THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Conformational Biases of Linear Motifs

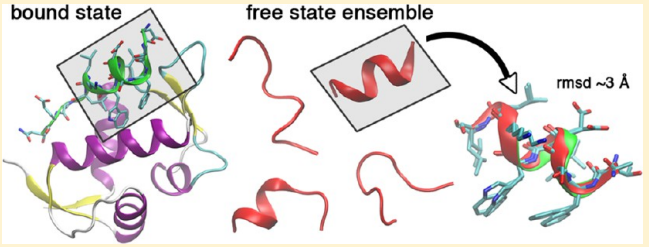
Elio A. Cino,[†] Wing-Yiu Choy,[‡,*] and Mikko Karttunen[†,*]

[†]Department of Chemistry and Waterloo Institute for Nanotechnology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

[‡]Department of Biochemistry, The University of Western Ontario, London, Ontario, Canada N6A 5C1

Supporting Information

ABSTRACT: Linear motifs (LMs) are protein−protein interaction sites, typically consisting of ∼4−20 amino acid residues that are often found in disordered proteins or regions, and function largely independent from other parts of the proteins they are found in. These short sequence patterns are involved in a wide spectrum of biological functions including cell cycle control, transcriptional regulation, enzymatic catalysis, cell signaling, protein trafficking, etc. Even though LMs may adopt defined structures in complexes with targets, which can be determined by conventional methods, their uncomplexed states can be highly dynamic and difficult to characterize. This hinders our understanding of the structure−function relationship of LMs. Here, the uncomplexed states of 6 different LMs are investigated using atomistic molecular dynamics (MD) simulations. The total simulation time was about 63 μs. The results show that LMs can have distinct conformational propensities, which often resemble their complexed state. As a result, the free state structure and dynamics of LMs may hold important clues regarding binding mechanisms, affinities and specificities. The findings should be helpful in advancing our understanding of the mechanisms whereby disordered amino acid sequences bind targets, modeling disordered proteins/regions, and computational prediction of binding affinities.



INTRODUCTION

The paradigm—still textbook knowledge—that all proteins have a well-defined globular structure has undergone a major change over the past 15 years.[1,2] It is now accepted that up to 45%[3] of proteins in eukaryotic cells contain significant intrinsically disordered regions (IDRs) or are intrinsically disordered proteins (IDPs), i.e., they do not have a well-defined three-dimensional structure. From a biological point of view, IDRs and IDPs are typically associated with regulatory and signaling functions, making them key players in health and sickness and, thus, important drug targets. From a physical point of view, IDRs and IDPs have very different entropic and conformational properties compared to globular proteins. Their conformational freedom also allows them to sample interactions—binding partners—over a larger area[4] and models such as "fly casting" and other have been suggested.[5,6]

IDPs and IDRs typically contain (short) linear motifs (LMs)—∼4−20 amino acid functional sites.[7−11] The idea of short linear motifs is credited to Nobel Laureate Tim Hunt, who recognized their importance and independence from the three-dimensional protein structure.[12] After the abundance and importance of IDPs and IDRs was discovered, LMs become a topic of intense research. Because of their short lengths, LMs do not typically adopt stable folds in the absence of binding partners. Interactions with targets can stabilize the conformation of LMs and it is often possible to determine complexed structures of LMs using conventional structure determination methods, such as X-ray crystallography and nuclear magnetic resonance spectroscopy. On the other hand, the structures and conformational tendencies of LMs in the absence of binding partners are less characterized. Such information is crucial for understanding the molecular basis of interactions involving LMs and developing targeted therapeutics.[13,14]

Perhaps the most common way to study LMs has been through bioinformatics analysis of protein sequences.[15,16] In recent years, however, a few Molecular Dynamics (MD) simulations have scrutinized the free (uncomplexed) states of LMs.[17−24] MD simulations have the ability to sample the conformational distribution, or ensemble, in a unique manner. Simulation length poses limitations, but current simulations covering microseconds and longer allow for sampling of the distribution. As an example, Lambrughi et al.[22] employed atomistic MD simulations in explicit solvent to investigate the free state structural ensemble of the disordered Sic1 kinase-inhibitory domain (KID). Their study showed that while the Sic1 KID conformational ensemble favors extended states, it is also populated with compact conformations with transient helical secondary structures. The simulations provided insights into the mechanisms stabilizing the LMs along its sequence and helped to identify potential target binding regions. Importantly, such simulation data would also allow for correlation analysis and, ideally, identification of possible collective effects.

Here, a total of 63 μs MD simulations were used to examine the uncomplexed state structures and dynamics of 6 different LMs, most of which have not been previously characterized. These LMs are present in proteins involved in different

Received: July 29, 2013
Revised: November 30, 2013
Published: December 3, 2013

Table 1. Summary of the MD Simulations[a]

LM and sequence	force-field(s)	length (μs)
exoenzyme S	CHARMM22*	2.0
Ac-GLLDALDLASK-NH_2	GROMOS96 53a6	2.0
	GROMOS96 54a7	2.0
	Amber99SB*-ILDN	2.0
amphiphysin 1	CHARMM22*	1.0
Ac-MSQTLPWDLWTTST-NH_2	GROMOS96 53a6	1.0
	GROMOS96 54a7	1.0
β-arrestin 2	CHARMM22*	1.0
Ac-DTNLIEFE-NH_2	GROMOS96 53a6	1.0
	GROMOS96 54a7	1.0
p21	CHARMM22*	3.0
Ac-QGRKRRQTSMTDFYHSKRRLIFSKRKP-Cterm		
p66	CHARMM22*	3.0
Ac-LGKANRQVSITGFFQRK-Cterm		
Fen1	CHARMM22*	3.0
Ac-SRQGSTQGRLDDFFKVTGSLSSAKRK-NH_2		
optimized LxxLL motif peptide[c]	CHARMM22*	2.0 × 2[b]
Ac-HPLLMRLLLSP-NH_2		
RIP140 L6	CHARMM22*	2.0 × 2[b]
Ac-VTLLQLLLGHK-NH_2		
RIP140 L7	CHARMM22*	2.0 × 2[b]
Ac-RTVLQLLLGNP-NH_2		
RIP140 L3	CHARMM22*	2.0 × 2[b]
Ac-SSHLKTLLKKS-NH_2		
Tcf4	CHARMM22*	2.0 × 2[b]
Ac-DLADVKSSLVNES-NH_2		
p53	CHARMM22*	5.0 × 2[b]
Ac-PLSQETFSDLWKLLPENNVL-NH_2		
p53 P27S mutant	CHARMM22*	5.0 × 2[b]
Ac-PLSQETFSDLWKLLSENNVL-NH_2		

[a]Region of the LMs used for rmsd and other calculations are underlined. Bolded residues within the underlined region have bound state structures with >10% burial of their solvent-accessible surface areas, relative to extended (A−X−A) conformations[67] and were used for the solvent-accessible area calculations. [b]Duplicate simulation with different random starting structure. [c]Ribosome display optimized LxxLL motif peptide sequence obtained from.[86]

biological processes, including enzymatic catalysis, protein transport, DNA replication and repair, cell signaling, transcriptional regulation, and apoptosis. The results together suggest that it may be common for LMs to have structural propensities toward bound state conformations in their uncomplexed states, albeit to different extents. Therefore, the free state structure and dynamics of LMs extracted from MD simulations could hold important clues regarding binding mechanisms, affinities and specificities.

METHODS

Six different LMs from the eukaryotic linear motif (ELM) database[8] with available high-resolution Protein Data Bank[25] structures in complex with binding partners were selected and MD simulations were performed on the LMs in the absence of their binding partners. The starting structures were *not* the bound state structure coordinates. Instead, starting structures were generated through a simulated annealing procedure using the crystallography and NMR system (CNS),[26] based solely upon the amino acid sequences of the LMs. Unstructured (non bound state-like) conformations from the annealing simulations were chosen as starting structures for the MD simulations. Where appropriate, acetyl and NH_2 groups were added to the N- and C-termini, respectively (Table 1).

The MD simulations were performed using GROMACS 4 software[27,28] and the CHARMM22*[29−31] force-field in explicit solvent under periodic boundary conditions. We also assessed CHARMM22* by comparing it with other force-fields: Benchmark simulations were run using the GROMOS96 53a6,[32,33] 54a7,[34] and Amber99SB*-ILDN[35−38] force-fields. While previous versions of the CHARMM force field have well-documented helical bias,[18,39−41] several of the backbone and side chain dihedral potentials have been recently modified to produce CHARMM22*,[31] which improves the helix−coil balance. These improvements have been validated in long time-scale protein folding simulations.[42,43] The same starting structure was used for all force-fields. TIP3P[44] water model was used for the CHARMM22* and Amber99SB*-ILDN simulations. SPC[45] (simple point charge) was used in the GROMOS96 simulations. Besides providing a comparison between the different force fields, the use of different water models (SPC and TIP3P) provides additional information about the robustness with respect to electrostatic properties of water; the dielectric constant of SPC water is about 60 whereas TIP3P has a value of about 94. The experimental value is about 78. A recent review of the properties of common water models is provided by Vega and Abascal.[46]

The simulations were set up as follows: Protonation states of all ionizable residues were assigned to their most probable state

