

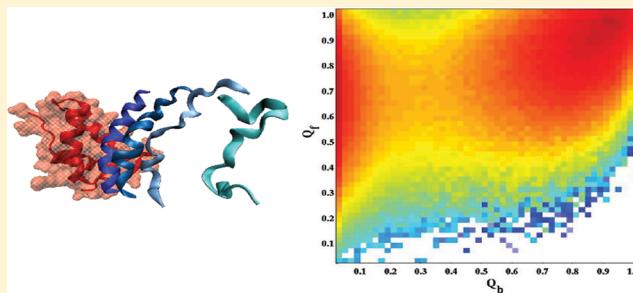
Native-Based Simulations of the Binding Interaction Between RAP74 and the Disordered FCP1 Peptide

Sushant Kumar,[†] Scott A. Showalter,^{*,‡} and William G. Noid^{*,‡}

[†]Huck Institutes of the Life Sciences, and [‡]Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, United States

 Supporting Information

ABSTRACT: By dephosphorylating the C-terminal domain (CTD) of RNA polymerase II (Pol II), the Transcription Factor IIF (TFIIF)-associating CTD phosphatase (FCP1) performs an essential function in recycling Pol II for subsequent rounds of transcription. The interaction between FCP1 and TFIIF is mediated by the disordered C-terminal tail of FCP1, which folds to form an α -helix upon binding the RAP74 subunit of TFIIF. The present work reports a structure-based simulation study of this interaction between the folded winged-helix domain of RAP74 and the disordered C-terminal tail of FCP1. The comparison of measured and simulated chemical shifts suggests that the FCP1 peptide samples 40–60% of its native helical structure in the unbound disordered ensemble. Free energy calculations suggest that productive binding begins when RAP74 makes hydrophobic contacts with the C-terminal region of the FCP1 peptide. The FCP1 peptide then folds into an amphipathic helix by zipping up the binding interface. The relative plasticity of FCP1 results in a more cooperative binding mechanism, allows for a greater diversity of pathways leading to the bound complex, and may also eliminate the need for “backtracking” from contacts that form out of sequence.



INTRODUCTION

Intrinsically disordered proteins (IDPs) defy the classical structure–function paradigm by performing a wide range of vital cellular functions, despite lacking a well-defined structure in isolation.^{1,2} IDPs function in assembling protein complexes,³ in mediating protein modifications,⁴ in chaperoning protein folding,⁵ and as entropic chains.^{6,7} Bioinformatic studies suggest that more than 30% of the proteins that are encoded by eukaryotic genomes contain extended disordered regions.² Moreover, disordered proteins have been implicated in the development of cancer, neurodegenerative diseases, and other protein misfolding diseases.⁸

Disordered proteins bind a wide range of ligands, including DNA,⁹ RNA,¹⁰ and other proteins.¹¹ Interactions with these ligands can induce disordered peptides to adopt a folded conformation upon binding.^{12,13} The structural plasticity associated with intrinsic disorder may confer unique functional advantages for molecular recognition by facilitating one-to-many and many-to-one binding,^{1,2} by accelerating molecular recognition,^{14,15} by optimizing the interfacial area available for binding,¹⁶ and by enabling binding with both low affinity and high specificity.¹²

Protein disorder appears to be particularly important for facilitating molecular recognition, especially in signal transduction and in regulating cell division,¹⁷ transcription, and translation.¹⁸ In particular, the FCP1–RAP74 interaction exemplifies the role of IDPs in regulating eukaryotic trans-

cription. During the course of transcription, the C-terminal domain (CTD) of RNA Polymerase II (Pol II) becomes extensively phosphorylated.¹⁹ The Transcription Factor IIF (TFIIF)-associating CTD phosphatase (FCP1)²⁰ performs an essential biological function in recycling Pol II for subsequent transcription by dephosphorylating the Pol II CTD.^{21,22} The association of TFIIF and FCP1 is mediated by the interaction of the folded winged-helix domain²³ that is at the C-terminal of the RAP74 subunit of TFIIF with the last 18 C-terminal residues of FCP1.^{20,24} This C-terminal FCP1 peptide is disordered in solution,²⁵ but folds to form a helix upon binding RAP74.²⁶

Many recent studies have investigated the remarkable disorder-to-order transition that couples IDP folding to ligand binding. Two limiting mechanisms are typically envisioned. In the “induced fit” mechanism,²⁷ binding proceeds before the IDP folds. Computational studies have suggested that this mechanism is favored when protein–protein interactions are relatively strong and long-ranged.²⁸ In the “conformational selection” mechanism,²⁹ binding proceeds via a folded or partially folded conformation that is sampled by the IDP prior to binding. Importantly, although IDPs may sample partially folded structures prior to binding,^{30–33} this preformed structure

Received: October 17, 2012

Revised: January 29, 2013

Published: February 6, 2013

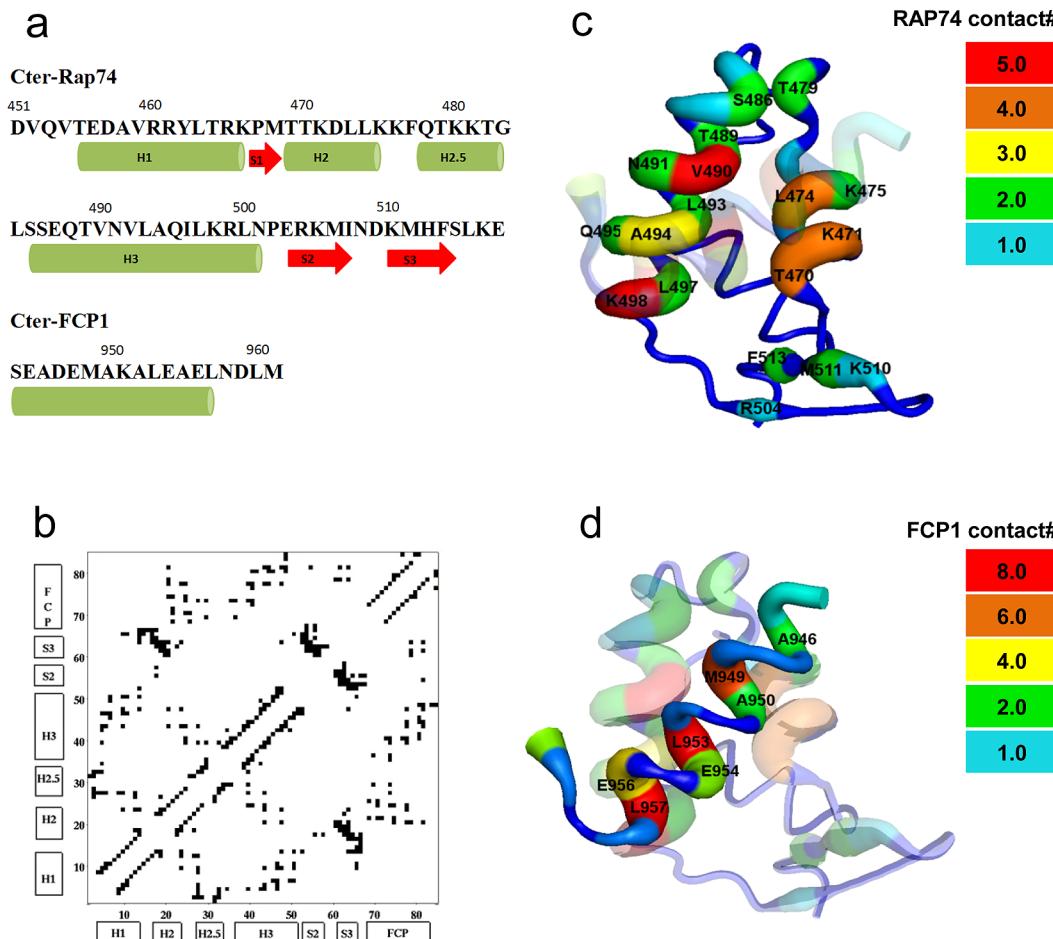


Figure 1. Analysis of the RAP74-FCP1 cocomplex crystal structure. (a) Sequence and secondary structure for the C-terminal WH domain of RAP74 (residues 451–517) and for the C-terminal (residues 944–961) of FCP1. (b) Contact map for the cocrystal structure. Structures of the complex highlighting the intermolecular contacts formed by RAP74 (c) and by FCP1 (d).

may not be necessary for the binding mechanism.^{34,35} Various intermediate mechanisms have also been discussed, including the extended conformational selection mechanism,³⁶ as well as the blending^{37,38} or concurrent superposition³⁹ of the two mechanisms. Moreover, the dominant mechanism may vary with changes in the protein sequence,⁴⁰ folding rate,⁴¹ or ligand concentration.⁴²

In addition to illustrating the biological function of protein disorder in regulating transcription, the interaction of RAP74²³ with the disordered C-terminal tail of FCP1²⁰ provides a useful model system for characterizing the structure, dynamics, and interactions of IDPs. Figure 1 presents the sequence, secondary structures, contact map, and three-dimensional structure^{26,25} for the RAP74-FCP1 complex. The RAP74 domain is an α/β winged-helix (WH) peptide that contains three larger helices (H1, H2, and H3), a short turn of a helix (H2.5), three strands (S1, S2, and S3), and two wings (W1 and W2), which are organized H1-S1-H2-H2.5-H3-S2-W1-S3-W2. The FCP1 C-terminal tail demonstrates many common features of disordered acidic transactivation domains and, when fused to the GAL4 DNA-binding domain, can function as such in vitro.⁴³ Because it is enriched in polar and acidic residues, while being relatively depleted of hydrophobic residues,⁴⁴ the C-terminal FCP1 peptide does not adopt a well-defined structure when free in solution.²⁵ However, when bound to the C-terminal folded WH domain of the RAP74 subunit of TFIIF,

the last 18 FCP1 residues fold to form an amphipathic helix.²⁶ Subsequently, we shall simply refer to these peptides as RAP74 and FCP1.

