Supporting Information for "Refining all-atom protein force fields for polar-rich, prion-like, low complexity intrinsically disordered proteins"

Wai Shing Tang,[†] Nicolas L. Fawzi,[¶] and Jeetain Mittal[‡,*]

[†]*Department of Physics, Brown University, Providence, RI 02912, USA*

[¶]*Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI 02912, USA*

[‡]*Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, PA 18015, USA*

[*]E-mail: jem309@lehigh.edu

Tables

Table S1: Protein sequences used in simulations

Protein	N	Sequence
FUS_{0-43}*	44	GMASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQ
$TDP43_{310-350}$	41	GMNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQSGPS
RNA Pol $II_{1927-1970}$	44	SPTYSPTSPKGSTYSPTSPGYSPTSPTYSLTSPAISPDDSDEEN
FUS_{41-84}	44	YSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGYGSTGGYGSS
FUS_{77-120}	44	STGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGSYGSSSQSS
$FUS_{120-163}$	44	SSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNS
FUS_{37-97}	61	SYSGYSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGQQ SSY
$hnRNPA2_{190-233}$	44	GRGGNFGFGDSRGGGGNFGPGPGSNFRGGSDGYGSGRGFGDGYN
$hnRNPA2_{265-308}$	44	GNQGGGYGGGYDNYGGGNYGSGNYNDFGNYNQQPSNYGPMKSGN

*a starting glycine is included in FUS_{0-43} as residue 0

Table S2: Summary of simulations performed using ff99SBws, where N is the number of residues in given protein or peptide.

Protein	N	Enhanced Sampling	Total simulation time (ns/replica)	Simulation time analyzed (ns/replica)
$TDP43_{310-350}$	41	PT-WTE	200	150
FUS_{0-43}	44	PT-WTE	200	150
FUS_{41-84}	44	PT-WTE	200	150
FUS_{77-120}	44	PT-WTE	200	150
$FUS_{120-163}$	44	PT-WTE	200	150
FUS_{37-97}	61	PT-WTE	200	150
$hnRNPA2_{190-234}$	44	PT-WTE	200	150
$hnRNPA2_{256-308}$	44	PT-WTE	200	150
RNA Pol $II_{1927-1970}$	44	PT-WTE	200	150

Table S3: Summary of simulations performed using ff03ws, where N is the number of residues in given protein or peptide.

Protein	N	Enhanced Sampling	Total simulation time* (ns/replica)	Simulation time analyzed* (ns/replica)
$TDP43_{310-350}$	41	PT-WTE	200	150
FUS_{0-43}	44	PT-WTE	200	150
FUS_{41-84}	44	PT-WTE	200	150
FUS_{77-120}	44	PT-WTE	200	150
$FUS_{120-163}$	44	PT-WTE	200	150
$hnRNPA2_{190-234}$	44	PT-WTE	200	150
$hnRNPA2_{256-308}$	44	PT-WTE	200	150
RNA Pol $II_{1927-1970}$	44	PT-WTE	200	150

Table S4: Summary of simulations performed using modified force fields. All simulations were performed for a total of 200 ns per replica, and analysis conducted on the last 150 ns.

	k_{ψ} (kJ/mol)	$k_{\psi,S}$ (kJ/mol)	$k_{\psi,T}$ (kJ/mol)	$k_{\psi,Q}$ (kJ/mol)	Simulated peptides
ff99SBws	2.00	2.00	2.00	2.00	FUS_{0-43}, $TDP\text{-}43_{310-350}$, RNA Pol $II_{1927-1970}$
ff99SBws-S	2.00	1.00	2.00	2.00	FUS_{0-43}
ff99SBws-T	2.00	2.00	1.00	2.00	FUS_{0-43}
ff99SBws-ST	2.00	1.00	1.00	2.00	FUS_{0-43}, $TDP\text{-}43_{310-350}$, RNA Pol $II_{1927-1970}$
ff99SBws-TQ	2.00	2.00	1.00	1.00	FUS_{0-43}, $TDP\text{-}43_{310-350}$, RNA Pol $II_{1927-1970}$
ff99SBws-STQ	2.00	1.00	1.00	1.00	FUS_{0-43}, $TDP\text{-}43_{310-350}$, RNA Pol $II_{1927-1970}$

Table S5: Root-mean-squared deviation (RMSD) of simulated $\Delta\delta C_\alpha$, $\Delta\delta C_\beta$, $\Delta\delta C_\alpha - \Delta\delta C_\beta$ (in ppm) from NMR ^{13}C chemical shifts of three IDPs for ff99SBws, modified ff99SBws variants (S,T,ST,TQ,STQ) (this study), and ff99SB-disp.

