Published in final edited form as:

Biochemistry. 2018 May 01; 57(17): 2549–2563. doi:10.1021/acs.biochem.7b00310.

Lysines in RNA polymerase II C-terminal domain contribute to TAF15 fibril recruitment

Abigail M. Janke¹, Da Hee Seo¹, Vahid Rahmanian², Alexander E. Conicella³, Kaylee L. Mathews³, Kathleen A. Burke¹, Jeetain Mittal², and Nicolas L. Fawzi^{*,1,3}

¹Department of Molecular Pharmacology, Physiology, and Biotechnology, Brown University, Providence, RI 02921, USA

²Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, PA 18015, USA

³Graduate Program in Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA

Abstract

Many cancer-causing chromosomal translocations result in transactivating protein products encoding FET family (FUS, EWSR1, TAF15) low complexity (LC) domains fused to a DNA-binding domain from one of several transcription factors. Recent work demonstrates that higher order assemblies of FET LC domains bind the carboxy-terminal domain of the large subunit of RNA polymerase II (RNA pol II CTD), suggesting FET oncoproteins may mediate aberrant transcriptional activation by recruiting RNA polymerase II to promoters of target genes. Here we use nuclear magnetic resonance (NMR) spectroscopy and hydrogel fluorescence microscopy localization and FRAP to visualize atomic details of a model of this process – RNA pol II CTD interactions with high molecular weight TAF15 LC assemblies. We report NMR resonance assignments of the intact degenerate repeat half of human RNA pol II CTD alone and verify its predominant intrinsic disorder by molecular simulation. By measuring NMR spin relaxation and dark-state exchange saturation transfer, we characterize the interaction of RNA pol II CTD with amyloid-like hydrogel fibrils of TAF15 and hnRNP A2 LC domains and observe that heptads far from the acidic C-terminal tail of RNA pol II CTD bind TAF15 fibrils most avidly. Mutation of CTD lysines in heptad position 7 to consensus serines reduced overall TAF15 fibril binding, suggesting that electrostatic interactions contribute to complex formation. Conversely, mutations of position 7 asparagine residues and truncation of the acidic tail had little effect. Thus weak, multivalent interactions between TAF15 fibrils and heptads throughout RNA pol II CTD collectively mediate complex formation.

*Corresponding Author: Nicolas L. Fawzi, Box G-E, Providence, Rhode Island 02912, Nicolas_Fawzi@brown.edu.

Author Contributions

A.M.J. and N.L.F. designed, performed, and analyzed experiments and wrote the manuscript. D.S. generated bacterial plasmids and K.A.B. assisted in assignment of RNA polymerase II CTD and NMR data collection. A.E.C. assisted in interpretation of DEST data and analysis of HNHA and 3D NOESY experiments. K.L.M. assisted in performing and analyzing HNHA and 3D NOESY experiments and writing the manuscript. V.R. and J.M. designed, performed, and analyzed computational/simulation experiments.

Supporting Information.

The following file is available free of charge on the ACS Publications website.

Supplementary methods, Figures S1–S8 (PDF)

Graphical Abstract

INTRODUCTION

Recent studies linking self-assembly of FET protein LC domains to enhanced recruitment of RNA pol II CTD and heightened transcriptional activation highlight a pathological role for CTD interactions in FET fusion protein-mediated sarcomas and leukemias^{1–5}. Almost half of all fusion proteins associated with sarcomas contain a SYGQ-rich transcriptional activation domain derived from a FET LC domain⁶. For example, the EWSR1-FLI1 chromosomal translocation is observed in at least 85% of Ewing's sarcoma patients; the remaining 15% of cases exhibit EWSR1 or FUS fused to ERG or EWSR1 fused to a different ETS transcription factor family DNA binding domain (i.e., ETV1, FEV, ETV4)⁶. Efforts to recover normal gene expression in Ewing's sarcoma family tumor (ESFT) cells, via suppression of EWS-FLI1 activated genes and ectopic reexpression of EWS-FLI1 repressed genes, impede tumor growth, suggesting that changes in transcription are the main drivers of pathogenesis⁷. In addition to Ewing's sarcoma, FET oncoproteins are implicated in varied subtypes of bone and soft tissue malignancies such as desmoplastic small round cell tumor (DSRCT) and acute myeloid (AML) and lymphoblastic (ALL) leukemia, which share particularly poor prognoses due to frequent metastasis, relapse following childhood onset, and chemo-resistance⁶.

Importantly, because transcriptional activation domains of FET proteins are interchangeable in multiple sarcoma subtypes, FET LC domains are hypothesized to form similar interactions⁶. One common feature of FET LC domains is their ability to drive self-association into liquid-liquid phase separated and amyloid-like assemblies^{3–5,8,9}. While it is hypothesized that liquid-liquid phase separation (LLPS) of FET proteins and other proteins with similar prion-like domains (hnRNPA2, TDP-43) seeds cytoplasmic inclusions in neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), the relationship between LLPS and fibrillar assemblies is poorly understood. Liquid-liquid phase separated FUS assemblies (i.e., “droplets”) convert into static circular assemblies within the nucleus of mammalian cells¹⁰, and FUS droplets can convert into fibrillar aggregates *in vitro*^{8,9,11}. Conversely, separate studies suggest that cross- β sheet amyloid-like fibrillization of FET proteins^{4,5,9,12} may underlie LLPS assemblies¹³ or act as stable core structures that nucleate LLPS¹⁴.

While the structural conformations and molecular interactions of FET LC domains in the context of fusion proteins bound to chromatin are unknown, previously studied hydrogels composed of amyloid-like fibrils of recombinant mCherry-tagged FET LC domains provide

a useful model for exploring molecular interactions between assemblies of FET LC domains and RNA pol II CTD¹. Indeed, the affinity of engineered FUS LC and TAF15 LC variants for wild-type hydrogels directly correlates with transcriptional activation⁵. While we have previously shown that liquid-liquid phase separated FUS LC assemblies are sufficient for recruitment of RNA pol II CTD³, here we sought to characterize the interaction of RNA pol II CTD with amyloid-like fibrils of recombinant TAF15 LC as a model for interactions between RNA polymerase II and oncogenic FET fusions.

The human RNA pol II CTD is composed of 52 tandem heptad repeats with the consensus sequence Y₁S₂P₃T₄S₅P₆S₇^{15,16}. Early NMR studies on an eight-consensus-repeat fragment of RNA pol II CTD suggested it adopts a unique conformation of overlapping β -turns¹⁷, while circular dichroism studies of the same fragment and full-length murine CTD reported predominant disorder in equilibrium with polyproline II helices and β -turns¹⁸. Recent work utilizing ¹³C direct-detect NMR and small angle X-ray scattering experiments to examine the effects of phosphorylation on the local and global structure of the *Drosophila melanogaster* CTD demonstrated that a twelve heptad, asparagine-rich section of RNA pol II CTD forms a compact random coil with individual heptads adopting structurally heterogeneous conformations^{19,20}. RNA pol II CTD heptad sequence degeneracies, a common feature of complex organisms, were shown to modulate proline isomerization and dephosphorylation, indicating diverse local heptad conformations can subtly tune interactions with regulatory factors^{19,20}. However the structure of the human CTD and specifically the lysine-rich region conserved in the last 26 heptads in mammals remains unexplored.

Interestingly, recombinant FET LC fibrils more readily bind RNA pol II CTD's degenerate lysine-rich repeat half (heptads 27-52, Figure 1A) than its tandem consensus repeat half (1-26), supporting a unique role for RNA pol II CTD's degenerate repeats in transcriptional activation¹. Additionally, RNA pol II CTD binds mCherry-tagged TAF15 LC hydrogel fibrils with greater affinity than FUS LC hydrogels and TAF15 LC acts as a more potent activator than FUS LC, despite similar sequences distinguished primarily by TAF15 LC's higher acidic residue content¹. This difference in transcriptional activation between FUS and TAF15 highlights an important question: how and to what extent do degenerate CTD residues (namely lysines, threonines, asparagines, and a single arginine) mediate interactions with assembled FET proteins¹?

Here we use solution NMR spectroscopy and simulations to measure residue-specific secondary structure population of the intact degenerate repeat half of the human RNA pol II CTD (CTD27-52). We then probe the interaction of CTD with recombinant TAF15 LC domain hydrogels as a model for FET fusion protein transcriptional interactions with CTD during promoter recruitment and subsequent formation of the pre-initiation complex. In order to address the significance of CTD's degeneracies we use solution NMR techniques (lifetime line broadening, dark state exchange saturation transfer) in combination with fluorescence localization and recovery after photobleaching experiments to probe interactions between CTD and TAF15 LC hydrogel fibrils with residue-specific resolution.

