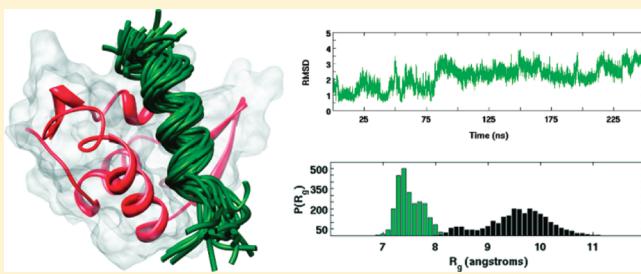


# Atomistic Simulations Reveal Structural Disorder in the RAP74-FCP1 Complex

Christopher Wostenberg,<sup>†</sup> Sushant Kumar,<sup>‡</sup> William G. Noid,<sup>†</sup> and Scott A. Showalter<sup>\*,†</sup><sup>†</sup>Department of Chemistry, The Pennsylvania State University, 104 Chemistry Building, University Park, Pennsylvania 16802, United States<sup>‡</sup>Huck Institutes for the Life Sciences, The Pennsylvania State University, Pennsylvania 16802, United States

Supporting Information

**ABSTRACT:** We report atomically detailed molecular dynamics simulations characterizing the interaction of the RAP74 winged helix domain with the intrinsically disordered C-terminal of FCP1. The RAP74–FCP1 complex promotes the essential dephosphorylation of RNA polymerase II prior to initiation of transcription. Although disordered in solution, the C-terminal of FCP1 forms an amphipathic helix when bound to RAP74. Our simulations demonstrate that this interaction also reorganizes and stabilizes RAP74. These simulations illuminate the significance of hydrophobic contacts for stabilizing disordered protein complexes, provide new insight into the mechanism of protein binding by winged helix domains, and also reveal “dynamic fuzziness” in the complex as FCP1 retains significant flexibility after binding. In conjunction with our recent NMR experiments identifying residual structure in unbound FCP1, these simulations suggest that FCP1 loses relatively little conformational entropy upon binding and that the associated coupled folding–binding transition may be less sharp than expected.



## INTRODUCTION

Intrinsically disordered proteins (IDPs), or intrinsically disordered regions (IDRs) within longer polypeptide chains, are an intriguing class of proteins that perform many vital cellular functions, despite lacking a well-defined equilibrium structure.<sup>1,2</sup> Because their sequences are relatively enriched in polar residues and depleted in hydrophobic residues, IDPs cannot form a stable hydrophobic core and instead sample an ensemble of disordered conformations *in vivo*.<sup>3</sup> The unique physical and sequence properties of IDPs have been suggested to confer a wide array of biological advantages for protein–protein interactions, including many-to-one signaling, high-specificity low-affinity binding, and rapid binding kinetics.<sup>4</sup>

Many IDPs fold to a stable structure when interacting with appropriate binding partners, but the resulting disordered protein complexes significantly differ from those involving well-folded proteins.<sup>5–7</sup> In contrast to ordered protein complexes, which involve a relatively small fraction of the protein surface and are stabilized by polar interactions of residues that are distant in the protein sequence, disordered protein complexes form extensive interfaces that are largely stabilized by hydrophobic interactions of residues that are close in sequence.<sup>6,8</sup> Nussinov and co-workers have suggested that the coupled folding–binding of IDPs results from interactions that reshape the free-energy landscape for disordered proteins.<sup>9</sup> Moreover, disordered proteins often retain considerable disorder after binding, resulting in somewhat “fuzzy” complexes.<sup>10</sup> The landscape reshaping paradigm has recently been adapted to the description of fuzzy

complex formation, which may reflect traces of landscape reshaping from the process of their formation.<sup>7</sup> The Sic1 ubiquitin ligase system, which has been extensively characterized by NMR spectroscopy,<sup>11–13</sup> offers unparalleled insight into the nature of fuzzy complexes and their free-energy landscapes. Recent atomistic simulations of IDP complexes have also contributed insight into the mechanisms of folding-upon-binding events and the possible significance of “fly-casting”<sup>14–16</sup> for this process.<sup>17–20</sup> Nevertheless, many questions remain regarding the general principles governing IDP binding<sup>21</sup> and about FCP1 binding in particular.

Bioinformatic studies have revealed that IDRs are especially prevalent in eukaryotic organisms,<sup>22</sup> where they function in transcription factors,<sup>4,23</sup> acidic transactivation domains,<sup>24</sup> interactions bridging the preinitiation complex and mediator,<sup>25</sup> and in other components of the eukaryotic transcription machinery.<sup>26</sup> Disordered regions can effectively link globular domains, and the relatively large surface area of disordered protein complexes optimizes molecular recognition while requiring a minimal number of amino acid residues to achieve specificity.<sup>26</sup> The RAP74–FCP1 interaction investigated in the present paper is one particularly interesting example of a disordered protein complex involved in regulating transcription.

Eukaryotic transcription of mRNA by RNA polymerase II (PolII) is tightly regulated by a collection of transcription factors,

**Received:** August 19, 2011

**Revised:** September 27, 2011

**Published:** October 11, 2011

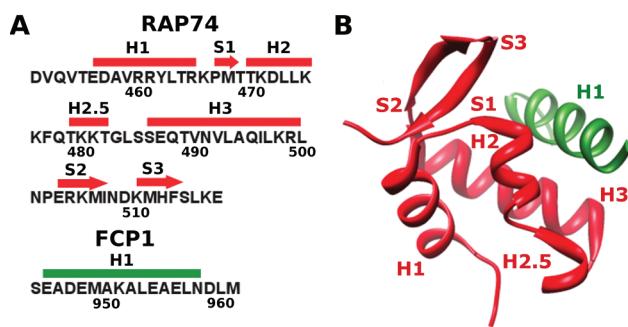
including transcription factor IIF (TFIIF), and also by the phosphorylation state of a repeated heptapeptide sequence in the disordered C-terminal domain (CTD) of PolII. The CTD repeated heptapeptide becomes hyperphosphorylated during the course of transcription.<sup>27</sup> The TFIIF-associating CTD Phosphatase (FCP1) performs an essential function in regulating transcription by dephosphorylating the CTD so that PolII can be recycled for another round of transcription.<sup>28,29</sup> This function is promoted by an interaction between the disordered C-terminal domain of FCP1 and the C-terminal winged helix (WH) domain of the TFIIF RAP74 subunit.<sup>30</sup>

The C-terminal domain of FCP1 bears striking similarity to acidic transactivation domains and can function as a transactivator *in vitro* when fused to the GAL4 DNA binding domain.<sup>31</sup> Similarly to other transactivation domains, the FCP1 C-terminal domain is enriched in acidic residues and contains relatively few hydrophobic residues, which are organized in several clusters. Consequently, the FCP1 tail cannot form a stable hydrophobic core and remains largely disordered in solution,<sup>32</sup> although it should be noted that the Showalter laboratory has obtained complete chemical shift assignments that have recently identified nascent helical character in the disordered ensemble for this tail.<sup>33,34</sup>

In the presence of the C-terminal WH domain of RAP74, the C-terminal 17 residues of FCP1 fold to form a well-defined amphipathic helix.<sup>32,35</sup> This helix binds to a hydrophobic groove lined by two helices of RAP74. Crystal and NMR structures have provided insight into molecular interactions between the proteins but do not provide a quantitative understanding of the spatiotemporal dynamics involved in binding and in maintaining the complex. The current literature contains no reports of experimental (e.g., NMR spin relaxation) or computational investigations into the dynamics of the RAP74–FCP1 interaction. Consequently, detailed simulation studies of the RAP74–FCP1 system should contribute additional insight into the spatiotemporal dynamics of binding IDPs and the interactions stabilizing these complexes.