$\Delta\delta C_\alpha - \Delta\delta C_\beta$	FUS_{0-43}	$TDP\text{-}43_{310-350}$	RNA Pol $II_{1927-1970}$
ff99SBws	0.90 ± 0.11	0.64 ± 0.09	0.68 ± 0.11
ff99SB-disp	0.90 ± 0.11	0.78 ± 0.07	0.74 ± 0.11
ff99SBws-S	0.86 ± 0.10		
ff99SBws-T	0.98 ± 0.13		
ff99SBws-ST	0.64 ± 0.07	0.71 ± 0.08	0.74 ± 0.08
ff99SBws-TQ	0.62 ± 0.08	0.66 ± 0.06	0.71 ± 0.10
ff99SBws-STQ	0.54 ± 0.07	0.51 ± 0.04	0.74 ± 0.08

$\Delta\delta C_\alpha$	FUS_{0-43}	$TDP\text{-}43_{310-350}$	RNA Pol $II_{1927-1970}$
ff99SBws	0.58 ± 0.07	0.43 ± 0.06	0.67 ± 0.11
ff99SB-disp	0.59 ± 0.07	0.54 ± 0.05	0.62 ± 0.09
ff99SBws-S	0.51 ± 0.06		
ff99SBws-T	0.61 ± 0.08		
ff99SBws-ST	0.41 ± 0.04	0.41 ± 0.04	0.62 ± 0.08
ff99SBws-TQ	0.36 ± 0.05	0.40 ± 0.04	0.64 ± 0.09
ff99SBws-STQ	0.37 ± 0.04	0.36 ± 0.03	0.62 ± 0.08

$\Delta\delta C_\beta$	FUS_{0-43}	$TDP\text{-}43_{310-350}$	RNA Pol $II_{1927-1970}$
ff99SBws	0.43 ± 0.05	0.42 ± 0.05	0.51 ± 0.06
ff99SB-disp	0.39 ± 0.05	0.39 ± 0.06	0.44 ± 0.11
ff99SBws-S	0.45 ± 0.06		
ff99SBws-T	0.44 ± 0.05		
ff99SBws-ST	0.37 ± 0.04	0.44 ± 0.05	0.47 ± 0.10
ff99SBws-TQ	0.40 ± 0.05	0.41 ± 0.06	0.42 ± 0.11
ff99SBws-STQ	0.34 ± 0.04	0.34 ± 0.06	0.47 ± 0.10

Protein sequences used in NMR experiments

FUS Low Complexity domain FUS_{1-163} (BMRB 26672)[1]

MASNDYTQQATQSYGAYPTQ PGQGYSQQSSQPYGQQSYSG YSQSTDTSGYGQSSYSSYGQ
SQNTGYGTQSTPQGYGSTGG YGSSQSSQSSYGQQSSYPGY GQQPAPSSTSGSYGSSSQSS
SYGQPQSGSYSQQPSYGGQQ QSYGQQQSYNPPQGYGQQNQ YNS
(163 residues)

TDP-43 Wild Type Low Complexity C-terminal domain $TDP43_{267-414}$ (BMRB 26823)[2]

*GHM*NRQLERSGRFGGNPGGF GNQGGFGNSRGGGAGLGNNQ GSNMGGGMNFGAFSINPAMM
AAAQAALQSSWGMMGMLASQ QNQSGPSGNNQNQGNMQREP NQAFGSGNNSYSGSNSGAAI
GWGSASNAGSGSGFNGGFGS SMDSKSSGWGM
(151 residues)

RNA Pol II C-terminal Domain 27-52 RNA Pol $II_{1770-1970}$ (BMRB 27063)[3]

*GHMS*PNYTPTSPNYSPTSPS YSPTSPSYSPTSPSYSPSSP RYTPQSPTYTPSSPSYSPSS
PSYSPASPKYTPTSPSYSPS SPEYTPTSPKYSPTSPKYSP TSPKYSPTSPTYSPTTPKYS
PTSPTYSPTSPVYTPTSPKY SPTSPTYSPTSPKYSPTSPT YSPTSPKGSTYSPTSPGYSP
TSPTYSLTSPAISPDDSDEE N
(201 residues)

hnRNPA2 low complexity domain 190-341 $hnRNPA2_{190-341}$ (BMRB 27123)[4]

*GHM*GRGGNFGFGDSRGGGGN FGPGPGSNFRGGSDGYGSGR GFGDGYNGYGGGPGGGNFGG
SPGYGGGRGGYGGGGPGYGN QGGGYGGGYDNYGGGNYGSG NYNDFGNYNQQPSNYGPMKS
GNFGGSRNMGGPYGGGNYGP GGSGGSGGYGGRSRY
(155 residues)

The residues in *italics* correspond to TEV-cleavage and cloning “scar” residues not present in the native sequences.

