JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Structural Model of the Tubular Assembly of the Rous Sarcoma Virus Capsid Protein

Jaekyun Jeon,[†,#] Xin Qiao,[†] Ivan Hung,[‡] Alok K. Mitra,[§] Ambroise Desfosses,[§] Daniel Huang,[†] Peter L. Gor'kov,[‡] Rebecca C. Craven,[∥] Richard L. Kingston,[§] Zhehong Gan,[‡] Fangqiang Zhu,[⊥] and Bo Chen*[,†]

[†]Department of Physics, University of Central Florida, Orlando, Florida 32816, United States

[‡]National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, United States

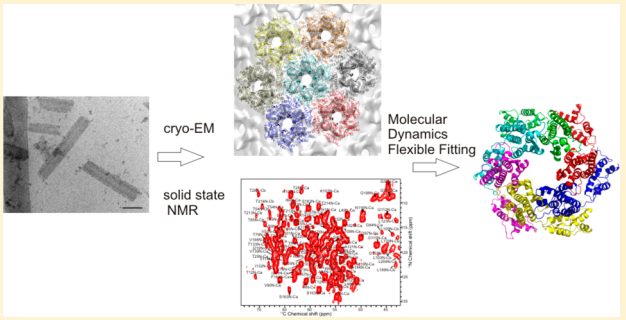
[§]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

[∥]Department of Microbiology and Immunology, Penn State University College of Medicine, Hershey, Pennsylvania 17033, United States

[⊥]Department of Physics, Indiana University−Purdue University Indianapolis, Indianapolis, Indiana 46202, United States

Supporting Information

ABSTRACT: The orthoretroviral capsid protein (CA) assembles into polymorphic capsids, whose architecture, assembly, and stability are still being investigated. The N-terminal and C-terminal domains of CA (NTD and CTD, respectively) engage in both homotypic and heterotypic interactions to create the capsid. Hexameric turrets formed by the NTD decorate the majority of the capsid surface. We report nearly complete solid-state NMR (ssNMR) resonance assignments of Rous sarcoma virus (RSV) CA, assembled into hexamer tubes that mimic the authentic capsid. The ssNMR assignments show that, upon assembly, large conformational changes occur in loops connecting helices, as well as the short 3_{10} helix initiating the CTD. The interdomain linker becomes statically disordered. Combining constraints from ssNMR and cryo-electron microscopy (cryo-EM), we establish an atomic resolution model of the RSV CA tubular assembly using molecular dynamics flexible fitting (MDFF) simulations. On the basis of comparison of this MDFF model with an earlier-derived crystallographic model for the planar assembly, the induction of curvature into the RSV CA hexamer lattice arises predominantly from reconfiguration of the NTD−CTD and CTD trimer interfaces. The CTD dimer and CTD trimer interfaces are also intrinsically variable. Hence, deformation of the CA hexamer lattice results from the variable displacement of the CTDs that surround each hexameric turret. Pervasive H-bonding is found at all interdomain interfaces, which may contribute to their malleability. Finally, we find helices at the interfaces of HIV and RSV CA assemblies have very different contact angles, which may reflect differences in the capsid assembly pathway for these viruses.



INTRODUCTION

The capsid of an infectious retroviral particle is formed by CA during the maturation process.[1] Despite limited sequence similarity, all orthoretroviral CAs share a common tertiary structure:[1] they consist of independently folded NTD and CTD,[2−9] each comprising mostly α helices, connected by a short and flexible linker. These CAs assemble into morphologically distinct capsids in different retroviruses, presumably by varying the basic curvature of the underpinning CA hexamer array as well as the location of the CA pentamers required for capsid closure.[10,11] Although the structures of many retroviral CAs have been determined,[2,9] direct structural analysis of authentic retroviral capsids at high resolution is extremely challenging due to their inherent polymorphism.[12−16] Consequently, our knowledge of CA in the assembled states is mostly derived from in vitro analysis. Specifically, HIV CA is the most studied, and structures of its hexameric assemblies were determined at atomic resolution.[17−21] In addition, an atomic-resolution model of isolated pentamers was obtained with mutant HIV CAs.[22] However, the dimer interface at helix 9 (H9) was disrupted in the mutant CA, depriving the model of the intercapsomer connection. CA pentamers were identified in $T = 1$ and 3 icosahedral particles assembled from RSV CA,[23,24] and a pseudoatomic model of the RSV CA hexamer was generated from crystallographic analysis.[25] However, in neither case is the resolution sufficient to determine the specifics of the intermolecular contacts involved in assembly. Moreover, the hexameric assemblies formed by the CAs of the N-tropic murine leukemia virus[7] and bovine leukemia virus[26] adopt

Received: November 18, 2016
Published: January 17, 2017

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

somewhat different quaternary arrangements from those of HIV CA. As shown by coarse-grained (CG) simulations[27−31] and experiments,[32−34] subtle changes at the intermolecular interface can have profound effects on capsid formation or alter the assembly pathway. Therefore, high-resolution structural and dynamical information is still needed to advance our understanding of orthoretroviral capsid assembly.

Aside from HIV CA, the 237-residue RSV CA is the most studied retroviral CA. ssNMR has been used to characterize various noncrystalline protein assemblies at a site-specific level,[35−44] including those formed by HIV CA.[45−52] Similar to HIV CA, RSV CA is twice the size of proteins routinely studied by ssNMR. In phosphate buffers, RSV CA assembles into a variety of structures that model the authentic capsid,[32] including tubes composed of capsid hexamers. Specifically, RSV CA tubes present an even tougher technical challenge for ssNMR studies than HIV CA tubes, as the RSV CA sequence is relatively enriched in the amino acids A, R, L, and P. In this Article, we present accurate assignments of congested ssNMR spectra that exploit the intrinsic correlation between NCACX and NCOCX spectra. This approach will be applicable to ssNMR studies of other large proteins. Combined with different labeling strategies, we achieved nearly complete de novo assignments of RSV CA (234 of 237 residues) in the tubular assembly.

Site-specific assignments of ssNMR chemical shifts (CSs) show that RSV CA can form tubes of various diameters with little change in backbone structure. The largest perturbations in CA structure and dynamics upon assembly are localized in the loops between helices and the short 3_{10} helix (residues 152−158) that borders the major homology region (MHR, residues 157−175). The interdomain linker of RSV CA appears to be statically disordered in the tubular assembly, in contrast to the dynamic interdomain linker observed for HIV CA tubular assemblies.[49]

Combining our ssNMR data with constraints derived from X-ray diffraction[25] and cryo-EM analyses, we obtained the first atomic-resolution structural model of the RSV CA tubular assembly by MDFF. Our model shows that the induction of curvature into the hexamer array is associated with large displacements at the NTD−CTD and CTD trimer interfaces. The model also suggests that the CTD dimer and trimer interfaces are intrinsically malleable, which may be a consequence of the pervasive H-bonding identified at all interfaces in the assembly. Hence, our modeling implicates the variable displacement of the CTD as the cause of structural polymorphism in RSV CA assemblies. In addition, helices at the interfaces exhibit different contact angles in the planar and tubular assemblies for HIV and RSV CA,[20,21] which suggests that they may asssemble via differing pathways.[29] Combined, our results provide valuable insights and intellectual guidance for further biophysical and biochemical studies of the assembly mechanism of retroviral CAs at a site-specific level.

