J|A|C|S ARTICLES

Published on Web 03/11/2009

Atomistic Details of the Disordered States of KID and pKID. Implications in Coupled Binding and Folding

Debabani Ganguly and Jianhan Chen*

Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506

Received November 17, 2008; E-mail: jianhanc@ksu.edu

Abstract: Intrinsically disordered proteins (IDPs) are a newly recognized class of functional proteins for which a lack of stable tertiary fold is required for function. Because of the heterogeneous and dynamical nature, molecular modeling is necessary to provide the missing details of disordered states of IDP that are crucial for understanding their functions. In particular, generalized Born (GB) implicit solvent, combined with replica exchange (REX), might offer an optimal balance between accuracy and efficiency for modeling IDPs. We carried out extensive REX simulations in an optimized GB force field to characterize the disordered states of a regulatory IDP, KID domain of transcription factor CREB, and its phosphorylated form, pKID. The results revealed that both KID and pKID, though highly disordered on the tertiary level, are compact and mainly occupy a small number of helical substates. Interestingly, although phosphorylation of KID Ser133 leads only to marginal changes in average helicities on the ensemble level, underlying conformational substates differ significantly. In particular, pSer133 appears to restrict the accessible conformational space of the loop region and thus reduces the entropic cost of KID folding upon binding to the KIX domain of CREB-binding protein. Such an expanded role of phosphorylation in the KID:KIX recognition was not previously recognized because of a lack of substantial conformational changes on the ensemble level and inaccessibility of the structural details from experiments. The results also suggest that an implicit solvent-based modeling framework, despite various existing limitations, might be feasible for accurate atomistic simulation of small IDPs in general.

Introduction

Intrinsically disordered proteins (IDPs), also known as natively disordered, natively unfolded, or intrinsically unstructured proteins, are a newly recognized class of functional proteins that rely on a lack of stable tertiary structures for function.[1−5] Substantial regions, and sometimes entire domains, of these proteins can exist as dynamic ensembles of disordered, but *not necessarily random*, structures under physiological conditions. Many IDPs undergo disorder−order transitions upon recognition of their biological partners, and it has been proposed that function can arise from ordered, partially disordered, and fully disordered states and from transitions among them.[2,4] Such recognition extends the protein structure−function paradigm. IDPs are highly abundant in biology. Sequence analysis has estimated that nearly one-third of the eukaryotic proteins contain continuous disorder segments of at least 50 residues.[2,4] The Database of Disordered Proteins (DisProt) currently contains 484 experimentally characterized IDPs with a range of functions.[6,7] In contrast to structured proteins where functions are often enzymatic (thus require precise organization of active sites), IDPs are frequently involved in cellular signaling and regulation.[4,8] In particular, cancer-associated and signaling proteins contain significantly more intrinsic disorder.[9] Intrinsic disorder might facilitate function in several ways.[3−5] Enthalpy gained from forming a large contact surface upon binding is effectively compensated by the entropy cost of the accompanying folding of IDPs, such that high specificity can be achieved at the same time with low affinity. This property is often desired for signaling and regulation. The unstructured nature of IDPs also provides a level of control that allows rapid turnover for accurate responses, structural plasticity for binding multiple partners, and inducibility by posttranslational modifications.[10−12] Intrinsic thermo-instability of IDPs might also provide a basis for conformational modulation in allosteric controls.[13]

(1) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321–331.
(2) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573–6582.
(3) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527–533.
(4) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
(5) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35–41.
(6) Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L. M.; Cortese, M. S.; Lawson, J. D.; Brown, C. J.; Sikes, J. G.; Newton, C. D.; Dunker, A. K. *Bioinformatics* **2005**, *21*, 137–140.
(7) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2007**, *35*, D786–D793.
(8) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *J. Mol. Recognit.* **2005**, *18*, 343–384.
(9) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323*, 573–584.
(10) Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K. *Biophys. J.* **2007**, *92*, 1439–1456.
(11) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2004**, *32*, 1037–1049.
(12) Bourhis, J. M.; Canard, B.; Longhi, S. *Curr. Protein Pept. Sci.* **2007**, *8*, 135–149.
(13) Lane, D. P.; Hupp, T. R. *Drug Discov. Today* **2003**, *8*, 347–355.

10.1021/ja808999m CCC: $40.75

Arguably, functions of IDPs are determined by distribution and redistribution of accessible conformations.[14] Residual structures often persist in the disordered states and such preformed structural elements might have important implications in coupled binding and folding of IDPs during function. On one hand, it has been hypothesized that preformed elements might serve as initial contact points and facilitate induced folding on the surface of cofactors.[15,16] On the other hand, increasing the amount of local structures in the free states might actually reduce the binding rate for IDPs if binding occurs through formation of nonspecific encounter complexes.[17,18] In either scenario, extent of residual structures directly affects the entropic cost of folding and in turn modulates the binding thermodynamics. A detailed structural and dynamical characterization of the disordered states is thus necessary for understanding the action of IDPs. A range of experimental techniques, especially nuclear magnetic resonance (NMR) spectroscopy, can be used to study the ensemble properties, sometimes at residue level, and to detect the presence of transient (tertiary) structures for disordered and partially disordered states of proteins.[19−22] Nonetheless, high-resolution structures can only be determined for folded IDPs in complex with specific biological partners and it remains very difficult to identify the conformational substates that collectively give rise to the observed ensemble properties of the disordered states from experiments alone. The substantial challenges of experimental characterization of IDPs present a unique opportunity for all-atom molecular mechanics modeling to make critical contributions. Atomistic simulation could provide the ultimate level of detail necessary for understanding the nature of the disordered states and some success has already been demonstrated in the study of unstructured amyloid-forming peptides.[23−25] Modeling IDPs also presents substantial new challenges. Explicit inclusion of water molecules in conventional explicit solvent-based modeling methodologies increases the system size about 10-fold and leads to prohibitive computational cost if one wants to sufficiently sample the broad manifolds of functionally relevant states of IDPs. Mainly optimized for folded native states of proteins, the current explicit solvent protein force fields are also known to have limitations such as excessive helical contents[26,27] and overstabilized peptide−peptide interactions.[28] These systematic biases further limit the ability of explicit solvent simulations to distinguish between weakly stable conformations of peptides necessary for modeling IDPs. New strategies must be developed and explored to overcome these limitations, particularly in efficient modeling of the solvent environment.

So-called implicit solvent-based protein modeling might offer an optimal balance between accuracy and efficiency required for simulating disordered states of IDPs. Implicit solvent is essentially a coarse-grained approach where only the protein is represented at the atomic level and the mean influence of solvent is captured by the free energy cost of solvating the protein (thus solvent is implicit).[29] Such an implicit treatment of solvent reduces the system size about 10-fold, albeit at the expense of losing the microscopic detail of water. Importantly, recent advances in generalized Born (GB) methodologies allow accurate estimation of the solvation free energy of biomolecules with only a modest increase in the associated computational cost.[30] For example, molecular dynamics (MD) simulations using the GBSW implicit solvent are only about 4−5 times more expensive than those in vacuum.[31] Interestingly, implicit solvent is also ideal for replica exchange (REX) simulations,[32] where the best efficacy is achieved with moderate system sizes and up to 10-fold enhancement in sampling efficiency is possible.[33−35] Substantial gain in computational efficiency also allows one to parametrize the implicit solvent protein force field through extensive peptide folding simulations,[36−39] an effective strategy that is largely unfeasible for explicit solvent modeling. We have previously optimized the GBSW implicit solvent force field by using a set of small helices and β hairpins with a wide range of thermodynamic stabilities.[36] This led to a consistent GB force field with proper balance of solvation and intramolecular interactions and in turn allowed a high level of accuracy in describing the peptide conformational equilibrium. The same consistent force field has been successfully applied to study the folding and unfolding of several small peptides and proteins.[36,40−44] In particular, recent successes in modeling the pH-induced unstable states of several proteins suggest that such carefully optimized implicit solvent protein force fields, coupled with REX enhanced sampling, might provide a viable computational framework for modeling the disordered states of small IDPs in general.