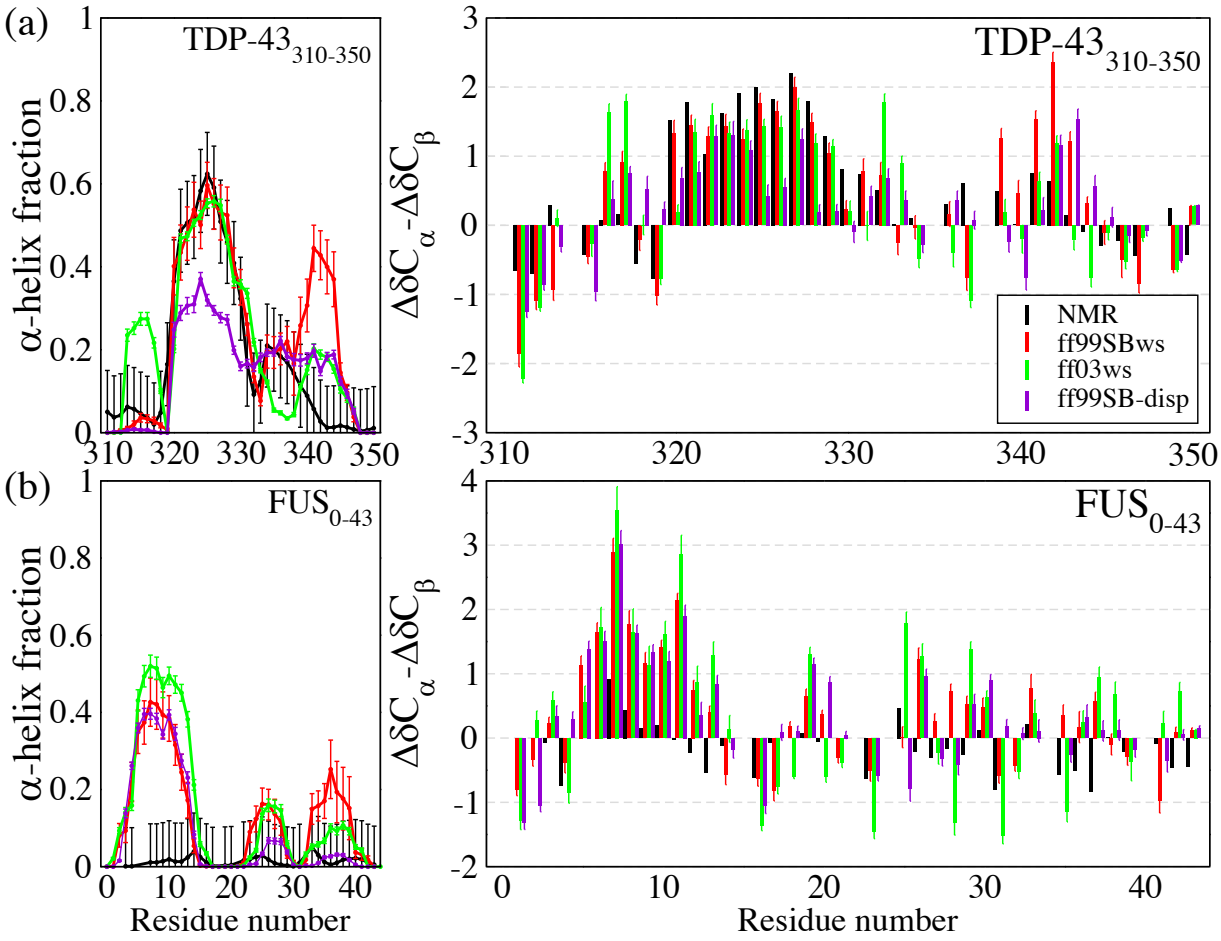

Figure S1. ***(a)*** α-helix propensities and ***(b)*** Chemical shifts $\Delta\delta C_\alpha - \Delta\delta C_\beta$ of TDP-$43_{310-350}$ and FUS_{0-43} using all-atom force fields ff99SBws, ff03ws, and ff99SB-disp, compared to experimental results. Chemical shifts are calculated from protein structure using SPARTA+ algorithm. Secondary structures are assigned by DSSP. Experimental values of secondary structure propensities are calculated from NMR chemical shifts using the δ2D program.

