**SUPPLEMENTARY MATERIAL**

**Supplementary Table 1. TDP-43 homodimer interactions critical for structure stability.** Listed here are the inter-monomer contacts that enhance dimer stability, measured from Molecular dynamics simulations and FoldX calculations. Column 1 (i.e. monomer 1) is the reference column. Column 2 (i.e. monomer 2) refers to interactions with monomer 1. Strong-interactions are indicated in bold.

### Dimer Type-1 (+180° angle)

| Monomer-1 | Monomer-2 |
|-----------|-----------|
| Seq/Residue | Seq/Residue |
| Arg6 | Arg83 |
| Val7 | Asp86 |
| Thr8 | Glu87 |
| Asp10 | Glu87, Thr88 |
| Asn12 | Thr88 |
| Asp13 | Glu87 |
| Cys39 | Thr88 |
| Leu41 | Val94, Lys95 |
| Arg44 | Arg98 |
| Arg83 | Glu9 |

### Dimer Type-2 (+45°)

| Monomer-1 | Monomer-2 |
|-----------|-----------|
| Seq/Residue | Seq/Residue |
| Ile5 | lesser |
| Thr8 | Asn81, Arg83 |
| Asp10 | lesser |
| Asn12 | Glu87 |
| Asp13 | Lys84, Ser91 |
| Glu17 | Lys97 |

### Dimer Type-3 (-135°)

| Monomer-1 | Monomer-2 |
|-----------|-----------|
| Seq/Residue | Seq/Residue |
| Arg6 | Tyr77 |
| Thr8 | Tyr73 |
| Asn12 | Cys39 |
The extreme N-terminal region of TDP-43 is required for TDP-43 biological activity. (A) Schematic representation of N-terminal GFP-tagged TDP-43 fragments co-transfected into HeLa cells with the CFTR mini-gene reporter construct. (B) Examination of CFTR exon 9 skipping by RT-PCR shows that, compared to exon 9-containing products in cells expressing GFP (top band), expression of GFP-TDP-43 significantly promotes exon 9 skipping (bottom band). In contrast, the N-terminal deleted TDP-43 fragments show a decreased ability to promote exon 9 exclusion. (C) Western blot analysis using anti-GFP antibody confirmed that the examined GFP-tagged TDP-43 products were expressed at similar levels (* represents a nonspecific band). (D) Similar to full-length TDP-43, the examined TDP-43 fragments localized to the nucleus (Scale bar 10 μM). (E) Like GFP-tagged N-terminal deleted TDP-43 fragments, untagged TDP-43 fragments lose their ability to promote exon 9 exclusion.
The formation of TDP-43 homodimers is independent of its DNA/RNA binding activity. (A) While low dose (100 µM) DSG treatment of cells stabilizes endogenous TDP-43 and DJ-1 homodimers without a perceivable reduction of their monomeric forms, high dose (1000 µM) DSG treatment results in a significant increase in TDP-43 and DJ-1 homodimers and a simultaneous reduction of TDP-43 and DJ-1 monomers. (B) Immunofluorescence studies show that endogenous TDP-43 remains in the nucleus after DSG treatment (Scale bar 10 µM). (C-D) Compared to the splicing regulatory activities of TDP-43WT, deletion of the N-terminus or C-terminus of TDP-43 results in TDP-43 loss-of-function, as assessed using the CFTR minigene reporter (C), even though all TDP-43 products remain in the nucleus (D). (E) Similar to TDP-43WT, TDP-435F-L, which has impaired DNA/RNA binding activity, forms homodimers in cells treated with DSG, suggesting that the formation of TDP-43 homodimers is independent of DNA/RNA binding.
Closed and open conformations of TDP-43 bound by nucleic acids. (A, B) The closed conformation of TDP-43 is very stable, as determined by energy analysis and molecular dynamics. Several amino acids (Phe147, Phe149 and Phe316) strongly interact to maintain the nucleic acids, like DNA and RNA, during binding. (C) The open conformation would allow for nucleic acids to bind into the crevice formed by RRM1 and RRM2. The closed conformation has tertiary and secondary structures very similar to the open conformation, but the loop domain between RRM2 and the C-terminus is altered significantly.
Structural models of the TDP-43 homodimer. Based on predictions from YASARA and PRIME on likely regions of dimerization, we derived several models of the TDP-43 homodimer to understand the functions of the N-terminal regions of TDP-43 in homodimer formation. These predicted homodimers were then subjected to a refinement protocol, and long equilibration molecular dynamics followed by fast timescale dynamics. FoldX was utilized to measure which dimerizations were most optimal. Three of the generated homodimer configurations had good binding energies on average. The overall orientation of these homodimer pairs can be described as having the N-terminus meet at a +45° (A), +180° (C), or -135° (D) angle, which have contact energies of -3.84 kcal/mol, -1.91 kcal/mol, and -1.34 kcal/mol, respectively. Since the homodimer structure with a +45° was the most stable, the key residues in each monomer that interact to stabilize this homodimer are shown (B). Residues Ile5, Thr8, Asp10, Asn12, Asp13, and Glu17 from monomer 1 (red) interact with Asn81, Arg83, Glu87, Lys84, Ser91, and Lys97 from monomer 2 (blue).
Supplementary Methods

