R_i \cdot \Delta R_j) \rangle \propto [\Gamma^{-1}]_{ij}. \quad (7)$$

The pseudoinverse of Kirchhoff matrix can be decomposed as

$$\Gamma^{-1} = U \Lambda^{-1} U^\top = \sum_{k=2}^N \frac{\mathbf{u}_k \cdot \mathbf{u}_k^\top}{\lambda_k}, \quad (8)$$

where  $\mathbf{U}$  is an orthogonal matrix whose columns  $\mathbf{u}_k$  are the eigenvectors of  $\mathbf{\Gamma}$ , and  $\mathbf{\Lambda}$  is a diagonal matrix with the diagonal  $\lambda_i$  being the eigenvalues of  $\mathbf{\Gamma}$ .

### 2.1.2 | Equally weighted multiscale Anisotropic Network Model (ewmANM)

In contrast to isotropic GNM, ANM can provide not only the amplitudes of residue fluctuations, but also their directions. In ewmANM<sup>33</sup> originated from mANM,<sup>31</sup> the dynamical properties of a molecule are determined by a Hessian matrix  $\mathbf{H}$  whose element is a submatrix with size of  $3 \times 3$ . The Hessian matrix associated with the  $n$ th kernel function can be written as

$$[\mathbf{H}_n]_{ij} = \begin{cases} [\mathbf{h}_n]_{ij}, & \text{if } i \neq j \\ -\sum_{jj \neq i} [\mathbf{h}_n]_{ij}, & \text{if } i = j \end{cases} \quad (9)$$

where the local  $3 \times 3$  Hessian matrix can be described as

$$[\mathbf{h}_n]_{ij} = -\frac{\Phi_n(R_{ij}^0, \eta_n)}{(R_{ij}^0)^2} \begin{bmatrix} (x_i - x_j)(x_i - x_j) & (x_i - x_j)(y_i - y_j) & (x_i - x_j)(z_i - z_j) \\ (y_i - y_j)(x_i - x_j) & (y_i - y_j)(y_i - y_j) & (y_i - y_j)(z_i - z_j) \\ (z_i - z_j)(x_i - x_j) & (z_i - z_j)(y_i - y_j) & (z_i - z_j)(z_i - z_j) \end{bmatrix}. \quad (10)$$

The MSF of the  $i$ th node and the cross-correlation between the  $i$ th and  $j$ th nodes can be calculated, respectively, by

$$\langle (\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j) \rangle \propto (H_{3i-2,3i-2}^{-1} + H_{3i-1,3i-1}^{-1} + H_{3i,3i}^{-1}) \quad (11)$$

$$\langle (\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j) \rangle \propto (H_{3i-2,3j-2}^{-1} + H_{3i-1,3j-1}^{-1} + H_{3i,3j}^{-1}). \quad (12)$$

The pseudoinverse of Hessian matrix can be decomposed as

$$\mathbf{H}^{-1} = \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^T = \sum_{k=1}^{3N} \frac{\mathbf{u}_k \cdot \mathbf{u}_k^T}{\lambda_k}, \quad (13)$$

where  $\mathbf{U}$  is an orthogonal matrix whose columns  $\mathbf{u}_k$  are the eigenvectors of  $\mathbf{H}$ , and  $\mathbf{\Lambda}$  is a diagonal matrix with the diagonal  $\lambda_i$  being the eigenvalues of  $\mathbf{H}$ . The cross-correlation is normalized as

$$C_{ij} = \frac{\langle (\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j) \rangle}{\sqrt{\langle (\Delta \mathbf{R}_i)^2 \rangle \cdot \langle (\Delta \mathbf{R}_j)^2 \rangle}}. \quad (14)$$

The values of  $C_{ij}$  range from  $-1$  to  $1$ . The positive values indicate that the residues move in the same direction, and the negative ones indicate that they move in the opposite direction. The higher the absolute value, the stronger the correlation between the two nodes. The zero value indicates no correlations between residue motions.

### 2.1.3 | Optimization of parameters in ewmENM

It is important to obtain a set of suitable parameters in ewmENM, which in this work is determined through maximizing the Pearson correlation coefficient (PCC) between experimental residue MSFs and theoretical residue MSFs calculated from ewmENM by systematically searching the parameter set  $\{\eta_1, \kappa_1; \eta_2, \kappa_2\}$  in the range of  $\{1-30, 1-12; 1-20, 1-12\}$  with step size  $1$ .<sup>32,33</sup> The optimization process is shown in Figure S1. The PCC is defined as

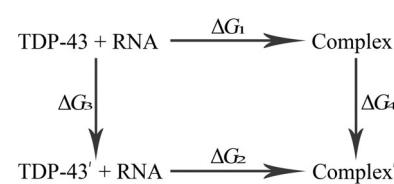
$$PCC = \frac{\sum_{i=1}^N (\text{MSF}_i^T - \overline{\text{MSF}^T})(\text{MSF}_i^E - \overline{\text{MSF}^E})}{\sqrt{\sum_{i=1}^N (\text{MSF}_i^T - \overline{\text{MSF}^T})^2} \sqrt{\sum_{i=1}^N (\text{MSF}_i^E - \overline{\text{MSF}^E})^2}}, \quad (15)$$

where  $\text{MSF}_i^T$  and  $\text{MSF}_i^E$  are the MSFs of the  $i$ th node calculated by ewmENM (Equation (6) or (11)) and from the NMR structure ensemble, respectively, and  $\overline{\text{MSF}^T}$  and  $\overline{\text{MSF}^E}$  are the corresponding mean values over all nodes. The PCC value ranges from  $-1$  to  $1$ , and a larger PCC value means the theoretically predicted residue flexibility by ewmENM is closer to the experimental values obtained from the NMR structure ensemble.

### 2.2 | ewmGNM-based thermodynamic cycle method

The thermodynamic cycle method based on the conventional GNM was proposed by us to identify the key residues involved in protein-ligand interactions and the coupled allosteric transitions.<sup>35,36</sup> Here, considering the advantage of ewmENM in reproducing function-related protein dynamics, we develop the ewmGNM-based thermodynamic cycle method to identify the key residues whose mutations result in a significant change in binding free energy of protein with ligand.