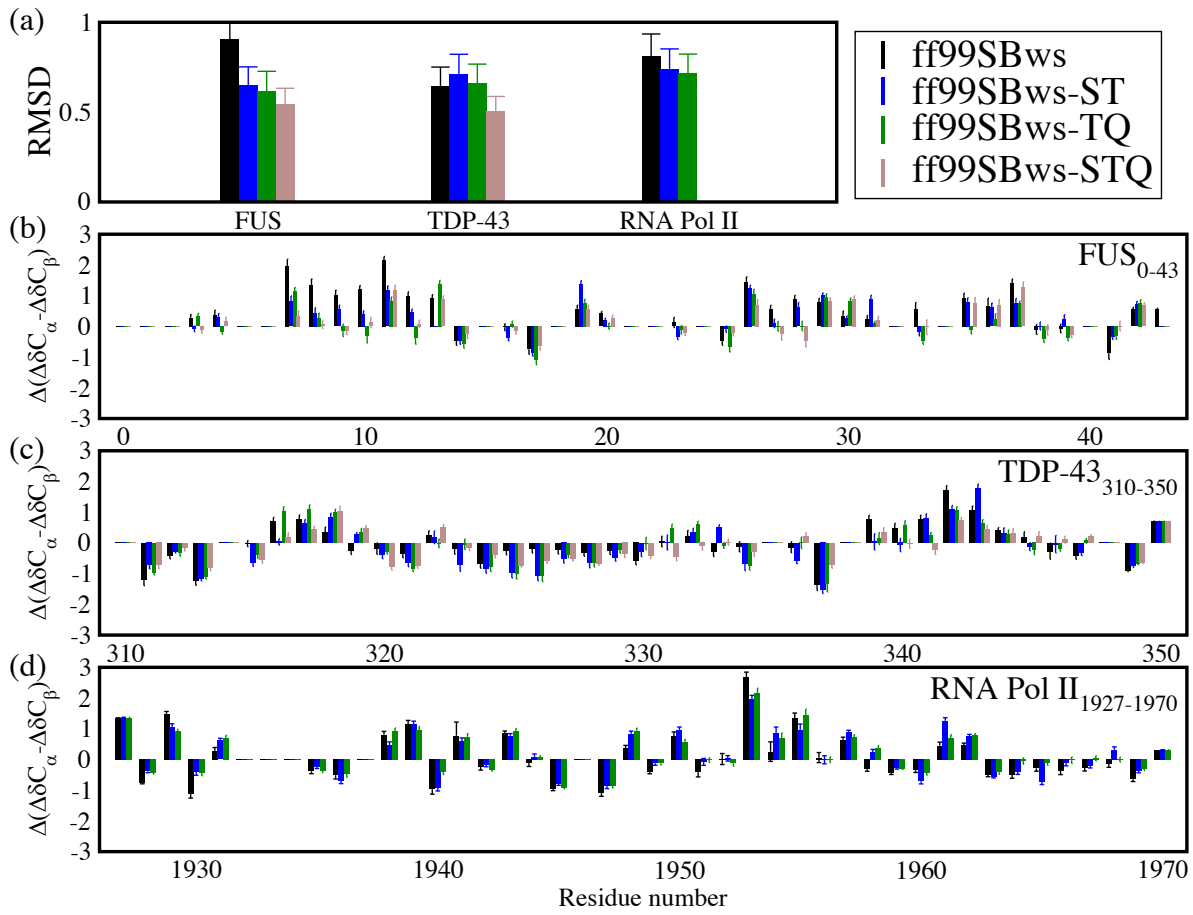

Figure S2. ***(a)*** Comparison of Root-mean-squared deviation (RMSD) of simulated $\delta C_\alpha - \delta C_\beta$ (in ppm) from NMR ^{13}C chemical shifts (RMSD) of simulated sequences and ***(b-d)*** chemical shift deviation from the experiment $\Delta(\Delta\delta C_\alpha - \Delta\delta C_\beta)$ with ff99SBws and modified force fields (see legend).