(14) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. *Chem. Rev.* **2008**, *108*, 1225–1244.
(15) Receveur-Brechot, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. *Proteins* **2006**, *62*, 24–45.
(16) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* **2004**, *338*, 1015–1026.
(17) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–U11.
(18) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. *PLoS Comput. Biol.* **2008**, *4*, e1000060.
(19) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(20) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
(21) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739–756.
(22) Bracken, C.; Iakoucheva, L. M.; Rorner, P. R.; Dunker, A. K. *Curr. Opin. Struct. Biol.* **2004**, *14*, 570–576.
(23) Baumketner, A.; Krone, M. G.; Shea, J.-E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6027–6032.
(24) Lei, H.; Wu, C.; Wang, Z.; Duan, Y. *J. Mol. Biol.* **2006**, *356*, 1049–63.
(25) Tran, H. T.; Mao, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2008**, *130*, 7380–92.
(26) Yoda, T.; Sugita, Y.; Okamoto, Y. *Chem. Phys.* **2004**, *307*, 269–283.
(27) Best, R. B.; Buchete, N. V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07–9.
(28) Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477–85.
(29) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1–20.
(30) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265–284.
(31) Im, W. P.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
(32) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(33) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83.
(34) Nymeyer, H. *J. Chem. Theory Comput.* **2008**, *4*, 626–636.
(35) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15340–5.
(36) Chen, J. H.; Im, W. P.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.
(37) Chen, J. H.; Brooks, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–148.
(38) Jang, S.; Kim, E.; Pak, Y. *Proteins* **2007**, *66*, 53–60.
(39) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. *J. Comput. Chem.* **2003**, *24*, 21–31.
(40) Chen, J. H.; Brooks, C. L. *Proteins* **2007**, *67*, 922–930.
(41) Chen, J. H.; Brooks, C. L. *Phys. Chem. Chem. Phys.* **2008**, *10*, 471–481.
(42) Khandogin, J.; Chen, J. H.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18546–18550.
(43) Khandogin, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16880–16885.
(44) Khandogin, J.; Raleigh, D. P.; Brooks, C. L. *J. Am. Chem. Soc.* **2007**, *129*, 3056–3057.

In this work, we explore REX-MD simulations in the consistent GBSW protein force field to characterize the disordered states of the kinase-inducible domain (KID) of transcription factor CREB (cAMP response element binding protein) and its phosphorylated form pKID. KID is one of the most extensively characterized IDPs and can serve as a paradigm for understanding IDPs and testing the proposed modeling framework. In cell, CREB regulates transcription in response to cAMP signaling partially by binding to the coactivator CBP (CREB-binding protein) or its paralogue p300.[17] This recognition requires phosphorylation of Ser133 of KID and involves direct interaction between pKID and the KIX domain of CBP.[45,46] When in complex with KIX, pKID adopts a helix−linker−helix structure that involves residues 119−146.[47] However, free KID and pKID lack stable tertiary structures in solution. NMR chemical shift analysis has estimated that helix αA (residues 120−129) is about 50−60% folded and helix αB (residues 134−144,) is only 10−15% formed.[48] Isothermal titration calorimetry measurements showed that unphosphorylated KID bound to KIX with about 2 orders of magnitude lower affinity than pKID.[49] Precise roles of phosphorylation in the KID:KIX recognition are debatable. In principle, phosphorylation might facilitate binding by either modulating the peptide conformational preferences or providing a site for additional intramolecular interactions. Because of the heterogeneous nature of the disorder states, only the ensemble-averaged properties such as the residue helicities of free KID and pKID could be analyzed. The results failed to show significant difference upon phosphorylation of Ser133.[48] The principal role of phosphorylation has therefore been attributed to mediating intermolecular interactions. In the pKID/KIX complex, pSer133 mainly forms two salt-bridge interactions with Tyr658 and Lys662 of KIX. The essential role of phosphate-mediated interactions is further supported by mutagenesis of Tyr658 to phenylalanine or Lys662 to alanine, which substantially reduces the KID:KIX binding affinity, up to more than 1000-fold with double mutations (a destabilization of about 4.4 kcal/mol).[50] Nonetheless, the magnitude of the stabilizing effects of these salt-bridges, estimated to be $\Delta\Delta G = -RT \ln K_d(\text{pKID})/K_d(\text{KID}) \approx -5RT$ (or ~ -3 kcal/mol at 300 K) is somewhat surprising.[51,52] While the contribution of a salt-bridge interaction to the stability of proteins and protein complexes depends on its specific nature and environment (e.g., buried or solvent-exposed), many are often destabilizing or contribute only slightly to protein stability.[53,54] Curiously, replacement of pSer with negative charge residues such as glutamic acid (KID S133E) substantially weakens the KIX binding and failed to support target gene activation in vivo,[50,55] even though Glu or Asp is believed to have similar ability to interact with Lys, Arg, or amide NH.[56] Therefore, it is possible that phosphorylation of Ser133 might play an expanded role in the KID:KIX recognition besides providing an interaction site for intermolecular interactions. In particular, conformational equilibria of IDPs are highly susceptible to amino acid replacements, and it is not clear how the binding affinity of unphosphorylated KID is affected by the mutations.[57] These potentially nontrivial consequences besides deleting the targeted contacts can substantially complicate the interpretation of the affinity measurements.

In principle, absence of significant difference in mean secondary structure properties between KID and pKID on the ensemble level does not rule out the possibility that phosphorylation might lead to reorganization of the accessible conformational substates without changing the apparent average helicity. Such conformational redistribution might affect both thermodynamics and kinetics of the coupled binding and folding. In particular, compensation between conformational entropy and interfacial contact enthalpy plays an important role in target recognition by IDPs. Existing mutagenesis studies clearly demonstrate that conformational preferences have a significant role in binding of constitutive and inducible activators to KIX.[49,50,58] A previous explicit solvent MD simulation study further suggested that phosphorylation of Ser133 might promote formation of a transient turn structure that resemble the KIX bound state.[57] However, the reported simulations were conducted by using a simulated annealing protocol in explicit solvent and only lasted 2.0 ns. This is too short to sufficiently sample relevant conformations in the disordered states (also reflected in the confined backbone ϕ/Ψ distributions that were observed[57]), substantially weakening the reliability of the proposed effects of phosphorylation. The current REX-MD simulations in implicit solvent offer much more extensive sampling and should provide a more realistic description of the disordered ensembles to allow a better assessment of the putative impacts of phosphorylation. Compared with existing experimental data, the current study also provides an opportunity to examine the feasibility of utilizing REX-MD simulations in implicit solvent to characterize disordered states of small IDPs at atomic level in general.

