Article

Conformational Ensembles of an Intrinsically Disordered Protein pKID with and without a KIX Domain in Explicit Solvent Investigated by All-Atom Multicanonical Molecular Dynamics

Koji Umezawa [1], Jinzen Ikebe [2], Mitsunori Takano [1], Haruki Nakamura [2] and Junichi Higo [2,*]

[1] Graduate School of Advanced Science and Engineering, Waseda University, Okubo 3-4-1, Shinjuku-Ku, Tokyo 169-8555, Japan; E-Mails: k.umezawa@aoni.waseda.jp (K.U.); mtkn@waseda.jp (M.T.)

[2] Institute for Protein Research, Osaka University, Suita, Osaka, 565-0871, Japan; E-Mails: jinzen@protein.osaka-u.ac.jp (J.I.); harukin@protein.osaka-u.ac.jp (H.N.)

* Author to whom correspondence should be addressed; E-Mail: higo@protein.osaka-u.ac.jp; Tel.: +81-6-6872-7252; Fax: +81-6-6872-8219.

Abstract: The phosphorylated kinase-inducible activation domain (pKID) adopts a helix–loop–helix structure upon binding to its partner KIX, although it is unstructured in the unbound state. The N-terminal and C-terminal regions of pKID, which adopt helices in the complex, are called, respectively, α_A and α_B. We performed all-atom multicanonical molecular dynamics simulations of pKID with and without KIX in explicit solvents to generate conformational ensembles. Although the unbound pKID was disordered overall, α_A and α_B exhibited a nascent helix propensity; the propensity of α_A was stronger than that of α_B, which agrees with experimental results. In the bound state, the free-energy landscape of α_B involved two low free-energy fractions: native-like and non-native fractions. This result suggests that α_B folds according to the induced-fit mechanism. The α_B-helix direction was well aligned as in the NMR complex structure, although the α_A helix exhibited high flexibility. These results also agree quantitatively with experimental observations. We have detected that the α_B helix can bind to another site of KIX, to which another protein MLL also binds with the adopting helix. Consequently, MLL can facilitate pKID binding to the pKID-binding site by blocking the MLL-binding site. This also supports experimentally obtained results.

1. Introduction

A traditional view related to protein function, is that a folded three-dimensional structure plays a fundamental role as a scaffold to hold the function. However, this view has been modified by discovery of intrinsically disordered proteins (IDPs), which are proteins (or protein regions) that lack a well-defined three-dimensional structure in the isolated state (*i.e.*, unbound/free state), existing as an ensemble of interconverting conformations. Many IDPs interact with partner molecules, transferring to the folded state (*i.e.*, bound state). The tertiary structures of IDPs in the bound state have been determined experimentally. This remarkable feature is called coupled folding and binding [1], which combines two major subjects—protein folding and molecular recognition—each of which has been studied individually in the protein science field.

Actually, IDPs differ from ordered proteins in several respects: They are found ubiquitously in transcriptional regulators of eukaryote [2] and they frequently undergo posttranscriptional modification [3] such as phosphorylation. Furthermore, some severe diseases are related to IDPs [4]. Consequently, they have been recognized as potential drug targets [5,6]. A biophysically interesting point is that the intrinsic disorder in the unbound state offers advantages over folded proteins [7]. For instance, molecular association is enhanced by the intrinsic disorder [8].

Experiments do not provide sufficient information related to early events in the coupled folding and binding, where the IDP and its partner molecule are separate or weakly interacting with one another. Consequently, theoretical [9] and computational [10–13] studies might be crucially important to reveal important aspects of their early events. One fascinating mechanism from these studies is a "fly casting" model [9], in which the disordered state allows IDP to capture a distant partner molecule because the disordered polypeptide has a greater interaction radius than a well ordered structure does. After capturing the partner, the disordered polypeptide is reeled to form the native complex. However, Huang *et al.* [12] have implied that the kinetic advantage derived from the greater interaction radius must be countervailed by its slow diffusion. They have suggested a picture in which the kinetic advantage may not be derived from the greater interaction radius but from a lowered free energy barrier. Then the IDP reaches the final bound form through fewer encounter complexes with its partner than an ordered protein does. Consequently, the association mechanism of IDP remains controversial. All of these binding schemes were derived from simplified protein models. We suggest that an all-atom computation can provide useful insight into this experimentally undetectable process.

Most computational studies [11–15] have been conducted using simplified protein models such as a Gō-like coarse-grained model, where one amino-acid residue is expressed usually by one sphere. This model has been used widely to investigate protein folding and molecular recognition because of its low computational cost. A typical Gō-like model postulates that natively formed residue–residue contacts (native interactions) in the native structure are energetically favored, even at a transition state. The

other contacts (non-native interactions) likely slow the folding rate. However, the non-native interactions might speed up the folding rate when they help the unfolded polypeptide to collapse, as occurs with competition with chain entropy [16]. If this scheme is correct, then non-native interactions might also positively support or facilitate the IDP–partner association. However, the protein models so far used are too simple to support a realistic discussion of the native and non-native factors. To elucidate these factors, an all-atom protein model is expected to be useful.

The all-atom model involves all the interaction factors. However, it is usually difficult to construct a statistically significant conformational ensemble of protein because the conformational space (potential energy surface) is constructed with a huge number of degrees of freedom and the conformation is frequently trapped in energy local minima during a simulation. Consequently, sampling the ensemble requires unrealistically long computation time by conventional simulations. We have overcome this difficulty using an enhanced sampling method: multicanonical molecular dynamics (McMD) [17,18]. The advantage of McMD is that the energy barriers among the energy local minima are overcome, as explained later. Recently, we developed a more efficient sampling method, trivial trajectory parallelization of multicanonical molecular dynamics (TTP-McMD) [19,20]. Using TTP-McMD, we have conducted an all-atom folding simulation of a 57 amino-acid residue protein [21] and the coupled folding and binding of NRSF and mSin3 [22], in explicit solvent. The NRSF is IDP and folds into the helix when binding to the partner mSin3. The simulation reproduced a conformational ensemble at room temperature, which contained the native-like complex structure in the largest population cluster (*i.e.*, the most thermodynamically stable cluster/the lowest free-energy cluster) as well as non-native complex structures in minor clusters. Therefore, the TTP-McMD is a useful computational technique to examine IDP systems.

The cAMP-response element-binding protein (CREB) induces transcription via an interaction with its co-activator CREB binding protein (CBP). The transcription factor CREB contains a kinase-inducible domain (KID) to bind the kinase-induced domain interacting domain (KIX) of CBP [23,24]. The binding affinity of KID with KIX depends on phosphorylation [25,26]: the affinity increases as Ser133 of KID is phosphorylated. Both KID and the phosphorylated KID (pKID) are IDPs [27], and the tertiary structure of the pKID–KIX complex was determined using NMR at 315 K (PDB ID: 1kdx [27]). The deposited 17 NMR models show that pKID adopts a helix–loop–helix structure on the KIX surface. Therefore the binding of pKID to KIX is cooperated with folding [28]. Sugase *et al.* studied the kinetics of pKID binding with KIX by NMR [29]. This system is suitable for all-atom computations because the pKID sequence deposited in PDB is short (28 residues).

We note some experimental features of this pKID–KIX system. First, the binding affinity of the C-terminal helix of pKID (called α_B; residues 134–145 in the original PDB file) is one order stronger than that of the N-terminal helix (α_A; residues 120–131), and formation of helix in α_B is necessary for the affinity maintaining, although the helix formation of α_A is not [25]. The NMR study [27] has shown that the orientation of α_A relative to the KIX framework is disordered, although that of α_B is well determined with contacting tightly with KIX. Contrarily, in the unbound state, α_A has a higher helix propensity than α_B [30]. These features should be confirmed through simulations.