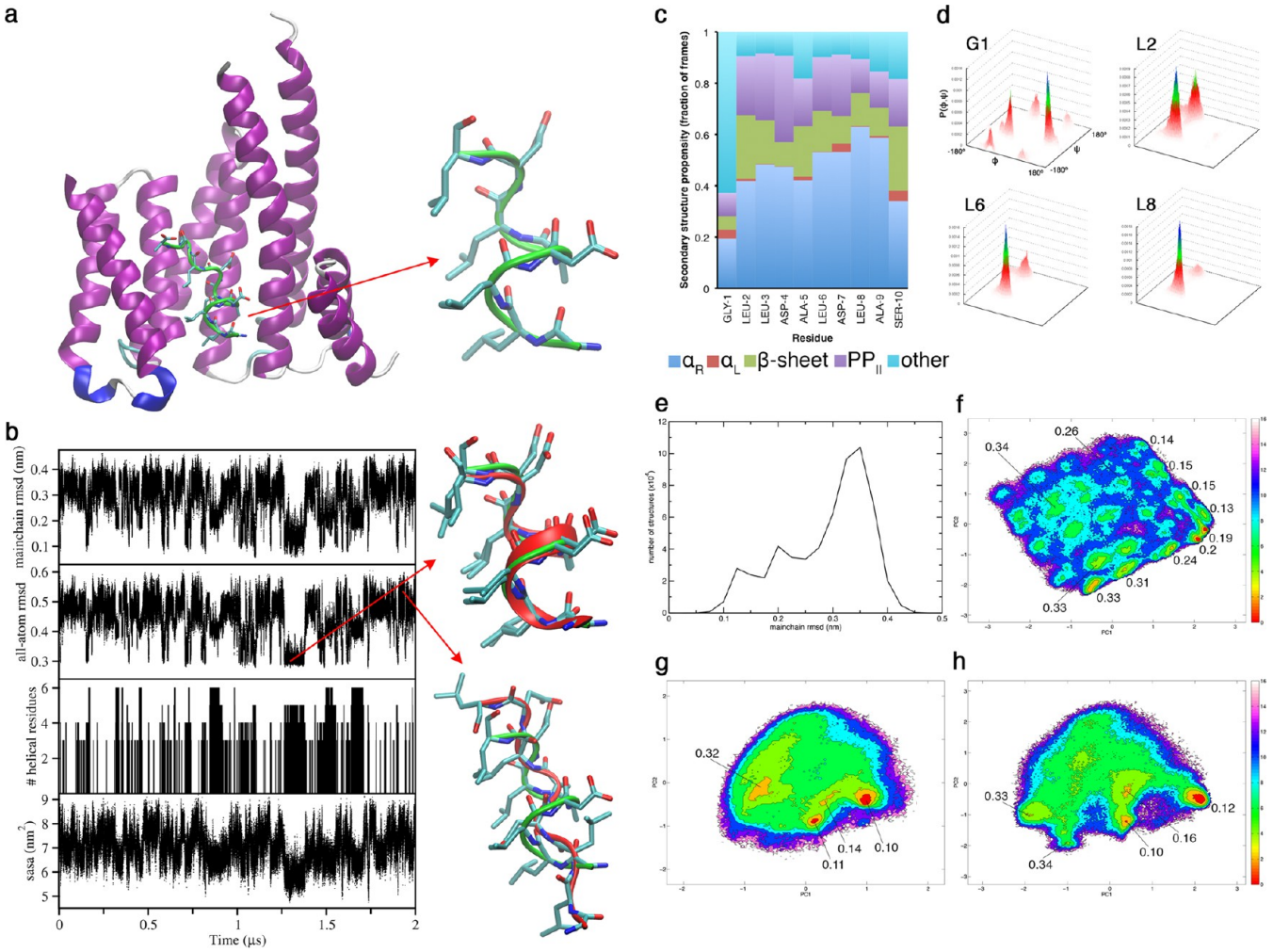


Figure 1. ExoS−14-3-3 complexed to uncomplexed state comparison. (a) X-ray structure of the 14-3-3 ExoS peptide complex (PDB id: 2C23)[65] identifying the region of the ExoS peptide (GLLDALD) used for analysis. (b) Main chain and all-atom rmsds to the complexed state and peptide helicity measurement. For the rmsd calculations, atoms of the seven residues at the N-terminus (GLLDALD), including 28 main chain (N, C_α, C′, and CO) and 49 non-hydrogen atoms were compared. Peptide helicity was the sum of residues (GLLDALD) in all helix types (α, 3_{10}, and π) as assigned by the DSSP algorithm. Solvent-accessible surface areas were calculated using the g_sas tool.[28,103,104] (c) Secondary structures were classified using standard ϕ, ψ dihedral angle spaces. α_R (right handed α helix): $-175° \le \phi \le -55°$, $-80° \le \psi \le -5°$; α_L (left-handed α helix): $5° \le \phi \le 75°$, $25° \le \psi \le 120°$. β-sheet: $-180° \le \phi \le -50°$, $80° \le \psi \le -170°$. PP_{II} (polyproline II helix): $-110° \le \phi \le -50°$, $120° \le \psi \le 180°$. Because of overlapping dihedral angle spaces, PP_{II} conformations were subtracted from the β-sheet region. (d) ϕ, ψ distributions for select residues. (e) Distribution of the main chain rmsds. (f−h) Gibbs free energy landscapes (kJ/mol) based on the first two principal components from (f) dPCA and (g) cPCA with least-squares fit to the main chain atoms of the bound state structure and (h) cPCA with least-squares fit to all atoms of the bound state structure. Values mapped onto the landscapes indicate the average main chain rmsds (to the bound state structure) of the structures contained in the basins. PCA was accomplished using the GROMACS tools g_covar and g_anaeig.[28]

at pH 7. Histidine residues were assigned a neutral form protonated on NE2 only. Each starting structure was centered in a rhombic dodecahedron of sufficient size so that the minimum distance to period images was at least 1.0 nm and solvated in (∼1400 to ∼5000) water molecules. The systems were neutralized and brought to 0.15 M salt with sodium (Na^+) and chloride (Cl^-) ions. Ion parameters specific to each force-field distribution were used; for a comparison of different water models and ion force-fields, see ref 47. Protein and nonprotein atoms were coupled to their own temperature baths, which were kept constant at 310 K using the Parrinello−Donadio−Bussi v-rescale algorithm.[48] This thermostat has been shown to be reliable in biomolecular simulations.[49,50] Pressure was maintained isotropically at 1 bar using the Parrinello−Rahman barostat.[51] A time step of 2 fs was used. Each system underwent energy minimization using the steepest descents algorithm. Initial atom velocities were taken from a Maxwellian distribution at 310 K. Bond lengths were constrained using LINCS.[52] A 1.0 nm cutoff was used for Lennard-Jones interactions. Dispersion corrections for energy and pressure were applied. Electrostatics were calculated by the particle-mesh Ewald (PME) method[53] with 0.12 nm grid-spacing and a 1.0 nm real-space cutoff. Charge groups were not used (single atom charge groups) to avoid potential artifacts.[49,50] For a comprehensive review of the importance and treatment of electrostatic interactions in biomolecular simulations, see ref 54. Each trajectory was run for 1−5 μs, for an aggregate time of 63 μs. Table 1 provides a summary of the simulations.

RESULTS

MD simulations were used to examine several LMs in their uncomplexed states. For each MD trajectory, the resemblance

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

of the LM to a complexed state structure was assessed directly using rmsd calculations. LM conformations were investigated by secondary structure analysis. Solvent-accessible surface areas were examined to assess the contribution of surface area burial to LM structure formation and stability. The free state ensembles for several of the LM systems were further explored using backbone (ϕ, ψ) dihedral angle principal component analysis (dPCA),[55−58] Cartesian principal component analysis (cPCA), and other measurements. This work focuses primarily on helical LMs, but also includes one that binds in an extended conformation. β-type LMs have been studied by MD simulations in several other works[17,18,59,60] and they are not discussed here.

1. Exoenzyme S−14-3-3. The 14-3-3 family of proteins are one of the most studied hubs for IDP interactions. They are primarily helical proteins capable of binding to hundreds of ligands, together which may comprise 0.6% of the human proteome.[61,62] It is estimated that over 90% of the 14-3-3 binding partners are completely or partially disordered.[61] Binding to 14-3-3 proteins is typically governed by phosphoserine and phosphothreonine containing motifs on the partners. Some partners bind, however, in a phosphorylation independent manner.[63−65] One such phosphorylation independent interactor is exoenzyme S (ExoS)—a toxin from the bacterium *Pseudomonas aeruginosa*.[64] Interaction of ExoS with 14-3-3 activates the ADP-ribosylating activity of the toxin, which can disrupt proper function of eukaryotic cells.[64]

The structure of the ExoS peptide−14-3-3 complex (PDB id: 2C23)[65] (Figure 1a) illustrates that the 11-mer ExoS peptide forms a short amphipathic helix in its N-terminal GLLDALDL portion, with the hydrophobic face pointing toward the 14-3-3 binding groove. MD simulations show that in its uncomplexed state, ExoS had clear propensity for adopting bound state-like conformations in the GLLDALDL region (Figure 1). Several transitions between high and low rmsds to the bound state occurred throughout the trajectory. In ∼8% of the frames the main chain rmsd was <0.15 nm, suggesting that although the uncomplexed state equilibrium favors the unfolded conformation, a population of bound state-like structures exists. This is similar to the other examples that will be presented here and what has been reported for other disordered peptides.[55,66]

Correlations between several combinations of variables were used to investigate the factors involved in formation of bound state-like structures (Table 2). Main chain rmsd was moderately correlated with ExoS peptide helicity (main chain rmsd vs the number of helical residues, r^2 = 0.37) (Figure 1b and Table 2). Dihedral angle analysis indicates that the ExoS peptide has moderate-high helical propensity throughout, peaking around the C-terminal half, L6-L8 region (Figure 1c,d). Analysis of the ExoS peptide−14-3-3 complex (PDB id: 2C23)[65] in the absence of 14-3-3 indicated six residues, primarily leucines, within the analyzed region (GLLDALDL) have >10% decreases in their solvent-accessible surface areas, relative to extended (A−X−A) conformations.[67] Decreases in solvent exposed surfaces of these residues were moderately correlated with main chain rmsd (r^2 = 0.33) (Figure 1b and Table 2). The effect of hydrophobic residues, in particular leucine, on helix formation observed here is similar to the results obtained from a simpler model by Dias et al.[68] The projection of the first dPCA principal component onto the trajectory was well correlated with main chain rmsd (r^2 = 0.64) and radius of gyration (r^2 = 0.74), a measure of compactness (Table 2). In comparison, lower correlations to main chain rmsd were obtained from cPCA (Table 2). Peptide compaction, however, was well correlated with PC2 from cPCA. Overall, the PCA illustrates that the ExoS peptide has a diverse structural ensemble, especially in the unfolded state. Despite this, the peptide does have some clear structural preferences, which are largely biased toward conformations with bound state resemblance. The finding is consistent with what has been observed for other disordered peptides.[55,66] The differences between the dPCA and cPCA may arise due to the mixing of internal and overall motions in the cPCA.[57] These effects can also account for the smoother appearance of the cPCA free energy landscapes (Figure 1g), compared to the rugged dPCA landscape (Figure 1f). We will return to this issue in the Discussion. Moreover, using different fits to the bound state structure (main chain, Figure 1g and all-atom, Figure 1h) did not make a large difference in landscape appearance or correlations to other variables (Table 2).

Table 2. ExoS Correlations

variable 1	variable 2	r^2
main chain rmsd	number of helical residues	0.37
main chain rmsd	solvent-accessible surface area (SASA)	0.33
main chain rmsd	radius of gyration	0.44
main chain rmsd	dPCA PC1	0.64
main chain rmsd	dPCA PC2	<0.1
radius of gyration	dPCA PC1	0.74
radius of gyration	dPCA PC2	<0.1
main chain rmsd	cPCA PC1 (main chain fit to bound state structure)	0.42
main chain rmsd	cPCA PC2 (main chain fit to bound state structure)	0.26
radius of gyration	cPCA PC1 (main chain fit to bound state structure)	<0.1
radius of gyration	cPCA PC2 (main chain fit to bound state structure)	0.84
main chain rmsd	cPCA PC1 (all-atom fit to bound state structure)	0.49
main chain rmsd	cPCA PC2 (all-atom fit to bound state structure)	0.17
radius of gyration	cPCA PC1 (all-atom fit to bound state structure)	0.15
radius of gyration	cPCA PC2 (all-atom fit to bound state structure)	0.73

Similar results were obtained using the GROMOS96 54a7 force-field (main chain rmsd <0.15 nm in 7.8% of the frames) (Supporting Information, Figure S1b). Bound state-like structure formation, however, occurred less frequently in the Amber99SB*-ILDN simulation and, to an even lesser extent, in the GROMOS96 53a6 simulation (main chain rmsd <0.15 nm in 2.2% and 0.9% of the frames, respectively) (Figure S1, parts c and a). The GROMOS96 53a6 result was an outlier here, and it is possible that this force-field may not be optimal for helical proteins/peptides. Recent modifications, such as GROMOS96 54a7, have improved the helical propensities of 53a6 to better match experimental data[34,69] and our results should contribute to the validation of these modifications.