Methods

Consistent Implicit Solvent Protein Force Field. A consistent GBSW implicit solvent force field[31,36] based on the CHARMM22/CMAP all-atom force field[59−62] was used in this work. GBSW is one of the latest GB models that employs a van der Waals (vdW)-based surface with a smooth dielectric boundary to allow stable force calculation. Born radii are calculated by a rapid volume integration scheme that includes a higher-order correction term to

(45) Parker, D.; Ferreri, K.; Nakajima, T.; Morte, V. J. L.; Evans, R.; Koerber, S. C.; Hoeger, C.; Montminy, M. R. *Mol. Cell. Biol.* **1996**, *16*, 694–703.
(46) Chrivia, J. C.; Kwok, R. P. S.; Lamb, N.; Hagiwara, M.; Montminy, M. R.; Goodman, R. H. *Nature* **1993**, *365*, 855–859.
(47) Radhakrishnan, I.; PerezAlvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741–752.
(48) Radhakrishnan, I.; Perez-Alvarado, G. C.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1998**, *430*, 317–322.
(49) Zor, T.; Mayr, B. M.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *J. Biol. Chem.* **2002**, *277*, 42241–42248.
(50) Parker, D.; Jhala, U. S.; Radhakrishnan, I.; Yaffe, M. B.; Reyes, C.; Shulman, A. I.; Cantley, L. C.; Wright, P. E.; Montminy, M. *Mol. Cell* **1998**, *2*, 353–359.
(51) Dal Peraro, M.; Alber, F.; Carloni, P. *Eur. Biophys. J. Biophys. Lett.* **2001**, *30*, 75–81.
(52) Mestas, S. P.; Lumb, K. J. *Nat. Struct. Biol.* **1999**, *6*, 613–614.
(53) Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K. *J. Phys. Chem. B* **1999**, *103*, 727–736.
(54) Yang, A.-S.; Honig, B. *Curr. Opin. Struct. Biol.* **1992**, *2*, 40–45.
(55) Shaywitz, A. J.; Dove, S. L.; Kornhauser, J. M.; Hochschild, A.; Greenberg, M. E. *Mol. Cell. Biol.* **2000**, *20*, 9409–9422.
(56) Mandell, D. J.; Chorny, I.; Groban, E. S.; Wong, S. E.; Levine, E.; Rapp, C. S.; Jacobson, M. P. *J. Am. Chem. Soc.* **2007**, *129*, 820–827.
(57) Solt, I.; Simon, I.; Tompa, P.; Fuxreiter, M. *FEBS J.* **2005**, *272*, 391–392.
(58) Parker, D.; Rivera, M.; Zor, T.; Henrion-Caude, A.; Radhakrishnan, I.; Kumar, A.; Shapiro, L. H.; Wright, P. E.; Montminy, M.; Brindle, P. K. *Mol. Cell. Biol.* **1999**, *19*, 5601–5607.
(59) Feig, M.; MacKerell, A. D.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 2831–2836.
(60) MacKerell, A. D. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(61) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
(62) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

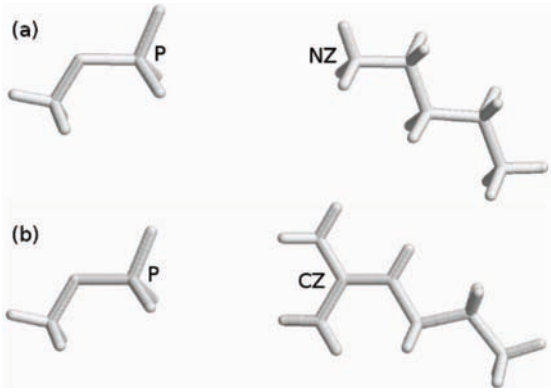

Figure 1. Molecular configurations of pairwise interactions of side-chain analogues: (a) pSer−Lys and (b) pSer−Arg.

the Coulomb field approximation.[63] Atomic input radii, the key physical parameters for defining the solvent boundary, have been carefully optimized together with peptide backbone torsion potentials to capture the delicate balance between solvation and intramolecular interactions.[36] The optimization also largely resolved the problem of systematically over stabilizing the hydrogen-bonding and salt-bridge interactions (in comparison to the explicit solvent results) that many GB models had suffered from.[37] We also note that, while parametrization of classical force fields traditionally relied primarily on experimental data and high-level quantum mechanics calculations of small molecules,[64] it only recently began to benefit more directly from experimental structural and thermodynamic properties of peptides and proteins. Through optimization, the consistent implicit solvent force field is able to predict both the structures *and* stabilities of a set of helical and β model peptides. In particular, stabilities of four sequentially related β-hairpins that range from largely unfolded to mostly folded at 300 K are quantitatively reproduced, indicating that the underlying interactions are properly balanced.

Potential of Mean Force (PMF) Calculations. Intrinsic radii for the phosphate group have been only been optimized on the basis of solvation free energies (r_p = 2.35 Å and r_o = 1.49 Å).[65] We previously showed that it is important to further examine the balance between intramolecular interactions and solvation, such as by comparing PMFs of representative pairwise interactions in explicit and implicit solvents.[36] An umbrella sampling protocol[36] was used to compute the PMFs between side chains of pSer and Arg or Lys in head-to-head configurations, shown in Figure 1. The solutes were constrained to move in a straight line along the reaction coordinate in specific orientations (see Figure 1) using the MMFP module in CHARMM.[66] The phosphate headgroup PO_4^{2-} was allowed to rotate freely along the O_γ−P bond, and the NH_3^+ group of Lys was allowed to rotate freely along the C_Z−N bond. For explicit solvent simulations, solutes were solvated by about 740 water molecules in periodic rectangular boxes. The particle mesh Ewald method was used to calculate long-range electrostatic interactions,[67] and the van der Waals interactions were smoothly switched off from 10 to 12 Å. Harmonic restraint potentials were placed every 0.5 Å along the reaction coordinate with a force constant of 5.0 kcal/mol/Å^2. For each umbrella sampling window, the system was first equilibrated for 60 ps, followed by 2 ns production at constant volume and temperature. The SHAKE algorithm[68] was applied to fix lengths of all hydrogen-involving bonds and the dynamics time step was 2 fs. Data were analyzed using the weighted histogram analysis method.[69] PMFs in implicit solvent were computed in a similar fashion, except that effects of solvent were described by GBSW. Convergence of the PMFs was examined by comparing results from the first and second halves of the data and was shown to be on the order of 0.2 kcal/mol. Additional implicit solvent PMFs were computed for collinear and coplanar dimer configurations previously studied by Mandell and co-workers[56] by directly translating the molecules along the reaction coordinates.

REX-MD Peptide Simulations. All REX-MD simulations were carried out using the Multiscale Modeling Tools in Structural Biology (MMTSB) Toolset[70] along with CHARMM. Briefly, multiple copies of the system were simulated simultaneously and independently at different temperatures. Exchanges of simulation temperatures are periodically attempted according to a Metropolis-type algorithm. In the course of the REX simulation, replicas can travel up and down the temperature space in a self-regularized fashion, which, in turn, induces a nontrivial random walk in temperature space and greatly reduces the probability of being trapped in states of local energy minima. In this work, a 28-residue segment of KID (residues 119−146, TD SQKRR EILSR RPSYR KILND LSSDA P), its phosphorylated form (pSer133) and a mutated sequence KID S133E were simulated. The termini were blocked with acetyl (Ace) and amine (NH_2) groups. For both KID and pKID, two independent runs were initiated from the folded (control simulations) and fully extended (folding simulations) conformations, respectively, to examine the convergence of the resulting structure ensembles. The folded conformations of both KID and pKID were derived from the pKID/KIX complex structure (PDB ID: 1kdx).[47] Only the folding simulation was carried out for KID S133E. In REX-MD, 12 replicas were placed in the range of 270−500 K, where the temperatures were distributed exponentially. Exchanges of simulation temperatures were attempted every 2 ps and the overall exchange ratios were about 0.3. The total length of each REX-MD simulation is 200 ns except for folding simulation of KID S133E, which only lasted 160 ns.