Computer-assisted modeling of TDP-43 structures

For modeling studies, the 414 amino acid sequence of human TDP-43 was taken from the NCBI Reference Sequence: NP_031401.1. The protein has four conserved domains to be modeled: N-terminal domain (residues 1-103), RRM1 domain (residues 105-170), RRM2 domain (residues 193-240), and the C-terminal domain (residues 270-414). The TDP-43 sequence was aligned based on many alignments, with each domain modeled as a separate unit built-in composite.

The modeling was built as a hybrid model from consensus between the programs PRIME (Prime v3.0, Schrodinger, LLC, New York, NY) (1, 2), YASARA SSP/Homology/PSSM Method (3-8), and TASSER (9-11). The variable loops and gaps were filled using knowledge-based homology and knowledge-based potentials with YASARA, or ab initio approach of ORCHESTRAR (12). Each missing loop was modeled using the Loop Search module implemented in Sybyl 8.0 or with YASARA loop modeler (3-8, 13, 14). Only loops with the highest homology and lowest root mean square deviations were selected for the final models. The side chains and rotamers were adjusted with knowledge-based potentials, simulated annealing with explicit solvent, and small equilibration simulations using YASARA’s refinement protocol and verified by WHAT-IF and PROCHECK (15). Fragments were divided into overlapping groups as follows: Fragment 1 (N-term domain) was built as two models consisting of residues 1 to 93 and 1 to 101, respectively. Fragment 2 (RRM domains) was built as two models consisting of residues 101 to 276 and residues 94 to 274, respectively. Fragment 3 (C-term domain) was built as two models consisting of residues 274 to 414 and 276 to 414, respectively. Combined fragments were overlaid using in-house superposition algorithms to determine optimal overlay and energies, which left the extraneous overlaid residues to be discarded as unnecessary. Fragment 1 results: both Yasara and Prime models were used to make a hybrid model of superior Z-scoring. Fragment 2 results: in this scenario YASARA model generated the best overall Z-score and was retained as the final model for this region. Fragment 3 results: YASARA model and Prime model performed similar in respective output and combined hybrid was not superior, thus each model was...
kept for possible conformations to consider of the C-terminal region. Finally
TASSER was considered for each fragment and the full-length protein. Full-
length TASSER models never generated ‘good’ Z-score, however individual
fragments gave reasonable structures retained for conformational comparison.

Refinement of the fragments was completed using YASARA’s refinement
module. These refinements started with the Secondary Structure Prediction
(SSP) feature of YASARA (3-6). Both homology and fold recognition were
considered and a final refinement with the entire model was completed using
YASARA for 250 ps of molecular dynamics (MD) using knowledge-based force
fields. Additionally, YASARA supports an extensive and large loop library for
modeling loops and gaps. The sequence and identity of each fragment was
reasonable. Impressively, the YASARA methods do seem to compare with the
results they demonstrated in the annual CASP8 and CASP9 protein-structure
prediction effort (3). The superposition and subsequent refinement of the
overlapping regions yields a complete model with two interesting conformations
for TDP-43: open and closed. We find that the closed conformation of TDP-43
has a few mini-gaps in the C-term domain that required additional modeling for
G298, N319, P320, L330, N355, and M359, which required Prime Side Chain
modeling to optimize the position and dihedrals. The final C-term domain was
subjected to energy optimization with PR conjugate gradient with an R-
dependent dielectric.

Both the open and closed conformations were verified with WHAT-IF and
PROCHECK (REF) and have valid conformations consistent with good phi-psi
space. Also, YASARA implements the latest features for modeling, such as,
packing (1D and 3D), dihedrals, PhiPsi, Backbone, and Packing3 (WHAT-IF
implementation), plus force field considerations. Atom consistency was checked
for all 414 amino acids, verifying correctness of chain name, dihedrals, angles,
torsions, non-bonds, electrostatics, atom-typing, and parameters. Model
manipulation was done with Maestro (Macromodel, version 9.8, Schrodinger,
LLC, New York, NY, 2010) , or Visual Molecular Dynamics (VMD) (16).

