# Position-, Disorder-, and Salt-Dependent Diffusion in Binding-Coupled-Folding of Intrinsically Disordered Protein†

**Xiakun Chu**[a] and **Jin Wang**[*,a,b]

[a]Department of Chemistry, State University of New York at Stony Brook, Stony Brook, NY 11794, USA.

[b]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin 130022, China.

## Abstract

Successful extensions of protein-folding energy landscape theory to intrinsically disordered proteins' (IDPs') binding-coupled-folding transition can enormously simplify this biomolecular process into diffusion along a limited number of reaction coordinates, and the dynamics subsequently is described by Kramers' rate theory. As the critical pre-factor, the diffusion coefficient $D$ has direct implications on binding kinetics. Here, we employ a structure-based model (SBM) to calculate $D$ in the binding–folding of an IDP prototype. We identify a strong position-dependent $D$ during binding by applying a reaction coordinate that directly measures the fluctuations in a Cartesian configuration space. Using the malleability of the SBM, we modulate the degree of conformational disorder in an isolated IDP and determine complex effects of intrinsic disorder on $D$ varying for different binding stages. Here, $D$ tends to increase with disorder during initial binding but shows a non-monotonic relationship with disorder in terms of a decrease-followed-by-increase in $D$ during the late binding stage. Salt concentration, which correlates with electrostatic interactions via Debye–H*ü*ckel theory in our SBM, also modulates $D$ in a stepwise way. The speeding up of diffusion by electrostatic interactions is observed during the formation of the encounter complex at the beginning of binding, while the last diffusive binding dynamics is hindered by non-native salt bridges. Because $D$ describes the diffusive speed locally, which implicitly reflects the roughness of the energy landscape, we are eventually able to portray the binding energy landscape, including that from IDPs' binding, then to binding with partial folding, and finally to rigid docking, as well as that under different environmental salt concentrations. Our theoretical results provide key mechanistic insights into IDPs' binding–folding that is internally conformation- and externally salt-controlled with respect to diffusion.
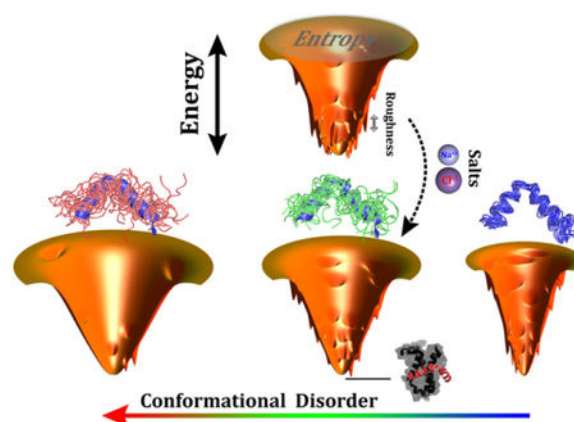
## Graphical Abstract

The topography of binding energy landscape of intrinsically disordered protein is hierarchically heterogeneous and modulated by conformational disorder and salt concentration.

[*]jin.wang.1@stonybrook.edu.

Conflicts of interest

There are no conflicts to declare.

# 1 Introduction

The energy landscape theory has become a standard framework for our mechanistic understanding of protein folding over the last three decades[1,2]. One remarkable postulate derived from the energy landscape theory is that despite the extreme complexity of the high-dimensional protein-folding process, which involves the intricate self-organization of hundreds or thousands of atoms, folding can be effectively described by a few (usually one or two) important collective reaction coordinates[3]. By means of such enormous simplifications, protein folding can be significantly simplified as a diffusive process along a limited number of coordinates. The low-dimensional free energy projection approach has been widely demonstrated to be effective in theory[4], simulations[5–7], and experiments[8,9]. The kinetics from this lowdimensional diffusive protein-folding model can be described by the Kramers' rate theory[10,11], where the stochastic kinetics is determined exponentially by the thermodynamic free energy profile and a diffusion coefficient $D$ as the pre-exponential factor in a diffusive dynamics formalism[2].

Quantifying the thermodynamic protein-folding free energy landscape experimentally has always been a challenging problem. Nevertheless, it can be plausibly achieved by carefully fitting the thermodynamic differential scanning calorimetry (DSC) thermogram to one-dimensional representations of folding free energy landscapes[12] or by analyzing the single-molecule (SM) spectroscopy trajectories along particular relevant reaction coordinates[13,14]. However, measuring $D$ is more elusive in experiments. The recently developed SM Förster resonance energy transfer (FRET) experiment technique seems to be a promising approach, by which $D$ at the free energy barrier top can be inferred by directly comparing the folding and transition path times, as these two critical kinetic quantities depend differently on the thermodynamic barrier height[15]. In addition, $D$ at the unfolded states can be estimated by probing the reconfiguration time based on the autocorrelation function of the end-to-end distance measurements[16]. Increasing experimental evidence implies that $D$ is position-dependent during protein folding[17], but it is still technically challenging in experiments to acquire reliable measurements of $D$ along the folding process. Recent SM force spectroscopy experiments have underlined the difficulties in measuring $D$, because the inherent artifacts led by the handle linkers and probes in force spectroscopy experimental

apparatus can signficiantly affect the results of $D$[18,19]. As an informative complement, molecular dynamics simulations clearly show that $D$ is highly configuration-, coordinate- and temperature-dependent and may have remarkable impacts on the protein-folding process[15,20–22].

$D$ has been established to be critically influential to the folding kinetics in terms of rate or flux[23]. One pronounced demonstration is to possibly shift the position of the kinetic transition state and barrier height[22,24–29]. Such implications in kinetics may have significant impacts for fast-folding proteins that often inherently possess marginal thermodynamic barriers. In this case, the folding rate is close to the "speed-limit"[30], which is primarily controlled by the diffusion. In addition, the pre-factor of the diffusive chain dynamics from Kramers' theory can be interpreted as an incorporation of the solvent external viscosity and protein internal friction[31,32]. The latter is directly related to the protein intrinsic diffusion coefficient, which is embedded in the roughness of the folding energy landscape[4,33]. It therefore provides a practical strategy to infer the topography of the energy landscape by measuring the protein internal diffusion coefficient[16].

Intrinsically disordered proteins (IDPs), which lack a stable three-dimensional structure isolated in solution, often (not always) undergo a "binding-coupled-folding" transition during their functioning process[34–39]. Recent experimental developments on measuring the internal friction of unfolded chains[16,40–42] have significantly improved our understanding of IDPs' intra-chain conformational dynamics. However, it is still a challenge to extend it to the binding process because of the anticipative difficulties that seemingly come along with the complex coupled binding– folding transition. From kinetic aspects, the most prominent functional advantage of IDP may be adequately seizing the "flycasting" mechanism, which can accelerate the binding process by increasing the capture radius through flexibility[43]. However, it is unclear how the conformational disorder impacts the diffusion at the late binding stage, where the coupled-folding often occurs concomitantly.