As illustrated in Figure 2,  $\Delta G_1$  and  $\Delta G_2$  are binding free energies of wild-type and mutant proteins with RNA, respectively. Because of the complexity of protein-RNA interactions, it is very difficult to calculate directly  $\Delta G_1$  and  $\Delta G_2$ . Therefore, a thermodynamic cycle is introduced to calculate the change in binding free energy. The two unphysical processes are supposed,  $\text{TDP-43} \rightarrow \text{TDP-43}'$  and



**FIGURE 2** Thermodynamic cycle diagram.  $\Delta G_1$  and  $\Delta G_2$  are the binding free energies of wild-type and mutant proteins TDP-43' binding with RNA, respectively, and  $\Delta G_3$  and  $\Delta G_4$  are the binding free energy changes caused by a residue perturbation to the protein TDP-43 and complex, respectively

complex → complex'. When a residual perturbation is introduced to the protein and complex, the free energy changes involved in the two processes are represented as  $\Delta G_3$  and  $\Delta G_4$ , respectively. In physics, free energy is a state variable that only depends on the initial and final states. Therefore, the change in binding free energy caused by a residue mutation is

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3. \quad (16)$$

Here, the residue mutation is introduced through reducing the force constants of all the springs connecting to the mutated residue. For a certain pair of the  $i$ th and  $j$ th residues, which form a contact both in the receptor and complex structures, the perturbation of the force constant  $\gamma_{ij}$ , we assume, will lead to the approximately same change of the potential energy (i.e.,  $\Delta U_3 \approx \Delta U_4$ ), so that the  $\Delta\Delta G$  is only related with the entropy change. According to Equation (16), the change of binding energy  $\Delta\Delta G$  is calculated as

$$\Delta\Delta G = -T(\Delta S_4 - \Delta S_3). \quad (17)$$

According to statistical physics, the vibration entropy of the system can be written as

$$\begin{aligned} S &= \frac{\langle H \rangle - F}{T} = \frac{1}{T} \left[ \frac{3}{2}(N-1)k_B T + k_B T \ln Z \right], \\ &= \frac{3}{2}(N-1)k_B + k_B \ln Z \end{aligned} \quad (18)$$

where  $\langle H \rangle$  is the average vibrational Hamiltonian,  $F = -k_B T \ln Z$  is the vibrational Helmholtz free energy,  $Z$  is the configurational integral part of the vibrational partition function given by  $Z = \int \exp(-H/k_B T) d\{\Delta R\}$ , and  $N$  is the number of nodes. Here, it should be pointed out that the symbol "Z" in Equation (18) is the entire partition function and only the vibrational part is used in this work. When a perturbation is introduced to the force constant of the spring connecting the  $i$ th and  $j$ th residues, the change of entropy can be written as

$$\Delta S = -\frac{\partial S}{\partial \gamma_{ij}} \Delta \gamma_{ij}, \quad (19)$$

where the negative sign represents the decrease of the force constant, and  $\Delta \gamma_{ij}$  is the change of the force constant when a perturbation is introduced. According to Equation (18), the derivation of the entropy  $S$  with respect to  $\gamma_{ij}$  is written by<sup>34,35</sup>

$$\frac{\partial S}{\partial \gamma_{ij}} = -\frac{1}{2T} \left( \langle (\Delta R_i)^2 \rangle + \langle (\Delta R_j)^2 \rangle - 2 \langle \Delta R_i \cdot \Delta R_j \rangle \right). \quad (20)$$

When introducing a mutation to a residue, all the springs connecting to this residue should be perturbed simultaneously. In this way, the  $\Delta\Delta G$  value caused by a residue mutation is the sum of the binding free energy changes for the perturbations of all the springs involved

in this residue. In the process of identifying key residue, every amino acid residue is mutated in the same way described above.

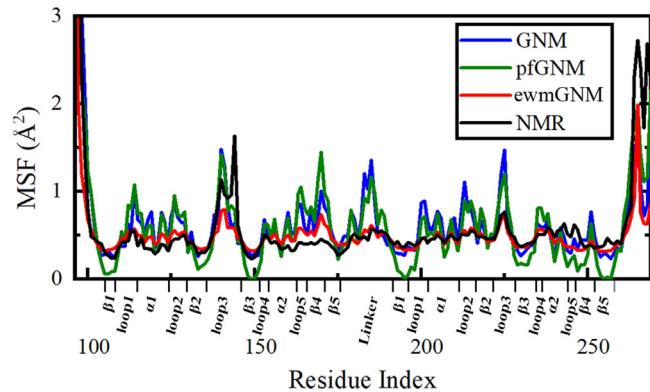
## 2.3 | Receptor and complex systems

The complex of human TDP-43 and RNA, determined by NMR spectroscopy, was obtained from the Protein Data Bank (PDB)<sup>38</sup> with PDB ID 4BS2<sup>9</sup> which has 20 models and each model has 174 protein residues and 12 RNA nucleotides. The representative structure with the average lowest RMSD from the others is used to construct the ewmENM model of the complex, and that of the protein receptor is built based on the structure with RNA eliminated from the complex.

## 3 | RESULTS AND DISCUSSIONS

### 3.1 | Comparison of experimental MSFs with theoretical MSFs from ewmGNM

The accurate prediction of residue MSFs provides an effective starting point to understand the dynamics of biomolecules. For the complex structure of TDP-43 with target RNA, the optimal ewmGNM model was constructed via maximizing the PCC between experimental and theoretical residue MSFs under a set of optimized parameters (see Section 2), with the results shown in Figure 3. In order to evaluate the performance of ewmGNM, Figure 3 also shows the corresponding optimal results from the conventional GNM and pfGNM. For the conventional GNM, three cutoff parameters for protein, RNA and interface regions were systematically searched in the range of [1–25 Å] with step size 1 Å; for pfGNM, no parameters need to be optimized. For clarity, only the results corresponding to the protein are shown in Figure 3. From Figure 3, all the three models obtain a high PCC value. Compared with the PCC (.81) from the conventional GNM (with