MATERIALS AND METHODS

Bacterial Plasmids. mCherry-TAF15 LC and mCherry-hnRNPA2 LC constructs were provided as gifts by Steven McKnight. CTD27-52 is a codon-optimized form of the degenerate repeat half of the CTD of DNA-directed RNA polymerase II subunit RPB1 (residues 1773-1970, recommended protein name: DNA-directed RNA polymerase II subunit RPB1; Uniprot P24928) incorporating a TEV cleavable N-terminal leader sequence and hexahistidine tag²¹ in pJ411 and synthesized by DNA2.0. After TEV cleavage of N-terminal expression sequence, constructs include a TEV cleavage and *NdeI* cloning artifact of a total of three amino acids in length: GHM. All other CTD DNA constructs were generated from CTD27-52 by either PCR site-directed mutagenesis or restriction enzyme subcloning. Variants generated are as follows: CTD N2A WT (N1775A, N1783A), CTD K5S WT (K1838S, K1859S, K1866S, K1873S, and K1887S), CTD 1961-1970, CTD27-37 (residues 1773-1852), CTD38-52 (residues 1853-1970), CTD43-52 (residues 1888-1970), CTD27-52 T1779A, P1785, P1802A, S1798A, S1822A, A1835T, Y1846A, S1857A, S1868A, S1878A, P1886A, Y1895A, P1911A, T1919A, Y1923A, T1926A, T1933A, T1943A, T1945A. Single-cysteine variants (S1966C) were also engineered in CTD27-52 WT, N2A, and K5S.

WebLogos

WebLogos for heptads 1-26 and 27-52 in Figure 1A were generated using Seq2Logo with type Shannon (basic probability) and no clustering or corrections based on low amino acid frequencies²². WebLogos are presented without bits scores (i.e. uniform height at every position) for clarity.

NMR Assignments

Assignments and relaxation data have been deposited to BMRB (accession number 27063). We used standard ¹H_N-detected triple resonance experiments (HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, a high resolution HNCA, and HNN) in combination with HSQCs of 22 CTD variants (19 missense variants and truncations CTD27-37, CTD38-52, and CTD43-52) to assign all non-overlapped backbone resonances of CTD27-52 (see **Bacterial Plasmids** for catalog of CTD variants). Missense mutations produce local perturbations in chemical structure that were observed as chemical shift perturbations in the HSQC spectrum of CTD27-52 (Figure S1). In summary, 58 residues out of CTD27-52's 145 non-proline residues correspond to non-overlapped, well-resolved peaks, and the remaining 87 residues have been localized to residue-type/sequence-motif-specific regions of the spectra. Cα, Cβ, ¹⁵N, ¹HN, and CO shifts for overlapped resonances, which were used to calculate (δCα – δCβ), SSP, and δ2D values, were assigned based on localization of residues within specific heptad sequence motifs to distinct regions of the ¹H-¹⁵N HSQC-based 3D triple resonance spectra.

Protein Expression

RNA polymerase II CTD27-52 and variants—Expression plasmids were transformed into BL21 Star (DE3) cells (Life Technologies) and grown overnight in starter cultures. Uniformly ¹⁵N-labeled protein (or ¹⁵N/¹³C-labeled) was expressed in M9 media with ¹³C

glucose and/or ^{15}N ammonium chloride as the sole carbon and nitrogen sources, respectively. One liter cultures were inoculated with 50 mL starter cultures and grown at 37 °C, 200 rpm to an OD600 between 0.6–0.8. Cultures were then induced with 0.5 mM IPTG and harvested after growing for 4 hours at 37°C, 200 rpm. Bacteria were pelleted at 4°C and stored at –80°C until purification.

mCherry-hnRNPA2/TAF15 LC—Protocol adapted from Han et al. and Kato et al., 2012^{4,5}. Expression plasmids were transformed into BL21 Star (DE3) cells (Life Technologies) and grown overnight in starter cultures. One liter LB cultures were inoculated with 50 mL starter cultures and grown at 37 °C, 200 rpm to an OD600 between 0.6–0.8. Cultures were then induced with 0.5 mM IPTG and harvested after growing overnight at 16°C, 200 rpm. Bacteria were pelleted at 4°C and stored at –80 °C until purification.

Protein Purification

RNA polymerase II CTD—Cell pellets were resuspended in 40 mL of pH 7.4 20 mM NaPi, 300 mM NaCl, 10 mM imidazole and lysed in an Avestin Emulsiflex C3. Cell lysate was cleared by centrifugation (20,000 rpm for 1 hour at 4 °C). Protein remained soluble in the supernatant following centrifugation. The supernatant was then filtered with a 0.22 µm syringe filter and loaded onto a 5 mL HisTrap HP column (GE Healthcare Life Sciences). Protein was eluted with a gradient of 10 to 300 mM imidazole in pH 7.4 20 mM NaPi 300 mM NaCl. Fractions containing protein (as determined by SDS-PAGE) were pooled and dialyzed overnight at room temperature into pH 7.4 20 mM NaPi, 300 mM NaCl, 10 mM imidazole. TEV protease was added to samples (at a ratio of roughly 2 mL of 0.3 mg/mL TEV protease per 30 mL of ~40 µM protein sample) at the start of dialysis in order to cleave the protein's hexahistidine tag. TEV protease was stored in pH 7.5 50 mM Tris-HCl, 1 mM EDTA, 5 mM DTT, 50% glycerol, and 0.1% Triton-X-100. Following dialysis and TEV cleavage, samples were again filtered with a 0.22 µm syringe filter and loaded onto a 5 mL HisTrap HP column. Cleaved protein was retrieved from the flow-through. Purity was confirmed to be >99% by SDS-PAGE, ratio of absorbance at 280 nm to 260 nm, and two-dimensional NMR. Protein was then concentrated using centrifugal filtration with a 10 kDa cutoff (Amicon, Millipore) and buffer exchanged at 4 °C into pH 7 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM βME, 0.5 mM EDTA. Samples were then flash frozen and stored at –80 °C.

mCherry-TAF15/hnRNPA2 LC—Protocol adapted from Han et al. and Kato et al., 2012^{4,5}: Cell pellets were resuspended in 40 mL of pH 7.5 (25 °C) 20 mM Tris-HCl, 500 mM sodium chloride, 10 mM imidazole, 5 mM DTT and an EDTA-free protease inhibitor tablet (Roche) and lysed in an Emulsiflex C3. Cell lysate was cleared by centrifugation (20,000 rpm for 1 hour at 4 °C). Protein remained soluble in the supernatant following centrifugation. The supernatant was then filtered with a 0.22 µm syringe filter and loaded onto a 5 mL HisTrap HP column. Protein was eluted with a gradient of 10 to 500 mM imidazole in pH 7.5 20 mM Tris-HCl, 500 mM NaCl, 5 mM DTT. Fractions containing protein (as determined by SDS-PAGE) were pooled and dialyzed overnight at 25 °C into pH 7.5 20 mM Tris-HCl, 200 mM sodium chloride, 0.5 mM EDTA, and 20 mM βME.

Hydrogel Formation

Protocol adapted from Han et al. and Kato et al., 2012^{4,5}: Following dialysis, mCherry-TAF15 LC and mCherry-hnRNPA2 LC protein were concentrated to roughly 100 mg/mL and 70 mg/mL, respectively, at room temperature using centrifugal filtration with a 10 kDa cutoff (Amicon, Millipore). 100 or 200 μ L aliquots were incubated in plastic 1.6 ml microcentrifuge tubes at 4°C and solidified overnight.

Transmission Electron Microscopy

Protocol adapted from Conicella et al., 2014²³: Samples for TEM were taken from aliquots of ~70 mg/mL mCherry-hnRNPA2 LC hydrogels and ~100 mg/mL mCherry-TAF15 LC hydrogels that were diluted to concentrations 10 mg/mL with pH 7 (at 4 °C) 20 mM Tris-HCl, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA. Aliquots were taken before and after sonication. Unsonicated hnRNPA2 and TAF15 samples shown in Figure 3 were 10 mg/mL and 3.5 mg/mL dilutions, respectively. Sonicated hnRNPA2 and TAF15 samples shown in Figure 3 were 0.2 mg/mL and 3.5 mg/mL dilutions, respectively. 4 μ L of each fibril dilution was spotted onto ultrathin carbon film on holey carbon support grids (product code 01824, Ted Pella, Reading, CA), washed three times with deionized water, stained with 5 μ L of 3% uranyl acetate (Electron Microscopy Sciences, Hatfield, PA), blotted, and air-dried. Sample grids were imaged with a Philips 410 transmission electron microscope.

Preparation of NMR Samples

RNA pol II CTD27-52 and variants—For the purposes of HSQC spectra comparison during assignment, uniformly ¹⁵N-labeled CTD samples were diluted to 50 μ M in pH 7 (at 4 °C) 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM BME, 0.5 mM EDTA and a 90% H₂O/10% D₂O mixture. For use in assignment experiments, uniformly ¹³C/¹⁵N-labeled CTD27-52 samples were diluted to either 520 μ M (for CBCA(CO)NH, HNCACB, HNCO, HN(CA)CO, and HNCA experiments) or 700 μ M (for the HNN experiment) in the same buffer and temperature conditions listed above. For measurement of backbone motions of CTD27-52 (¹⁵N R₂, ¹⁵N R₁, and heteronuclear (¹H)-¹⁵N nuclear Overhauser effects), ¹⁵N-labeled CTD27-52 samples were diluted to 250 μ M in the same buffer and temperature conditions listed above. For the 3D NOESY experiment and measurement of 3J_{HNH α} scalar coupling constants of CTD27-52, ¹⁵N-labeled samples were diluted to 200 μ M in the same buffer and temperature conditions listed above. Extinction coefficients calculated by ProtParam²⁴ were used to estimate sample concentrations.

