JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

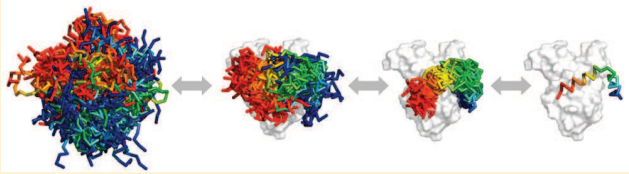
Mechanism of Folding and Binding of an Intrinsically Disordered Protein As Revealed by ab Initio Simulations

Mateusz Kurcinski, Andrzej Kolinski, and Sebastian Kmiecik*

Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

Supporting Information

ABSTRACT: A complex of the phosphorylated kinase-inducible domain (pKID) with its interacting domain (KIX) is a model system for studies of mechanisms by which intrinsically unfolded proteins perform their functions. These mechanisms are not fully understood. Using an efficient coarse-grained model, ab initio simulations were performed of the coupled folding and binding of the pKID to the KIX. The simulations start from an unbound, randomly positioned and disordered pKID structure. During the simulations the pKID chain and its position remain completely unrestricted, while the KIX backbone is limited to near-native fluctuations. Ab initio simulations of such large-scale conformational transitions, unaffected by any knowledge about the bound pKID structure, remain inaccessible to classical simulations. Our simulations recover an ensemble of transient encounter complexes in good agreement with experimental results. We find that a key folding and binding step is linked to the formation of weak native interactions between a preformed nativelike fragment of a pKID helix and KIX surface. Once that nucleus forms, the pKID chain may condense from a largely disordered encounter ensemble to a natively bound and ordered conformation. The observed mechanism is reminiscent of a nucleation−condensation model, a common scenario for folding of globular proteins.



INTRODUCTION

Proteins are dynamic molecules and may sample a huge ensemble of sometimes vastly different conformations, while participating in molecular recognition events. Over the past 50 years, the understanding of molecular recognition processes has been dominated by the "induced fit" hypothesis.[1] According to this hypothesis it is the binding interaction that drives a protein from an unbounded to a bounded conformation. In contrast, a number of recent experimental studies suggest a "conformational selection" mechanism postulating that all protein conformations preexist in the ensemble sampled by the free protein in the absence of the binding partner.[2] Noteworthy, binding mechanisms can be very complex, and the distinction between "conformational selection" and "induced fit" may be blurred.[3] For instance, the mechanistic path may depend on many factors (e.g., the mechanism of ligand binding changes with varied ligand and protein concentrations[3]) and thus can be accessed using detailed knowledge of the kinetics and thermodynamics of the reaction mechanism.

Many regulatory proteins are conformationally disordered under physiological conditions and form ordered structures only upon target binding.[4−6] Intrinsically disordered proteins (IDPs) remain difficult to study, both by experiment[7] and by simulation,[8] due to their high heterogeneity. The recent review of known disorder-based complexes suggests that the general model for the interaction of IDPs with their partners should be considered as a two-stage process,[9] with the first stage leading to the formation of disordered complexes and a slower second stage resulting in the formation of ordered associations. Based on theoretical analysis, it has been proposed that folding and binding processes may be enhanced by the "fly casting" mechanism.[10] In this scenario, the disordered protein binds weakly at a relatively large distance to its target and folds while approaching the binding site. It has been hypothesized that a disordered protein can have a greater capture radius for a native binding site; therefore the binding rate can be significantly enhanced in comparison with the fully folded protein.[10]

Molecular simulations of IDPs are challenging or impossible to perform using contemporary simulation techniques, especially if they are carried out in an ab initio fashion (no experimental information about the simulated system is used). The challenge can be characterized as efficient treatment of substantial conformational changes combined with a sufficiently accurate model to capture the relevant properties of the system.[8] This problem becomes even more complex when IDP binding simulation is sought. As highlighted in the recent reviews of the results of the critical assessment of predicted interactions (CAPRI),[11,12] a community-wide experiment aimed to assess the state of the art in protein−protein docking, the major unsolved problem in the field is the docking of proteins with a substantial backbone conformation change. These challenges can be addressed using efficient sampling strategies combined with coarse-grained models.[13]

Most (or perhaps all) of the published simulations of coupled folding and binding of pKID-KIX used simulation models in which pKID was biased toward the nativelike state. This bias is usually introduced either by the use of natively biased force

Received: January 26, 2014

   dx.doi.org/10.1021/ct500287c | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

field[14−19] or by starting simulations from (and sampling limited to) the native complex arrangement.[20,21] In this work, we go a step further by using a coarse-grained simulation model in an ab initio (unbiased) fashion (i.e., with full flexibility of pKID allowed and no experimental info on the pKID position and conformation used). In our simulations, we observe pKID binding to native and to nonnative sites on the KIX surface, which agrees with experimental findings.[22] To describe the mechanism of native binding, we analyze in detail trajectories which ended up in a near-native arrangement. Our analysis focuses on the transient encounter complexes involving barely bound and highly disordered ensembles and a multitude of nonnative interactions on the path to native binding. The proposed simulation and analysis scheme can be applied to other systems with available experimental clues on residues involved in the binding interface.