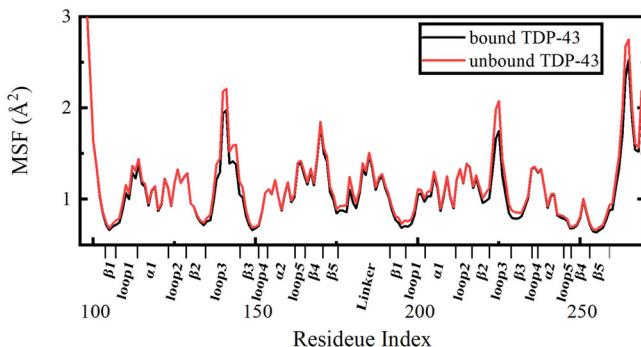
**FIGURE 3** Residue mean square fluctuations (MSFs) of TDP-43 in RNA-bound state calculated by the conventional GNM, pfGNM, and ewmGNM under the optimal set of parameters. For comparison, the MSFs from NMR structure ensemble are also shown

cutoff values 10, 14, and 12 Å for protein, RNA, and interface regions, respectively) and that (.80) from pfGNM, ewmGNM ( $\eta_1 = 20$ ,  $\kappa_1 = 7$ ;  $\eta_2 = 13$ ,  $\kappa_2 = 10$ , see Figure S2 for kernel functions) achieves a highest PCC value (.86). The better performance of ewmGNM in flexibility reproducing than the conventional GNM and pfGNM is consistent with our previous results on RNAs and proteins.<sup>32,33</sup>

From Figure 3, it is worth mentioning that for the loosely packed segments such as the loops and linker, their MSFs from the conventional GNM and pfGNM are much higher than the experimental data. For GNM, the main reason lies in the uniform cutoff distance, which is too small and insufficient to restrict the dynamics of the loosely packed segments, resulting in their over-high flexibility. For pfGNM, similarly, the spring constant represented by the inversely second power of inter-node distance is too small for characterizing the interactions within the loosely packed regions, whereas it is too large for the tightly folded regions, such as  $\beta$ -sheet segments (Figure 3), resulting in their over-low flexibility. However, the predicted flexibility from ewmGNM is in good agreement with the experimental data regardless of the tightly and loosely packed regions, which mainly attributes to its consideration of the multiscale effect of inter-residue interactions through using the sum of multiple kernel functions to represent the force constant, making up for the deficiencies of the conventional GNM and pfGNM.

### 3.2 | Effects of RNA binding on the dynamics of TDP-43

The RNA binding to TDP-43 stabilizes the relative positions between RRM1 and RRM2 in TDP-43, facilitating its alternative splicing function exertion.<sup>9</sup> In order to explore the effects of RNA binding on the protein dynamics, we computed (according to Equation (6)) and compared the residue MSFs of TDP-43 in RNA-bound and RNA-free states, with the results shown in Figure 4. From Figure 4, RNA binding induces a certain loss of flexibility in some segments including loop1 and loop3 in each RRM, C-terminal, as well as linker residues adjacent to  $\beta$ 5 in RRM1, and  $\beta$ 1-3 in RRM2.



**FIGURE 4** Residue mean square fluctuations of TDP-43 in RNA-bound and RNA-free states calculated by ewmGNM

Compared with loop1s, loop3s have an evident decrease in flexibility upon RNA binding. Both loop1s and loop3s in RRMs protrude to RNA, constituting a part of the highly positively charged RNA binding groove lined with Trp113 (loop1), Lys145 (loop3), Arg197 (loop1), and Arg227 (loop3)<sup>9</sup> which mainly form electrostatic and stacking interactions (such as Trp113-G1, Lys145-G3, Lys145-A6, Arg197-A7, and Arg227-A7) with RNA, contributing the majority of electrostatic attraction to TDP-43 binding with RNA.<sup>7,8</sup> Loop3s' and loop1s' flexibility losses upon RNA binding, as well as the high enrichments of positively charged residues in loop3s (4/10 for RRM1 and 2/7 for RRM2) and loop1s (1/6 for both RRMs) indicate that they are very critical in attracting RNA, suggesting their important roles in anchoring RNA into TDP-43 binding groove and in the induced fit in TDP-43-RNA recognition. Additionally the larger flexibility loss in loop3s than in loop1s, as well as the more residues and higher enrichment of positively charged residues in the former than in the latter indicate that loop3s provide more interactions with RNA than loop1s, consistent with the experimental observation.<sup>8</sup> Furthermore from Figure 4, RRM1 loop3's more extensive flexibility loss than RRM2 loop3's upon RNA binding suggests that the former contributes more to the interactions of TDP-43 with RNA than the latter, which can partially explain the experimental finding that RRM1 contributes a major role in the binding affinity of TDP-43 with RNA compared with RRM2.<sup>15,16</sup>

For C-terminal, some residues in it have a flexibility loss upon RNA binding to some extent, consistent with the NMR experimental observation that the residues in C-terminal have a large chemical-shift perturbation upon RNA binding.<sup>9</sup> C-terminal constitutes a part of the binding interface, forming specific hydrogen-bonding interactions with RNA, such as Asn259-U8, Glu261-U8, and Lys263-G9,<sup>18</sup> indicating that C-terminal is essential for the specific recognition between TDP-43 and RNA.

For the linker residues (connecting RRM1 and RRM2) adjacent to  $\beta$ 5 in RRM1, they have a certain decrease in flexibility upon RNA binding. Researchers found that Lys176 (in  $\beta$ 5), Leu177 (linker), and Asn179 (linker) form hydrogen bonds with U4 to read out U4's entire Watson Crick face.<sup>9</sup> Additionally, gel filtration experiment observed that upon RNA binding, linker forces RRMs to a suitable orientation and thereby exerts its ability to bind RNA using a specific sequence.<sup>11</sup> From above, the linker plays indispensable roles in the conformational rearrangement of RRM1 and RRM2 and in the specific interactions of TDP-43 with RNA.