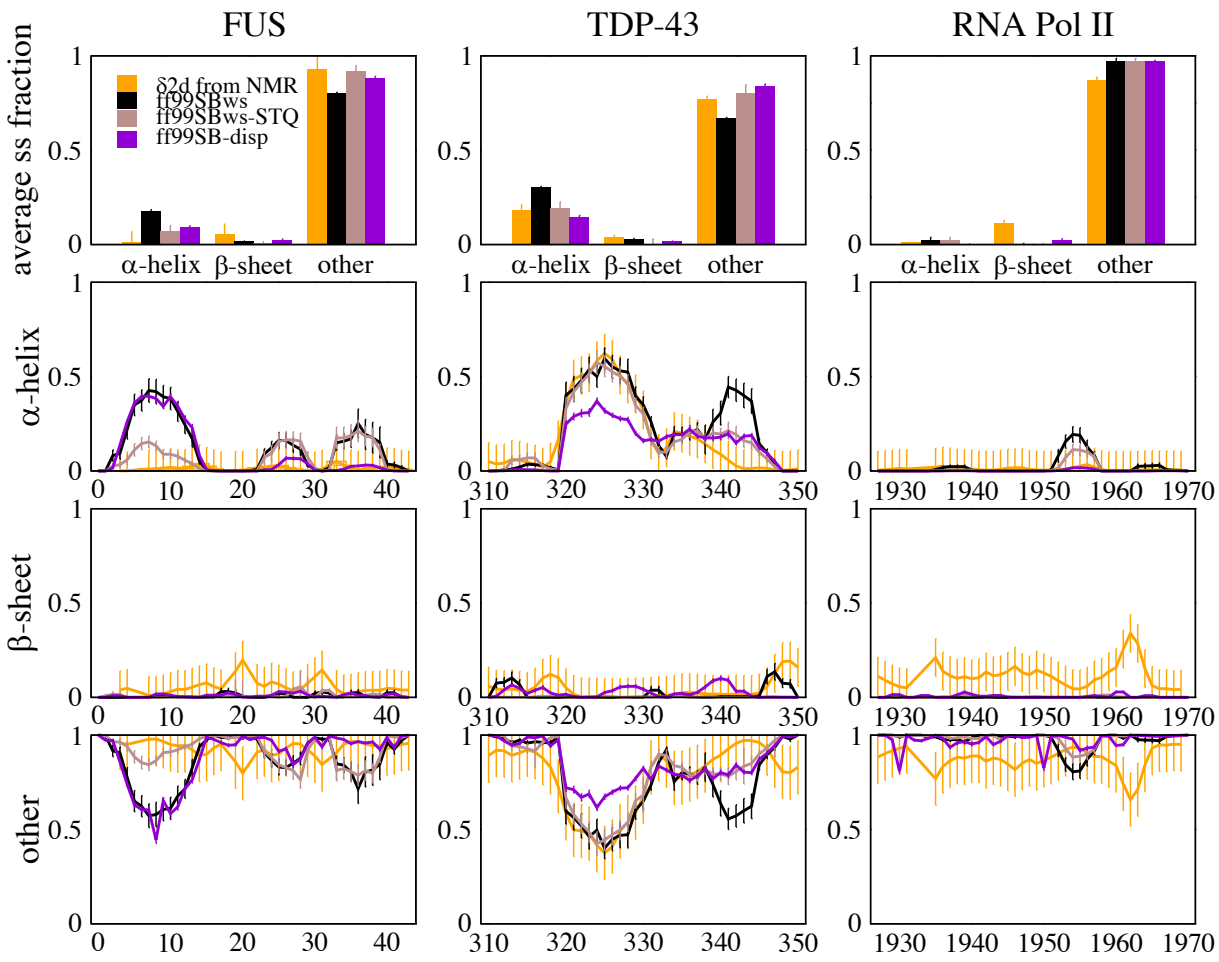

Figure S3. Secondary structure propensities of FUS_{0-43}, $TDP\text{-}43_{310-350}$, and $RNA\ Pol\ II_{1927-1970}$ with ff99SBws, ff99SBws-STQ (this study) and ff99SB-disp, compared to *δ*2D secondary structure propensities calculated from experimental NMR chemical shifts.

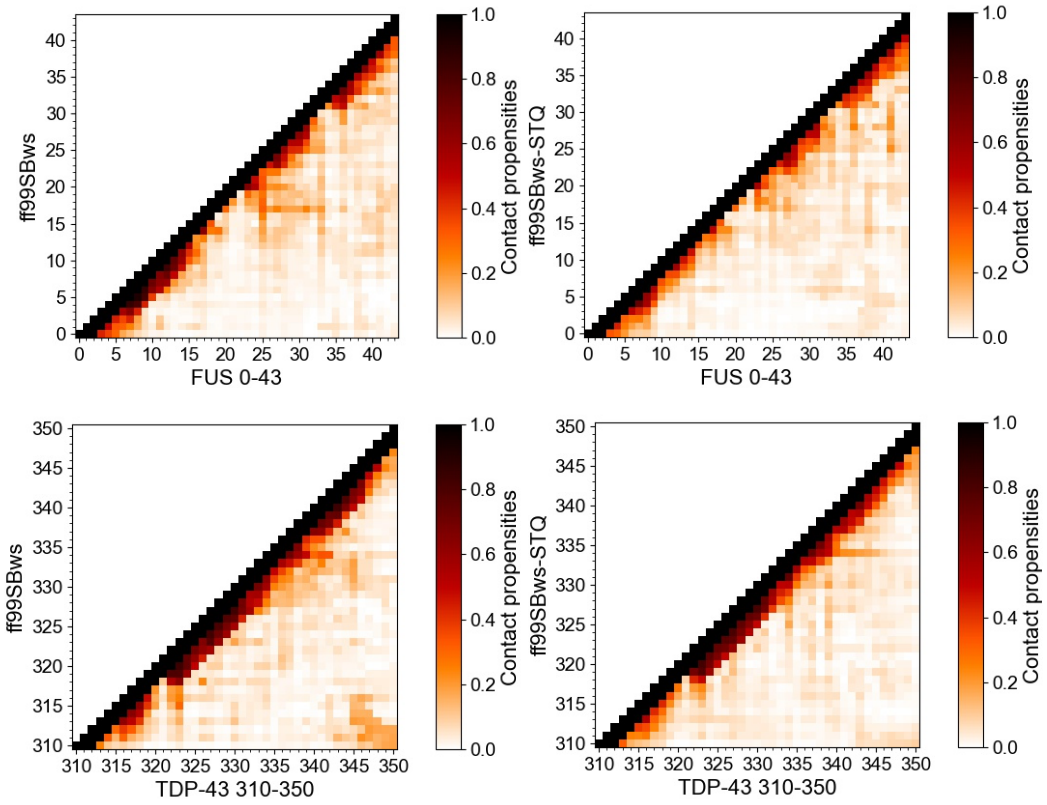

Figure S4. Intrachain residue-to-residue contact propensities in $FUS_{0\text{-}43}$ and $TDP\text{-}43_{310\text{-}350}$ with ff99SBws and ff99SBws-STQ (this study). Two residues are considered in contact when any heavy atoms from each of the residue are within 6 Angstrom in simulation. Contact propensities are the proportion of frames in simulation where two residues are in contact, with contact propensity = 1 being always in contact.