MATERIALS AND METHODS

Methods for RSV CA protein expression and purification, as well as for in vitro tube assembly, were adaptations of previously published protocols,[32] as described in the Supporting Information. Details of transmission electron microscopy (TEM), ssNMR, and MDFF simulations are given in the Supporting Information.

RESULTS AND DISCUSSION

Near-Complete Resonance Assignment Combining Various Isotopic Labelings. RSV CA forms tubes of varying diameter (40−190 nm) in phosphate buffer solutions. Thick tubes (80−190 nm diameter, sample 1) were obtained when the protein solution was concentrated to 20 mg/mL in one step, before dialysis into phosphate buffer. Narrower tubes (40−90 nm diameter, sample 2) were obtained when the protein solution was concentrated to 20 mg/mL more gradually prior to dialysis. TEM images of negatively stained specimens are shown in Figure 1A and B. The tubes are multilayered, and

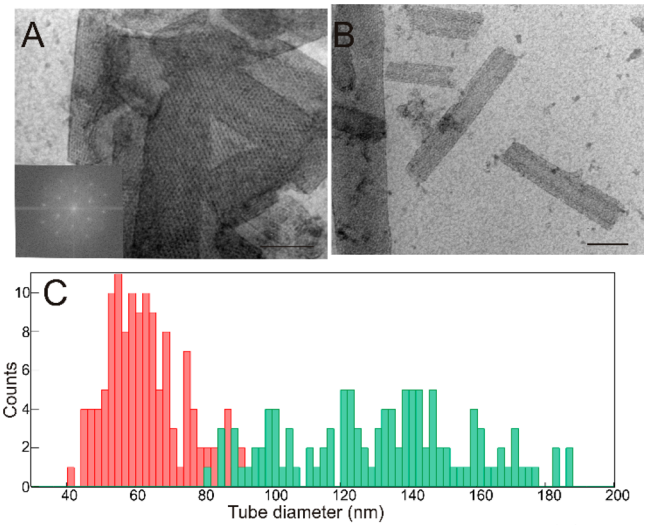


Figure 1. RSV CA tubular assemblies of various diameters. (A) Negatively stained TEM image of an assembly sample with mostly large tubes. The bottom left inset in (A) is the Fourier transformation of the tubes in the image. It shows sharp and discrete spots that can be indexed on a primitive hexagonal lattice. The dimensions of the lattice ($a = b = 9.6$ nm) are consistent with known dimensions of the RSV CA hexamers.[25] (B) Negatively stained TEM image of an assembly sample showing mostly narrower tubes. Panel (C) is the histogram of the two samples shown in (A) and (B), colored in green and red, respectively. The scale bars are 100 nm.

both samples exhibit considerable variations in tube diameter, as shown in Figure 1C. Despite the broad size distribution, both samples generate ssNMR spectra with sharp resonances comparable to those of HIV CA assemblies.[45,47−50] This indicates that the tubes have a regular structure, consistent with the periodicity apparent in the TEM images (inset, Figure 1A). Uniformly ^{13}C and ^{15}N labeled RSV CA (U-CA) assembly samples were prepared for NMR characterization. 3D NCACX, NCOCX, and CANCO spectra of sample 1 were acquired for sequential assignments, as listed in Table S1 in the Supporting Information. All resonances in 3D NCACX spectra are well-resolved, as shown in Figure 2B and C and Figure S1A. The average line width is ∼0.6 ppm along the ^{13}C dimension and 0.8 ppm along the ^{15}N dimension, shown by the 1D slices in the 2D NCA spectra in Figure 4A. However, 3D NCOCX spectra along the ^{15}N dimension between 116 and 120 ppm are too congested for direct residue-specific assignments (RSA), shown by the 2D planes in Figures 2A and S1A. To alleviate signal congestion, sparsely labeled samples were prepared by 1,3-^{13}C glycerol (1,3G-CA) and 2-^{13}C glycerol (2G-CA) expression.[53] The spectra of these samples are less congested,

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

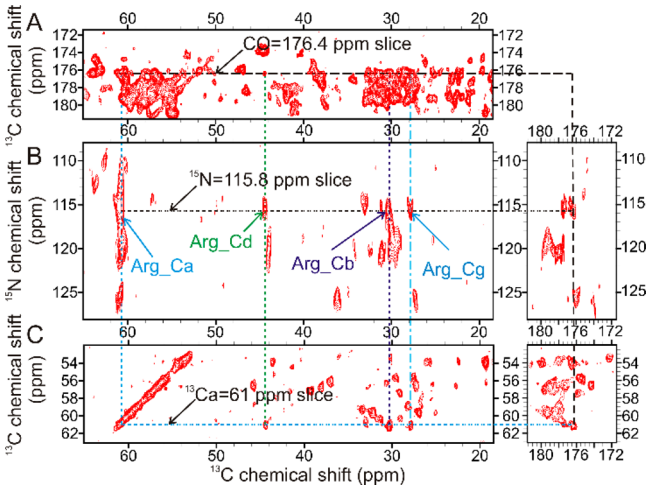


Figure 2. Assignments of congested NCOCX by exploiting well-resolved NCACX spectra. (A) The ^{15}N = 119.1 pm plane of the NCOCX spectra is highly congested, with a large number of resonances at CO frequency (vertical axis) 176.4 ppm, indicated by the horizontal dashed line. Panels (B) and (C) are the aliphatic and CO regions of the better resolved ^{13}C = 61 ppm and ^{15}N = 115.8 ppm planes of the NCACX spectra, respectively. They reveal intraresidue correlations between carbons and Ca of 61 ppm resonance in (B) or with the amide nitrogen of 115.8 ppm in (C). By inspection, only resonances at 44.5, 30.4, and 27.8 ppm are present in both NCACX and NCOCX spectra and correlated with CO = 176.4 ppm and Ca = 61 ppm. Therefore, they are assigned as an R according to the characteristic resonances of amino acids.

with line widths comparable to those of the U-CA sample, as shown in Figure S1B.