MATERIALS AND METHODS

Simulation Model. CABS is a coarse-grained protein model, extensively tested in protein structure prediction[23−25] and simulations of protein dynamics: folding mechanisms[26−30] and near-native fluctuations.[31,32] CABS has also been used for flexible protein−protein docking[33,34] and, in conjunction with other computational tools and experimental data, for the assembly of the first realistic model of the human telomerase enzyme.[35] The flexible docking tests included a study of the binding mechanism of the peptide TRAP220 coactivator to the Retinoid X Receptor[34] in which the modeled system was treated in a manner similar to that of pKID-KIX in the present work. Namely, during the docking simulation, the peptide coactivator chain (11 amino acids in length) was treated as fully flexible and allowed for unbiased detection for the binding site, while the receptor structure (238 amino acids in length) was restricted to near-native fluctuations (see movie S1 in the Supporting Information showing example simulation trajectory). The resulting mechanism agreed well with the experimental data.

As CABS has been described in great detail before,[36] here we only outline its main features. CABS uses reduced representation of protein chains limited to up to four atoms per residue (see Figure 1): Cα carbon, Cβ carbon (except for glycine), a united pseudoatom representing the amino acid side chain (except for glycine and alanine), and a pseudoatom representing the center of a peptide bond. In addition, CABS employs discrete space representation, by limiting Cα atom positions only to a cubic lattice. The distance between lattice beads is arbitrarily set to 0.61 Å, which is small enough to prevent any lattice-associated inaccuracies but large enough to allow for a significant boost in calculations at the same time. Importantly, the resolution of CABS-generated models enables reconstruction to realistic all-atom models.[29,37] The CABS force field is purely statistical, derived from a representative set of known protein structures. Both electrostatic and hydrophobic interactions between the side chains are indirectly included in the mean force statistical potential, and the solvent effects are accounted for in an implicit manner. This potential depends on the types, distance, and orientation of the two interacting residues (it was derived via the Boltzmann inversion method from regularities and preferences observed in known protein structures). Our earlier studies demonstrated that statistical potentials are not only useful in protein structure predictions but are also capable of reproducing the folding pathways.[26−29] Conformational sampling proceeds by small, local transitions controlled by the very efficient replica exchange Monte Carlo algorithm. The long series of such moves may faithfully reproduce the realistic protein dynamics of large-time-scale processes.[26−29] Energy minimization is handled by simulated temperature annealing. Both energy and temperature parameters in the CABS model are dimensionless and thus cannot be precisely converted to known units. Energy serves as a structure-dependent scoring function, while temperature is simply a parameter that controls molecules' mobility. The degree of protein flexibility can be adjusted in CABS from almost completely rigid to fully unrestrained, both locally and globally. In particular, it allows docking a fully flexible protein/peptide ligand to the receptor fluctuating around near-native conformations. All of these enhancements are aimed at modeling bigger systems in longer time-scales without compromising accuracy.

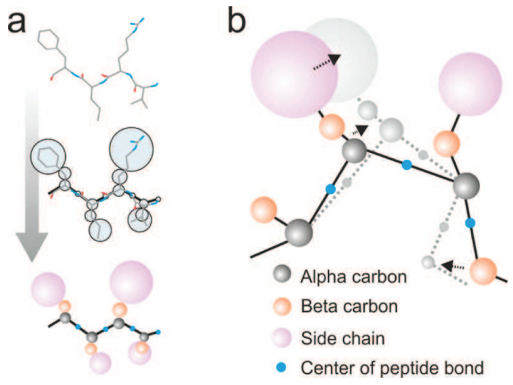


Figure 1. Overview of the simulation model (CABS): (a) coarse-grained representation; (b) example move.

Simulations Setup. In our simulations, we employed an efficient replica exchange version of the CABS Monte Carlo sampling scheme with simulated temperature annealing. In a single simulation, we used 20 replicas uniformly distributed on the temperature scale, with all replicas starting from random conformations. To ensure exhaustive sampling of the conformational space, we ran 75 independent simulations. In these simulations, different temperature ranges were applied in search of optimal conditions at which folding and binding occurs. For a single simulation, an initial setup was constructed in the following manner:

(i) The KIX structure was taken from model 1 from the 1KDX Protein Data Bank (pdb) entry.

(ii) The pKID structure was randomly generated and randomly placed on the surface of a 50 Å radius sphere centered at the KIX center of gravity. No restraints were imposed on the pKID molecule, leaving it fully flexible. Weak distance restraints were imposed on the KIX structure, thus keeping it in near-native conformation, but also allowing some degree of flexibility (note that CABS dynamics was reported to be a good predictor of near-native flexibility[31,32]). Therefore, no information about the pKID molecule's native conformation or its binding location was used.

(iii) For simplicity, CABS was designed to handle protein molecules with the 20 standard amino acids only. To simulate the effect of phosphorylated Ser133 in the pKID molecule, we substituted serine in position 133 with glutamic acid, which is a common approach for mimicking serine phosphorylation.