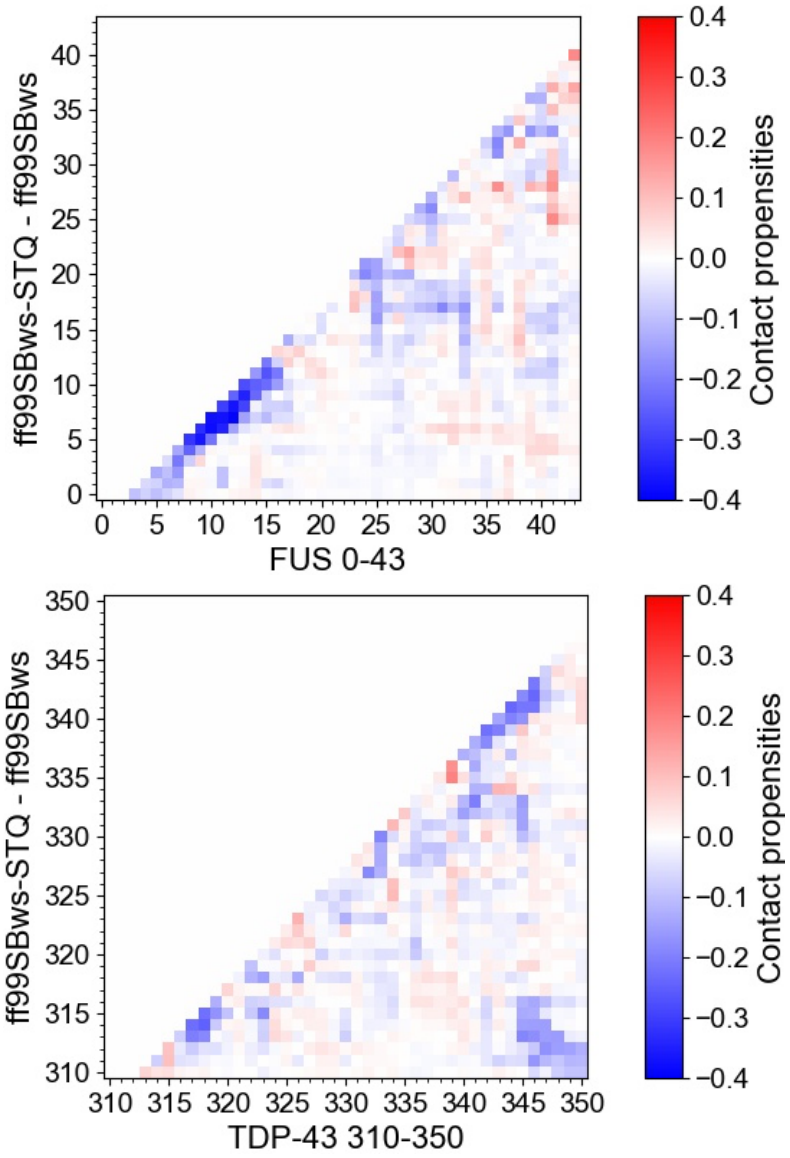

Figure S5. Change in Intrachain residue-to-residue contact propensities in FUS_{0-43} and $TDP\text{-}43_{310-350}$ from ff99SBws to ff99SBws-STQ (this study) Two residues are considered in contact when any heavy atoms from each of the residue are within 6 Angstrom in simulation. Contact propensities are the proportion of frames in simulation where two residues are in contact, with contact propensity = 1 being always in contact.

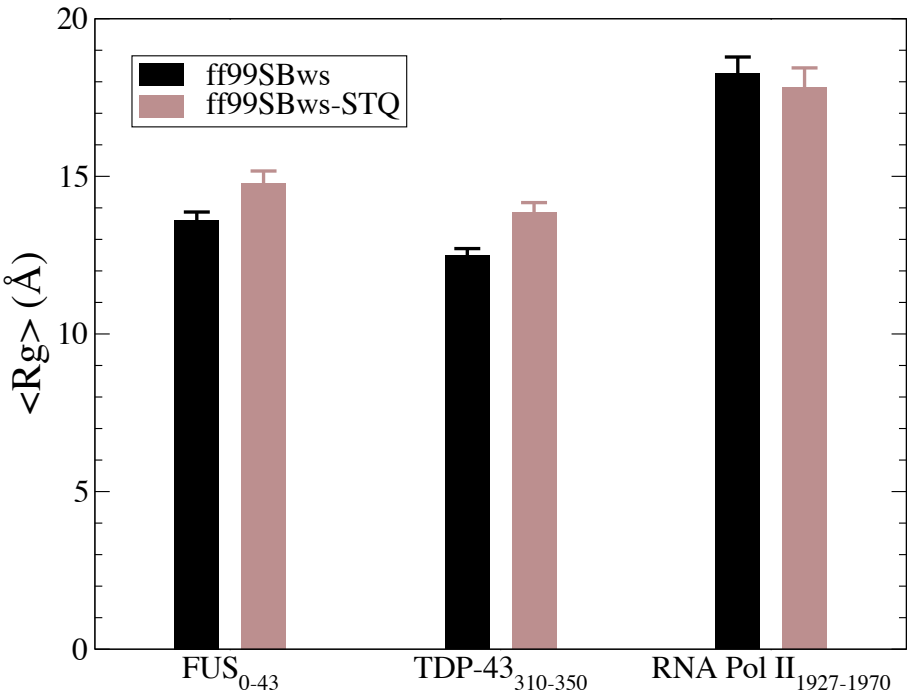

Figure S6. Average radius of gyration (Rg) in FUS_{0-43}, $TDP\text{-}43_{310-350}$ and RNA Pol $II_{1927-1970}$ with ff99SBws and ff99SBws-STQ (this study).