As described in this paper, we performed TTP-McMD simulations of pKID in the presence and absence of KIX. We denote the residues 120–131 as "α_A residues" and the residues 134–145 as "α_B residues" whether these residues form helices in simulation snapshots or not. Similarly, when the α_A

and α_B residues are expressed as elements, they are denoted respectively as an "α_A region" and "α_B region." We show that the obtained conformational ensemble from the simulation agrees with the experimental features described above, and that the α_A and α_B regions have different mechanisms of coupled folding and binding.

2. Experimental Section

2.1. Setting Simulation Systems (pKID and pKID–KIX Systems)

We designated the simulation system of pKID in the absence of KIX as the "pKID system" and that in the presence of KIX as the "pKID–KIX system." In the NMR experiment on the pKID–KIX complex [27], the pKID sequence was longer than that deposited in PDB (residues 119–146) because unstructured regions are not deposited. We used the deposited pKID region for the simulation, which is the minimum sequence of pKID binding with KIX (residues 586–666).
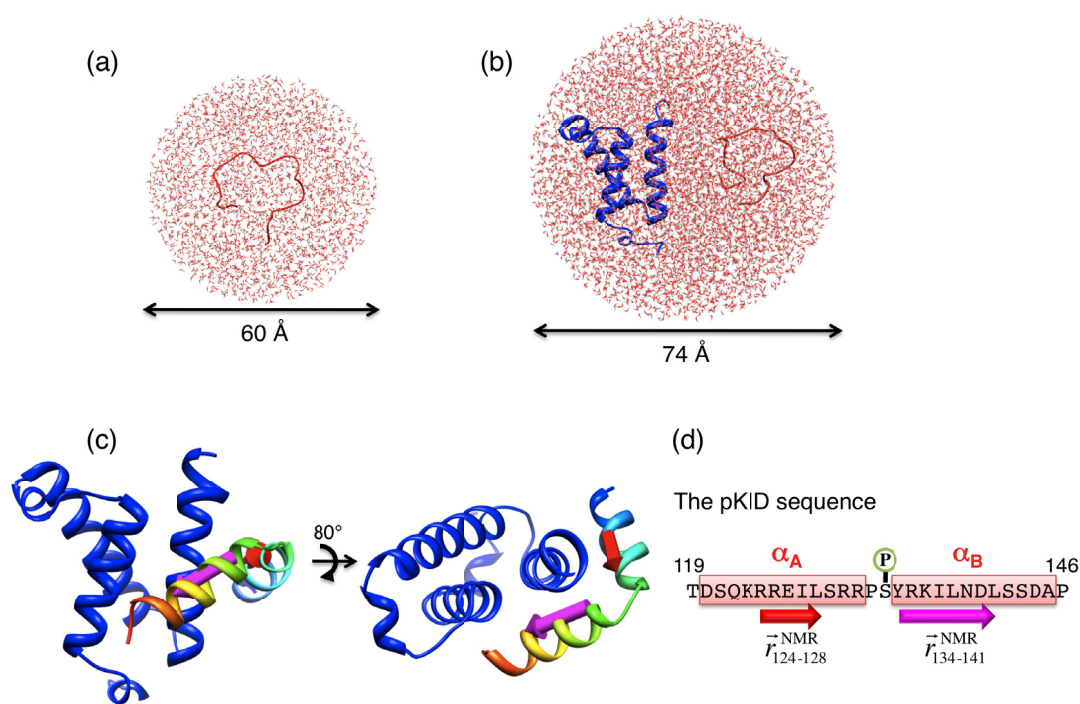
We prepared the pKID system as follows. The coordinates of pKID were taken from the first model out of the 17 NMR ones. The pKID was immersed in a solvent sphere (called sphere 1; radius = 30 Å), which consists of water molecules with the density of 1 g/mL equilibrated at 300 K in advance. The mass center of pKID was set to the center of sphere 1, and water molecules overlapping with pKID were removed. To neutralize the net charge of the system and make consistency with the ionic strength of the NMR experiment, nine water molecules were selected randomly and replaced with five Cl^- and four Na^+ ions. Finally, the pKID system consisted of 3473 water molecules, pKID, and nine ions. Although pKID was taken from the NMR model at this stage, the structure was randomized completely using a high-temperature simulation, as described later.

Next, we prepared the pKID–KIX system as follows. The coordinates of the two proteins were taken from the first NMR model again. They were immersed in a solvent sphere (called sphere 2; radius = 37 Å). The sphere 2 radius was sufficiently large to contain the entire KIX structure, as described later. The center of sphere 2 was set to the mass-center position of pKID of the NMR model. It is noteworthy that the center of sphere 2 is fixed in space (*i.e.*, the center of sphere 2 is not fixed to pKID) when pKID is moving in the simulation: The translation and rotation of pKID are not restrained. Water molecules overlapping the proteins were removed. Eighteen randomly selected water molecules were replaced with nine Cl^- and nine Na^+ ions. Then, the pKID–KIX system came to consist of 6166 water molecules, pKID, KIX, and 18 ions. This complex was dissociated by a high-temperature simulation, as shown later.

We used the AMBER-based hybrid force field for the proteins. This force field is the combination of AMBER force field parm94 (E_{94}) [31] and parm96 (E_{96}) [32] with a mixing rate ω as $E = (1 - \omega) \times E_{94} + \omega \times E_{96}$. We set $\omega = 0.75$ because our previous works indicated that $\omega = 0.75$ reproduces the optimal secondary structure preference for some peptides [33,34]. The TIP3P model [35] was used for the water molecules. After energy minimization, we performed a high-temperature (700 K) canonical MD simulation for each of the pKID and pKID–KIX systems to generate the initial simulation conformations for the following TTP-McMD. Figure 1 shows that this temperature was sufficiently high to demolish the native conformation of pKID and to dissociate the native complex. Although pKID was able to move freely in the solvent spheres, the structure of KIX was weakly

restrained in the simulation of the pKID–KIX system, as described later. We used the simulation program PRESTO ver. 3 [36]. The canonical MD and TTP-McMD simulations were done in the following condition: 1.0 fs time step, SHAKE algorithm [37] to constrain covalent bonds between heavy atoms and hydrogen atoms, and the cell-multipole expansion method [38] to compute the long-range electrostatic interactions. Throughout the simulation, we maintained the volume of sphere 1 and 2 by supplementing a harmonic potential on atoms flying out of the sphere to pull them into the sphere. Additionally, for the pKID–KIX system, the positions of the main-chain atoms (N, Cα, C, and O) of KIX residues (591–594, 597, 608–611, 617–621, 623–640, 646–648, 660–661) were restrained on those of the NMR model by a weak harmonic potential to maintain the KIX structure around the NMR model. No restraint was applied on pKID to diffuse it freely in sphere 2. Figures 1(a) and 1(b) are the last snapshots of the high-temperature canonical MD simulations of the pKID and pKID-KIX systems, respectively, that are used for the initial structures in TTP-McMD simulations.

Figure 1. Initial structures for TTP-McMD simulations of the pKID system (**a**) and the pKID–KIX system (**b**). Blue and red ribbons respectively represent KIX and pKID. Small molecules surrounding the proteins are solvent molecules. Arrows below the spheres indicate the sizes of solvent spheres. (**c**) First NMR model, where KIX is represented by blue ribbon and pKID by rainbow (blue N-terminal; red C-terminal). Red and magenta arrows indicate vectors $\vec{r}^{\,NMR}_{124\text{-}128}$ and $\vec{r}^{\,NMR}_{134\text{-}141}$, respectively, which are defined in Section 3.2. (**d**) pKID sequence, where α_A and α_B residues are highlighted and character P in the circle represents the phosphate group. Arrows below the sequence indicate the starting and ending residues to define the vectors $\vec{r}^{\,NMR}_{124\text{-}128}$ and $\vec{r}^{\,NMR}_{134\text{-}141}$. The structure images were created using Chimera viewer software [39].