The resulting complex demonstrates several common features of IDP complexes. For instance, the hydrophobic face of the amphipathic FCP1 helix covers an exposed hydrophobic groove on the surface of RAP74 between H2 and H3.⁴⁵ Simultaneously, the polar and acidic FCP1 residues complement the polar and basic residues on the RAP74 surface that decorate the hydrophobic binding pocket.⁴⁶ However, the RAP74-FCP1 interaction is also quite intriguing because WH domains, such as RAP74, typically feature basic regions that bind DNA via the recognition helix, H3,⁴⁷ or possibly via H1.⁴⁸ Instead, the hydrophobic H2/H3 cleft of RAP74 binds the FCP1 peptide via an interaction that is remarkably similar to that observed in the DNA-mediated association of the WH domains from E2F3 and from DP2⁴⁹ and in the Cul1-Rbx1-Skp1-Fbox^{Skp2} quaternary protein complex.⁵⁰

Motivated by these considerations, our laboratories have recently initiated a joint investigation of the RAP74-FCP1 interaction using both experimental and computational techniques. Showalter and co-workers have extended and applied carbon direct-detection nuclear magnetic resonance (NMR) experiments⁵¹ for IDPs in order to achieve almost complete chemical shift assignments for unbound FCP1.^{52,53} NMR and circular dichroism experiments with denaturants,

osmolytes, and crowders revealed the presence of nascent helical structure within the FCP1 unbound disordered ensemble.⁵⁴ Atomically detailed molecular dynamics simulations of RAP74 and of the RAP74-FCP1 complex suggested that binding to FCP1 induces subtle reorganization in the RAP74 binding pocket adjacent to H2.5.⁵⁵ These simulations highlighted the role of the hydrophobic residues Leu953 and Leu957 for anchoring FCP1 in place and, moreover, demonstrated that FCP1 retains considerable flexibility even after binding RAP74, i.e., that the RAP74-FCP1 complex is quite “fuzzy.”⁵⁶ Subsequent NMR spin relaxation measurements further supported the conclusions that bound FCP1 remains quite disordered after binding, and that FCP1 slightly promotes the ordering of the RAP74 binding pocket near H2.5.⁵⁷

The present work extends these previous studies by employing molecular dynamics (MD) simulations to investigate this binding interaction and the mechanism by which it induces FCP1 folding. Several recent simulation studies have employed atomically detailed models with either explicit^{58–61} or implicit solvent^{34,62,63} to characterize IDPs and their interactions. However, these models remain prohibitively expensive to adequately explore the relevant configurational space or characterize the free energy surface that governs the binding process.⁶⁴ Consequently, we investigated the RAP74-FCP1 interaction with simplified native-based (G_0) models that are completely determined by the structure of the bound complex.⁶⁵ Native-based models, which assume that protein topology determines the folding mechanism, have contributed both qualitative and quantitative insight into the folding of globular proteins.^{66–68} Such a simplified model certainly provides a highly limited description of the ensemble of unfolded conformations sampled by unbound FCP1. Nevertheless, previous studies have demonstrated that, in many cases, the structure of the bound state determines the mechanism for the formation of IDP complexes.^{69,70} Moreover, previous simulation studies with similar native-based models have contributed remarkable insight into several IDP systems.^{71–78}

Our simulations suggest that binding proceeds when the C-terminal turn of FCP1 contacts the C-terminal region of RAP74 H3. Interactions with RAP74 then induce further folding of FCP1 by zipping up the amphipathic helix. The simulations suggest that protein disorder promotes a more cooperative binding transition and facilitates efficient binding by allowing for a greater diversity of binding paths and also by providing the flexibility necessary to avoid “backtracking” when contacts form out of sequence.⁷⁵

METHODS

Native Contacts. We employed the Structure-based Models in Gromacs (SMOG) web-server⁷⁹ (<http://smog.ucsd.edu/>) to construct coarse-grained (CG) native-based (G_0) models for RAP74 and FCP1 from the published crystal structure of the RAP74/FCP1 complex (PDBID: 1J2X).²⁶ Residues 451–517 were included in the RAP74 model, while residues 944–961 were included in the FCP1 model. The CG models represented each residue with a single site that was associated with the corresponding α carbon. The SMOG webserver identifies native contacts according to the shadow contact map algorithm.⁸⁰ This algorithm first identifies each pair of residues for which one heavy atom from the first residue is within 0.6 nm of a heavy atom from the second residue and then eliminates all pairs for which a third residue casts an

intervening shadow between. The remaining pairs that are separated in sequence by three or more residues define the native contact list.

The shadow contact map algorithm identified $M_f = 16$ intramolecular native contacts stabilizing FCP1, $M_r = 166$ intramolecular contacts stabilizing RAP74, and $M_b = 51$ intermolecular contacts stabilizing the FCP1-RAP74 complex. Figure 1b presents the complete contact map, while the Supporting Information explicitly enumerates these contacts. Figure 1c,d presents the three-dimensional crystal structure for the FCP1-RAP74 complex. The thickness and colors in Figure 1c,d indicate the total number of intermolecular native contacts formed by each residue in the complex.

The contact map includes the stabilizing intermolecular contacts that have been identified in previous experimental reports.^{26,25} In particular, the contact map highlights the role of Met949, Leu953, and Leu957 for anchoring FCP1 in the RAP74 binding pocket.⁵⁵ The contact map also highlights the RAP74 residues Thr470, Lys471 and Leu474 in H2, as well as Ser486, Val490, Ala494, and Lys 498 in H3. The contact map includes the Asp947-Lys471 and Asp959-Lys498 salt bridges that have been proposed to anchor the N- and C-terminal regions of the FCP1 peptide.^{25,26}

Native Contact Analysis. In order to analyze the folding of FCP1, we define, for each of the $M_f = 16$ intramolecular FCP1 native contacts, a function, $\Delta_{fi}(\mathbf{R})$, that quantifies the formation of contact i . For native contact i and configuration \mathbf{R} , $\Delta_{fi}(\mathbf{R})$ equals 1 if the distance between the corresponding pair of CG sites is less than or equal to 1.20 times the distance between the pair in the crystal structure. Otherwise, $\Delta_{fi}(\mathbf{R}) = 0$. $Q_f(\mathbf{R}) = (1/M_f) \sum_{i=1}^{M_f} \Delta_{fi}(\mathbf{R})$ quantifies the structural similarity of FCP1 in configuration \mathbf{R} to the folded, helical configuration observed in the cocrystal structure. $\langle Q_f(\mathbf{R}) \rangle$ is the corresponding thermodynamic average.

Following Onuchic and co-workers,^{81,82} we define the variance

$$\langle \delta^2 Q_f(\mathbf{R}) \rangle_{Q_f} \equiv \frac{1}{M_f} \sum_{i=1}^{M_f} [\langle \Delta_{fi}(\mathbf{R}) \rangle_{Q_f} - Q_f]^2 \quad (1)$$

to quantify the heterogeneity within the subensemble of FCP1 configurations that have a given similarity, Q_f , to the folded structure. In eq 1, the sum is performed over the M_f intramolecular FCP1 native contacts, and $\langle \Delta_{fi}(\mathbf{R}) \rangle_{Q_f}$ is the probability for contact i to have formed within the subensemble of configurations for which $Q_f(\mathbf{R}) = Q_f$. Given the subensemble with native content Q_f , the variance achieves a minimum value of 0 when all contacts have equal probability for forming (i.e., $\langle \Delta_{fi}(\mathbf{R}) \rangle_{Q_f} = Q_f$ for all $i = 1, \dots, M_f$). Conversely, given this subensemble, the variance achieves a maximum value of $Q_f(1 - Q_f)$ when the same $M_f Q_f$ contacts are always formed and the remaining $M_f(1 - Q_f)$ contacts have never formed. Onuchic and co-workers^{81,82} defined a normalized “route measure” for folding:

$$R_f(Q_f) = \langle \delta^2 Q_f(\mathbf{R}) \rangle_{Q_f} / [Q_f(1 - Q_f)] \quad (2)$$

The inverse of the route measure provides a quantitative measure of the diversity of configurations leading to the native state as a function of Q_f . Similarly, in order to analyze the binding process in terms of the $M_b = 51$ intermolecular native contacts, we define analogous functions $Q_b(\mathbf{R})$ and $R_b(Q_b)$ to

characterize the extent of binding and the route measure of binding routes, respectively.

Simulation Models. The energy scale for the CG model is determined by an arbitrary constant ϵ_0 , which then determines a dimensionless temperature scale $\tau = k_B T / \epsilon_0$. The potential for the C- α native-based models employed functional forms and parameters that have been previously published.⁶⁷ The bonded geometries of RAP74 and FCP1 are maintained by the bonded potential:

$$U_{\text{bonded}}(\mathbf{R}) = \sum_i^{\text{bonds}} \frac{1}{2} k_{\text{bond}} (b_i - b_{i;0})^2 + \sum_i^{\text{angles}} \frac{1}{2} k_{\text{angle}} (\theta_i - \theta_{i;0})^2 + \sum_i^{\text{dihedrals}} \epsilon_0 \left[\cos(\psi_i - \psi_{i;0}) + \frac{1}{2} \cos(3(\psi_i - \psi_{i;0})) \right] \quad (3)$$

where the first sum is over each bond with length, b_i , the second sum is over each angle, θ_i , formed by two consecutive bonds, and the third sum is over each dihedral angle, ψ_i , formed by three consecutive bonds. The constants $b_{i;0}$, $\theta_{i;0}$, and $\psi_{i;0}$ are the equilibrium bond length, bond angle, and torsional angles, respectively. These constants are defined, for each degree of freedom, by their value in the crystal structure. In the C- α models, the harmonic bond and angle constants are defined as $k_{\text{bond}} = 2 \times 10^4 \epsilon_0 / \text{nm}^2$ and $k_{\text{angle}} = 40 \epsilon_0 / \text{deg}^2$.

