Insights into the Aggregation Mechanism of RRM domains in TDP-43: A Molecular Dynamics Study

Wei Liu

Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875

Chaoqun Li (✉ lichaoqun@hdc.edu.cn)

Hebei Key Laboratory of Heterocyclic Compounds, College of Chemistry, Chemical Engineering and Materials, Handan University, Handan, 056005, Hebei Province

Jiankai Shan

Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875

Yan Wang

Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875

Guangju Chen

Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875

Research Article

Keywords: TDP-43 protein, RRM1 and RRM2 domain, molecular dynamics simulation, aggregation

Posted Date: December 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-124204/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Insights into the aggregation mechanism of RRM domains in TDP-43: A molecular dynamics study

Wei Liu¹, Chaoqun Li^{2, *}, Jiankai Shan¹, Yan Wang¹ and Guangju Chen^{1, *}

1. *Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875, China*

2. *Hebei Key Laboratory of Heterocyclic Compounds, College of Chemistry, Chemical Engineering and Materials, Handan University, Handan, 056005, Hebei Province, China.*

To whom correspondence should be addressed. Tel: +86-0310-6260302. *E-mail address:*

lichaoqun@hdc.edu.cn (C. Li); gjchen@bnu.edu.cn (G. Chen).

ABSTRACT

The transactive response DNA-binding protein 43 (TDP-43) is associated with several diseases such as Amyotrophic lateral sclerosis (ALS) and Frontotemporal lobar degeneration (FTLD) due to pathogenic aggregations. In this work, we examined the dimer, tetramer and hexamer models built from the RRM domains of TDP-43 using molecular dynamics simulations in combination with the protein-protein docking. Our results show that the parallel β -sheet layers in the RRM1 domains in these oligomer models were formed energetically favorable, which provide the potential binding sites in the aggregation process. The closeness of the parallel β -sheet layers from small oligomer models gradually expanded to large ones through the communication of the interaction between $\alpha 1/\alpha 2$ helices of RRM domains, which enhanced the binding affinities and interactions in these aggregation models to further promote the aggregation. Using the repeatable-superimposing method based on the tetramer models, we propose a new aggregation mechanism of RRM domains in TDP-43 which mediates the formation of large aggregation structure with the repeated, helical and rope-like characteristics. These new insights help to understand the amyloid-like aggregation phenomena of TDP-43 protein in ALS and FTLD diseases.

Keywords: TDP-43 protein; RRM1 and RRM2 domains; molecular dynamics simulation; aggregation.

Introduction

Many human neurodegenerative diseases are correlated with the abnormal accumulation of proteins in the neuron tissues¹⁻³. Recent evidence has further connected progressive disorder diseases of Amyotrophic lateral sclerosis (ALS), Frontotemporal lobar degeneration (FTLD), Alzheimer's disease (AD) to the aggregation or polymer formation of proteins⁴⁻⁷. The transactive response RNA-binding protein 43 (TDP-43) is not only a primarily nuclear RNA-binding proteins but also one of the potential aggregation proteins⁸. It has been evidenced that aggregation in nervous tissues of wild-type TDP-43 is a potential risk in ALS and FTLD⁹⁻¹¹. Therefore, investigation of TDP-43 aggregation mechanism has gained momentum in recent years¹²⁻¹⁴.

TDP-43 was first discovered in 1995¹⁵ and is implicated in mRNA metabolism, such as mRNA transport, mRNA stability, pre-mRNA splicing and miRNA processing⁸. TDP-43 contains 414 amino acids¹⁵ and consists of one N-terminal domain (NTD), two repeated RNA recognition motif domains (RRM domains), one C-terminal domain (CTD) and linking loops¹⁶. The domain schematic is shown in Figure 1a. Specifically, NTD (residues 1-101) includes one α -helix, six β -strands and several loops¹⁷, while CTD (residues 274-414) contains a helix-turn-helix structure and unstructured loops¹⁸. RRM domains feature two independent domains including two α -helices (α 1, α 2) and five β -strands (β 1- β 5), namely RRM1 (residues 102-176) and RRM2 (residues 192-267), with a α -linker loop between the two domains¹⁹ (Figure 1b). The aggregation phenomena mostly mediated by independent NTD and CTD have been extensively investigated *in vivo* and *in vitro*²⁰⁻²⁴. However, previous studies are not much focus on the RRM domains. Recently, some other studies have shown that the interaction of TDP-43 with nucleic acids are mainly achieved by RRM domains to result in aberrant assembly, which plays an important role in pathological aggregation of TDP-43²⁵⁻²⁷. Moreover, it has been documented that RRM domains have the amyloid-like aggregation characteristics¹³. The amyloid-like structures

in RRM1 and RRM2, such as a parallel β -sheet layer, exposes one side of β -sheet of RRM and facilitates protein-protein aggregation or protein-RNA interaction, which further supports the effectiveness of RRM1 and RRM2 in promoting TDP-43 aggregation to result in an amyloid-like structure formation^{1,13,28}. In addition, RNA was shown to be an important factor in preventing aggregation of TDP-43^{29–32}.

Molecular dynamics (MD) simulations of two mutated structures (D169G and K263E) of RRM domains have been reported, which predicted the mutation leading aggregation tendency³³. Protein-protein docking studies indicated that the cysteine residues in NTD contribute to aggregation propensity³⁴. Atomistic MD simulations studies helped elucidate the structural stability and the unfolding dynamics of NTD and RRMs, revealing that non-native unfolded intermediate ensembles induced oligomerization and aggregation of TDP-43^{35,36}. MD studies have also shown that the 341–366 segment of CTD in was able to drive aggregation of the entire TDP-43 protein³⁷. However, theoretical studies on the aggregation structures of wild RRM domains of TDP-43 are scarce so far. Especially, detailed aggregation processes by which small oligomers promote large oligomers leading to an amyloid-like aggregation of TDP-43 are elusive but necessary to understand the aggregation mechanisms of RRM1 and RRM2. To gain a deeper understanding of the structural and dynamical basis for the aggregation mechanism and the oligomerization process of RRM domains in TDP-43, we used protein-protein docking to build the initial structures of dimer, tetramer and oligomer models and carried out MD simulations with free energy calculations at atomic levels. Our main objectives were: (1) To establish as many small oligomer models as possible of RRM1 and RRM2, calculate and compare their stabilities; (2) To elucidate the possible aggregation mechanism by which small oligomers promote large oligomers. This study will help better understand the RRM domains amyloid-like aggregation mechanism. Our models may be used for developing novel drugs in ALS and FTLN diseases.

Results

Model building and characteristics of monomer models of RRM domains

Because of the importance of RRM1 and RRM2 in TDP-43 aggregation^{13,27,38}, we have chosen them to investigate the aggregation mechanism. Given the absence of the intact experimental structures of RRM domains, we built the model structure of RRM1-RRM2 together with the linking loop (designated as monomer RRM12 model) using the individual RRM1 and RRM2 NMR structures (PDB codes: 2CQG and 1WF0), and the linking loop between the two RRMs in the RNA-bound RRM domains structure (PDB code: 4BS2). The structural construction detail is shown in Supplementary Section S1. The model system was explicitly solvated in a rectangular TIP3P water box with the solvent shell extended to 8Å in all directions of this model. 19 Na⁺ and 16 Cl⁻ counterions were added to achieve electroneutrality and to satisfy the experimental ionic strength of 75 mM for the RRM12 model¹³. This counterion process was used to other aggregation models.