Finally,  $\beta$ 1-3 segments in RRM2 have a slight flexibility loss upon RNA binding. They constitute part of the main binding surface and form hydrogen-bonding and hydrophobic interactions with RNA, such as Phe194-U8, Phe221-G9, and Phe229-U8.<sup>9</sup> Additionally, the segments RNP1 and RNP2 mainly located in  $\beta$ 3 and  $\beta$ 1 and rich in conserved aromatic/hydrophobic residues form stacking interactions with RNA, such as Phe194-U8 and Phe231-G9, contributing the specific recognition of TDP-43 with RNA.<sup>14</sup>

It should be pointed out that here we assume the protein structure does not change too much upon RNA binding, which is not the case in the real situation. We want to know how the results are after

the conformational changes are considered? To this aim, we performed an 80 ns MD simulation to model the free state structure by allowing the protein to relax after removing the RNA (see the MD simulation procedure in Supporting Information). Figure S4 shows the change of the RMSD of the backbone atoms of the protein as a function of time. The system reaches an equilibrium state after 20 ns. The last equilibrated snapshot was used to build the ewmENM of the free protein, from which the residue MSFs were obtained, as shown in Figure S5 with the results from the bound TDP-43 also shown for comparison. From Figure S5, RNA binding induces a similar or a little more evident loss of flexibility in the segments including loop1 and loop3 in each RRM, C-terminal, as well as linker residues, compared with the result from the pseudo unbound protein without relaxation (Figure 4), which is understandable as for the former, the conformational changes are considered in the protein by relaxing it after removing RNA to some extent.

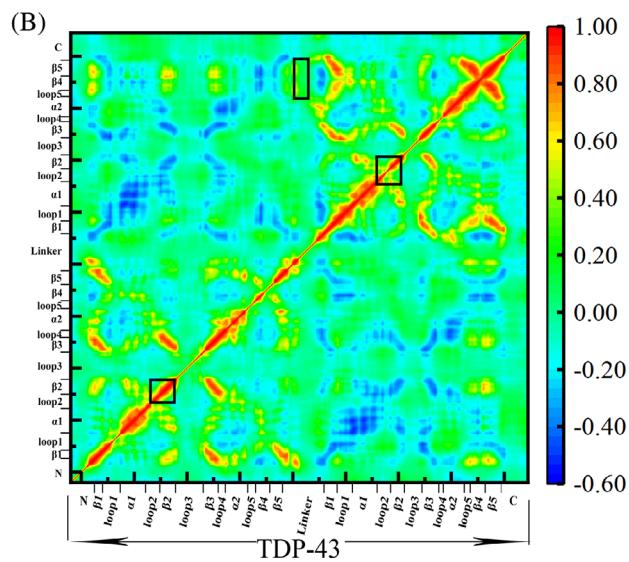
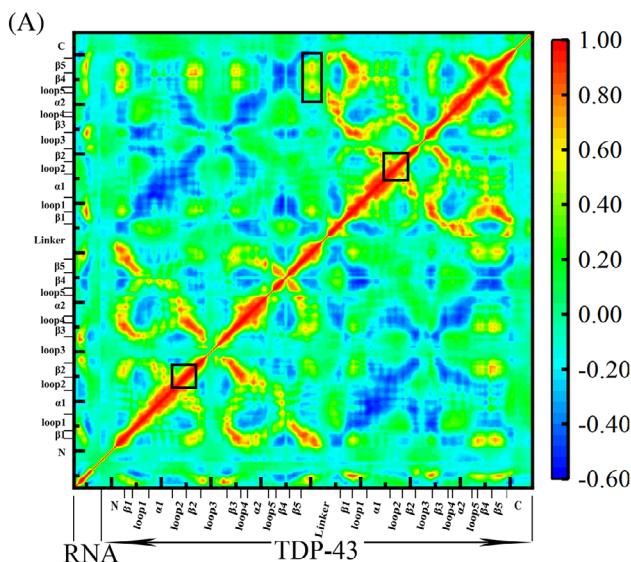
Additionally, it can be seen that the differences of the segments mentioned above in flexibility between the two states are not very big. To detect whether the differences are statistically significant, we built the ewmENM models for all the 20 available structures in the NMR ensemble and calculated the residue MSFs of TDP-43 in RNA-bound and RNA-free states to statistically analyze the effect of RNA binding on the protein dynamics. Table S1 shows the average MSFs of several residues of RRM1 loop3 (residues Lys140, Thr141, and Ser144), linker (residues Asn179 and Glu186), RRM2 loop3 (residue Pro225), and C terminal (residue His264) that have a certain flexibility change when RNA binding in the two states and their corresponding *p*-values. From the results, the differences of the segments or residues in flexibility between the two states are statistically significant.

In summary, RNA binding to TDP-43 makes some segments including loop1, loop3, C-terminal, linker, and  $\beta$ 1-3 in RRM2 lose their flexibility to some extent due to their formation of electrostatic,

hydrogen-bonding and hydrophobic interactions with RNA. Loop1s and loop3s play important roles in the electrostatic attraction and induced fit in TDP-43-RNA interactions, linker and C-terminal are important for the conformational rearrangement and adjustment and specific interactions, and  $\beta$ 1-3 in RRM2 constituting part of the RNA binding surface are necessary to the specific recognition of TDP-43 with RNA.

### 3.3 | Movement coupling between residues in complex and in receptor TDP-43

In order to detect the functional movement couplings in the complex of TDP-43 with RNA, we calculated residue fluctuation cross-correlations according to Equation (14) via ewmANM, with the result shown in Figure 5A. The optimized ewmANM model has a PCC of .85 ( $\eta_1 = 22$ ,  $\kappa_1 = 9$ ;  $\eta_2 = 11$ ,  $\kappa_2 = 8$ , see Figure S3 for kernel functions). From Figure 5A, for protein, although not very evident, the two RRMs can be identified with more positive correlations within each RRM and more negative ones between RRMs. Generally, the positive correlations occur among  $\beta$ 1- $\beta$ 3 and between  $\beta$ 4 and  $\beta$ 5 in each RRM, consistent with their spatial relationship with the former and the latter stacking together with each other, respectively, which facilitates the intra-RRM stability. For the interface region between TDP-43 and RNA, strong couplings mainly occur between some segments in RRM1 (RRM2) and the front (back) part of RNA, consistent with their binding positions. The interface residues in loop1s, loop3s, and linker adjacent to  $\beta$ 5 are strongly coupled with their individual interacting nucleotides in RNA, which is attributed to the formed strong interactions between them.<sup>7</sup> Additionally, in each RRM the residues in  $\beta$ 1 and  $\beta$ 3 constituting parts of the conserved RNP2 and RNP1, respectively,<sup>8</sup> and residues in  $\beta$ 4 and  $\beta$ 5 in RRM2 are also strongly correlated with RNA, respectively, which is due to their participating in