The nonbonded contribution to the total potential includes contributions from each pair of CG sites that are separated by more than three bonds:

$$U_{\text{nb}}(\mathbf{R}) = \sum_{J>I+3}^{\text{native}} \epsilon \left[5 \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^n - 6 \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^m \right] + \sum_{J>I+3}^{\text{nonnative}} \epsilon_0 \left(\frac{\sigma_0}{R_{IJ}} \right)^{12} \quad (4)$$

The first sum is over all native contacts, while the second sum is over all nonnative pairs, i.e., every other pair of residues that is separated by more than three bonds. For each pair of CG sites, $\{I, J\}$, R_{IJ} is the distance between the pair in configuration \mathbf{R} , and σ_{IJ} is the distance between the pair in the cocrystal structure. In the C- α native-based models, $n = 12$, $m = 10$, $\sigma_0 = 0.4 \text{ nm}$, and all nonbonded interactions were truncated at 2.5 nm.

In the following, we present results for three C- α models that are distinguished by the value of ϵ employed to model native contacts and also by the simulated dimensionless temperature, τ . (1) In the weakly interacting, strongly folding (WISF) model, $\epsilon = 0.58 \epsilon_0$ for all intermolecular RAP74-FCP1 native contacts, $\epsilon = \epsilon_0$ for all other native contacts, and $\tau = 0.83$. Because this temperature is at the collapse temperature for FCP1, the unbound FCP1 peptide samples relatively folded conformations with $\langle Q_f \rangle = 0.63$. Given these parameters, the WISF model ($K_D = 0.14 \mu\text{M}$) qualitatively reproduces the experimentally observed ($K_D = 0.5 \mu\text{M}$) binding affinity between the two proteins.⁸³ (2) In the strongly interacting, weakly folding (SIWF) model $\epsilon = 2 \epsilon_0$ for all intramolecular RAP74 native contacts, $\epsilon = \epsilon_0$ for all other native contacts, and $\tau = 1.33$. Because this temperature is above the collapse temperature for FCP1, the unbound FCP1 peptide samples relatively unfolded conformations with $\langle Q_f \rangle = 0.42$. RAP74 remains folded at this high temperature, though, because the

strength of the intramolecular RAP74 interactions are doubled in the SIWF model. Moreover, because the intermolecular FCP1-RAP74 interactions are sufficiently strong, the SIWF model ($K_D = 0.40 \mu\text{M}$) also reproduces the experimentally observed binding affinity.⁸³ (3) In the strongly interacting, strongly folding (SISF) model, $\epsilon = \epsilon_0$ for all native contacts and $\tau = 0.83$. As in the case of the WISF model, $\langle Q_f \rangle = 0.63$ for the unbound peptide. However, in the SISF model, binding is essentially irreversible. The calculation of dissociation constants is described below.

In addition, we also briefly present simulations of the free FCP1 peptide with an atomically detailed native-based model, which was constructed with the SMOG webserver in an analogous manner.^{79,84} The only differences between the C- α and atomistic native-based models are that (1) the atomistic model explicitly included sites for each heavy atom in both proteins; (2) in the atomistic model, the attractive nonbonded Lennard-Jones potentials for native (intramolecular FCP1) contacts were modeled with $\epsilon = \epsilon_0$, $n = 12$, and $m = 6$; (3) in the atomistic model, the repulsive nonbonded potentials for nonnative interactions were modeled with $\sigma_0 = 0.25 \text{ nm}$ and were also rescaled by a factor of 0.01; and (4) the atomistic model also employed a weak improper torsion potential to model the planarity of the amide planes in the backbone of each polypeptide. The parameters for this model have been previously published.⁸⁴

Simulation Details. All MD simulations were performed with the Gromacs 4.0.5 software package.⁸⁵ The energy scale was set to $\epsilon_0 = 1 \text{ kJ mol}^{-1}$, which then defines a corresponding dimensional temperature scale $T_0 = \epsilon_0 / k_B = 120.27 \text{ K}$. Simulations employed “Gromacs” temperatures $T_G = \tau T_0$, where τ is the reported dimensionless temperature. All masses were set to 1 amu and the stochastic dynamics algorithm was employed to propagate canonical dynamics with a 0.5 fs time step. All simulations of the FCP1-RAP74 interaction were performed in a periodic cubic box with sides of 12.0 nm. Sampling for free energy calculations was obtained by employing harmonic umbrella potentials to bias sampling within 23 windows.⁸⁶ For $j = -2, \dots, 20$, the j th window employed a bias potential $V_{\text{bias},j}(\mathbf{R}) = (k_{\text{bias}}/2)(R_{\text{com}} - R_j)^2$, where $k_{\text{bias}} = 500 \epsilon_0 / \text{nm}^2$, R_{com} is the distance between the FCP1 and RAP74 centers of mass in configuration \mathbf{R} , $R_j = j \times 0.1 \text{ nm} + R_0$, and $R_0 = 1.17 \text{ nm}$ is the distance between the protein mass centers in the cocrystal structure. Each window was independently simulated for 40 ns using the cocrystal structure as an initial configuration. The first 2 ns of each window were treated as equilibration and discarded, while the remaining 38 ns were employed in subsequent analysis.

Data Analysis. Chemical shifts for the α , β , and carbonyl carbons were calculated from atomically detailed simulations of the FCP1 peptide with the SHIFTS 4.3 software package^{87,88} according to the protocol of Li and Brüschweiler.⁸⁹ Additional data analysis employed GROMACS utilities, Pymol,⁹⁰ MATLAB,⁹¹ and in-house software. Thermodynamic averages, pair potentials of mean force, and free energy surfaces were estimated from biased simulations after applying the weighted histogram analysis method.^{92,93} From these simulations, we calculated the probability, $P_R(R_{\text{com}})$, for the two protein mass centers to be separated by a distance R_{com} , as well as joint probability distributions for Q_f and Q_b , and for Q_f and R_{com} . The dimensionless pair potential of mean force⁹⁴ was calculated $W(R_{\text{com}})/k_B T = -\ln[P_R(R_{\text{com}})/R_{\text{com}}^2] + C$. The last grid point for the calculations corresponded to $R_\infty = 3.17 \text{ nm}$, and C was

determined so that $W(R_\infty) = 0$. The thermodynamic average, \bar{U} , for the total potential, U , was calculated as a function of R_{com} , $\bar{U}(R_{\text{com}}) = \langle U(R) \rangle_{R_{\text{com}}}$, and its distance-dependence was then analyzed as $\Delta U(R_{\text{com}}) = \bar{U}(R_{\text{com}}) - \bar{U}(R_\infty)$. Similar analysis was performed for contributions to U from the FCP1 intramolecular potential, U_b and from the FCP1–RAP74 intermolecular potential, U_b . The contribution from configurational entropy to the pair potential of mean force was computed $\Delta S(R_{\text{com}})/k_B = -[W(R_{\text{com}})/k_B T - \Delta U(R_{\text{com}})/k_B T]$. Two dimensional free energy surfaces were computed by direct Boltzmann inversion of the corresponding joint probability distributions. Binding constants were calculated in units of $M = \text{mol/L}$ according to^{76,78}

$$K_D/M = \frac{1660}{V_0} \frac{p_{ub}^2}{1 - p_{ub}} \quad (5)$$

where V_0 is the volume expressed in units of \AA^3 and p_{ub} is the probability for the complex to sample unbound configurations, i.e., configurations with $Q_b < 0.1$.

RESULTS

Characterization of the Disordered FCP1 Ensemble. In order to obtain a preliminary analysis of the FCP1 unfolded ensemble, we performed simulations of unbound FCP1 with an atomically detailed native-based model^{79,84} at a range of temperatures. For each simulated ensemble, we computed chemical shifts for the α , β , and carbonyl carbons^{87–89} for comparison with chemical shifts measured in NMR experiments.⁵² The measured chemical shifts for the β and carbonyl carbons do not discriminate between the simulated ensembles. Chemical shifts for β carbons provide relatively little information regarding the formation of α -helices.⁹⁵ Consequently, the experimentally measured chemical shifts for the β carbon quantitatively agree with those calculated for the simulated ensembles, irrespective of simulated temperature. At the same time, because the native-based model does not accurately describe the peptide amide plane, the measured carbonyl chemical shifts compare poorly with those calculated for each ensemble. In contrast, the α carbon chemical shifts do discriminate between the simulated ensembles for unbound FCP1.

Figure 2 presents the root mean squared deviation (RMSD), σ , between the measured and calculated α carbon chemical

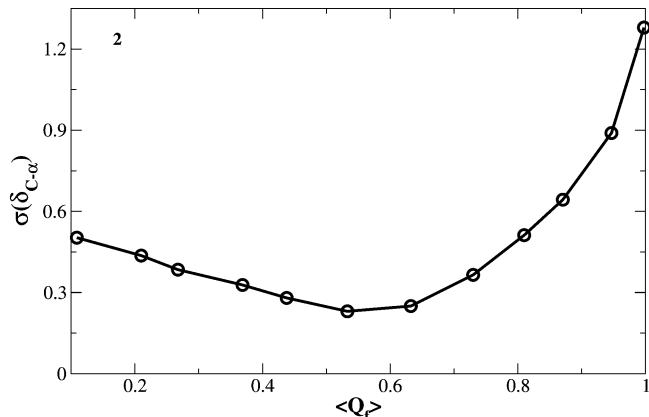


Figure 2. Comparison of measured and simulated chemical shifts as a function of helical content within the simulated ensemble, $\langle Q_f \rangle$.

shifts, $\delta_{C-\alpha}$ as a function of the average helical content, $\langle Q_f \rangle$, for the simulated FCP1 ensemble. The measured α carbon chemical shifts are most accurately reproduced ($R^2 \geq 0.98$) for ensembles with $0.42 < \langle Q_f \rangle < 0.63$, while the best agreement is observed for the ensemble with $\langle Q_f \rangle = 0.53$. These results are consistent with prior experimental⁵⁴ and computational studies⁹⁶ that have suggested that the disordered FCP1 ensemble samples nascent helical structure prior to binding RAP74.