The RRM12 model was subjected to 50 ns MD simulation. The root-mean-square deviation (RMSD) values of all backbone atoms referenced to the corresponding starting structure over the MD trajectory were examined to determine if the system had attained equilibrium (Figure 2a). It is often considered that small RMSD values of one simulation indicate a stable state of the system, also suggest that the newly constructed RRM12 model can satisfactorily reproduce the experimental structures. The RRM12 model reached equilibrium after 40 ns, and the energies were stable during the remainder of the simulation. Next, the equilibrated conformation of this model was extracted by recording 5000 snapshots at 2 ps time-interval of the trajectory between 40 ns and 50 ns of simulation time. The corresponding average structure of RRM12 model is shown in Figure 2b. In this average structure, RRM1 and RRM2 are connected by a small α -helix as the flexible linker. Both domains display a similar conformation, i.e. the five β -strands of β_1 , β_2 , β_3 , β_4 , β_5 forming a curved sheet layer. This calculated monomer RRM12 model will be prepared to

construct the oligomer aggregation models for the investigations of the aggregation mechanisms of RRM domains of TDP 43 protein.

The structure and interaction characteristics of dimer models of RRM domains

Structure analysis of dimer models

To explain the aggregation process of RRM domains of TDP-43, the initial dimer models were produced from the average conformation of the calculated RRM12 model by protein-protein docking method using ZDOCK (<http://zdock.umassmed.edu/>)³⁹. Five dimer models were selected from the top 250 docking results with the high ZDOCK scores. The details of the docking procedure are available in Supplementary Section S2. Finally, MD runs of 50 ns for five models were carried out by following the same protocol above.

Based on the similar trajectory analysis, MD simulations for all five dimer models reached equilibrium after 40 ns, RMSD of the five dimer models are shown in Figure S1a-e and the corresponding energies were found to be stable during the remainder of the simulation. Therefore, the equilibrated conformations were extracted between 40 ns and 50 ns of simulation times (Figure 3). Based on the monomer RRM12 model (Figure 2b), their two long sides and two short sides are respectively called “shoulder- α ”, “shoulder- β ”, and “head-1”, “head-2”. The shoulder- α represents the side consisted of two α 2-helices in RRM1 and RRM2, and the shoulder- β side consisted of two β -sheet layers in RRM1 and RRM2, respectively. Both of two side represent a linking loop (or the α -helix-linker) between RRM1 and RRM2; The head-1 and head-2 represent the side consisted of α 1, α 2 helices in RRM1, and the side consisted of β -sheet layer tail in RRM2, respectively (Figure 2b). The five dimer models are termed as $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$, $S_{\beta}H_{1\beta}/D$ and H_1H_1/D models with three classified groups of shoulder-to-shoulder, i.e. $S_{\alpha}S_{\beta}/D$ and $S_{\beta}S_{\beta}/D$ models, shoulder-to-head, i.e. $S_{\beta}H_{1\alpha}/D$ and $S_{\beta}H_{1\beta}/D$ models, and head-to-head, i.e. H_1H_1/D model, respectively (Figure 3).

It can be seen from the average structure in Figure 3a that the $S_\alpha S_\beta/D$ model presents that the α -helix linker between RRM1 and RRM2 in the shoulder- α side of one monomer RRM12 model (cyan in Figure 3a) contacts diagonally to two curve β -sheet layers of RRM1 and RRM2 domains in the shoulder- β side of the other monomer RRM12 model (green in Figure 3a). In Figure 3b, the $S_\beta S_\beta/D$ model shows that the curve β -sheet layer of RRM1 in the shoulder- β side of one monomer RRM12 model (cyan in Figure 3b) inserts into the interstice of two β -sheet layers in the shoulder- β side of the other monomer RRM12 model (green in Figure 3b) in parallel, and *vice versa*. In these two models there are two interfaces between two monomer models due to the combination along two shoulder sides of two monomer models, named shoulder-to-shoulder. The $S_\beta H_{1\alpha}/D$ model in Figure 3c shows that the α_2 -helix of head-1 side of one monomer RRM12 model (cyan in Figure 3c) inserts into the interstice with some angle of two β -sheet layers in the shoulder- β side of the other monomer RRM12 model (green in Figure 3c). The $S_\beta H_{1\beta}/D$ model (Figure 3d) depicts that the curve β -sheet layer of head-1 side in one monomer RRM12 model (cyan in Figure 3d) inserts vertically into the interstice of two β -sheet layers in the shoulder- β side of the other monomer RRM12 model (green in Figure 3d). In these two models there still are two interfaces between two monomer models due to the combination of one shoulder side in one monomer model with a head side in the other monomer model, so called the shoulder-to-head. The $H_1 H_1/D$ model (Figure 3e) illustrates that the α_2 -helix of head-1 side in one monomer RRM12 model combines to that in the other monomer RRM12 model, which shows clearly one interface between two monomer models due to the combination of head-to-head.

Analyses of binding free energies and interactions of the dimer models

To address the aggregation abilities of the five dimer models, $S_\alpha S_\beta/D$, $S_\beta S_\beta/D$, $S_\beta H_{1\alpha}/D$, $S_\beta H_{1\beta}/D$ and $H_1 H_1/D$, their binding free energies were investigated by using the MM-PBSA method based on their corresponding MD simulations. All energy terms and the total

binding free energies for these models are given in Table 1. The binding free energies of $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$, $S_{\beta}H_{1\beta}/D$ and H_1H_1/D dimer models are $-84.07 \text{ kcal}\cdot\text{mol}^{-1}$, $-73.14 \text{ kcal}\cdot\text{mol}^{-1}$, $-80.49 \text{ kcal}\cdot\text{mol}^{-1}$, $-70.76 \text{ kcal}\cdot\text{mol}^{-1}$ and $-43.83 \text{ kcal}\cdot\text{mol}^{-1}$, respectively. This result supports the experimental results that RRM1 and RRM2 themselves can cause the aggregation of TDP-43 in the absence of RNA. Moreover, based on their different energies of the five dimer models, their aggregation abilities with two interfaces between two monomer RRM12 models are evidently greater than that with one interface. As such, aggregation mechanisms related to the two monomer RRM12 models along the shoulder sides are more favorable.

To further explore the interaction mechanisms between the two RRM12 monomer models in these dimers with the different binding modes, the percentages of occurrences of all possible hydrogen bond and hydrophobic interactions located at the interfaces for the five dimer models were extracted from the MD simulations and are shown in Table 2. The criteria include a donor–acceptor distance of $<3.5 \text{ \AA}$ and a donor–proton–acceptor angle of $>120^\circ$ for a hydrogen bond, while a C–C distance of $<4.5 \text{ \AA}$ for a hydrophobic interaction. It can be seen from Table 2 for the five dimers that the total occupancies of hydrogen bond and hydrophobic interaction are 523.81% and 584.64% for $S_{\alpha}S_{\beta}/D$ model, 422.26% and 427.00% for $S_{\beta}S_{\beta}/D$, 455.24% and 812.28% for $S_{\beta}H_{1\alpha}/D$, 425.80% and 380.84% for $S_{\beta}H_{1\beta}/D$, and, 209.30% and 186.46% for H_1H_1/D , respectively. Obviously, the total occupancies for the hydrogen bond or hydrophobic interaction for $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$ and $S_{\beta}H_{1\beta}/D$ models with two interfaces between two monomer RRM12 models are larger than those for H_1H_1/D one with one interface, which are consistent with the binding free energy calculations. Some binding sites have been found at the corresponding interfaces with the binding type of shoulder or head for these models. Moreover, the ionic interaction analysis for $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$, $S_{\beta}H_{1\beta}/D$ and H_1H_1/D models were calculated and are shown in Figure 4. Similarly, the criterion for an ionic interaction

includes a distance of $<5.0 \text{ \AA}$ between the HN group of positive Arg or Lys and the O atom of negative Asp or Glu. It can be seen that the ionic interaction mainly occurred at the residue-pairs of Asp174-Lys176, Glu156-Arg171 and Arg165-Asp174 with the total occupancy of 261.9% for $S_{\alpha}S_{\beta}/D$ model, the residue-pairs of Glu200-Lys140 and Arg165-Glu261 with the total occupancy of 141.0% for $S_{\beta}S_{\beta}/D$ model, the residue-pairs of Lys136-Asp174 and Lys145-Asp174 with the total occupancy of 177.6% for $S_{\beta}H_{1\alpha}/D$ model, the residue-pairs of Arg171-Glu261 and Glu246-Lys140 with the total occupancy of 187.2% for $S_{\beta}H_{1\beta}/D$ model. Conversely, there is no ionic interaction for H_1H_1/D model. These results are consistent with the analyses of binding energies and interactions discussed above and the detail can be found in Supplementary Section S3. Figure 3 shows the binding interfaces in the dimer models, and as expected the rank of the binding areas in this interface is consistent with the order of the binding free energies and interactions in this interface. In conclusion, the greater the aggregation ability, the larger the binding area.