A single simulation of 20 replicas took around 15 hours on a single CPU. From the resulting data, we calculated the average energy as a function of temperature. Numerical differentiation of the energy produced a plot of heat capacity as a function of temperature (see Figure S1 in the Supporting Information) from which the temperature of the phase transition was found. Obviously, the heat capacity calculated in such a way is a crude approximation, but it serves well in our case for finding an optimal temperature range for simulations.

Simulation Preanalysis and Selection. The structural analysis of the resulting trajectories showed that pKID binds to multiple sites of the KIX surface, which is in accordance with competition and mutagenesis approaches combined with NMR titrations and ^{15}N relaxation dispersion experiments.[22] According to these experimental data, pKID binds natively and nonnatively to additional hydrophobic sites on the KIX surface, including the site for the other KIX-interaction domain (located at the KIX surface opposite to the pKID site).

Furthermore, we analyzed energetic properties of all of the resulting complexes using CABS energy values for: the pKID-KIX complex, or pKID alone, or pKID-KIX interaction. We found that with use of CABS energy values it is not possible to discriminate pKID-KIX complexes, which are closest to the native. More specifically, some of the nonnative complexes being distant from the native binding site (with complex root mean square deviation (RMSD) around 10 or 15 Å to the native) were scored on the same level as those closest to the native. According to experimental data,[22] the native interaction is of high affinity relative to other nonnative sites being of low affinity. Interestingly, the effect of high-affinity binding comes from the fine details of pKID-KIX interactions. Namely, it is due to intermolecular interactions of the phosphate moiety attached to Ser133, since nonphosphorylated KID binds natively to KIX, but with low affinity.[38] Therefore, the native pKID-KIX structure seems to correspond to the local energy minimum which is flattened in the CABS energy landscape (also note that phosphorylated Ser133 is mimicked in our simulations by glutamic acid).

Since our goal was to get an insight into the mechanism of native pKID-KIX binding, for further analysis we attempted to select the simulations, which ended up in a near-native pKID-KIX arrangement. The selection was based on ranking using the highest fraction of near-native complex structures (i.e., RMSD below 5 Å vs model 1 from the 1KDX pdb entry) among 10% of the terminal trajectory frames. Based on these criteria, 10 trajectories were chosen, each one of 1000 snapshots and having at least 50 near-native structures in the final 100 (all of the snapshots of the selected trajectories are characterized in Figure 2). In the simulations selected, the nativelike and nonnativelike complexes can be roughly distinguished using CABS energy values (see Figure 2). Noteworthy, with use of replica exchange sampling, it is not possible to identify the correct folding and binding mechanisms from a replica trajectory. This is because a single replica performs a random walk in temperature space. However, if temperature changes are small over the observed folding and binding events, as in our simulations, the replica trajectories can be considered as correct approximations of the system dynamics.

Finally, the analysis of the folding and binding mechanism presented in this report is largely qualitative. A more detailed qualitative analysis of the energy landscape would be feasible after combining CG protein models with all-atom molecular dynamics (MD) into a multiscale protocol (such as in ref 30), or using physics-based CG simulation models.[13]

RMSD Calculations. In this work, we used RMSD calculated on Cα atoms of the entire pKID-KIX complex as a measure of structural similarity between simulated pKID-KIX complexes and the native complex. Since the KIX molecule was restricted only to small, near-native fluctuations, the KIX structure variations did not affect the RMSD values.

RESULTS

pKID-KIX Complex: A Model System for IDP Binding Studies. To investigate the mechanism of coupled IDP folding and binding, we used a protein docking methodology utilizing the CABS protein dynamics model[26] (see Materials and Methods). We used this approach to simulate the binding of the phosphorylated kinase-inducible domain of the transcription factor CREB to the KIX domain of the CREB binding protein.

The pKID domain is a 28 residue peptide which folds upon binding to the KIX domain.[22,39,40] NMR spectra of pKID in an unbounded form have the characteristics of a disordered peptide with a slight propensity toward helix formation in the KIX binding region.[39] In the pKID-KIX complex (pdb code: 1KDX), pKID is folded into two helices designated as αA and αB (residue numbers 120−129 and 133−145, respectively).[39] These two helices are arranged in the complex structure at an angle of about 90° and are essentially wrapped around the C-terminal α helix (α3) of KIX. KIX is composed of three α helices (from the N- to C-terminus: α1, α2, and α3). KIX α1 and α3 helices form together a hydrophobic binding patch for the pKID αB helix, while the second pKID α helix (αA) interacts with a different side of the KIX α3 helix.