2.2. Trivial Trajectory Parallelization of Multicanonical Molecular Dynamics (TTP-McMD)

Before introducing TTP-McMD, we describe the conventional McMD, *i.e.*, the single-run McMD. The single-run McMD is a canonical simulation at a temperature T (a constant-temperature method [40] was used as thermostat in this study) using the multicanonical energy E_{mc} instead of the original potential energy E of the system, as

$$E_{mc} = RT \ln n(E) = E + RT \ln P_c(E,T) \tag{1}$$

where $n(E)$ is the density of states of the system, R signifies the gas constant, $P_c(E,T)$ the probability distribution function of the canonical ensemble at T. If the simulation is sufficiently long, then the single-run McMD provides a flat distribution along the axis of E because the probability distribution is formally given by the following equation.

$$P_{mc}(E) = \frac{n(E)\exp(-E_{mc}/RT)}{Z_{mc}} = \frac{n(E)\exp(-E/RT)}{P_c(E,T)Z_{mc}} = const \tag{2}$$

In that equation, Z_{mc} is the partition function for McMD, which can be regarded simply as a factor to normalize the distribution function P_{mc}. To derive Equation (2), we used a formulation of statistical mechanics: $P_c(E,T) = n(E)\exp(-E/RT)/Z_c$, where Z_c is a factor to normalize $P_c(E,T)$ (*i.e.*, the partition function of the system). This flat energy distribution guarantees that the conformation can overcome the energy barriers and visit low energy conformational regions during the simulation. We refer to the entire conformational ensemble obtained from McMD as a "multicanonical ensemble." Then we can reconstruct a canonical energy distribution P_c at any target temperature T_{tag} from P_{mc} as

$$P_c(E,T_{tag}) = \frac{n(E)\exp(-E/RT_{tag})}{Z_c} = \frac{P_{mc}(E)\exp(E_{mc}/RT)\exp(-E/RT_{tag})}{Z_c/Z_{mc}} \tag{3}$$

To derive Equation (3), we used Equation (2) in the following form: $n(E) = Z_{mc}P_{mc}(E)\exp(E_{mc}/RT)$. The canonical conformational ensemble at T_{tag} is constructed by assigning the probability $P_{mc}(E,T_{tag})$ to all conformations in the multicanonical ensemble.

McMD uses the probability density function $P_c(E,T)$ in Equation (1), which is generally unknown *a priori* (before simulation). Then, we must construct it self-consistently through iterative simulation runs, during which $P_c(E,T)$ converges to a precise function. TTP-McMD takes an advantage over the single-run McMD [19,20]. TTP-McMD is technically equivalent with performing independent multiple runs of single-run McMD starting from various initial conformations. The multiple trajectories generated are integrated simply into an ensemble. It is noteworthy that low-energy conformations (low-energy basins) distribute widely in the conformational space, which are spaced by high-energy barriers. Then, a single-run McMD takes a long flight while overcoming the barriers to reach the low-energy basins. On the other hand, the multicanonical algorithm tries to ensure the flat distribution along the energy axis (Equation (2)) so that the conformation is expected to exist evenly in both the low-energy and high-energy regions. This evenness might cause a difficulty of single-run McMD, *i.e.*, no convergence. The connected multiple runs (TTP-McMD), which are spread in the conformational space, are equivalent to a long trajectory, each part of which flights and searches the low-energy basins. Consequently, TTP-McMD provides the convergence of conformational ensemble faster than the single-run McMD.

The initial conformations for TTP-McMD were those shown in Figure 1(a) and (b), which were generated from the high-temperature canonical MD simulations, as described above. In this study, we performed 256 multiple runs for each system. To obtain the accurate $P_c(E,T)$ in Equation (1), we performed the simulations iteratively. First, we performed a canonical simulation at 700 K and generated the probability $P_c(E,700K)$ for each system. This probability distribution is restricted around the peak position (designated as E_{high}) of $P_c(E,700K)$, and is accurate only in the narrow region. We designate this energy range as $[E_0, E_{high}]$ ($E_0 < E_{high}$), where E_0 is a lower limit of the accurately sampled energy region. Second, we performed the first TTP-McMD at 700 K with the multicanonical energy E_{mc}, where $P_c(E,700K)$ obtained above was used for $P_c(E,T)$ in Equation (1). In the simulation, we set artificial energy walls at E_0 and E_{high} so that the conformation did not escape out of the range $[E_0, E_{high}]$. This simulation produced a flat energy distribution $P_{mc}(E)$ only in $[E_0, E_{high}]$. Then, using Equation (3) we reconstructed the probability $P_c(E,200K)$. Here, E_{mc} defined by $E + RT \ln P_c(E,200K)$ is effective only for multicanonical runs at 200 K. The reason for this temperature reset (*i.e.*, 700 K $\rightarrow$ 200 K) is explained later. We again extrapolated $P_c(E,200K)$ to an energy range as $[E_1, E_{high}]$ ($E_1 < E_0$), and performed multicanonical runs at 200 K to produce a flat energy distribution $P_{mc}(E)$ in this extrapolated energy range. Multicanonical runs were performed iteratively until the sampled energy range reaches an energy (E_{low}) that corresponds to a temperature lower than a room temperature. After reaching E_{low}, we performed the final TTP-McMD to generate a flat energy distribution in $[E_{low}, E_{high}]$. We used 256 computing cores (intel Xeon X5365 3.0 GHz), each core executed one run of TTP-McMD. In the equilibration stage (*i.e.*, the stage to estimate E_{mc} before the final sampling stage), the simulations were done for 21 ns and 23 ns in each of 256 trajectories (total of 5.4 μs and 5.9 μs) of the pKID and pKID-KIX system, respectively. The final runs were done for 8.51 ns in each of 256 trajectories (total time of 2.18 μs) and stored the snapshots every 5 ps for subsequent conformational analyses. The computation times for the equilibration stage were 33 and 76 days for the pKID and pKID-KIX system, respectively. Those for the final sampling stage were 13 and 28 days for the pKID and pKID-KIX system, respectively.

As described above, we reset the simulation temperature from 700 K to 200 K. The multicanonical energies at the two temperatures are, respectively, $E_{mc}(700K) = 700 \times R \ln n(E)$ and $E_{mc}(200K) = 200 \times R \ln n(E)$. Mathematically, the multicanonical ensembles from $E_{mc}(700K)$ and $E_{mc}(200K)$ are expected to be equivalent:

$$P_{mc} = \exp[E_{mc}(700K)/700K \times R] = \exp[E_{mc}(200K)/200K \times R].$$

Consequently, the temperature reset is theoretically non-sense. However, the two simulations produce practically different ensembles. We use a polynomial function to approximate $P_c(E,T)$ in Equation (1). The function form $P_c(E,200K)$ is smoother than $P_c(E,700K)$. Then, $P_c(E,200K)$ is more suitable than $P_c(E,700K)$ to define E_{mc}.
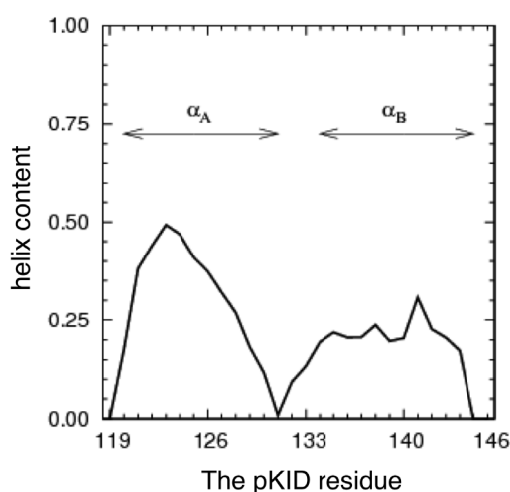
3. Results and Discussion

The conformational ensembles at 315 K of the pKID and pKID–KIX systems were investigated. In Section 3.1, we analyzed the pKID system and showed that pKID is intrinsically disordered in the unbound state. In Section 3.2, we investigated the pKID–KIX system and showed that the free energy landscape of pKID in the bound state is rugged. Furthermore, the orientation of α_A of pKID relative to the KIX framework fluctuates more than α_B does. In Section 3.3, we discuss the mechanism of the coupled folding and binding for this system. In the final section, differences between the unbound and bound states are discussed.