**FIGURE 5** Residue fluctuation cross-correlations of TDP-43 in RNA-bound state (A) and RNA-free state (B) calculated by ewmANM

the formation of binding interface.<sup>7,14</sup> The above interactions involved between TDP-43 and RNA have been mentioned in Section 3.2. Effects of RNA binding on the dynamics of TDP-43. Generally, the observation that these residues are strongly coupled with RNA is well consistent with the result that they have a certain decrease in fluctuation upon RNA binding.

In order to detect the effect of RNA binding on the motional coupling within the protein, Figure 5B shows the fluctuation cross-correlation of TDP-43 in RNA-free state. Compared with TDP-43 in RNA-free state (Figure 5B), the protein in RNA-bound state (Figure 5A) has a stronger and more extensive motional coupling regardless of the positive and the negative ones. Evidently, generally the enhanced and extended positive correlations occur among  $\beta$ -sheets in each RRM, implying that RNA mediates the interactions among  $\beta$ -sheets (consistent with the binding mode where RNA lies on the extended surface constituted by  $\beta$ -sheets (Figure 1B), which facilitates the formation of intra-RRM hydrophobic core<sup>14</sup> and stabilizes the RRM structures. Additionally, the enhanced and extended couplings between  $\beta$ 1-3 in RRM1 and  $\beta$ 4-5 in RRM2 just located at RRM1-RRM2 interface are also observed, suggesting RNA binding strengthens the inter-RRM interactions, which helps form the hydrophobic core composed by  $\beta$ 2 and  $\beta$ 4<sup>9</sup> and stabilize the RNA binding interface, critical for RNA binding efficiency. Still, the couplings between  $\beta$ 1 and  $\beta$ 3 in RRM1,  $\beta$ 4-5 in RRM2 and the linker are strengthened, respectively to some extent, hinting the mediating role of the linker in inter-RRM interactions, which also helps stabilize the RNA binding interface.

In summary, RNA binding to TDP-43 strengthens the intra-RRM interactions mainly among  $\beta$ -sheets and the inter-RRM interactions partially through the linker's mediating role, which facilitates the formation of intra-RRM and inter-RRM hydrophobic cores and the stabilization of RNA binding interface. Additionally, loop1, loop3, and

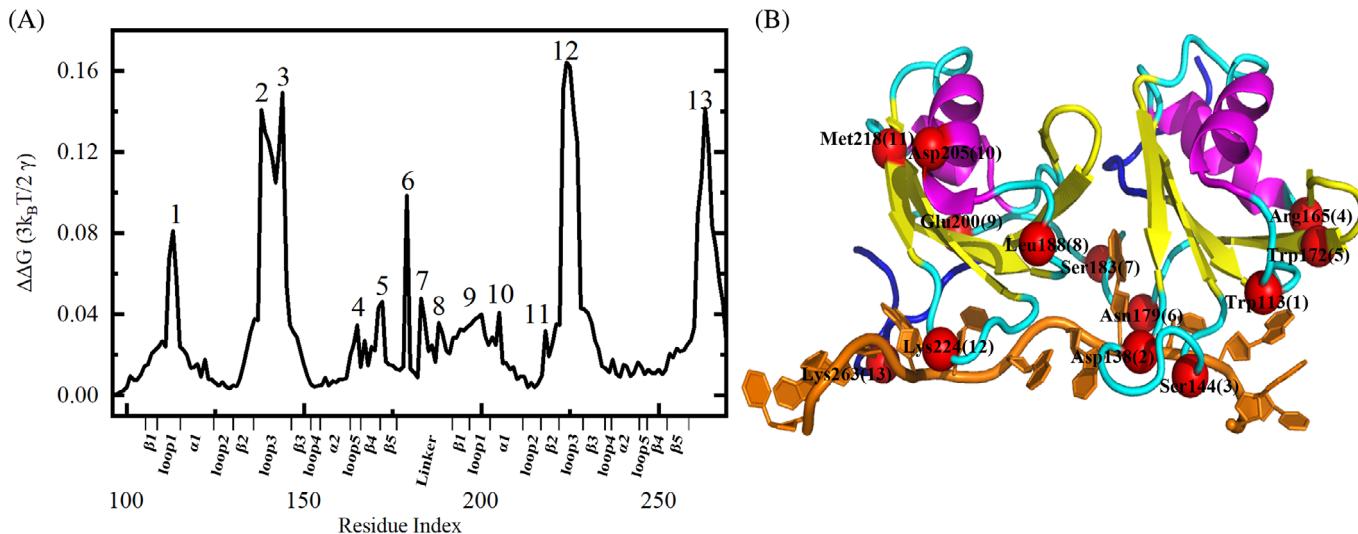
$\beta$ -sheet segments provide the main interactions with RNA, presenting strong dynamical correlations with RNA.

### 3.4 | Identification of functional key residues in TDP-43

The ewmGNM-based thermodynamic cycle method (see Section 2) was used to identify TDP-43 key residues associated with RNA binding and conformational rearrangement coupled with the binding. Each residue in TDP-43 was perturbed, and those residues whose mutations induce a larger change in the binding free energy  $\Delta\Delta G$  are taken as the key residues. The results are shown in Figure 6A. Here, the residue with the maximal  $\Delta\Delta G$  value in each cluster is defined as the central residue of the corresponding cluster. From Figure 6A, 13 residue clusters are identified whose central residues are Trp113, Asp138, Ser144, Arg165, Trp172, Asn179, Ser183, Leu188, Glu200, Asp205, Met218, Lys224, and Lys263, as shown in Figure 6B. Based on the central residues' positions in the protein structure, the identified key residues are compared with the available experimental and theoretical data and their functional information will be discussed in detail below.