The mechanism of pKID and KIX binding has been a subject of numerous simulation studies. As highlighted in the Introduction, the major obstacle to simulating the IDP binding process is efficient treatment of large-time-scale dynamics, while preserving sufficient model accuracy. The gold standard of accuracy for protein simulations is all-atom explicit solvent MD, recently used for the characterization of pKID-KIX.[20] Due to time-scale limitations, the MD simulations of pKID-KIX binding were carried out as high-temperature unfolding: thus, an assumption was necessary that the mechanisms of high-temperature unfolding and room-temperature folding are similar. The problem of sampling efficiency may be overcome by the application of Go-modeling as demonstrated in pKID-KIX studies.[14−18] The Go-like models provided valuable insights into characteristics of IDPs, e.g., into advantages of IDPs in molecular recognition through the fly casting mechanism.[10] The major assumption under the Go model is that interactions found in the native state also prevail in the studied mechanisms. However, the effect of nonnative interactions on protein folding has been demonstrated in many experiments and is expected to be even more pronounced for IDPs than for globular proteins. This has been demonstrated by studies using a simple Go-like model with a nonnative hydrophobic interaction component,[16,17] which suggested that nonnative interactions might accelerate the binding rate.

Encounter Complexes in pKID and KIX Binding. To enable thorough understanding of the mechanisms of native binding, we jointly analyzed 10 representative simulation trajectories, which ended up in near-native association (for details, see Materials and Methods). These simulations are

characterized in Figure 2. As presented in the figure, the trajectories show multiple minima (representing near-native

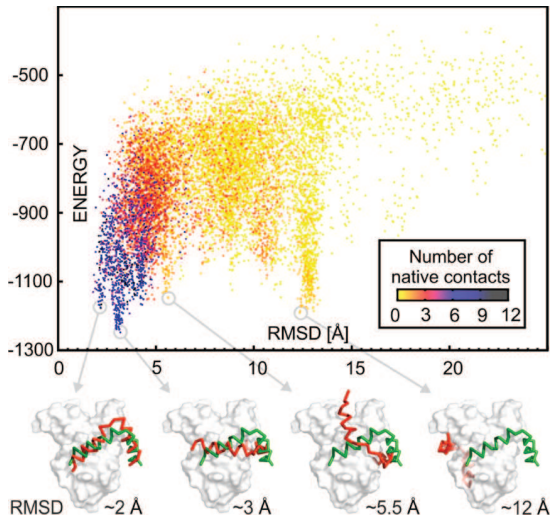


Figure 2. Characteristics of the simulation trajectories. Each point in the plots represents pKID-KIX energy value (vertical axis), RMSD value (to the native pKID-KIX complex; horizontal axis), colored according to the number of native contacts between pKID and KIX. Additionally, example low-energy KIX-pKID models are shown (pKID chain is shown in red, together with its native conformation in green).

associations or a nonnative complex on the opposite surface of the pKID binding site) and a multitude of medium- and high-energy transient states. These kinds of transient states are characteristic for diffusional encounter complexes,[41] an ensemble of molecular arrangements at the end point of diffusional association which involve short-range interactions and nondiffusional motion in search of favorable orientations and contacts.

In Figure 3a, we further analyze the simulation trajectories using the distribution of RMSD values. The histogram analysis shown allowed us to define the following ensembles of complexes (reflected in three significant clusters of peaks): (i) encounter complexes 1 (En1) [representing nonnative or partially native complexes, assembled in the neighborhood of the native site (with RMSD values in the range of 6.5−11.5 Å, constituting 37% of the trajectories)]; (ii) encounter complexes 2 (En2) [representing native or partially native complexes, assembled at least partially in the native site (with RMSD values lower than 6.5 Å, constituting 44% of the trajectories)]; (iii) unbounded/nonnative complexes [representing unbounded or partially native complexes, assembled unlike the pKID native orientation (with RMSD values larger than 11.5 Å, constituting 19% of the trajectories)].

In subsequent text, we will focus on the analysis of encounter complexes En1 and En2, the two dominating ensembles, which comprise transition paths between the very first stage of binding and near-native pKID and KIX arrangements.

Characteristics of Encounter Complex 1. In Figure 3, we present a detailed insight into the complexity of En1 and En2 ensembles, on the level of residue−residue contacts. The figure confronts En1 and En2 contact maps calculated for pKID and KIX interactions (b) and intra-pKID interactions (c). In En1, the average contact frequency between any pKID and KIX residue is at a very low level (Figure 3b, bottom map). This is indicated by a few contacts marked in red/orange (contact frequency around 0.1) and many contacts colored in yellow (contact frequency around 0.05). As shown in the map, En1 is clearly dominated by a multitude of nonnative contacts. Interestingly, these contacts are mostly formed by hydrophobic KIX residues, which are involved in native binding (see vertical yellow lines in the map, in KIX 598−608 and 647−662 regions, and native contact positions marked by green circles). Except for nonnative binding, a tendency to form some of the native contacts can already be observed. In particular, two native contacts of Leu138 with Ala654, and with Tyr650, seem to form an anchoring spot for further binding, as these belong to the most frequent contacts in the En1 set (Leu138-Ala654 being the most frequent; the list of the most frequent contacts in En1 is presented in Table S1 in the Supporting Information). In addition to interacting with KIX, Leu138 is engaged in the formation of the most persistent fragment of pKID secondary structure in the En1 ensemble: the αB helix fragment between 136 and 141 residue. This is indicated by the most frequent intrachain contacts of pKID (Lys136-Asn139 and Leu138-