For proteins of comparable size and spectral quality, sequential assignments of the majority (~75%) of the residues were previously achieved by connecting the RSA in 3D spectra using a "backbone walk" process.[47−51,53] The principle of the backbone walk is illustrated in Figure S2, where resonances of adjacent residues from T101 to G112 of RSV CA are connected in a pairwise fashion by matching the protein sequence and the signal transfer pathway in 3D spectra. However, because the RSV CA sequence is enriched in residues such as A, R, L, and P, even the NCOCX spectra of sparsely labeled samples are not sufficiently resolved for unambiguous and accurate RSA, as shown by Figure S1B. As a consequence, the number of residues that could be sequentially assigned was limited to 82, even when employing the computer-automated sequential assignment program MCASSIGN2.[54] In fact, accurate RSA in congested spectra is a universal challenge for ssNMR studies of large proteins.

To resolve this issue, we exploit the intrinsic correlation between NCACX and NCOCX spectra to enhance the efficiency and accuracy of RSA. Normally, owing to the relatively large Cα CS dispersion (~25−30 ppm), resonances in NCACX spectra are better resolved than those in NCOCX spectra, as shown in Figures 2 and S1. Although all carbons in the same residue are correlated with a different amide nitrogen in NCACX and NCOCX spectra, they exhibit identical carbon resonances in both. Therefore, the well-resolved NCACX spectra can be used to guide the RSA of the congested NCOCX spectra. The implementation of this strategy is shown in Figure 2. As shown in Figure 2A, a large number of resonances are present at CO = 176.4 ppm (vertical axis) in the extracted ^{15}N = 119.1 ppm plane of the NCOCX spectra. Among all Cα resonances (signals higher than 50 ppm) at this CO frequency in this ^{15}N plane, we focus on assignment of the Cα signal at 61 ppm. What needs to be determined is which of the side-chain carbons (signals lower than ~45 ppm) arise from the same residue. Without additional constraints, the problem is intractable as there are too many candidates due to the congested signals. However, the intraresidue correlation between all carbons is recorded in NCACX spectra. Therefore, the side-chain carbon signals for the 61 ppm of Cα resonance at CO = 176.4 ppm in the ^{15}N = 119.1 pm plane of the NCOCX spectra can be identified unambiguously by inspecting the Cα = 61 ppm plane of the NCACX spectra, which is well-resolved, as shown in Figure 2B. By inspection, only resonances at 44.5, 30.4, and 27.8 ppm along the ^{15}N = 115.8 ppm slice of the $^{13}C\alpha$ = 61 ppm plane in NCACX spectra exhibit a correlation with the CO resonance at 176.4 ppm. This suggests that these resonances arise from the same residue with the Cα signal at 61 ppm and CO signal at 176.4 ppm. Inspection of the corresponding ^{15}N = 115.8 ppm plane of NCACX spectra confirms this intraresidue correlation, as shown in Figure 2C. The residue is assigned as an R according to the characteristic resonances of amino acids. With this approach, the accuracy and certainty of RSA were greatly improved, and the sequentially assigned residues were increased to 128. This strategy is quite general and could be applied to achieve RSA of other proteins with congested 3D ssNMR spectra.

Further improvement of the sequential assignments requires the discrimination of similar residues to reduce ambiguity in RSA. Two samples were prepared with selective ^{13}C and ^{15}N labeling of either all L or all R residues. This helped resolve the ambiguity between R, L, P and K, as shown in Figure 3A and B.

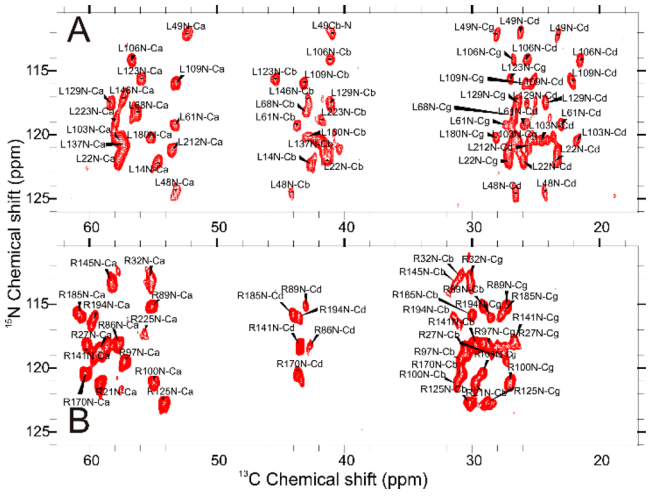


Figure 3. (A) 2D NCACX spectra of the RSV CA tubular assembly sample with ^{13}C, ^{15}N labeling at all L residues and R residues in (B). Spectra acquired at 14.1 T with MAS at 13.5 kHz. The spectra were recorded with 50 ms dipolar-assisted rotational resonance (DARR) mixing to generate the cross-peaks and processed with 30 Hz Gaussian broadening along each dimension.

Subsequently, the program MCASSIGN2 was used to automate sequential assignments on the basis of the refined RSA.[54] The uniqueness and correctness of the assignments were evaluated by comparing 50 independent sequential assignment results, as shown by Figure S3A. In summary, 197 residues were sequentially assigned with 100% consistency. Another 23 residues were assigned with <100% consistency due to similar

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

resonances and residue types. However, this inconsistency does not influence structural inferences based on the assignments. Moreover, two sets of resonances were found for both T207 and Q158 (the latter located in the middle of MHR), but only one set of signals were sequentially assigned, as shown in Figure S3B and C. There are another five residues (P91, R145, P149, P152, and L180) with weak or missing signals in either NCACX or NCOCX, which lead to reduced consistency in assignments as shown in Table S2.

Only rigid segments contribute to these spectra, because anisotropic dipolar interaction was employed to record spatial correlations between atoms. Conversely, scalar interactions can be utilized to selectively visualize dynamic segments, and hence an INEPT-TOBSY spectrum of the U-CA sample was recorded,[55] as shown in Figure S3D. Seven residues were identified in this spectra, and their CSs indicate random-coil conformation. Their residue types coincide with the last seven residues of the protein, and given the remaining assignments, these signals must originate from this region.

Collectively, we sequentially assigned 234 residues of the 237-residue RSV CA, as shown by Table S2. Three residues (A150, A229, and P230) are missing in all NMR spectra. Given the proximity of A229 and P230 to the flexible C-terminus, it is very likely that their dynamics falls in the invisible intermediate regime. Considering the rigidity of its adjacent residues, the absence of A150 is probably associated with static structural disorder which causes diminished resonance intensity.