Data Analysis. Most analysis shown was based on structures sampled at 302 K, except that the residue helicities were computed at 285 K for direct comparison to the NMR secondary chemical shift analysis, which was carried out at 288 K.[48] On the basis of the convergence analysis (see Results and Discussion), structures sampled during the last 80 ns of the folding simulations were used to construct the disordered ensembles, each of which contains 40 000 snapshots. The secondary structure was analyzed based on hydrogen bonding patterns using the COOR SECS module in CHARMM. Contact maps were computed using MMTSB/contact.pl, where the side chains are considered to be in contact if the shortest distance between the heavy atoms is no greater than 4.2 Å. Structure ensembles were clustered both using a hierarchical clustering algorithm based on mutual C_α root-mean-square-deviation (rmsd) (as implemented in MMTSB/cluster.pl) and by sorting the structures based on the helical substates (defined by the helix starting position and length). Given the extensive conformational sampling provided by very long REX-MD simulations, we directly estimated the backbone conformational entropy of a given residue from the 2D ϕ/Ψ dihedral angle distributions using the Shannon definition of entropy[71] as $S = -R\sum p_i \ln p_i$, where R is the ideal gas constant and p_i is the fractional probability of bin i.

The backbone conformational entropy of a short loop segment can then be estimated by summing up single-residue entropies, with

(63) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606–10614.
(64) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
(65) Banavali, N. K.; Roux, B. *J. Phys. Chem. B* **2002**, *106*, 11026–11035.
(66) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(67) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(69) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(70) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377–395.
(71) Shannon, C. E. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.

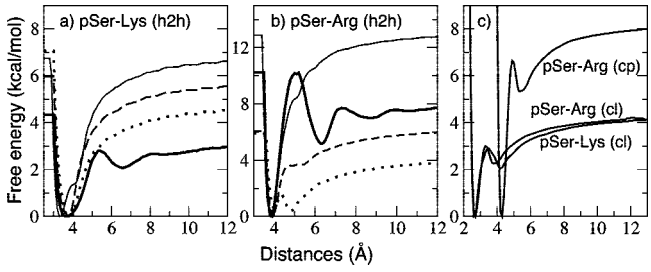

Figure 2. PMFs of pairwise interactions between pSer and Lys or Arg side-chain analogues in TIP3P and GBSW. In panels (a) and (b), thick lines plot the TIP3P results of head-to-head (h2h) interactions (see Figure 1) and the other traces plot the GBSW PMFs with various combination of input radii, (r_p, r_o) = (2.35, 1.49 Å) (thin lines), (2.35, 1.40 Å) (dashed lines), and (2.0, 1.4 Å) (dotted lines). Panel (c) plots the PMFs in GBSW with the optimized phosphate radii (r_p = 2.0 Å and r_o = 1.4 Å) for interactions in colinear (cl) and coplanar (cp) configurations. Corresponding interaction strengths in explicit solvent are (in kcal/mol) −10.6 (pSer−Arg, cp), −4.7 (pSer−Arg, cl), and −4.2 (pSer−Lys, cl).[66]

the assumption that backbone conformational equilibria are essentially uncoupled for residues in a disordered loop. We also considered the coupling of neighboring residues and estimated two-residue backbone conformational entropies from 4D backbone torsion distributions. Both approaches were used to analyze the loop entropies of KID and pKID and the results were very similar. In addition, the dependence of calculated entropy on the bin size was examined by repeating the calculations with 36°, 30°, and 15° bins in all dihedral angle dimensions. All structural visualizations presented in this work were prepared with the VMD program.[72]

Results and Discussion

Optimized Input Radii for the Phosphate Group. PMFs of the pairwise interactions between pSer and Lys or Arg side chain analogs in the head-to-head configurations (see Figure 1) are summarized in Figure 2. We will focus mainly on examining the interaction strength as defined by the relative free energy minimum. Note that desolvation peaks are not present in the GBSW implicit solvent due to the nature of the underlying vdW-like solute surface. This might alter the folding kinetics (such as speeding up conformational diffusion) but should not lead to systematic bias in the equilibrium properties that we examine in this work. Clearly, the previously optimized radii (r_p = 2.35 Å and r_o = 1.49 Å) significantly overstabilized both salt-bridge interactions compared with the explicit solvent results. By systematically adjusting both radii, agreement with the explicit solvent results can be improved. An optimal choice was found to be r_p = 2.0 Å and r_o = 1.4 Å, which still overestimates the stability of pSer−Lys head-to-head interaction while underestimating that of pSer−Arg head-to-head interaction. The inability of GBSW to reproduce stabilities of both pairs in explicit solvent might reflect some intrinsic limitations of implicit solvent, as granularity of solvent and local ordering of water molecules might play a role in interactions of heavily charged side chains. Another possible source of discrepancies is different preferred rotation of PO_4^{2-} (and Lys NH_3^+) groups in GBSW and TIP3P solvents. As such, we computed additional PMFs in GBSW for three rigid colinear and coplanar configurations. Shown in Figure 2c, the results seem to agree very well with previous explicit solvent calculations by Mandell and co-workers,[56] indicating that the optimized radii capture the balance between solvation and intramolecular interactions quite well.

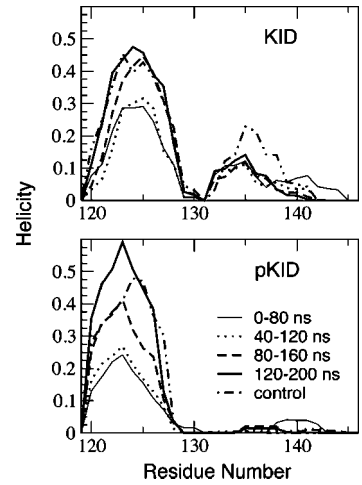

Figure 3. Residue helicity of KID (top panel) and pKID (bottom panel) at 285 K, calculated from 200 ns folding REX-MD simulations using various time intervals. The control traces shown are residue helicities computed from the last 80 ns of 200 ns control REX-MD simulations initiated from folded NMR strutures.

Convergence of the Structural Ensembles. Convergence of the REX-MD simulations was examined both by studying the dependence of structural properties on the simulation length and by comparing results from independent folding and control simulations initiated from fully extended and folded initial conformations, respectively. The ensemble-averaged residue helicities at 285 K, shown in Figure 3, start to stabilize during the second half of the 200 ns folding simulations (e.g., comparing traces computed from the 80−160 and 120−200 ns segments). Importantly, mean residue helicities computed from the last 80 ns agree very well between folding and control simulations that were inititiated from completely different initial conformations, providing a reliable confirmation of the overall convergence. It appears that the convergence of pKID simulations was poorer compared to that of KID, which is likely related to the more compact and rigid nature of free pKID (see below). Most analysis in the following will be based on the structural ensembles near the room temperature (302 K). Convergence of these structural ensembles was further examined by analyzing the nature of conformatioanl substates sampled. Such analysis provides a stringent test of the convergence of a structure ensemble.[73] For helical proteins like KID and pKID, conformational substates can be effectively identified by analyzing the distributions of helical substates, defined by the locations and lengths of parital helices in the disordered ensembles. As shown in Figure S1 (see Supporting Information), similar helical substates were sampled in both control and folding simulations and with comparable probabilities. Examination of the substate distribution as a function of simulation length (see Figures 5 and S1) indicates that the folding simulations are particularly well converged, both in the nature of major helical substates and in their relative occupancies in the disordered ensembles. Accordingly, the following analysis is based mainly on structural ensembles derived from the last 80 ns of the folding simulations.