To compare the energy and interaction differences of the aggregation mechanisms and to further provide the quantitative information of energy contributions of the residues, decompositions of the corresponding binding free energies into the key residues were performed for the five dimer models, and are shown in Figure 5a, b (only for $S_{\alpha}S_{\beta}/D$ and H_1H_1/D dimer models). For the two models with the energy contributions of $> 2 \text{ kcal}\cdot\text{mol}^{-1}$ (i.e. the decomposition energy values less than $-2 \text{ kcal}\cdot\text{mol}^{-1}$) (Figure 5), the distinct differences of energy decompositions occur mainly at several crucial residues. Clearly, the number of critical interaction sites in the $S_{\alpha}S_{\beta}/D$ dimer model is more than the H_1H_1/D model. The number and locations of binding residue support the free energy calculations, and the results of hydrogen bond and hydrophobic interaction analyses. Other decomposition details of the remaining dimer models can be found in Figure S2.

Structural modelling and energy analysis of tetramer models of RRM domains

To investigate the larger aggregation phenotype mediated by RRM1 and RRM2, two tetramer models of $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ based on the most stable $S_{\alpha}S_{\beta}/D$ dimer model in the present work were built by using ZDOCK with the top energy scoring, and simulated by using the same method described above. The equilibrated conformations of $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ are shown in Figure 6a, b, and RMSD values are shown in Figure S1f. The $S_{\alpha}S_{\beta}1/T$ tetramer model depicts that both head-1 sides associate with both head-1 sides, i.e. two dimers are bound together with the type of head-to-head along the shoulder horizontal direction. However, the $S_{\alpha}S_{\beta}2/T$ tetramer model (Figure 6b) shows that both head-2 sides associate with both head-2 side of the $S_{\alpha}S_{\beta}/D$ dimer model in the direction of head-to-head opposite to the first dimer, i.e. head-2 to head-2 along the tilt head direction. The binding free energies (Table S1 and Table S2) in the head-1-to-head-1 or head-2-to-head-2 interfaces between two $S_{\alpha}S_{\beta}/D$ dimer models for $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ tetramer models are respectively $-62.52 \text{ kcal}\cdot\text{mol}^{-1}$ and $-55.43 \text{ kcal}\cdot\text{mol}^{-1}$. The binding free energies are $-89.30 \text{ kcal}\cdot\text{mol}^{-1}$ and $-99.99 \text{ kcal}\cdot\text{mol}^{-1}$ in the shoulder-to-shoulder interfaces of $S_{\alpha}S_{\beta}/D$ dimer (cyan and green in Figure 6a) in the tetramer $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ models, respectively. These are more energetically favorable than the average binding free energy of $-84.07 \text{ kcal}\cdot\text{mol}^{-1}$ in the $S_{\alpha}S_{\beta}/D$ model, which suggests that the affinity of shoulder-to-shoulder interface in the tetramer models is increased compared to the corresponding dimer model favoring the formation of tetramer aggregation. The details of others tetramer models from the $S_{\beta}S_{\beta}/D$ and $S_{\beta}H_{1\alpha}/D$ dimer models are also analyzed, and are shown in Section S4, Figure S1f, and Supplementary Figure S3.

To investigate the relationship between structural dynamics and interactions, the dynamical fluctuations of every residue of the RRM12, $S_{\alpha}S_{\beta}/D$ and $S_{\alpha}S_{\beta}1/T$ models were determined by RMSF values (Figure 7a). RMSF values of the dimer $S_{\alpha}S_{\beta}/D$ model and tetramer $S_{\alpha}S_{\beta}1/T$ model are decreased compared to the RRM12 model, as a direct result of new hydrogen bond, hydrophobic and ionic interaction formations, which further supports

the interaction analysis (Table 2, Table S2 and Figure 4). Aggregations of the dimer $S_{\alpha}S_{\beta}/D$ model and tetramer $S_{\alpha}S_{\beta}1/T$ models are mainly achieved by parallel β -sheet layer of two RRM1 domains and the connecting flexible loops. As expected, the residues in $\beta 1$, $\beta 3$ and $\beta 5$ of parallel β -sheet layer of RRM1 domain have the lower RMSF values. The increased affinity of shoulder-to-shoulder interface in the tetramer $S_{\alpha}S_{\beta}1/T$ model may result from the allosteric communication from head-to-head interface to shoulder-to-shoulder interface. To address further the allosteric communication, the motion correlations of backbone atoms over 10 ns simulation time in the aggregation process from the dimer $S_{\alpha}S_{\beta}/D$ model to tetramer $S_{\alpha}S_{\beta}1/T$ model were calculated, and the corresponding cross correlation map is shown in Figure 7b. The range of the residue motion correlations was labeled from low anticorrelation (blue) to high correlations (red). Motions of $\alpha 1$ and $\alpha 2$ helices of the RRM1 domains in the head-to-head interface significantly correlate with the motions of the parallel β -sheet layer ($\beta 1$, $\beta 3$ and $\beta 5$) of the same RRM1 domains in the shoulder-to-shoulder interface. These results imply the allosteric communication in the RRM1 domains from the $\alpha 1$ and $\alpha 2$ helices in the head-to-head interface to the parallel β -sheet layer in the shoulder-to-shoulder interface.

The large oligomer aggregation characteristics

To explore the large oligomer aggregation mechanism and to explain the pathogenic aggregation phenomenon in vivo¹³, we further built large aggregation models based on the small aggregation models. $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ tetramer models display the similar structures, apart from the different association modes of head-1-to-head-1 or head-2-to-head-2 from two $S_{\alpha}S_{\beta}/D$ dimer models. Especially, each $S_{\alpha}S_{\beta}/D$ dimer in the two tetramers still shows the similar conformations to the corresponding isolated $S_{\alpha}S_{\beta}/D$ dimer. By using the superimposing method of PyMol program (<http://www.pymol.org>), the hexamer model was built (as called $S_{\alpha}S_{\beta}/H$ model) from $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ tetramer model and is shown in Figure 8 with the scheme of the model. In detail, we use first an $S_{\alpha}S_{\beta}/D$ dimer of the