Site-Specific Changes of Local Structures and Dynamics upon Tubular Assembly. The narrow line width and uniqueness of the assignments for most of the residues indicate that RSV CA adopts a uniform structure, in spite of the varying diameter of the tubes in the sample. Furthermore, comparison of samples 1 and 2 informs us about the correlation between molecular structure and the curvature of RSV CA hexameric lattice, as their average tube diameters differ by a factor of 2, as shown in Figure 1E. Although detailed interpretation must await the sequential assignments of sample 2, overlay of the 2D NCA spectra shows identical resonances and line widths for nearly all α carbon and amide nitrogen nuclei, as shown in Figure 4A. This suggests that large tubular curvature variations cause little significant perturbation of the secondary structure and dynamics of RSV CA, as previously observed for HIV CA assemblies.[47,52]

Regions experiencing local structural perturbation upon assembly can be identified by comparing the RSV CA CSs in the tubular assembly with those in the soluble state.[2] The residue specific ΔCS values are shown in Figures 4B and 6A, and the overall distribution of these values in Figure S3E. Overall, 147 residues out of the 237-residue protein display $\Delta CS \leq 0.5$ ppm. On the basis of our assignments, quantitative dynamics and secondary structural information at a site-specific level was derived using program TALOSN,[56] shown in Figure 4C. As a comparison, the TALOSN prediction based on solution NMR CSs of monomeric RSV CA is shown in Figure S4B,[2] which gave excellent agreement with the experimental values[2] and validates the procedure. TALOSN also predicts the change of torsion angles upon assembly, at a site-specific level, shown in Figure S4A. Collectively, the analysis suggests that RSV CA retains nearly identical secondary structures and dynamics upon assembly, as previously observed for HIV CA.[47−52]

Nonetheless, the analysis highlights large local structural rearrangements upon assembly in four regions. First, large

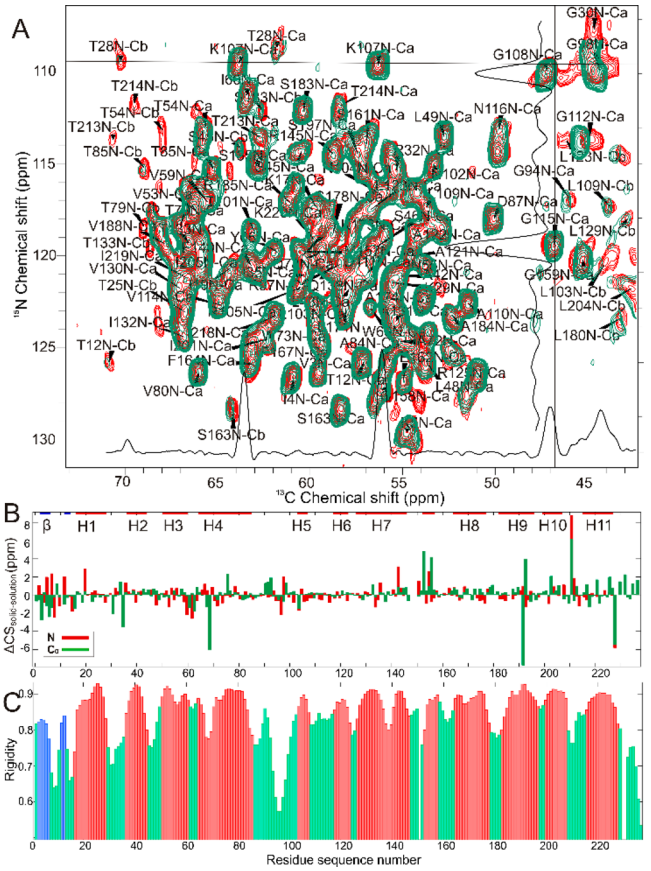


Figure 4. Dynamics and secondary structure of the tubular RSV CA assembly derived from ssNMR assignments. (A) Assignments of the 2D NCACX spectrum of the tubular assembly of the U-CA. 1D slices at the denoted positions along ^{13}C and ^{15}N positions are shown at the bottom and right. The red spectrum was recorded with the sample showing tube diameter between 40 and 80 nm, with the histogram of the tube diameter colored in red in Figure 1E. The spectrum was acquired with 2.5 ms DARR mixing and 13.5 kHz MAS at 21 T. The blue spectrum was recorded with the sample showing tube diameter between 80 and 190 nm, with the histogram of the tube diameter colored in green in Figure 1E. The spectrum was acquired with 6 ms DARR mixing and 13.5 kHz MAS at 14 T. Spectra were processed with 40 Hz exponential and 60 Hz Gaussian broadening functions along both dimensions. (B) Chemical shift differences of Ca (green) and amide nitrogen (red) sites of RSV CA in the tubular assembly and the soluble state. (C) Site-specific dynamics and secondary structure of the RSV CA tubular assembly derived from our NMR assignments by TALOSN. Residues in β strand, random coil, and α helices are colored blue, green, and red, respectively. The height of the bar represents the S^2 order parameter. Residues with $S^2 < 0.6$ are highly dynamic.

changes are observed at the N-terminus. Unfortunately, meaningful comparison of the solution and ssNMR data in this region is impossible, as the solution NMR observations were made on a CA variant carrying an N-terminal polyhistidine tag. The presence of the tag disrupts formation of the naturally occurring β-hairpin structure[57] that is critical for assembly. The TALOSN predictions demonstrate that this region forms a structured β-hairpin in the tubular assembly, similar to HIV CA.[21,22,58,59] Second, the long loop between helices H4 and H5 displays a large change of torsion angles, shown in Figure S4A, but retains its high mobility in the assembled state. Third, the interdomain linker (residues 147−151) undergoes both a large change of torsion angles and an

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

attendant loss of flexibility. This loss of dynamics agrees with the observation of weak or missing signals for residues R145, P149, A150, and P152 in 3D spectra. Collectively the data indicate that the RSV CA interdomain linker adopts statically disordered states in the final tubular assembly, associated with variations in the relative positioning of the CTDs and NTDs. Finally, the 3_{10} helix that initiates the CTD also exhibits large changes in torsion angles and acquires α helical character upon assembly, as shown in Figures 4C and S4A.

The structural perturbations that accompany RSV CA assembly differ in several notable ways from those that accompany HIV CA assembly. First, the β-hairpin in the RSV CA tubular assembly exhibits a rigidity comparable to the α helices in the body of the protein, while that in the HIV CA tubular assembly appears more flexible.[49] Second, in contrast to the statically disordered interdomain linker in RSV CA, this linker region was shown to be mobile in the HIV CA tubular assembly by ssNMR dipolar tensor measurements.[49] This is in agreement with our TALOSN prediction based on the CS of HIV CA, as shown in Figure S4C. Third, while the 3_{10} helix preceding the MHR in HIV CA also undergoes large structural rearrangements upon tubular assembly, it becomes more extended[46,47,50] and does not acquire α helical character, as it does in RSV CA.