We then employed unbiased MD simulations to characterize the folding of unbound FCP1 peptide in the C– α native-based model. Figure 3 presents the average extent of folding, $\langle Q_f \rangle$, the

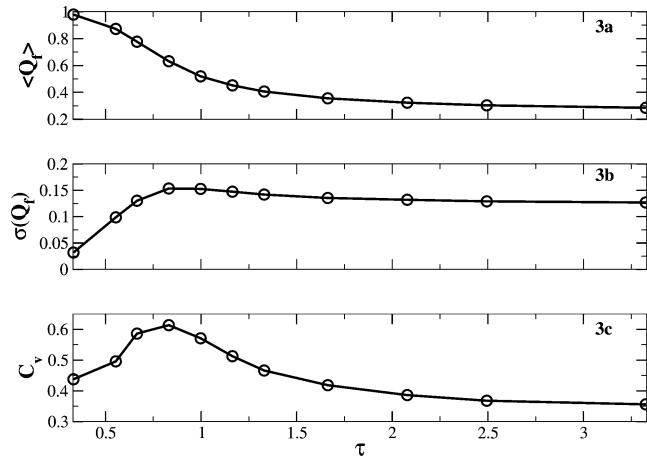


Figure 3. Analysis of folding in the C– α native-based model for FCP1 as a function of dimensionless temperature, τ . Row a presents the average content of native helical structure; Row b presents the RMSD in native helicity; Row c presents the constant volume specific heat curve.

RMSD describing fluctuations in Q_f , $\sigma(Q_f)$, and the heat capacity, C_V , as a function of the dimensionless temperature, τ . As expected, the C– α model for the disordered FCP1 peptide does not demonstrate a cooperative folding transition. Instead, $\langle Q_f \rangle$ gradually decreases from 1.0 to 0.3 as τ increases from 0.33 to 3.3. The collapse temperature, which is defined by the peak of the C_V curve, is $\tau_c = 0.83$ and corresponds to $\langle Q_f \rangle = 0.63$.

While Figure 2 indicates that the experimental chemical shifts are best described by an ensemble for which $0.4 < \langle Q_f \rangle < 0.6$, Figure 3 demonstrates that these ensembles correspond to temperatures $1.33 > \tau > 0.83$ in the C– α model. In order to investigate the significance of preformed structure in the FCP1 ensemble for the FCP1–RAP74 interaction, subsequent calculations focus on two models: (1) strongly interacting, but weakly folded (SIWF) model at a relatively high temperature $\tau = 1.33$ (corresponding to $\langle Q_f \rangle = 0.42$ for unbound FCP1); and (2) weakly interacting, but strongly folded (WISF) model at a relatively low temperature $\tau = 0.83$ (corresponding to $\langle Q_f \rangle = 0.63$ for unbound FCP1). In the WISF model, the strength, ε , of attractive intermolecular interactions, corresponding to RAP74–FCP1 native contacts, is rescaled by a factor of 0.58 relative to the SIWF model. The WISF and SIWF models roughly correspond to the maximum and minimum extent of native structure within the FCP1 unbound ensemble that is most consistent with the experimentally determined chemical shifts. Moreover, the dissociation constants for the WISF and SIWF models ($K_D = 0.14 \mu\text{M}$ and $0.40 \mu\text{M}$, respectively) indicate that both models

are reasonably consistent with the experimentally determined binding affinity for the RAP74–FCP1 system ($K_D = 0.5 \mu\text{M}$).⁸³ Finally, the Supporting Information section provides results for a second WISF model in which the strength of intermolecular interactions are further weakened (ε is reduced from $0.58\varepsilon_0$ to $0.50\varepsilon_0$). Although in some cases it provides complementary information, this model will not be discussed in detail because it does not reproduce the experimentally determined binding affinity.

Pair Potential of Mean Force. Figure 4 presents, for the WISF and SIWF models, the dimensionless pair potential of

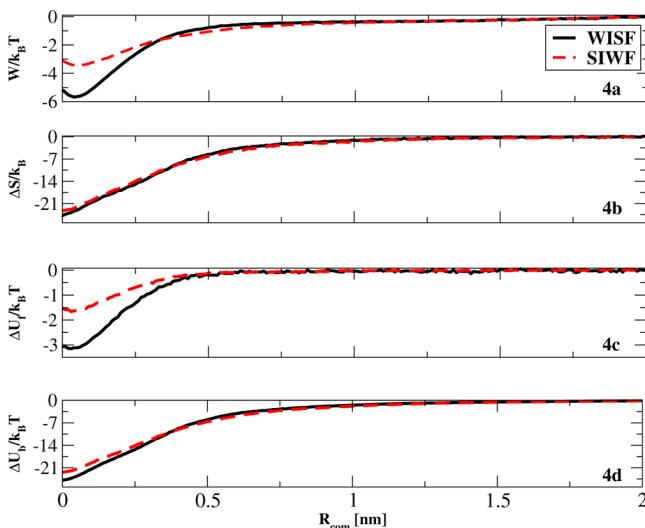


Figure 4. Row a presents the pair potential of mean force between RAP74 and FCP1 as a function of distance between their mass centers, R_{com} . Row b presents the contribution to the pair potential of mean force from configurational entropy. Rows c and d present the distance-dependent changes in folding and binding energy, respectively.

mean force (ppmf), $W/k_B T$, as a function of the distance, R_{com} , between the FCP1 and RAP74 centers of mass. In Figure 4 and subsequently, R_{com} has been reduced by $R_0 = 1.17 \text{ nm}$, which corresponds to the distance between the protein centers in the crystal structure, i.e., $R_{\text{com}} = 0 \text{ nm}$ when FCP1 is bound in the RAP74 hydrophobic groove.

The thermodynamics of FCP1 binding are remarkably similar in the SIWF and the WISF models. Figure 4 suggests that both models demonstrate a “downhill” binding scenario without a significant free energy barrier. (A more detailed analysis would be necessary to rigorously determine the exact transition state ensemble and associated free energy barrier.⁹⁷) In both models, FCP1 loses approximately $24k_B$ in configurational entropy upon binding RAP74. The distance dependence of this entropy loss appears almost identical for the two models. In both models, interactions with RAP74 only induce significant FCP1 folding when the proteins are within 0.5 nm of contact. Moreover, because both models have similar binding affinities, they demonstrate almost identical distance-dependent binding energies. This binding energy almost exactly balances the associated entropy loss.

Figure 4 also indicates several minor differences in the thermodynamics of binding for the two models. The ppmf for binding is approximately $2k_B T$ more favorable in the WISF model than in the SIWF model. This difference appears to primarily arise from the difference in favorable folding energies. The WISF and SIWF models gain approximately $3k_B T$ and

$1.5k_B T$ in favorable folding energy, respectively, corresponding to the folding of 3 and 1.5 additional native contacts upon binding RAP74. It seems somewhat surprising that interactions with RAP74 have relatively greater influence upon FCP1 folding in the WISF model, because, in the WISF model, unbound FCP1 demonstrates significantly greater structure and interacts with RAP74 via comparatively weaker pair potentials. Presumably, this result arises as a consequence of the relatively stronger intramolecular folding interactions in the WISF model. Finally, it is interesting to note that if the well depth of the pair potentials that correspond to native contacts is slightly reduced from $\varepsilon = 0.58\varepsilon_0$ in the WISF model to $\varepsilon = 0.50\varepsilon_0$, then a free energy barrier for binding emerges as FCP1 approaches RAP74. This barrier emerges because, in this case, although FCP1 loses configurational entropy less rapidly as it approaches RAP74, it also gains less favorable energetic stabilization, and this stabilization occurs later in the binding mechanism. (See Supporting Information.)

Two-Dimensional Free Energy Surfaces. Column a of Figure 5 presents two-dimensional intensity plots of the free energy surfaces for the RAP74–FCP1 interaction in the WISF and SIWF models as a function of Q_f and Q_b , which describe the extent of FCP1 folding and binding, respectively. The two-dimensional free energy surfaces present similar descriptions of the FCP1–RAP74 interaction in the two models. In both models, FCP1 samples a highly heterogeneous ensemble prior to binding. Binding does not proceed via the most folded or least folded state. Rather, as Q_b increases from 0 to 0.2 during the initial stages of binding, the ensemble of FCP1 unfolded structures narrows around the center of the Q_f distribution. This subensemble then forms a large number of binding contacts and Q_b increases from 0.2 to 0.5 without a concomitant increase in FCP1 folding, as might be expected for a conformational selection mechanism. In the final stage of binding, interactions with RAP74 drive additional folding of FCP1, as Q_b and Q_f increase together in a linear manner, as might be expected for an induced fit mechanism.

Figure 5a also demonstrates several small differences in the free energy surfaces for binding in the two models. By construction, unbound FCP1 samples more helical structure in the WISF model than in the SIWF model. Consequently, in the WISF model, the binding transition (i.e., for $0.2 < Q_b < 0.5$) occurs via more folded FCP1 configurations, subsequent induced folding occurs over a more narrow range in Q_f , and ultimately results in a complex with a more complete binding interface and FCP1 helix. Additionally, the final stage of induced folding demonstrates the slightly different balance of forces driving this process in the two models. In the WISF model intramolecular interactions appear to drive FCP1 folding (i.e., induced folding occurs slightly above the diagonal with $Q_f > Q_b$), while in the SIWF model intermolecular interactions appear to drive FCP1 folding, (i.e., induced folding occurs slightly below the diagonal with $Q_f < Q_b$).