3.1. The Conformational Ensemble in the pKID System (Unbound State)

The TTP-McMD simulation for the pKID system is explored evenly in an energy range [−35000.0 kcal/mol, −23480.0 kcal/mol] that corresponds to a temperature range [300 K, 700 K]. From the multicanonical ensemble, we constructed a conformational ensemble at 315 K (denoted as "315K-ensemble") consisting of 9922 conformations. It is noteworthy that 315 K is the temperature of the NMR experiment [27]. The average of energy E at 315 K was −34272.0 kcal/mol. The secondary structure content at each residue position is shown in Figure 2. Apparently, the α_A region preferred helix more than the α_B region does. This result agrees with the NMR observation [30]. However, the helix content rate was small: The content for α_A was below 50% and that for α_B was about 20%, and the unbound pKID was fluctuating among the helix and non-helix conformations at room temperature. Furthermore, the relative positioning between the α_A and α_B regions fluctuated highly in the 315K-ensemble. Consequently, the simulation confirmed that pKID in the unbound state is intrinsically disordered.
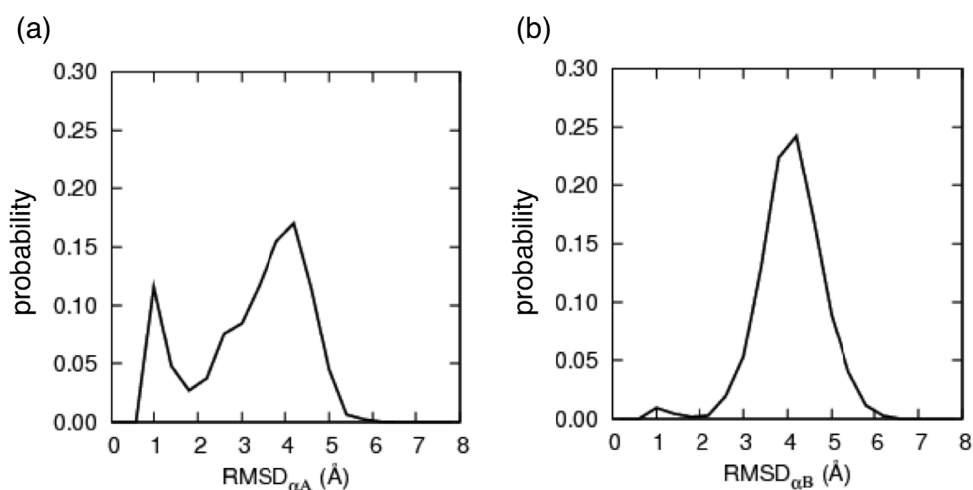
Figure 2. α-helix content rate per residue in the pKID system computed from conformations in the 315K-ensemble. DSSP program [41] was used to assign the α-helix to each residue. The x-axis represents the residue ordinal number of pKID in the original PDB file [27]. Arrows indicate the α_A and α_B regions.

We investigated whether the conformation in the unbound state showed similarity with that in the native bound state (*i.e.*, NMR structure). Picking up a sampled conformation from the 315 K-ensemble,

the root-mean-square deviation (*RMSD*) was calculated for each of the α_A and α_B regions as follows: After superimposing Cα atoms of the α_A region onto those in each of the 17 NMR models, 17 *RMSD* values were calculated and the smallest *RMSD* of the 17 values, denoted as $RMSD_{\alpha A}$, was selected. Similarly, the smallest *RMSD*, denoted as $RMSD_{\alpha B}$, was calculated for the α_B region. The Cα-atomic *RMSD* for the entire pKID (residues 121–144) is denoted as $RMSD_{all}$, where raw coordinates of pKID were used for the RMSD computation without superposition. Figure 3 presents the probability distributions of $RMSD_{\alpha A}$ and $RMSD_{\alpha B}$ for all conformations in the 315K-ensemble. A non-negligible peak was found in the α_A region at small $RMSD_{\alpha A}$: $RMSD_{\alpha A} \approx 1.0$ Å. Contrarily, the α_B region showed a small peak at $RMSD_{\alpha B} \approx 1.0$ Å. These results suggest that the α_A and α_B regions have different mechanisms of coupled folding and binding: α_A might bind with KIX with a population-selection mechanism, where the structured fraction of α_A is used for binding to KIX. In contrast, α_B might bind to KIX with an induced-folding mechanism, where α_B is bindable to KIX with various conformations and then the native complex is formed.

Figure 3. Probability distributions of **(a)** $RMSD_{\alpha A}$ and **(b)** $RMSD_{\alpha B}$ in the pKID system at 315 K.
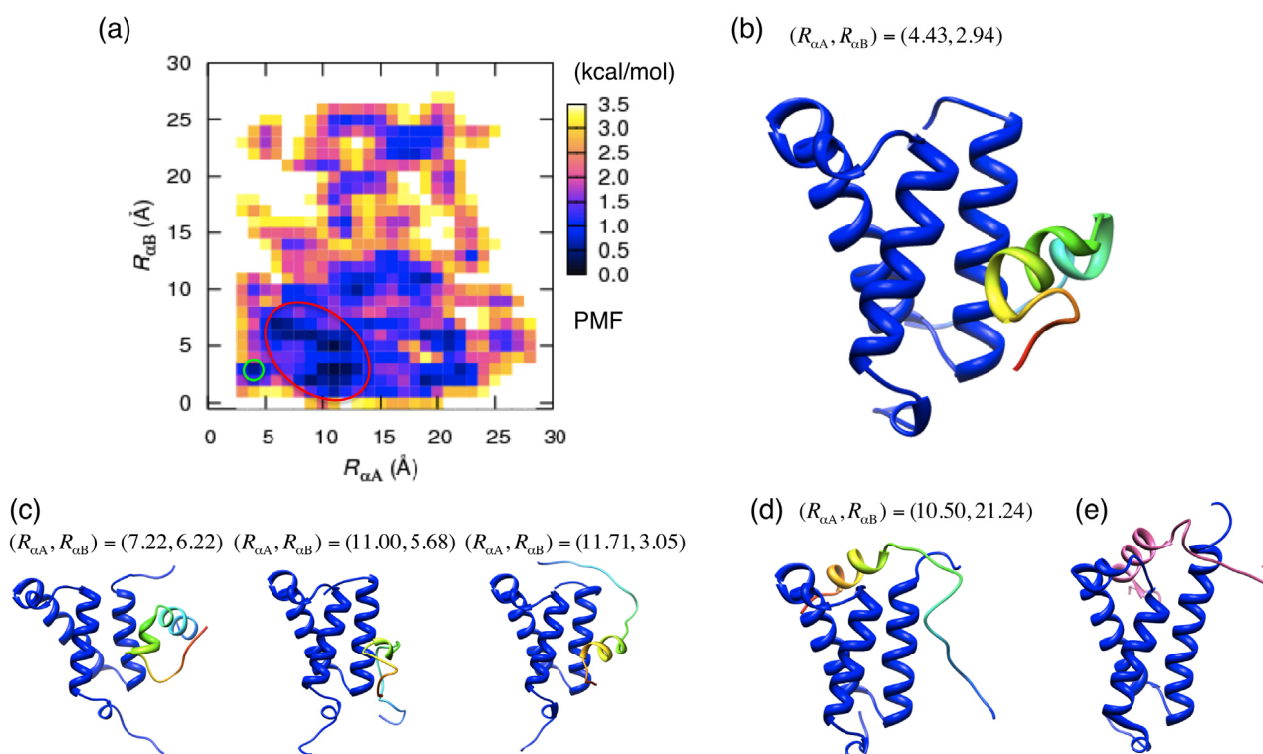