By quantifying the binding energy landscapes through molecular simulations[44,45], we found that introducing conformational disorders into the dissociative proteins increases the degree of funnelness of the binding energy landscape, intriguingly through decreasing the roughness of the binding energy landscape. The results uncovered an unprecedented effect of conformational disorder on IDPs' binding, wherein flexibility can accelerate the diffusive binding dynamics by smoothing the binding energy landscape surface. It is worth noting that previous approaches calculated the average of the roughness based on the overall topography of the energy landscape, taking no account of the heterogeneity at different hierarchical layers of the energy landscapes[44–46]; thus, they lack a description of the local diffusive dynamics at specific stages during binding. Because $D$ encodes information about the energy landscape roughness, stepwise measuring $D$ along IDPs' binding–folding process can complement the quantification of the energy landscape and finally solidify our understanding of IDPs' binding–folding from an energy landscape point of view.

Here, we implemented coarse-grained structure-based models (SBMs) to investigate the diffusive binding dynamics of an IDP prototype[47,48]. Given their validity and computational efficiency for investigating protein folding, SBMs have been successfully extended to study

IDPs' binding-coupled-folding[49–52], relying on the hypothesis that binding occurs on a funneled binding energy landscape[48]. We found that the binding $D$ is strongly position-dependent along the binding process by applying a reaction coordinate that directly measures the fluctuations in Cartesian configuration space. $D$ in general decreases when approaching bound states, similar to the protein-folding case[16,27,28,40,53]. The binding $D$ is correlated to the folding or collapsed degree in IDP, while the conformational disorder smooths the initial diffusive binding but impedes the diffusive formation of the binary complex in the last stage to a moderate degree. Electrostatic interactions are found to play a dual role with positive and negative effects on diffusion at the transition and post-transition stage. Finally, with quantitative knowledge of $D$ as the binding process progresses, we are able to gain profound insights into the topography of the energy landscape for IDP binding.

## 2    Materials and methods

### 2.1    Structure-based models

We use SBMs to explore the binding-coupled-folding of a well-studied IDP system that includes the phosphorylated kinase inducible domain (pKID) of the cAMP response element-binding (CREB) protein, which is an IDP when isolated in solution[54], and the KIX domain of the coactivator CREB-binding protein (CBP), which is a natively structured three-helix bundle protein[55,56] (Figure S1, ESI†). Here, we utilize the SBM at the one-bead $C_a$ level, initially taking only into account interactions in the native structure; thus, the folding–binding process is significantly accelerated in the SBM simulation at the cost of completely removing the non-native interactions. A typical potential of the plain SBM can be expressed as follows[47]:

$$V_{SBM}^{(KIX, pKID, pKID-KIX)} = V_{Bond} + V_{Angle} + V_{Dihedral} + V_{Native} + V_{Non-native}$$

, where the first three terms describe local interactions, including bond stretching, angle bending, and dihedral rotations; the last two terms are non-local interactions, including native variant Lennard-Jones (LJ) interactions and non-native purely volume-repulsive interactions[47]. The native contact map is built using the Contacts of Structural Units (CSU) software[57], which generates the numbers of intra-chain contacts for KIX and pKID (159 and 25), as well as the number of inter-chain contacts between KIX and pKID (50). The default parameters widely applied in previous research[44,48,50], are used here, unless explicitly specified.

To describe the binding process, we use the fraction of native contacts ($Q$) as a reaction coordinate. $Q$ has been found to be an ideal reaction coordinate in SBM simulations to capture the transition states and distinguish native and non-native states[58]. To make $Q$ continuous, we instead used a switching function that is included in the PLUMED software[59]: for $r_{ij}$  $1.2r_{ij0}$, $Q=1$, while for $r_{ij} > 1.2r_{ij0}$, $Q$ decays smoothly with the following expression:

$$Q = 1 - tanh(\frac{r_{ij} - 1.2r_{ij_0}}{r_s})$$

, where $r_s$ controls the steepness of the decaying function and is usually set to 0.01 nm. We have further normalized $Q$ by dividing the number of contacts in the native structure to make $Q$ be in the range from 0 (completely non-native) to 1 (fully native). To describe different dynamics, different parts of $Q$ are used; for example, $Q_F$ is used for pKID's folding and $Q_B$ is used for pKID-KIX binding.

As an alternative coordinate, we also introduce $d_{RMS}$ [60], which is defined as the difference in the distance $r_{ij}$ between $N_{pairs}$ native pairs

$$d_{RMS} = \sqrt{\frac{1}{N_{pairs}} \sum_{i < j - 3} \left( r_{ij} - r_{ij_0} \right)^2}$$

, to measure the native similarity of folding. $d_{RMS}$ is equal to 0 for the native structure and increases with unfolding. $d_{RMS}$ preserves the magnitude of Cartesian space and is in units of length (*nm* here).

Upon binding, pKID undergoes a large-scale folding transition to form two perpendicularly placed helical structures ($\alpha A$ and $\alpha B$), connected by a post-translational phosphorylated Serine133[55]. To achieve different folding degrees for pKID in isolated states, a pre-factor ($a$) is applied to the dihedral and native LJ potentials in the SBM potential of pKID ($V_{SBM}^{pKID}$) (Figure S9, ESI†), as was previously used to control the conformational flexibility of IDP in SBM simulations[51,61,62].

To investigate the effects of salt concentration, an additional electrostatic potential, modeled by Debye-Hückel theory, is added to the aforementioned plain SBM with the following expression:

$$V_{Ele} = K_{Coulomb} B(\kappa) \sum \frac{q_i q_j exp\left(-\kappa r_{ij}\right)}{\varepsilon_r r_{ij}}$$

, where $K_{Coulomb}$ is a constant, $B(\kappa) \approx 1$ in dilute solutions, $q_i$ is the point charge of the residue $i$, $\varepsilon_r$ is the dielectric constant, and $\kappa$ is the reciprocal of the Debye radius, which is directly related to the environmental salt concentration. To simplify, only Lysine and Arginine are modeled to take one positive charge, and Glutamic and Aspartic acids, as well as the phosphorylated serine, are modeled to take one negative charge. The typical parameters used in the Debye-Hückel model in SBM simulations can be found elsewhere[63–66]. We have re-scaled the strength of LJ and electrostatic interactions for the oppositely charged residues that already form the native contacts to have a reasonable energetic balance[63] (Figure S11, ESI†).

All simulations are performed using Gromacs (Version 4.5.7)[67] following the standard protocol proposed by SMOG web tools[68]. Reduced units are used, so the time, mass, and energy scales are set to 1, except that the length scale is in units of $nm$[68]. The temperature is converted to energy units by multiplying by the Boltzmann constant. A Langevin dynamics with time step 0.0005 is applied. The friction coefficient is set to 1.0. All non-local interactions are cut-off at 3.0 nm. A periodic cubic box with dimensions of 10 nm × 10 nm × 10 nm is used to generate an effective concentration of molecules of 1.66 mM (only one pKID and KIX are placed in each box). For each replica-exchange molecular dynamics (REMD) simulation[69], 24 replicas are used, with a temperature range basically covering the binding–folding transition temperature, and the neighbor replica attempts to exchange every 2000 steps. The Weighted Histogram Analysis Method (WHAM) is then used to collect the data for all temperatures and generate the free energy landscapes[70].