MD was completed on each model for conformational sampling, using methods
previously described in the literature (17-20). Briefly, each TDP-43 system
(monomer or homodimer) was minimized with relaxed restraints using either
Steepest Descent or Conjugate Gradient PR, and equilibrated in solvent with
physiological salt conditions (17-20). Following equilibration, each system was allowed to run MD calculations from 25 to 100 nanoseconds in length. The primary purpose of MD for this study was conformational variability that may occur with different dimerization pairs.

Protocol for refinement:
(1) Simulated annealing with explicit water molecules and ions
(2) Energy minimization
(3) MD for 500 ps to relax to the force field (both AMBER03 and YASARA2 were tested)

MD studies were also completed on all models for various conditions, time-lengths (100s nanoseconds), and opening/closure/structural stability. Those findings did lead to indication of important sites for mutants to test for domain stability. PRIME and YASARA also gave output for likely dimerizations. FoldX was utilized as a plugin within YASARA to achieve protein-protein comparisons (21-23). In summary, the FoldX algorithm calculates protein-protein interactions, or protein-DNA interactions, whereby FoldX calculates $\Delta \Delta G$ of interaction: $\Delta \Delta G_{ab} = \Delta G_{ab} - (\Delta G_a + \Delta G_b) + \Delta G_{kon} + \Delta S_{sc}$. Here, $\Delta G_{kon}$ reflects the effect of electrostatic interactions on the kon and $\Delta S_{sc}$ is the loss of translational and rotational entropy upon making the complex.

Mutants and Deletions: Both models and simulations for the closed conformation of TDP-43 with mutations at the N-terminus were completed. Residues 6-9 were mutated to Gly to abolish activity with the inter-domain pairing to Phe35. A model with residues 2 to 9 deleted was also created for the 414 amino acids model, resulting in a deletion construct for simulation.

REFERENCE

1 Jacobson, M.P., Pincus, D.L., Rapp, C.S., Day, T.J., Honig, B., Shaw, D.E. and Friesner, R.A. (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins*, 55, 351-367.

2 Jacobson, M.P., Friesner, R.A., Xiang, Z. and Honig, B. (2002) On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol*, 320, 597-608.
3 Krieger, E., Joo, K., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. and Karplus, K. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins, 77 Suppl 9, 114-122.
4 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25, 3389-3402.
5 Hooft, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. Nature, 381, 272.
6 Hooft, R.W., Sander, C., Scharf, M. and Vriend, G. (1996) The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. Comput Appl Biosci, 12, 525-529.
7 King, R.D. and Sternberg, M.J. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Sci, 5, 2298-2310.
8 Qiu, J. and Elber, R. (2006) SSALN: an alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs. Proteins, 62, 881-891.
9 Zhou, H. and Skolnick, J. (2007) Ab initio protein structure prediction using chunk-TASSER. Biophys J, 93, 1510-1518.
10 Zhou, H. and Skolnick, J. (2009) Protein structure prediction by pro-Sp3TASSER. Biophys J, 96, 2119-2127.
11 Zhou, H., Pandit, S.B. and Skolnick, J. (2009) Performance of the ProSp3-TASSER server in CASP8. Proteins, 77 Suppl 9, 123-127.
12 Dolan, M.A., Keil, M. and Baker, D.S. (2008) Comparison of Composer and ORCHESTRAR. Proteins, 72, 1243-1258.
13 Muckstein, U., Hofacker, I.L. and Stadler, P.F. (2002) Stochastic pairwise alignments. Bioinformatics, 18 Suppl 2, S153-160.
14 Canutescu, A.A., Shelenkov, A.A. and Dunbrack, R.L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci, 12, 2001-2014.
15 Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. Journal of Applied Crystallography, 26, 283-291.
16 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. J Mol Graph, 14, 33-38, 27-38.
17 Caulfield, T. and Devkota, B. (2012) Motion of transfer RNA from the A/T state into the A-site using docking and simulations. *Proteins*.

18 Caulfield, T. and Medina-Franco, J.L. (2011) Molecular dynamics simulations of human DNA methyltransferase 3B with selective inhibitor nanaomycin A. *J Struct Biol*, 176, 185-191.

19 Caulfield, T.R., Devkota, B. and Rollins, G.C. (2011) Examinations of tRNA Range of Motion Using Simulations of Cryo-EM Microscopy and X-Ray Data. *J Biophys*, 2011, 219515.

20 Caulfield, T.R. (2011) Inter-ring rotation of apolipoprotein A-I protein monomers for the double-belt model using biased molecular dynamics. *J Mol Graph Model*, 29, 1006-1014.

21 Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res*, 33, W382-388.

22 Schymkowitz, J.W., Rousseau, F., Martins, I.C., Ferkinghoff-Borg, J., Stricher, F. and Serrano, L. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc Natl Acad Sci U S A*, 102, 10147-10152.

23 Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J. and Rousseau, F. (2011) A graphical interface for the FoldX forcefield. *Bioinformatics*, 27, 1711-1712.