3.2. Free Energy Landscape and Orientation of α_A in the pKID–KIX System

The TTP-McMD simulation of the pKID–KIX system produced a flat energy distribution in an energy range [−64440.0 kcal/mol, −42500.0 kcal/mol], which corresponds to a temperature range [300 K, 700 K]. From the multicanonical ensemble, we derived the 315K-ensemble (8840 conformations). The free energy landscape for the α_A and α_B regions on binding to KIX was constructed as a function of two variables $R_{\alpha A}$ and $R_{\alpha B}$ defined as follows: $R_{\alpha A} = \left| \vec{R}_{\alpha A} - \vec{R}_{\alpha A}^{NMR} \right|$ and $R_{\alpha B} = \left| \vec{R}_{\alpha B} - \vec{R}_{\alpha B}^{NMR} \right|$, where $\vec{R}_{\alpha A}$ and $\vec{R}_{\alpha B}$ respectively represent the position vectors of the centroids of the α_A and α_B regions in a sampled conformation, and $\vec{R}_{\alpha A}^{NMR}$ and $\vec{R}_{\alpha B}^{NMR}$ those in the reference structure (*i.e.*, the NMR model), respectively. Using these distances as the reaction coordinates, we calculated the potential of mean force (PMF) as

$$PMF = -RT \ln P(R_{\alpha A}, R_{\alpha B}),$$ (4)

where the probability $P(R_{\alpha A}, R_{\alpha B})$ was represented as the number of conformations counted in a fraction $[i \pm 0.5$ Å, $j \pm 0.5$ Å$]$ ($i, j = 0, 1, 2, \cdots$) of $[R_{\alpha A}, R_{\alpha B}]$ over the total conformations (8840) in the 315-K ensemble, R is the gas constant, and T is the temperature (315 K). Figure 4(a) shows the free energy landscape *PMF*, which was complex and ragged. We found a low free-energy fraction circled by green circle around $(R_{\alpha A}, R_{\alpha B}) = (4$ Å, 3 Å$)$. We designate this fraction as the "native-like low-free-energy fraction." This native-like fraction comprised conformations with $RMSD_{all} < 7.1$ Å. The nearest-native structure, shown in Figure 4(b), had $RMSD_{all}$ of 5.65 Å, whose position in Figure 4(a) is: $(R_{\alpha A}, R_{\alpha B}) = (4.43$ Å, 2.94 Å$)$. This structure forms a partially disordered α_A helix and a well-ordered α_B helix.

Figure 4. (**a**) Free energy landscape (potential of mean force (*PMF*) defined by Equation (4)) constructed on the plane of $R_{\alpha A}$ and $R_{\alpha B}$. The lowest *PMF* was set to zero. The green and red circles are described in the text. (**b**) Smallest-$RMSD_{all}$ (nearest-native) structure of pKID in the pKID–KIX system. KIX is represented by the blue ribbon, and pKID by rainbow (blue N-terminal and red C-terminal). Values for $R_{\alpha A}$ and $R_{\alpha B}$ shown near the structures indicate its position in panel **a**. (**c**) Structures taken from the red-circle region of panel **a**. (**d**) pKID structure bound at the MLL binding site of KIX via the α_B region. (**e**) MLL (magenta) structure bound to KIX (blue) taken from PDB ID: 2agh.

We also found a non-native broad fraction with low free energy in Figure 4(a) (red circle): $[R_{\alpha A}, R_{\alpha B}] = [5–13$ Å, 0–7 Å$]$. We designate this fraction "non-native low-free-energy fraction", where pKID attached KIX with various binding poses as exemplified in Figure 4(c). Because the two fractions were distinguishable in Figure 4(a), a free energy barrier exists between them. The non-native low-free-energy fraction suggests that pKID can bind to KIX with a structural diversity, and the

generated various encounters reach the final native-like fraction across the free energy barrier. This result suggests that the high flexibility of pKID might help the pKID–KIX association because pKID binds to KIX without adopting a well-ordered structure. The final native-complex formation is completed after forming the various non-native complexes. Later in this report, we describe our examination of why the non-native low-free-energy fraction was larger than the native-like low-free-energy fraction in the free-energy landscape and why the α_A helix is partially disordered, even in the native-like low-free-energy fraction (Figure 4(b)).

We found pKID bound to another site of KIX in 323 snapshots of the 315K-ensemble (Figure 4(d)). This site is a binding site for the activation domain of the mixed lineage leukemia (MLL) transcription factor [42]. The MLL–KIX complex structure (Figure 4(e)) shows that a segment of MLL adopts helix and binds to KIX, and the other parts are unstructured. In all of these snapshots, the α_B region adopted helix to bind to the MLL binding site with the α_A region unstructured. The orientation of the α_B helix cylinder was approximately the same as that of the MLL segment in the MLL–KIX complex structure [43]. Consequently, the α_B region corresponds to the MLL segment. In fact, both the MLL-binding site and the genuine α_B-binding site on the KIX surface consist of hydrophobic amino acids. Furthermore, the hydrophobic residues in the α_B region and the MLL segment have similarity; they contains ϕ-x-x-ϕ-ϕ motif (ϕ = hydrophobic residue and x = any residue), which is conserved in many KIX binding proteins (see Figure 9 of reference [43]). It is particularly interesting that in the presence of MLL, pKID binds to KIX with the two-fold higher affinity than pKID in the absence of MLL [44]. Our simulation results suggest that MLL might facilitate the pKID binding to the genuine binding site via blocking the MLL binding site.