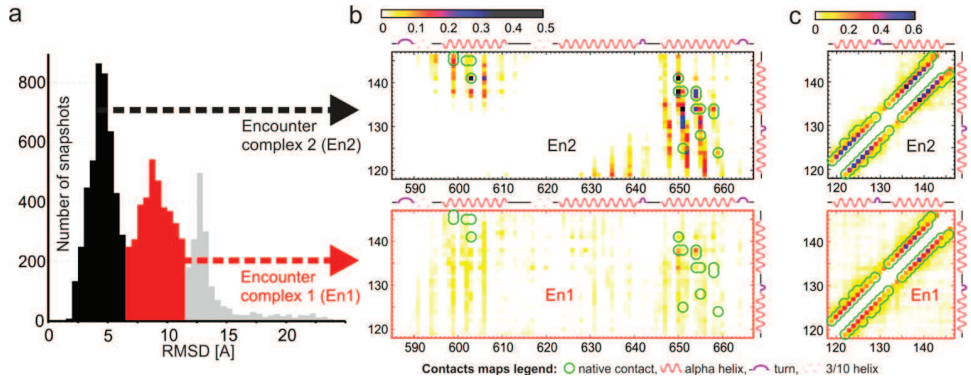


Figure 3. Characteristics of dominant structural ensembles by residue−residue contact maps. (a) Distribution of RMSD values of the pKID-KIX complex (to the native pKID-KIX complex). The distribution suggests the presence of three ensembles marked as Encounter 1 (En1, bars colored in black), Encounter 2 (En2, red) and unbounded/non-native complexes (gray). (b) Residue contact maps for pKID and KIX interactions. (c) Residue contact maps for intra-pKID interactions. The contact maps in b and c show contact frequencies for En2 (upper row) and En1 (lower row) complexes. Contact frequencies are denoted by colors (see color legends). Residue numbers and secondary structures are marked on map borders. Native contacts are outlined in green circles. In the c maps, short-range contacts (up to i, $i + 2$) are omitted for clarity. Contact maps were derived from distances between the gravity centers of the side chains using a 5 Å cutoff.

dx.doi.org/10.1021/ct500287c | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Leu141) present in about 40% of En1 snapshots (see Figure 3c, bottom panel). Other contacts (i, i + 3, reflecting the formation of αA and αB helices) have contact frequency in the range of 0.3−0.1. The magnitude and pattern of intra-pKID contacts suggest that the pKID is generally disordered and that the αA and the αB regions are partly helical, with the αB region being more helical than αA.

The important role of Leu138 is also revealed in the analysis of the involvement of pKID residues in En1 complex formation; see Figure 4. As shown in the figure, Leu138

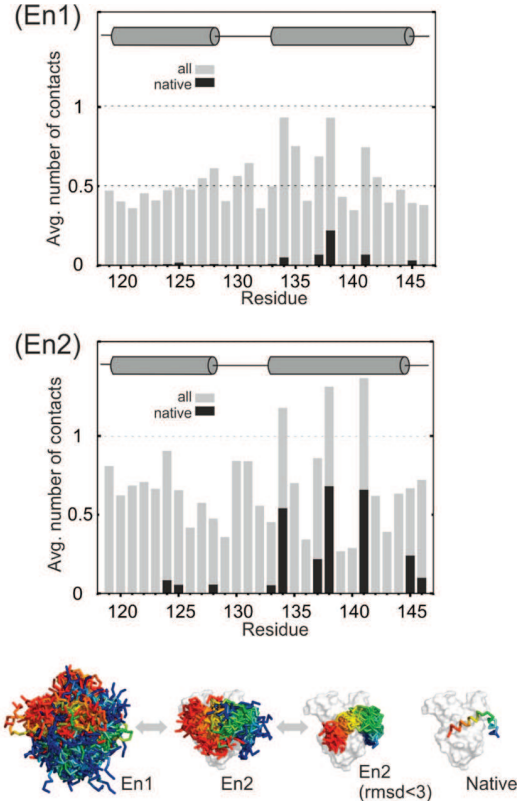


Figure 4. pKID involvement in complex assembly. The plots present an average number of all contacts (gray bars) and native contacts (black bars) for pKID residues upon binding to KIX. Two plots are shown, exclusively for the encounter complexes: En1 (upper plot) and En2 (lower plot). At the bottom, structural visualization of the following ensembles is presented: En1, En2, near-native from En2 (RMSD < 3), and the native pKID-KIX complex structure.

forms native contacts with KIX significantly more frequently than any other pKID residue. The pattern of the average number of nonnative contacts per residue seems to reflect the disordered and highly flexible nature of pKID, since it is uniformly distributed around 0.4 among all of the residues. The only exceptions are Tyr134, Leu138, and Leu141 (together with their neighboring residues) that make a noticeably larger number of contacts, with a small fraction being native.