$S_{\alpha}S_{\beta}2/T$ model overlapping on same one in $S_{\alpha}S_{\beta}1/T$ model, then, by removing the repetitive $S_{\alpha}S_{\beta}/D$ dimer. Based on this superimposing method above, the two possible paths for establishing the repeating octamer model from the $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ tetramer models are shown in Figure 9a, in which an octamer model from the $S_{\alpha}S_{\beta}1/T$ tetramer model via the hexamer model can be obtained along two different overlapping directions on the $S_{\alpha}S_{\beta}1/T$ tetramer model, i.e. by using the left $S_{\alpha}S_{\beta}/D$ dimer of the $S_{\alpha}S_{\beta}2/T$ tetramer model overlapping with the right $S_{\alpha}S_{\beta}/D$ dimer of the $S_{\alpha}S_{\beta}1/T$ tetramer model (path1 in Figure 9a) or by using the right $S_{\alpha}S_{\beta}/D$ dimer overlapping with the left $S_{\alpha}S_{\beta}/D$ dimer (path2 in Figure 9a). The two octamer models built from these two paths are shown in Figure 9b, which possess the same repeating structure as the $S_{\alpha}S_{\beta}1/T$ tetramer model in equal aggregation mode. Consequently, the large aggregation models were established using the similar approaches. For example, the dodecamer model was built by repeatable-superimposing method (Figure 9c), featuring a repeating, helical and rope-like aggregation mode. The rope-like structural characteristics were detected in previous experiments³⁸. The $S_{\alpha}S_{\beta}/H$ hexamer model derived from path1 was simulated for 50 ns and the equilibrated conformation is shown in Figure S4. The small RMSD values support the stable of this hexamer model in aggregation (Figure S1g). The binding free energies for head-1-to-head-1 and head-2-to-head-2 interfaces in hexamer $S_{\alpha}S_{\beta}/H$ model are $-70.52 \text{ kcal}\cdot\text{mol}^{-1}$ and $-91.10 \text{ kcal}\cdot\text{mol}^{-1}$, respectively. Moreover, the binding free energy of $-120.93 \text{ kcal}\cdot\text{mol}^{-1}$ for the shoulder-to-shoulder interface of the center $S_{\alpha}S_{\beta}/D$ dimer (cyan and green in Figure S4) of hexamer $S_{\alpha}S_{\beta}/H$ model is more energetically favorable over that of $-89.30 \text{ kcal}\cdot\text{mol}^{-1}$ of the $S_{\alpha}S_{\beta}1/T$ dimer model. Our results suggest that the affinity of shoulder-to-shoulder interface in the center $S_{\alpha}S_{\beta}/D$ dimer domains in hexamer $S_{\alpha}S_{\beta}/H$ model experienced an increase by the formation of new head-1-to-head-1 and head-2-to-head-2 interfaces. The details of binding free energies are shown in Table S3. Our demonstrated stable hexamer model implies that studying aggregation process from tetramer to large oligomer aggregation using the repeatable-superimposing method is feasible, which generates a

repeating, helical and rope-like structure that likely mimics physiologic aggregation mode.

DISCUSSION

RRM1 dominated RRM domains dimerization

Possible aggregation mechanisms of RRM domains were elucidated based on the structural characteristics and interactions of $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$, $S_{\beta}H_{1\beta}/D$ and H_1H_1/D dimer models. The energy decomposition and interaction support the hypothesis that the aggregation of the dimer $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$ dimer models are mainly achieved by parallel β -sheet layer of two RRM1 domains and the connected flexible loops. In contrast, RRM2 domains are not a big play which have the weaker interactions and affinities in the aggregation. These observations are consistent with the experimental result which identified area of RRM1 binding to RNA⁴⁰. Binding of RNA to RRM1 and RRM2 could prevent TDP-43 aggregation²⁹⁻³². β_2 and β_3 strands in both RRM1 and RRM2 provide mainly hydrophobic interactions, except that the contribution of RRM1 is higher than that of RRM2 for RNA target binding^{19,41}. For example, the interactions and affinities for protein aggregations in $S_{\alpha}S_{\beta}/D$ model occurred mainly at Leu109 of β_1 ; Leu147 of β_3 ; Glu156, Val159, Met162 of α_2 ; Asp165 of β_4 ; Arg171, Asp174, Cys175, Lys176, Leu177 of β_5 in RRM1, which are associated with the domains for RNA binding, directly or indirectly. In particular, β_1 , β_3 and β_5 of two RRM1 domains form the typical interacting surface on the parallel β -sheet layer for RNA binding. The similar results were obtained in the dimer $S_{\beta}S_{\beta}/D$ and $S_{\beta}H_{1\alpha}/D$ models (Table S4). The binding sites located at the parallel β -sheet layers of the RRM12 dimers are almost in the same positions for RNA binding domains, suggesting that TDP-43 binding to RNA prior would prevent or slow down TDP-43 aggregations due to the same binding sites. These results imply that the parallel β -sheet layer of two RRM1 domains plays an important role in the dimer aggregation, and RNA binding can protect the pathogenic aggregation of TDP-43, consistent with experimental

results^{13,42,43}. In addition, since the parallel β -sheet layer as the main structural feature of amyloid structure is widely accepted⁴⁴, it is plausible that TDP-43 may form the large amyloid-like aggregation as a result of the interaction of the parallel β -sheet layer in two RRM1 domains.

Importantly, the distance between two β sheets in two RRM1 domains for the $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$ and $S_{\beta}H_{1\alpha}/D$ models (17.1 Å, 16.4 Å and 19.8 Å) and parallel stacking of the two β sheets are comparable with amyloid-like structure (Figure 10a). In contrast, the distances for the $S_{\beta}H_{1\beta}/D$ and H_1H_1/D models are longer (22.5 Å and 23.9 Å) (Figure 10a) and the two β sheets of two RRM1 domains are not parallel (Figure 3d, e). In this case, the parallel β -sheet structure would not form. As a result, the $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$ and $S_{\beta}H_{1\alpha}/D$ models are likely favored to form large oligomer through amyloid-like aggregation.

Analyses of the oligomer aggregation

The larger models were built by taking advantage of the dimer models and the periodic nature of oligomer aggregation. The amyloid aggregation forms a cross- β structure, in which β -strands are oriented perpendicularly to the aggregation axis (Figure 11) and are assembled into β -sheet layers that run the length of the aggregation. In the tetramer models of $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$, the interactions in the shoulder-to-shoulder interface mainly occurred at Leu109 of β_1 , Phe147 of β_3 , Glu156, Val159, Met162, Ser163, Gln164 of α_2 , Asp165 of β_4 , Arg171, Asp174, Cys175, Lys176, Leu177 of β_5 in RRM1, which are similar with those in the $S_{\alpha}S_{\beta}/D$ dimer model. The details of the interactions in shoulder-to-shoulder interface for $S_{\alpha}S_{\beta}1/T$, $S_{\alpha}S_{\beta}2/T$ models are shown in Table S5. As expected, the residues in β_1 , β_3 and β_5 of parallel β -sheet layers of two RRM1 domains have the lower RMSF values.

The distances between the two β -sheet layers of two RRM1 domains in the $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$ models are 15.8 Å and 16.4 Å (Figure 10a), smaller than 17.1 Å in the $S_{\alpha}S_{\beta}/D$

model. The closeness of the parallel β -sheet layer induces decreased free energies in the shoulder-to-shoulder interface in the tetramer model. $\alpha 1$ and $\alpha 2$ helices of RRM1 domains in the head-to-head interface correlate with the parallel β -sheet layer (including $\beta 1$, $\beta 3$ and $\beta 5$) in the same domains in the shoulder-to-shoulder interface with large cross-correlate coefficients (Figure 7), which represents the allosteric communication from $\alpha 1$ and $\alpha 2$ to the parallel β -sheet layer (i.e. $\beta 1$, $\beta 3$ and $\beta 5$). Such communication is related to the structural changes around $\alpha 1$, $\alpha 2$ and the parallel β -sheet layer. Tetramer aggregation caused the movement of $\alpha 1$ and $\alpha 2$ in the head-to-head interface. The distances of two $\alpha 1$ and two $\alpha 2$ in both RRM1 domains decreased from 37.0 Å to 36.6 Å and from 47.4 Å to 46.7 Å, respectively (Figure 10b). Because $\alpha 1$ is connected to $\beta 1$, and $\alpha 2$ is connected to $\beta 3$ and $\beta 5$, the movements of $\alpha 1$ and $\alpha 2$ is definitely related to the closeness of the parallel β -sheet layer, as expected from the correlations of $\alpha 1$ and $\alpha 2$ to the parallel β -sheet layer. The closeness of the parallel β -sheet layer indicates that the parallel β -sheet layers of RRM1 of TDP-43 in the dimer model are also retained in the tetramer model which are perpendicular to the aggregation axis, likely facilitating the formation of amyloid-like aggregation.

Furthermore, the large aggregation models were constructed by using the same repeatable-superimposing method described in the “**The large oligomer aggregation characteristics**”, in which the structural character with the repeatable, helical and rope-like as a mimicked aggregation mode has been appeared by this oligomer model. The results obtained from the tetramer models may also satisfy the large aggregation models. This is because the parallel β -sheet layer still exists in the large aggregation, such as hexamer, octamer, dodecamer and so on, and orients perpendicularly to the aggregation axis. These interactions between the parallel β -sheet layers favor the formation of large aggregation.