The orientations of the α_A and α_B regions with respect to the KIX framework were investigated, respectively, using inner products, $I_{\alpha A}$ and $I_{\alpha B}$. The inner product $I_{\alpha A}$ was defined as

$$I_{\alpha A} = \vec{e}_{124\text{-}128} \bullet \vec{e}_{124\text{-}128}^{\,\text{NMR}}, \tag{5}$$

where vectors $\vec{e}_{124\text{-}128}$ and $\vec{e}_{124\text{-}128}^{\,\text{NMR}}$ respectively represent the unit vectors of vectors $\vec{r}_{124\text{-}128}$ and $\vec{r}_{124\text{-}128}^{\,\text{NMR}}$: see Figure 1(c). The vector $\vec{r}_{124\text{-}128}$ is pointing from the Cα-atomic position of the 124th residue to that of the 128th residue in a sampled conformation. The vector $\vec{r}_{124\text{-}128}^{\,\text{NMR}}$ is defined in the same way for the NMR structure. For the orientation of α_B residues, $I_{\alpha B}$ was defined as
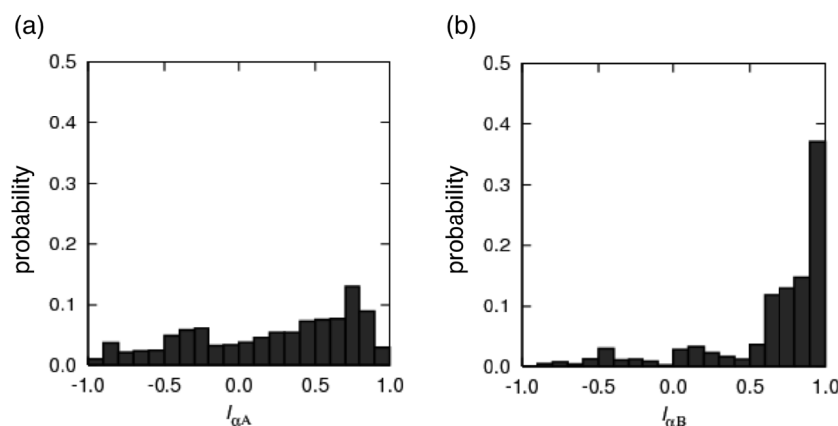
$$I_{\alpha B} = \vec{e}_{134\text{-}141} \bullet \vec{e}_{134\text{-}141}^{\,\text{NMR}}. \tag{6}$$

The unit vectors $\vec{e}_{134\text{-}141}$ and $\vec{e}_{134\text{-}141}^{\,\text{NMR}}$ were calculated similarly as $\vec{e}_{124\text{-}128}$ and $\vec{e}_{124\text{-}128}^{\,\text{NMR}}$ by replacing the residue numbers 124–134 and 128–141. When the inner products $I_{\alpha A}$ and $I_{\alpha B}$ are 1, the orientations are aligned as in the native bound form.

The inner products were calculated for 2476 conformations whose distances $[R_{\alpha A}, R_{\alpha B}]$ satisfy $R_{\alpha A} \leq 13$ Å and $R_{\alpha B} \leq 7$ Å, which involves the native-like (green circle in Figure 4(a)) and non-native (red circle) low-free-energy fractions. The histogram for $I_{\alpha B}$ (Figure 5(b)) has the largest peak at 1, which indicates that the α_B region attaches to KIX with the same orientation as that in the final bound form in both low-free-energy fractions. We discussed the factors stabilizing this orientation later in Section 3.4. Recall that the α_B region less adopts helix than the α_A region in the isolated state (Figure 2) and that the α_B region seldom adopts native-complex form in the isolated state (Figure 3(b)).

Consequently, it is likely that the α_B region binds to KIX with the right orientation and then the helix is formed. In contrast, the α_A region has a large variety in the orientation (Figure 5(a)). These results agree well with the experimental observation: The α_B orientation ordered well and the α_A orientation does less in the native complex form [27]. Furthermore, the simulation results are consistent with the experimental report that the α_B region has a stronger binding affinity than the α_A region to bind to KIX [25].

Figure 5. Orientational probability distribution of (**a**) α_A and (**b**) α_B regions. Definitions for $I_{\alpha A}$ and $I_{\alpha B}$ are given in the text (Equations 5 and 6).
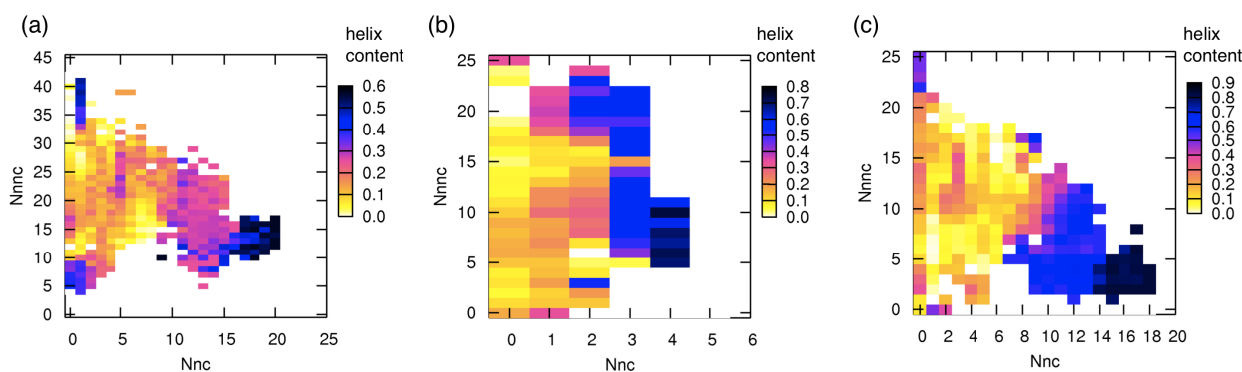

3.3. Correlation between Intermolecular Residue Contacts and Folding of pKID

The folding of α_A and α_B regions upon binding was investigated by measuring residue contacts between pKID and KIX. An intermolecular contact was determined as the distance between the centers of side chains of two residues: If this distance is below 6.5 Å, then we judge that the two residues are contacting. We calculated the contacts ("native contacts") in the 17 NMR models, where a contact is assigned to a residue pair if the pair is contacting in at least 8 of the 17 models. We found 30 native contacts: 8 between α_A and KIX residues, 18 between α_B and KIX residues, and 4 between the other residues in pKID and KIX residues. We designate contacts other than the native contacts found in simulation snapshots as "non-native contacts".

We counted the quantities of the native and non-native contacts, respectively denoted as Nnc and Nnnc, for each conformation. Figure 6 presents the helix content rate in pKID at 315 K projected on the Nnc-Nnnc plane. Coupled folding and binding for α_B is explained well from Figure 6(c): In complexes where the native contacts are less formed, various non-native contacts are formed. With increasing native contacts, the non-native contacts decrease. Finally the full native contacts are formed with a few (*ca.* 5) non-native contacts. This result suggests that coupled folding and binding of the α_B region accords to the induced-fit mechanism, where the α_B region varies the conformation in the encounter complex to reach the native complex. Because the helix content rate of α_B is small in the unbound state (Figure 3(b)), the induced-fit mechanism has an advantage over the population-selection mechanism. In fact, the formation of the non-native contacts before the native contacts might facilitate the pKID–KIX association, as discussed earlier in the Introduction: the α_B region can bind to KIX via non-native contacts without waiting until a helix is formed.

Figure 6. Helix content of pKID at 315 K projected on plane of *N*nc and *N*nnc. Results are shown for the entire pKID (**a**), for the α_A region (**b**), and for the α_B region (**c**).

However, the α_A folding depends less on *N*nc than the α_B folding did. A few native contacts (*ca.* 4, which is 50% of the entire native contacts) were able to stabilize the binding of α_A to KIX. This result derives from the high directional fluctuations of α_A, as shown in Figure 5(a). This result agrees qualitatively with those of the NMR experiment [27], where the flexibility of α_A is greater than α_B. One can recognize that α_A has directional diversity by viewing the NMR models (PDB ID: 1kdx). Furthermore, the fragile contacts between α_A and KIX qualitatively support and agree with the experiment [25] showing that the binding affinity assigned to α_A is weaker than that to α_B. We were unable to specify which of the induced fit or population selection characterizes the α_A binding better, although Figure 3(a) supports the population selection. The large flexibility of α_A prevents us from answering which is likely. However, we emphasize that the dynamics [27] and the weak binding affinity [25] of α_A are characterized by its large flexibility.