The transition state ensembles are characterized by calculating the conditional probability $p(TP|x)$ for the system on a transition path along the reaction coordinate $x$ with Bayesian expression[71,72]:

$$p\left(TP\middle|x\right) = p\left(x\middle|TP\right)p\left(TP\right)/p_{eq}\left(x\right)$$

, where $p(x|TP)$ and $p_{eq}(x)$ are the probabilities of the system with $x$ being at the transition path and equilibrium ensemble, respectively; $p(TP)$ is the probability of the system being on a transition path. In principle, the highest probability for the transition state lay along the transition path. If $p(x|TP)$ is able to reach the theoretical maximum (0.5), $x$ can be regarded as a good reaction coordinate to describe the system diffusion process. To calculate $p(TP|x)$, six independent long constant-temperature simulations ($1 \times 10^9$ steps) starting from different dissociative conformations are performed. All trajectories are then collected and a total of ~ 1100 transitions are observed by monitoring the (un)binding along either $d_{RMS-B}$ or $Q$ reaction coordinates. This good sampling, especially on the transition path, provides reliable estimation of $p(TP|x)$.

## 2.2    Diffusion coefficient calculations

We use restraining molecular simulations to calculate the diffusion coefficient along the binding reaction[22,24–26,73,74]. The biasing potential is implemented with a harmonic potential centered at a set interested binding point $x$ with strength $K_x$, where $x$ is the reaction coordinate. Here, we use binding $d_{RMS-B}$ instead of $Q$ as the reaction coordinate because $Q$ preserves 0 constantly at unbound states, leading to difficulty calculating the $D$ there. By fulfilling a quasi-harmonic diffusive dynamics approximation, the coordinate-dependent diffusion coefficient ($D(d_{RMS-B})$) can be calculated as:

$$D\left(d_{RMS-B}\right) = \frac{\Delta d_{RMS-B}^2}{\tau_{corr}\left(d_{RMS-B}\right)}$$

, where $\Delta d^2_{RMS-B}$ is the mean-squared fluctuation in $d_{RMS-B}$, and $\tau_{corr}(d_{RMS-B})$ is the relaxation time for the autocorrelation function of $d_{RMS-B}$, $C_{d_{RMS-B}}(t)$. In practice, $\tau_{corr}(d_{RMS-B})$ can be estimated by fitting $C_{d_{RMS-B}}(t)$ to an exponential decay (single or multiple) to obtain the relaxation time or by integrating $C_{d_{RMS-B}}(t)$ based on the knowledge that $\tau_{corr}(d_{RMS-B}) = \int_0^\infty C_{d_{RMS-B}}(t)dt$. We use the latter to obtain the relaxation time and integrate $C_{d_{RMS-B}}(t)$ only up to the first zero-crossing[73,75]. Note that the strength of the biasing potential $K_{d_{RMS-B}}$ should be high enough to make the landscape locally harmonic, fulfilling the need for quasi-harmonic approximation, but should not be too high when the underlying topology of the local energy landscape is completely distorted (Figure S6, ESI†)[22,24–26,74]. Practically, the value of $K_{d_{RMS-B}}$ is determined by performing a series of simulations with different values of $K_{d_{RMS-B}}$ and choosing the one from a range when $D(d_{RMS-B})$ is independent of $K_{d_{RMS-B}}$, so that $D(d_{RMS-B})$ can probe the underlying landscape, rather than the artificially biasing potential (Figure S6, ESI†). There is another alternative way to calculate $D$ using a Bayesian approach[20,21,27,28]. It has been demonstrated that for small dipeptide dynamics and SBM folding, the values of $D$ estimated from these two approaches are similar[20,21].

Constant long simulations with biasing potential ($2 \times 10^8$ MD steps for each simulation) are performed for each $d_{RMS-B}$ with different strengths of $K_{d_{RMS-B}}$. To have sufficient statistics for $D$ calculations, each long trajectory is then divided into 100 same-length pieces of segment to ensure the starting points to calculate the $C_{d_{RMS-B}}(t)$ are random and thus the resulting $C_{d_{RMS-B}}(t)$ from different pieces of trajectory segments are irrelevant. We note that such post-segmented trajectory ($2 \times 10^6$ MD steps for each) is long enough to ensure a good estimation of $\tau_{corr}$, and then $D$ (Figure S8, ESI†).

The effective free energy is calculated by taking into account the diffusion coefficient correction of kinetics with the following expression[22,24]:

$$F_{eff}(d_{RMS-B}) = F(d_{RMS-B}) - kTln(D(d_{RMS-B})/D_u)$$

, where $D_u$ is the diffusion coefficient at unbound state, where $d_{RMS-B}$ is sufficiently large with no interactions between two protein chains. Here, we choose $d_{RMS-B}$ having a thermodynamic free energy minimum equal to 4.00 $nm$.

The diffusion coefficient $D(x)$ of the system at $x$ is related to the energy landscape roughness with the following relationship[33,76]:

$$D(x) = D_0 exp[-(\Delta E(x)/kT)^2]$$

where $D_0$ is the diffusion coefficient in the absence of roughness and $E(x)$ is the energy roughness at $x$. The above expression is valid by assuming the energy roughness is random with a Gaussian distribution[76], and it gives a direct connection between the diffusion coefficient and energy roughness.

The mean $D$ of a certain binding stage can be estimated from integration of $D(x)$:

$$< D > = \int D(x)dx / \int dx$$

, where $D(x)$ is the position-dependent diffusion coefficient at reaction coordinate $x$.

## 3   Results and discussion

### 3.1   Diffusion coefficient modulates the binding

Using the SBM, we quantified the free energy landscape along folding and binding reaction coordinates at the binding temperature (Fig. 1A and 1B). The binding temperature, analogous to the folding temperature, is defined as the temperature when the bound and unbound states are equally populated, so it can be practically extracted from the peak of the heat capacity curve. Simulations at the binding temperature can provide sufficient sampling of (un)binding transition events and are thus conducted in our work. Our results are valid based on the assumption that changing the temperature does not qualitatively change the binding mechanism, as a universal phase boundary between monomers and oligomers appears to exist in different ranges of conditions[77]. We have also performed the simulations at different effective temperatures and salt concentrations and found that the binding mechanisms remain similar (See section "Effects of salt concentration on diffusive binding"). The free energy landscapes clearly show a coupled folding of pKID with binding to its target KIX (Fig. 1A and 1B). In detail, pKID remains unfolded over a broad range of conformational fluctuations when it is far away from KIX. Then, in the free energy barrier region, pKID initiates binding with anchoring KIX through unfolded conformations. Finally, pKID folds with strongly coupled binding after overcoming the barrier. This "binding prior to folding" mechanism, deduced from our 2D free energy landscape, is in line with previous simulations[49,51] and experiments[78], serving as the basis for the following investigations of the diffusion coefficient.