The FCP1–RAP74 complex is also intriguing because this interaction is mediated by the basic WH domain of RAP74. The WH motif is a variation on the all  $\alpha$ -helical helix-turn-helix motif that generally functions as a DNA binding domain.<sup>36</sup> The WH domain has a compact  $\alpha/\beta$  structure consisting of three  $\alpha$  helices (H1, H2, and H3), three  $\beta$  strands (S1, S2, and S3), and two wings (W1 and W2) arranged in order H1–S1–H2–H3–S2–W1–S3–W2; RAP74 contains an additional short helix (termed H2.5) between H2 and H3 (annotated in Figure 1). The canonical mode of DNA binding by WH domains involves extensive polar interactions mediated by the recognition helix, H3.<sup>36</sup> The WH domain in the RAP30 chain of TFIIF also displays chemical shift changes in H1 upon DNA binding, suggesting an alternate binding mode involving this helix, long-range conformational change induced by ligand binding, or both.<sup>37</sup>

The structure of FCP1 in complex with RAP74 (Figure 1) reveals a binding site created by the shallow groove between H2 and H3, adjacent to the canonical DNA binding interface. Previous studies have shown that the hydrophobic patch bordered by H2 and H3 mediates protein–protein interactions, particularly when the domains are bound to DNA.<sup>36</sup> This hypothesis is supported by the DNA mediated heterodimerization of the WH domains from DP2 and E2F4, in which H3 from each protein interacts simultaneously with the major groove of DNA and with the H2/H3 interface of the adjacent protein.<sup>38</sup> The



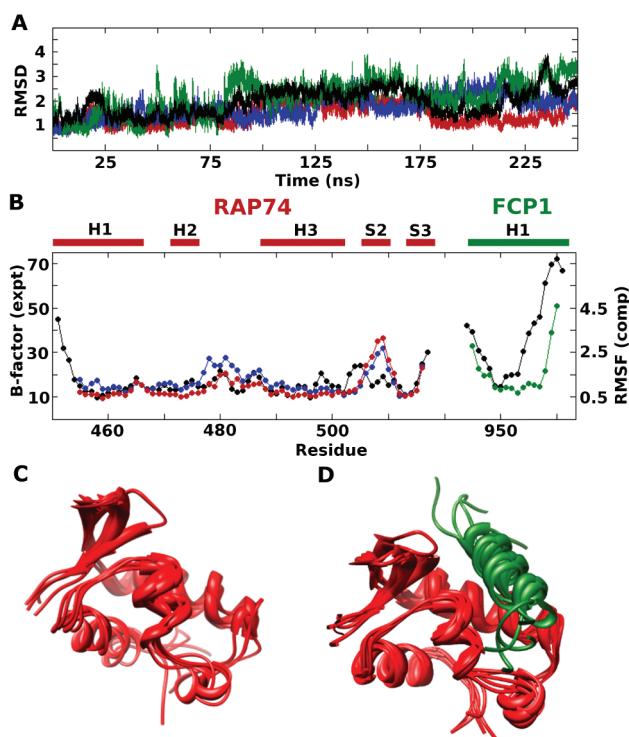
**Figure 1.** Sequence schematic and a ribbon diagram representing the crystal structure of FCP1 (green) bound to RAP74 (red) (pdb 1J2X). (A) Secondary structures are annotated over the primary sequence of each protein. (B) The ribbon diagram shows that FCP1 lays across helix 2 (H2) and helix 3 (H3) of RAP74.

quaternary Cul1–Rbx1–Skp1–Fbox<sup>Skp2</sup> complex also features a protein–protein interaction mediated by the H2/H3 interface.<sup>39</sup> To our knowledge, no other high-resolution structures of protein–protein complexes involving a WH domain have been experimentally determined in the absence of DNA. The role of the H2/H3 interface in mediating protein–protein interactions is further supported by mutagenesis studies of the T4 transcription factor MotA<sup>40</sup> and by the structure of p53 bound to MDM2 (which bears topological similarity to a WH).<sup>41</sup> Consequently, simulation studies of the RAP74–FCP1 complex should identify interactions stabilizing this complex and reveal fundamental insight into the potentially novel protein–protein interactions involving WH domains.

Motivated by these considerations, we have performed atomically detailed explicit solvent molecular dynamics studies of the RAP74 WH in the apo-state and in complex with FCP1. Our *in silico* studies of apo-RAP74 and the RAP74–FCP1 complex elucidate the atomistic interactions stabilizing the complex, their effect upon protein dynamics, and the mode of binding FCP1 by RAP74. These simulations demonstrate that this binding not only induces FCP1 folding but also reorganizes and stabilizes the RAP74 WH domain. Most significantly, though, these simulations reveal that FCP1 retains substantial conformational flexibility in the RAP74–FCP1 complex, that is, the complex is fuzzy. The significant disorder of bound FCP1, coupled with our previous NMR studies revealing partial order in unbound FCP1, suggests a relatively limited entropic cost associated with the folding-upon-binding of IDPs.

## MATERIALS AND METHODS

Two MD simulations were performed with the AMBER 11.0 software package<sup>42</sup> using the ff99SB<sup>43,44</sup> force field and the SPC water model<sup>45</sup> under particle-mesh Ewald periodic boundary conditions.<sup>46</sup> Both simulations were derived from the crystal structure of RAP74 in complex with FCP1 (pdb 1J2X):<sup>35</sup> RAP74 apo and RAP74–FCP1 holo. The RAP74 apo simulation modeled residues 451–517 from human RAP74; the holo simulation also modeled residues 944–961 from human FCP1. Two and seven chloride ions were added to neutralize the net positive charge of the proteins in the holo and apo simulations, respectively. The resulting systems were solvated such that no solute atom was within 10 Å of a box edge, requiring 6019 waters for the holo simulation and 5340 for the apo simulation. The starting



**Figure 2.** RMSD and RMSF analysis of the RAP74 and RAP74–FCP1 simulations. (A) RMSD traces show the overall stability of the two simulations, the RAP74 apo simulation (blue) and the holo simulation (black). The isolated proteins from the holo simulation are also plotted, FCP1 (green) and RAP74 (red), which shows that the high RMSD in the holo simulation comes from deviations in FCP1. (B) Experimental B-factors (black) and computational RMSF by the residue (colored as per panel A). Ribbon bundles of the apo state (C) and the holo state (D) show that there is no loss of secondary structure during the simulations. All bundles were created by taking structures from the simulation every 50 ns.