#### 3.4.1 | Residues at loop regions

As shown in Figure 6, there are five residue clusters 1, 2, 3, 9, and 12 at loop regions. Clusters 1 and 9 are at loop1s in RRM1 and RRM2, respectively. For the former, Lukavsky et al. reported that Trp113 and Leu111 in cluster 1 form stacking and hydrogen-bonding interactions with G1, and hydrogen-bonding interactions with G4, respectively, whose substitutions with Ala abolish the capacity of TDP-43 to bind to RNA.<sup>9</sup> The experiment found that Pro112His mutation in cluster



**FIGURE 6** Key residue clusters identified by ewmGNM-based thermodynamics cycle method. (A)  $\Delta\Delta G$  values in response to residue mutations with key residue clusters with relatively high  $\Delta\Delta G$  values marked by the numbers 1–13. (B) Locations of the central residues for the 13 key residue clusters

1 moderately reduces the RNA-binding capacity of TDP-43 and the theoretical study by FoldX indicates the mutation destabilizes the protein.<sup>11</sup> As for cluster 9 at loop1 in RRM2, researchers show that the site-directed mutagenesis of Glu200 to Ala has a 3-fold decrease in RNA binding affinity.<sup>9</sup> As for loop3 segments, the site-directed mutagenesis of Leu139 to Ala in cluster 2 located in the binding cleft between TDP-43 and RNA leads to a 3-fold decrease in affinity of TDP-43 with RNA.<sup>9</sup> RRM1 Lys145 in cluster 3 forms electrostatic and hydrogen-bonding interactions with G3,<sup>11</sup> and its post-translational acetylation largely inhibits the RNA-binding activity of TDP-43. Still RRM1 Lys145 in cluster 3 and RRM2 Arg227 in cluster 12 bridging the two  $\beta$ 2- $\beta$ 3 loops, respectively both form strong electrostatic and stacking interactions with A6 and A7, respectively, thereby largely stabilizing the complex structure.<sup>9</sup>

Combining the results here and above, loop3s in TDP-43 have a flexibility loss upon RNA binding to some extent and their key residues' mutations cause a relatively large change in binding free energy, which indicates they play important roles in RNA recognition and interactions through induced fit and forming electrostatic and stacking interactions with RNA.

### 3.4.2 | Residues at $\beta$ -sheets

As shown in Figure 6, there are three residue clusters 4, 5, and 11 at  $\beta$ -sheets. For Arg171 and Asp174 in cluster 5 which specifically recognize G3 through hydrogen-bonding interactions, their double substitutions (Arg171Ala/Asp174Ala) disrupt the specific readout of G3, leading to a 20-fold reduction in RNA-binding affinity, and the single-point Arg171Ala mutant leads to a 6-fold reduction in affinity, indicating their important roles in RNA recognition and binding, and complex stability.<sup>8</sup> Additionally, the NMR study has identified Phe221 in cluster 11 as a key aromatic residue for the direct hydrophobic interaction with RNA G9, and for its accurately positioning the nucleotide on the  $\beta$ -sheet surface.<sup>9</sup>

In a whole, the results above suggest that the key residues identified in  $\beta$ -sheets play an important role in RNA specific recognition and binding affinity through forming hydrogen-bonding, hydrophobic, and stacking interactions with RNA.

### 3.4.3 | Residues at linker and C-terminal regions

As shown in Figure 6, there are three residue clusters 6, 7, and 8 at linker and one residue cluster 13 at C-terminal. Linker and C-terminal, with high flexibility, both make a significant contribution to RNA binding.<sup>9</sup> For linker, it is stabilized by RNA binding and becomes a part of the binding interface. The NMR experimental study shows that Leu177 and Asn179 in cluster 6 form hydrogen-bonding interactions with U4, and the mutation of Asn179Ala affects the binding affinity to some extent.<sup>8</sup> The mutation of Lys181Glu in cluster 6 reducing the electrostatic potential in nucleotide binding site disrupts RNA binding.<sup>18</sup> As for C-terminal, positively charged Lys263 in cluster 13 forms

strong electrostatic interaction with G9, and Glu261 in cluster 13 forms specific hydrogen bonds with U8 and G9.<sup>9</sup> The site-directed mutagenesis of Lys263 to Glu completely abolishes the RNA binding.<sup>18</sup> Through relaxation experiments, Lukavsky et al. found the residues in linker and C-terminal have a large chemical-shift perturbation upon RNA binding, indicating they are essential for RNA binding.<sup>9</sup>

In short, the residues in linker and C-terminal regions play an important role in binding recognition dynamics, forming a part of the binding interface through conformational adjustments, which further facilitates the specific binding and strengthens the affinity between TDP-43 and RNA through the formation of electrostatic and hydrogen-bonding interactions.

Finally, for the residue clusters 4, 8, and 10 (centered at Arg165, Leu188, and Asp205, respectively), we have not yet found the reports on their importance for RNA binding. However, cluster 8 is strategically located at the linker connecting RRs, and presumably plays an underlying role in orienting RRs for RNA binding. For cluster 4 located at  $\beta$ 4 that expands the  $\beta$ -sheet surface accessible for RNA binding as an additional  $\beta$ -sheet in TDP-43 RRs compared with the canonical RRM,<sup>39</sup> thus it possibly contributes much to the RNA-binding affinity. For cluster 10 at helix 1 in RRM2, MD simulations show that helix1 in RRM2 forms long-range contacts with  $\beta$ 4,<sup>14</sup> which possibly hints that the identified residues at helix1 would influence RNA binding through affecting  $\beta$ 4 dynamics.