Derivation of a Hybrid Structural Model by MDFF. *MDFF Model Construction.* As shown in Figures S4A and 4B, a number of additional regions exhibit large changes in predicted torsion angles, most of which are the loops linking adjacent helices. To better understand the implications of these changes, an assembly model comprising seven hexamers on the tube surface was constructed by MDFF simulations that combine ssNMR constraints with cryo-EM data, as shown in Figures 5 and 6. Complementary to the torsion angles derived from ssNMR CS, the cryo-EM image reconstruction of an RSV CA tube at 24 Å resolution[60] determines the approximate locations of individual protein domains in the assembly, as shown in Figure 5. The pseudoatomic model of the RSV CA hexamer derived from crystallographic analysis (PDB entry 3TIR) was used as the initial structure in our MDFF simulations, and domain restraints on each NTD hexamer and CTD dimer were applied in the early stages of the simulations (see Supporting Information) to preserve the internal conformations of each domain as well as the NTD−NTD and CTD−CTD dimer interfaces. These domain restraints were removed in the last 5 ns of the MDFF simulation to allow further relaxation of the domain conformations under only the restraints imposed by the cryo-EM density map and the NMR CS data. Figure 6A shows a CA monomer in our final model. When compared to the monomer structure defined in 3TIR, the 42 monomers in our final model have RMSD (Cα atoms) values ranging from 3.4 to 4.5 Å. In contrast, the 42-monomer conformations share a higher degree of similarity among themselves: when compared to the average conformation over all the monomers, the average RMSD of the individual monomers is 1.5 Å, which is smaller than the value of 2.6 Å for the model of HIV CA tubular assembly (PDB entry 3J34)[20] obtained with a cryo-EM map at much higher resolution.

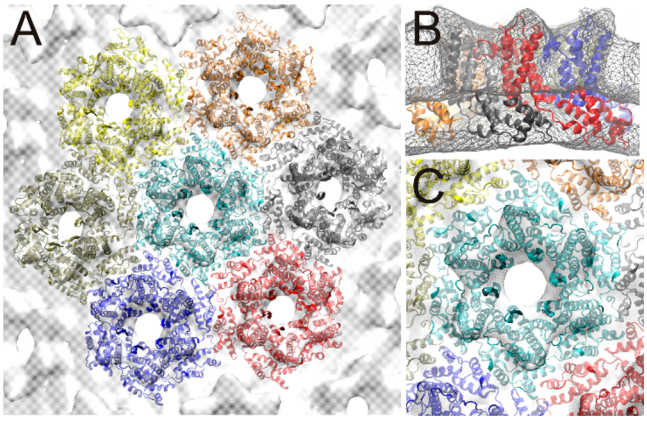


Figure 5. Atomic model obtained from MDFF displayed along with the cryo-EM electron density map. (A) Seven hexamers on the tube surface. (B) Side view of a single hexamer, with each of the six monomers shown in a different color. (C) Top view of the central hexamer with its six neighbors partly shown.

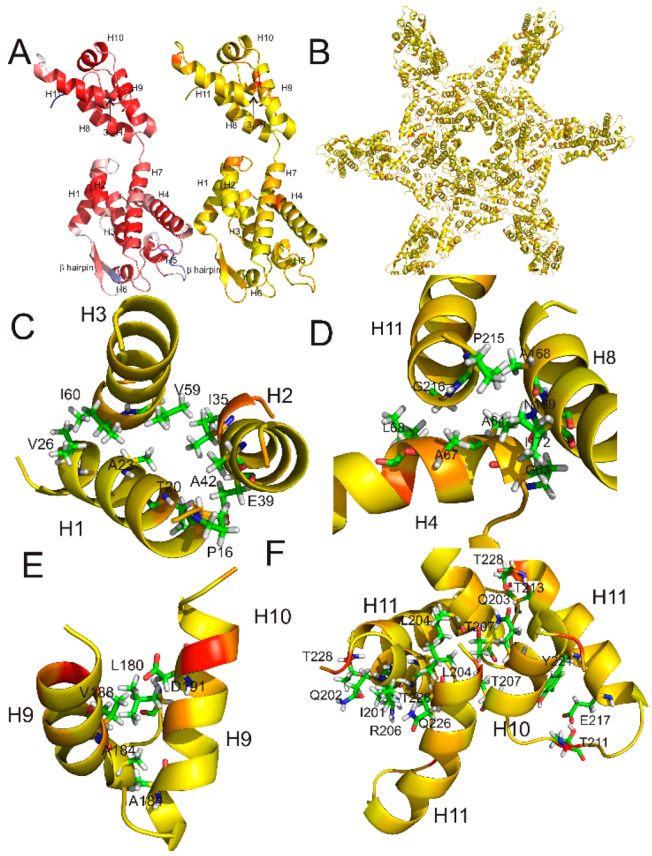


Figure 6. Structural model of the RSV CA tubular assembly. (A) Monomer of the RSV CA in the tubular assembly. The left is color coded according to the backbone dynamics (from red to blue for rigid to mobile), and the right is color coded according to largest changes of the backbone (Ca, N, CO) ΔCS of each residue (from yellow to red for the least to the largest changes), indicating changes of backbone structure relative to that in the soluble state. (B) RSV CA hexamer with adjacent subunits on the tube, color coded in the same manner as the left panel of (A). (C−F) Close-up views of the NTD−NTD, NTD−CTD, dimeric, and trimeric interfaces, respectively. Only those helices involved in close contacts are labeled. Backbones are color coded in the same manner as the left panel of (A). Residues inferred at close contact at each interface are shown in stick mode.

MDFF Model Validation. The final model shows that the RSV CA tube is stabilized by four intermolecular interfaces, similar to those previously identified in RSV and HIV CA assemblies,[19,20,22−25,58] as shown in Figure 6C−F. The NTD−

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

Table 1. Crossing Angles of Helices at Each Interface (Angles were measured in pymol by anglebetweenhelices.py script; standard deviations were calculated by averaging all subunits in the model)

angle	RSV CA tube, this work	RSV CA flat sheet, 3TIR.pdb	HIV CA tube, 3J34.pdb	HIV CA flat sheet, 4XFX.pdb
H1 and H2	146.2° ± 2.3°	158.9° ± 0°	152.6° ± 5.5°	149.5° ± 0°
H2 and H3	157.4° ± 1.7°	152.9° ± 0°	145.0° ± 6.6°	138.0° ± 0°
H4 and H8	85.7° ± 4.3°	35.6° ± 0°	74.8° ± 9.2°	71.5° ± 0°
H9 and H9	61.2° ± 6.2°	52.9° ± 0°	68.6° ± 11.8°	66.2° ± 1.1°
H10 and H11	156.3° ± 7.3°	130.0° ± 26.9°[a]	148.8° ± 22.8°	148.7° ± 17.4°[a]
H10 and H10	65.8° ± 5.4°	52.3° ± 0°	39.9° ± 14.3°	62.3° ± 0°

[a]There are three pairs of H10 and 11 at each trimer interface. They do exhibit different crossing angles, but the crossing angles between the same pair of H10 and 11 at different trimer interface in the crystal model are identical.