Characteristics of Encounter Complex 2. The contact map characterization of the En2 complex shows two patches of contacts located just where the binding site is (Figure 3b, top map). As presented in the map, most of the native contacts occur with the highest observed frequencies. At the same time the En2 complex is stabilized by frequently occurring nonnative contacts. The moderate frequency of the most pronounced contacts (in a range of 0.35−0.2) and abundance of various nonnative interactions both demonstrate the transient character of the complex being formed. Native contacts identified as anchoring in En1 still belong to the most pronounced ones; however, in En2, pKID Leu141 seems to have the major anchoring role. Namely, two native contacts of Leu141 belong to four of the most frequent contacts in En2: Tyr134-His651, Leu141-Tyr650, Leu138-Tyr650, and Leu141-Leu603 (the list of the top frequent contacts in En2 is presented in Table S1 in the Supporting Information). Thus, Leu141 stabilizes the interaction of Leu138 with Tyr650 (already pronounced in En1) and simultaneously anchors the pKID to another helix in KIX by contacts with Leu603. These interactions seem to be supported by other native hydrophobic contacts made by Tyr134, Ile137, and Leu138 with KIX Ala654. As demonstrated in Figure 3c, the pKID chain is significantly more helical in En2 than in the En1 complex. This folding upon binding effect is also demonstrated by the absence of cross-helix contacts (present in En1). The essential role of the hydrophobic residues, Tyr134, 138Leu, and Leu141, in native binding is also highlighted in Figure 3 (lower panel). These residues located on one face of the αB helix come into contact with KIX most frequently. This is largely due to their native contacts, with the number of exclusively native contacts much higher than for any other pKID residue (Figure 4).

In the bottom of Figure 4, we compared structural visualizations of the En1 and En2 complexes found in our simulations, together with the native complex structure. In addition, an ensemble of En2 complexes with RMSD lower than 3 Å is also presented. In these nearest to the native complexes, the orientation of the αA helix is less defined than that of αB. This is consistent with the NMR characterization of the pKID/KIX complex structure[39] in which αB helix interactions and orientation relative to KIX are very well defined, while those of αA are not (note that, consequently, the definition of αA native contacts may be not so accurate as that of αB).

In accordance with the above facts, native contacts of pKID αA with KIX are the least pronounced on the maps of native contact occurrence during example simulations (see Figure S2 in the Supporting Information). As shown in Supporting Information Figure S2, transitions from En1 to En2 complexes (reflected in the shift of RMSD values from around 10 to 5 Å) are involved in the cooperative formation of most of the native interactions. This contrasts with En1 complexes, which interact simultaneously usually with one or two clusters of native contacts (formed by residues 124−128 in αA, or by residues 133−138 or 141 or 145−146 in αB).

DISCUSSION

In agreement with NMR observations,[22] our simulations indicate that (1) pKID forms an ensemble of transient encounter complexes upon native binding to the KIX domain and (2) these encounter complexes are stabilized predominantly by hydrophobic contacts (this is in contrast to association of globular proteins, which is driven by electrostatic interactions[42]).

We show that the generated encounter complexes may be separated into two structurally different ensembles, En1 and En2, the first being distant from the native pKID arrangement and the second more nativelike (which comprises ordered, near-native complexes but is mostly disorderedlike). The transition between the En1 and the En2 ensembles reflects a key step for the native association. Most of the native

dx.doi.org/10.1021/ct500287c | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

hydrophobic interactions in the binding interface form later, and rather cooperatively, after the key transition event (see Figure S2 in the Supporting Information). Such a binding pathway in which native interactions form late is consistent with the previous studies on the pKID/KIX binding mechanism using NMR[22] or a topology-based Go-model,[17] and also with studies of the binding of other IDPs using stopped-flow spectroscopy,[43−45] or Φ value analysis.[46]

More importantly, the picture seen in our simulations agrees very well with the experimental identification of pKID residues that are essential for the binding. All of the pKID residues that we found responsible for making initial native contacts (lying in the region of the αB helix: Tyr134, Leu138, Ile137, and Leu141; see Figure 4) have been found in an NMR relaxation−dispersion experiment as critical for productive binding.[22] Moreover, the NMR data suggested that the αB helix region is partly helical (up to 30%) in the low-affinity complex (counterpart of the En1 ensemble in our simulations), which agrees quantitatively with our findings. An important role of the above-mentioned hydrophobic residues has also been demonstrated in mutation analyses.[38,40,47] For instance, Ile137, Leu138, and Leu141 have been identified to be required for pKID to interact with KIX.[40,47] A critical role of Leu141 has also been suggested based on the results of mutating KIX Tyr650[38] and Φ value analysis obtained from Go-type modeling.[17] Moreover, our simulations confirm the important role of secondary structure formation in binding to KIX, as demonstrated by an NMR and biochemical study.[38] This biochemical study concluded that the minimal requirement for interaction (of KID, phosphorylated or unphosphorylated, and C-Myb activation domains) with the native site of KIX is binding-coupled stabilization of an amphipathic helix (αB in pKID).[38] Furthermore, our observation that a structural element of the αB helix forms initial native contacts during the binding is also consistent with other studies emphasizing the important role of the preformed structural elements in partner recognition by IDPs.[48−50] As shown in the experiment,[38] phosphorylation of Ser133 converts KID from a low-affinity binder (interacting with the native binding site) into a high-affinity binder via the formation of additional intermolecular interactions. This effect is not accounted for in our simulations (due to flattened energy curves, as already discussed in Materials and Methods) and our results suggest that a stabilizing effect of phosphorylated Ser133 (mimicked in the simulations by Glu133) may be important in the final complex but not in early encounter complexes.