Aggregation analysis of intact TDP-43

The aggregation possibility of the intact TDP-43 protein was further studied. The intact TDP-43 model was constructed (Figure 12a, detail in Section S5), NTD and CTD connect RRM12 domain by two long and flexible linkers, and the average centroid distances between NTD and RRM12, and between CTD and RRM12 are 52.4 Å and 51.5 Å, respectively. The average angle among the average centroid of NTD, the RRM12 and CTD is 130.5°. Since NTD and CTD are in the opposite direction on a nearly parallel line, and are relatively far from RRM12, the RRM domains mediated aggregations are not occupied by NTD and CTD, which retain enough aggregation space to expose binding motifs. The aggregated dimer of intact TDP-43 was constructed using the same superimposing approach of the five dimer models. For example, the overlaid full TDP-43 dimer structure based on $S_{\alpha}S_{\beta}/D$ is shown in Figure 12b. It can be seen that NTD and CTD do not affect TDP-43 dimer aggregation because the large space are retained to expose the aggregation binding motifs. The other full TDP-43 dimer structures are shown in Figure S5. Similarly, the tetramer, hexamer, octamer, dodecamer and large oligomer aggregation model of TDP-43 can be constructed based the $S_{\alpha}S_{\beta}1/T$ model. In summary NTD and CTD may not prevent the aggregation of TDP-43 because the unoccupied binding motifs for aggregations. Those results predict that the TDP-43 may have the similar aggregation ability with RRM domains.

CONCLUSION

We performed molecular dynamics simulations, binding free energy calculations and interaction analyses of the monomer, dimer, tetramer and larger oligomer of RRM domains in TDP-43 to reveal their aggregation mechanism and aggregation process from small oligomer to large aggregation. Amongst five dimer $S_{\alpha}S_{\beta}/D$, $S_{\beta}S_{\beta}/D$, $S_{\beta}H_{1\alpha}/D$, $S_{\beta}H_{1\beta}/D$ and H_1H_1/D models of TDP-43, the $S_{\alpha}S_{\beta}/D$ model was found most stable. Aggregation of the

S_αS_β/D dimer model is mainly achieved by interactions of the parallel β-sheet layers of two RRM1 domains, indicating the characteristics of amyloid-like structure. TDP-43 binding to RNA results in preventing or slowing down TDP-43 aggregations due to the same sites for RNA binding and dimer aggregation. In the tetramer models, the movement of α1 and α2 in the head-to-head interface induced the closeness of the parallel β-sheet layers in RRM1 domains in the shoulder-to-shoulder interface, causing decreased binding free energy in the shoulder-to-shoulder interface for the tetramer and large oligomer aggregations. In the large aggregation models built using the same repeatable-superimposing method based on the tetramer model, RRM domains-mediated structural characteristics of the repeating, helical and rope-like aggregation mode may drive amyloid-like aggregation. The parallel β-sheet layers also exist in the large aggregation and are perpendicular to the aggregation axis, the interactions of which favor aggregation. In the same way, intact TDP-43 and RRM domains may have the similar ability and mode for the large amyloid-like aggregations due to sufficient space and exposed aggregation motifs. Our results imply that the RRM1 domains could play an important role in TDP-43 aggregation process. The mimicked aggregation mode is helpful for understanding the amyloid-like aggregation mechanism of TDP-43, and providing theoretical basis for ALS and FTLN disease prevention.

METHODS

Model building of TDP-43 aggregation models

Five initial dimer models were produced from the average structure of the monomer RRM12 model by using the protein-protein docking^{39,45,46}, and the details of construction process can be found in supplementary Section 1. The tetramer models were construction based on the average structure of the most stable dimer S_αS_β/D model using the same method. The large aggregation models, such as hexamer, octamer, dodecamer and so on, were constructed using the superimposing-creating method, and the details of construction

process can be found in “**The large oligomer aggregation characteristics**”. The model system was explicitly solvated by using the transferable intermolecular potential 3P (TIP3P) water inside a rectangular box large enough to ensure the solvent shell extended to 8Å in all directions of this model. 19 Na⁺ and 16 Cl⁻ counterions were added to achieve electroneutrality and to satisfy the experimental ionic strength of 75 mM for the RRM12 model.

Molecular dynamics simulation and ZDOCK

MD simulations for the monomer, dimer, tetramer and hexamer models, including minimization, heating and equilibration, were performed by using a classical AMBER parm99⁴⁷, which was run with sander module of AMBER16 package⁴⁸. The details of the MD procedure are given in Supplementary Section 2. The Cpptraj and Ptraj utilities of the AMBER16 program were used to obtain root-mean-square deviation (RMSD), average structures, hydrogen bond and hydrophobic interactions *etc.*⁴⁹. ZDOCK’s algorithm based on Fast Fourier Transform searches all possible binding modes by changing the translational and rotational positions of proteins, and each mode was evaluated by a comprehensive energy scoring function which includes some terms of shape complementarity, electrostatics and statistical potential^{50–52}. ZDOCK software was used to search for possible binding conformations of RRM12 domain (version 3.0.2, <http://zdock.umassmed.edu/>)^{39,50–52}. The details of the docking procedure are available in Supplementary Section 2.

Binding free energy analyses

The molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method was used to perform the binding free energy analyses in AMBER16 package^{53,54}. The binding free energy ($\Delta G_{\text{binding}}$) was calculated based on followed equation (1):

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{ligand}} + G_{\text{receptors}}) \quad (1)$$

,where G_{complex} , G_{ligand} , and $G_{\text{receptors}}$ are the free energies of complex, ligand and receptor calculated from the snapshots of the MD trajectories, respectively. The binding free energy was calculated as follows (2):

$$\begin{aligned} \Delta G_{\text{binding}} &= \Delta H_{\text{binding}} - T\Delta S \\ &= \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \end{aligned} \quad (2)$$

Because $T\Delta S$ term usually is costly in terms of computational resource and standard deviation is too large from normal-mode, which could be neglected when receptor and ligand have similar conformations in the binding free energy calculation⁵⁵. The details of the MM-PBSA method can be found in Supplementary Section 6.

Data availability

The structure of RRM12 model of TDP-43 is available from the Protein Data Bank under the following codes: 2CQG, 1WF0 and 4BS2.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (no. 22073010 and 21573020); Colleges and Universities in Hebei Province Science and Technology Research Youth Talent Support Program (Grant Number BJ2019204).

References

1. Moreno, F. *et al.* A novel mutation P112H in the TARDBP gene associated with frontotemporal lobar degeneration without motor neuron disease and abundant neuritic amyloid plaques. *Acta Neuropathol. Commun.* **3**, 19, (2015).
2. Josephs, K. A. *et al.* TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol.* **127**, 811–824, (2014).
3. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10 Suppl**, (2004).
4. Shiina, Y., Arima, K., Tabunoki, H. & Satoh, J.-i. TDP-43 dimerizes in human cells in culture. *Cell. Mol. Neurobiol.* **30**, 641–652, (2010).
5. Davis, S. A., Gan, K. A., Dowell, J. A., Cairns, N. J. & Gitcho, M. A. TDP-43 expression influences amyloid β plaque deposition and tau aggregation. *Neurobiol. Dis.* **103**, 154–162, (2017).
6. Roychaudhuri, R., Yang, M., Hoshi, M. M. & Teplow, D. B. Amyloid beta-protein assembly and Alzheimer disease. *J. Biol. Chem.* **284**, 4749–4753, (2009).
7. Sun, Y., Arslan, P. E., Won, A., Yip, C. M. & Chakrabartty, A. Binding of TDP-43 to the 3'UTR of its cognate mRNA enhances its solubility. *Biochemistry* **53**, 5885–5894, (2014).
8. Buratti, E. & Baralle, F. E. TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem. Sci.* **37**, 237–247, (2012).
9. Arnold, E. S. *et al.* ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *PNAS.* **110**, E736-45, (2013).
10. Afroz, T. *et al.* Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. *Nat. Commun.* **8**, 45, (2017).
11. Seyfried, N. T. *et al.* Multiplex SILAC analysis of a cellular TDP-43 proteinopathy model reveals protein inclusions associated with SUMOylation and diverse