Comparison of Ensemble-Averaged Properties with Experimental Results. Disordered states of proteins are largely out of reach of conventional high-resolution structural characterization and only certain ensemble-averaged structural properties, mainly including the radius of gyration and residual

(72) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–&.

(73) Lyman, E.; Zuckerman, D. M. *Biophys. J.* **2006**, *91*, 164–172.

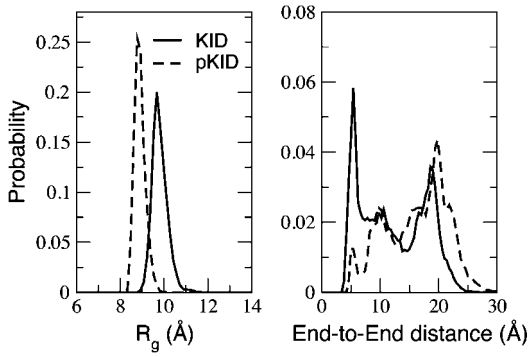

Figure 4. Probability distributions of (a) radius of gyration and (b) end-to-end distance at 302 K. End-to-end distance is defined as the distance between C_α atoms of two termini.

secondary structures, can be determined from experiments. At present, only mean residue helicities of free KID and pKID have been estimated on the basis of NMR secondary chemical shift analysis.[48] Defined as observed chemical shifts less the corresponding random coil values, secondary chemical shifts of certain backbone nuclei, especially those of H_α and C_α, depend mainly (but not only) on the backbone ϕ/Ψ dihedral angles,[74] with up to 0.80 correlation to the helical content.[75,76] Assuming that both helices αA and αB are fully folded upon binding to KIX, the helical population of free KID was estimated from ratio of secondary chemical shifts of free peptides and folded peptides in complex to be 50−60% for αA and about 10% for αB at 288 K.[48] Upon phosphorylation, the helical content of αA remains similar, while that of αB increases slightly to about 15%. The helicity increase was further supported by observation of several short-range NOEs.[48] In comparison, the average helicities calculated from the simulated ensembles at 285 K are 37% ± 2% and 5% ± 5% for helix αA and αB of KID, 38% ± 4% and 1% ± 1% for helix αA and αB of pKID, respectively. The reported uncertainties were estimated from differences between values computed from the last 80 ns of the control and folding simulations. These values are in good agreement with the NMR results, even though the helical population of αB of pKID seems to be underestimated. Inability of the current simulation to predict a slight increase in the αB helical population most likely reflects some limitations of the GBSW implicit solvent force field, possibly in the balance of charge−charge and nonpolar interactions.[41] On the other hand, significant uncertainties in the quantitative interpretation of NMR secondary chemical shifts should not be neglected.[75,76] We note that putatively the most functionally relevant conformational changes upon phosphorylation occur in the loop segment and the inability to predict the slight increase in the αB helicity does not appear to be critical for the analysis that follows. We also want to stress that the ability of well-converged de novo simulations to semiquantitatively reproduce these ensemble properties on residue level is an important, albeit insufficient, indication that the simulated ensembles are likely realistic. We have also examined the distributions of radius of gyration (R_g) and end-to-end distance, shown in Figure 4. The results indicate that both KID and pKID are quite compact, with an average R_g of 9.8 Å for KID and 8.9 Å for pKID. One might be tempted to interpret the narrow R_g distribution of pKID as it being substantially more structured. However, the distributions of end-to-end distance indicate that conformational ensembles of pKID and KID are similarly diverse. Experimental data on the radius of gyration of either KID or pKID are not available at present and these predictions will need to be validated by future measurements.

Helical Substates of the Disordered KID and pKID. Availability of the disordered structure ensembles at atomic level allows for in-depth characterization of the conformational substates to understand putative conformational consequences of phosphorylation that might be obscured on the ensemble level. Figure 5 plots the 2D probability distributions of the helical substates sampled during the last 80 ns of the folding simulations. Clearly, both KID and pKID sample a similar, if not common, set of helical substates, which is reflected in a resemblance of the overall shapes of the two distributions. The most dominant helical substates for both KID and pKID are essentially full-length helix αA (KID substates (a) and (b) and pKID substate (a), see Figure 6). Interestingly, the packing of disordered helix αB segment (residues 134−144) is also quite similar for KID and pKID (Figures 6, substate a of KID and pKID). However, KID appears to sample a more diverse set of helical substates compared to pKID. There are four major helical substates within αA in the KID ensemble, defined by the 0.01 contours, whereas there are only three such states for pKID αA. Representative snapshots of these major substates and their occupancies are shown in Figure 6. Partial helices in the pKID ensemble appear to have an increased tendency to begin at "native-like" locations (as in the KID/KIX complex). For example, about 31% of the partial helices begin at Asp120 or Ser121 in pKID, compared with only about 22% in KID. Interestingly, even though the current simulations appear to underestimate the helical content in αB segment, the results indicate that helix αB predominantly extends from Tyr134 to Ser143 in pKID whereas it extends from Arg131 to Asp144 on average in KID. Clearly, phosphorylation of Ser133 leads to substantial redistributions of the underlying helical substates in the disordered ensembles, despite having very little effects on the overall helicity in αA and small effects in αB on either peptide or residue level. More importantly, the nature of the conformational redistribution indicates that phosphorylation might promote a small number of "native-like" helical conformations, albeit at a very low level, even before encountering KIX.

Automatic clustering is routinely applied to identify important conformational substates, such as in protein structure prediction and refinement.[77,78] In principle, greater care needs to be taken in clustering disordered structure ensembles, particularly in choosing the most proper metrics for measuring pairwise distances between conformations. Such metrics need to provide a robust measurement of essential differences between structures in the presence of large noise that arises due to the heterogeneous nature of a disordered state. Mutual C_α rmsd commonly used in clustering analysis might not be a good distance metric for analyzing a disordered structure ensemble. For example, a hierarchical clustering algorithm implemented in MMTSB/cluter.pl was applied to cluster the simulated ensembles based on mutual C_α rmsd values. Resulting clusters are shown in Figure S2 (Supporting Information). While these clusters do

(74) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436–9444.
(75) Mielke, S. P.; Krishnan, V. V. *J. Biomol. NMR* **2004**, *30*, 143–153.
(76) Sibley, A. B.; Cosman, M.; Krishnan, V. V. *Biophys. J.* **2003**, *84*, 1223–1227.
(77) Zhang, Y.; Skolnick, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7594–7599.
(78) Bradley, P.; Misura, K. M. S.; Baker, D. *Science* **2005**, *309*, 1868–1871.

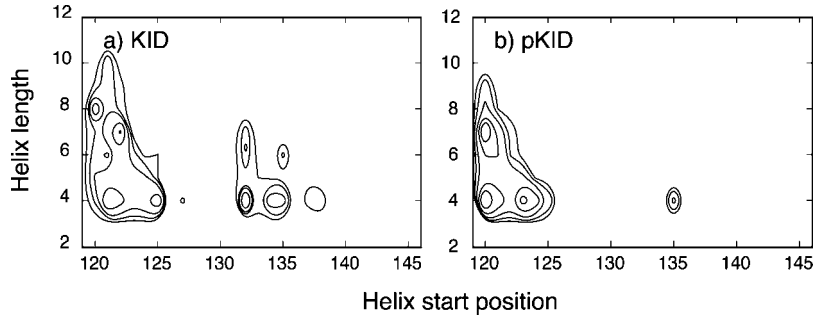

Figure 5. Contour plots of the 2D probability distributions of helix starting position and length for (a) KID and (b) pKID at 302 K form the last 80 ns of the folding simulations. Contours are drawn at 0.002, 0.005, 0.01, 0.02, 0.04, and 0.06 levels.