configurations were energy-minimized and equilibrated as previously reported.<sup>47</sup> After an initial equilibration period, each system was simulated for 250 ns with a 2.0 fs time step in the isothermal–isobaric (*NPT*) ensemble.<sup>48</sup> Snapshots from each trajectory were stored to disk every 1.0 ps. The trajectories were analyzed with the AMBER software package, using the ptraj program,<sup>42</sup> with VMD,<sup>49</sup> with MATLAB (The MathWorks, Natick, MA), with GROMACS utilities,<sup>50</sup> and also with in-house software. Molecular graphics images were created using the UCSF Chimera package.<sup>51</sup>

## RESULTS AND DISCUSSION

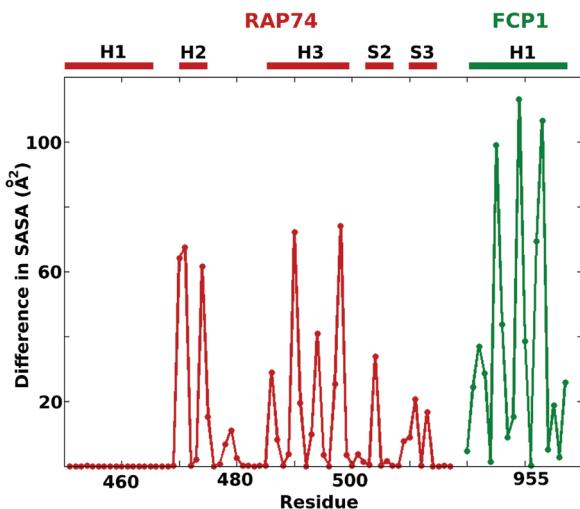
**Atomic displacements in the simulations.** The backbone root-mean-square deviation (RMSD) from the crystal structure was calculated to determine the protein stability during each simulation. The unstructured tails of RAP74 (N-terminal residues 451–454 and C-terminal residue 517) and FCP1 (residues 944 and 961) were excluded from the calculation. The apo simulation (Figure 2A, blue) reaches a maximum backbone RMSD near 3.0 Å, which is slightly higher than expected for a single globular domain. Nevertheless, the  $\text{Ca}$  RMSF (Figure 2B) and ribbon bundle (Figure 2C) demonstrate that RAP74 retains its secondary structure and that the relatively high RMSD reflects the mobility of the wing subdomain and helix 2.5. The holo

simulation (Figure 2A, black) has a slightly higher RMSD with a maximum of approximately 3.9 Å. The corresponding RMSF and ribbon bundle demonstrate (Figure 2B,D) greater flexibility in the RAP74 wing subdomain (upper left), as observed for the apo-state. In both simulations, the average backbone RMSD of RAP74 is well below 2.0 Å (1.7 and 1.4 Å in the apo and the holo trajectories, respectively), indicating the overall stability of the trajectories.

Previous studies have demonstrated that the ff99SB parameters used in these simulations accurately reproduce the structure, dynamics, and, in particular, magnitude of loop dynamics for small globular systems;<sup>43,47,52</sup> however, it is still important to cross-validate the current simulations with experimental data. Given that no reports of NMR spin relaxation measurement exist for this system, we compare the crystallographic B-factors reported for the complex and the RMS fluctuations observed in the trajectories. Qualitatively, it is clear that the RMSF of RAP74 tracks the experimental B-factors well in all regions except the wing subdomain, which appears to be overly flexible in both trajectories (Figure 2B). The largest B-factors found in the complex are almost all found in FCP1. While our trajectory appears to reproduce the extent of motion in the N-terminal portion of the peptide well, the RMSF of the C-terminal region is much lower than the B-factors, suggesting either the presence of dynamics occurring on time scales longer than the simulation or heterogeneity in the crystal.

**Burial of the Solvent-Accessible Surface Area (SASA) in the RAP74–FCP1 Complex.** Upon binding RAP74, the disordered C-terminus of FCP1 folds into an amphipathic helix that presents its hydrophobic face to cover a long but relatively shallow hydrophobic groove on the RAP74 surface.<sup>32,35</sup> Nussinov and co-workers have suggested that intrinsic disorder confers a functional advantage by allowing a larger fraction of the binding protein to contribute to interface formation.<sup>8</sup> Analysis of the co-crystal structure supports this hypothesis. To further characterize this interface, we calculated the average SASA of each residue in the RAP74–FCP1 complex using configurations sampled from the holo simulation. We then compared these averages with the corresponding averages for residues in each isolated protein after removing its binding partner from the sampled configurations. A per residue comparison of SASA between residues in the isolated proteins and in the complex (Figure 3) demonstrates that less than 15% of residues experience a significant decrease of SASA in the complex. During the simulation, the complex buries an average of 1256 Å<sup>2</sup> of SASA, including 614 Å<sup>2</sup> from RAP74 and 642 Å<sup>2</sup> from FCP1, neglecting changes in FCP1 SASA due to forming an  $\alpha$ -helix. FCP1 illustrates the ability of IDPs to achieve high specificity through burial of a large surface area while utilizing a minimal number of amino acid residues to do so.<sup>8</sup> The majority of the buried surface area arises from the hydrophobic face of FCP1 covering the hydrophobic groove of RAP74 and, in particular, burial of Met949, Leu953, and Leu957.

These changes in SASA also reflect reorganization of the RAP74 binding pocket. A total of eight residues from RAP74 become solvent-inaccessible in the bound state: Thr470, Lys471, and Leu474 from H2; Ser486, Val490, Ala494, and Lys498 from H3; and Arg504 from S2. Comparison of the SASAs for the holo and apo simulations suggests that the RAP74 hydrophobic groove expands to accommodate FCP1. This expansion exposes several residues for interacting with FCP1, including Thr470 (in H2), which has been implicated in an important polar contact with Glu954,<sup>32</sup> as well as Leu474 (in H2), Val490, Val492, and