3.4. Change of Accessible Surface Area on Binding

We demonstrated the change of accessible surface area (ASA) on binding. We denote ASAs for the α_A and α_B regions respectively as $ASA_{\alpha A}$ and $ASA_{\alpha B}$. The distributions of $ASA_{\alpha A}$ and $ASA_{\alpha B}$ are presented in Figure 7. The average of $ASA_{\alpha B}$ was 1136 Å^2 for the pKID system and 889 Å^2 for the pKID–KIX system. Consequently, the reduction of ASA upon binding is clear for α_B (Figure 7(b)). This reduction was caused mainly by hydrophobic contacts formed in the complex. However, for α_A the reduction was small (Figure 7(a)): the average $ASA_{\alpha A}$ was 1294 Å^2 in the pKID system and 1150 Å^2 in the pKID–KIX system. This small reduction occurs because the α_A region has a large helix probability in the pKID system (Figure 2) and the α_A-KIX interaction is weak (Figure 5(a)). These results are consistent with experimental observations [25], showing that α_A and α_B regions are characterized, respectively, by weak and strong binding affinities.

The averages of $ASA_{\alpha A}$ and $ASA_{\alpha B}$ over the 17 NMR models (complex form) are, respectively, 1093±72 Å^2 and 723±52 Å^2. Consequently, the computed ASA is larger than the experimental value because the computed 315K-ensemble consisted of various complex conformations, as shown in the non-native low-free-energy fraction (Figure 2(c)) and even the conformations in the native-like low-free-energy fraction partially disordered as in Figure 4(b).

Figure 7. Normalized probabilities of the accessible surface area (**a**) for α_A ($ASA_{\alpha A}$) and (**b**) for α_B residues ($ASA_{\alpha B}$). Dashed and solid lines respectively represent the probabilities for the pKID and pKID–KIX systems.
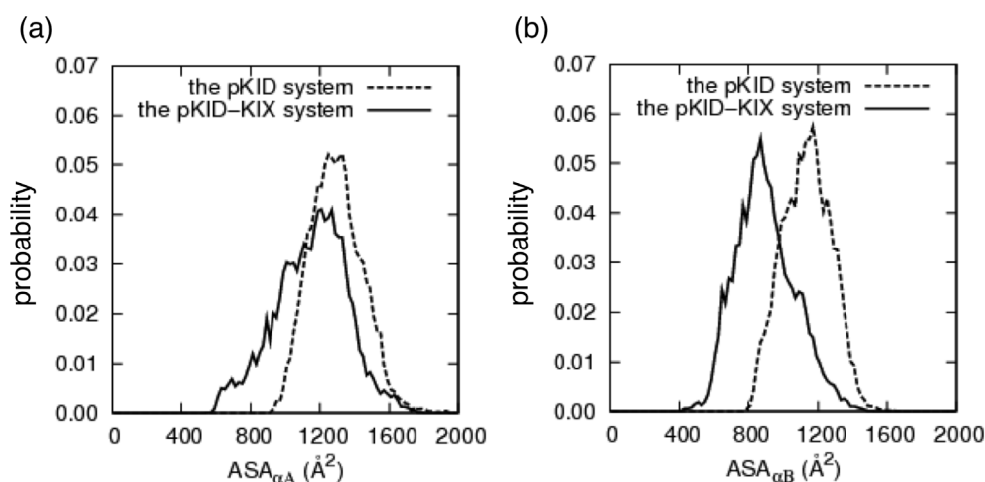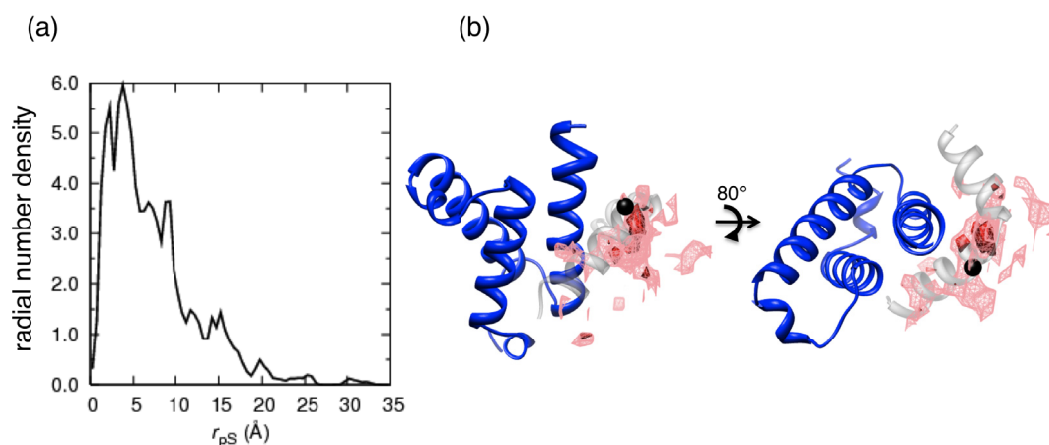

Figure 8. Distribution of the Cα-atom of the phosphorylated SER133 (pS133) in the pKID–KIX system at 315 K. (**a**) Radial (r_{pS}) distribution function defined as: $r_{pS} = \left| \vec{r}_{pS} - \vec{r}_{pS}^{\,NMR} \right|$ where $\vec{r}_{pS}$ denotes the Cα-atom position of pS133 in the sampled snapshots and $\vec{r}_{pS}^{\,NMR}$ one in the reference structure (NMR model 1). (**b**) Spatial distribution of the Cα-atomic position of pS133, where red and magenta contours respectively represent the iso-density regions with density larger than 0.0032 Å^{-3} and 0.0001 Å^{-3}. The black sphere is the position in the NMR Model 1.

A hydrophobic interaction is a non-directional interaction. However, the α_B residues were able to bind KIX with the same orientation as in the native complex form (Figure 5(b)). This directional conservation is inferred from a spatial distribution of the phosphorylated SER (pS133), which is located in the middle of the α_A and α_B (Figure 1(d)). The pS133 interacted with a hydrophilic residue (Y658) and a positively charged residue (K662) of KIX. Figure 8(a) depicts that pS133 restrained the hydrophobic interaction of the α_B helix with KIX as in the native complex form. Figure 8(b) demonstrates the spatial distribution of the Cα-atomic position of pS133, where the iso-density contours were computed for conformations with $R_{\alpha A} \leq 13$ Å and $R_{\alpha B} \leq 7$ Å. Although the high-

density site was slightly deviated from that in the NMR structure, the anchor effect of pS133 is clearly shown. Consequently, phosphorylation plays an important role for specifying the α_B-helix orientation (Figure 5(b)). We presume that this phosphorylation-induced restraint on the α_B helix increases the binding affinity more than non-phosphorylated KID [25,26].

4. Conclusions

We performed the TTP-McMD simulations for the pKID folding and binding with KIX in explicit solvents. Although the overall property of pKID in the unbound state was disordered, pKID has the nascent helix propensity in the α_A and α_B regions in the computed conformational ensemble. The propensity for α_A was stronger than that for α_B, which agrees with an experiment described in the literature [30].

In the presence of KIX, the free-energy landscape at 315 K involved two low-free-energy fractions: The native-like low-free-energy fraction and non-native low-free-energy fraction. Because the α_B region can bind to KIX with various non-native contacts (various encounter complexes), the α_B region might provide fast association with KIX [8]. This landscape proposes an induced-fit mechanism for coupled folding and binding of the α_B region: various encounter complexes are possible in the early stage, and the complex passes through the free-energy barrier to reach the native-like low-free-energy fraction. The well-oriented binding of the α_B region was controlled by the phosphorylated SER133 located in the middle of the α_A and α_B regions. This control supports the higher binding affinity of pKID than KID, as observed experimentally [25,26]. In contrast to the α_B region, the α_A region exhibited high flexibility, which agrees qualitatively with that found in the NMR structure [27]. An earlier experiment [25] has demonstrated that the helix formation of α_A is not important for binding to KIX. Consequently, the simulation supports the high binding affinity of α_B and the low binding affinity of α_A.