- polyubiquitin chains. *Mol. Cell Proteomics. MCP* **9**, 705–718, (2010).
12. He, R.-Y. *et al.* Characterization and real-time imaging of the FTLD-related protein aggregation induced by amyloidogenic peptides. *Chem. Commun. (Cambridge, England)* **51**, 8652–8655, (2015).
 13. Zacco, E., Martin, S. R., Thorogate, R. & Pastore, A. The RNA-Recognition Motifs of TAR DNA-Binding Protein 43 May Play a Role in the Aberrant Self-Assembly of the Protein. *Front. Mol. Neurosci.* **11**, 372, (2018).
 14. Chou, C.-C. *et al.* TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD. *Nat. Neurosci.* **21**, 228–239, (2018).
 15. Ou, S. H., Wu, F., Harrich, D., García-Martínez, L. F. & Gaynor, R. B. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.* **69**, 3584–3596 (1995).
 16. Ayala, Y. M. *et al.* Structural determinants of the cellular localization and shuttling of TDP-43. *J. Cell. Sci.* **121**, 3778–3785, (2008).
 17. Mompeán, M. *et al.* Point mutations in the N-terminal domain of transactive response DNA-binding protein 43 kDa (TDP-43) compromise its stability, dimerization, and functions. *J. Biol. Chem.* **292**, 11992–12006, (2017).
 18. Wang, I.-F. *et al.* The self-interaction of native TDP-43 C terminus inhibits its degradation and contributes to early proteinopathies. *Nat. Commun.* **3**, 766, (2012).
 19. Kuo, P.-H., Doudeva, L. G., Wang, Y.-T., Shen, C.-K. J. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808, (2009).
 20. Tsoi, P. S. *et al.* The N-Terminal Domain of ALS-Linked TDP-43 Assembles without Misfolding. *Angew. Chem.* **129**, 12764–12767, (2017).
 21. Wang, A. *et al.* A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J.* **37**, (2018).

22. Budini, M., Romano, V., Quadri, Z., Buratti, E. & Baralle, F. E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum. Mol. Genet* **24**, 9–20, (2015).
23. Zhang, Y.-J. *et al.* The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum. Mol. Genet* **22**, 3112–3122, (2013).
24. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by α -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure (London, England : 1993)* **24**, 1537–1549, (2016).
25. Ihara, R. *et al.* RNA binding mediates neurotoxicity in the transgenic Drosophila model of TDP-43 proteinopathy. *Hum. Mol. Genet* **22**, 4474–4484, (2013).
26. Wang, Y.-T. *et al.* The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* **288**, 9049–9057, (2013).
27. Lu, Y., Lim, L. & Song, J. RRM domain of ALS/FTD-causing FUS characteristic of irreversible unfolding spontaneously self-assembles into amyloid fibrils. *Sci. Rep.* **7**, 1043, (2017).
28. Mompeán, M., Baralle, M., Buratti, E. & Laurents, D. V. An Amyloid-Like Pathological Conformation of TDP-43 Is Stabilized by Hypercooperative Hydrogen Bonds. *Front. Mol. Neurosci.* **9**, 125, (2016).
29. Zacco, E. *et al.* RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *J. Mol. Biol.* **431**, 1671–1688, (2019).
30. Huang, Y.-C. *et al.* Inhibition of TDP-43 aggregation by nucleic acid binding. *PloS one* **8**, e64002, (2013).
31. Chen, H.-J. *et al.* RRM adjacent TARDBP mutations disrupt RNA binding and enhance TDP-43 proteinopathy. *Brain* **142**, 3753–3770, (2019).
32. Dang, M. *et al.* ATP is a cryptic binder of TDP-43 RRM domains to enhance stability

- and inhibit ALS/AD-associated fibrillation. *Biochem. Biophys. Res. Commun.*, (2019).
33. Bhandare, V. V. & Ramaswamy, A. The proteinopathy of D169G and K263E mutants at the RNA Recognition Motif (RRM) domain of tar DNA-binding protein (tdp43) causing neurological disorders: A computational study. *J. Biomol. Struct. Dyn.* **36**, 1075–1093, (2018).
 34. Bozzo, F. *et al.* Structural insights into the multi-determinant aggregation of TDP-43 in motor neuron-like cells. *Neurobiol. Dis.* **94**, 63–72, (2016).
 35. Prakash, A., Kumar, V., Meena, N. K., Hassan, M. I. & Lynn, A. M. Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *J. Biomol. Struct. Dyn.* **37**, 178–194, (2019).
 36. Prakash, A., Kumar, V., Meena, N. K. & Lynn, A. M. Elucidation of the structural stability and dynamics of heterogeneous intermediate ensembles in unfolding pathway of the N-terminal domain of TDP-43. *RSC Adv.* **8**, 19835–19845, (2018).
 37. Mompeán, M. *et al.* “Structural characterization of the minimal segment of TDP-43 competent for aggregation”. *Arch. Biochem. Biophys.* **545**, 53–62, (2014).
 38. Garnier, C. *et al.* Zinc binding to RNA recognition motif of TDP-43 induces the formation of amyloid-like aggregates. *Sci. Rep.* **7**, 6812, (2017).
 39. Pierce, B. G. *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics (Oxford, England)* **30**, 1771–1773, (2014).
 40. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443–1449, (2013).
 41. Maris, C., Dominguez, C. & Allain, F. H.-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131, (2005).
 42. Shodai, A. *et al.* Aberrant assembly of RNA recognition motif 1 links to pathogenic

- conversion of TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* **288**, 14886–14905, (2013).
43. Takagi, S. *et al.* RNP2 of RNA recognition motif 1 plays a central role in the aberrant modification of TDP-43. *PloS one* **8**, e66966, (2013).
 44. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68, (2017).
 45. Liu, J., He, X. & Zhang, J. Z. H. Improving the scoring of protein-ligand binding affinity by including the effects of structural water and electronic polarization. *Int. J. Mol. Sci.* **53**, 1306–1314, (2013).
 46. Villa, A., Wöhnert, J. & Stock, G. Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch. *Nucleic acids res.* **37**, 4774–4786, (2009).
 47. Li, C., Zhu, Y., Wang, Y. & Chen, G. Molecular dynamics simulation on the conformational transition of the mad2 protein from the open to the closed state. *Int. J. Mol. Sci.* **15**, 5553–5569, (2014).
 48. Li, C. *et al.* DNA Structural Distortions Induced by a Monofunctional Trinuclear Platinum Complex with Various Cross-Links Using Molecular Dynamics Simulation. *J. Chem. Inf. Model* **60**, 1700–1708, (2020).
 49. Meng, H., Li, C., Wang, Y. & Chen, G. Molecular dynamics simulation of the allosteric regulation of eIF4A protein from the open to closed state, induced by ATP and RNA substrates. *PloS one* **9**, e86104, (2014).
 50. Chronopoulou, E. G., Vlachakis, D., Papageorgiou, A. C., Ataya, F. S. & Labrou, N. E. Structure-based design and application of an engineered glutathione transferase for the development of an optical biosensor for pesticides determination. *Biochim. Biophys. Acta Gen. Subj.* **1863**, 565–576; 10.1016/j.bbagen.2018.12.004 (2019).
 51. Healy, E. F., Little, C. & King, P. J. A model for small heat shock protein inhibition

- of polyglutamine aggregation. *Cell Biochem. Biophys.* **69**, 275–281; 10.1007/s12013-013-9795-1 (2014).
52. Tarus, B., Straub, J. E. & Thirumalai, D. Probing the initial stage of aggregation of the Aβ(10-35)-protein: assessing the propensity for peptide dimerization. *J. Mol. Biol.* **345**, 1141–1156, (2005).
53. Xue, T., Song, C., Wang, Q., Wang, Y. & Chen, G. Investigations of the CLOCK and BMAL1 Proteins Binding to DNA: A Molecular Dynamics Simulation Study. *PloS one* **11**, e0155105, (2016).
54. Song, C., Wang, Q., Xue, T., Wang, Y. & Chen, G. Molecular dynamics simulations on the conformational transitions from the GA 98 (GA 88) to GB 98 (GB 88) proteins. *J. Mol. Recognit.* **29**, 580–595, (2016).
55. Wang, E. *et al.* End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **119**, 9478–9508, (2019).