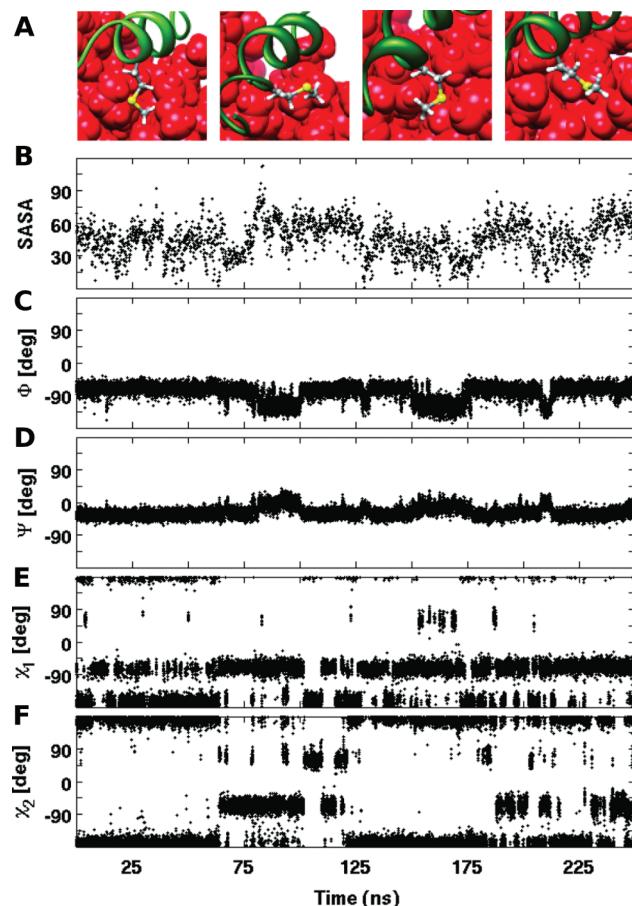


**Figure 3.** The difference in the SASA shows the specific residues from RAP74 and FCP1 that are buried upon binding. In FCP1 (green), the major residues that are buried are Met949, Leu953, and Leu957, which are all along the hydrophobic face of the amphipathic  $\alpha$ -helix. In RAP74 (red), the major residues that are buried are Thr470, Lys471, and Leu474 from helix 2, Ser486, Val490, Ala494, and Lys498 from helix 3, and Arg504 from strand 2. Secondary structure elements are depicted above the graph.

Leu493 (all in H3), which form contacts with key hydrophobic residues in FCP1: Met949, Leu953, and Leu957.

**Interactions Stabilizing the RAP74–FCP1 Complex.** Complexes involving two ordered proteins tend to be stabilized by polar interactions, whereas complexes involving IDPs are often depleted in polar contacts.<sup>6</sup> As described above, the hydrophobic face of the amphipathic FCP1 helix lies across the RAP74 hydrophobic binding groove, but acidic residues near the FCP1 helix termini are positioned in close proximity with basic residues flanking the RAP74 binding pocket. The present simulations supplement previous experimental structural studies by microscopically characterizing the key interactions that stabilize the RAP74–FCP1 complex and by providing a dynamic description of the complex.

**Hydrophobic Interactions.** Both crystallographic<sup>35</sup> and NMR structures<sup>32</sup> indicate that FCP1 buries three key hydrophobic residues in the shallow hydrophobic groove of RAP74: Met949, Leu953, and Leu957. Mutation of these residues to alanine significantly decreases the complex affinity, further demonstrating that these residues are critical for proper binding.<sup>35</sup> These three residues form an extensive network of interaction with RAP74, as evidenced by the multiple methyl–methyl NOEs recorded;<sup>32</sup> this result stands in contrast to the unusually large crystallographic B-factors reported for Leu953 and Leu957, which suggests that the C-terminal portion of RAP74-bound FCP1 may be poorly ordered in the crystal environment. Table S1 (Supporting Information) documents the average heavy atom distances separating RAP74 and FCP1 residues identified in the initial structural reports as forming significant contacts. When the heavy atom distances in Table S1 (Supporting Information) are shortened to account for the fact that NOEs measure H–H distances, our RAP74–FCP1 simulation is seen to preserve all contacts mediated by the methyl groups of Met949, Leu953, and Leu957. As described below, our atomically detailed simulation of the RAP74–FCP1 complex clarifies the significance and stability of these contacts.



**Figure 4.** The reversible escape of Met 949 from the RAP74 binding groove is largely due to jumps in  $\chi_2$ . (A) Representative snapshots taken at 50, 100, 150, and 200 ns zoomed in to show Met 949 (colored by atom) from FCP1 (green ribbon) and its excursions out from and into the RAP74 binding groove (red van der Waals spheres). (B) The SASA of Met949 as a function of time shows increased solvent exposure when  $\chi_2$  is approximately  $\pm 90^\circ$ . The remaining panels display Met 949 torsion angles  $\varphi$  (C),  $\psi$  (D),  $\chi_1$  (E), and  $\chi_2$  (F) as a function of time. The angle  $\chi_3$  is unconstrained and not shown for clarity.

Closer inspection of the complex reveals that the RAP74 binding pocket is shallow in the vicinity of Met949 and that this residue appears to cap the interaction interface on the N-terminal end of FCP1. The backbone of the neighboring FCP1 residues retains a stable  $\alpha$ -helical conformation, and both RMSF and B-factors report relatively small fluctuations in the Met949 backbone. However, the simulation suggests that the Met949 side chain fluctuates dramatically, repeatedly escaping from and re-entering the RAP74 hydrophobic groove during the trajectory (indicated in Figure 4A by snapshots at 50, 100, 150, and 200 ns). The escape of its side chain significantly increases the Met949 SASA and correlates with reversible transitions in the rotameric state of the torsion angle  $\chi_2$  (Figure 4B and F, respectively) that occur on the nanosecond time scale but do not appear to correlate strongly with any other fluctuations in Met949 torsions.

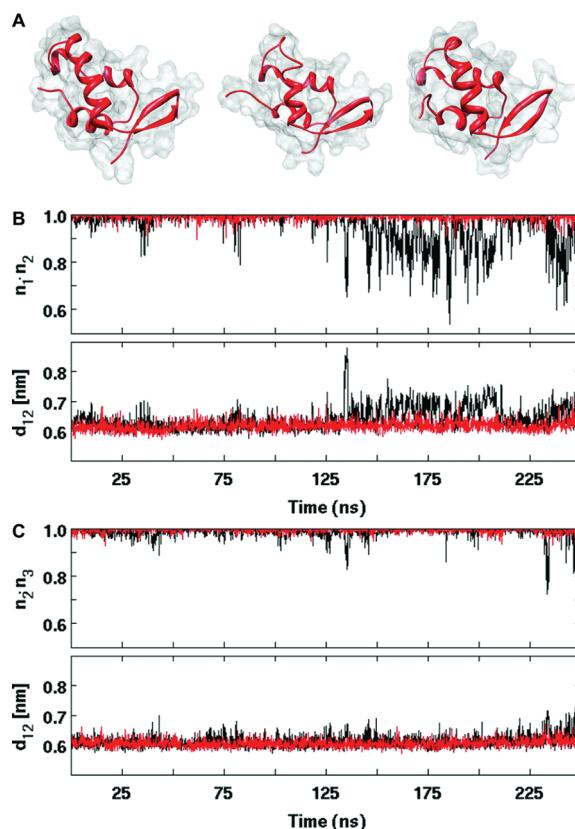
In contrast, RAP74 residues Thr470, Leu474, Leu493, and Ala494 form a pocket that locks the Leu953 side chain in place and reduces the fluctuations in Leu953  $\chi_1$  and  $\chi_2$  torsions (see Supporting Information, Figures S1 and S3). The side-chain carbons of Leu953 remain within 5 Å of each RAP74 methyl,

forming the binding pocket over 40% of the time. Because Leu957 forms half as many van der Waals contacts as Leu953, Leu957 demonstrates greater conformational freedom, and its  $\chi_1$  and  $\chi_2$  torsions sample significant fluctuations, which reposition its methyl groups in the RAP74 groove (see Supporting Information, Figure S2; a less constrained leucine residue is shown in Figure S3 for contrast). Despite its backbone flexibility, the side chain of Leu957 does not escape from the binding groove as Met949 does. The very low RMSF near Leu953 and Leu957, compared with the anomalously high B-factors in this region, suggests that these residues may undergo motions in vitro that are too slow to characterize with MD. Future NMR relaxation–dispersion studies are needed to support this hypothesis. Nevertheless, our MD simulations clearly support the mutagenesis studies by demonstrating the importance of hydrophobic contacts and, in particular, Leu953 and Leu957 for stabilizing the RAP74–FCP1 complex.