2. Amphiphysin 1 and β-Arrestin 2−Clathrin. Clathrin is a protein that forms coated vesicles for transporting cargos via endo- and exocytosis and within cells.[70] The N-terminal β-propeller domain (TD) of the clathrin heavy chain interacts with numerous accessory proteins including the amphiphysin (a protein essential in endocytosis) and arrestin proteins (GPCR regulators).[71,72] There are interaction sites located at the top

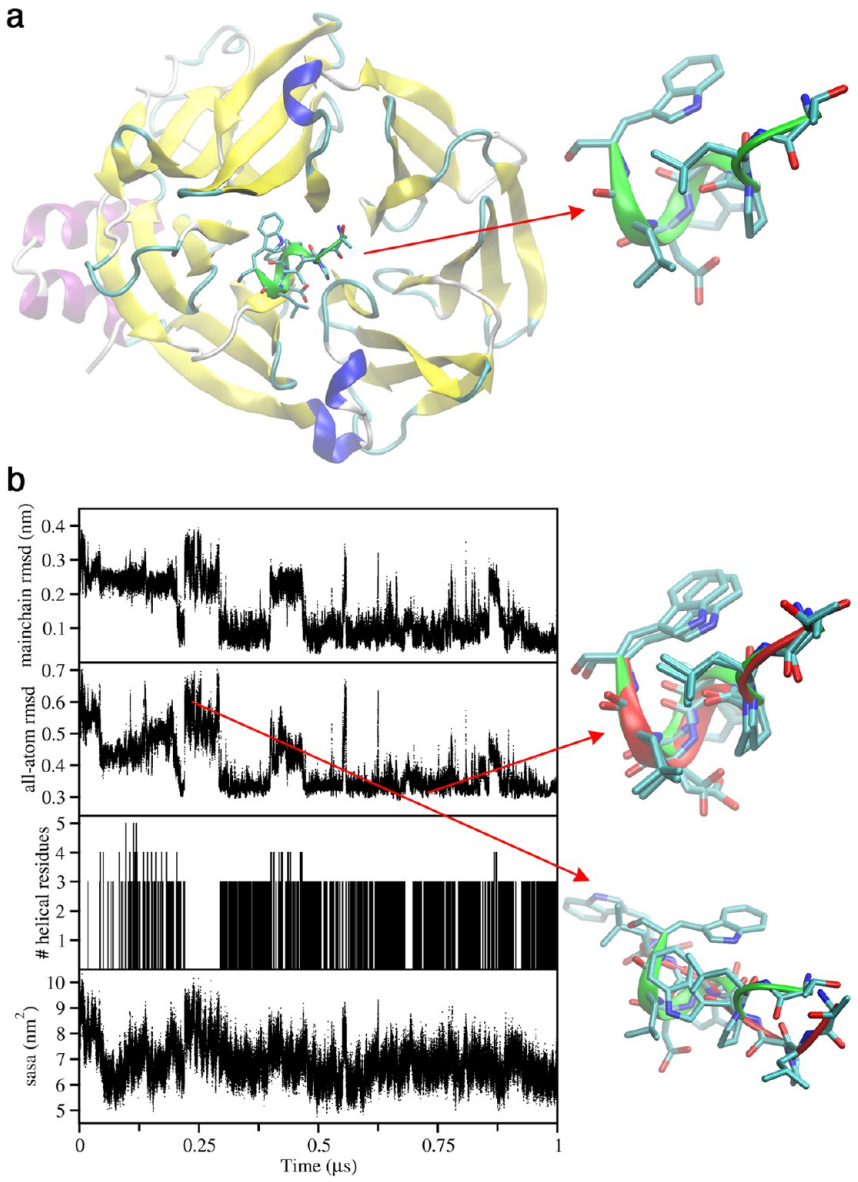


Figure 2. Amphiphysin 1−clathrin complexed to uncomplexed state comparison. a) X-ray structure of the amphiphysin 1 peptide−clathrin complex (PDB id: 1UTC)[72] identifying the region of the amphiphysin 1 peptide used for rmsd calculations. (b) Main chain and all-atom rmsds to the complexed state and peptide helicity measurement. For the rmsd calculations, seven residues (TLPWDLW), including 28 main chain (N, C_α, C′, and CO) and 66 non-hydrogen atoms were compared. Peptide helicity was the sum of residues in a 3_{10} helix as assigned by the DSSP algorithm. Solvent-accessible surface areas were calculated using the g_sas tool.[28,103,104]

and side surfaces of the clathrin TD, which bind to distinct motifs. First, the "W box" motif of amphiphysin 1 (Table 1), which binds the top face of the TD propeller (Figure 2a), is examined, followed by the "clathrin box" region of β-arrestin 2 (Table 1), which binds to a groove on the side of the TD.[72−74] Both motifs can be found in naturally disordered regions of their respective proteins.[72,73]

In the amphiphysin 1 W-box peptide−clathrin complex structure (PDB id: 1UTC),[72] the W-box peptide adopts a compact organization, containing a single 3_{10} helical turn, which initiates at a proline (Figure 2a). In the absence of clathrin, our MD simulations show that the amphiphysin 1 W-box peptide had obvious tendency for folding into bound state-like structures in the TLPWDLW region (Figure 2b). As the trajectory progressed, revisiting of higher rmsd conformations became less frequent, with structures resembling the bound state conformation being preferred. In ∼61% of the frames the main chain rmsd was <0.15 nm, suggesting that the uncomplexed state equilibrium favors the compact bound state-like conformation rather than extended structures. Analysis of the amphiphysin 1 W-box peptide−clathrin complex structure (PDB id: 1UTC)[72] in the absence of clathrin shows that six consecutive and primarily hydrophobic residues within the analyzed region (T**LPWDLW**) have a >10% decrease in their solvent-accessible surface area, relative to being in an extended (A−X−A) conformation.[67] The solvent-accessible surface area of these residues was correlated with main chain rmsd to some extent (r^2 = 0.24). Comparable results were obtained using the GROMOS96 53a6 and 54a7 force-fields (Figure S2, parts a and b), although the bound state-like conformation was more stable with 54a7 compared to 53a6. Like the CHARMM22* simulation, the GROMOS96 54a7

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

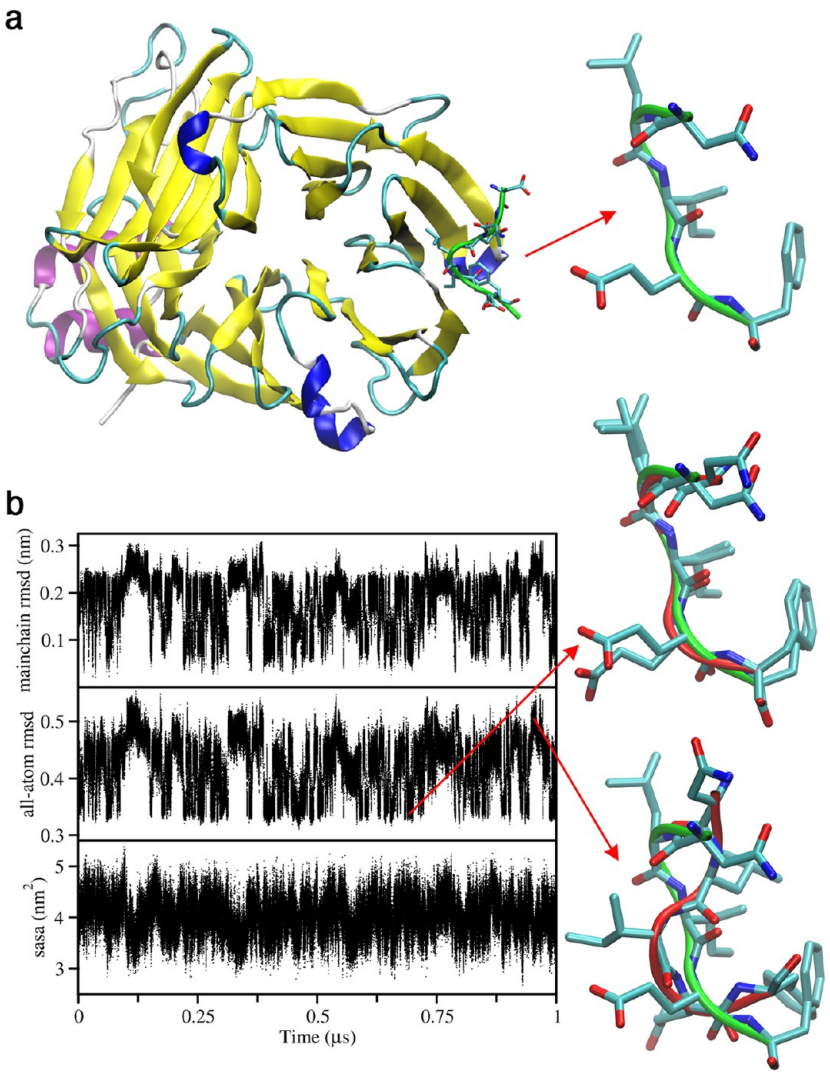


Figure 3. β-arrestin 2−clathrin complexed to uncomplexed state comparison. (a) X-ray structure of the β-arrestin 2 peptide−clathrin complex (PDB id: C9L)[74] identifying the region of the β-arrestin 2 peptide used for rmsd calculations. (b) Main chain and all-atom rmsds to the complexed state. For the rmsd calculations, 5 residues (NLIEF), including 20 main chain (N, C_α, C′, and CO) and 44 non-hydrogen atoms were compared. Solvent-accessible surface areas were calculated using the g_sas tool.[28,103,104]

trajectory eventually yielded a stable low rmsd conformation and ∼54% of the total frames contained structures with main chain rmsds <0.15 nm (Figure S2b). Although the CHARMM22* and GROMOS96 54a7 trajectories suggest that the W-box peptide favors bound state-like conformations, longer simulations and experimental studies may help to evaluate the apparent high percentage of preformed structures in solution.

In the β-arrestin 2 "clathrin box" peptide−clathrin complex structure (PDB id: 1C9L),[74] the clathrin box peptide adopts an extended conformation (Figure 3a). MD simulations of the clathrin box peptide (Table 1) in its uncomplexed state revealed frequent, yet transient conversions to bound state-like structures in the NLIEF region (Figure 3b). In ∼29% of the frames the main chain rmsd was <0.15 nm, implying that the uncomplexed state equilibrium contains a considerable population of bound state-like conformations. Analysis of the β-arrestin 2−clathrin bound state structure (PDB id: 1C9L),[74] in the absence of clathrin indicated that three residues within the analyzed region (**NLIEF**) had a >10% decreases in solvent-accessible surface area, relative to being in an extended (A−X−A) conformation.[67] This parameter was only weakly correlated with main chain rmsd (r^2 < 0.1). The results from the GROMOS96 54a6 and 54a7 force-field simulations (Figure S3, parts a and b) were consistent with the CHARMM22* result. Bound state-like structures occurred more frequently in the 54a7 trajectory (main chain rmsd < 0.15 nm in 51 vs 26% of the frames), however.

3. PIP-Box (p21, p66, Fen1)−PCNA. The proliferating cell nuclear antigen (PCNA) is the "sliding clamp" that surrounds DNA and acts as a processivity factor for DNA polymerase δ.[75] PCNA interacting proteins often have roles in DNA replication and repair.[76] The "PIP box" is a conserved PCNA binding motif that is often found in disordered regions of PCNA interacting proteins.[76−78] These motifs associate on the outer surface of PCNA, and form a short amphipathic helix with hydrophobic side chains facing inward toward PCNA[79,80] (Figure 4a).