The ewmGNM-based thermodynamic cycle method is efficient for identifying key residues while MD simulation can provide more details and serve as a validation of the predictions. Next, taking the identified key residue Lys263 in cluster 13 for example which is not at the binding interface and whose mutation to Glu is observed experimentally to abolish the RNA binding, we mutated Lys263 to Glu to analyze the effect of the mutation on the RNA binding through MD simulation. Besides on the mutant complex, we performed MD simulation on the wild-type one for comparison. The two independent MD simulations were carried out for 80 ns under an NPT ensemble, respectively. The MD simulation procedure is similar to that used for relaxing the protein after removing RNA (see Supporting Information). Figure S6 shows the changes of RMSDs of backbone atoms of the two complexes as a function of time, respectively. The two systems reach an equilibrium state after 40 ns. Next, the Molecular mechanics Poisson Boltzmann surface area (MM-PBSA) approach was performed to calculate the binding energies for the two systems. In the MM-PBSA calculations, we used the CHARMM 36 all-atom force field. A dielectric constant of 1 was used for the solute and 80 for the aqueous solvent, and the ionic strength was set to 0.15 M. The equilibrium trajectory after 10 ns was used for the calculation. For the binding free energy calculation, although the relatively good results can be obtained for some cases of protein-protein and protein-ligand interactions, for the highly charged systems such as protein-RNA/DNA interactions the results are still not ideal, partial reason for which is from the inaccurate entropy change calculation.<sup>40</sup> Therefore, here only the binding energy is considered. Table S2 gives the binding energies and their components including the electrostatic and van der Waals energies, and polar and nonpolar solvation energies. From

Table S2, compared with the wild-type complex with binding energy being  $-751.64$  kcal/mol, the Lys263Glu mutant complex has a much higher binding energy of  $-26.97$  kcal/mol, which indicates that the positively charged Lys263 is important for the binding interaction between TDP-43 and RNA, and the mutation mainly weakens the electrostatic interaction. From Figure S7 that displays the interfaces of the two equilibrated complexes, it can be observed that the residue mutation of Lys263 to Glu has a big effect on the structures of residue segments of RRM1 loop3 and linker that are located at the interface, which further affects the contributions of these residues to the binding energy as shown in Figure S8 (the binding energy contribution of per residue in the two states). The above results indicate that the mutation of residue Lys263 to Glu weakens the interaction between TDP-43 and RNA through affecting the interface structure. Additionally, in 2021, Zheng et al. used MD simulations to explore the molecular basis of RNA recognition by TDP-43, and their results indicate that several residue individual mutations including Asp105, Phe147, Phe149, and Arg171 destabilize the interface structure and disrupt RNA-binding activity,<sup>41</sup> of which Phe147 and Arg171 are located at our predicted key residue clusters 3 and 5, respectively. This work is also a support to our prediction.

#### 4 | CONCLUSION

Transactive response DNA-binding protein 43 (TDP-43), an alternative-splicing regulator, plays multiple roles in pre-mRNA maturation and microRNA processing through specifically binding to long UG-rich RNAs. RNA binding causes TDP-43 to undergo a conformational rearrangement, strengthening the binding affinity between TDP-43 with target RNA.

Considering the advantage of ewmENM in reproducing function-related protein dynamics, we extend ewmENM to interpret the effect of RNA binding on the dynamics of TDP-43. First, we construct the optimal ENM of the complex with ewmGNM. Compared with the conventional GNM and pfGNM, ewmGNM achieves an evident improvement in flexibility prediction with the PCC of the experimental and theoretical MSFs being .86, higher than .81 and .80 from the conventional GNM and pfGNM, respectively. The advantage of ewmGNM is mainly attributed to the consideration of the multiscale effect of inter-residue interactions, which makes its good performance in flexibility prediction of residues regardless of loosely packed and tightly folded regions.

Then, the analyses on the effect of RNA binding on protein dynamics show that generally RNA binding causes a flexibility loss in loop1s and loop3s to some extent, suggesting their important roles in the electrostatic attraction for RNA and the induced fit in TDP-43-RNA recognition. A certain flexibility loss also occurs to linker and C-terminal, which play important roles in the conformational rearrangement and adjustment and in the specific binding. Also,  $\beta$ 1-3 in RRM2 have a slight flexibility loss, which form part of the RNA

binding surface, important for the specific interactions. Additionally, a more extensive flexibility loss in loop3 of RRM1 than in loop3 of RRM2 upon RNA binding, as well as the more residues and higher enrichment of positively charged residues in the former than in the latter suggest that loop3 of RRM1 are more important for RNA binding affinity, which partially explains the experimental observation that RRM1 contributes more to the interactions with RNA compared with RRM2.

Additionally, the cross-correlation analyses by ewmANM reveal that the enhanced and extended movement couplings within TDP-43 occur upon RNA binding, which suggests that RNA binding strengthens not only the intra-RRM interactions mainly among  $\beta$ -sheets, facilitating RRM structure stability, but also the inter-RRM interactions mainly between  $\beta$ 1,  $\beta$ 3 in RRM1, and  $\beta$ 4-5 in RRM2, important for stabilizing the RNA binding interface. Additionally, loop1s, loop3s, and  $\beta$ -sheet segments provide the main interactions with RNA, presenting strong dynamical correlations with RNA.

Finally, we extend the ewmGNM-based thermodynamic cycle approach to identify the key residues in TDP-43, whose mutations significantly alter the binding free energy of TDP-43 with RNA. The identified key residues can be grouped into three parts according to their positions in the protein structure. The key residues at loop segments rich in positively charged amino acids and highly flexible mainly contribute electrostatic attractions for RNA and induced fit in TDP-43-RNA interactions. The residues at  $\beta$ -sheets provide a perfect binding surface for RNA and generally participate in the direct specific interactions with RNA. Those at linker and C-terminal with high mobility are dynamically coupled with RNA binding, important for the conformational rearrangement and adjustment, as well as for the complex structure stability.

Although based on a coarse-grained elastic network model, ewmENM approach is featured with the analytical solution and timesaving virtues compared with atomic MD simulations, and importantly it takes the multiscale effect of inter-residue interactions into account, reproducing good details in protein flexibility prediction. Considering the approach's simplicity and the consistence of the analytical results with experimental data, we believe that ewmENM and ewmGNM-based thermodynamic cycle methods represent a promising approach, complementary to the atomic-level simulations, to investigate the binding and allosteric dynamics of protein and RNA interactions, which can be readily extended for other macromolecular interaction systems.

#### ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (31971180 and 11474013).

#### CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

#### AUTHOR CONTRIBUTIONS

Xueqing Deng, Chunhua Li, and Shihao Wang designed the research. Xueqing Deng performed ewmENM calculations and Weikang Gong

and Yang Liu carried out the conventional GNM and pfGNM calculations. Xueqing Deng and Zhongjie Han carried out the MD simulation calculations. Xueqing Deng and Chunhua Li performed data analyses and wrote the manuscript.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Xueqing Deng  <https://orcid.org/0000-0003-0991-6966>

Yang Liu  <https://orcid.org/0000-0002-3332-439X>

Chunhua Li  <https://orcid.org/0000-0002-0895-3506>

## REFERENCES

1. Yu H, Lu S, Gasior K, et al. HSP70 chaperones RNA-free TDP-43 into anisotropic intranuclear liquid spherical shells. *Science*. 2021;371(6529):eabb4309.
2. Buratti E, Baralle FE. TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem Sci*. 2012;37(6):237-247.
3. Ling S, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron*. 2013;79(3):416-438.
4. Ou SH, Wu F, Harrich D, García-Martínez LF, Gaynor RB. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J Virol*. 1995;69(6):3584-3596.
5. Lee EB, Lee VM, Trojanowski JQ. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat Rev Neurosci*. 2011;13(1):38-50.
6. Josephs KA, Whitwell JL, Weigand SD, et al. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol*. 2014;127(6):811-824.
7. Kuo P, Doudeva LG, Wang Y, Shen CJ, Yuan HS. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res*. 2009;37(6):1799-1808.
8. Kuo P, Chiang C, Wang Y, Doudeva LG, Yuan HS. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res*. 2014;42(7):4712-4722.
9. Lukavsky PJ, Daujotye D, Tollervey JR, et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol*. 2013;20(12):1443-1449.
10. Auweter SD, Fasan R, Reymond L, et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J*. 2006;25(1):163-173.
11. Furukawa Y, Suzuki Y, Fukuoka M, et al. A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Sci Rep*. 2016;6:20576.
12. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*. 1994;372(6505):432-438.
13. Wang X, Tanaka Hall TM. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol*. 2001;8(2):141-145.
14. Prakash A, Kumar V, Meena NK, Hassan MI, Lynn AM. Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *J Biomol Struct Dyn*. 2019;37(1):178-194.
15. Zacco E, Graña-Montes R, Martin SR, et al. RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *J Mol Biol*. 2019;431(8):1671-1688.
16. Buratti E, Baralle FE. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem*. 2001;276(39):36337-36343.
17. Bhardwaj A, Myers MP, Buratti E, Baralle FE. Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Res*. 2013;41(9):5062-5074.
18. Chen H, Topp SD, Hui HS, et al. RRM adjacent TARDDBP mutations disrupt RNA binding and enhance TDP-43 proteinopathy. *Brain*. 2019;142(12):3753-3770.
19. Chiang C, Grauffel C, Wu L, et al. Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Sci Rep*. 2016;6:21581.
20. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett*. 1996;77(9):1905-1908.
21. Bahar I, Wallqvist A, Covell DG, Jernigan RL. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry*. 1998;37:1067-1075.
22. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des*. 1997;2:173-181.
23. Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins*. 2000;40(3):512-524.
24. Zheng W. Coarse-grained modeling of conformational transitions underlying the processive stepping of myosin V dimer along filamentous actin. *Proteins*. 2011;79(7):2291-2305.
25. Yang L, Song G, Jernigan RL. Protein elastic network models and the ranges of cooperativity. *Proc Natl Acad Sci USA*. 2009;106(30):12347-12352.
26. Hinsen K. Analysis of domain motions by approximate normal mode calculations. *Proteins*. 1998;33(3):417-429.
27. Opron K, Xia K, Wei G. Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis. *J Chem Phys*. 2014;140(23):234105.
28. Xia K, Opron K, Wei G. Multiscale multiphysics and multidomain models—flexibility and rigidity. *J Chem Phys*. 2013;139(19):194109.
29. Opron K, Xia K, Burton Z, Wei G. Flexibility-rigidity index for protein-nucleic acid flexibility and fluctuation analysis. *J Comput Chem*. 2016;37(14):1283-1295.
30. Nguyen DD, Xia K, Wei G. Generalized flexibility-rigidity index. *J Chem Phys*. 2016;144(23):234106.
31. Xia K, Opron K, Wei G. Multiscale Gaussian network model (mGNM) and multiscale anisotropic network model (mANM). *J Chem Phys*. 2015;143(20):204106.
32. Wang S, Gong W, Deng X, Liu Y, Li C. Exploring the dynamics of RNA molecules with multiscale Gaussian network model. *Chem Phys*. 2020;538:110820.
33. Gong W, Liu Y, Zhao Y, Wang S, Han Z, Li C. Equally weighted multiscale elastic network model and its comparison with traditional and parameter-free models. *J Chem Inf Model*. 2021;61(2):921-937.
34. Su JG, Xu XJ, Li CH, Chen WZ, Wang CX. Identification of key residues for protein conformational transition using elastic network model. *J Chem Phys*. 2011;135(17):174101.
35. Su JG, Du HJ, Hao R, et al. Identification of functionally key residues in AMPA receptor with a thermodynamic method. *J Phys Chem B*. 2013;117(29):8689-8696.
36. He J, Han Z, Farooq QUA, Li C. Study on functional sites in human multidrug resistance protein 1 (hMRP1). *Proteins*. 2021;89:659-670.
37. Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys Rev Lett*. 1998;80(12):2733-2736.

38. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-242.
39. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 2005;272(9):2118-2131.
40. Duan L, Liu X, Zhang JZ. Interaction entropy: a new paradigm for highly efficient and reliable computation of protein-ligand binding free energy. *J Am Chem Soc.* 2016;138(17):5722-5728.
41. Sun H, Chen W, Chen L, Zheng W. Exploring the molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43 using molecular dynamics simulation and free energy calculation. *J Comput Chem.* 2021;42(23):1670-1680.

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Deng X, Wang S, Han Z, Gong W, Liu Y, Li C. Dynamics of binding interactions of TDP-43 and RNA: An equally weighted multiscale elastic network model study. *Proteins.* 2022;90(2):589-600. doi:10.1002/prot.26255