**Polar Interactions.** In addition to contacts involving the three key hydrophobic residues (i.e., Met949, Leu953, and Leu957), alanine scanning mutations performed by Kamada et al. suggested that two polar contacts in FCP1 are essential for effective binding to RAP74.<sup>35</sup> Both the crystal and solution NMR structures indicate that Asp947 at the N-terminal of the FCP1 helix forms a salt bridge with Lys471 in H2 of RAP74.<sup>32,35</sup> Mutation of either Asp947 to alanine or Lys471 to glutamate disrupts the RAP74–FCP1 interaction.<sup>35,53</sup> Our MD simulations further substantiate the importance of the Lys471–Asp947 salt bridge (Supporting Information, Table S1). Intriguingly, despite the many basic residues lining the RAP74 binding groove and the acidic nature of FCP1, this is the only salt bridge directly confirmed by mutagenesis to significantly influence binding affinity. Upon the basis of NOEs observed between Lys498 and Leu960, Nguyen et al. also suggested that the neighboring Asp959 at the C-terminus of FCP1 may form a salt bridge with Lys498 in RAP74 H3.<sup>32</sup> However, this interaction appears to be poorly preserved in the holo simulation, suggesting that this salt bridge may be relatively weak.

**FCP1 Alters RAP74 Structure and Dynamics.** The RAP74–FCP1 interaction not only induces FCP1 to fold but also appears to significantly alter the structure and fluctuations in the folded RAP74 WH domain. Except for fluctuating contacts that involve the terminal residues, the RAP74–FCP1 complex neither breaks nor forms stable intramolecular hydrogen bonds or salt bridges that are not present in apo-RAP74. However, detailed analysis of the simulations reveals that interactions with FCP1 expose key interacting residues in RAP74, as described above, and also lead to a reorganization and stabilization of RAP74 secondary structures.

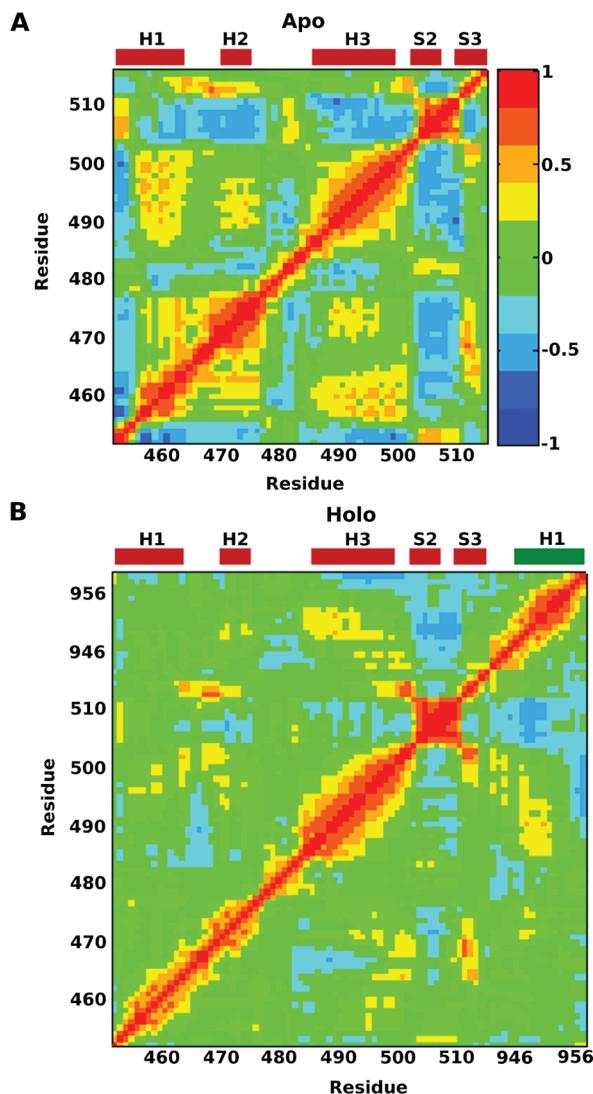
**Secondary Structure Analysis.** The FCP1 binding cleft on RAP74 is composed primarily of three  $\alpha$ -helix secondary structures, suggesting that helix dipole interactions, which significantly stabilize folded helix bundles, may play a significant electrostatic role in stabilizing the complex.<sup>54</sup> When bound to RAP74, the FCP1 helix dipole is equally antiparallel with H2 (average angle of  $-52 \pm 10^\circ$ ) and parallel with the larger dipole of H3 (average angle of  $+56 \pm 10^\circ$ ), contradicting this premise. However, FCP1 has little effect upon the structure and stability of H2. The distributions for the H2 end-to-end distance and backbone dipole moment are essentially unchanged between the apo and holo simulations. Moreover, the angle between H2 and H3 remains approximately  $60^\circ$  in both simulations, although this angle fluctuates slightly less in the holo simulation (standard deviation reduces from  $7$  to  $4^\circ$ ).



**Figure 5.** The N-terminal turn of helix 3 is unstable in apo-RAP74. (A) Representative snapshots from 100, 125, and 240 ns reveal the reversible unfolding of H3. Panels (B) and (C) present time traces from simulations of apo-RAP74 (black) and the RAP74/FCP1 complex (red). Panel (B) corresponds to the first (486–489) and second (490–493) turns; panel (C) corresponds to the second and third (494–497) turns. In each case, the top and bottom graphs describe the angle formed by successive turns and the corresponding distance, respectively. The direction of each turn was calculated from the average of the helix directors for the four alpha carbons. The distance between turns was calculated from centers of geometry for the alpha carbons. Disorder in H3 of apo-RAP74 is localized to the first turn.

In contrast, interactions with FCP1 significantly stabilize the canonical DNA recognition helix H3. In the apo state, the first turn of H3 (involving Ser486, Glu487, Gln488, and Thr489) fluctuates in and out of alignment to form a kink at an angle of roughly  $50^\circ$  with respect to the rest of the helix. This kink correlates with an expansion of  $2 \text{ \AA}$  between the alpha carbons of the first and second turns of H3 (Figure 5). However, in the presence of bound FCP1, H3 retains a rigid helical structure with stable hydrogen bonding between successive turns and a uniform orientation along the helix.