Column b of Figure 5 presents intensity plots of the two-dimensional free energy surfaces for the FCP1–RAP74 interaction as a function of Q_b and R_{com} . This calculation reinforces the conclusion from Figure 4 that the binding transition occurs when FCP1 is within 0.5 nm of the RAP74 binding pocket. Figure 5 also reinforces the conclusion that both models demonstrate a similar binding mechanism, although the free energy surface is slightly more diffuse in the SIWF model as FCP1 samples larger fluctuations away from RAP74 in the bound state. These larger fluctuations presumably

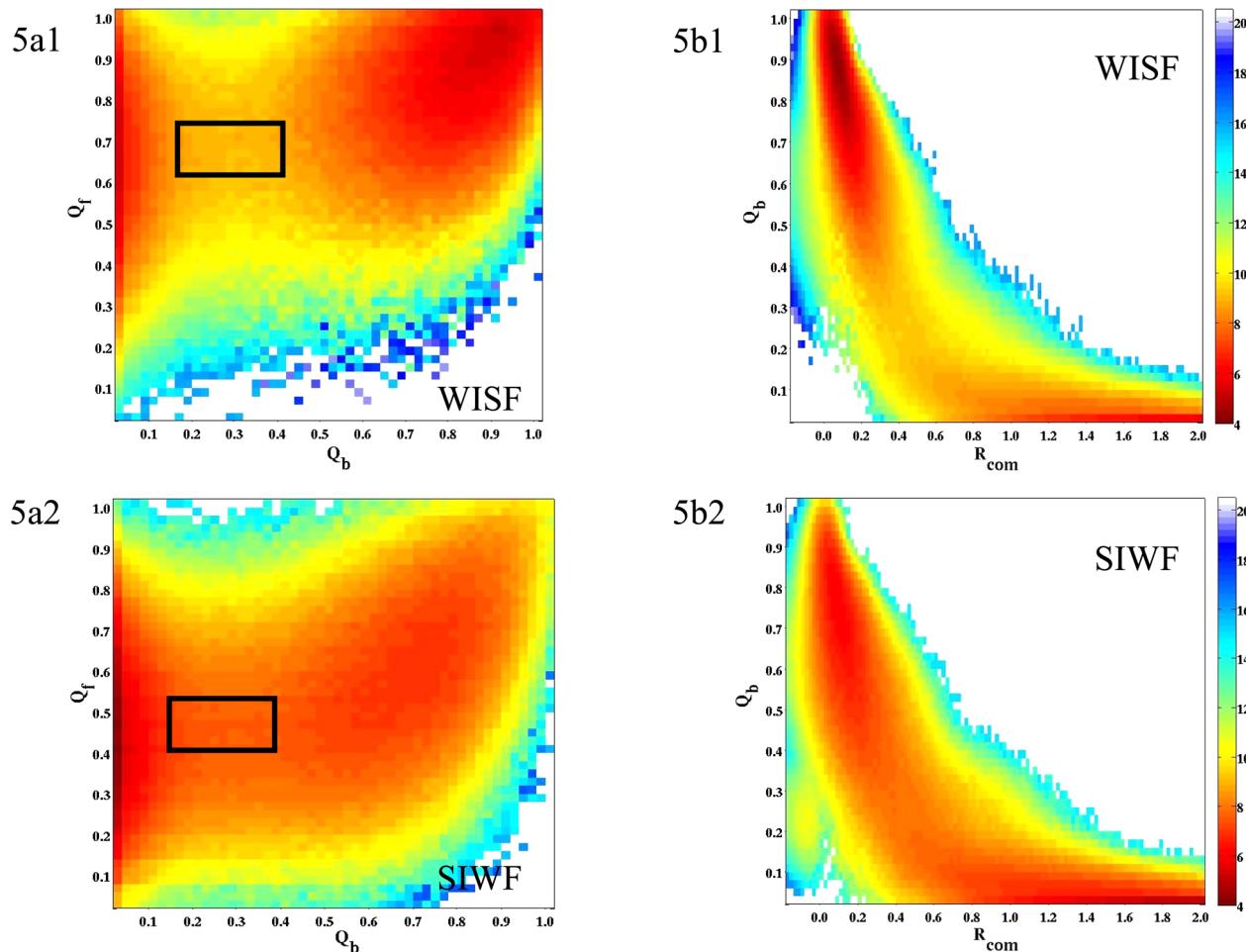


Figure 5. Intensity plots of two-dimensional free energy surfaces describing the RAP74–FCP1 interaction. Column a presents the free energy as a function of Q_f and Q_b , while column b presents the free energy as a function of Q_b and R_{com} . Rows 1 and 2 present results for the WISF and SIWF models, respectively. As shown by the scales, colors denote free energies in units of $k_B T$.

result from the higher temperature employed in the SIWF model.

Figure 6 analyzes the binding transition in the WISF model by characterizing the subensemble of configurations corresponding to the transition region of the free energy surface that is enclosed by the box superimposed upon Figure 5a. (Analysis of the SIWF model leads to qualitatively similar conclusions with the only difference being that the intermolecular contacts are less well developed in the WISF model.) Figure 6a enumerates the average number of intermolecular native contacts formed by each residue. Figure 6b presents the probability of observing each native contact in this subensemble. In these transition configurations, the interactions between the FCP1 C-terminal region and the RAP74 H3 are quite well developed, while the FCP1 N-terminal region has formed very few contacts. In particular, Leu957 and Leu953 on the hydrophobic face of the FCP1 amphipathic helix, as well as Val490 in the RAP74 hydrophobic binding groove, appear to play key roles in stabilizing the transition structures. Polar residues, such as Lys498 in RAP74 and, to a lesser extent, Glu954 and Glu956 in FCP1, also stabilize the transition complex. Finally, additional analysis (not shown) of these transition structures indicates the formation of several non-native contacts that are not explicitly stabilized by the native-centric G_0 model potential. Not surprisingly, the transition

structures bring into contact the polar residues that flank the hydrophobic cleft of the RAP74 binding site and the hydrophobic face of FCP1, suggesting that nonnative polar interactions may also contribute to the binding mechanism. In addition, and somewhat more surprisingly, the binding transition also appears to induce a subtle shift of S3 into closer proximity with both H1 and H2.

Binding Mechanism. In order to further investigate the binding interaction, we have analyzed the formation of native intermolecular contacts. Our analysis disregarded the loop region of FCP1 (residues 958–961), since it does not appear to form significant contacts with RAP74. The remaining residues of FCP1 are decomposed into three sequential groups: (1) the FCP1 N-terminal turn, Ser944–Ala946, which forms seven contacts with H2 and H2.5 of RAP74; (2) the FCP1 central helical region, Asp947–Glu954, which binds across the RAP74 hydrophobic groove and forms 25 contacts with H2 and H3; and (3) the FCP1 C-terminal turn, Ala955–Leu957, which forms 13 contacts, including 10 with the C-terminal end of RAP74 H3.

Figure 7a presents the average number of intermolecular FCP1–RAP74 native contacts formed by the N-terminal (Figure 7a1), central helix (Figure 7a2), and C-terminal (Figure 7a3) regions of FCP1, as a function of the distance, R_{com} between protein centers. The solid black and dashed red

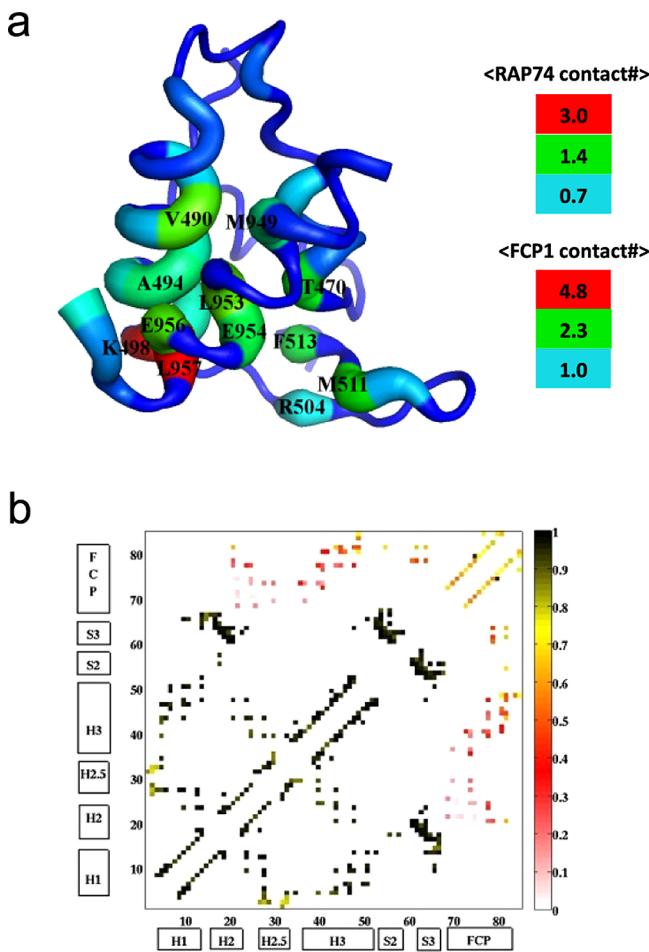


Figure 6. Analysis of transition structures in the WISF model as identified by the box in Figure 5a1. Figure 6a enumerates the observed intermolecular native contacts, while Figure 6b presents the probability for each native contact to form.

curves correspond to the WISF and SIWF models, respectively. The dotted blue curve corresponds to a third model: strongly interacting, strongly folded (SISF) model. The SISF model is simulated at the same low temperature ($\tau = 0.83$) as the WISF model, so that unbound FCP1 is relatively folded ($\langle Q_f \rangle = 0.63$), but also interacts with RAP74 as strongly as in the SIWF model.