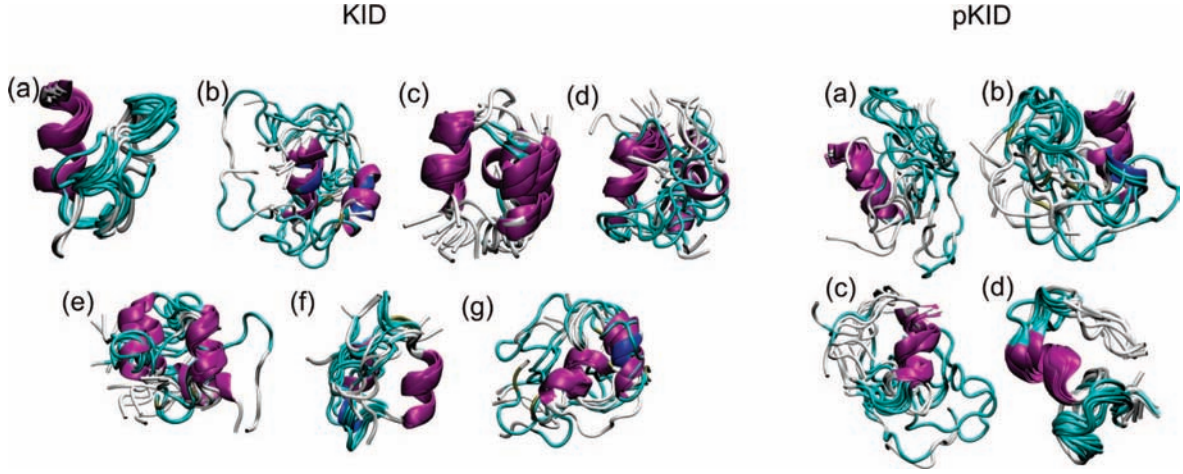

Figure 6. Randomly selected snapshots of major helical substates of the disordered structure ensembles of KID and pKID at 302 K. All snapshots are aligned using the backbone atoms of the common helical segment specific to a particular helical substate. The helical substate is identified by the helix starting position (the first number) and length (the second number). For KID: (a) 120,8 (9%) (b) 121,4 (13%) (c) 122,7 (8%) (d) 125,4 (6%) (e) 132,4 (5%) (f) 132,6 (2%) (g) 134,4 (3%); for pKID: (a) 120,7 (17%) (b) 120,4 (14%) (c) 123,4 (13%) (d) 135,4 (1%). These substates correspond to the centers of the peaks in the 2D probability distributions in Figure 5, and the occupancies provided in parenthesis include contributions from nearby helical substates within the same peak (defined by the 0.01 contours). Note that conformations with multiple partial helices might contribute to more than one substate.

have some resemblance to the helical substates identified above (see Figure 6), the clustering is clearly more sensitive to the packing of disordered helix αB segment with partial helices in helix αA segment and the resulting clusters contain mixtures of various helical substates. Automatic clustering was also attempted on the basis of contact maps, and the results were similar. We argue that it might be more relevant to identify accessible conformational substates by directly analyzing the nature of partial helices. This exercise illustrates that one needs to carefully identify the most (functionally) relevant metrics for clustering a disordered structure ensemble.

Transient Tertiary and Short-Range Contacts. Ensemble-averaged contact maps were computed to identify possible persistent tertiary contacts that might be present in the disordered ensembles. Summarized in Figure 7, the results show several similarities and differences in tertiary contacts present in the disordered states of KID and pKID. First, charged residues appear to play a dominant role in transient packing of largely disordered αB segment and partial helices of αA segment for both KID and pKID. This is consistent with general recognitions that IDPs typically lack hydrophobic cores important for defining the structures of folded proteins and that hydrophobic residues of IDPs tend to play roles in intermolecular interactions. Among the contacts observed Arg124 and Arg125 are most frequently involved, mainly by interacting with Tyr134 and Ser142 (and pSer133 in pKID). Ser121 is also involved by making contacts with Arg135 and Lys136. Second, effects of phosphorylation

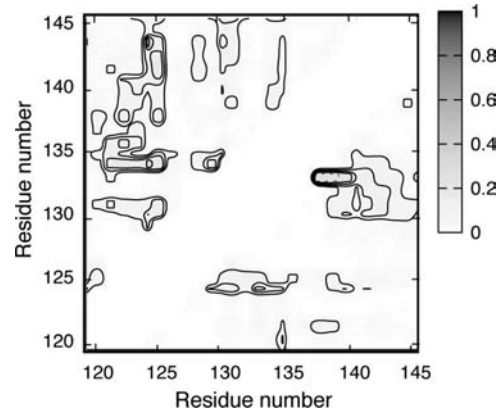

Figure 7. Residue−residue tertiary contacts between residues separated by more than four residues of KID (upper triangle) and pKID (lower triangle) at 302 K. These contact maps were calculated from the last 80 ns of the folding simulations. Contours are drawn at 0.05, 0.1, 0.2, 0.4, and 0.8 levels.

on tertiary contact pattern appear to be localized in the loop region (residues 130−135), where pSer133 makes contacts with the nearby charged residues such as arginines and lysines, mainly Arg130, Arg135 and Lys136. These are short-range contacts and thus not shown in the tertiary contact map. In addition, contacts induced by pSer133 also appear to enhance

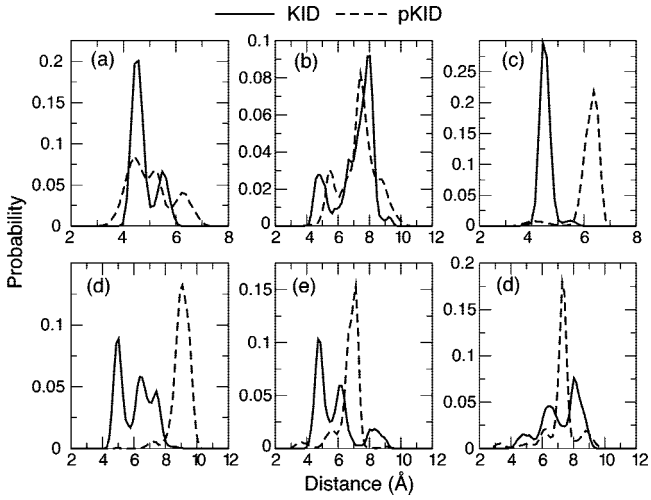

Figure 8. Probability distributions of interproton distances calculated from the last 80 ns of folding trajectories of KID and pKID at 285 K. (a) $I127H_\alpha-S129H_N$, (b) $I127H_\alpha-R130H_N$, (c) $R131H_\alpha-S133H_N$, (d) $R131H_\alpha-Y134H_N$, (e) $Y134H_\alpha-I137H_N$, (f) $I137H_\alpha-D140H_N$. Theoretical NOE distances calculated with r^{-6} averaging are (in Å): (a) 4.4/4.6, (b) 4.8/6.6, (c) 4.2/5.8, (d) 5.4/8.8, (e) 3.7/5.7, and (f) 5.0/5.5 for KID and pKID, respectively.

the packing of C-terminal tail with the loop region, which appears to exist with a much lower probability for KID.