MD simulations of PIP-box containing regions of p21, p66, and Fen1 in the absence of PCNA revealed that the peptides explored bound state-like conformations, but to different extents (Figures 4b−d). In the p21 trajectory, 8.4% of the frames had main chain rmsds of <0.15 nm to the bound state

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

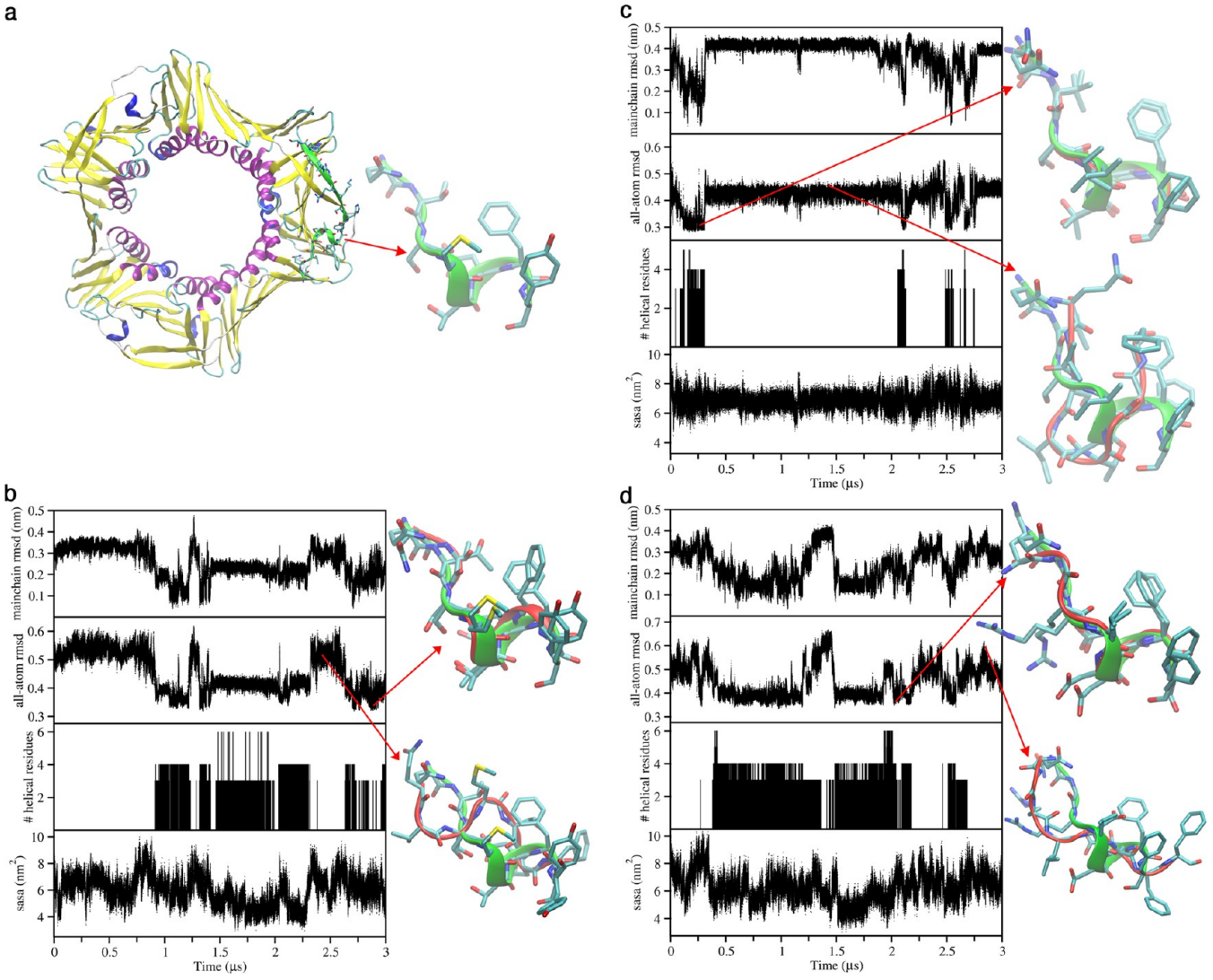


Figure 4. PIP-box–PCNA complexed to uncomplexed state comparison. (a) X-ray structure of the p21 PIP-box peptide–PCNA complex (PDB id: 1AXC)[80] identifying the region of the p21 peptide used for rmsd calculations. (b) p21, (c) p66, and (d) Fen1 main chain and all-atom rmsds to the complexed states. For the rmsd calculations, the PIP-box regions of the p21 (QTSMTDFY), p66 (QVSITGFF) and Fen1 (QGRLDDFF) peptides were compared to their corresponding PCNA bound state structures, PDB ids: 1AXC, 1U76, and 1U7B.[79,80] Eight residues, including 32 main chain (N, C_α, C′, and CO) and (68 p21, 63 p66, and 70 Fen1) non-hydrogen atoms were compared. Peptide helicity was the sum of residues in all helix types (α, 3_{10}, and π) as assigned by the DSSP algorithm. Solvent-accessible surface areas were calculated using the g_sas tool.[28,103,104]

structure compared to 6.5 and 20.1% of the p66 and Fen1 trajectory frames. The main chain rmsds were moderately correlated with peptide helical content (p21, r^2 = 0.25; p66, r^2 = 0.46; Fen1, r^2 = 0.19). Analysis of the p21, p66 and Fen1–clathrin complex structures (PDB ids: 1AXC, 1U76, and 1U7B)[79,80] in the absence of clathrin indicated that seven consecutive residues in corresponding positions within the PIP-boxes (p21, **QTSMTDFY**; p66, **QVSITGFF**; and Fen1, **QGRLDDFF**) have >10% decreases in their solvent-accessible surface areas, relative to extended (A–X–A) conformations.[67] These residues have a total combined accessible surface areas of ~7.4 nm^2 (p21), ~7.1 nm^2 (p66) and ~8.1 nm^2 (Fen1) in folded states. Although conformations with exposed surface areas around these values occurred simultaneously with bound state-like main chain rmsds, these two parameters were not highly correlated (r^2 <0.2) in all cases. Interestingly, in the p21 and Fen1 simulations, conformations with bound state-like rmsds and accessible surface areas considerably lower than their bound state conformations were present. These low surface areas arose from self-interaction of the C-terminal ends of the relatively long peptides with themselves. The shorter p66 peptide did not show similar behavior. The effects of self-interaction diminish the rmsd vs accessible area correlation in these cases. For instance, if only the adjacent residues to analyzed region are included in the solvent-accessible area calculation, the Fen1 main chain rmsd vs sasa correlation increases to r^2 =0.4. In addition to this work, an earlier NMR and computational study of p21 had determined that bound state like conformations populate a portion of the uncomplexed structural ensemble.[81] Considering the results here, it is likely that preformed structures are common features of PIP-box motifs in general.

4. LxxLL Motif Proteins. Proteins containing the LxxLL protein–protein interaction motif are abundant in cells and are often involved in transcriptional regulation. Many proteins harboring LxxLL motifs interact with ligand binding domains (LBDs) of nuclear receptors (for review see ref 82). Receptor-interacting protein-140 (RIP140) contains multiple LxxLL

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943–15957

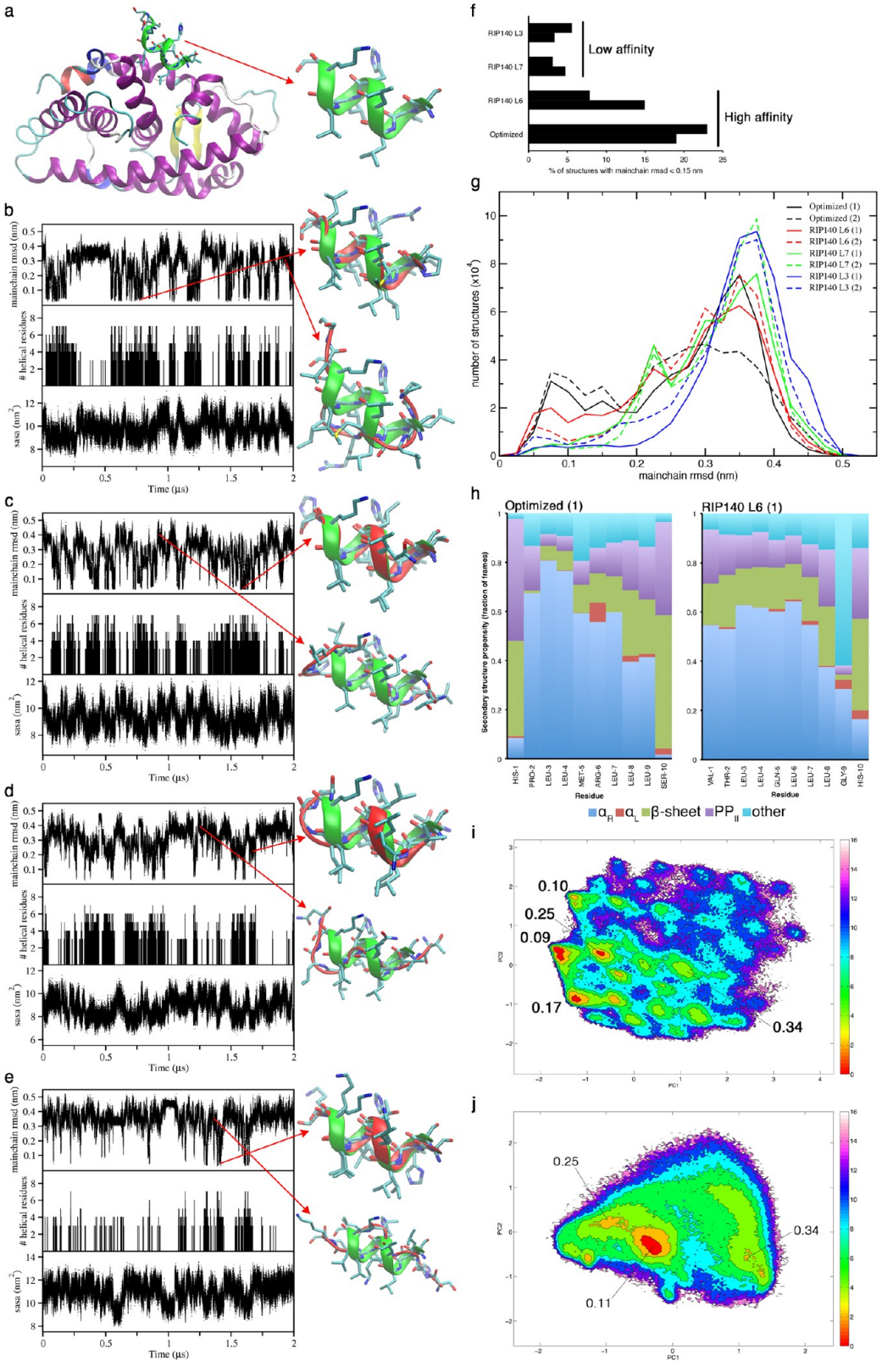


Figure 5. LxxLL motif complexed to uncomplexed state comparison. (a) X-ray structure of the RIP140 L5 LxxLL motif in complex with the LBD of estrogen-related receptor-γ (PDB id: 2GPO)[85] identifying the region of the RIP140 L5 peptide used for rmsd calculations. (b−e) Main chain rmsds to the RIP140 L5 complexed state structure (b, optimized peptide; c−e, RIP140 L6, L7, and L3 peptides, respectively). For the rmsd calculations, the corresponding main chain atoms of the different motifs were compared to the same 9-residue (SLLLHLLKS) segment of the L5 motif, which includes 36 main chain (N, C_α, C′, and CO) atoms. Peptide helicity was the sum of residues in all helix types (α, 3_{10}, and π) as assigned by the DSSP algorithm. Solvent-accessible surface areas were calculated using the g_sas tool.[28,103,104] (f) Percentage of all structures with main chain rmsds <0.15 nm. The two data points for each peptide (labeled 1, 2) were obtained from independent MD simulations initiated from different starting structures (Figure S4). (g) Distribution of the main chain rmsds. (h) Secondary structures classified using standard ϕ, ψ dihedral angle spaces. α_R (right handed

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

Figure 5. continued

α helix): $-175° \le \phi \le -55°$, $-80° \le \psi \le -5°$. α_L (left-handed α helix): $5° \le \phi \le 75°$, $25° \le \psi \le 120°$. β-sheet: $-180° \le \phi \le -50°$, $80° \le \psi \le -170°$. PP_{II} (polyproline II helix): $-110° \le \phi \le -50°$, $120° \le \psi \le 180°$. Because of overlapping dihedral angle spaces, PP_{II} conformations were subtracted from the β-sheet region. (i, j) Gibbs free energy landscapes (kJ/mol) based on the first two principal components from (i) dPCA and (j) cPCA with least-squares fit to the main chain atoms of the L5 bound state structure. Values mapped onto the landscapes indicate the average main chain rmsds (to the bound state structure) of the structures contained in the basins. PCA was accomplished using the GROMACS tools g_covar and g_anaeig.[28]

motifs (L1-L9), with differential affinities for LBDs.[83,84] The X-ray structure of the RIP140 L5 LxxLL motif in complex with the LBD of estrogen-related receptor-γ (PDB id: 2GPO)[85] shows that the region encompassing the LxxLL motif forms a short amphipathic helical structure with hydrophobic side chains facing the LBD (Figure 5a).