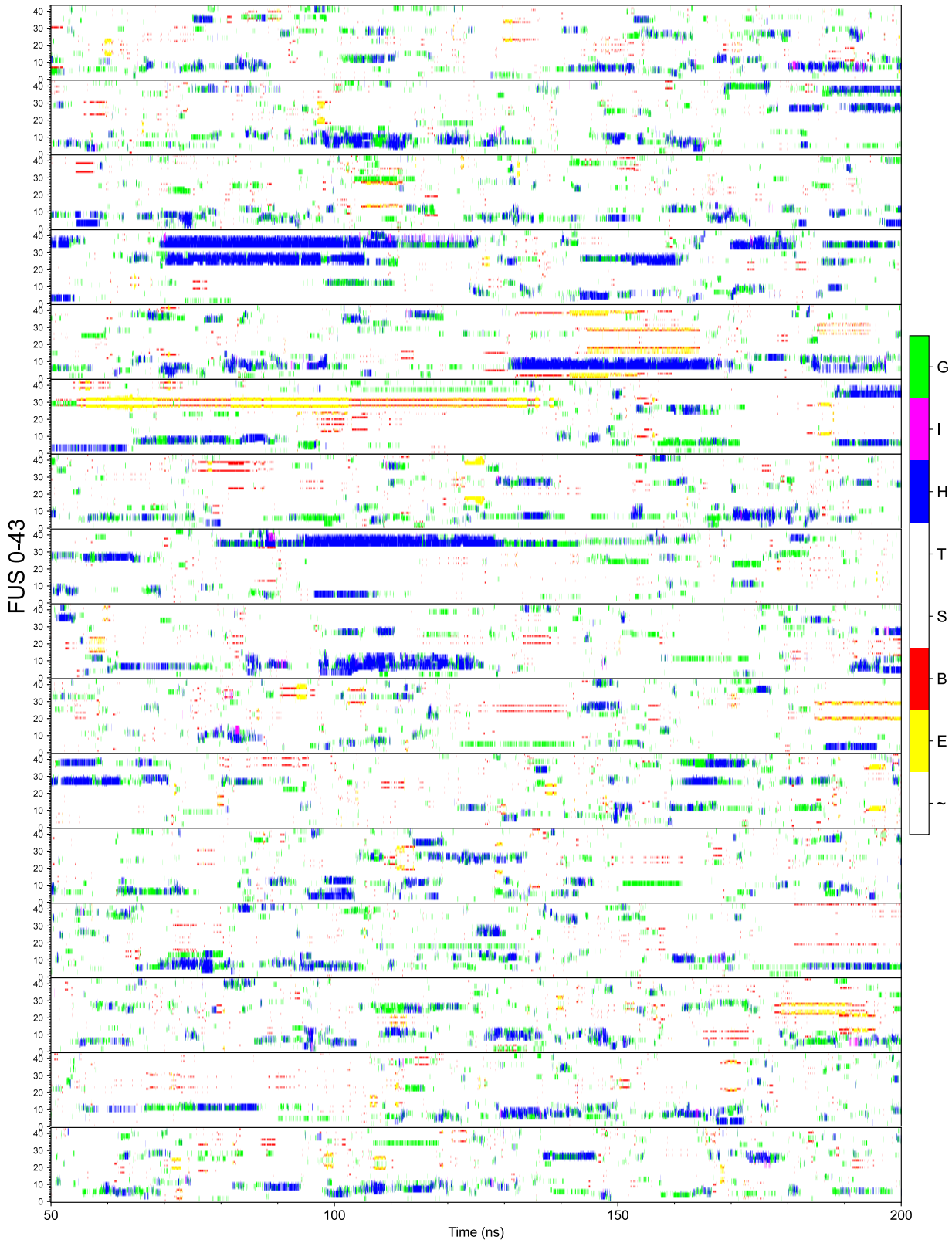

Figure S7. DSSP secondary structure as a function of simulation time for each demultiplexed replica moving through the temperature space of $FUS_{0\text{-}43}$ with ff99SBws-STQ (this study).