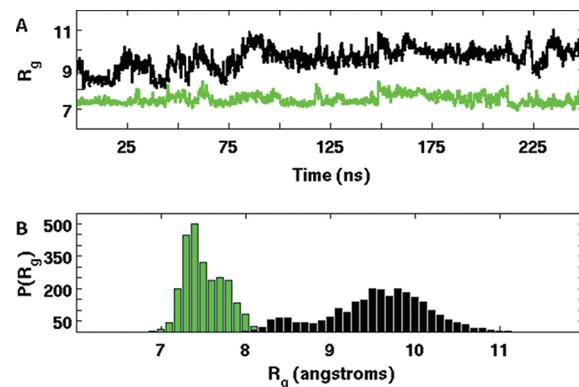
FCP1 also induces several changes in the structure and fluctuations of regions that are spatially separated from the RAP74 binding pocket. In the holo simulation, H1 tends to sample more compact conformations and also tilts to form an angle that is more nearly (anti-) parallel with respect to H2, giving rise to a periodicity in the difference of SASA between the apo-RAP74 and RAP74–FCP1 complex simulations (Figure S5, Supporting Information). FCP1 also dramatically stabilizes the interaction between S1 and S3. In the apo state, the twist vectors of S1 and S3 form an angle of  $50^\circ$  with very large fluctuations. In contrast,



**Figure 6.**  $\text{Ca}$  correlation matrixes reveal decreased dynamics of RAP74 in the FCP1 bound state. (A) Correlation matrix for the unbound RAP74 simulation, shown on a consistent scale with the RAP74 portion of the RAP74–FCP1 complex below. The color bar on the right indicates strong positive correlation (red), strong negative correlation (blue), and noncorrelated motion (green). (B) The holo state shows positive correlation between helix 3 (485–499) and the middle of FCP1 (950–955) and negative correlation between the winged domain of RAP74 (505–513) and the middle of FCP1 (950–955). Secondary structures are annotated with bars above each panel.

the holo state, the twist vectors are almost antiparallel with much smaller fluctuations. The RAP74 reorganization also significantly reduces the solvent exposure of strands S2 and S3, including residues Met511 and Phe513, both of which make several contacts with FCP1 in the crystal structure. The Supporting Information contains an expanded analysis of helical interactions in the apo-RAP74 WH and in the RAP74–FCP1 complex.

**Covariance Analysis.** Covariance analysis provides further evidence that FCP1 significantly stabilizes and rigidifies RAP74. Figure 6 presents covariance matrices calculated from the fluctuations of alpha carbons in apo and holo simulations. RAP74 clearly samples many more and much larger correlated motions in the apo state (Figure 6A) than in the holo state (Figure 6B).



**Figure 7.** Radius of gyration ( $R_g$ ) distributions indicate disorder in bound FCP1. (A)  $R_g$  for full FCP1 (944–961) (black) and for the helical residues (944–957) only (green) as a function of time. (B) The histograms of the same  $R_g$  data demonstrate that the FCP1 helix undergoes minor fluctuations about its mean, while the full peptide experiences significant excursions, dominated by the disordered C-terminus.

The covariance matrix for the holo state reveals that the fluctuations of the central FCP1 residues (950–955) are characterized by positive correlations with RAP74 H3 (residues 485–499) and strong negative correlation with the RAP74 wing domain (residues 505–513). These positive correlations with H3 likely result from the many intermolecular interactions described above. However, despite the presence of many intermolecular interactions between H2 (residues 470–476) and FCP1, the covariance matrix indicates relatively little correlated motion involving these regions in the holo simulation (Figure 6B).

**Structural Disorder of FCP1 in complex with RAP74.** Published crystal and NMR structures provide a static description of the RAP74–FCP1 complex and demonstrate that the disordered FCP1 C-terminus forms a helix when bound to RAP74. Our simulations indicate that the RAP74 binding groove stabilizes the orientation of FCP1 at an angle of roughly  $55^\circ$  with respect to both H2 and H3 with fluctuations of  $5\text{--}10^\circ$ . The groove also maintains a stable helix orientation for FCP1 as the average angle between the helix directors in successive turns is less than  $14^\circ$  with a standard deviation of a few degrees. However, the simulations also demonstrate that FCP1 retains significant disorder in this complex; in addition to the conformational flexibility of Met949 described above, the tails of FCP1 remain quite disordered, and the hydrogen bonding between FCP1 turns is relatively unstable.

The radius of gyration for the first 14 FCP1 helical residues fluctuates about  $7.2 \pm 0.1$  Å with occasional fluctuations to a larger radius of  $7.8$  Å. However, the final four residues of FCP1 (Gln958, Asp959, Leu960, and Met961) sample much larger fluctuations (Figure 7). As discussed earlier, Met949, Leu953, and Leu957 anchor the center and C-terminal regions of the FCP1 helix. The alpha carbons for these residues remain separated by a distance of  $6.2$  Å with a standard deviation of less than  $0.3$  Å, reflecting stable hydrogen bonding in this region and thus a stable  $\alpha$ -helix. In contrast, the N-terminal side of the FCP1 helix is significantly less well anchored to the binding groove, although the previously discussed salt bridge mediated by the side chain of Asp947 does help hold this end of the protein in place. The distance between the first and second turns of the helix samples significant fluctuations that correlate with an expanded radius of gyration for the helix. For much of the simulation, Ala946 and

Ala950 anchor the N-terminal turns of the FCP1 helix by making very stable contacts that only rarely expand. However, in the last 10 ns, these two residues separate, and the first turn of the FCP1 helix escapes so that the corresponding alpha carbons are separated by 10 Å. At the same time, the backbone dipole of FCP1, which had already decreased by roughly 10% from its initial value during the first 150 ns, precipitously drops during the final 10 ns of simulation.

## CONCLUSIONS

The present work reports the first simulations investigating the WH domain of the larger (RAP74) subunit of TFIIF<sup>55</sup> and its interaction with FCP1.<sup>28</sup> FCP1 performs an essential function by dephosphorylating the C-terminal domain of RNA polymerase II prior to reinitiation of transcription.<sup>29</sup> The catalytic activity of FCP1 is promoted when its C-terminal tail binds the RAP74 WH domain.<sup>56</sup> Upon binding the hydrophobic groove of RAP74, the intrinsically disordered FCP1 tail folds to form an amphipathic helix.<sup>32,35</sup> Our atomically detailed simulations of RAP74 in its apo and holo states employed a force field and simulation procedures previously verified to accurately reproduce the dynamics of domains the size of RAP74.<sup>43,47,52</sup> Moreover, these simulations are consistent with a considerable body of experimental results. Consequently, our study supplements previous experimental studies by providing exquisite insight into the interactions and fluctuations characterizing this complex.