MD simulations of RIP140 L6, L7 and L3 LxxLL motif peptides,[84] as well as a ribosome display optimized LxxLL motif peptide,[86] show that the peptides have propensities toward bound state conformations, but to different extents (Figure 5b–e and Figure S4). The main chain rmsds were fairly well correlated with helical content (optimized peptide, $r^2 = 0.54$). Analysis of the RIP140 L5 LxxLL–LBD structure (PDB id: 2GPO)[85] in the absence of the LBD indicated that eight consecutive residues within the analyzed region (**SLLLHLLKSQ**) have a >10% decrease in their solvent-accessible surface areas, relative to extended (A–X–A) conformations.[67] Decreases in solvent exposed surfaces of residues in the corresponding positions were moderately correlated with main chain rmsd (optimized peptide $r^2 = 0.33$) (Figure 5b).

The LxxLL motifs examined here have differential affinities for LBDs, with the optimized peptide having the highest affinity, followed by the L6, L7, and L3 motifs of RIP140.[84,86] Parts f and g of Figure 5 show that the optimized LxxLL motif peptide sampled bound state-like conformations more frequently than the RIP140 peptides, which could indicate that its higher binding affinity is at least partially due to a lower entropic barrier.

Several factors may explain why the optimized peptide had higher preference for bound state-like structures. The optimized LxxLL peptide contains prolines flanking the N- and C-terminal ends of the LM—a common attribute of helical LMs.[86,87] An N-terminal proline has been shown to promote helix initiation, while a C-terminal proline encourages helix termination, effectively marking the LM boundaries.[86,87] Secondary structure analysis revealed that the residues flanking the helix (H1 and S10) of the optimized peptide were less helical compared to the corresponding positions in L6 (V1 and H10) (Figure 5h). Meanwhile, the region flanked by these residues showed moderately higher helical propensity in the optimized peptide compared to L6. It was also interesting that the L6 motif sampled bound state-like conformations more frequently than the L7 motif, despite having nearly identical sequences. One possible explanation is the presence of a leucine at the −1 position in L6 peptide, compared to a less hydrophobic valine in L7. A preference for leucine at the −1 position is a common attribute of high affinity LxxLL motifs.[84] The optimized peptide contains leucines at this position, and also at the +6 position. Among the LxxLL motif proteins identified to date, none appear to have leucines at both −1 and +6. These residues may also contribute to increased sampling of bound state-like structures, which could modulate binding affinity.

In addition to helical content and solvent-accessible surface areas, correlations between several other combinations of variables were used to investigate the factors involved in formation of bound state-like structures (Table 3). For the optimized peptide, dPCA PC1 and main chain rmsd were fairly well correlated ($r^2 = 0.63$), illustrating that the folding process is the most dominant motion. The motion captured by the dPCA PC1 was strongly correlated with peptide compactness ($r^2 = 0.76$). cPCA PC1 was considerably less correlated to main chain rmsd ($r^2 = 0.22$). Similar to the ExoS example, the PC1 vs PC2 landscape from dPCA (Figure 5i) was more rugged than that from cPCA (Figure 5j). Despite this, both analyses clearly indicate small (~6 kJ/mol) free energy barriers across a large rmsd range.

Table 3. LxxLL Optimized Peptide Correlations

variable 1	variable 2	r^2
main chain rmsd	number of helical residues	0.54
main chain rmsd	solvent-accessible surface area (SASA)	0.33
main chain rmsd	radius of gyration	0.32
main chain rmsd	dPCA PC1	0.63
main chain rmsd	dPCA PC2	<0.1
radius of gyration	dPCA PC1	0.76
radius of gyration	dPCA PC2	<0.1
main chain rmsd	cPCA PC1 (main chain fit to bound state structure)	0.22
main chain rmsd	cPCA PC2 (main chain fit to bound state structure)	<0.1

5. Tcf4–β-Catenin. Interactions between Tcf family proteins and β-catenin play crucial roles in the canonical Wnt signaling pathway,[88] which regulates gene transcription. In the nucleus, Tcf proteins associate with β-catenin, leading to the transcriptional activation of numerous genes, many of which are associated with cancers.[88] As a result, there are efforts to characterize and design drugs that specifically target Tcf–β-catenin interactions.[89−91] Tcf4 interacts with β-catenin though its intrinsically disordered N-terminal catenin binding domain (CDB). The X-ray structure of the Tcf4 peptide–β-catenin complex (PDB id: 1JPW)[92] shows that three distinct regions of Tcf4 participate in the interaction. Here, the helical region, located toward the C-terminal end of the CDB, is examined (Figure 6a). Mutations to hydrophobic residues in this region lead to significant decreases in β-catenin binding affinity.[89]

MD simulations of the helix-forming region of the CDB in the absence of β-catenin showed that the peptide transiently formed bound state like conformations on several occasions throughout the trajectory (Figure 6b). Structures with main chain rmsds <0.15 nm were found in only ~1.7% of the frames, suggesting that the free state ensemble is only slightly populated with bound state-like conformations. The main chain rmsds were moderately correlated with peptide helical content ($r^2 = 0.37$). Analysis of the bound state CDB–β-catenin structure (PDB id: 1JPW)[92] in the absence of β-catenin

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

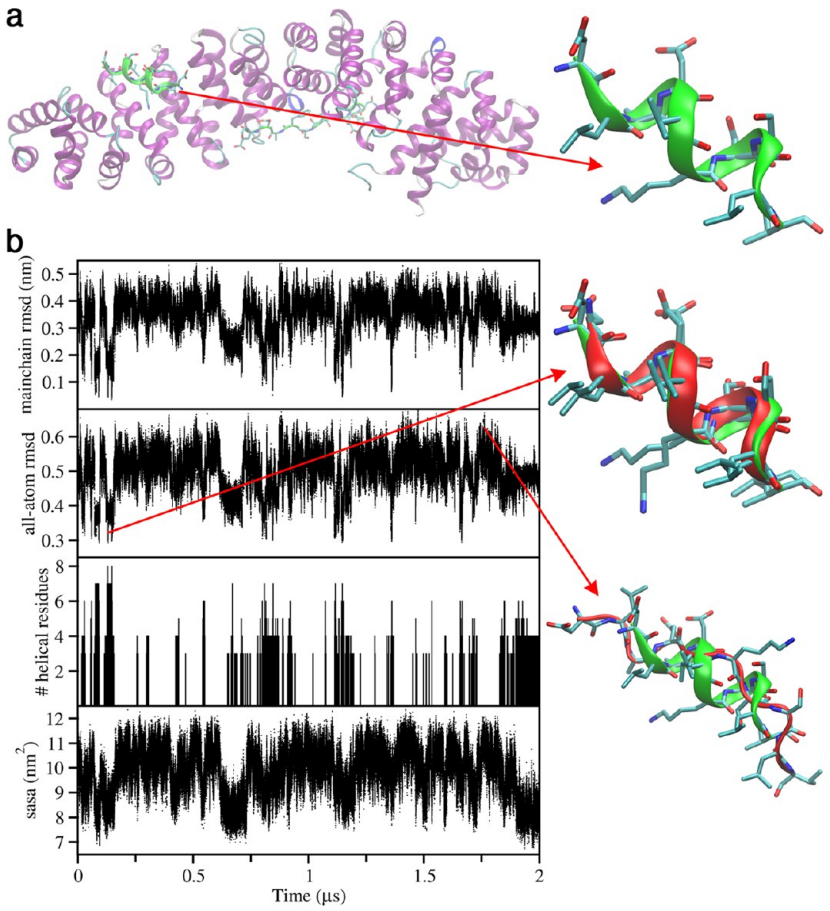


Figure 6. Tcf4–β-catenin complexed to uncomplexed state comparison. (a) X-ray structure of the Tcf4 peptide–β-catenin complex (PDB id: 1JPW)[92] identifying the region of the Tcf4 peptide used for rmsd calculations. (b) Main chain and all-atom rmsds to the complexed state. For the rmsd calculations, 10 residues (DLADVKSSLV), including 40 main chain (N, C_α, C′, and CO) and 72 non-hydrogen atoms were compared. Peptide helicity was the sum of residues in all helix types (α, 3_{10}, and π) as assigned by the DSSP algorithm. Solvent-accessible surface areas were calculated using the g_sas tool.[28,103,104]

indicated that 9 consecutive residues within the analyzed region (D**LADVKSSLV**) have >10% decreases in their solvent-accessible surface areas, relative to extended (A–X–A) conformations.[67] Decreases in solvent exposed surfaces of these residues were moderately correlated with main chain rmsd (r^2 = 0.31) (Figure 6b). A duplicate simulation initiated from a different starting structure yielded highly similar results, with main chain rmsds <0.15 nm found in ∼1.2% of the frames (Figure S5).

6. p53–MDM2. The tumor suppressor p53 has been termed nothing less than the "guardian of the genome".[93] It is an important transcription factor with a number of roles, including cell cycle control, stress response and apoptosis.[94] p53 activity is regulated through interactions with the MDM2 protein.[94] Interactions between the disordered N-terminal transactivation domain (TAD) of p53 and MDM2 have been extensively studied; agents that can specifically disrupt the TAD-MDM2 interaction are thought to be potential cancer therapeutics.[95−99]

In a complex with MDM2 (PDB id: 1YCR),[99] the p53 TAD forms an amphipathic helix with its hydrophobic side chains pointing toward MDM2 (Figure 7a). MD simulations of the p53 wild-type and P27S mutant TAD show sampling of conformations with close resemblance of their bound state structures (Figure 7b,c and Figure S6). Peptide helical content tended to increase correspondingly with the main chain rmsd decreasing (p53 WT, r^2 = 0.44, 0.40 (duplicate); p53 P27S r^2 = 0.33, 0.45). Analysis of the bound state TAD structure (PDB id: 1YCR)[99] in the absence of MDM2 indicated that eight consecutive residues within the analyzed region (T**FSDLWKLL**) have >10% decreases in their solvent-accessible surface areas, relative to extended (A–X–A) conformations.[67] Decreases in solvent exposed surfaces of these residues were correlated to some extent with main chain rmsd (p53 WT, r^2 = 0.20, 0.17; p53 P27S r^2 = 0.17, 0.18). In the WT p53 peptide simulation (first duplicate), a relatively long-lived β-turn conformation was observed (Figure 7, parts b and f–h). Such a conformation was not observed in the duplicate WT or P27S mutant simulation.