CTD27-52 + mCherry-TAF15 LC fibrils—For binding experiments, 100 or 200 μ L aliquots of ~100 mg/mL mCherry-TAF15 LC hydrogels were diluted and resuspended on ice to final volumes of 400 μ L with pH 7 (at 4 °C) 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM BME, 0.5 mM EDTA. Diluted samples were not homogenous after resuspension and were sonicated on ice for 6 repeats of 8-second cycles at 12% power with 52-second breaks in between each cycle. Sonicated mCherry-TAF15 LC hydrogel stocks were immediately diluted to 8 mg/mL with pH 7 (at 4 °C) 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM BME, 0.5 mM EDTA and stored on ice while preparing NMR samples, which consisted of 250 μ M CTD27-52 (WT, N2A, or K5S) + 2mg/mL of sonicated mCherry-TAF15 LC fibril

diluted in pH 7 (at 4 °C) 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM BME, 0.5 mM EDTA and a 90% H₂O/10% D₂O mixture. Note that fresh hydrogel stocks were sonicated for each CTD variant tested.

CTD27-52 + mCherry-hnRNPA2 LC fibrils—For binding experiments, 100 μ L aliquots of ~70 mg/mL mCherry-hnRNPA2 LC hydrogels were first diluted and resuspended on ice with 100 μ L of pH 7 (at 4 °C) 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM BME, and 0.5 mM EDTA. Diluted hydrogel samples were not homogenous after resuspension and were sonicated on ice for 5 repeats of 10-second cycles at 12% power with 50-second breaks in between. NMR samples consisted of 250 μ M CTD27-52 + 9mg/mL of sonicated mCherry-hnRNPA2 LC fibril diluted in pH 7 20 mM Tris-HCl, 200 mM sodium chloride, 20 mM BME, 0.5 mM EDTA and a 90% H₂O/10% D₂O mixture.

Solution NMR Experiments

All NMR experiments were recorded at 4 °C in pH 7 20 mM Tris, 200 mM sodium chloride, 20 mM β ME, 0.5 mM EDTA using a Bruker Avance NMR spectrometer operating at either 850 or 500 MHz ¹H Larmor frequency equipped with a Bruker TCI z-axis gradient cryogenic probe. Experimental sweep widths, acquisition times, and number of transients were optimized for the necessary resolution, experiment time, and signal-to-noise for each experiment type but acquisition parameters were kept constant for parallel experiments. See Supplemental Information for acquisition parameters for triple resonance assignment experiments (CBCA(CO)NH, HNCACB, HNCO, HN(CA)CO, a high resolution HNCA, and HNN), ¹H-¹⁵N HSQC spectra, ¹⁵N R_1 , temperature-compensated ¹⁵N R_2 , ¹H-¹⁵N heteronuclear NOE, ³J_{HNHA}, and dark-state exchange saturation transfer (DEST) experiments.

DEST Model Fitting

Dynamic and kinetic parameters describing RNA pol II CTD27-52 WT and K5S interaction with sonicated TAF15 LC fibrils were derived from experimental R_2 and DEST data and fit with DESTfit in Matlab as previously described^{23,25,26}. DESTfit was run with the two-state and pseudo-two-state fit types and no parameters were fixed. To better estimate R_2^{tethered} and K_3 parameter uncertainty, 100 Monte Carlo simulations were performed. Synthetic data sets for Monte Carlo simulations were generated via Gaussian sampling of experimental R_2 and DEST data using the estimated uncertainty of the observed DEST profile intensity data.

Simulations

We performed parallel tempering metadynamics in well-tempered ensembles (PTMetaD-WTE) simulations of 44-residue fragment of RNA pol II CTD (residues 1927-1970) in explicit water with ff99SBws and TIP4P/2005 using GROMACS-4.6.7^{27,28} and Plumed 2.2²⁹. NMR chemical shift deviations in the equilibrium structural ensembles of the RNA pol II CTD peptide are calculated using an empirical chemical shift deviation prediction algorithm, SPARTA+³⁰. For further description of simulation methods see Conicella et al., 2016³¹.

Fluorescence microscopy and FRAP experiments

Quantification of RNA pol II CTD fluorescence localization to mCherry-TAF15 LC hydrogels was adapted from protocols developed by the McKnight lab^{1,4,5,32}. For fluorescence localization and fluorescence recovery after photobleaching (FRAP) experiments, mCherry-TAF15 LC was purified as described above and concentrated to 100 mg/mL. 3x2 μ L aliquots were pipetted into single wells of a 24-well MatTek glass-bottom plate. A few blank wells without hydrogels were filled with Tris buffer before parafilming the lid onto the plate and storing at 4 °C to prevent evaporation of samples before fibrillization. Note that in contrast to recently published methods²⁷, we did not sonicate mCherry-TAF15 LC before hydrogel formation. Hydrogels solidified at 4°C overnight.

Fluorescently labeled CTD variants were prepared using two approaches: 1) For experiments comparing N2A, K5S, and WT in different NaCl concentrations, a single cysteine was introduced (S1966C) in N2A, K5S, and WT CTD27-52 plasmids in preparation for Alexa Fluor-488 maleimide labeling. CTD 27-52 S1966C variants were purified from LB media as described above and buffer exchanged into pH 7.4 20 mM NaPi 300 mM NaCl, 1 mM DTT to eliminate any disulfides before snap freezing and storage at -80 °C. Protein samples were desalted using 2 mL 7000 MWCO spin desalting columns (Zeba, ThermoFisher) into pH 7.5 20 mM Tris, 200 mM NaCl to remove DTT immediately before maleimide conjugation.

Protein samples were incubated with Alexa Fluor 488 C5 maleimide according to manufacturer's instructions and excess label was removed by desalting samples twice with 2 mL 7000 MWCO Zeba spin desalting columns. 2) For experiments comparing WT and CTD 1961-1970, ¹⁵N-labeled protein was purified as describe above. Protein samples were then buffer exchanged into pH 6.5 100 mM NaPi and 150 mM NaCl using 0.5 mL 7000 MWCO Zeba spin desalting columns and incubated with DyLight 488 NHS Ester dye at pH 6.5 for 16 hours at 4 °C in order to preferentially label the α -amino N-terminal group rather than lysine amino groups³³. Manufacturer's instructions for fluorescent labeling were otherwise followed, and excess label was removed by desalting samples twice with 2 mL 7000 MWCO Zeba spin desalting columns.

300 μ L of 1 μ M CTD 27-52 WT-Alexa 488, N2A-Alexa 488, K5S-Alexa 488, DyLight 488 labeled CTD27-52 WT, or DyLight 488 labeled CTD 1961-1970 was then added to each well of a 24-well glass bottom MatTek plate which contained three 2 μ L hydrogels of mCherry-TAF15 LC per well. Samples were imaged on an LSM 710 confocal microscope (Zeiss) with either a 5x (for full hydrogel slices) or 20x objective (for FRAP experiments, using a 488-nm laser line). For FRAP experiments, circular regions 8 μ m in diameter positioned approximately 0.2 mm from the edge of ~2 mm diameter hydrogels were bleached with 100 iterations of full laser power and fluorescence recovery was imaged at 2.1% laser intensity. FRAP recovery curves for each trial were computed using the FRAP_Calculator v3 macro for ImageJ (NIH) and fit to single exponential models to calculate halftime recoveries. Calculated halftime recoveries were averaged from three trials and translational diffusion coefficients were then estimated given average halftime recoveries, bleached regions of a defined radius, and percentage bleached³⁴. For confocal slices of full hydrogels, intensity profiles for mCherry-TAF15 (594 nm) and fluorescently tagged CTD (488 nm) were calculated for horizontal slices of confocal images using

ImageJ. Three hydrogels were imaged per CTD condition at stated time points, and samples were stored at 4 °C between image collection.

RESULTS

Human RNA polymerase II CTD heptads 27-52 is predominantly disordered

We first assigned the backbone amide NMR resonances observed in the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum of CTD27-52. Its highly repetitive and proline-rich sequence (Figure 1A) made these assignments particularly challenging (Figure 1B), causing resonance overlap and interruptions in traditional triple resonance (H/C/N) assignment experiments, respectively, thus complicating sequential assignments (deposited as BMRB 27063). Although we attempted carbon-detected experiments³⁵, significant overlap of peaks in the ^{13}C , ^{15}N CON spectrum, the high concentration of protein required, and our eventual goal of using ^1H -observe experiments led us to pursue an alternate approach. Ultimately, we used standard $^1\text{H}_\text{N}$ -detected triple resonance experiments (see Methods) in combination with HSQCs of a series of CTD variants to assign all non-overlapped backbone resonances of CTD27-52 (Figure S1, see methods for catalog of CTD variants).