NTD interface, involving the first three helices of CA, bundles NTDs together into individual hexamers, which are further stabilized by the NTD−CTD interface between helix H4 of each NTD and helices H8 and H11 from the CTD of the neighboring subunit within the hexamer. The CTDs sit below the NTD-hexamers and form dimeric (helix H9) and trimeric (helices H10 and H11) interactions with their immediate neighbors, thereby creating the hexamer lattice.

Close inspection of the interfaces in the MDFF model shows that the interacting residues exhibit large CS changes upon assembly, summarized in Table S3. This consistency is highlighted by Figure 6C−F, color coded according to the maximum backbone ΔCS (Cα, CO, N) of each residue. Furthermore, the interfaces in our model agree with prior genetic and structural analyses of RSV CA assembly. In Figure 6C, A38 is located right in the center of the three-helix bundle of the NTD−NTD interface. Its mutation to V with a longer side chain will enhance the binding between all the hydrophobic residues surrounding this interface, which agrees with the observed promotion effect of A38 V mutant.[34] The NTD−CTD interface consists of the capping of G63 and A64 at H4 by N169 and I172 at H8. This is consistent with earlier reports that mutations of residues around this interface, such as L171 V, R170Q, and F167Y, are detrimental to core formation and infectivity, some of which can be restored by a second mutation at the NTD, such as P65Q.[33,34] Moreover, the N169 side chain forms H-bond with the backbone amide group of A64, shown in Figure S7B, exactly as previously predicted from low-resolution crystallographic analysis.[25] The dimer interface is stabilized by interactions between D191 and V188 across neighboring H9 and L180, and those between W153, L180, and V188, as well as two A184 facing each other, as shown in Figure 6E. This description matches with a previous crystallographic structure of the dimer[61] and is consistent with recent mutagenesis results that show mutations of W153, L180, and V188 are detrimental to both in vivo viral core formation and in vitro capsid assembly.[62]

Because the overall CS differences associated with tubular assembly of CA are quite small, it is important to establish that the ssNMR constraints actually impart meaningful information to the final model. To address this issue, we repeated the MDFF simulations, replacing the torsion angles derived from ssNMR of the tubular CA assembly with those for soluble RSV CA. Overall, the final model obtained using solution NMR constraints exhibits 3.074 Å RMSD from the model obtained with our ssNMR constraints. However, we find large differences in the intermolecular interfaces that contradict prior experimental observations, indicating that the model is incorrect. For example, helix H9 forms a large crossing angle at the dimer interface, as shown in Figure S6A. This places residues A184 at a distance of 8.5 Å from each other, in disagreement with the high-resolution crystal structure of the dimer.[61] Moreover, the model obtained using solution NMR constraints does not exhibit consistent contacts at the NTD−CTD interface. In addition, a number of contacts present in the model obtained with ssNMR constraints are absent in the model obtained using solution NMR constraints, including those between W153, V188, and L180, which were confirmed to be critical for both in vivo and in vitro capsid assembly.[62] Collectively, this suggests that the few localized regions experiencing large ΔCS upon assembly identified in ssNMR data do transmit critical structural information and inform the 3D model of the assembled state.

MDFF Model Predictions: Induction of Curvature, Polymorphism, and Differences between HIV and RSV CA Assemblies. The mechanism for induction of curvature into the hexamer array is suggested by comparing the MDFF model of the hexamer tubes with the crystallographic model of the hexamer sheets.[25] The induction of curvature requires reconfiguration of intermolecular interfaces, because the planar array has exact hexagonal symmetry, which must be relaxed to create the tube. Table 1 summarizes the crossing angles of contacting helices at interfaces in the RSV planar assembly (PDB ID: 3TIR)[25] and our tubular model. Considering the mean crossing angles, the NTD−CTD and CTD trimer interfaces differ significantly in the two assemblies, while the NTD−NTD and CTD−CTD dimer interfaces are relatively conserved. Additionally, we calculated the variance in the crossing angles for the interfaces in the MDFF model (Table 1). The NTD−NTD interface exhibits relatively small variations. In contrast, the variances observed for the CTD dimer and CTD trimer interfaces are 3−4-fold larger. Collectively these observations suggest that lattice curvature arises almost entirely from the variable displacement of the CTDs that surround each hexameric turret, as well as that CTD−CTD dimerization and trimerization interfaces involved in connecting the turrets are somewhat plastic. It is important to note that the 3_{10} helix that initiates the CTD undergoes some assembly-dependent reconfiguration, and hence the CTD does not function as an entirely rigid entity.

The ability of the interfaces to reconfigure may be explained by the pervasive hydrogen-bonding interactions observed at the interfaces in the MDFF model, as shown in Figure S7. This is consistent with prior structural analysis of HIV CA hexamer arrays.[21] Within the NTD−CTD interface, we highlight the formation of a hydrogen bond (H-bond) between the N169 side chain and the backbone amide group of A64, as shown in Figures 6D and S7B. No explicit H-bond involving N169 was present in the model that initiated our MDFF simulations. However, this H-bond was predicted to occur based on

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

crystallographic analysis of domain motions observed in the planar hexamer array.[25] The CTD−CTD dimer interface is very similar to that observed in the low-pH crystal structure of the dimer (PDB entry 3G21),[61] with only 2.938 Å RMSD from 3G21. However, the H-bond between D179 and D191 in 3G21 is disrupted in our model, due to the deprotonation of the side chains, as shown in Figure S6B.