The TAD peptides examined here have varying affinities for MDM2.[96] The p53 P27S mutant binds MDM2 with ∼20-fold higher affinity than the wild-type p53.[96] Although the TAD is classified as being intrinsically disordered, experimental and computational investigations have suggested that it can adopt compact helical structures that closely resemble the MDM2 bound state.[20,21,23,96] The p53 P27S peptide is considerably more helical in its uncomplexed state compared to the wild-type peptide.[96] Furthermore, the large difference in MDM2 binding affinity arises mainly due to an increased entropy gain for the P27S peptide.[96] Our MD data is in qualitative agreement with this result, and shows that the P27S peptide forms low rmsd (main chain <0.15 nm) conformations more

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

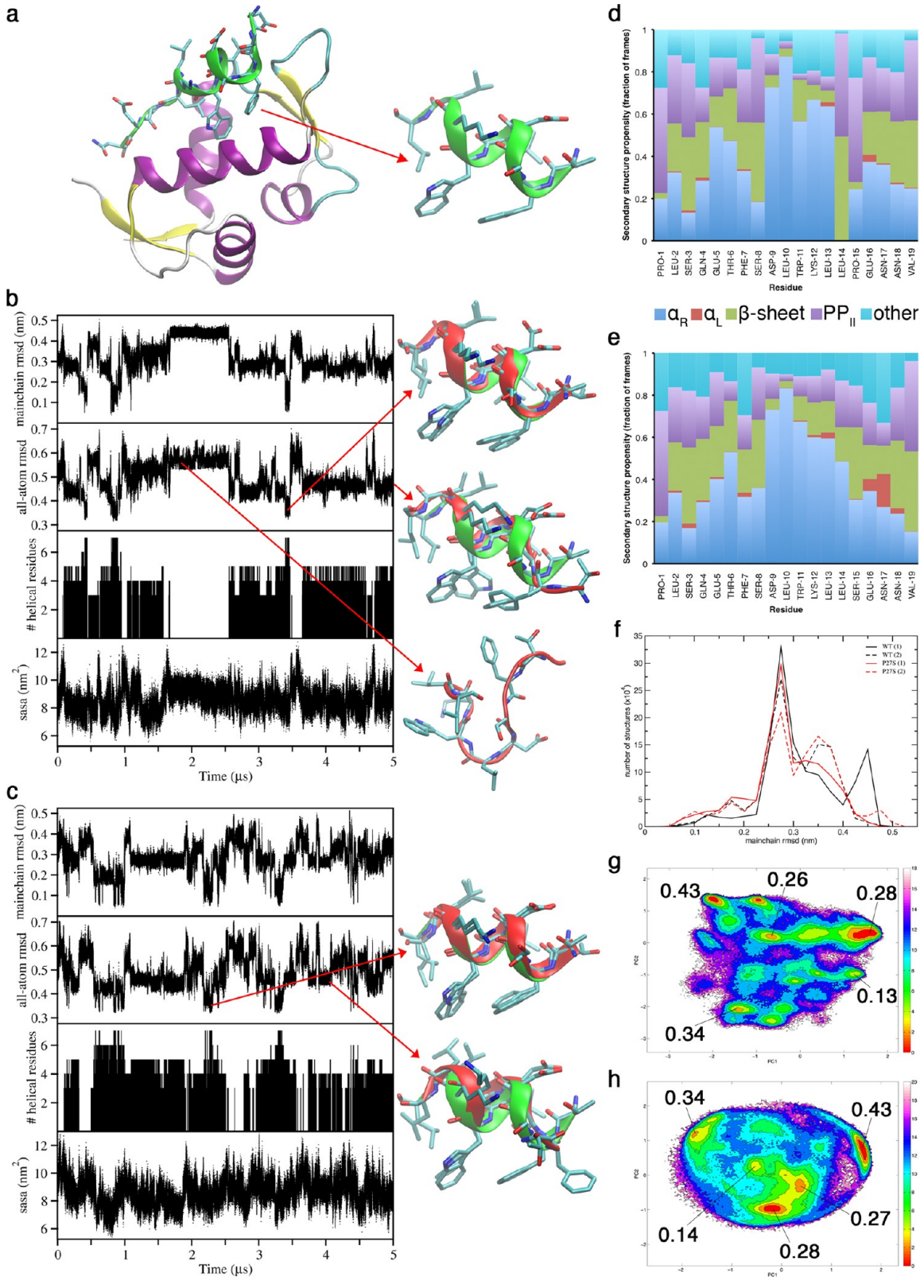


Figure 7. p53 TAD complexed to uncomplexed state comparison. (a) X-ray structure of the p53 TAD in complex with MDM2 (PDB id: 1YCR)[99] identifying the region of the TAD used for rmsd calculations. (b, c) Main chain and all-atom rmsds to the complexed state for the (b) WT and (c) P27S peptide. For the rmsd calculations, nine residues (TFSDLWKLL), including 36 main chain (N, C_α, C′, and CO) and 79 non-hydrogen atoms were compared to the corresponding atoms in the MDM2 bound state structure.[99] Peptide helicity was the sum of residues in all helix types (α, 3_{10}, and π) as assigned by the DSSP algorithm. Solvent-accessible surface area was calculated using the g_sas tool.[28,103,104] (d, e) Secondary structures classified using standard ϕ, ψ dihedral angle spaces. α_R (right handed α helix): $-175° \le \phi \le -55°$, $-80° \le \psi \le -5°$. α_L (left-handed α helix): $5° \le \phi \le 75°$, $25° \le \psi \le 120°$. β-sheet: $-180° \le \phi \le -50°$, $80° \le \psi \le -170°$. PP_{II} (polyproline II helix): $-110° \le \phi \le -50°$, $120° \le \psi \le 180°$ for the (d) WT and (e) P27S peptide. Because of overlapping dihedral angle spaces, PP_{II} conformations were subtracted from the β-sheet region. (f)

dx.doi.org/10.1021/jp407536p | J. Phys. Chem. B 2013, 117, 15943−15957

Figure 7. continued

Distribution of the main chain rmsds. (g, h) Gibbs free energy landscapes (kJ/mol) of the WT peptide based on the first two principal components from (g) dPCA and (h) cPCA with least-squares fit to the main chain atoms of the bound state structure. Values mapped onto the landscapes indicate the average main chain rmsds (to the bound state structure) of the structures contained in the basins. PCA was accomplished using the GROMACS tools g_covar and g_anaeig.[28]

frequently than the wild type (6.3, 5.7 vs 3.3, 3.5)% of the frames). The percentages of <0.15 nm main chain rmsd structures reported here for the wild type p53 are lower than the ∼20% prefolded population determined experimentally.[23,100] However, as shown in parts b and c of Figure 7, the intermediate (main chain <0.3 nm) rmsd states have appreciable helical content and bound state resemblance. Using the less stringent criteria (main chain rmsd <0.3 nm), the percentage of frames was 61.7, 48.7 and 53.3, 55.6 for the P27S mutant and wild type, respectively. The main chain rmsd distributions (Figure 7f) show that while conformations with high bound state resemblance (rmsd <0.15 nm) populate the ensemble to some extent, the majority of structures have an intermediate rmsd <0.3 nm.

Secondary structure analysis (Figure 7, parts d and e) showed that the helical propensity of the leucine preceding the proline residue was significantly lower for the wild type compared to P27S. A likely explanation for these trends is that the backbone NH of the substituted serine could form a hydrogen bond with the backbone carbonyl oxygen of W23, leading to the extension of the helix. This specific hydrogen bond was found to occur in 23.6, 16.0% of the P27S trajectory frames (using a NH−O distance cutoff of 0.35 nm and O−NH angle cutoff of 30°). The LxxLL optimized peptide example discussed earlier (Figure 5) shows analogous behavior of prolines decreasing the helical propensities of preceding residues. In the LxxLL example, however, helix termination at proline is thought to enhance target-binding affinity, whereas for the p53 TAD, helix extension results in a more favorable interaction.

Correlations between additional variables were assessed to further investigate the factors involved in formation of bound state-like structures (Table 4). Once again, the dPCA PC1 was

Table 4. p53 Correlations

variable 1	variable 2	r^2 (WT)
main chain rmsd	radius of gyration	<0.1
main chain rmsd	dPCA PC1	0.41
main chain rmsd	dPCA PC2	0.12
radius of gyration	dPCA PC1	<0.1
radius of gyration	dPCA PC2	<0.1
main chain rmsd	cPCA PC1 (main chain fit to bound state structure)	0.22
main chain rmsd	cPCA PC2 (main chain fit to bound state structure)	0.26

one of the parameters best correlated with main chain rmsd to the bound state structure (p53 WT, r^2 = 0.41). The cPCA PC1 and PC2 were correlated with rmsd to lesser extents (Table 4). Furthermore, peptide compactness was not highly correlated with formation of bound state-like structures.

DISCUSSION

LMs are commonly found at important target binding regions of disordered proteins and regions. They are responsible for crucial biological functions, such as transcriptional regulation, cell signaling, and protein trafficking. Accurate detection and characterization of these motifs is essential to understand their functions. Experimental characterization of LM can be challenging, especially in cases where the populations of "folded" molecules in a solution is very low.[5,60] Several recent works have demonstrated the use of MD simulations to identify and assess conformational preferences of a number of these short amino acid sequences.[5,17,18,22,55,60,66,101,102] Here, MD simulations were used to examine LMs from six different proteins in their uncomplexed states. The results suggest that it may be fairly common for the uncomplexed states of LMs to have structural propensities toward bound state conformations. Furthermore, it is demonstrated that the specific conformational biases encoded by different amino acid sequences could hold important clues regarding binding mechanisms, affinities and specificities. This work emphasizes that MD simulations can be useful tools for identifying and studying LMs. However, it should be stressed that even longer simulations and/or more replicates are required to more clearly establish the roles of preformed structures in regulating the binding affinities, and to accurately determine the free energy landscapes of LMs.

In addition to assessing bound state resemblance, efforts were taken to further characterize the free state ensembles of the LMs. One property that was examined for all systems was the solvent-accessible surface areas. LMs are primarily identified by regions of increased hydrophobic content, relative to their surroundings,[7] which can promote local structure formations. Indeed, several of the examples here showed that main chain rmsd values had moderate correlations with solvent exposure of primarily hydrophobic residues. Therefore, a hydrophobic collapse mechanism appears to be important for the folding and stability of some LMs.

Better correlations with main chain rmsd values were typically obtained from dPCA and cPCA. Both dPCA and cPCA usually identified the dominant folding motions of LMs in the first couple principal components; however, dPCA components were better correlated with main chain rmsd. dPCA typically yielded a single principal component with good correlation to main chain rmsd, while cPCA often had two components with lower rmsd correlations. The free energy landscapes generated by the two methods also had clear differences. dPCA landscapes were more rugged, and effectively indicated the specific ϕ, ψ angle basins of the peptides. Despite these differences, the major structural basins were well illustrated in both the dPCA and cPCA landscapes, and they served as complementary methods for analyzing the structural ensembles of the LMs. Overall, the findings from these analyses align well with other data from other disordered peptides,[55,66] which show heterogeneous structural ensembles, which often contain some specific conformational preferences. The shallow free energy barriers between energy minima gives disordered peptides efficient routes for structural interconversion, which is thought to be crucial to their biological functions. The dPCA method was introduced by Mu et al. in 2005.[57] dPCA uses internal coordinates (dihedral angles as the name suggests) and

thus is not sensitive to the overall motion of the protein. In dPCA, the periodicity of the angles needs to be taken into account and Mu et al. resolved this by applying a mapping to a metric space. They then compared the method to cPCA and demonstrated its ability to resolve more details from the free-energy landscape by identifying the folding pathways for penta-alanine in explicit water. Although some doubts of the method were raised,[56] they appear to have been resolved[58] and dPCA has been successfully used to study, e.g., IDPs[57] and RNA.[105] Despite dPCA's success, it is, however, important to notice that all PCA methods have potential issues regarding the uniqueness of the states. Whether or not they are present has to be established by comparison with other methods and experiments. Here, dPCA proved successful in revealing more details of the free-energy landscape. The dPCA results helped to understand the details of the main chain rmsd and radius of gyration, and thus to explain the details of LM folding pathways.