These calculations indicate that productive binding primarily proceeds via a sequential zipping mechanism that starts with the FCP1 C-terminal turn, continues through the central helix, and finally is completed at the FCP1 N-terminal turn. Although the FCP1 N-terminal turn does form a few weak contacts with RAP74 at longer distances, these contacts do not appear to lead to productive binding. Either relatively weak intermolecular interactions (WISF model) or protein flexibility (SIWF model) permit binding to proceed without requiring that these contacts disappear from the simulated ensemble. However, if the intermolecular interactions are strong and FCP1 is also highly folded (SISF model), then the intermolecular contacts that form via the FCP1 N-terminal region must break for productive binding. Moreover, in the SISF model, the binding mechanism requires unfolding of intramolecular native contacts that form in FCP1 prior to complete binding. (See Supporting Information.)

Figure 7a indicates that when the proteins are separated by more than 0.5 nm, the WISF and SIWF models appear to form very similar contacts. However, as the proteins approach contact, the WISF and SIWF models demonstrate greater differences. When the proteins are separated by less than 0.5 nm, the C-terminal of FCP1 begins to form more intermolecular contacts in the WISF than in the SIWF model. Moreover, when the proteins are approximately 0.2 nm from contact, the helical and N-terminal regions of FCP1 also make more intermolecular contacts with RAP74 in the WISF than in the SIWF model.

Figure 7b further investigates the cooperativity of the FCP1-RAP74 binding by presenting a scatter plot of the data from Figure 7a. For each point in Figure 7b, the ordinate presents the average number of intermolecular contacts formed by the FCP1 C-terminal turn, $\langle n_C \rangle_{R_{\text{com}}}$, when the protein centers are separated by R_{com} . The circles and crosses present, for a given protein separation, the corresponding number of intermolecular contacts formed by the central helical region, $\langle n_H \rangle_{R_{\text{com}}}$, and by the N-terminal region, $\langle n_N \rangle_{R_{\text{com}}}$, of FCP1, respectively. Figures 7b1, 7b2, and 7b3 present this analysis for the WISF, SIWF, and SISF models, respectively. Clearly, binding is most sequential in the SISF model. In this model, the FCP1 C-terminal turn forms two-thirds of its intermolecular contacts before the central region of FCP1 undergoes significant binding. In contrast, in the WISF and SIWF models, binding proceeds in a much more cooperative fashion, as contacts form in the C-terminal and central regions almost simultaneously. Similarly, FCP1 also folds more cooperatively in the WISF and SIWF models, but more sequentially in the SISF model. Interestingly, if the strength of intermolecular interactions in the WISF model is further reduced (Supporting Information), then contacts between the two proteins develop still later in the binding process with the consequence that binding proceeds even more cooperatively.

Routing of the FCP1-RAP74 Interaction. Figure 8 presents the route measure,^{81,82} R , to quantify the diversity of paths by which FCP1 couples its folding to the binding of RAP74. When the route measure, R , is plotted as a function of the number, n , of contacts formed, a value of 1 indicates that all configurations with exactly n contacts have formed the same n contacts, i.e., the process is completely routed through that particular set of contacts. Conversely, as $R(n)$ decreases toward 0, the more different arrangements of the n contacts are sampled.

Figure 8a presents the route measure for FCP1 folding, R_f , as a function of the number, n_f , of native intramolecular folding contacts. Figure 8a indicates that the folding of FCP1 is only very weakly routed in the SIWF and WISF models, i.e., that a very large number of different paths lead to folding. Moreover, the diversity of paths remains roughly constant throughout the folding process. Furthermore, for the SIWF and WISF models, the Supporting Information demonstrates that the routing in the folding mechanism appears quite insensitive to the presence of RAP74. In contrast, in the SISF model, for which unbound FCP1 is relatively stable and also strongly interacts with RAP74, FCP1 folding is much more strongly routed in the presence of RAP74.

Figure 8b presents the route measure for the binding process, R_b , as a function of n_b , the number of intermolecular binding contacts. As is the case for folding, binding is more strongly routed in the SISF model than in the SIWF and WISF models.

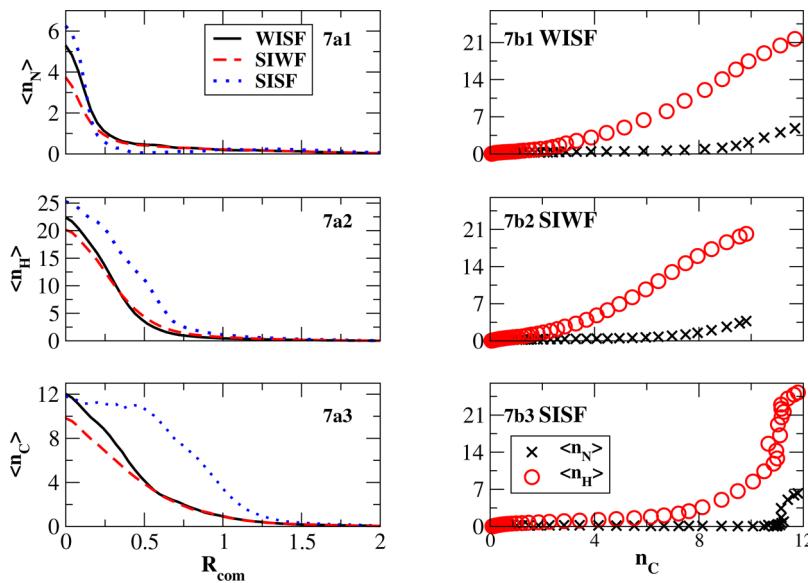


Figure 7. Formation of native intermolecular binding contacts. In column a, panels 1, 2, and 3 present, as a function of protein separation, the average number of intermolecular contacts formed by the N-terminal turn, central helical region, and C-terminal turn of FCP1, respectively. The solid black, dashed red, and dotted blue curves correspond to the WISF, SIWF, and SISF models, respectively. Column b presents the data from column a as a scatter plot using, as an ordinate, the average number of intermolecular contacts formed by the FCP1 C-terminal turn at a given interprotein separation, i.e., $\langle n_C \rangle_{R_{\text{com}}}$. The black crosses and red circles are the corresponding averages for the number of intermolecular contacts formed by the N-terminal turn, $\langle n_N \rangle_{R_{\text{com}}}$, and by the central helical region, $\langle n_H \rangle_{R_{\text{com}}}$, of FCP1, respectively. In column b, rows 1, 2, and 3 correspond to the WISF, SIWF, and SISF models, respectively.

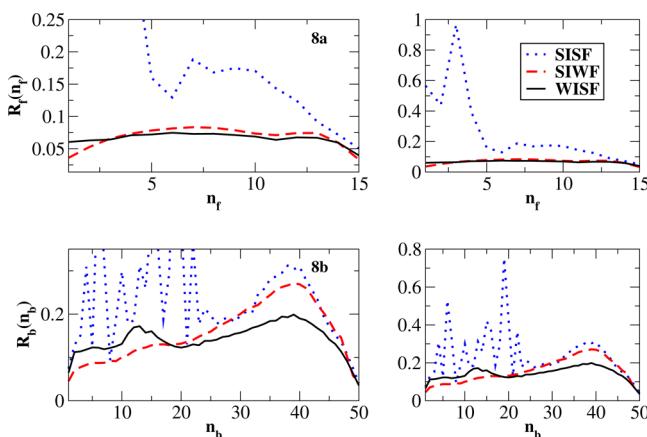


Figure 8. Route measure for FCP1 folding (Row a) and for the binding interaction (Row b). In each row, the left and right panels present the same data on different scales. The solid black, dashed red, and dotted blue curves correspond to the WISF, SIWF, and SISF models, respectively.

However, Figure 8b also indicates that, in the SIWF and WISF models, the binding mechanism becomes increasingly routed as binding proceeds (i.e., as n_b increases). A moderately narrow bottleneck appears when $Q_b \approx 0.8$ ($n_b \approx 40$), at which point the relative diversity of binding paths has been reduced by a factor of 2–4. This routing is more significant in the SIWF model than for the WISF model, which suggests that intermolecular interactions may be more significant than intramolecular interactions for restricting the number of paths to the final bound complex.

■ DISCUSSION AND CONCLUSIONS

The present work continues our recent investigations of the interaction between the disordered C-terminal peptide of FCP1 and the folded C-terminal winged-helix domain of the RAP74 subunit of transcription factor TFIIF. Circular dichroism and NMR experiments in the presence of various cosolutes demonstrated that the disordered FCP1 peptide samples nascent helical structure in its unbound state.⁵⁴ Atomically detailed, explicit solvent MD simulations of apo-RAP74 and of the FCP1-RAP74 complex suggested that FCP1 modestly impacts the region of the RAP74 binding pocket that is near H2.5.⁵⁵ These simulations also demonstrated that the RAP74-FCP1 complex is quite “fuzzy,” i.e., that bound FCP1 retains considerable conformational flexibility.⁵⁶ More recent NMR spin relaxation experiments provided further support for these conclusions by quantifying the dynamic fluctuations of both proteins before and after binding.⁵⁷ In combination, these studies suggested that FCP1 may lose surprisingly little conformational entropy upon binding RAP74.

The present work extends these studies by reporting the first molecular simulations investigating the binding mechanism underlying the RAP74-FCP1 interaction. These initial simulations employed simplified native-based models. These models provide a highly limited description of interactions in the disordered ensemble and, in particular, neglect electrostatic and nonnative interactions. Nevertheless, similar native-based models have provided a successful description of several IDP systems.^{69–78} Moreover, these models provide the efficiency that is necessary for characterizing the free energy surface guiding the RAP74-FCP1 interaction.

Preliminary simulations employed an atomically detailed native-based model to characterize the disordered ensemble sampled by free FCP1. The comparison of simulated and measured chemical shifts⁵² suggested that the FCP1 peptide

contains 40–60% native helical structure prior to binding RAP74, which agrees with prior experimental⁵⁴ and computational⁹⁶ studies. In order to investigate the role of this preformed structure for the binding interaction, subsequent simulations considered two C- α native-based models. In the WISF model, the unbound FCP1 ensemble contained on average 60% native helicity, but the FCP1–RAP74 intermolecular interactions were relatively weak. In the SIWF model, the unbound FCP1 ensemble contained on average only 40% native helicity, but the FCP1–RAP74 intermolecular interactions were relatively strong. Both models were reasonably consistent with the experimentally determined binding affinity for the complex.