It is particularly interesting that the α_B region bound to another shallow hydrophobic concave of KIX than the genuine pKID-binding site, where α_B adopted helix. This hydrophobic concavity is the MLL binding site of KID, to which a segment of MLL binds with adopting helix [42]. It has been pointed out in an earlier report that the hydrophobic residue pattern of the α_B and MLL segments have a similar hydrophobic amino-acid residue pattern [43]. In the presence of MLL, pKID binds KIX with the two-fold higher affinity than pKID alone [44]. We presume that MLL might facilitate the pKID binding to the genuine binding site by blocking the MLL binding site.

The current study demonstrated the importance of hydrophobic interactions between pKID and KIX. Because the multicanonical simulation is an efficient sampling method, the multicanonical trajectory can overcome high potential energy barriers in the conformational space. At the cost of this high performance, the trajectory does not provide information of time series. A conventional MD simulation may provide another important aspect, such as electrostatic interaction, on the complex formation.

Finally, it is noteworthy that the non-native low-free-energy fraction (red circle in Figure 4(a)) is larger than the native-like low-free-energy fraction (green circle) in the free-energy landscape and that the α_A helix is partially disordered, even in the native-like low-free-energy fraction (Figure 4(b)). These points disagree with the NMR experimentally obtained results [27]. Presumably, this disagreement

results from the truncation of pKID, which are unstructured in the NMR experiment. It is important to remember that the computed pKID segment is the only part deposited to PDB. The unstructured region might stabilize the α_A helix more, which might result in an increase of the native-like low-free-energy fraction.

Acknowledgments

References

1. Dyson, H.J.; Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
2. Nishikawa, K. Natively unfolded proteins: An overview. *Biophysics* **2009**, *5*, 53–58.
3. Dunker, A.K.; Brown, C.J.; Lawson, J.D.; Iakoucheva, L.M.; Obradović, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573–6582.
4. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the d2 concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246.
5. Dunker, A.K.; Uversky, V.N. Drugs for 'protein clouds': Targeting intrinsically disordered transcription factors. *Curr. Opin. Pharmacol.* **2010**, *10*, 782–788.
6. Metallo, S.J. Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* **2010**, *14*, 481–488.
7. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Showing your id: Intrinsic disorder as an id for recognition, regulation and cell signaling. *J. Mol. Recognit.* **2005**, *18*, 343–384.
8. Pontius, B.W. Close encounters: Why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem. Sci.* **1993**, *18*, 181–186.
9. 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. USA* **2000**, *97*, 8868–8873.
10. Levy, Y.; Onuchic, J.N.; Wolynes, P.G. Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **2007**, *129*, 738–739.
11. 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*, doi:10.1371/journal.pcbi.1000060.
12. 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*, 1143–1159.
13. Huang, Y.; Liu, Z. Nonnative interactions in coupled folding and binding processes of intrinsically disordered proteins. *PLoS One* **2010**, *5*, doi:10.1371/journal.pone.0015375.
14. Ganguly, D.; Chen, J. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins* **2011**, *79*, 1251–1266.

15. Huang, Y.; Liu, Z. Smoothing molecular interactions: The "kinetic buffer" effect of intrinsically disordered proteins. *Proteins* **2010**, *78*, 3251–3259.

16. Oliveberg, M.; Wolynes, P.G. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **2005**, *38*, 245–288.

17. Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B* **1997**, *101*, 817–824.

18. Hansmann, U.H.E.; Okamoto, Y.; Eisenmenger, F. Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.* **1996**, *259*, 321–330.

19. Higo, J.; Kamiya, N.; Sugihara, T.; Yonezawa, Y.; Nakamura, H. Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. *Chem. Phys. Lett.* **2009**, *473*, 326–329.

20. Ikebe, J.; Umezawa, K.; Kamiya, N.; Sugihara, T.; Yonezawa, Y.; Takano, Y.; Nakamura, H.; Higo, J. Theory for trivial trajectory parallelization of multicanonical molecular dynamics and application to a polypeptide in water. *J. Comput. Chem.* **2011**, *32*, 1286–1297.

21. Ikebe, J.; Standley, D.M.; Nakamura, H.; Higo, J. Ab initio simulation of a 57-residue protein in explicit solvent reproduces the native conformation in the lowest free-energy cluster. *Protein Sci.* **2011**, *20*, 187–196.

22. Higo, J.; Nishimura, Y.; Nakamura, H. A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. *J. Am. Chem. Soc.* **2011**, *133*, 10448–10458.

23. Mayr, B.; Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 599–609.

24. Karamouzis, M.V.; Konstantinopoulos, P.A.; Papavassiliou, A.G. Roles of CREB-binding protein (CBP)/p300 in respiratory epithelium tumorigenesis. *Cell Res.* **2007**, *17*, 324–332.

25. 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*, 42241–42248.

26. Johannessen, M.; Delghandi, M.P.; Moens, U. What turns CREB on? *Cell Signal.* **2004**, *16*, 1211–1227.

27. Radhakrishnan, I.; PÈrez-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*, 741–752.

28. Wright, P.E.; Dyson, H.J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.

29. Sugase, K.; Dyson, H.J.; Wright, P.E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **2007**, *447*, 1021–1025.

30. Radhakrishnan, I.; PÈrez-Alvarado, G.C.; Dyson, H.J.; Wright, P.E. Conformational preferences in the ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett.* **1998**, *430*, 317–322.

31. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

32. Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. *The Development/Application of a "Minimalist" For Organic/Biochemical Molecular Mechanic Force Field Using a Combination of ab Initio Calculations and Experimental Data*; Springer: Dordrecht, The Netherlands, 1997; pp. 83–96.

33. Kamiya, N.; Watanabe, Y.S.; Ono, S.; Higo, J. Amber-based hybrid force field for conformational sampling of polypeptides. *Chem. Phys. Lett.* **2005**, *401*, 312–317.

34. Ikebe, J.; Kamiya, N.; Ito, J.-I.; Shindo, H.; Higo, J. Simulation study on the disordered state of an Alzheimer's β amyloid peptide aβ(12–36) in water consisting of random-structural, β-structural, and helical clusters. *Protein Sci.* **2007**, *16*, 1596–1608.

35. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

36. Morikami, K.; Nakai, T.; Kidera, A.; Saito, M.; Nakamura, H. PRESTO (protein engineering simulator): A vectorized molecular mechanics program for biopolymers. *Comput. Chem.* **1992**, *16*, 243–248.

37. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

38. Ding, H.-Q.; Karasawa, N.; Goddard, W.A. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **1992**, *97*, 4309–4315.

39. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

40. Evans, D.J.; Morriss, G.P. The isothermal/isobaric molecular dynamics ensemble. *Phys. Lett. A* **1983**, *98*, 433–436.

41. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.

42. De Guzman, R.N.; Goto, N.K.; Dyson, H.J.; Wright, P.E. Structural basis for cooperative transcription factor binding to the CBP coactivator. *J. Mol. Biol.* **2006**, *355*, 1005–1013.

43. Lee, C.W.; Arai, M.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. *Biochemistry* **2009**, *48*, 2115–2124.

44. Goto, N.K.; Zor, T.; Martinez-Yamout, M.; Dyson, H.J.; Wright, P.E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). *J. Biol. Chem.* **2002**, *277*, 43168–43174.