The narrow chemical shift dispersion observed in the CTD27-52 HSQC (Figure 1B) and relatively uniform ^{15}N NMR spin relaxation parameters (^{15}N R_2 , R_1 , heteronuclear Overhauser effect (NOE)) (Figure 1C), which are sensitive to protein backbone reorientational motions on the picosecond to nanosecond timescale, are indicative of predominant disorder. Although backbone amide transverse relaxation is uniformly higher at higher field (850 MHz compared to 500 MHz ^1H Larmor frequency) (Figure 1C), ^{15}N R_2 measurements at 850 MHz and 500 MHz are linearly correlated with no obvious outliers (Figure S2A), providing no evidence that locally elevated ^{15}N R_2 arises due to chemical exchange effects from transient population of structured conformations. Secondary structure propensity (SSP) scores³⁶ and differences in secondary $\text{C}\alpha$ and $\text{C}\beta$ shifts ($\delta\text{C}\alpha - \delta\text{C}\beta$) near zero are consistent with a primarily disordered conformation; SSP values of -1 and 1 predict β -sheet and α -helix secondary structures, respectively, whereas $\delta\text{C}\alpha - \delta\text{C}\beta$ values near -4 and 4 are predicted for 100% β -sheet and 100% α -helix conformations, respectively. Using $\text{C}\alpha$, $\text{C}\beta$, ^{15}N , ^1HN , and CO shifts as inputs to the $\delta 2\text{D}$ algorithm for predicting secondary structure population, CTD27-52 is predicted to be predominantly random coil (62%) with minor populations of dihedral angles consistent with polyproline II helix (25%) and extended β -sheet (13%) (Figure 1C). Note that $\text{C}\alpha$, $\text{C}\beta$, ^{15}N , ^1HN , and CO shifts for degenerate repeats that map to distinct but overlapped positions in the ^1H - ^{15}N HSQC spectra were included for ($\delta\text{C}\alpha - \delta\text{C}\beta$), SSP, and $\delta 2\text{D}$ calculations (see methods).

Simulations of RNA pol II CTD 1927-1970 extend structural characterization of human degenerate repeats

In order to further characterize the structure of CTD and account for the possibility of transient structured conformations that are difficult to observe by NMR experiment alone due to conformational averaging and resonance overlap, we employed a physics-based molecular simulation approach. This approach combines parallel tempering metadynamics

and water and protein force fields shown to efficiently sample the equilibrium conformational ensemble for disordered proteins in reasonable agreement with experimental observables (local secondary structure, radii of gyration)^{37–40} as we demonstrated for TDP-43³¹. We simulated explicit solvent molecular dynamics ensembles of an RNA pol II CTD subpeptide (amino acids 1927-1970) containing heptads 49-52 and the acidic tail, choosing this representative subpeptide for tractability of the simulations. Both the heptad repeats and the acidic tail are primarily unstructured, as observed by experiment (see above). Values for NMR parameters calculated from simulated ensembles are in reasonably good agreement with experimentally observed $^3J_{\text{HNH}\alpha}$ scalar coupling constants and differences in secondary $\text{C}\alpha$ and $\text{C}\beta$ shifts (Figure 2A–B) as well as NMR spin relaxation parameters (^{15}N R_1 , R_2 , NOE) (Figures S2B), validating the use of the simulated structural ensemble to examine heterogeneous and transient conformations of CTD heptad repeats. Interestingly, DSSP⁴¹ analysis of secondary structure elements shows that residues 1954-1956 adopt helical structure in ~25% of the ensemble (Figure 2C). Residues 1954-1956 correspond to the first three residues ('YSL') of heptad 51. The proline to leucine substitution at position 3 is a rare degeneracy only seen at position 1956 in the human CTD, and although it is conserved in mammals its function is unknown. A map of α -helical secondary structure was created by binning adjacent stretches of α -helical torsion angles⁴², and a helical structure centered at 1955 for a contiguous span of four residues (a complete helical turn) was observed in 13% of simulated ensembles (Figure 2D). The helix at this position extends to up to eight residues in length in minor populations. Only small populations of α -helix were observed for other consecutive stretches (~8% of a short helix centered at 1938). Using the same torsion angle binning technique, the remainder of the simulated subpeptide is primarily random coil (Figure S2C) and no heightened propensity for β -sheet was observed for any contiguous span of the CTD (Figure S2D).

Short, low-population secondary structures are difficult to detect by NMR because observables like scalar couplings are population-weighted averages and hence report primarily on the most populated (e.g., disordered) structure (Figure 2B). Hence, we also measured ^1H - ^1H NOEs, whose intensities increase non-linearly as a function of proximity and can report on close contacts formed in transiently populated states⁴³, to attempt to confirm the presence of partial helical structure. Due to the low complexity sequence and resulting high degree of resonance overlap in all spectra except the ^1H - ^{15}N HSQC, only NOEs to ^1HN positions are clearly resolvable via 3D spectra. In a four-day 3D NOESY experiment (^1H - ^1H NOE - ^1H - ^{15}N HSQC), we did not observe evidence for i to $i+3$ NOEs indicative of helical structure (for example, between $\text{H}\alpha_i$ and HN_{i+3}) (Figure S2E). However, even in this spectrum, overlap in $^1\text{H}_{\text{aliphatic}}$ chemical shift resonances (i.e. $\text{H}\alpha$ for all residues are not resolved, $\text{H}\beta$ and other side chain ^1H positions are partially resolved) caused by primarily random coil structures and sequence repetition precludes observation of all potential NOEs. Future experiments at higher concentration to increase signal-to-noise and with shorter subpeptides containing this region to decrease overlap would be necessary to definitively quantify and validate the presence of transient helix near L1956.

NMR transverse relaxation enhancements (R_2) and dark-state exchange saturation transfer (DEST) reveal residue-specific dynamics of RNA polymerase II CTD 27-52 in complex with hnRNPA2 and TAF15 fibrils

With the resonances assigned and predominant global disorder confirmed, we could then probe with atomic detail the interaction of CTD27-52 with hydrogels of mCherry-TAF15 LC. To generate homogenous samples for solution NMR, mCherry-TAF15 hydrogel fibrils were briefly sonicated, similar to amyloid-seeding protocols^{44,45}. Hydrogels of hnRNPA2 LC which retain CTD much more weakly than TAF15¹ were used as a control. While TAF15 and hnRNPA2 fibril clusters were disrupted by sonication, the fibrils themselves were not destroyed but rather truncated into small “seeds” (Figures 3A,B).

Addition of sonicated hnRNPA2 or TAF15 fibrils to CTD27-52 does not produce chemical shift differences in the CTD27-52 spectra (Figures S3A, S4A). However, in the presence of sonicated fibrils we reliably observe decreased peak intensity and increased transverse relaxation, R_2 , of CTD27-52 which can be explained by transient binding of CTD27-52 to the high molecular weight fibrillar segments²⁵. R_2 values are maximal for heptads 27-33 of RNA pol II CTD27-52 (Figures 3B–C, S4B), suggesting that the N-terminal half of CTD27-52 more frequently or tightly binds fibrils, possibly because the acidic C-terminal tail of CTD27-52 remains unbound. Notably, sonicated TAF15 fibrils produce markedly higher R_2 values than sonicated hnRNPA2 fibrils: adding 2 mg/mL of TAF15 fibrils to CTD27-52 produces maximum R_2 values of 14 s^{-1} , whereas 9 mg/mL hnRNPA2 fibrils is necessary to produce similar relaxation enhancements (Figure 3C). However, similar R_2 vs. CTD residue number profiles suggest similar CTD27-52 binding modes to both TAF15 and hnRNPA2 fibrils. Although even the weak CTD27-52 binding to hnRNPA2 fibrils was unexpected because no recruitment was observed by fluorescence microscopy¹, mCherry-hnRNPA2 has a similar sequence composition and forms similar amyloid-like gels as FET proteins^{4,5} and pull-downs suggest hnRNPA2 interacts with RNA pol II⁴⁶.

The observation of R_2 values that are nearly field-independent (Figure S3C–D, S4C–E) support a model where R_2 arises from binding of CTD to the surface of large ($> 1 \text{ MDa}$) hnRNPA2 or TAF15 fibrils, rather than from direct observation of a bound state^{25,26}. The maximum value of R_2 then provides a lower bound for the first order kinetic on-rate for binding which must be reversible as the CTD in the sample is stable and not consumed by irreversible binding within 1 s. Addition of supernatant extracted from centrifuged sonicated fibrils results in no observable R_2 , demonstrating that residual soluble species within sonicated hydrogels are not responsible for observed R_2 (Figures 3C, 3D). Separately, addition of 2 mg/mL monomeric mCherry-TAF15 LC also does not increase R_2 (Figure 3D), further demonstrating that R_2 arises from transient fibril binding.

We next sought to test the two following questions: 1) Given that the highest R_2 values occur near N1775 and N1782, are asparagine CTD residues critical for TAF15 recruitment? 2) Given that TAF15 LC contains more aspartates/glutamates (23) than either hnRNPA2 LC (6) or FUS LC (2), do opposite charge interactions between lysines in CTD27-52 and acidic residues in TAF15 explain heightened CTD recruitment to TAF15 compared to hnRNPA2 and FUS? Therefore, we made two CTD27-52 variants substituting either two N-terminal asparagines (N1775, N1782) with alanines (N2A) or five central lysines (K1838, K1859,

K1866, K1873, and K1887) with consensus serines (K5S). Note that chemical shift perturbations in the HSQC spectra of CTD27-52 variants N2A (Figure S5A) and K5S (Figure S5B) as compared to that of WT are localized within heptads containing mutations, suggesting no global structural differences are caused by these CTD variants.