We applied the same interface analysis to HIV CA hexamers, arranged in planar arrays (PDB entry 4XFX) and tubular assembly (PDB entry 3J34),[20,21] as shown in Table 1. Similar to RSV CA, HIV CA assemblies have a well-conserved NTD−NTD interface, which does not change significantly between tube and sheet and shows relatively low variance in helical crossing angles in either case. The most striking difference between RSV and HIV CA occurs at the NTD−CTD interface, where transition from sheet to tube is associated with virtually no mean change in crossing angle for HIV and a very large mean change for RSV. This may reflect the relative paucity of direct contacts at the NTD−CTD interface in RSV. The mean helix crossing angles at all interfaces differ between HIV and RSV CA, which may have implications for assembly. Recent coarse-grained simulations of HIV CA assembly showed that a 10° variation of the crossing angles altered the balance of assembly intermediates, shifting the predicted HIV CA assembly pathway from trimer-dominant to tetramer-dominant.[29,30]

CONCLUSION

In this Article, we demonstrate the NMR assignment strategy correlating signals seen in NCACX and NCOCX experiments to enable accurate RSA of relatively large proteins. When combined with nonuniform labeling strategies, we achieved near-complete assignments of the 237-residue RSV CA in the tubular assembly. Localized structural perturbations upon RSV CA assembly are concentrated in the loops between helices, the interdomain linker region, and the 3_{10} helix preceding the MHR. MDFF simulations were used to develop a structural model for the tubular assembly, effectively combining the ssNMR data with cryo-EM based constraints. The MDFF model shows that the induction of curvature into the hexamer array is associated with structural rearrangements at the NTD−CTD and CTD trimer interfaces. In addition, the model suggests that the CTD−CTD dimerization and trimerization interfaces are inherently malleable, perhaps due to the pervasive H-bonding observed at these interfaces.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b11939.

Summary of sample preparation; acquisition of ssNMR spectra; results of ssNMR assignments; cryo-EM conditions and MDFF simulations (PDF)

AUTHOR INFORMATION

Corresponding Author

*bo.chen@ucf.edu

ORCID

Fangqiang Zhu: 0000-0001-6202-2262

Bo Chen: 0000-0001-5589-2075

Present Address

#Laboratory of Chemical Physics, NIDDK, NIH, Maryland, MD.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge discussions and suggestions from Robert Tycko and the CA plasmid from Matthew England and consultation on protein purification and assembly. This work is supported by the AFOSR YIP award, funding no. FA9550-13-1-0150, the National High Magnetic Field Laboratory through NSF DMR-1157490, and the State of Florida.

REFERENCES

(1) Vogt, V. Retroviral virions and genomes. In *Retroviruses*; Coffin, J. M., Hughes, S. H., Varmus, H. E., Eds.; Cold Spring Harbor Press: New York, 1997; pp 22−70.

(2) Campos-Olivas, R.; Newman, J. L.; Summers, M. F. *J. Mol. Biol.* **2000**, *296*, 633−649.

(3) Cornilescu, C. C.; Bouamr, F.; Yao, X.; Carter, C.; Tjandra, N. *J. Mol. Biol.* **2001**, *306*, 783−797.

(4) Gamble, T. R.; Yoo, S.; Vajdos, F. F.; von Schwedler, U. K.; Worthylake, D. K.; Wang, H.; McCutcheon, J. P.; Sundquist, W. I.; Hill, C. P. *Science* **1997**, *278*, 849−853.

(5) Gitti, R. K.; Lee, B. M.; Walker, J.; Summers, M. F.; Yoo, S.; Sundquist, W. I. *Science* **1996**, *273*, 231−235.

(6) Khorasanizadeh, S.; Campos-Olivas, R.; Summers, M. F. *J. Mol. Biol.* **1999**, *291*, 491−505.

(7) Mortuza, G. B.; Haire, L. F.; Stevens, A.; Smerdon, S. J.; Stoye, J. P.; Taylor, I. A. *Nature* **2004**, *431*, 481−485.

(8) Worthylake, D. K.; Wang, H.; Yoo, S. H.; Sundquist, W. I.; Hill, C. P. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 85−92.

(9) Deshmukh, L.; Schwieters, C. D.; Grishaev, A.; Ghirlando, R.; Baber, J. L.; Clore, G. M. *J. Am. Chem. Soc.* **2013**, *135*, 16133−16147.

(10) Ganser, B. K.; Li, S.; Klishko, V. Y.; Finch, J. T.; Sundquist, W. I. *Science* **1999**, *283*, 80−83.

(11) Li, S.; Hill, C. P.; Sundquist, W. I.; Finch, J. T. *Nature* **2000**, *407*, 409−413.

(12) Briggs, J. A. G.; Gruenewald, K.; Glass, B.; Foerster, F.; Kraeusslich, H. G.; Fuller, S. D. *Structure* **2006**, *14*, 15−20.

(13) Briggs, J. A. G.; Wilk, T.; Welker, R.; Krausslich, H.-G.; Fuller, S. D. *EMBO J.* **2003**, *22*, 1707−1715.

(14) Benjamin, J.; Ganser-Pornillos, B. K.; Tivol, W. F.; Sundquist, W. I.; Jensen, G. J. *J. Mol. Biol.* **2005**, *346*, 577−588.

(15) Frank, G. A.; Narayan, K.; Bess, J. W.; Del Prete, G. Q.; Wu, X.; Moran, A.; Hartnell, L. M.; Earl, L. A.; Lifson, J. D.; Subramaniam, S. *Nat. Commun.* **2015**, *6*, 5854.

(16) Yu, Z.; Dobro, M. J.; Woodward, C. L.; Levandovsky, A.; Danielson, C. M.; Sandrin, V.; Shi, J.; Aiken, C.; Zandi, R.; Hope, T. J.; Jensen, G. J. *J. Mol. Biol.* **2013**, *425*, 112−123.

(17) Ganser-Pornillos, B. K.; Cheng, A.; Yeager, M. *Cell* **2007**, *131*, 70−79.

(18) Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M. *Cell* **2009**, *137*, 1282−1292.

(19) Byeon, I.-J. L.; Meng, X.; Jung, J.; Zhao, G.; Yang, R.; Ahn, J.; Shi, J.; Concel, J.; Aiken, C.; Zhang, P.; Gronenborn, A. M. *Cell* **2009**, *139*, 780−790.

(20) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. *Nature* **2013**, *497*, 643−646.

(21) Gres, A. T.; Kirby, K. A.; KewalRamani, V. N.; Tanner, J. J.; Pornillos, O.; Sarafianos, S. G. *Science* **2015**, *349*, 99−103.

(22) Pornillos, O.; Ganser-Pornillos, B. K.; Yeager, M. *Nature* **2011**, *469*, 424−7.