Based on our simulation data, we propose the following mechanism for the productive binding of pKID to KIX. In the highly disordered encounter complex (En1), a few native contacts begin to form between pKID αB residues (particularly by Leu138 and also by Ile137) and the KIX surface. Except for the formation of these binding contacts (present in up to 10% En1 complexes), Leu138 and Ile137 constitute the most helical region of pKID (helical in about 40% En1 complexes). The key binding and folding step involves significant strengthening of the above-mentioned contacts, which is primarily enhanced by the formation of native contacts of Leu141 and also native contacts of Tyr134. As highlighted in Results (Characteristics of Encounter Complex 2), Leu141 seems to play a major anchoring role in binding to KIX, being also involved in αB helix formation. Overall, the key transition state involves a network of weak native interactions (present in 20−30% of En2) formed by the 134−141 fragment of pKID αB (helical in about 50% of En2). Following the formation of the transition state, the remaining native contacts may be formed.

CONCLUSION

In summary, the binding and folding scenario described here resembles the nucleation−condensation mechanism whose variations appear to describe the overall features of the folding of most protein domains.[51−53] A similar concept of the binding and folding scenario has been recently proposed based on protein engineering and kinetic experiments (Φ value analysis) of the ACTR/NCBD system,[46] or stopped-flow spectroscopy study of reaction between disordered peptides and PDZ domains.[44] Overall, it is likely that this mechanism is common for the coupled folding and binding of IDPs.

ASSOCIATED CONTENT

Supporting Information

Figure S1 showing heat capacity as a function of CABS temperature and Figure S2 showing the progress of the pKID-KIX complex structure formation in two example trajectories, Table S1 listing the most frequent pKID-KIX interactions in the encounter complexes, and movie S1 showing the example trajectory of pKID-KIX folding and binding. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +48 22 822 02 11 ext. 310. Fax: +48 22 822 02 11, ext. 320. E-mail: sekmi@chem.uw.edu.pl.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge funding from the Foundation for Polish Science TEAM project [TEAM/2011-7/6] cofinanced by the EU European Regional Development Fund operated within the Innovative Economy Operational Program; the Polish National Science Centre [Grant NN301071140]; and the Polish Ministry of Science and Higher Education [Grant IP2011 024371].

REFERENCES

(1) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98−104.

(2) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−96.

(3) Hammes, G. G.; Chang, Y. C.; Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13737−41.

(4) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **2010**, *23*, 105−116.

(5) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 415−427.

(6) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321−31.

(7) Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23−30.

(8) Rauscher, S.; Pomes, R. Molecular simulations of protein disorder. *Biochem. Cell Biol.* **2010**, *88*, 269−90.