The results presented in this work have broad implications in a number of important areas. For instance, advancing our understanding of the mechanisms whereby disordered amino acid sequences bind targets, identifying and characterizing LM folding pathways and intermediates, modeling of disordered proteins/regions, and computational prediction of binding affinities.

■ ASSOCIATED CONTENT

Supporting Information

ExoS−14-3-3, amphiphysin 1−clathrin, β-arrestin 2−clathrin, LxxLL motif, Tcf4−β-catenin, and p53 TAD complexed to uncomplexed state comparison. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: (W.-Y.C.) jchoy4@uwo.ca.

*E-mail: (M.K.) mkarttu@gmail.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Computational resources were provided by SharcNet (www.sharcnet.ca), the SciNet HPC Consortium and Compute Canada. Financial support was provided by the Ontario Early Researcher Award Program, Canadian Institutes of Health Research (CIHR; MOP 74679), University of Waterloo and Natural Sciences and Engineering Research Council of Canada (NSERC).

■ REFERENCES

(1) Babu, M. M.; Kriwacki, R. W.; Pappu, R. V. Versatility From Protein Disorder. *Science* **2012**, *337*, 1460−1461.

(2) Tompa, P. Unstructural Biology Coming of Age. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419−425.

(3) Tompa, P. Intrinsically Disordered Proteins: A 10-Year Recap. *Trends Biochem. Sci.* **2012**, *37*, 509−516.

(4) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. J. Protein Dynamics and Conformational Disorder in Molecular Recognition. *Mol. Recognit.* **2010**, *23*, 105−116.

(5) Khan, H.; Cino, E. A.; Brickenden, A.; Fan, J.; Yang, D.; Choy, W. Y. Fuzzy Complex Formation Between the Intrinsically Disordered Prothymosin α and the Kelch Domain of Keap1 Involved in the Oxidative Stress Response. *J. Mol. Biol.* **2013**, *425*, 1011−1027.

(6) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868−8873.

(7) Davey, N. E.; Van Roey, K.; Weatheritt, R. J.; Toedt, G.; Uyar, B.; Altenberg, B.; Budd, A.; Diella, F.; Dinkel, H.; Gibson, T. J. Attributes of Short Linear Motifs. *Mol. Biosyst.* **2012**, *8*, 268−281.

(8) Dinkel, H.; Michael, S.; Weatheritt, R. J.; Davey, N. E.; Van Roey, K.; Altenberg, B.; Toedt, G.; Uyar, B.; Seiler, M.; Budd, A.; Jödicke, L.; Dammert, M. A.; Schroeter, C.; Hammer, M.; Schmidt, T.; Jehl, P.; McGuigan, C.; Dymecka, M.; Chica, C.; Luck, K.; Via, A.; Chatr-Aryamontri, A.; Haslam, N.; Grebnev, G.; Edwards, R. J.; Steinmetz, M. O.; Meiselbach, H.; Diella, F.; Gibson, T. J. ELM—The Database of Eukaryotic Linear Motifs. *Nucleic Acids Res.* **2012**, *40*, D242−D251.

(9) Das, R. K.; Mao, A. H.; Pappu, R. V. Unmasking Functional Motifs Within Disordered Regions of Proteins. *Sci. Signal.* **2012**, *5*, pe17.

(10) Fuxreiter, M.; Tompa, P.; Simon, I. Local Structural Disorder Imparts Plasticity on Linear Motifs. *Bioinformatics* **2007**, *23*, 950−956.

(11) Mészáros, B.; Dosztányi, Z.; Simon, I. Disordered Binding Regions and Linear Motifs—Bridging the Gap Between Two Models of Molecular Recognition. *PLoS One* **2012**, *7*, e46829.

(12) Hunt, T. Protein Sequence Motifs Involved in Recognition and Targeting: A New Series. *Trends Biochem. Sci.* **1990**, *15*, 305.

(13) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y. Y.; Romero, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. Rational Drug Design Via Intrinsically Disordered Protein. *Trends Biotechnol.* **2006**, *24*, 435−442.

(14) Uversky, V. N. *Expert.* Intrinsically Disordered Proteins and Novel Strategies for Drug Discovery. *Opin. Drug. Discov.* **2012**, *7*, 475−488.

(15) Davey, N. E.; Edwards, R. J.; Shields, D. C. Estimation and Efficient Computation of the True Probability of Recurrence of Short Linear Protein Sequence Motifs in Unrelated Proteins. *BMC Bioinformatics* **2010**, *11*, 14.

(16) Haslam, N. J.; Shields, D. C. Profile-Based Short Linear Protein Motif Discovery. *BMC Bioinformatics* **2012**, *13*, 104.

(17) Cino, E. A.; Wong-Ekkabut, J.; Karttunen, M.; Choy, W. Y. Microsecond Molecular Dynamics Simulations of Intrinsically Disordered Proteins Involved in the Oxidative Stress Response. *PLoS One* **2011**, *6*, e27371.

(18) Cino, E. A.; Choy, W. Y.; Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory. Comput.* **2012**, *8*, 2725−2740.

(19) Cino, E. A.; Karttunen, M.; Choy, W. Y. Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins. *PLoS One* **2012**, *7*, e49876.

(20) Espinoza-Fonseca, L. M.; Trujillo-Ferrara, J. G. Transient Stability of the Helical Pattern of Region F19-L22 of the N-terminal Domain of p53: A Molecular Dynamics Simulation Study. *Biochem. Biophys. Res. Commun.* **2006**, *343*, 110−116.

(21) Espinoza-Fonseca, L. M. Leucine-Rich Hydrophobic Clusters Promote Folding of the N-terminus of the Intrinsically Disordered Transactivation Domain of p53. *FEBS. Lett.* **2009**, *583*, 556−560.

(22) Lambrughi, M.; Papaleo, E.; Testa, L.; Brocca, S.; De Gioia, L.; Grandori, R. Intramolecular Interactions Stabilizing Compact Conformations of the Intrinsically Disordered Kinase-Inhibitor Domain of Sic1: A Molecular Dynamics Investigation. *Front. Physiol.* **2012**, *3*, 435.

(23) Lee, H.; Mok, K. H.; Muhandiram, R.; Park, K. H.; Suk, J. E.; Kim, D. H.; Chang, J.; Sung, Y. C.; Choi, K. Y.; Han, K. H. Local Structural Elements in the Mostly Unstructured Transcriptional Activation Domain of Human p53. *J. Biol. Chem.* **2000**, *275*, 29426−29432.

(24) Staneva, I.; Huang, Y.; Liu, Z.; Wallin, S. Binding of Two Intrinsically Disordered Peptides to a Multi-Specific Protein: A Combined Monte Carlo and Molecular Dynamics Study. *PLoS Comput. Biol.* **2012**, *8*, e1002682.

(25) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(26) Brunger, A. T. Version 1.2 of the Crystallography and NMR System. *Nat. Protoc.* **2007**, *2*, 2728−2733.
(27) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects From Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.
(28) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(29) MacKerell, A. A.; Bashford, B.; Bellott; Dunbrack, R. R.; Evanseck, J. J.; Field, M. M.; Fischer, F.; Gao, G.; Guo, G.; Ha, H.; Joseph-McCarthy, J. -M.; Kuchnir, K.; Kuczera, K.; Lau, F. F. T.; Mattos, M.; Michnick, M.; Ngo, N.; Nguyen, D. D.; Prodhom, P.; Reiher, W. W.; Roux, R.; Schlenkrich, S.; Smith, J. J.; Stote, S.; Straub, S.; Watanabe, W.; Wiórkiewicz-Kuczera, W. -K.; Yin, Y.; Karplus, K. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(30) Mackerell, A. A.; Feig, F.; Brooks, C. C. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(31) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations With Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(32) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Chem. Theory Comput.* **2004**, *25*, 1656−1676.
(33) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F.; van Gunsteren, W. F. Validation of the 53A6 GROMOS Force Field. *Eur. Biophys. J.* **2005**, *34*, 273−284.
(34) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(35) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(36) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.
(37) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99Sb Protein Force Field. *Proteins* **2010**, *78*, 1950−1958.
(38) Sorin, E. J.; Pande, V. S. Exploring the Helix-Coil Transition Via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472−2493.
(39) Best, R. B.; Mittal, J. Free-Energy Landscape of the GB1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins* **2011**, *79*, 1318−1328.
(40) Freddolino, P. P.; Liu, L.; Gruebele, G.; Schulten, S. *Biophys. J.* **2008**, *94*, L75.
(41) Mittal, J.; Best, R. B. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys. J.* **2010**, *99*, L26−L28.
(42) Lindorff-Larsen, L. -L.; Maragakis, M.; Piana, P.; Eastwood, M. M.; Dror, R. R.; Shaw, D. D. Systematic Validation of Protein Force Fields Against Experimental Data. *PLoS One* **2012**, *7*, e32131.
(43) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-Level Description of Ubiquitin Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915−5920.
(44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(45) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullmann, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; p 331.
(46) Vega, C.; Abascal, J. L. Simulating Water with Rigid Non-Polarizable Models: A General Perspective. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663−19688.
(47) Patra, M.; Karttunen, M. Systematic Comparison of Force Fields for Microscopic Simulations of Nacl in Aqueous Solutions: Diffusion, Free Energy of Hydration, and Structural Properties. *J. Comput. Chem.* **2004**, *25*, 678−689.
(48) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(49) Wong-ekkabut, J.; Miettinen, M. S.; Dias, C.; Karttunen, M. Static Charges Cannot Drive a Continuous Flow of Water Molecules Through a Carbon Nanotube. *Nat. Nanotechnol.* **2010**, *5*, 555−557.
(50) Wong-Ekkabut, J.; Karttunen, M. Assessment of Common Simulation Protocols for Simulations of Nanopores, Membrane Proteins & Channels. *J. Chem. Theory Comput.* **2012**, *8*, 2905−2911.
(51) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(52) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Johannes, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(53) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(54) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chem. Rev.* **2013**, in press.
(55) Espinoza-Fonseca, L. M.; Ilizaliturri-Flores, I.; Correa-Basurto, J. Backbone Conformational Preferences of An Intrinsically Disordered Protein in Solution. *Mol. Biosyst.* **2012**, *8*, 1798−1805.
(56) Hinsen, K. Comment On: "Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis". *Proteins* **2006**, *64*, 795−797.
(57) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins* **2005**, *58*, 45−52.
(58) Mu, Y.; Nguyen, P.; Stock, G. Reply to the Comment on "Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis". *Proteins* **2006**, *64*, 798−799.
(59) Cao, Z.; Liu, L.; Wang, J. Why the OPLS-AA Force Field Cannot Produce the B-Hairpin Structure of H1 Peptide in Solution When Comparing with the GROMOS 43A1 Force Field? *J. Biomol. Struct. Dyn.* **2011**, *29*, 527−539.
(60) Cino, E. A.; Killoran, R. C.; Karttunen, M.; Choy, W. Y. Binding of Disordered Proteins to a Protein Hub. *Sci Rep.* **2013**, *3*, 2305.
(61) Bustos, D. M. The Role of Protein Disorder in the 14-3-3 Interaction Network. *Mol. Biosyst.* **2012**, *8*, 178−184.
(62) Jin, J.; Smith, F. D.; Stark, C.; Wells, C. D.; Fawcett, J. P.; Kulkarni, S.; Metalnikov, P.; O'Donnell, P.; Taylor, P.; Taylor, L.; Zougman, A.; Woodgett, J. R.; Langeberg, L. K.; Scott, J. D.; Pawson, T. Proteomic, Functional, and Domain-Based Analysis of in Vivo 14−3-3 Binding Proteins Involved in Cytoskeletal Regulation and Cellular Organization. *Curr. Biol.* **2004**, *14*, 1436−1450.
(63) Dougherty, M. K.; Morrison, D. K. Unlocking the Code of 14−3-3. *J. Cell. Sci.* **2004**, *117*, 1875−1884.
(64) Henriksson, M. L.; Francis, M. S.; Peden, A.; Aili, M.; Stefansson, K.; Palmer, R.; Aitken, A.; Hallberg, B. A Non-phosphorylated 14−3-3 Binding Motif on Exoenzyme S That Is Functional in Vivo. *Eur. J. Biochem.* **2002**, *269*, 4921−4929.
(65) Yang, X.; Lee, W. H.; Sobott, F.; Papagrigoriou, E.; Robinson, C. V.; Grossmann, J. G.; Sundström, M.; Doyle, D. A.; Elkins, J. M. Structural Basis for Protein-Protein Interactions in the 14−3-3 Protein Family. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17237−17242.