Free energy calculations for the two models suggest very similar mechanisms for the RAP74–FCP1 binding interaction and its coupling to FCP1 folding. FCP1 samples a highly heterogeneous ensemble prior to encountering RAP74. In the initial stages of the encounter, before the binding interface is 20% formed, the distribution of structures within the FCP1 ensemble narrows about its center. Once the proteins are within approximately 0.5 nm of contact, binding proceeds without further folding of FCP1. The simulations suggest that the dominant binding mechanism begins with the FCP1 C-terminal turn contacting the C-terminal residues of RAP74 H3. Once these contacts have formed, folding and binding then proceed simultaneously as the FCP1 helix zips up from the C-terminal region. The central region of FCP1 next binds to H3 and H2 of RAP74 and, finally, the N-terminal turn of the FCP1 helix binds the RAP74 binding pocket near H2.5. It is intriguing to note that RAP74 makes 6 more intermolecular contacts with the C-terminal region than with the N-terminal region of FCP1, which may possibly explain why productive binding proceeds from the FCP1 C-terminal region. This clearly motivates subsequent simulations with G $\ddot{\text{o}}$ models that treat favorable nonnative interactions.

Despite the differences in the unbound FCP1 ensembles and in the energetic strength of intermolecular interactions for the two models, the two models also demonstrate very similar thermodynamics for the binding interaction. In both models, the pair potential of mean force suggests that association proceeds without a noticeable free energy barrier. Moreover, the two models demonstrate nearly identical distance-dependent entropy losses associated with binding. Furthermore, despite differences in the energetic strength of intermolecular interactions, the two models demonstrate very similar distance-dependent gains in favorable binding energy.

Further analysis identifies several minor differences in the binding processes for the WISF and SIWF models. Because both models demonstrate similar binding affinity and the WISF model includes relatively stronger intramolecular FCP1 interactions, interactions with RAP74 are more effective in inducing FCP1 folding in the WISF model than in the SIWF model. In addition, the two-dimensional free energy surfaces demonstrate that the later stages of binding are driven by intramolecular interactions in the WISF model and by intermolecular interactions in the SIWF model. Finally, although both the WISF and SIWF models sample a large number of paths leading to productive binding, the final stages of binding appear to be slightly more strongly routed in the SIWF model.

In order to further investigate the role of protein disorder for the FCP1–RAP74 interaction, we performed additional simulations with a third model, the SISF model, for which

unbound FCP1 is relatively folded and also interacts strongly with RAP74. As observed in previous simulations with native-based models, binding and folding occur more cooperatively for disordered than strongly folded proteins.⁷⁸ In the SISF model, the C-terminal turn of FCP1 almost completely binds before the central and N-terminal regions begin to bind. In contrast, by weakening the strength of intermolecular interactions, the C-terminal and central helical regions of FCP1 can bind almost simultaneously. Furthermore, in the SISF model, FCP1 folding and binding both demonstrate “backtracking,” i.e., native contacts that form early in the binding process must completely disappear from the ensemble in order to reach the final bound structure.⁷⁵ In contrast, in the WISF and SIWF models, these contacts do not disrupt the mechanism for coupled folding-binding. Finally, binding and folding are much more strongly routed in the SISF model than in either the SIWF or WISF models. These observations provide further support for the proposed role of intrinsic disorder in allowing a greater fraction of encounters to lead to productive binding.¹⁵ Given the topology of the RAP74 binding pocket, folded peptides that interact strongly with RAP74 may be more likely to start down an unproductive binding path or become kinetically trapped in a rugged free energy surface. While previous studies have emphasized the role of nonnative interactions for introducing ruggedness in the free energy surface for binding,^{72,98} the present study only considered native interactions.

The present results are consistent with previous native-based simulations that demonstrated the importance of the induced-folding mechanism for IDP interactions.^{72,75,77,98} Both the SIWF and WISF models sample relatively folded conformations in the unbound FCP1 ensemble. However, in both models, the transition from free to bound FCP1 occurs from the center of the FCP1 distribution. Moreover, the mechanism for FCP1 folding and RAP74 binding appear to be relatively insensitive to the difference in helical content of the SIWF and WISF ensembles.

The present results may also provide fundamental insight into the biological function and mechanism of WH domains for regulating cellular processes and, in particular transcription, via protein–protein interactions. Intriguingly, although early studies suggested that the C-terminal WH domain of RAP74 may not bind DNA,²³ a very recent study demonstrated that the yeast homologue binds nonspecific DNA with modest affinity.⁹⁹ Moreover, although early studies suggested that WH domains primarily function in binding DNA,⁴⁷ several studies have now demonstrated that WH and similar domains can also function in protein–protein interactions.^{25,26,49,50,100,101} In particular, kinetic studies have suggested that the disordered Activation Factor 1 (AF1) of the Androgen Receptor (AR) also binds the same hydrophobic cleft of the RAP74 WH domain.¹⁰¹ This study further speculated that the essential function of FCP1 may be regulated via a competition for the RAP74 hydrophobic cleft between FCP1, AR–AF1, and possibly DNA.¹⁰¹ Thus, it is intriguing to consider whether the unique properties conferred by intrinsic disorder, such as the increased cooperativity of binding, the minimization of backtracking, and the capacity for one-to-many binding, may feature prominently in regulating transcription via regulating the phosphatase activity of FCP1. Significantly, relatively little is known regarding the location of the C-terminal WH domain of RAP74 with respect to either RNA Pol II or DNA during the course of transcription.¹⁰² Nevertheless, it has been proposed that long-range and large-scale allostery may act as a key

regulator of transcription.¹⁰³ Consequently, it is particularly intriguing to speculate whether the large scale conformational change and commensurate release of conformational frustration that result from the coupled folding-binding of the FCP1 and AR-AF1 disordered proteins may play a significant role in enabling such large-scale allostery. Such speculation certainly requires considerable scrutiny and motivates further investigation.

In summary, the present work provides the first simulation study investigating the mechanism underlying the interaction between the folded, C-terminal WH domain of the RAP74 subunit of TFIIF and the disordered C-terminal peptide of FCP1. These simulations suggest that productive binding proceeds when the C-terminal region of this FCP1 peptide contacts the C-terminal regions of RAP74 H3. The FCP1 helix then zips up sequentially along the binding interface. These results motivate future studies with more detailed models to investigate the effects of, e.g., nonnative and electrostatic interactions upon this interaction. More generally, these studies provide additional insight into the role of protein flexibility for allowing a greater diversity of binding paths and for avoiding backtracking and kinetic trapping. Finally, these studies may provide insight into the biological function of intrinsic disorder for regulating transcription.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information Section provides a detailed contact map for the simulation, as well as additional analysis of the FCP1 folding mechanism and the unstable encounter complex. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sas76@psu.edu (S.A.S.); wnoid@chem.psu.edu (W.G.N.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

W.G.N. gratefully acknowledges the support of fellowships from the Penn State Institute for Cyberscience and the Sloan Research foundation, as well as startup funds from the Pennsylvania State University. S.A.S. acknowledges the support of an NSF CAREER award (MCB-0953918). W.G.N. also gratefully acknowledges the assistance of Dr. Jaegil Kim and Prof. John Straub in implementing the free energy calculations. Numerical calculations used support and resources from Research Computing and Cyberinfrastructure, a unit of Information Technology Services at Penn State. The table of contents figure was made with VMD. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