Figure legends

Figure 1. Compositions of TDP-43 protein. (a) The RRM12 domains are connected to NTD (in red) and CTD (in magenta) by the N-loop and C-loop, respectively. RRM12 domains contain the RRM1 (in purple) domain and RRM2 (in orange) domain that are connected by a α -linker. (b) The domain organization of the RRM12 domains are represented by the five β -strands (β 1, β 2, β 3, β 4, β 5) and two mutually vertical α -helices (α 1, α 2) of each RRM1 (in purple) domain and RRM2 (in orange) domain, respectively.

Figure 2. (a) RMSD of heavy atoms with respect to the initial structure of the RRM12 system. (b) The average structure of the RRM12 system. The RRM1 domain and the RRM2 domain are colored in purple and orange, respectively.

Figure 3. Average structures and the corresponding non-transparent surfaces of (a) dimer $S_{\alpha}S_{\beta}/D$, (b) $S_{\beta}S_{\beta}/D$, (c) $S_{\beta}H_{1\alpha}/D$, (d) $S_{\beta}H_{1\beta}/D$ and (e) H_1H_1/D systems, respectively. Two monomers in the five dimer models are colored in cyan and green, respectively.

Figure 4. The ionic interaction occupancies between two monomer RRM12 from the simulations of the dimer models of (a) $S_{\alpha}S_{\beta}/D$, (b) $S_{\beta}S_{\beta}/D$, (c) $S_{\beta}H_{1\alpha}/D$ and (d) $S_{\beta}H_{1\beta}/D$, respectively.

Figure 5. MM-PBSA energy decompositions between two monomer RRM12 representing ($\text{kcal}\cdot\text{mol}^{-1}$) into the residues of RRM12 domains for the dimer (a) $S_{\alpha}S_{\beta}/D$ and (b) H_1H_1/D models.

Figure 6. Average structures of the tetramer models of (a) $S_{\alpha}S_{\beta}1/T$ and (b) $S_{\alpha}S_{\beta}2/T$. The shoulder-to-shoulder, head-1-to-head-1 and head-2-to-head-2 interfaces were marked by dash lines.

Figure 7. The residues fluctuations for the (a) the RRM12 (black line), $S_{\alpha}S_{\beta}/D$ (red line) and $S_{\alpha}S_{\beta}1/T$ (blue line) models. (b) Dynamical cross-correlation maps calculated

from the first 10 ns of the $S_\alpha S_\beta 1/T$ model. The key subregions are marked by squared in black.

Figure 8. The recipe to construct the hexamer model of $S_\alpha S_\beta/H$ by superimposing two tetramer models of $S_\alpha S_\beta 1/T$ and $S_\alpha S_\beta 2/T$.

Figure 9. (a) The scheme to construct the periodic octamer model by superimposing the tetramer models of $S_\alpha S_\beta 1/T$ and $S_\alpha S_\beta 2/T$ in different orders. (b) The full equal structures of the two hexamer models constructed along two different overlapping directions. (c) The structure of the dodecamer model constructed using the similar constructing pathways.

Figure 10. Time-dependences of distances (a) between two β -sheets in two RRM1 domains for the dimer models of the $S_\alpha S_\beta/D$ (black), $S_\beta S_\beta/D$ (red), $S_\beta H_{1\alpha}/D$ (blue), $S_\beta H_{1\beta}/D$ (magenta), $H_1 H_1/D$ (green) and the tetramer models of the $S_\alpha S_\beta 1/T$ (purple) and $S_\alpha S_\beta 2/D$ (deep blue); (b) the distances of two $\alpha 1$ in both RRM1 domains and two $\alpha 2$ in both RRM1 domains in dimer model of $S_\alpha S_\beta/D$ (oranges) and the tetramer model of $S_\alpha S_\beta 1/T$ (light blue).

Figure 11. The aggregation axis existed in the large oligomer model. The inset represents that the parallel β -sheet layer orients perpendicularly to the aggregation axis.

Figure 12. (a) Average structure of the complete TDP-43. The average centroid distances between NTD and RRM12, between CTD and RRM12, and the average angle among the average centroid of NTD, RRM12 and CTD were labeled. (b) The full TDP-43 dimer structure constructed based on $S_\alpha S_\beta/D$ model. NTD: N-terminal domain, CTD: C-terminal domain.

Table 1. Components of the MM-PBSA free energies (kcal·mol⁻¹) for the S_αS_β/D, S_βS_β/D, S_βH_{1α}/D, S_βH_{1β}/D and H₁H₁/D dimer models

Model	S_αS_β/D	S_βS_β/D	S_βH_{1α}/D	S_βH_{1β}/D	H₁H₁/D
ΔE_{ele}	-399.04	-258.23	-274.65	-296.07	-142.78
ΔE_{vdw}	-127.30	-121.59	-142.82	-125.85	-63.07
ΔE_{int}	0.00	0.00	0.00	0.00	0.00
$\Delta G_{np/solv}$	-17.27	-17.58	-17.76	-17.59	-7.91
$\Delta G_{pb/solv}$	459.53	324.26	354.73	368.75	169.93
ΔG_{np}	-144.57	-139.17	-160.58	-143.44	-70.98
ΔG_{pb}	60.49	66.03	80.09	72.68	27.15
$\Delta H_{binding}$	-84.07	-73.14	-80.49	-70.76	-43.83

$$\Delta G_{np} = \Delta E_{vdw} + \Delta G_{np/solv}, \Delta G_{pb} = \Delta E_{ele} + \Delta G_{pb/solv}$$

$$\Delta H_{binding} = \Delta G_{np} + \Delta G_{pb} + \Delta E_{int}$$

Table2. Occupancies (%) of hydrogen bonds and hydrophobic interactions between different RRM12 monomer for the S_αS_β/D, S_βS_β/D, S_βH_{1α}/D, S_βH_{1β}/D and H₁H₁/D models