We further analyzed the mean solvent-accessible surface areas (SASA) of all side chains, shown in Figure S3 (Supporting Information). Consistent with the highly disordered nature of both KID and pKID on tertiary level, there are only a few residues with substantial changes on SASA (>25 $Å^2$). These residues are mostly polar except Leu128 and Pro132. Interestingly, interfacial residues in the loop region (Leu128 and Tyr134) become much more exposed in pKID; other (hydrophobic) inartificial residues are similarly exposed in KID and pKID, including Leu141 previously predicted to be a key residue in KIX binding.[18] Arg131 was previously proposed to be in contact with pSer133 to stabilize complex-like loop conformations.[57] However, current simulations show that it is not frequently in contact with pSer133 and actually becomes lightly more exposed on average. Increased solvent exposure of Arg131 could support its putative role in electrostatic steering when forming the encounter complexes with KIX, which was proposed based on the observation of the larger on rate for Arg131 in a recent NMR relaxation study.[17]

Six weak or very weak NOEs were resolved for pKID that were not observed for KID.[48] In Figure 8, we compare probability distributions of the distances between all six corresponding pairs of protons. Theoretical NOE distances computed as $r = (1/N\sum_i r_i^{-6})^{-1/6}$ are mostly on the order of 6 Å for pKID, which could give rise to the observed weak NOE signals. However, five of these six pairs of protons have shorter theoretical NOE distances for KID, which is inconsistent with the absence of corresponding NOE peaks in KID spectra. Closer inspection shows that all six proton pairs involve either i,$(i + 2)$ or i,$(i + 3)$ contacts, which are often associated with helix formation. Therefore, inability to predict that these NOE signals should be observed for pKID but not for KID is actually recapitulating the observation of lower αB helical content for pKID from the current simulations (see Figure 3). Observation of more short interproton distances in the simulated structure ensembles might also reflect a systematic bias toward compact structures of the implicit solvent protein force field currently employed.[41] Such bias arises both from remaining limitations in the implicit treatment of nonpolar solvation and from various imperfections in the underlying protein force field. On the other hand, we note that KID appears to sample a more heterogeneous set of helical substates and that the associated conformational exchanges might lead to substantial line broadening, preventing one from resolving these potential NOE signals.

Loop Entropy. Analysis of the conformational substates, tertiary contact patterns, and side-chain solvent exposure prompted us to further examine possible consequences of phosphorylation on loop flexibility. Rigidification of the loop in pKID was also proposed previously by Solt et al.[57] Given the apparent good convergence of the simulated ensembles, the loop entropy was directly estimated from the backbone dihedral angle probability distributions of the loop residues. Helix αB appears to start from Arg135 instead of Tyr134 for pKID in the disordered states, and we thus expand the definition of the loop region to include residues 130 through 135. The results are summarized in Table 1. Both the total loop entropies and the difference between pKID and KID have a slight dependence on the bin size of the probability distribution. Furthermore, two-residue backbone conformational entropy estimated from the 4D joint backbone dihedral angle distributions are always smaller than those estimated from summation of single-residue distributions, suggesting that backbone conformation of neighboring residues are coupled. The single-residue-based estimation appears to systematically underestimate the entropy change upon phosphorylation. Clearly, there is a significant reduction in the loop flexibility upon phosphorylation of KID Ser133. The associated entropy change is estimated to be $\Delta S/R \approx -1.9 \pm 0.4$ (from the average and standard deviation of $\Delta S^{(2)}/R$ computed with three bin sizes). This entropic contribution is significant, which could account for ~40% of the $\sim -5RT$ total free energy of stabilization upon phosphorylation assuming that bound conformations of KID and pKID are similar.[49] This contribution might complement the salt-bridge interactions between pSer133 and KIX Tyr658 and Lys662, but it was not previously recognized due to a lack of experimentally detectable conformational changes upon phosphorylation on the ensemble level. Several important limitations of the above estimation of loop entropy contribution to binding need to be emphasized. In principle, one needs to consider the whole-protein conformational entropies of pKID and KID in both bound and unbound states for a rigorous estimation of the conformational entropy's contribution to binding. However, this is largely out of reach of the current state-of-the-art simulation methodologies, limited by both the convergence issues and the force field accuracy. Simply inferring the role of conformational entropy in coupled folding and binding from the loop region alone appears to be a necessary compromise. The NMR experiments have also demonstrated that observable structural impacts of phosphorylation were largely localized around the loop region.[48] We argue that simply analyzing the loop conformational changes can provide meaningful insights into how phosphorylation-induced conformational changes might play a role in coupled folding and binding of pKID to KIX. We further note that entropy cost of folding generally plays an important role in balancing the enthalpy gain of binding to achieve low affinity simultaneously with high specificity in IDP recognitions. As such, it is feasible that phosphorylation might modulate the binding thermodynamics readily through controlling the conformational entropy.

2D backbone dihedral angle distributions of key loop residues are shown in Figure 9. Those of Arg130 and Arg131 are very

Table 1. Loop Entropies Estimated from Backbone Dihedral Angle Distributions at 302 K[a]

	$\Sigma_{i=130,135} S_i^{(1)}/R$			$\Sigma_{i=65,67} S_{2i,2i+1}^{(2)}/R$			$\Delta S^{(1)}/R$		$\Delta S^{(2)}/R$	
bin size	KID	pKID	S133E	KID	pKID	S133E	pKID	S133E	pKID	S133E
15°	53.8	52.7	54.3	52.2	50.7	52.8	−1.1	+0.5	−1.5	+0.6
30°	54.8	53.6	55.5	53.2	51.2	54.0	−1.2	+0.7	−2.0	+0.8
36°	55.4	53.8	56.2	53.7	51.4	54.6	−1.6	+0.8	−2.3	+0.9

[a] R, ideal gas constant; $S^{(1)}$, single-residue backbone entropy; $S^{(2)}$, double-residue backbone entropy; ΔS, total loop entropy change upon phosphorylation or S133E replacement.

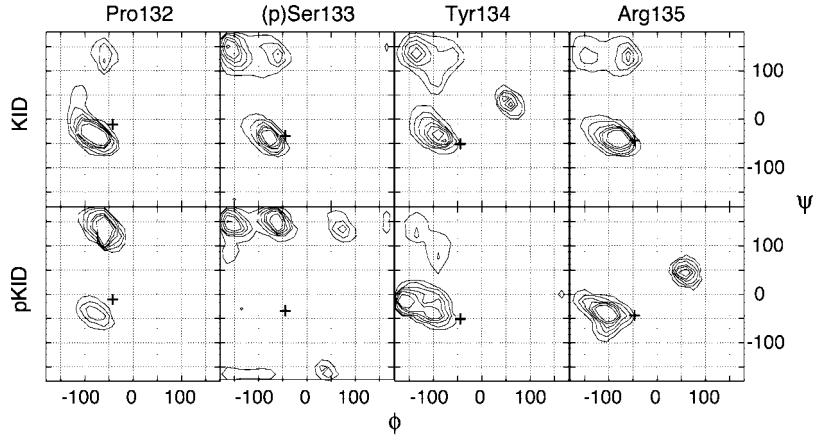

Figure 9. 2D backbone dihedral angle distributions of selected loop residues at 302 K with corresponding backbone torsion angles in the pKID/KIX complex marked with "+". Distributions were computed on 24 × 24 grids, and contours were drawn at 0.001, 0.004, 0.01, 0.02, 0.03, and 0.05 levels.

similar between pKID and KID and thus not shown. Residues Pro132, Ser133, and Tyr134 have the largest flexibility changes, each contributing $\sim -0.4R$ entropy reduction upon phosphorylation. In contrast to what was previously suggested by Solt et al.,[57] the loop does not appear to predominantly adopt "native-like" conformations (as in the KID/KIX complex). In absence of KIX, compact (disordered) conformations dominate (e.g., see Figures 4 and 6), and the probability of observing "native-like" loop conformations, which would lead to much less compact conformations, should be small. The previous simulations were initiated from folded structures in explicit solvent and only lasted 2 ns. The amount of sampling was thus very limited and could result in a strong bias toward the initial folded conformations. We argue that structural ensembles obtained from the current simulations are likely more realistic due to the improved force field and extensive sampling. On the other hand, we also recognize that virtually all modern protein force fields (implicit or explicit) do tend to overstabilize protein−protein interactions and lead to overly compacted structures.[28,41] Nonetheless, many experimental studies have demonstrated that water is poor solvent for peptides and even highly denatured states of proteins are generally compact.[25] Future experimental characterization of the hydrodynamic radii might provide direct evidence regarding the degree of compactness of the disordered states of KID and pKID.