■ REFERENCES

- (1) Wright, P.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321–331.
- (2) Dunker, A.; Lawson, J.; Brown, C.; Williams, R.; Romero, P.; Oh, J.; Oldfield, C.; Campen, A.; Ratliff, C.; Hipps, K.; et al. *J. Mol. Graphics Modell.* **2001**, *19*, 26–59.
- (3) Fuxreiter, M.; Tompa, P.; Simon, I.; Uversky, V. N.; Hansen, J. C.; Asturias, F. J. *Nat. Chem. Biol.* **2008**, *4*, 728–737.
- (4) Xie, H.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. *J. Proteome Res.* **2007**, *6*, 1917–1932.
- (5) Tompa, P.; Csermely, P. *FASEB J.* **2004**, *18*, 1169–1175.
- (6) Dunker, A.; Brown, C.; Lawson, J.; Iakoucheva, L.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573–6582.
- (7) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527–533.
- (8) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, *37*, 215–246.
- (9) Spolar, R.; Record, M. *Science* **1994**, *263*, 777–784.
- (10) Ivanyi-Nagy, R.; Lavergne, J.-P.; Gabus, C.; Ficheux, D.; Darlix, J.-L. *Nucleic Acids Res.* **2008**, *36*, 712–725.
- (11) Meszaros, B.; Simon, I.; Dosztanyi, Z. *Phys. Biol.* **2011**, *8*, 035003.
- (12) Dyson, H. J.; Wright, P. *Curr. Opin. Struc. Biol.* **2002**, *12*, 54–60.
- (13) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struc. Biol.* **2009**, *19*, 31–38.
- (14) Shoemaker, B.; Portman, J.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868–8873.
- (15) Huang, Y.; Liu, Z. *J. Mol. Biol.* **2009**, *393*, 1143–1159.
- (16) Gunasekaran, K.; Tsai, C.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28*, 81–85.
- (17) Wang, Y.; Fisher, J. C.; Mathew, R.; Ou, L.; Otieno, S.; Sublet, J.; Xiao, L.; Chen, J.; Roussel, M. F.; Kriwacki, R. W. *Nat. Chem. Biol.* **2011**, *7*, 214–221.
- (18) Dyson, H. J.; Wright, P. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
- (19) Venters, B. J.; Pugh, B. F. *Crit. Rev. Biochem. Mol. Biol.* **2009**, *44*, 117–141.
- (20) Archambault, J.; Pan, G.; Dahmus, G.; Cartier, M.; Marshall, N.; Zhang, S.; Dahmus, M.; Greenblatt, J. *J. Biol. Chem.* **1998**, *273*, 27593–27601.
- (21) Kobor, M.; Archambault, J.; Lester, W.; Holstege, F.; Gileadi, O.; Jansma, D.; Jennings, E.; Kouyoumdjian, F.; Davidson, A.; Young, R.; et al. *Mol. Cell* **1999**, *4*, 55–62.
- (22) Cho, H.; Kim, T.; Mancebo, H.; Lane, W.; Flores, O.; Reinberg, D. *Genes Dev.* **1999**, *13*, 1540–1552.
- (23) Kamada, K.; De Angelis, J.; Roeder, R.; Burley, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3115–3120.
- (24) Chambers, R.; Wang, B.; Burton, Z. F.; Dahmus, M. *J. Biol. Chem.* **1995**, *270*, 14962–14969.
- (25) Nguyen, B.; Abbott, K.; Potempa, K.; Kobort, M.; Archambault, J.; Greenblatt, J.; Legault, P.; Omichinski, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5688–5693.
- (26) Kamada, K.; Roeder, R.; Burley, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2296–2299.
- (27) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.
- (28) Okazaki, K.-i.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11182–11187.
- (29) Tsai, C.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181–1190.
- (30) Bienkiewicz, E.; Adkins, J.; Lumb, K. *Biochemistry* **2002**, *41*, 752–759.
- (31) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* **2004**, *338*, 1015–1026.
- (32) Baker, J. M. R.; Hudson, R. P.; Kanelis, V.; Choy, W.-Y.; Thibodeau, P. H.; Thomas, P. J.; Forman-Kay, J. D. *Nat. Struct. Mol. Biol.* **2007**, *14*, 738–745.
- (33) Lacy, E.; Filippov, I.; Lewis, W.; Otieno, S.; Xiao, L.; Weiss, S.; Hengst, L.; Kriwacki, R. *Nat. Struct. Mol. Biol.* **2004**, *11*, 358–364.
- (34) Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 2088–2089.
- (35) Marsh, J. A.; Dancheck, B.; Ragusa, M. J.; Allaire, M.; Forman-Kay, J. D.; Petri, W. *Structure* **2010**, *18*, 1094–1103.
- (36) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539–546.
- (37) Espinoza-Fonseca, L. M. *Biochem. Biophys. Res. Commun.* **2009**, *382*, 479–482.
- (38) Wlodarski, T.; Zagrovic, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19346–19351.

- (39) Long, D.; Brüschweiler, R. *PLoS Comput. Biol.* **2011**, *7*, e1002035.
- (40) Onitsuka, M.; Kamikubo, H.; Yamazaki, Y.; Kataoka, M. *Proteins* **2008**, *72*, 837–847.
- (41) Zhou, H.-X. *Biophys. J.* **2010**, *98*, L15–L17.
- (42) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 13737–13741.
- (43) Licciardo, P.; Ruggiero, L.; Lania, L.; Majello, B. *Nucleic Acids Res.* **2001**, *29*, 3539–3545.
- (44) Uversky, V. *Protein Sci.* **2002**, *11*, 739–756.
- (45) Meszaros, B.; Tompa, P.; Simon, I.; Dosztanyi, Z. *J. Mol. Biol.* **2007**, *372*, 549–561.
- (46) Ganguly, D.; Otieno, S.; Waddell, B.; Iconaru, L.; Kriwacki, R. W.; Chen, J. *J. Mol. Biol.* **2012**, *674*–684.
- (47) Gajiwala, K.; Burley, S. *Curr. Opin. Struc. Biol.* **2000**, *10*, 110–116.
- (48) Groft, C.; Uljon, S.; Wang, R.; Werner, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9117–9122.
- (49) Zheng, N.; Fraenkel, E.; Pabo, C.; Pavletich, N. *Genes Dev.* **1999**, *13*, 666–674.
- (50) Zheng, N.; Schulman, B.; Song, L.; Miller, J.; Jeffrey, P.; Wang, P.; Chu, C.; Koeppl, D.; Elledge, S.; Pagano, M.; et al. *Nature* **2002**, *416*, 703–709.
- (51) Bermel, W.; Bertini, I.; Felli, I. C.; Piccioli, M.; Pierattelli, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 25–45.
- (52) Showalter, S. A. *Biomol. NMR Assignments* **2009**, *3*, 179–181.
- (53) O'Hare, B.; Benesi, A. J.; Showalter, S. A. *J. Magn. Reson.* **2009**, *200*, 354–358.
- (54) Lawrence, C. W.; Bonny, A.; Showalter, S. A. *Biochem. Biophys. Res. Commun.* **2011**, *410*, 461–465.
- (55) Wostenberg, C.; Kumar, S.; Noid, W. G.; Showalter, S. A. *J. Phys. Chem. B* **2011**, *115*, 13731–13739.
- (56) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2–8.
- (57) Lawrence, C. W.; Showalter, S. A. *J. Phys. Chem. Lett.* **2012**, *3*, 1409–1413.
- (58) Chen, H.-F. *PLoS One* **2009**, *4*, e6516.
- (59) Allen, W. J.; Capelluto, D. G. S.; Finkielstein, C. V.; Bevan, D. R. *J. Phys. Chem. B* **2010**, *114*, 13201–13213.
- (60) Higo, J.; Nishimura, Y.; Nakamura, H. *J. Am. Chem. Soc.* **2011**, *133*, 10448–10458.
- (61) Knott, M.; Best, R. B. *PLoS Comput. Biol.* **2012**, *8*, e1002605.
- (62) Ganguly, D.; Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 5214–5223.
- (63) Zhang, W.; Ganguly, D.; Chen, J. *PLoS Comput. Biol.* **2012**, *8*, e1002353.
- (64) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787–3791.
- (65) Nymeyer, H.; Garcia, A.; Onuchic, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921–5928.
- (66) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171–180.
- (67) Clementi, C.; Nymeyer, H.; Onuchic, J. *J. Mol. Biol.* **2000**, *298*, 937–953.
- (68) Wolynes, P. G. *Q. Rev. Biophys.* **2005**, *38*, 405–410.
- (69) Levy, Y.; Wolynes, P.; Onuchic, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 511–516.
- (70) Levy, Y.; Cho, S.; Onuchic, J.; Wolynes, P. *J. Mol. Biol.* **2005**, *346*, 1121–1145.
- (71) Wang, J.; Lu, Q.; Lu, H. P. *PLoS Comput. Biol.* **2006**, *2*, 842–852.
- (72) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. *PLoS Comput. Biol.* **2008**, *4*, e1000060.
- (73) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–1025.
- (74) Huang, Y.; Liu, Z. *Proteins* **2010**, *78*, 3251–3259.
- (75) Wang, J.; Wang, Y.; Chu, X.; Hagen, S. J.; Han, W.; Wang, E. *PLoS Comput. Biol.* **2011**, *7*, e1001118.
- (76) Ganguly, D.; Chen, J. *Proteins* **2011**, *79*, 1251–1266.
- (77) Ganguly, D.; Zhang, W.; Chen, J. *Mol. BioSyst.* **2012**, *8*, 198–209.
- (78) De Sancho, D.; Best, R. B. *Mol. BioSyst.* **2012**, *8*, 256–267.
- (79) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Nucleic Acids Res.* **2010**, *W657–W661*.
- (80) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. *J. Phys. Chem. B* **2012**, *116*, 8692–8702.
- (81) Plotkin, S. S.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6509–6514.
- (82) Chavez, L. L.; Onuchic, J. N.; Clementi, C. *J. Am. Chem. Soc.* **2004**, *126*, 8426–8432.
- (83) Yang, A.; Abbott, K. L.; Desjardins, A.; Di Lello, P.; Omichinski, J. G.; Legault, P. *Biochemistry* **2009**, *48*, 1964–1974.
- (84) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins* **2009**, *75*, 430–441.
- (85) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (86) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002.
- (87) Xu, X.-P.; Case, D. *J. Biomol. NMR* **2001**, *21*, 321–333.
- (88) Xu, X.-P.; Case, D. A. *Biopolymer* **2002**, *65*, 408–423.
- (89) Li, D.-W.; Brüschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246–248.
- (90) *The PyMOL Molecular Graphics System*, version 1.2r3pre; Schrödinger, LLC (<http://www.pymol.org/>).
- (91) MATLAB, version 7.10.0; The MathWorks Inc.: Natick, MA, 2010.
- (92) Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (93) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40–57.
- (94) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- (95) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171–180.
- (96) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. *Biochemistry* **2012**, *51*, 2224–2231.
- (97) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, *125*, 054108.
- (98) Chu, X.; Wang, Y.; Gan, L.; Bai, Y.; Han, W.; Wang, E.; Wang, J. *PLoS Comput. Biol.* **2012**, *8*, e1002608.
- (99) Kilpatrick, A. M.; Koharudin, L. M. I.; Calero, G. A.; Gronenborn, A. M. *Proteins* **2011**, *80*, 519–29.
- (100) Finnin, M. S.; Cicero, M. P.; Davies, C.; Porter, S. J.; White, S. W.; Kreuzer, K. N. *EMBO J.* **1997**, *16*, 1992–2003.
- (101) Lavery, D. N.; McEwan, I. J. *Biochemistry* **2008**, *47*, 3352–9.
- (102) Eichner, J.; Chen, H.-T.; Warfield, L.; Hahn, S. *EMBO J.* **2010**, *29*, 706–16.
- (103) Tsai, C.-J.; Nussinov, R. *Biochem. J.* **2011**, *439*, 15–25.