S_αS_β/D				Occupancy %			
Hydrogen bonds							
Asp 174	OD*...HZ*-NZ	176 Lys	95.70	Gln 182	HE21-NE2...O	144 Ser	57.28
Arg 165	OD*...HZ*-NZ	174 Asp	80.62	Ser 163	O...HE-NE	197 Arg	55.16
Cys 175	HH22-NH2...OD*	176 Lys	74.06	Gln 184	O...H*-N*	171 Arg	48.74
Met 162	O*...H*-NZ	176 Lys	70.82	Asp 174	O*...HZ*-NZ	145 Lys	43.72
Glu 156	O...HZ*-NZ	171 Arg	70.06	Asp 174	O*...HZ*-NZ	136 Lys	39.44
Glu 156	OE*...HH12-NH2	171 Arg	59.81	Hydrophobic interactions			
Arg 165	H-N...O	174 Asp	38.38	Met 162	C*...C*	147 Leu	99.84
Lys 263	OE*...HE-NE	263 Lys	34.36	Val 159	CG*...CG	178 Pro	93.96
Hydrophobic interactions				Phe 127	CZ...CD2	229 Phe	91.52
Val 159	CG*...CD1	109 Leu	92.42	Leu 177	CD1...CD*	109 Leu	89.10
Pro 262	CD...CB	262 Pro	82.82	Val 159	CG*...CD1	107 Ile	86.10
Pro 262	CD...CB	262 Pro	76.94	Met 162	C*...CD1	107 Ile	81.50
Leu 177	CD1...CD2	109 Leu	73.72	Trp 172	CZ3...CB	228 Ala	71.18
Val 159	CG1...CD	147 Leu	70.56	Leu 177	CB*...CE2*	147 Leu	70.82
Leu 177	CD1...C*	147 Leu	71.96	Val 159	CG1...CE2	149 Phe	64.00
Leu 177	CD*...CD1	147 Leu	69.82	Trp 155	CD...CB	178 Pro	32.92
His 264	CD2...CE	263 Lys	46.40	Met 167	CG...CD1	226 Phe	31.34
S_βS_β/D				S_βH_{1β}			
Occupancy %				Occupancy %			
Hydrogen bonds				Hydrogen bonds			
Glu 261	OE...HH*-OE**	165 Arg	97.90	Arg 165	HH11-NH1...OG	180 Ser	91.66
Trp 172	H-N...O	183 Ser	80.90	Arg 171	H*-N*...OE*	261 Glu	90.28
Trp 172	HE1-NE1...OD*	185 Asp	69.32	Arg 227	HZ*-NZ...OE*	261 Glu	90.02
Arg 171	HH21-NH2...OG	258 Ser	65.20	Arg 171	H*-N*...OE*	261 Glu	85.04
Arg 171	HH22-NH2...O	259 Asn	60.78	Arg 165	H*-N*...O*	258 Ser	68.80
Lys 140	HZ*-NZ...OE*	200 Glu	48.16	Hydrophobic interactions			
Hydrophobic interactions				Leu 109	CD1...CD1	229 Phe	96.42
Trp 113	CE3...CE2	229 Phe	99.66	Pro 187	CD...CD11	139 Leu	83.52
Phe 194	C*...C*	113 Trp	90.42	Val 159	CG2...CE1	147 Phe	76.30
Phe 147	CZ...CG	147 Leu	89.60	Trp 172	CB...C*	188 Leu	63.30
Phe 147	CE...CD2	109 Leu	75.80	Trp 172	CD...CD	194 Phe	61.30
His 143	CD2...CD2	229 Phe	71.52	H₁H₁/D			
S_βH_{1α}				Occupancy %			
Hydrogen bonds				Hydrogen bonds			
Asp 185	OD*...HH1-NH1	165 Arg	82.78	Met 167	O...HE21-NE2	164 Gln	87.90
Gln 164	OE1...HE2-NE2	256 His	64.90	Glu 122	O...HH11-NH1	165 Arg	80.10
Gln 182	HE22-NE2...O	111 Leu	63.22	Met 167	H-N...OE1	164 Gln	41.30
				Hydrophobic interactions			
				Trp 123	CD1...CZ3	172 Trp	99.20
				Ile 168	CD1...CD3	127 Phe	87.26

(a)

(b)

Figure 1

(a)

(b)

Figure 2

Figure 3

(a)

(b)

(c)

(d)

Figure 4

(a)

(b)

Figure 5

(a)

(b)

Figure 6

(a)

(b)

Figure 7

$S_\alpha S_\beta 1/T$	2MRR	1MRR	RRM1	RRM2		
	2MRR	1MRR	RRM1	RRM2		
$S_\alpha S_\beta 2/T$			RRM1	RRM2	RRM2	RRM1
			RRM1	RRM2	RRM2	RRM1

Figure 8

(a)

Figure 9

(a)

(b)

Figure 10

Aggregation axis

Figure 11

(a)

(b)

Figure 12

Figures

Figure 1

Compositions of TDP-43 protein. (a) The RRM12 domains are connected to NTD (in red) and CTD (in magenta) by the N-loop and C-loop, respectively. RRM12 domains contain the RRM1 (in purple) domain and RRM2 (in orange) domain that are connected by a α -linker. (b) The domain organization of the RRM12 domains are represented by the five β -strands ($\beta 1, \beta 2, \beta 3, \beta 4, \beta 5$) and two mutually vertical α -helices ($\alpha 1, \alpha 2$) of each RRM1 (in purple) domain and RRM2 (in orange) domain, respectively.

Figure 2

(a) RMSD of heavy atoms with respect to the initial structure of the RRM12 system. (b) The average structure of the RRM12 system. The RRM1 domain and the RRM2 domain are colored in purple and orange, respectively.

Figure 3

Average structures and the corresponding non-transparent surfaces of (a) dimer $S_{\alpha}S_{\beta}/D$, (b) $S_{\beta}S_{\beta}/D$, (c) $S_{\beta}H_{1\alpha}/D$, (d) $S_{\beta}H_{1\beta}/D$ and (e) H_1H_1/D systems, respectively. Two monomers in the five dimer models are colored in cyan and green, respectively.

Figure 4

The ionic interaction occupancies between two monomer RRM12 from the simulations of the dimer models of (a) S_aS_β/D , (b) $S_\beta S_\beta/D$, (c) $S_\beta H_{1\alpha}/D$ and (d) $S_\beta H_{1\beta}/D$, respectively.

Figure 5

MM-PBSA energy decompositions between two monomer RRM12 representing ($\text{kcal}\cdot\text{mol}^{-1}$) into the residues of RRM12 domains for the dimer (a) S_aS_β/D and (b) $H_{1\beta}H_{1\beta}/D$ models.

Figure 6

Average structures of the tetramer models of (a) $S_\alpha S_\beta 1/T$ and (b) $S_\alpha S_\beta 2/T$. The shoulder-to-shoulder, head-1-to-head-1 and head-2-to-head-2 interfaces were marked by dash lines.

Figure 7

The residues fluctuations for the (a) the RRM12 (black line), $S_\alpha S_\beta/D$ (red line) and $S_\alpha S_\beta 1/T$ (blue line) models. (b) Dynamical cross-correlation maps calculated from the first 10 ns of the $S_\alpha S_\beta 1/T$ model. The key subregions are marked by squared in black.

$S_{\alpha}S_{\beta}1/T$	2MRR	1MRR	RRM1	RRM2
	2MRR	1MRR	RRM1	RRM2
$S_{\alpha}S_{\beta}2/T$			RRM1	RRM2
			RRM1	RRM2

Figure 8

The recipe to construct the hexamer model of $S_{\alpha}S_{\beta}/H$ by superimposing two tetramer models of $S_{\alpha}S_{\beta}1/T$ and $S_{\alpha}S_{\beta}2/T$.

Figure 9

(a) The scheme to construct the periodic octamer model by superimposing the tetramer models of SaSβ1/T and SaSβ2/T in different orders. (b) The full equal structures of the two hexamer models constructed along two different overlapping directions. (c) The structure of the dodecamer model constructed using the similar constructing pathways.

Figure 10

Time-dependences of distances (a) between two β-sheets in two RRM1 domains for the dimer models of the SaSβ/D (black), SβSβ/D (red), SβH_{1α}/D (blue), SβH_{1β}/D (magenta), H₁H₁/D (green) and the tetramer models of the SaSβ1/T (purple) and SaSβ2/D (deep blue); (b) the distances of two α1 in both RRM1 domains and two α2 in both RRM1 domains in dimer model of SaSβ/D (oranges) and the tetramer model of SaSβ1/T (light blue).

Aggregation axis

Figure 11

The aggregation axis existed in the large oligomer model. The inset represents that the parallel β -sheet layer orients perpendicularly to the aggregation axis.

Figure 12

(a) Average structure of the complete TDP-43. The average centroid distances between NTD and RRM12, between CTD and RRM12, and the average angle among the average centroid of NTD, RRM12 and CTD were labeled. (b) The full TDP-43 dimer structure constructed based on SaS β /D model. NTD: N-terminal domain, CTD: C-terminal domain.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial20201207.docx](#)