We then plotted the one-dimensional free energy landscape along binding $d_{RMS-B}$ (Fig. 1E). It apparently shows a two-state binding transition, with a barrier located at $d_{RMS-B}$=1.49 $nm$, which appropriately separates the bound (0.01 $nm$) and unbound states ( 4.0 $nm$) (Fig. 1C). The barrier heights for binding and unbinding from The thermodynamic free energy landscapes are respectively 1.42 $kT_B$ and 4.43 $kT_B$. Projection to another well-known SBM-optimized reaction coordinate $Q_B$ [58], which is the fraction of native binding contacts, also leads to a typical two-state binding process, but the free energy barriers for both binding and

unbinding change drastically (5.00 $kT_B$ and 2.65 $kT_B$) (Fig. 1D and 1F). Despite such differences in the free energy barrier, the positions of transition states that are obtained from the thermodynamic free energy landscape and kinetic transition path analyses[71,72] (Fig. 1G and 1H), which are performed by analyzing hundreds of (un)binding pathways within additional constant-temperature simulations, are very similar, based on both of the reaction coordinates. This indicates the validity of these two reaction coordinates to identify the characteristics of the transition states. In addition, it is interesting to note that $d_{RMS-B}$ provides a better description of the binding kinetics because the conditional probability of the transition state being on the transition path identified by $d_{RMS-B}$ is greater than that found along $Q_B$ (0.45±0.03 versus 0.35±0.01), and is very close to the theoretical maximum 0.5. This implies that for IDP binding, $d_{RMS-B}$, which preserves the fluctuations in Cartesian space, seems to be a better binding reaction coordinate than $Q_B$, in contrast to protein-folding cases[27].

Because $d_{RMS-B}$ offers a better description of the kinetics than $Q_B$ does based on the transition path conditional probability calculation, $D$ is calculated along $d_{RMS-B}$ within restraining simulations. We extracted the values of $D(d_{RMS-B})$ at $K_{d_{RMS-B}}$ = 500, when $D(d_{RMS-B})$ roughly remains constant in a range of biasing potential strengths for each $d_{RMS-B}$, and the distributions of $d_{RMS-B}$ are quasi-harmonic (Figure S6, ESI†). With quantified $D(d_{RMS-B})$ (Fig. 2), several interesting insights can be gained: (1) $D(d_{RMS-B})$ almost monotonically decreases as binding proceeds, in line with intuition and experiments on protein collapse[16]. The difference between unbound and bound states is 30-fold, showing a strong position-dependent $D$. (2) The initial slightly increasing $D(d_{RMS-B})$ from unbound states proceeding to transition states ($d_{RMS-B} \sim 2$ $nm$) implies that interfacial interactions may drive the binding using the "fly-casting" mechanism[43]. (3) The sharply decrease of $D(d_{RMS-B})$ that occurs immediately after passing the transition states is similar to that observed in protein folding, where the formation of a native compact structure may increase the local barrier, hindering the diffusive dynamics[16,24]. (4) The decrease in $D(d_{RMS-B})$ starts synchronously with binding after the transition states, giving a hint that the binding transition states are quite loose and the binding collapse may occur after the transition states, consistent with the structural analyses of the transition state ensemble (Figure S5, ESI†), previous simulations[49], and recent experiments[79]. Furthermore, our findings imply that the considerable amount of interfacial disorders, which are retained after binding in terms of "fuzziness"[80], are capable of facilitating the diffusive dynamics existing in the binding complex, as theoretically proposed[43] and also observed in protein-protein/DNA recognition processes experimentally and computationally[81–83].

We are then able to have an effective kinetic free energy landscape by incorporating the position-dependent $D$ into the thermodynamic free energy landscape (Fig. 2). The positions of transition states obtained from the thermodynamic and kinetic free energy landscapes are quite similar, consistent with the results from the transition path analysis (Fig. 1). Furthermore, as expected, the effective kinetic free energy remains quite similar to the thermodynamic one before the transition state, but changes dramatically after the transition state, finally resulting in a rising height shift of 1.85 $kT_B$ and right barrier position shift at bound states. Therefore, the effective free energy barrier height for binding is almost the

same as that obtained in the thermodynamic case, but differs by ~2 $kT_B$ in the post-transition-state regions. This implies that for IDPs with such a binding mechanism, i.e., collapsed complex forms after crossing the barrier, the position-dependent $D$ plays a major role at the last stage, which falls into the fast inactivated "downhill" regime, where the binding kinetics is fully determined by the diffusive dynamics. It would be very interesting to see how $D$ influences the binding kinetics and effective free energy landscape when a compact transition state ensemble is formed during IDPs' binding–folding.

### 3.2   Effects of conformational disorder on diffusive binding

The different degrees of conformational disorder in pKID are realized by changing the pre-factor $a$ of intra-chain interaction strength in SBMs (Figure S9, ESI†). With $a$=5.0, pKID is almost fully folded, with negligible conformational disorder, while with $a$=0.1, pKID is quite flexible and extended with very few folded conformations. The monotonic relationship between the SBM parameter $a$ and structured degree serves as the basis for the following investigations of the effects of conformational disorder on $D$. The two helical segments ($aA$ and $aB$) of pKID were experimentally found to have different propensities for forming helical structures at unbound states[54]. Further calibrating the strengths of intra-chain interactions at the corresponding segments in the SBM potential for experimental measurements could improve the precision of the current SBM[49,51]; nevertheless, it was not performed in the current work. The uniform and gradual introduction of conformational disorder into pKID could help focus on characterizing the general effects of disorder on IDPs' binding[62].

We observed a similar tendency for $D$ along $d_{RMS-B}$ compared with that of the default SBM parameter ($a$=1.0) for different degrees of conformational disorder (Fig. 3A). Then, we divided the binding process into three stages: the binding state ensemble (BSE), transition state ensemble (TSE), and unbinding state ensemble (USE), based on the transition path analysis at $a$=1.0 (Fig. 1). From Fig. 3A, we can see that while approaching the bound state (decreasing $d_{RMS-B}$), $D$ slightly increases at USE, then starts to decrease at TSE, and finally decreases sharply at BSE for all degrees of conformational disorder. We applied SBMs with a wide range of flexibility-control parameter $a$, which basically covers binding from completely unfolded to partially folded, and finally fully folded, monomers. Therefore, our results here imply that such $d_{RMS-B}$–dependent $D$ behavior, which is also similar to that observed in protein folding[27,28], may be applicable to general protein-protein binding cases. In addition, $D(d_{RMS-B})$ appears to increase with additional conformational disorder, in line with intuition and experimental findings that loosely unstructured conformation favors fast diffusion[16].