Inability of S133E to Induce Significant Conformational Changes. To further validate that observed pSer133-induced conformational changes in the disorder state of KID are functionally relevant, we simulated a mutated sequence of KID where Ser133 was replaced by a glutamic acid. KID S133E has been shown to bind KIX at <25% strength compared to pKID,[50,55] even though Glu can interact with Lys, Arg, or backbone amide with similar strengths compared to the phosphate group.[56] Summarized in Figure 10, the results suggest that the disordered ensemble of KID S133E is very similar to that of KID in terms of both residue helicities and helical

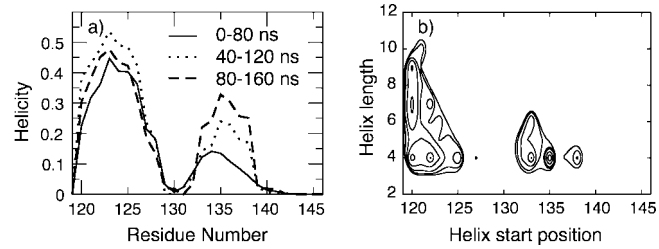

Figure 10. (a) Residue helicities of KID S133E at 285 K calculated from various segments of the folding REX-MD simulation and (b) the 2D probability distribution of helix starting position and length at 302 K computed from the last 80 ns of the same simulation. Contours are drawn at 0.002, 0.005, 0.01, 0.02, 0.04, and 0.06 levels.

substates. The averaged helix populations of helix αA and αB are 36% ± 6% and 14% ± 4%, respectively. The uncertainties were estimated from differences between results computed from 40−120 and 80−160 ns segments. KID S133E appears to sample a similarly diverse set of helical substates compared to KID (see Figure 5a). Furthermore, the loop entropy estimated from backbone dihedral distributions suggests that the loop of KID S133E is somewhat more flexible than that of KID, leading to a slight increase of loop entropy of $\sim 0.8 \pm 0.2R$ (see Table 1). Inability of S133E to induce changes in helical substates and reduction in loop flexibility similar to those observed with pSer133 could explain why binding of KID S133E to KIX is not nearly as strong as that of pKID. These results further support that conformational changes as a result of phosphorylation might play a relevant role in coupled folding and binding of KID to KIX.

Concluding Discussion

We characterized the disordered states of KID and pKID at atomic level based on extensive REX-MD simulations in a consistent implicit solvent protein force field previously opti-

mized to provide a realistic description of peptide conformational equilibria. Examination of the dependence of ensemble-averaged structural properties and conformational substates on simulation length and initial structures indicates that the simulations are very well converged. Calculated mean residue helicities are in good agreement with the NMR secondary chemical shift analysis. The simulation also shows that helicity of αA is near 50% and does not change upon phosphorylation of Ser133 and that of αB is much smaller, on the 10% level, and is more significantly affected by phosphorylation of Ser133. Both predictions agree with the NMR results, indicating that simulated structure ensembles of free KID and pKID might be realistic. However, the current simulations fail to reproduce the slight increase in the helical population of αB and the simulated structure ensembles might be too compact. These limitations are likely due to several known drawbacks in the current implicit solvent protein force fields in ref 41. Nonetheless, the ability of well-converged de novo simulations to "predict" the important features of the disordered states of a 28-residue peptide is nontrivial and represents a very encouraging progress toward accurate simulation of IDP conformational equilibria.

In-depth analysis has been carried out to identify the major conformational substates in the disordered ensemble and associated properties such as transient tertiary contacts, solvent exposure of side chains and loop conformational entropies. Detailed pictures of free KID and pKID emerging from such analysis are generally inline with previous NMR characterizations. The simulations further reveal that, while being highly disordered on the tertiary level, both KID and pKID appear to sample a limited number of compact helical substates. Comparison of the nature of these accessible substates of KID and pKID points to an expanded role of phosphorylation in the KID:KIX recognition. In addition to providing a new site for intermolecular interactions, phosphorylation of Ser133 also leads to a nontrivial redistribution of the helical substates within the conformational ensembles. Importantly, such redistribution occurs without leading to significant changes of the average secondary structure propensities on the ensemble level and thus has escaped experimental detection. In particular, there appears to be a substantial reduction in the flexibility of the loop connecting two helical segments, which could provide an entropic contribution of about $-1.9R$ toward the stability of the pKID/KIX complex. We argue that such an expanded role of phosphorylation in the KID:KIX recognition is probably more consistent with the general mechanism where reversible phosphorylation regulates protein function by modulating local and global conformational preferences.[79,80] Interestingly, simulation of KID S133E mutant demonstrates that replacing pSer133 with Glu fails to induce similar conformational changes to those observed with pKID, providing an explanation to why binding of KID S133E to KIX is not nearly as strong as that of pKID.[50,55]

The consequences of the observed conformational changes upon phosphorylation in binding kinetics are not obvious. While phosphorylation appears to promote a smaller number of helical substates that are similar to folded helices in the KID/KIX complex, the dominant loop conformations are dissimilar from those in the complex. On one hand, one might expect an elevated kinetic barrier for binding-induced folding, as the peptide needs to break the non-native contacts to arrive at the folded structure. On the other hand, the conformational consequence of phosphorylation appears to be largely localized near the loop region. Both KID and pKID are highly unstructured on the tertiary level and the exposed surface areas of key interfacial hydrophobic residues are essentially unchanged in the helix αB region that has been identified as a key area involved in the rate-limiting step of forming the intermediate by recent NMR relaxation, as well as coarse-grained computer modeling studies.[17,18] Absence of changes in solvent exposure of these key residues upon phosphorylation suggests that formation of the encounter complexes and subsequent evolution to the intermediate might not be affected. Interestingly, we observed substantial increases in solvent exposures of a few residues in the loop region (Leu128, Arg131, and Tyr134), which are either directly involved in forming intermolecular contacts (Leu128 and Tyr134) or suggested to play a role in electrostatic steering for forming the encounter complexes (Arg131). These observations suggest that phosphorylation might also play a role in enhancing the binding kinetics.

Acknowledgment. This work was supported by an Innovative Research Award from the Terry C. Johnson Center for Basic Cancer Research and institutional funds from Kansas State University. This work is contribution 09-163-J from the Kansas Agricultural Experiment Station.

Supporting Information Available: Complete ref 56 and four figures showing the convergence of the simulations on the helical substates, automatic clustering of the pKID structure ensemble, solvent exposure of side chains of KID and pKID and correlation of two-residue and single-residue entropies. This material is available free of charge via the Internet at http://pubs.acs.org.

JA808999M

(79) Stock, A. M.; Robinson, V. L.; Goudreau, P. N. *Annu. Rev. Biochem.* **2000**, *69*, 183–215.
(80) Johnson, L. N.; Lewis, R. J. *Chem. Rev.* **2001**, *101*, 2209–2242.