(9) Uversky, V. N. Multitude of binding modes attainable by intrinsically disordered proteins: A portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* **2011**, *40*, 1623−34.
(10) 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*, 8868−73.
(11) Vajda, S.; Kozakov, D. Convergence and combination of methods in protein-protein docking. *Curr. Opin. Struct. Biol.* **2009**, *19*, 164−70.
(12) Lensink, M. F.; Wodak, S. J. Docking and scoring protein interactions: CAPRI 2009. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3073−3084.
(13) Sieradzan, A. K.; Liwo, A.; Hansmann, U. H. Folding and self-assembly of a small protein complex. *J. Chem. Theory Comput.* **2012**, *8*, 3416−22.
(14) Huang, Y.; Liu, Z. Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding−Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. *J. Mol. Biol.* **2009**, *393*, 1143−59.
(15) Ganguly, D.; Chen, J. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1251−66.
(16) Huang, Y.; Liu, Z. Nonnative interactions in coupled folding and binding processes of intrinsically disordered proteins. *PLoS One* **2010**, *5*, No. e15375.
(17) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* **2008**, *4*, No. e1000060.
(18) Huang, Y.; Liu, Z. Smoothing molecular interactions: the ″kinetic buffer″ effect of intrinsically disordered proteins. *Proteins* **2010**, *78*, 3251−9.
(19) Dadarlat, V. M.; Skeel, R. D. Dual role of protein phosphorylation in DNA activator/coactivator binding. *Biophys. J.* **2011**, *100*, 469−77.
(20) Chen, H. F. Molecular dynamics simulation of phosphorylated KID post-translational modification. *PLoS One* **2009**, DOI: 10.1371/journal.pone.0006516.
(21) Espinoza-Fonseca, L. M. Thermodynamic aspects of coupled binding and folding of an intrinsically disordered protein: A computational alanine scanning study. *Biochemistry* **2009**, *48*, 11332−4.
(22) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021−5.
(23) Kmiecik, S.; Gront, D.; Kolinski, A. Towards the high-resolution protein structure prediction. Fast refinement of reduced models with all-atom force field. *BMC Struct. Biol.* **2007**, *7*, No. 43.
(24) Kolinski, A.; Bujnicki, J. M. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins* **2005**, *61* (Suppl 7), 84−90.
(25) Blaszczyk, M.; Jamroz, M.; Kmiecik, S.; Kolinski, A. CABS-fold: Server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Res.* **2013**, *41*, W406−11.
(26) Kmiecik, S.; Kolinski, A. Characterization of protein-folding pathways by reduced-space modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12330−5.
(27) Kmiecik, S.; Kolinski, A. Folding pathway of the B1 domain of protein G explored by multiscale modeling. *Biophys. J.* **2008**, *94*, 726−36.
(28) Kmiecik, S.; Kolinski, A. Simulation of chaperonin effect on protein folding: A shift from nucleation−condensation to framework mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 10283−9.
(29) Kmiecik, S.; Gront, D.; Kouza, M.; Kolinski, A. From coarse-grained to atomic-level characterization of protein dynamics: Transition State for the Folding of B Domain of Protein A. *J. Phys. Chem. B* **2012**, *116*, 7026−32.
(30) Wabik, J.; Kmiecik, S.; Gront, D.; Kouza, M.; Koliński, A. Combining Coarse-Grained Protein Models with Replica-Exchange All-Atom Molecular Dynamics. *Int. J. Mol. Sci.* **2013**, *14*, 9893−905.
(31) Jamroz, M.; Orozco, M.; Kolinski, A.; Kmiecik, S. Consistent View of Protein Fluctuations from All-Atom Molecular Dynamics and Coarse-Grained Dynamics with Knowledge-Based Force-Field. *J. Chem. Theory Comput.* **2013**, *9*, 119−25.
(32) Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic Acids Res.* **2013**, *41*, W427−31.
(33) Kurcinski, M.; Kolinski, A. Steps towards flexible docking: Modeling of three-dimensional structures of the nuclear receptors bound with peptide ligands mimicking co-activators' sequences. *J. Steroid Biochem. Mol. Biol.* **2007**, *103*, 357−60.
(34) Kurcinski, M.; Kolinski, A. Theoretical study of molecular mechanism of binding TRAP220 coactivator to Retinoid X Receptor alpha, activated by 9-cis retinoic acid. *J. Steroid Biochem. Mol. Biol.* **2010**, *121*, 124−9.
(35) Steczkiewicz, K.; Zimmermann, M. T.; Kurcinski, M.; Lewis, B. A.; Dobbs, D.; Kloczkowski, A.; Jernigan, R. L.; Kolinski, A.; Ginalski, K. Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 9443−8.
(36) Kolinski, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Polym.* **2004**, *51*, 349−71.
(37) Gront, D.; Kmiecik, S.; Kolinski, A. Backbone building from quadrilaterals: A fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. *J. Comput. Chem.* **2007**, *28*, 1593−7.
(38) Zor, T.; Mayr, B. M.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **2002**, *277*, 42241−8.
(39) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell* **1997**, *91*, 741−52.
(40) Shaywitz, A. J.; Dove, S. L.; Kornhauser, J. M.; Hochschild, A.; Greenberg, M. E. Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Mol. Cell. Biol.* **2000**, *20*, 9409−22.
(41) Gabdoulline, R. R.; Wade, R. C. Biomolecular diffusional association. *Curr. Opin. Struct. Biol.* **2002**, *12*, 204−13.
(42) Tang, C.; Iwahara, J.; Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* **2006**, *444*, 383−6.
(43) Bachmann, A.; Wildemann, D.; Praetorius, F.; Fischer, G.; Kiefhaber, T. Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3952−7.
(44) Haq, S. R.; Chi, C. N.; Bach, A.; Dogan, J.; Engstrom, A.; Hultqvist, G.; Karlsson, O. A.; Lundstrom, P.; Montemiglio, L. C.; Stromgaard, K.; Gianni, S.; Jemth, P. Side-chain interactions form late and cooperatively in the binding reaction between disordered peptides and PDZ domains. *J. Am. Chem. Soc.* **2012**, *134*, 599−605.
(45) Karlsson, O. A.; Chi, C. N.; Engstrom, A.; Jemth, P. The transition state of coupled folding and binding for a flexible beta-finger. *J. Mol. Biol.* **2012**, *417*, 253−61.
(46) Dogan, J.; Mu, X.; Engstrom, A.; Jemth, P. The transition state structure for coupled binding and folding of disordered protein domains. *Sci. Rep.* **2013**, *3*, No. 2076.
(47) Shih, H. M.; Goldman, P. S.; DeMaggio, A. J.; Hollenberg, S. M.; Goodman, R. H.; Hoekstra, M. F. A positive genetic selection for disrupting protein-protein interactions: Identification of CREB mutations that prevent association with the coactivator CBP. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13896−901.
(48) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **2004**, *338*, 1015−26.

(49) Sivakolundu, S. G.; Bashford, D.; Kriwacki, R. W. Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol.* **2005**, *353*, 1118−28.
(50) Knott, M.; Best, R. B. A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. *PLoS Comput. Biol.* **2012**, *8*, No. e1002605.
(51) Fersht, A. R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10869−73.
(52) Daggett, V.; Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **2003**, *28*, 18−25.
(53) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **1995**, *254*, 260−88.