Interestingly, careful examination of the last binding stage within the BSE at low $a$ (high degree of folding) shows an increase in $D$ turnover as $d_{RMS-B}$ decreases. This leads to an opposite result with respect to the relationship between conformational disorder and $D(d_{RMS-B})$, where a rigid protein chain can diffuse quickly. To quantitatively identify such an effect, we calculated the mean $D(d_{RMS-B})$ for the three typical aforementioned stages and also the entire process (WSE) (Fig. 3B). The mean $D$s of USE, TSE, and WSE increase monotonically with decreasing $a$. This indicates that as the conformational disorder

increases, the diffusion at the unbinding stage, transition stage, and for the whole binding process that the protein chain can achieve is faster. However, at the last binding stage (BSE), increasing conformational disorder initially decreases the diffusion rate, probably because of the boost in local barriers for the energy landscape, led by non-native conformational topology. Finally, the conformational disorder increases the diffusion rate because of the loose binding, the same result that occurred in other binding stages. Such dual effects of conformational disorder on $D$ at different stages point to a complex binding diffusive process that the IDP inherently possesses.

To explain the non-monotonic relation between conformational disorder and $D$, in particular at high values of $a$, we calculated the free energy landscapes and evolution of the two helices $aA$ and $aB$ along the binding (Fig. 4). It is interesting to see that the binding transition pathways are different at different degrees of conformational disorder. At low $a$ (0.1 and 1.0), the two helices intend to bind primarily through two rectangular edges, along with an additional intermediate pathway from $Q_{BaA} \sim 0.2, Q_{BaB} \sim 0.0$ to $Q_{BaA} \sim 0.2, Q_{BaB} \sim 0.8$. These distinct pathways are separated from each other without apparent connections. These two parallel binding pathways were also observed in previous work, when a similar SBM approach was applied[49]. At high $a$ (3.0 and 5.0), many binding pathways emerge that can potentially connect the states reciprocally on the free energy landscapes. Such binding scenarios, along with different conformational disorders, can also be observed from the binding evolution of the two helices to KIX (Fig. 4B). Binding of the two helices starts to be distinct after passing the TSE, and at $a = 1.0$, the differences are the most prominent, with binding of $aB$ accomplished prior to that of $aA$. This is probably due to the fact that $aB$ has more binding contacts formed in the native structure than $aA$ does. Decreasing or increasing $a$ can modulate the two helices to bind synchronically; however, they have different mechanistic factors. At low $a$, entropy dominates the binding, so the energetic stabilization term has less effect on managing the pathway than that in $a = 1.0$. At high $a$, the stabilization of the binding complex largely comes from the intra-chain energetic term of pKID, so the inter-chain interactions become less dominant with respect to control of the binding pathway. The binding coupling between the two helices at high $a$ with low conformational disorder can lead to multiple binding pathways, which increases the possible connections of states on the free energy landscape, facilitating the escape from local energetic or topological traps, which directly corresponds to the diffusion coefficient. The other factor that increases the diffusion coefficient at a high degree of conformational disorder may be contributed by the loose binding complex formed, according to previous experiments[40–42].

### 3.3 Effects of salt concentration on diffusive binding

Adding electrostatic interactions into the plain SBM will inevitably introduce non-native interactions beyond the repulsive volume term existing in the SBM. Intuitively, the non-native competing interactions decrease the stability, as also is observed in our simulations (Figure S13, ESI†), but realistically may have more complicated impacts on kinetics[84]. To determine how the electrostatic interactions modulate $D$, we change the environmental salt concentration, which controls the length of the Debye radius, thus leading to different strengths of electrostatic interactions.

Binding with electrostatic interactions also leads to a similar tendency of $D$ along $d_{RMS-B}$ (Fig. 5A). With different salt concentrations, which lead to different strengths of electrostatic interactions, the change in $D(d_{RMS-B})$ is not significant and the effects seem to vary at different binding stages. We then calculated the mean $D$ at different binding stages (Fig. 5B). The results show that electrostatic interactions have very little influence on the diffusion rate at USE, which corresponds to the completely unbinding stage, but facilitates diffusion at TSE, implying that the "flycasting" mechanism may be effective at TSE for diffusive binding with long-range steering electrostatic interactions, although they may be non-native[50]. At the BSE, where pKID has overcome the barrier accompanying the coupled-folding occurring in a downhill manner, the electrostatic interactions slow down the binding diffusion, mostly because they are formed along with the increasing compactness of the binary complex, where the short-ranged nonnative electrostatic interactions will increase the roughness of the energy landscape. Finally, the mean rate of diffusive dynamics shows a slight increase with increasing electrostatic interactions, as a net gain from compensation between TSE and BSE. These findings are similar to those obtained from different conformational disorders of pKID and temperatures (Figure S17 and S18, ESI†), implying that a salt-dependent diffusion coefficient may persist for general protein-protein binding cases.

To explain such salt-dependent diffusive binding dynamics, we further calculated the number of non-native salt bridges progressively formed during the binding process (Fig. 6). From both the structural characteristic and free energy landscape (Figure S14, ESI†), the two helices $aA$ and $aB$ show unsynchronized steps with binding to KIX. The $aA$ helix contains more charged residues than the $aB$ helix does. This gives a hint that the electrostatic interactions may play more important roles in binding of the $aA$ than the $aB$ helix. In Fig. 6, there are more non-native salt bridges formed between KIX and $aA$ helix than $aB$ throughout the binding process. In particular, the number of salt bridges formed by $aA$ with KIX initially increases in both the USE and TSE stages, then decreases slightly at the beginning of the BSE Stage, followed by a sharp increase at $d_{RMS-B} \sim 0.5\,nm$, and finally decreases to 0 at the completely bound state. The number of salt bridges formed by $aB$ simply increases at the USE and TSE stage, and finally decreases in the BSE stage to 0 at the bound state. At different salt concentrations, stronger electrostatic interactions lead to a higher number of non-native salt bridges (Figure S16, ESI†). The distinct dependence of salt bridge formation in the two helices can shed light on the explanation of the relationship between $D(d_{RMS-B})$ and the salt concentration. Binding of pKID is steered by the long-range non-native sparsely formed interactions, which primarily rely on the highly charged $aA$ helix in accordance with the "fly-casting" mechanism[43] at the USE and TSE stages (Figure S15, ESI†), whereby the diffusive binding dynamics is thought to be facilitated. When pKID and KIX are close in space to the BSE from $d_{RMS-B} \sim 1.0$ $nm$, the $aB$ helix is triggered to specifically anchor with KIX through short-range native binding contacts prior to the $aA$ helix. This is based on the fact that $aB$ has a larger number of native binding contacts with KIX than $aA$. Therefore, the nonnative salt bridges between the $aB$ helix with KIX are weakened. For the $aA$ helix, the native binding shows only a slight change, but the number of non-native salt bridges decreases first because the native binding of the $aB$ helix aids in eliminating the nonnative interactions. Then, the non-native salt bridge increases,

mostly because the non-native interactions between $_{a}A$ and KIX can guide the search for binding sites after $aB$ accomplishes the binding (Figure S15, ESI†). Further proceeding, the binding of pKID to KIX in the BSE ($d_{RMS-B}$ from ~ 0.5$nm$ to ~ 0$nm$) mostly relates to a slight adaption of the $aB$ helix by native contacts on the surface of KIX to native binding sites with further elimination of the non-native salt bridges, as well as native binding of the $aA$ helix by breaking the strongly pre-formed non-native salt bridges. This will certainly lead to slow diffusive binding dynamics (Figure S15, ESI†). In summary, although the salt concentration exhibits little impact on the overall binding diffusion coefficient, it manipulates the diffusive dynamics step-by-step during binding by forming periodical intermittent non-native salt bridges and heterogeneous binding pathways.