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013

(23) Cardone, G.; Purdy, J. G.; Cheng, N.; Craven, R. C.; Steven, A. C. *Nature* **2009**, *457*, 694−U3.
(24) Hyun, J.-K.; Radjainia, M.; Kingston, R. L.; Mitra, A. K. *J. Biol. Chem.* **2010**, *285*, 15056−15064.
(25) Bailey, G. D.; Hyun, J.-K.; Mitra, A. K.; Kingston, R. L. *J. Mol. Biol.* **2012**, *417*, 212−223.
(26) Obal, G.; Trajtenberg, F.; Carrion, F.; Tome, L.; Larrieux, N.; Zhang, X.; Pritsch, O.; Buschiazzo, A. *Science* **2015**, *349*, 95−98.
(27) Chen, B.; Tycko, R. *Biophys. J.* **2011**, *100*, 3035−3044.
(28) Grime, J. M. A.; Voth, G. A. *Biophys. J.* **2012**, *103*, 1774−1783.
(29) Qiao, X.; Jeon, J.; Weber, J.; Zhu, F.; Chen, B. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 2353−2367.
(30) Qiao, X.; Jeon, J.; Weber, J.; Zhu, F.; Chen, B. *Data in Brief* **2015**, *5*, 506−512.
(31) Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A. *Nat. Commun.* **2016**, *7*, 11568.
(32) Purdy, J. G.; Flanagan, J. M.; Ropson, I. J.; Craven, R. C. *J. Mol. Biol.* **2009**, *389*, 438−451.
(33) Bowzard, J. B.; Wills, J. W.; Craven, R. C. *J. Virol.* **2001**, *75*, 6850−6856.
(34) Butan, C.; Lokhandwala, P. M.; Purdy, J. G.; Cardone, G.; Craven, R. C.; Steven, A. C. *J. Virol.* **2010**, *84*, 6377−6386.
(35) Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. *Science* **2008**, *319*, 1523−1526.
(36) Loquet, A.; Sgourakis, N. G.; Gupta, R.; Giller, K.; Riedel, D.; Goosmann, C.; Griesinger, C.; Kolbe, M.; Baker, D.; Becker, S.; Lange, A. *Nature* **2012**, *486*, 276−279.
(37) Vasa, S.; Lin, L.; Shi, C.; Habenstein, B.; Riedel, D.; Kuehn, J.; Thanbichler, M.; Lange, A. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E127−E136.
(38) Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R. *Cell* **2013**, *154*, 1257−1268.
(39) Sharma, M.; Yi, M.; Dong, H.; Qin, H.; Peterson, E.; Busath, D. D.; Zhou, H.-X.; Cross, T. A. *Science* **2010**, *330*, 509−512.
(40) Hu, F.; Luo, W.; Hong, M. *Science* **2010**, *330*, 505−508.
(41) Sborgi, L.; Ravotti, F.; Dandey, V. P.; Dick, M. S.; Mazur, A.; Reckel, S.; Chami, M.; Scherer, S.; Huber, M.; Boeckmann, A.; Egelman, E. H.; Stahlberg, H.; Broz, P.; Meier, B. H.; Hiller, S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 13237−13242.
(42) Jehle, S.; Rajagopal, P.; Bardiaux, B.; Markovic, S.; Kuehne, R.; Stout, J. R.; Higman, V. A.; Klevit, R. E.; van Rossum, B.-J.; Oschkinat, H. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1037−U1.
(43) Cady, S. D.; Schmidt-Rohr, K.; Wang, J.; Soto, C. S.; DeGrado, W. F.; Hong, M. *Nature* **2010**, *463*, 689−U127.
(44) Wang, S.; Munro, R. A.; Shi, L.; Kawamura, I.; Okitsu, T.; Wada, A.; Kim, S.-Y.; Jung, K.-H.; Brown, L. S.; Ladizhansky, V. *Nat. Methods* **2013**, *10*, 1007−12.
(45) Byeon, I.-J. L.; Hou, G.; Han, Y.; Suiter, C. L.; Ahn, J.; Jung, J.; Byeon, C.-H.; Gronenborn, A. M.; Polenova, T. *J. Am. Chem. Soc.* **2012**, *134*, 6455−6466.
(46) Chen, B.; Tycko, R. *Protein Sci.* **2010**, *19*, 716−730.
(47) Han, Y.; Ahn, J.; Concel, J.; Byeon, I.-J. L.; Gronenborn, A. M.; Yang, J.; Polenova, T. *J. Am. Chem. Soc.* **2010**, *132*, 1976−1987.
(48) Han, Y.; Hou, G.; Suiter, C. L.; Ahn, J.; Byeon, I.-J. L.; Lipton, A. S.; Burton, S.; Hung, I.; Gor'kov, P. L.; Gan, Z.; Brey, W.; Rice, D.; Gronenborn, A. M.; Polenova, T. *J. Am. Chem. Soc.* **2013**, *135*, 17793−17803.
(49) Lu, M.; Hou, G.; Zhang, H.; Suiter, C. L.; Ahn, J.; Byeon, I.-J. L.; Perilla, J. R.; Langmead, C. J.; Hung, I.; Gor'kov, P. L.; Gan, Z.; Brey, W.; Aiken, C.; Zhang, P.; Schulten, K.; Gronenborn, A. M.; Polenova, T. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 14617−14622.
(50) Bayro, M. J.; Chen, B.; Yau, W.-M.; Tycko, R. *J. Mol. Biol.* **2014**, *426*, 1109−1127.
(51) Liu, C.; Perilla, J. R.; Ning, J.; Lu, M.; Hou, G.; Ramalho, R.; Himes, B. A.; Zhao, G.; Bedwell, G. J.; Byeon, I.-J.; Ahn, J.; Gronenborn, A. M.; Prevelige, P. E.; Rousso, I.; Aiken, C.; Polenova, T.; Schulten, K.; Zhang, P. *Nat. Commun.* **2016**, *7*, 10714.
(52) Lu, J.-X.; Bayro, M. J.; Tycko, R. *J. Biol. Chem.* **2016**, *291*, 13098.
(53) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98−102.
(54) Hu, K.-N.; Qiang, W.; Tycko, R. *J. Biomol. NMR* **2011**, *50*, 267−276.
(55) Hardy, E. H.; Verel, R.; Meier, B. H. *J. Magn. Reson.* **2001**, *148*, 459−464.
(56) Shen, Y.; Bax, A. *Methods Mol. Biol.* **2015**, *1260*, 17−32.
(57) Kingston, R. L.; Fitzon-Ostendorp, T.; Eisenmesser, E. Z.; Schatz, G. W.; Vogt, V. M.; Post, C. B.; Rossmann, M. G. *Structure* **2000**, *8*, 617−628.
(58) Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y. Z.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M. *Cell* **2009**, *137*, 1282−1292.
(59) Ganser-Pornillos, B. K.; Cheng, A.; Yeager, M. *Cell* **2007**, *131*, 70−79.
(60) Jaballah, S. A.; Bailey, G. D.; Desfosses, A.; Hyun, J.; Mitra, A. K.; Kingston, R. L. *bioRxiv* **2017**, 098400.
(61) Bailey, G. D.; Hyun, J. K.; Mitra, A. K.; Kingston, R. L. *Structure* **2009**, *17*, 737−748.
(62) England, M. R. Ph.D. Thesis, The Pennsylvania State University, State College, PA, 2015.

DOI: 10.1021/jacs.6b11939
J. Am. Chem. Soc. 2017, 139, 2006−2013