To account for variability observed in hydrogel preparations, multiple trials mixing each CTD27-52 variant and sonicated TAF15 fibril sample were performed (Figure S4B); variant-dependent trends in R_2 are consistent across all trials and magnetic field strengths (Figure S4F). CTD27-52 N2A displays R_2 values indistinguishable from that of CTD27-52 wild-type (Figure 3D), suggesting N1775 and N1782 sidechains do not directly mediate binding. However, CTD27-52 K5S shows R_2 values ~30% less than wild type (Figure 3D), suggesting lysine contacts do contribute to binding. This decrease in R_2 values cannot be directly attributed to a decrease in the population of fibril-bound CTD K5S as compared to that of WT, because differences in transverse relaxation enhancements are dependent on not only bound populations but also binding/unbinding rates as well as possibly on chemical shift differences in the bound state. As a result, additional experiments to measure the contribution of lysines to binding were designed (see below).

In order to further investigate the residue-by-residue dynamics of CTD27-52 and TAF15 fibril association and the specific roles of asparagine and lysine degeneracies, we employed dark-state exchange saturation transfer (DEST) NMR. As previously demonstrated^{25,26}, DEST NMR allows quantitative description of the average motions of individual RNA pol II CTD27-52 residues bound to TAF15 fibrils despite the NMR-prohibitive size of the complex. DEST takes advantage of both the extremely slow tumbling of the bound state and the equilibrium with the free (NMR-visible) form. Briefly, selective saturation of slowly relaxing ^{15}N longitudinal magnetization in the bound state is transferred to the monomeric, NMR-visible state by exchange (i.e. unbinding) and observed as peak intensity decreases of the resolved free CTD27-52. The greater the transverse relaxation rate in the bound state at the specific residue due to greater extent of binding, the greater the resulting residue-specific attenuation of monomeric CTD27-52 resonances within the DEST spectrum. By applying two different radiofrequency (RF) fields at a series of frequency offsets, DEST profiles for each resolved CTD27-52 resonance are generated. A kinetic binding model accounting for NMR spin relaxation can then be fit simultaneously to DEST and R_2 data to extract global and residue-specific kinetic and dynamical parameters^{25,26}.

A simple two-state model with a single fibril-bound state does not allow simultaneous fitting of R_2 and DEST data, as demonstrated by the poor fit of experimental R_2 (Figure 4A, gray). It should be noted this simplest model attempts to capture all the bound conformations with a single average residue-specific parameter for R_2 in the bound state; no heterogeneity in conformations can be accommodated. As previously described for A β peptides binding A β protofibrils^{23,25}, to account for the surely diverse conformational ensemble of CTD bound to TAF15 fibrils, a simple phenomenological extension to a two-state model, accounting for the fact that a single residue may be in direct contact with the TAF15 fibril in some bound-conformation while tethered to the surface via the binding of other residues in other bound states, results in excellent fits of R_2 and DEST data (Figure 4A, 4B). Greater saturation transfer for residues in the N-terminal region including N1782 and R1810 (Figure

4B) reflects more stable association with fibrils compared to residues in the acidic C-terminal region (T1957, D1964). Intermediate saturation transfer values are observed in the central region (K1887, K1936), consistent with the patterns in R_2 .

The fitted apparent first-order binding ($k_{\text{on}}^{\text{app}}$) and unbinding (k_{off}) rates are consistent with transient binding by CTD27-52 (Figure 5A). The best-fit $k_{\text{on}}^{\text{app}}$ of $13.7 \pm 1.7 \text{ s}^{-1}$ agrees with the maximal R_2 value observed for the sample ($12.3 \pm 1.3 \text{ s}^{-1}$ for residue R1810) which serves as a lower bound for $k_{\text{on}}^{\text{app}}$. Similarly, the population of CTD27-52 in the fibril-bound state, p_B , is $2.9\% \pm 0.8\%$ and k_{off} is estimated as $470 \pm 150 \text{ s}^{-1}$, suggesting weak global affinity consistent with the small p_B .

The best-fit residue-specific K_3 and R_2^{tethered} parameters derived from the pseudo-two-state model provide insight into each individual residue's relative probability to directly bind to the fibril surface and motions when tethered to the surface of a fibril, respectively. In this model, the ensemble of fibril-bound CTD27-52 states is described by a probability of each residue (i) being either in direct contact with the fibril surface or tethered by binding of residues elsewhere on the chain, as described by effective equilibrium constant $K_3(i) = p_{\text{direct-contact}}(i) / p_{\text{tethered}}(i)$. Here, K_3 ranges from near zero at the acidic tail, which rarely binds TAF15 fibrils, to values of 0.3 to 1 in the first half of CTD27-52, which avidly binds (Figure 5C). The average ^{15}N transverse relaxation rate of each residue when directly contacting, $R_2^{\text{direct-contact}}$, or tethered, $R_2^{\text{tethered}}(i)$, to the surface provides insight into the local motions experienced by the chain in each state. Constrained to be the same for each residue due to our simplifying assumption that residues in direct contact with the surface of the fibril will take on similarly slow tumbling motions of the fibril species, $R_2^{\text{direct-contact}}$ is extremely fast ($12,500 \text{ s}^{-1} \pm 820 \text{ s}^{-1}$) (Figure 5B), as observed for amyloid β monomers bound to high molecular weight amyloid protofibrils²⁵. Residue-specific $R_2^{\text{tethered}}(i)$ is low at the acidic tail, implying fast motion due to long tether lengths, and high (200 to 2000 s^{-1}) across the N-terminal and central sections due to contacts formed by nearby heptads. Intriguingly, K_3 is maximal near R1810 and R_2^{tethered} is maximal at Q1814 suggesting this region may mediate contacts with transcriptional activation domains as well as its recently reported functional contact with SMN protein in the R-loop pathway⁴⁷.

Because mutation of 5 central lysines to serines modestly reduced recruitment of CTD27-52 as observed by R_2 experiments, we next investigated interactions between CTD27-52 K5S and TAF15 fibrils via DEST NMR. Residues near lysine-to-serine substitutions in CTD27-52 K5S show modestly decreased saturation transfer in the presence of TAF15 fibrils compared to the same residues in CTD27-52 WT (Figure S6), suggesting that substitution of five lysines with consensus serines may impede assembly by removing lysine-localized interactions between heptads and TAF15 fibrils. While this observation is consistent with decreased binding, instability of CTD27-52 K5S + TAF15 samples over the length of the DEST experiment and low overall signal-to-noise resulted in large uncertainty in extracted parameters and precluded quantitative comparison with WT. Therefore, we used hydrogel localization and FRAP assays to quantitatively compare WT with K5S and other variants (see below). Conversely, no significant DEST attenuation of CTD27-52 in the presence of 9 mg/mL sonicated hnRNPA2 fibrils was observed (data not shown). Because

R_2 values are comparable to 2 mg/ml TAF15, lack of DEST in the presence of hnRNPA2 fibrils may be due to decreased population or altered residence time in the fibril-bound state.

Fluorescence localization and FRAP experiments support contributions of lysine electrostatic interactions to TAF15 hydrogel binding

Given that NMR samples of CTD27-52 variants and sonicated mCherry-TAF15 LC hydrogel fibrils were relatively unstable and inherently variable due to differences in sonicated hydrogel preps (Figure S4B), we sought to verify the contribution of lysine residues and other electrostatic interactions to CTD-TAF15 assembly using fluorescence localization and recovery after photobleaching (FRAP) experiments. Using a hydrogel localization assay adapted from McKnight and colleagues^{1,4,5}, we quantified CTD27-52 variants' recruitment to, and penetrance within, 2 μ l intact mCherry-TAF15 LC hydrogels. A serine-to-cysteine substitution (S1966C) was introduced in all three variants (WT, K5S, and N2A) in order to label each variant with a single Alexa Fluor dye via maleimide-thiol conjugation, and a solution of 1 μ M Alexa Fluor tagged CTD was added to preformed mCherry-TAF15 LC hydrogels. A fluorescence time course (without a wash phase) was performed in triplicate for each variant and salt concentration, and profiles of fluorescence intensities across the diameter of each hydrogel illustrate differences in fluorescence localization for each CTD variant (Figure S7A,B). Because Alexa Fluor-labeled CTD fluorescence intensity may change along the z-axis of each mCherry-TAF15 LC hydrogel, slices were taken at the same position within each hydrogel, as demonstrated by comparable profiles of mCherry fluorescence intensity (Figure S7C,D).