## 4 Conclusions

One functional advantage of IDPs is that (un)binding is efficiently fast thanks to the inherent disorder. To explain this effect, theory has successfully introduced a "fly-casting" mechanism[43], by which a flexible chain possesses a larger capture radius than a rigid one, to facilitate binding at the very beginning stage. The simulations and experiments also have unambiguously observed the phenomena of binding acceleration by conformational disorder[85–88]. Previously, the "fly-casting" mechanism was interpreted in terms of free energy, where conformational disorder tends to lower the binding energy barrier[43], which contributes to the exponential term in Kramers' theory. The pre-factor as diffusion coefficient was often ignored. Incorporating the effects of the diffusion coefficient by careful assessments of the "fly-casting" mechanism on binding rates has given contradictory results, where the influence of the "fly-casting" mechanism on the binding kinetics led by the conformational disorder is overestimated[61]. This occurs because the diffusion of a flexible chain is likely slowed down by its extended conformation because of the larger hydrodynamic radius[38]. Based on our quantified $D$, one may draw another conclusion: as from Fig 3, when the two chains are completely dissociative at $d_{RMS-B} > 4.0$ with $Q_B$=0, $D(d_{RMS-B})$ increases monotonically with added conformational disorder. The seeming contradiction primarily occurs because $d_{RMS-B}$ reflects the Cartesian coordinates of the residues evolved in the binding native contacts. This not only describes the binding process, but also captures the intra-chain dynamics. The conformational disorder has more flexible dynamics on each residue of the isolated IDP. This may result in a higher diffusion along $d_{RMS-B}$, compared with a rigid chain, where the diffusion of $d_{RMS-B}$ only derives from binding.

Nevertheless, the effect of the "fly-casting" mechanism on diffusive binding dynamics is faithfully observed within our salt concentration-dependent simulations. The isolated IDP has an inherent diffusive characteristic, which is insensitive to ionic strength. However, the environment with low salt concentration has a weaker Debye screening effect on electrostatic interaction. Thus, the capture radius of the charged protein chain is relatively larger than that with high salt concentration. The initially formed electrostatic interactions, which are mostly non-specific and nonnative, can drive the two chains close to each other in space. This indirectly accelerates the diffusive binding until the transition state stage. Therefore, we state that the "fly-casting" mechanism can not only reduce the thermodynamic free energy barrier, but can also accelerate local diffusion at the beginning

of binding. In general, IDPs hold higher percentages of charged residues than globular proteins[38,89], and the vicinity of IDP binding interfaces are rich in complementary charges[90]. These facts likely enhance the effects of the "fly-casting" mechanism on IDPs' binding kinetics in practice.

It is still controversial whether the pre-formed structures in IDPs aid or hinder binding[91–95]. From a diffusion perspective, we have addressed the complex non-monotonic role of conformational disorder on IDP binding at the last stage, from initial to final complex. A large degree of conformational disorder can prevent the orientational restraints and steric hindrance as topological frustrations[39]. This leads to loosely formed or fuzzy complexes[80] that consequently accelerate the local diffusion. Decreasing conformational disorder may partition the energy landscape into a handful of disconnected binding pathways, where the diffusion is likely hindered by limited directional constraint movements. Further decreasing conformational disorder into completely folded proteins, in some cases, boosts the possibility that a rigid protein can smoothly slide in the vicinity of the binding sites to accomplish docking, particularly when the binding interfaces are extended[96]. This potentially increases the chance of escaping the traps on the energy landscape with the emergence of additional pathways. Our results regarding the impacts of conformational disorder on diffusive binding reveal an unprecedented complex binding mechanism for IDPs.

The quantified position-, disorder-, and salt-dependent diffusion coefficient, which inherently encodes the roughness of the energy landscape, provides a practical way to infer the topography of the binding energy landscape hierarchically. This moves beyond previous investigations[44–46], where the roughness was mostly estimated by implicit averaging from the overall energy landscape, lacking detail at each layer of the funnel. The topography of the energy landscape is quantified by the quantity $\Lambda = \delta E/(\Delta E\sqrt{2S})$, where $\delta E$, $E$, and $S$ are the average energy gap, roughness, and entropy, respectively. Combined with our previous work on quantification of the energy landscape[44–46], we are now able to provide a more precise portrait of the binding energy landscape (Fig. 7): the roughness at the top of the binding funnel is moderate, then with binding proceeding to the middle of the funnel, the roughness tends to slightly decrease because of transiently formed attracting intermolecular interactions; finally, the roughness increases significantly at the bottom of the funnel when the compact binary complex forms. This funnel picture is in accord with the disconnectivity graph energy landscape representation of a protein-folding SBM[97,98], implying that even a native-centric model without any energetic frustration still has remarkable topology-led energy roughness, which varies stepwise differently during the binding/folding process.

The increase in intrinsic disorder in binding protein has a distinct influence on increasing entropy, but decreasing energy roughness, in modulation of the binding energy landscape. This results in a higher $_\Lambda$, representing a more funneled binding energy landscape[45]. Incorporating current studies based on quantifications of the diffusion coefficient, we therefore offer a profound understanding of the roles of intrinsic disorder on modulating the binding energy roughness (Fig. 7): a fully unfolded protein with a high degree of conformational disorder eliminates the roughness triggered by topological frustrations throughout the binding process, portraying a smooth binding energy landscape; decreasing

the intrinsic disorder in a protein chain increases the binding roughness at every layer of the funnel, where the bottom layer of the binding funnel appears to be more affected as a result of the higher topological frustration that the compact complex potentially has; further decreasing the conformational disorder continuously increases the roughness at the funnel top, but smoothens the bottom of the funnel by means of multiple emerging binding pathways that assist to smooth the local topological barriers on the energy landscape. Overall, the global roughness of the binding energy landscape monotonically decreases with increasing conformational disorder, consistent with our previous work, where a completely different approach based on calculation of density of states was applied[45].