(66) Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. Structural Ensemble of An Intrinsically Disordered Polypeptide. *J. Phys. Chem. B* **2013**, *117*, 118−124.
(67) Samanta, U.; Bahadur, R. P.; Chakrabarti, P. Quantifying the Accessible Surface Area of Protein Residues in Their Local Environment. *Protein Eng.* **2002**, *15*, 659−667.
(68) Dias, C. L.; Karttunen, M.; Chan, H. S. Hydrophobic Interactions in the Formation of Secondary Structures in Small Peptides. *Phys. Rev. E* **2011**, *84*, 041931.
(69) Huang, W.; Lin, Z.; van Gunsteren, W. F. Validation of the GROMOS 54A7 Force Field with Respect to B-Peptide Folding. *J. Chem. Theory. Comput.* **2011**, *7*, 1237−1243.
(70) Schmid, S. L. Clathrin-Coated Vesicle Formation and Protein Sorting: An Integrated Process. *Annu. Rev. Biochem.* **1997**, *66*, 511−548.
(71) Lafer, E. M. Clathrin-Protein Interactions. *Traffic* **2002**, *3*, 513−520.
(72) Miele, A. E.; Watson, P. J.; Evans, P. R.; Traub, L. M.; Owen, D. J. Two Distinct Interaction Motifs in Amphiphysin Bind Two Independent Sites on the Clathrin Terminal Domain Beta-Propeller. *Nat. Struct. Mol. Biol.* **2004**, *11*, 242−248.
(73) Kang, D. S.; Kern, R. C.; Puthenveedu, M. A.; von Zastrow, M.; Williams, J. C.; Benovic, J. L. Structure of An Arrestin2-Clathrin Complex Reveals a Novel Clathrin Binding Domain That Modulates Receptor Trafficking. *J. Biol. Chem.* **2009**, *284*, 29860−29872.
(74) ter Haar, E.; Harrison, S. C.; Kirchhausen, T. Peptide-In-Groove Interactions Link Target Proteins to the Beta-Propeller of Clathrin. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1096−1100.
(75) Krishna, T. S.; Kong, X. P.; Gary, S.; Burgers, P. M.; Kuriyan, J. Crystal Structure of the Eukaryotic DNA Polymerase Processivity Factor PCNA. *Cell* **1994**, *79*, 1233−1243.
(76) Maga, G.; Hubscher, U. Proliferating Cell Nuclear Antigen (PCNA): A Dancer with Many Partners. *J. Cell. Sci.* **2003**, *116*, 3051−3060.
(77) Shell, S. S.; Putnam, C. D.; Kolodner, R. D. The N Terminus of Saccharomyces Cerevisiae Msh6 is an Unstructured Tether to PCNA. *Mol. Cell* **2007**, *26*, 565−578.
(78) Warbrick, E. PCNA Binding Through a Conserved Motif. *Bioessays* **1998**, *20*, 195−199.
(79) Bruning, J. B.; Shamoo, Y. Structural and Thermodynamic Analysis of Human PCNA with Peptides Derived From DNA Polymerase-Delta P66 Subunit and Flap Endonuclease-1. *Structure* **2004**, *12*, 2209−2219.
(80) Gulbis, J. M.; Kelman, Z.; Hurwitz, J.; O'Donnell, M.; Kuriyan, J. Structure of the C-Terminal Region of P21(Waf1/Cip1) Complexed with Human PCNA. *Cell* **1996**, *87*, 297−306.
(81) Yoon, M. -K.; Venkatachalam, V.; Huang, A.; Choi, B. -S.; Stultz, C. M.; Chou, J. J. Residual Structure Within the Disordered C-Terminal Segment of P21Waf1/Cip1/Sdi1 and Its Implications for Molecular Recognition. *Protein Sci.* **2009**, *18*, 337−347.
(82) Plevin, M. J.; Mills, M. M.; Ikura, M. The Lxxll Motif: A Multifunctional Binding Sequence in Transcriptional Regulation. *Trends. Biochem. Sci.* **2005**, *30*, 66−69.
(83) Heery, D. M.; Kalkhoven, E.; Hoare, S.; Parker, M. G. A Signature Motif in Transcriptional Co-Activators Mediates Binding to Nuclear Receptors. *Nature* **1997**, *387*, 733−736.
(84) Heery, D. M.; Hoare, S.; Hussain, S.; Parker, M. G.; Sheppard, H. Core LXXLL Motif Sequences in Creb-Binding Protein, SRC1, and RIP140 Define Affinity and Selectivity for Steroid and Retinoid Receptors. *J. Biol. Chem.* **2001**, *276*, 6695−6702.
(85) Wang, L.; Zuercher, W. J.; Consler, T. G.; Lambert, M. H.; Miller, A. B.; Orband-Miller, L. A.; McKee, D. D.; Willson, T. M.; Nolte, R. T. X-Ray Crystal Structures of the Estrogen-Related Receptor-Gamma Ligand Binding Domain in Three Functional States Reveal the Molecular Basis of Small Molecule Regulation. *J. Biol. Chem.* **2006**, *281*, 37773−37781.
(86) Fuchs, S.; Nguyen, H. D.; Phan, T. T.; Burton, M. F.; Nieto, L.; de Vries-van Leeuwen, I. J.; Schmidt, A.; Goodarzifard, M.; Agten, S. M.; Rose, R.; Ottmann, C.; Milroy, L. G.; Brunsveld, L. Proline Primed Helix Length As a Modulator of the Nuclear Receptor-Coactivator Interaction. *J. Am. Chem. Soc.* **2013**, *135*, 4364−4371.
(87) Theillet, F. -X.; Kalmar, L.; Tompa, P.; Han, K. -H.; Selenko, P.; Dunker, A. K.; Daughdrill, G. W.; Uversky, V. N. The Alphabet of Intrinsic Disorder I. Act Like a Pro: On the Abundance and Roles of Proline Residues in Intrinsically Disordered Proteins. *Intrinsically Disordered Proteins* **2013**, *1*, 0−1.
(88) MacDonald, B. T.; Tamai, K.; He, X. Wnt/Beta-Catenin Signaling: Components, Mechanisms, and Diseases. *Dev. Cell* **2009**, *17*, 9−26.
(89) Fasolini, M.; Wu, X.; Flocco, M.; Trosset, J. Y.; Oppermann, U.; Knapp, S. Hot Spots in Tcf4 for the Interaction with Beta-Catenin. *J. Biol. Chem.* **2003**, *278*, 21092−21098.
(90) Sukhdeo, K.; Mani, M.; Zhang, Y.; Dutta, J.; Yasui, H.; Rooney, M. D.; Carrasco, D. E.; Zheng, M.; He, H.; Tai, Y. T.; Mitsiades, C.; Anderson, K. C.; Carrasco, D. R. Targeting the Beta-Catenin/Tcf Transcriptional Complex in the Treatment of Multiple Myeloma. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7516−7521.
(91) Tian, W.; Han, X.; Yan, M.; Xu, Y.; Duggineni, S.; Lin, N.; Luo, G.; Li, Y. M.; Han, X.; Huang, Z. Structure-Based Discovery of a Novel Inhibitor Targeting the B-Catenin/Tcf4 Interaction. *Biochemistry* **2012**, *51*, 724−731.
(92) Poy, F.; Lepourcelet, M.; Shivdasani, R. A.; Eck, M. J. Structure of a Human Tcf4—B-Catenin Complex. *Nat. Struct. Mol. Biol.* **2001**, *8*, 1053−1057.
(93) Lane, D. P. p53, Guardian of the Genome. *Nature* **1992**, *358*, 15−16.
(94) Moll, U. M.; Petrenko, O. The Mdm2-p53 Interaction. *Mol. Cancer Res.* **2003**, *1*, 1001−1008.
(95) Pazgier, M.; Liu, M.; Zou, G.; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zella, D.; Tarasov, S. G.; Lu, W. Structural Basis for High-Affinity Peptide Inhibition of P53 Interactions with MDM2 and MDMX. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4665−4670.
(96) Brown, C. J.; Dastidar, S. G.; Quah, S. T.; Lim, A.; Chia, B.; Verma, C. S. C-Terminal Substitution of MDM2 Interacting Peptides Modulates Binding Affinity by Distinctive Mechanisms. *PLoS One* **2011**, *6*, e24122.
(97) Czarna, A.; Popowicz, G. M.; Pecak, A.; Wolf, S.; Dubin, G.; Holak, T. A. High Affinity Interaction of the P53 Peptide-Analogue with Human Mdm2 and Mdmx. *Cell Cycle* **2009**, *8*, 1176−1184.
(98) Dömling, A. Hot, Hotter, Hottest. *Cell Cycle* **2009**, *8*, 1112−1113.
(99) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. Structure of the MDM2 Oncoprotein Bound to the P53 Tumor Suppressor Transactivation Domain. *Science* **1996**, *274*, 948−953.
(100) Lee, S. H.; Kim, D. H.; Han, J. J.; Cha, E. J.; Lim, J. E.; Cho, Y. J.; Lee, C.; Han, K. H. Understanding Pre-Structured Motifs (Presmos) in Intrinsically Unfolded Proteins. *Curr. Protein. Pept. Sci.* **2012**, *13*, 34−54.
(101) Arrigoni, A.; Grillo, B.; Vitriolo, A.; De Gioia, L.; Papaleo, E. C-terminal Acidic Domain of Ubiquitin-Conjugating Enzymes: A Multi-Functional Conserved Intrinsically Disordered Domain in Family 3 of E2 Enzymes. *J. Struct. Biol.* **2012**, *178*, 245−259.
(102) Espinoza-Fonseca, L. M. Reconciling Binding Mechanisms of Intrinsically Disordered Proteins. *Biochem. Biophys. Res. Commun.* **2009**, *382*, 479−482.
(103) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199−203.
(104) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. The Double Cubic Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies. *J. Comput. Chem.* **1995**, *16*, 273−284.
(105) Riccardi, L.; Nguyen, P. H.; Stock, G. Free-Energy Landscape of RNA Hairpins Constructed via Dihedral Angle Principal Component Analysis. *J. Phys. Chem. B* **2009**, *113*, 16660−16668.