As would be expected for an IDP, the RAP74–FCP1 complex significantly differs from interfaces formed by ordered proteins, which typically cover a relatively small fraction of the proteins and which are largely stabilized by interactions between polar and charged residues.<sup>6,8</sup> Instead, the RAP74–FCP1 complex forms an extensive interface with the RAP74 hydrophobic binding pocket covering approximately 1200 Å<sup>2</sup> of SASA, including the hydrophobic surface of FCP1. Our simulations confirm the experimental observation that, despite the presence of many charged and polar residues at the FCP1–RAP74 interface, polar interactions play a relatively small role in stabilizing the complex. The only significant exception is the very stable Lys471–Asp947 salt bridge, which binds the N-terminal side of the FCP1 helix to the N-terminal side of RAP74 H2. Although experimental studies have suggested that Glu954–Thr470 and Asp959–Lys498 form important contacts, these contacts are not very stable in our simulations.<sup>32,35</sup> It is possible that these interactions were not well modeled in the simulations; alternatively, these weak polar interactions may contribute to binding affinity through means that are not immediately obvious from observations of the final bound state, such as by distinguishing between parallel and anti-parallel binding of FCP1. Intriguingly, helix dipole interactions, which significantly stabilize folded helix bundles,<sup>54</sup> would appear to destabilize the interaction of RAP74 and FCP1 as the FCP1 dipole is oriented parallel with the relatively long H3.

Our atomistic simulations demonstrate that the RAP74–FCP1 interaction not only induces FCP1 folding but also impacts the structure and fluctuations of RAP74. The RAP74 hydrophobic groove expands and exposes key residues to bind FCP1. The presence of FCP1 stabilizes H3, which samples kinked conformations in apo-RAP74, tightens the wing domain, and also dampens the collective dynamics of the WH domain. Intriguingly, experiments have demonstrated that binding of the central IDR of FCP1 (centFCP1-P04) to RAP74 induces an expansion of the H2/H3 cleft and reorganizes interactions between H2.5

and H1.<sup>57</sup> Our *in silico* results reinforce the conclusions drawn from these experimental studies and suggest that reorganization of the WH domain fine-tunes the binding interaction.

The observations from our atomistic simulations are significant for providing the first quantitative description of the impact that protein binding may have on the structure and dynamics of WH domains. An alternative DNA binding mode for WH domains that directly involves H1 has been suggested by NMR chemical shifts recorded for the RAP30 WH domain in complex with DNA.<sup>37</sup> We note that binding FCP1 also appears to realign H1 on the surface of RAP74 opposite the FCP1 binding groove, which may alter potential interactions between H1 and other proteins or nucleic acids in the preinitiation complex. While low-resolution footprinting experiments have suggested a possible location for H1,<sup>58</sup> its precise orientation and interactions in the preinitiation complex remain ambiguous.

Most significantly, the simulations also reveal that the FCP1–RAP74 is quite fuzzy.<sup>10</sup> Although the RAP74 hydrophobic groove stabilizes the FCP1 helix and maintains a stable orientation between the two proteins, FCP1 retains significant conformational flexibility in the complex. The N-terminal turns of the FCP1 helix fluctuate significantly and completely separate during the last 10 ns of the simulation, which may be consistent with their large B-factors in the crystal structure. The Met949 side chain reversibly escapes from and re-enters the RAP74 binding pocket, suggesting that interactions near this N-terminal turn of the helix are not maximally stabilizing. Intriguingly, phosphorylation of nearby Ser942 and Ser944 by CK2 stabilizes an extra turn of the helix in FCP1 and strengthens the RAP74–FCP1 interaction, indicating that fine-tuning the interaction strength in this region may play a critical regulatory role.<sup>59</sup>

Clearly, the current definition of fuzzy complexes is rather vague. Consequently, it is important to compare the fluctuations in the RAP74–FCP1 complex and, in particular, the dynamics of side chains, such as Met949, with those that would be expected for a folded protein complex. Several recent studies have employed methyl groups to probe side-chain dynamics in the binding interfaces of folded protein complexes. In these experimental studies, binding-induced changes range from being very minimal in the barnase–barstar complex<sup>60</sup> to being much more dramatic, as in the globally reduced side-chain flexibility of a ternary FKBP12–mTOR complex.<sup>61</sup> In light of these studies, our simulations appear to suggest that FCP1 samples significantly larger conformational fluctuations than might be expected for folded protein complexes. Nevertheless, direct experimental validation is required to clarify this point and to establish quantitative bounds for distinguishing the fuzziness of disordered and folded protein complexes.

In closing, we note that the fuzziness observed in our simulations, in conjunction with previous and ongoing studies in our laboratories, are beginning to reveal a mechanism for the binding of FCP1 and, in particular, to suggest that the disorder–order transition associated with FCP1–RAP74 interactions is more gradual than initially expected. Preformed structural elements have been widely recognized for their significant roles in defining the mechanism of IDP interactions with globular proteins.<sup>62</sup> For example, NMR studies have reported preformed motifs embedded within disordered interaction regions of many proteins, including viral surface antigens<sup>63</sup> and the N-terminal transactivation domain of p53.<sup>64,65</sup> Recently, the idea of conformational fuzziness in IDP complexes has emerged as a further mechanistic signature of IDP binding that reduces the entropic barrier to binding.

In the specific case of FCP1, recent NMR experiments performed by the Showalter laboratory have demonstrated that unbound FCP1 samples helical conformations *in vitro* that are sufficiently stable to yield  $\alpha$ -helical assignment through the chemical shift index in residues 945–950.<sup>34</sup> This region corresponds well with the region of FCP1 having the lowest B-factors in the cocrystal structure.<sup>55</sup> In addition, preliminary studies with native-based  $G_0$  potentials suggest that FCP1 binding proceeds through an extended conformational selection mechanism.<sup>9</sup> The distribution of FCP1 conformations gradually narrows during the early stages of binding, but the binding transition is crossed when RAP74 selects a partially folded FCP1 conformation (S.K and W.G.N, unpublished). Clearly, measurements of backbone NMR spin relaxation for both RAP74 and FCP1 in solution should provide much needed quantitative insight into the extent of FCP1 helix stabilization in the unbound state and confirmation of the results presented here for the bound state.

Taken together with ongoing experimental studies in our laboratory and consistent with these themes from the literature, the present simulations suggest that FCP1 may lose relatively little conformational entropy upon binding RAP74. Consequently, the entropic cost for the coupled folding–binding transition of some IDPs may be smaller than previously anticipated, and the order–disorder transition associated with IDP folding may be less pronounced than implied by static crystal structures.

## ■ ASSOCIATED CONTENT

**5 Supporting Information.** Extended methods; four figures of FCP1 Leu and Met residue solvent exposure and torsion angles; figure of the difference in solvent exposure between the unbound and bound states of RAP74; extensive figures documenting changes in the RAP74 structure induced by FCP1 binding. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: sas76@psu.edu.

## ■ ACKNOWLEDGMENT

We thank Chad Lawrence for helpful discussion and for assistance revising the manuscript. This work was supported by a grant from the NSF to S.A.S. (MCB-0953918), by a fellowship awarded to W.G.N. by the Penn State Institute for Cyberscience, and by startup funds from the Pennsylvania State University.