Electrostatic interactions, which are more common in IDPs than globular proteins[38,89,90], also have heterogeneous roles in modulation of the topography of the binding energy landscape (Fig. 7): removing the ions in the solution increases the strengths of electrostatic interactions by eliminating the ionic screening effect. This facilitates diffusive binding by smoothing the energy roughness at the middle of the funnel via non-native long-range transiently formed electrostatic interactions. Conversely, the nonnative salt bridges formed during the approach to the binding complex lead to energetic roughness at the bottom of the funnel. Such dual effects of electrostatic interactions on the roughness are compensated during the binding process and lead to a slightly decreasing global roughness of the binding energy landscape in the end. In addition, electrostatic interactions in IDPs have been found to contribute to the stabilization of the binding complex[99–101], which can deepen the bottom of the funnel and increase the energy gap. Conversely, the intra-chain electrostatic interactions of some IDPs, which in particular have high net or bipolar charge, are capable of collapsing the conformation of isolated IDPs, depending on the environment[65,102–106]. These collapsed structures in IDPs, either native or non-native, are expected to have distinct effects on the energy gap and entropy. This adds baffling complexity into the effects of electrostatic interactions on the binding energy landscape, whereby the interpretation seems to proceed on a case-by-case basis.

The other advantage of intrinsic disorder in protein recognition is the capability of binding with high specificity but low affinity[35,36,39,107–109]. However, it is always practically challenging to measure the specificity, which is conventionally defined as the relative affinities between all possible binding targets. We have verified the validity of transferring the conventional specificity to intrinsic specificity, which describes the distributions of different binding modes and is determined by the energy landscape[110–112]. Therefore, measuring specificity is feasible by quantifying the topography of the binding energy landscape. Based on our previous work[45], the intrinsic binding specificity is found to increase with conformational disorder, along with decreasing energy roughness and increasing entropy, while the first factor dominates to eventually increase the degree of funnelness of the binding energy landscape. In the current work, we delineate this effect by means of the diffusion coefficient, whereby the conformational disorder increases the flexibility of the protein chain and therefore leads to dynamic interaction patterns in the complex concomitant with multiple binding pathways. This smooths the binding roughness and finally results in high binding specificity. By the same token, we can conclude a lower specificity for rigid docking than highly flexible binding. However, gradually increasing the partially stabilizing or pre-formed structures in IDPs' isolated states has distinct effects on

energy roughness during different binding process, so the specificity of binding is anticipated to be complicated to determine and be understood. The difficulty encountered here calls for future upgrades or completely new developments of theory, with deduction basically relying on each layer of the funnel hierarchically, rather than integrally.

Summarizing, we have addressed the position-, disorder-, and salt-dependent diffusion coefficient behavior of IDPs' binding coupled with folding to targets. Our study has filled the gap in current research regarding IDPs, where major efforts are focused/based on using a free energy approach. With the quantified diffusion coefficient, the shape of the binding energy landscape can be inferred. Our theoretical investigations provide a profound understanding of the topography of the binding energy landscape in IDPs' binding–folding and may provide guidance for biophysical experiments regarding the affinity/specificity engineering of IDPs by means of the energy landscape.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

## Notes and references

1. Bryngelson JD and Wolynes PG, Proc. Natl. Acad. Sci. U.S.A, 1987, 84, 7524–7528. [PubMed: 3478708]

2. Bryngelson JD and Wolynes PG, J. Phys. Chem, 1989, 93, 6902–6915.

3. Socci N, Onuchic JN and Wolynes PG, J. Chem. Phys, 1996, 104, 5860–5868.

4. Bryngelson JD, Onuchic JN, Socci ND and Wolynes PG, Proteins: Struct., Funct., Bioinf, 1995, 21, 167–195.

5. Shakhnovich E, Karplus M et al., nature, 1994, 369, 248. [PubMed: 7710478]

6. Das P, Moll M, Stamati H, Kavraki LE and Clementi C, Proc. Natl. Acad. Sci. U.S.A, 2006, 103, 9885–9890. [PubMed: 16785435]

7. Zheng W and Best RB, J. Phys. Chem. B, 2015, 119, 15247–15255. [PubMed: 26601695]

8. Oliveberg M and Wolynes PG, Q. Rev. Biophys, 2005, 38, 245–288. [PubMed: 16780604]

9. Neupane K, Manuel AP and Woodside MT, Nature Phys, 2016, 12, 700.

10. Kramers HA, Physica, 1940, 7, 284–304.

11. Berezhkovskii A and Szabo A, J. Chem. Phys, 2005, 122, 014503.

12. Naganathan AN, Perez-Jimenez R, MuÃ oz V and Sanchez-Ruiz JM, Phys. Chem. Chem. Phys, 2011, 13, 17064–17076. [PubMed: 21769353]

13. Pirchi M, Ziv G, Riven I, Cohen SS, Zohar N, Barak Y and Haran G, Nat. Commun, 2011, 2, 493. [PubMed: 21988909]

14. Woodside MT and Block SM, Annu. Rev. Biophys, 2014, 43, 19–39. [PubMed: 24895850]

15. Chung HS, Piana-Agostinetti S, Shaw DE and Eaton WA, Science, 2015, 349, 1504–1510. [PubMed: 26404828]

16. Nettels D, Gopich IV, Hoffmann A and Schuler B, Proc. Natl. Acad. Sci. U.S.A, 2007, 104, 2655–2660. [PubMed: 17301233]

17. Cellmer T, Henry ER, Hofrichter J and Eaton WA, Proc. Natl. Acad. Sci. U.S.A, 2008, 105, 18320–18325. [PubMed: 19020085]

18. Woodside MT, Lambert J and Beach KS, Biophys. J, 2014, 107, 1647–1653. [PubMed: 25296317]

19. Foster DA, Petrosyan R, Pyo AG, Hoffmann A, Wang F and Woodside MT, Biophys. J, 2018, 114, 1657–1666. [PubMed: 29642035]

20. Best RB and Hummer G, Phys. Rev. Lett, 2006, 96, 228104. [PubMed: 16803349]

21. Hummer G, New J Phys, 2005, 7, 34.

22. Xu W, Lai Z, Oliveira RJ, Leite VB and Wang J, J. Phys. Chem. B, 2012, 116, 5152–5159. [PubMed: 22497604]

23. Naganathan AN, Doshi U and MuÃ oz V, J. Am. Chem. Soc, 2007, 129, 5673–5682. [PubMed: 17419630]

24. Chahine J, Oliveira RJ, Leite VB and Wang J, Proc. Natl. Acad. Sci. U.S.A, 2007, 104, 14646–14651. [PubMed: 17804812]

25. Oliveira RJ, Whitford PC, Chahine J, Leite VB and Wang J, Methods, 2010, 52, 91–98. [PubMed: 20438841]

26. Oliveira RJ, Whitford PC, Chahine J, Wang J, Onuchic JN and Leite VB, Biophys. J, 2010, 99, 600–608. [PubMed: 20643080]