As previously observed⁵, high avidity binding of the CTD27-52 in solution corresponds to CTD27-52 fluorescence localization around the outer ring of each mCherry-TAF15 LC hydrogel. Throughout a three-day time course of confocal slices after addition of Alexa Fluor-labeled CTD, CTD27-52 WT and N2A primarily localize to the exterior surface of the TAF15 hydrogels. Confocal slices through mCherry-TAF15 hydrogels show a circular ring of green fluorescence when Alexa-Fluor labeled CTD27-52 WT or N2A is trapped at the hydrogel exterior. Conversely, K5S exhibits increased penetrance of CTD fluorescence to the center of the hydrogel and reduced fluorescence intensity around the exterior. Although this experiment is not at equilibrium, the distinct kinetic pattern of K5S localization presumably arises because it is less easily trapped by high avidity binding of TAF15 hydrogel fibrils than WT and N2A (Figure 6A, S7). Similarly, incubation of CTD27-52 WT in increasing salt concentrations (50, 100, 400 mM sodium chloride) decreases CTD fluorescence intensity around the circumference of each TAF15 hydrogel (Figure 6B and Figure S7B), supporting the importance of electrostatic interactions to CTD:TAF15 complex formation. Increasing the concentration of salt ions screens out electrostatic interactions between CTD variants and TAF15, therefore Alexa-Fluor labeled CTD fluorescence within each TAF15 hydrogel may decrease.

Results from FRAP experiments largely mirror fluorescence localization trends. Compared to WT, CTD K5S displayed faster recovery after photobleaching a circular bleached region with a radius of 4 μ m (recovery half life of 35.6 ± 0.8 s for K5S compared to 58.5 ± 1.1 s for WT) and thus a faster estimated translational diffusion coefficient³⁴ (approximately

0.176 ± 0.002 μm²/s for K5S compared to 0.117 ± 0.002 μm²/s for WT), again consistent with lower affinity of K5S for mCherry-TAF15 LC hydrogel fibrils. By contrast, N2A exhibits a half life and diffusion coefficient similar to WT (Figure 6C). Altering the salt concentration has modest effects on diffusion; compared to our standard condition of 200 mM NaCl, CTD WT in 50 mM NaCl has a half life of 69.1 ± 1.1 seconds and a diffusion rate of 0.099 ± 0.002 μm²/s, whereas CTD WT in 400 mM NaCl exhibits a half life of 53.0 ± 0.9 seconds and a diffusion coefficient of 0.129 ± 0.002 μm²/s (Figure 6D). Given these observed differences in fluorescence localization and FRAP due to changes in salt concentration, we next attempted to use NMR to measure R_2 of CTD27-52 WT in the presence of sonicated TAF15 fibrils and buffers with varying salt concentrations. Unfortunately, our standard sample concentrations of 250 μM CTD27-52 mixed with sonicated TAF15 fibrils rapidly precipitate in 100 mM NaCl buffer, precluding measurement of R_2 . The heightened sensitivity to incubation with RNA pol II CTD of our sonicated TAF15 fibril samples compared to intact hydrogels may be due to the ability of CTD27-52 monomers (highly concentrated in NMR samples versus dilute when incubated with TAF15 hydrogels) to bridge multiple sonicated hydrogel fibrils, seeding additional interactions between TAF15 fibril species within NMR samples as compared to within intact hydrogels, which are held apart by steric fibril-fibril contacts forming the hydrogel. Thus, NMR sample stability at 200 mM NaCl may be due to screening out either excessive CTD-TAF15 or TAF15-TAF15 interactions. This salt concentration threshold may be near 200 mM NaCl, because increasing the salt concentration above 200 mM in our NMR sample preps does not significantly affect the interaction between CTD27-52 WT and TAF15 fibrils: R_2 values of CTD27-52 WT + 2 mg/mL TAF15 fibrils in 400 mM NaCl were within error of R_2 values in 200 mM NaCl (Figure S8A).

Taken together, quantification of differences in RNA pol II CTD27-52 green fluorescence intensity profiles and FRAP for three variants (WT, N2A, and K5S) and three NaCl concentrations (50, 100, 400 mM) corroborate residue-specific NMR observables and support a model in which charge-charge electrostatic contributions from CTD lysine residues contribute to TAF15 fibril complex formation. Because removing five of the CTD's nine positively charged residues (8 lysines and R1810) both reduces apparent binding to sonicated TAF15 fibrils (R_2 decrease by ~30%) and increases diffusion within TAF15 hydrogels by roughly 50%, we suggest that the increased prevalence of positively charged degeneracies in the human CTD may serve to modulate interactions with assemblies such as TAF15.

The acidic C-terminal tail (residues 1961-1970) of the human CTD does not contribute to binding of TAF15 hydrogels but prevents aggregation of sonicated fibrils

Because NMR relaxation and DEST experiments showed less binding of the acidic C-terminal tail of human RNA pol II (residues 1961-1970) to sonicated TAF15 fibrils compared to heptad repeats, we next sought to further interrogate the role of residues 1961-1970 and whether they are actively repelled from TAF15 fibrils. Unfortunately, NMR samples of 250 μM CTD27-52 1961-1970, a CTD variant lacking the last ten residues, rapidly aggregate upon the addition of sonicated TAF15 fibrils at concentrations as low as 0.2 mg/mL. Interestingly, CTD27-52 1961-1970 is stable in NMR sample buffer at

concentrations up to 1 mM, indicating that the combination of CTD 1961-1970 and sonicated TAF15 fibrils induces aggregation. Additionally, removing the acidic C-terminal tail does not affect transverse relaxation rates of samples of CTD alone (Figure S8B) or produce any long-range chemical shifts in the HSQC spectrum (Figure S8C), suggesting its removal has no significant effect on the structure of the CTD. Based on these observations, it is possible that removing the acidic C-terminal tail increases association between CTD monomers and TAF15 sonicated fibrils. However, when dilute CTD 1961-1970 is incubated with intact TAF15 hydrogels, CTD 1961-1970 fluorescence localization and recovery after photobleaching are almost identical to that of WT (the half life and diffusion rate of CTD 1961-1970 were 56.6 ± 7.1 seconds and $0.096 \pm 0.012 \mu\text{m}^2/\text{s}$, respectively, compared to 58.0 ± 0.3 seconds and $0.094 \pm 0.001 \mu\text{m}^2/\text{s}$ for WT) (Figure S8D–E). For these experiments, CTD WT and 1961-1970 were selectively labeled at the N-terminus with Alexa Fluor NHS ester dyes. These data support a model in which residues 1961-1970 play a negligible role in complex formation between RNA pol II CTD and intact TAF15 hydrogel fibrils. Taken together, our NMR and fluorescence microscopy data show no evidence that the acidic C-terminal tail of the human CTD actively binds TAF15 fibrils.

DISCUSSION

The data presented here suggest the intact lysine-rich degenerate repeat half of the human RNA polymerase II CTD is predominantly disordered. This is in agreement with recent conclusions from proteolysis, small angle x-ray scattering, and ^{13}C -detect NMR studies that suggest the human and *Drosophila* RNA pol II CTDs form compact random coils^{19,20}. Our molecular simulations suggest that heptad 52, which contains the rare P1956L degeneracy, forms a transient α -helix, although we were unable to corroborate this experimentally. While a transient helix near P1956L may play a role in the CTD's interactions with binding partners, it is clear based on both our NMR and simulations that the CTD is predominantly disordered.

These assignments of the backbone resonances of CTD27-52 will enable future residue-specific analysis of CTD complexes such as with Mediator and TFIIF associated with transcriptional and co-transcriptional interactions as well as the location and effect of various post-translational modifications on CTD structure and interactions. The ^1H - ^{15}N HSQC spectrum of the degenerate repeat half of the human CTD is remarkably well-resolved given its sequence repetition: 60 residues out of CTD27-52's 148 non-proline residues correspond to non-overlapped, well-resolved peaks, and the remaining 88 residues have been localized to residue-type-specific regions of the spectrums. Though the data presented here are only for the unmodified CTD, recent work on CTD phosphorylation identification within specific heptad repeats via genetic and mass spectrometric techniques suggest that multiple phosphorylations within a single CTD heptad are rarely observed in mammalian cells, and many heptads remain unphosphorylated, underscoring the physiological relevance of the unmodified CTD heptads^{48,49}.

The structural characterization of CTD27-52 in complex with TAF15 LC fibrils presented here may provide insight into the role of the human CTD's degenerate repeats in transcription initiation. NMR and hydrogel binding assays using several CTD variants and at

different salt concentrations suggest that degenerate lysines at position 7 in the human RNA pol II CTD mediate interaction with TAF15 fibril assemblies and that the acidic tail does not participate in the interactions. Interestingly, although removal of residues 1961-1970, comprising the acidic C-terminal tail of the CTD (ISPDDSDDEEN), does not alter binding of CTD to intact TAF15 hydrogels based on FRAP experiments, NMR samples of CTD

1961-1970 and sonicated TAF15 LC fibrils rapidly precipitate. Taken together, these results suggest that the acidic C-terminal tail may in some conditions dampen interactions between the CTD and sonicated TAF15 fibrils, possibly because its negative charge repels aspartic acids throughout the TAF15 LC domain as well as other CTD monomers in this *in vitro* model environment.