## ■ REFERENCES

- Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* **2010**, *1804*, 1231.
- Tompa, P. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419.
- Uversky, V. N. *Protein Sci.* **2002**, *11*, 739.
- Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.
- Gunasekaran, K.; Tsai, C. J.; Nussinov, R. *J. Mol. Biol.* **2004**, *341*, 1327.
- Mészáros, B.; Tompa, P.; Simon, I.; Dosztányi, Z. *J. Mol. Biol.* **2007**, *372*, 549.
- Mészáros, B.; Simon, I.; Dosztányi, Z. *Phys. Biol.* **2011**, *8*, 035003.
- Gunasekaran, K.; Tsai, C. J.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28*, 81.
- Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539.
- Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2.
- Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.
- Mittag, T.; Orlicky, S.; Choy, W. Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17772.
- Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. *Structure* **2010**, *18*, 494.
- Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868.
- Levy, Y.; Onuchic, J. N.; Wolynes, P. G. *J. Am. Chem. Soc.* **2007**, *129*, 738.
- Huang, Y.; Liu, Z. *J. Mol. Biol.* **2009**, *393*, 1143.
- Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 2088.
- Ganguly, D.; Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 5214.
- Chen, H. F. *PLoS One* **2009**, *4*, e6516.
- Higo, J.; Nishimura, Y.; Nakamura, H. *J. Am. Chem. Soc.* **2011**, *133*, 10448.
- Liu, J.; Faeder, J. R.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19819.
- Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635.
- Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2006**, *45*, 6873.
- Ferreira, M. E.; Hermann, S.; Prochasson, P.; Workman, J. L.; Berndt, K. D.; Wright, A. P. *J. Biol. Chem.* **2005**, *280*, 21779.
- Tsai, C. J.; Nussinov, R. *Biochem. J.* **2011**, *439*, 15.
- Fuxreiter, M.; Tompa, P.; Simon, I.; Uversky, V. N.; Hansen, J. C.; Asturias, F. J. *Nat. Chem. Biol.* **2008**, *4*, 728.
- Meinhart, A.; Kamenski, T.; Hoeppner, S.; Baumli, S.; Cramer, P. *Genes Dev.* **2005**, *19*, 1401.
- Archambault, J.; Chambers, R. S.; Kobor, M. S.; Ho, Y.; Cartier, M.; Bolotin, D.; Andrews, B.; Kane, C. M.; Greenblatt, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14300.
- Cho, H.; Kim, T. K.; Mancebo, H.; Lane, W. S.; Flores, O.; Reinberg, D. *Genes Dev.* **1999**, *13*, 1540.
- Archambault, J.; Pan, G. H.; Dahmus, G. K.; Cartier, M.; Marshall, N.; Zhang, S.; Dahmus, M. E.; Greenblatt, J. *J. Biol. Chem.* **1998**, *273*, 27593.
- Licciardo, P.; Ruggiero, L.; Lania, L.; Majello, B. *Nucleic Acids Res.* **2001**, *29*, 3539.
- Nguyen, B. D.; Abbott, K. L.; Potempa, K.; Kobort, M. S.; Archambault, J.; Greenblatt, J.; Legault, P.; Omichinski, J. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5688.
- Showalter, S. A. *Biomol. NMR Assign.* **2009**, *3*, 179.
- Lawrence, C. W.; Bonny, A.; Showalter, S. A. *Biochem. Biophys. Res. Commun.* **2011**, *410*, 461.
- Kamada, K.; Roeder, R. G.; Burley, S. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2296.
- Gajiwala, K. S.; Burley, S. K. *Curr. Opin. Struct. Biol.* **2000**, *10*, 110.
- Groft, C. M.; Uljon, S. N.; Wang, R.; Werner, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9117.
- Zheng, N.; Fraenkel, E.; Pabo, C. O.; Pavletich, N. P. *Genes Dev.* **1999**, *13*, 666.
- Zheng, N.; Schulman, B. A.; Song, L.; Miller, J. J.; Jeffrey, P. D.; Wang, P.; Chu, C.; Koepp, D. M.; Elledge, S. J.; Pagano, M.; Conaway, R. C.; Conaway, J. W.; Harper, J. W.; Pavletich, N. P. *Nature* **2002**, *416*, 703.
- Finnin, M. S.; Cicero, M. P.; Davies, C.; Porter, S. J.; White, S. W.; Kreuzer, K. N. *EMBO J.* **1997**, *16*, 1992.
- Kussie, P. H.; Gorina, S.; Marechal, V.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science* **1996**, *274*, 948.
- Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

- (43) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
- (44) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915.
- (45) Jorgensen, W. L.; Chandrasekhar, J.; D., M. J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (46) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (47) Showalter, S. A.; Bruschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961.
- (48) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- (49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
- (50) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605.
- (51) Pettersen, E. F.; G., T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605.
- (52) Wostenberg, C.; Quarles, K. A.; Showalter, S. A. *Biochemistry* **2010**, *49*, 10728.
- (53) Kobor, M. S.; Simon, L. D.; Omichinski, J.; Zhong, G.; Archambault, J.; Greenblatt, J. *Mol. Cell. Biochem.* **2000**, *20*, 7438.
- (54) Sheridan, R. P.; Levy, R. M.; Salemme, F. R. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4545.
- (55) Kamada, K.; De Angelis, J.; Roeder, R. G.; Burley, S. K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3115.
- (56) Chambers, R. S.; Wang, B. Q.; Burton, Z. F.; Dahmus, M. E. *J. Biol. Chem.* **1995**, *270*, 14962.
- (57) Yang, A.; Abbott, K. L.; Desjardins, A.; Di Lello, P.; Omichinski, J. G.; Legault, P. *Biochemistry* **2009**, *48*, 1964.
- (58) Eichner, J.; Chen, H. T.; Warfield, L.; Hahn, S. *EMBO J.* **2010**, *29*, 706.
- (59) Abbott, K. L.; Renfrow, M. B.; Chalmers, M. J.; Nguyen, B. D.; Marshall, A. G.; Legault, P.; Omichinski, J. G. *Biochemistry* **2005**, *44*, 2732.
- (60) Zhuravleva, A.; Korzhnev, D. M.; Nolde, S. B.; Kay, L. E.; Arseniev, A. S.; Billeter, M.; Orekhov, V. Y. *J. Mol. Biol.* **2007**, *367*, 1079.
- (61) Sapienza, P. J.; Mauldin, R. V.; Lee, A. L. *J. Mol. Biol.* **2011**, *405*, 378.
- (62) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* **2004**, *338*, 1015.
- (63) Chi, S. W.; Kim, D. H.; Lee, S. H.; Chang, I.; Han, K. H. *Protein Sci.* **2007**, *16*, 2108.
- (64) Lowry, D. F.; Stancik, A.; Shrestha, R. M.; Daughdrill, G. W. *Proteins* **2008**, *71*, 587.
- (65) Espinoza-Fonseca, L. M. *FEBS Lett.* **2009**, *583*, 556.