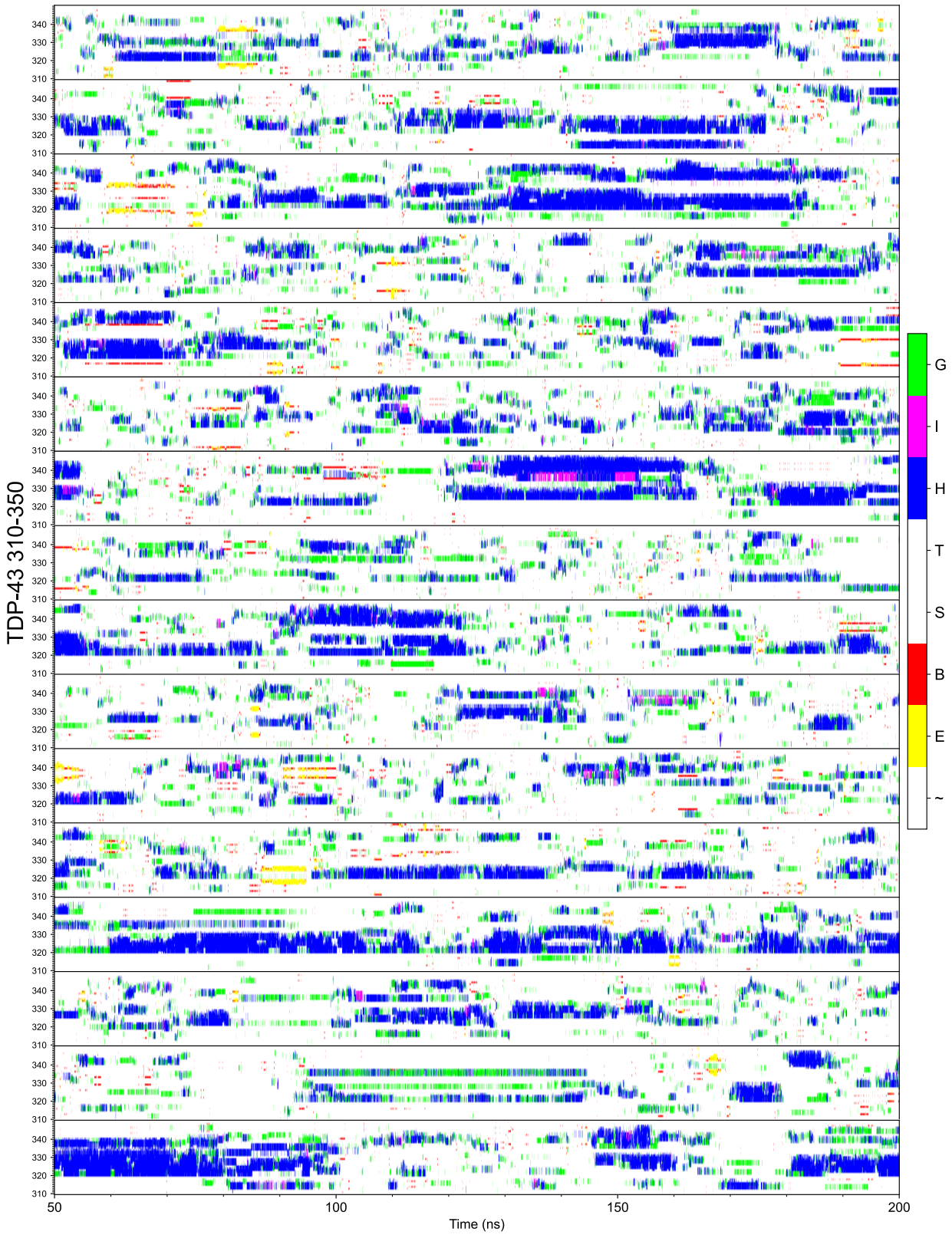

Figure S8. DSSP secondary structure as a function of simulation time for each demultiplexed replica moving through the temperature space of TDP-43$_{310\text{-}350}$ with ff99SBws-STQ (this study).

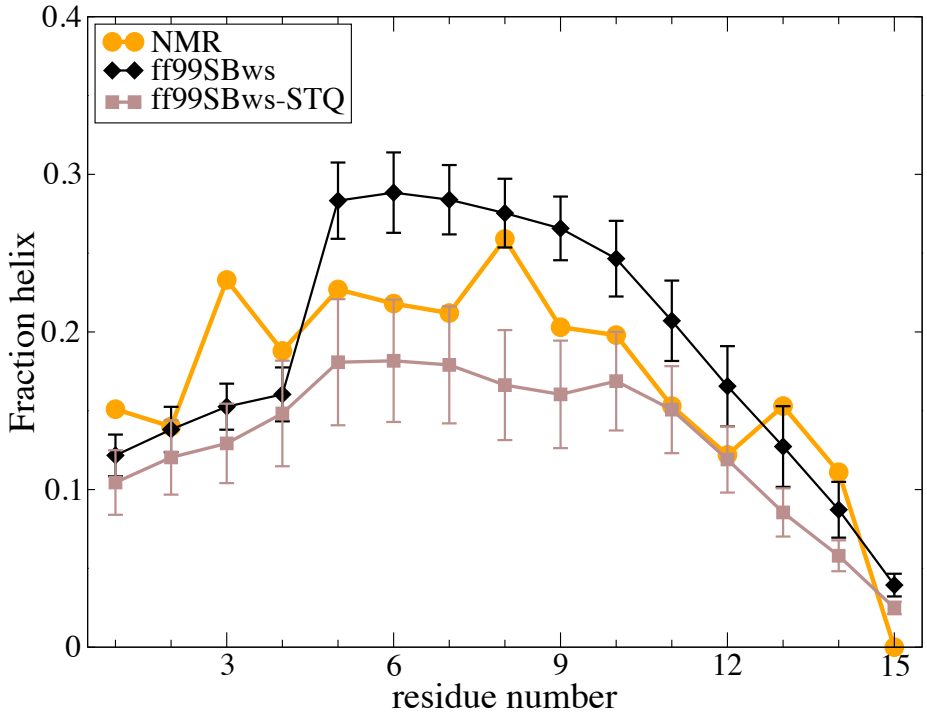

Figure S9. Fraction Helix of $(AAQAA)_3$ with ff99SBws and ff99SBws-STQ (this study) compared to NMR experimental data[5].

References

1. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **60**, 231–241 (2015).

2. 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* **24**, 1537–1549 (2016).

3. Janke, A. M. *et al.* Lysines in the RNA Polymerase II C-Terminal Domain Contribute to TAF15 Fibril Recruitment. *Biochemistry* **57**, 2549–2563 (2018).

4. Ryan, V. H. *et al.* Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol. Cell* **69**, 465-479.e7 (2018).

5. Shalongo, W., Dugad, L. & Stellwagen, E. Distribution of Helicity within the Model Peptide Acetyl(AAQAA)3amide. *J. Am. Chem. Soc.* **116**, 8288–8293 (1994).