These results suggest FET fusion proteins bind RNA pol II CTD via multivalent interactions: numerous tyrosines, serines, and threonines within consensus YSPTSPS heptads may form polar contacts with FET SYGQ repeats, and R1810 and lysine degeneracies may form electrostatic interactions with FET aspartic and glutamic acids. In contrast with previous fluorescence microscopy studies which found no association between hnRNPA2 LC fibrils and RNA pol II CTD¹, NMR relaxation enhancements suggest that RNA pol II CTD can bind to hnRNPA2 LC fibrils in a similar, residue-specific manner as TAF15 LC fibrils, albeit much more weakly.

It remains to be answered whether amyloid-like fibrils of FUS or EWSR1 LC domains interact in the same residue-specific way with the CTD's degenerate repeats as observed for amyloid-like fibrils of recombinant TAF15 and hnRNPA2 LC domains. Unlike mCherry-TAF15 LC and mCherry-hnRNPA2 LC, preparations of mCherry-FUS LC hydrogels readily dissolve upon dilution in our hands and are unstable to sonication, despite testing a range of 4°C incubation concentrations from roughly 100 mg/mL to 500 mg/mL. In addition, attention must now be given to how the RNA pol II CTD's interactions with FET fibrils described here compare to its interactions within liquid-liquid phase separated LC domain assemblies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources

No competing financial interests have been declared. Research reported in this publication was supported in part by the National Institute Of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under Award Numbers R01GM118530 (to N.L.F.), a subproject as part of an Institutional Development Award (IDeA) from NIGMS (P20 GM104937), and Medical Research Grant #20133966 from the Rhode Island Foundation (to N.L.F.). A.M.J. was supported in part by Karen T. Romer Undergraduate Teaching and Research Awards from Brown University. A.E.C. and K.L.M. were supported in part by an NIGMS training grant to the graduate program in Molecular Biology, Cell Biology and Biochemistry (MCB) at Brown University (T32 GM07601) and a BIBS Graduate Award in Brain Science from the Brown Institute for Brain Science Reisman Fund (to A.E.C.) and a Sidney Frank Fellowship (to K.L.M.). Work at Lehigh University was supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), Division of Material Sciences and Engineering, under Award DE-SC0013979 (to J.M.). Use of the high-performance computing capabilities of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (NSF) Grant TG-MCB-120014, is gratefully acknowledged. This research used resources of the National Energy Research

Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research is based in part on data obtained at the Brown University Structural Biology Core Facility supported by the Division of Biology and Medicine, Brown University, and at the Brown Genomics Core Facility supported by NIGMS P30GM103410, NCRR P30RR031153, P20RR018728 and S10RR02763, National Science Foundation EPSCoR 0554548. This research is based in part upon work conducted using the Rhode Island NSF/EPSCoR Proteomics Share Resource Facility, which is supported in part by the National Science Foundation EPSCoR Grant No. 1004057, National Institutes of Health Grant No. 1S10RR020923, S10RR027027, a Rhode Island Science and Technology Advisory Council grant, and the Division of Biology and Medicine, Brown University. Experimental DEST fitting in Matlab was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

We thank Steven McKnight and Masato Kato for gifts of mCherry fusion plasmids and helpful suggestions. We thank Geoff Williams and the Leduc Bioimaging Facility at Brown University for confocal and transmission electron microscopy assistance and resources, respectively. We thank Michael Clarkson and Mandar Naik for assistance with NMR experiments, Gerwald Jogl for use of sonication equipment, and Christoph Schorl, Veronica Ryan and Anastasia Murthy for helpful advice.

ABBREVIATIONS

HSQC	heteronuclear single quantum coherence
DEST	dark-state exchange saturation transfer
CTD	C-terminal domain
RNA pol II	RNA polymerase II
hnRNPA2	heterogeneous nuclear ribonucleoprotein A2
TAF15	TATA-box binding protein associated factor 15
FUS	fused in sarcoma
EWSR1	Ewing sarcoma breakpoint region 1 protein

References

1. Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell*. 2013; 155:1049–1060. [PubMed: 24267890]
2. Schwartz JC, Ebmeier CC, Podell ER, Heimiller J, Taatjes DJ, Cech TR. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes & development*. 2012; 26:2690–2695. [PubMed: 23249733]
3. Burke KA, Janke AM, Rhine CL, Fawzi NL. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular cell*. 2015; 60:231–241. [PubMed: 26455390]
4. Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell*. 2012; 149:768–779. [PubMed: 22579282]
5. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012; 149:753–767. [PubMed: 22579281]
6. Riggi N, Cironi L, Suva ML, Stamenkovic I. Sarcomas: genetics, signalling, and cellular origins. Part 1: The fellowship of TET. *J Pathol*. 2007; 213:4–20. [PubMed: 17691072]

7. Kovar H. Dr. Jekyll and Mr. Hyde: The Two Faces of the FUS/EWS/TAF15 Protein Family. *Sarcoma*. 2011; 2011:837474. [PubMed: 21197473]
8. Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoyanov S, Mahamid J, Saha S, Franzmann TM, Pozniakovski A, Poser I, Maghelli N, Royer LA, Weigert M, Myers EW, Grill S, Drechsel D, Hyman AA, Alberti S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. 2015; 162:1066–1077. [PubMed: 26317470]
9. Schwartz JC, Wang X, Podell ER, Cech TR. RNA seeds higher-order assembly of FUS protein. *Cell Rep*. 2013; 5:918–925. [PubMed: 24268778]
10. Murakami T, Qamar S, Lin JQ, Schierle GS, Rees E, Miyashita A, Costa AR, Dodd RB, Chan FT, Michel CH, Kronenberg-Versteeg D, Li Y, Yang SP, Wakutani Y, Meadows W, Ferry RR, Dong L, Tartaglia GG, Favrin G, Lin WL, Dickson DW, Zhen M, Ron D, Schmitt-Ulms G, Fraser PE, Shneider NA, Holt C, Vendruscolo M, Kaminski CF, St George-Hyslop P. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron*. 2015; 88:678–690. [PubMed: 26526393]
11. Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, O'Meally R, Dignon GL, Conicella AE, Zheng W, Best RB, Cole RN, Mittal J, Shewmaker F, Fawzi NL. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J*. 2017; doi: 10.15252/embj.201696394
12. Sun Z, Diaz Z, Fang X, Hart MP, Chesi A, Shorter J, Gitler AD. Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol*. 2011; 9:e1000614. [PubMed: 21541367]
13. Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, Tycko R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*.
14. Wheeler JR, Matheny T, Jain S, Abrisch R, Parker R. Distinct stages in stress granule assembly and disassembly. *Elife*. 2016; doi: 10.7554/eLife.18413
15. Corden JL, Cadena DL, Ahearn JM Jr, Dahmus ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A*. 1985; 82:7934–7938. [PubMed: 2999785]
16. Allison LA, Moyle M, Shales M, Ingles CJ. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*. 1985; 42:599–610. [PubMed: 3896517]
17. Cagas PM, Corden JL. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins*. 1995; 21:149–160. [PubMed: 7777490]
18. Bienkiewicz EA, Moon Woody A, Woody RW. Conformation of the RNA polymerase II C-terminal domain: circular dichroism of long and short fragments. *J Mol Biol*. 2000; 297:119–133. [PubMed: 10704311]
19. Gibbs EB, Lu F, Portz B, Fisher MJ, Medellin BP, Laremore TN, Zhang YJ, Gilmour DS, Showalter SA. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat Commun*. 2017; 8:15233. [PubMed: 28497798]
20. Portz B, Lu F, Gibbs EB, Mayfield JE, Rachel Mehaffey M, Zhang YJ, Brodbelt JS, Showalter SA, Gilmour DS. Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat Commun*. 2017; 8:15231. [PubMed: 28497792]
21. Peti W, Page R. Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr Purif*. 2007; 51:1–10. [PubMed: 16904906]
22. Thomsen MC, Nielsen M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res*. 2012; 40:W281–287. [PubMed: 22638583]
23. Conicella AE, Fawzi NL. The C-terminal threonine of Abeta43 nucleates toxic aggregation via structural and dynamical changes in monomers and protofibrils. *Biochemistry*. 2014; 53:3095–3105. [PubMed: 24773532]
24. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExpASY server. *Methods Mol Biol*. 1999; 112:531–552. [PubMed: 10027275]

25. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM. Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature*. 2011; 480:268–272. [PubMed: 22037310]
26. Fawzi NL, Ying J, Torchia DA, Clore GM. Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nat Protoc*. 2012; 7:1523–1533. [PubMed: 22814391]
27. Berendsen HJC, Vanderspoel D, Vandrunen R. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput Phys Commun*. 1995; 91:43–56.
28. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*. 2008; 4:435–447. [PubMed: 26620784]
29. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun*. 2009; 180:1961–1972.
30. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR*. 2010; 48:13–22. [PubMed: 20628786]
31. Conicella AE, Zerbe GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure*. 2016; 24:1537–1549. [PubMed: 27545621]
32. Kato M, Lin Y, McKnight SL. Cross-beta polymerization and hydrogel formation by low-complexity sequence proteins. *Methods*. 2017; doi: 10.1016/j.ymeth.2017.06.011
33. Selo I, Negroni L, Creminon C, Grassi J, Wal JM. Preferential labeling of alpha-amino N-terminal groups in peptides by biotin: application to the detection of specific anti-peptide antibodies by enzyme immunoassays. *J Immunol Methods*. 1996; 199:127–138. [PubMed: 8982354]
34. Poitevin E, Wahl P. Study of the translational diffusion of macromolecules in beads of gel chromatography by the FRAP method. *Biophys Chem*. 1988; 31:247–258. [PubMed: 2466497]
35. Bastidas M, Gibbs EB, Sahu D, Showalter SA. A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts in Magnetic Resonance Part A*. 2015; 44:54–66.
36. Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci*. 2006; 15:2795–2804. [PubMed: 17088319]
37. Best RB, Zheng W, Mittal J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J Chem Theory Comput*. 2014; 10:5113–5124. [PubMed: 25400522]
38. Zerbe GH, Best RB, Mittal J. Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. *J Phys Chem B*. 2015; 119:14622–14630. [PubMed: 26498157]
39. Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett*. 2010; 104:190601. [PubMed: 20866953]
40. Deighan M, Bonomi M, Pfendtner J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *J Chem Theory Comput*. 2012; 8:2189–2192. [PubMed: 26588950]
41. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983; 22:2577–2637. [PubMed: 6667333]
42. Iglesias J, Sanchez-Martinez M, Crehuet R. SS-map: Visualizing cooperative secondary structure elements in protein ensembles. *Intrinsically Disord Proteins*. 2013; 1:e25323. [PubMed: 28516013]
43. Fawzi NL, Phillips AH, Ruscio JZ, Doucleff M, Wemmer DE, Head-Gordon T. Structure and dynamics of the Abeta(21-30) peptide from the interplay of NMR experiments and molecular simulations. *J Am Chem Soc*. 2008; 130:6145–6158. [PubMed: 18412346]
44. O’Nuallain B, Williams AD, Westermarck P, Wetzel R. Seeding specificity in amyloid growth induced by heterologous fibrils. *J Biol Chem*. 2004; 279:17490–17499. [PubMed: 14752113]

45. Stathopoulos PB, Scholz GA, Hwang YM, Rumfeldt JA, Lepock JR, Meiering EM. Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci.* 2004; 13:3017–3027. [PubMed: 15459333]
46. Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Molecular cell.* 2007; 26:867–881. [PubMed: 17588520]
47. Zhao DY, Gish G, Braunschweig U, Li Y, Ni Z, Schmitges FW, Zhong G, Liu K, Li W, Moffat J, Vedadi M, Min J, Pawson TJ, Blencowe BJ, Greenblatt JF. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature.* 2016; 529:48–53. [PubMed: 26700805]
48. Schuller R, Forne I, Straub T, Schreieck A, Texier Y, Shah N, Decker TM, Cramer P, Imhof A, Eick D. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular cell.* 2016; 61:305–314. [PubMed: 26799765]
49. Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA, Buratowski S. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular cell.* 2016; 61:297–304. [PubMed: 26799764]

Figure 2. Simulated conformational ensemble of an RNA pol II CTD subpeptide comprising residues 1927-1970 is primarily disordered except for partial helix structure near degenerate residue L1956

A) Experimental secondary chemical shifts ($\delta C\alpha - \delta C\beta$) and **B)** $^3J_{\text{HNH}\alpha}$ scalar coupling constants are in reasonable agreement with values derived from simulation, suggesting the forcefield does not overstabilize structure. Values are plotted as mean $\pm$ std. dev for experiment and mean $\pm$ SEM derived from 10 equal divisions of the converged ensemble. **C)** Although the DSSP algorithm predicts predominant random coil structure across residues 1927-1970, residues 1954-1956 sample partial helical structure in 25% of conformation ensembles using the same serial simulations conducted to compute NMR relaxation parameters (open symbols). Filled symbols are from PT-WTE simulations. **D)** A map of α -helical secondary structure created by binning adjacent stretches of DSSP-defined secondary structure shows helical structure in 13% of simulated ensembles for a contiguous span of 4 residues (x-axis) at positions 1954-1957 (y-axis) near L1956 degeneracy at heptad position 3 usually occupied by proline. Smaller populations of longer five-residue (2.6%) and six-residue (0.3%) α -helix are also observed near this position.

Figure 3. Heightened NMR relaxation reveals residue-specific binding of RNA polymerase II CTD 27-52 in complex with hnRNPA2 and TAF15 fibrils

mCherry-TAF15 LC (A) and mCherry-hnRNPA2 LC (B) hydrogels fibrils (left) are fractured into short “seeds” by sonication (right) as visualized by transmission electron microscopy. An increase in transverse relaxation (R_2) suggests CTD27-52 avidly binds to sonicated TAF15 LC fibrils (C, black) with higher affinity but similar residue-specificity as compared to sonicated hnRNPA2 LC fibrils (C, purple) where 9 mg/mL of sonicated hnRNPA2 fibrils are required to produce R_2 values comparable to those produced by adding 2 mg/mL of sonicated TAF15 fibrils. Neither addition of supernatant extracted from centrifuged sonicated hnRNPA2 and TAF15 fibrils (C–D, maroon) nor addition of monomeric mCherry-TAF15 LC (D, brown) show significant R_2 , confirming that observed R_2 arises from CTD interaction with high molecular weight fibrillar species. While CTD27-52 N2A (D, blue) binds TAF15 fibrils with similar avidity as wild-type, the CTD27-52 K5S variant (D, red) exhibits a ~30% reduction in R_2 , suggesting that electrostatic interactions mediated by lysines in the CTD contribute to complex formation. Positions of mutated asparagine (N2A, blue) and lysine (K5S, red) residues are labeled with filled circles; gray open circles refer to unmutated lysines. Data are plotted as values $\pm$ standard deviation. For RNA pol II WT, N2A, and K5S + TAF15 fibrils, the average of multiple R_2 trials and the corresponding standard deviations are plotted. See Figure S4 for all CTD variant and TAF15 R_2 trials. RNA pol II WT + 2 mg/mL sonicated TAF15 fibril data are plotted in panels both C and D for clarity.

Figure 4. Dark-state exchange saturation transfer (DEST) NMR probes the binding of CTD27-52 WT to sonicated TAF15 fibrils

A) Values of ¹⁵N R_2 calculated from tethered/direct-contact pseudo-two-state model (blue), but not the simplest two-state model (gray), best-fit to the observed ¹⁵N R_2 (black) and DEST data for CTD27-52 WT + 2 mg/mL of sonicated TAF15 fibrils. Small deviations between fitted and experimental data are expected due to the simple phenomenological nature of the model. **B)** Example residue-specific experimental DEST profiles (circles) as a function of offset from the ¹⁵N carrier frequency and applied saturation field show decreasing width towards the C-terminus of CTD27-52, consistent with lower saturation transfer; best-fit values are calculated from the tethered/direct-contact model (lines).

Figure 5. DEST-derived residue-specific binding parameters from a tethered/directly-bound model show N-terminal and central heptads of CTD27-52 mediate binding to TAF15 fibrils

A) Binding model for fibril-bound CTD27-52; p_M and p_B refer to monomeric and bound populations, respectively. Extremely high $R_2^{direct-contact}$, constrained to be the same for all residues, is consistent with values for high molecular weight assemblies. **B)** Residue-specific $R_2^{tethered}(i)$, the average ^{15}N transverse relaxation rate constant of each residue when tethered to the surface of fibrils (i.e., not in direct contact), is high (200 to 2000 s^{-1}) across the N-terminal and central sections due to contacts formed by nearby heptads and low at the acidic tail, implying faster motions due to long tether lengths. **C)** Equilibrium constant K_3 , the ratio of the population of direct-contact states to the population of tethered states, ranges from near zero at the acidic tail, which rarely binds TAF15 fibrils, to values of 0.3 to 1 in the first half of CTD27-52, which avidly binds.

Figure 6. Contributions of lysine residues and electrostatic interactions to TAF15 hydrogel binding by hydrogel fluorescence localization and FRAP experiments

A) Confocal slices of mCherry-TAF15 LC hydrogels incubated with 1 μ M solutions of Alexa Fluor 488 tagged RNA pol II CTD variants in 200 mM NaCl. Time points refer to time after incubating hydrogels with RNA pol II. WT and N2A in 200 mM NaCl show tighter binding and corresponding localization to the outer circumference of TAF15 hydrogels; K5S exhibits more uniform localization across the hydrogel diameter compared to WT and N2A. **B)** Modest differences in CTD WT localization are observed when TAF15 hydrogels are incubated in varying NaCl concentrations (see Figure S7 for quantification of fluorescence localization across multiple hydrogel trials for time point 37 hours). **C)** Fluorescent recovery after photobleaching experiments demonstrate that K5S diffuses more quickly, suggesting weaker binding to TAF15 fibrils. **D)** WT in 400 mM NaCl also diffuses more quickly than WT in 50 mM NaCl, but diffusion constants at 200 mM NaCl and 400 mM NaCl are comparable. Three FRAP trials were conducted and plotted as average values $\pm$ std. dev. Best fit parameters for a single exponential model for FRAP recovery were calculated for each individual experiment and translational diffusion coefficients were estimated (see Methods). **E)** Images from confocal slices during representative FRAP trials for each variant and salt condition tested.