27. Best RB and Hummer G, Proc. Natl. Acad. Sci. U.S.A, 2010, 107, 1088–1093. [PubMed: 20080558]

28. Best RB and Hummer G, Phys. Chem. Chem. Phys, 2011, 13, 16902–16911. [PubMed: 21842082]

29. Lai Z, Zhang K and Wang J, Phys. Chem. Chem. Phys, 2014, 16, 6486–6495. [PubMed: 24605364]

30. Kubelka J, Hofrichter J and Eaton WA, Curr. Opin. Struct. Biol, 2004, 14, 76–88. [PubMed: 15102453]

31. Ansari A, Jones CM, Henry ER, Hofrichter J and Eaton WA, Science, 1992, 256, 1796–1798. [PubMed: 1615323]

32. Hagen SJ, Curr. Protein Pept. Sci, 2010, 11, 385–395. [PubMed: 20426733]

33. Wolynes PG, Proc. Natl. Acad. Sci. U.S.A, 1997, 94, 6170–6175. [PubMed: 9177189]

34. Wright PE and Dyson HJ, J. Mol. Biol, 1999, 293, 321–331. [PubMed: 10550212]

35. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW et al., J. Mol. Graph. Model, 2001, 19, 26–59. [PubMed: 11381529]

36. Tompa P, Trends Biochem. Sci, 2002, 27, 527–533. [PubMed: 12368089]

37. Dyson HJ and Wright PE, Curr. Opin. Struct. Biol, 2002, 12, 54–60. [PubMed: 11839490]

38. Uversky VN, Protein Sci, 2002, 11, 739–756. [PubMed: 11910019]

39. Zhou H-X, Trends Biochem. Sci, 2012, 37, 43–48. [PubMed: 22154231]

40. Waldauer SA, Bakajin O and Lapidus LJ, Proc. Natl. Acad. Sci. U.S.A, 2010, 107, 13713–13717. [PubMed: 20643973]

41. Soranno A, Buchli B, Nettels D, Cheng RR, Müller-Späth S, Pfeil SH, Hoffmann A, Lipman EA, Makarov DE and Schuler B, Proc. Natl. Acad. Sci. U.S.A, 2012, 109, 17800–17806. [PubMed: 22492978]

42. Soranno A, Holla A, Dingfelder F, Nettels D, Makarov DE and Schuler B, Proc. Natl. Acad. Sci. U.S.A, 2017, 114, E1833–E1839. [PubMed: 28223518]

43. Shoemaker BA, Portman JJ and Wolynes PG, Proc. Natl. Acad. Sci. U.S.A, 2000, 97, 8868–8873. [PubMed: 10908673]

44. Chu X, Gan L, Wang E and Wang J, Proc. Natl. Acad. Sci. U.S.A, 2013, 110, E2342–E2351. [PubMed: 23754431]

45. Chu X and Wang J, PLoS Comput. Biol, 2014, 10, e1003782. [PubMed: 25144525]

46. Wang J, Oliveira RJ, Chu X, Whitford PC, Chahine J, Han W, Wang E, Onuchic JN and Leite VB, Proc. Natl. Acad. Sci. U.S.A, 2012, 109, 15763–15768. [PubMed: 23019359]

47. Clementi C, Nymeyer H and Onuchic JN, J. Mol. Biol, 2000, 298, 937–953. [PubMed: 10801360]

48. Levy Y, Wolynes PG and Onuchic JN, Proc. Natl. Acad. Sci. U.S.A, 2004, 101, 511–516. [PubMed: 14694192]

49. Turjanski AG, Gutkind JS, Best RB and Hummer G, PLoS Comput. Biol, 2008, 4, e1000060. [PubMed: 18404207]

50. Wang J, Wang Y, Chu X, Hagen SJ, Han W and Wang E, PLoS Comput. Biol, 2011, 7, e1001118. [PubMed: 21490720]

51. Ganguly D and Chen J, Proteins: Struct., Funct., Bioinf, 2011, 79, 1251–1266.

52. De Sancho D and Best RB, Mol. Biosyst, 2012, 8, 256–267. [PubMed: 21892446]

53. Li P, Oliva FY, Naganathan AN and MuÃ oz V, Proc. Natl. Acad. Sci. U.S.A, 2009, 106, 103–108. [PubMed: 19118204]

54. Radhakrishnan I, Pérez-Alvarado GC, Dyson HJ and Wright PE, FEBS Lett, 1998, 430, 317–322. [PubMed: 9688563]

55. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR and Wright PE, Cell, 1997, 91, 741–752. [PubMed: 9413984]

56. Zor T, De Guzman RN, Dyson HJ and Wright PE, J. Mol. Biol, 2004, 337, 521–534. [PubMed: 15019774]

57. Sobolev V, Sorokine A, Prilusky J, Abola EE and Edelman M, Bioinformatics, 1999, 15, 327–332. [PubMed: 10320401]

58. Cho SS, Levy Y and Wolynes PG, Proc. Natl. Acad. Sci. U.S.A, 2006, 103, 586–591. [PubMed: 16407126]

59. Tribello GA, Bonomi M, Branduardi D, Camilloni C and Bussi G, Comput. Phys. Commun, 2014, 185, 604–613.

60. Zagrovic B, Snow CD, Khaliq S, Shirts MR and Pande VS, J. Mol. Biol, 2002, 323, 153–164. [PubMed: 12368107]

61. Huang Y and Liu Z, J. Mol. Biol, 2009, 393, 1143–1159. [PubMed: 19747922]

62. Umezawa K, Ohnuki J, Higo J and Takano M, Proteins: Struct., Funct., Bioinf, 2016, 84, 1124–1133.

63. Azia A and Levy Y, J. Mol. Biol, 2009, 393, 527–542. [PubMed: 19683007]

64. Givaty O and Levy Y, J. Mol. Biol, 2009, 385, 1087–1097. [PubMed: 19059266]

65. Chu X, Wang Y, Gan L, Bai Y, Han W, Wang E and Wang J, PLoS Comput. Biol, 2012, 8, e1002608. [PubMed: 22807669]

66. Chu X, Liu F, Maxwell BA, Wang Y, Suo Z, Wang H, Han W and Wang J, PLoS Comput. Biol, 2014, 10, e1003804. [PubMed: 25188490]

67. Berendsen HJC, van der Spoel D and van Drunen R, Comput. Phys. Commun, 1995, 91, 43–56.

68. Noel JK, Whitford PC, Sanbonmatsu KY and Onuchic JN, Nucleic Acids Res, 2010, 38, W657–W661. [PubMed: 20525782]

69. Sugita Y and Okamoto Y, Chem. Phys. Lett, 1999, 314, 141–151.

70. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH and Kollman PA, J. Comput. Chem, 1992, 13, 1011–1021.

71. Hummer G, J. Chem. Phys, 2004, 120, 516–523. [PubMed: 15267886]

72. Best RB and Hummer G, Proc. Natl. Acad. Sci. U.S.A, 2005, 102, 6732–6737. [PubMed: 15814618]

73. Hummer G, García AE and Garde S, Phys. Rev. Lett, 2000, 85, 2637. [PubMed: 10978126]

74. Chen T and Chan HS, PLoS Comput. Biol, 2015, 11, e1004260. [PubMed: 26016652]

75. Oâ neill P, Nicolaides D, Honnery D, Soria J et al., 15th Australasian Fluid Mechanics Conference, 2004, pp. 13–17.

76. Zwanzig R, Proc. Natl. Acad. Sci. U.S.A, 1988, 85, 2029–2030. [PubMed: 3353365]

77. Yang S, Levine H and Onuchic JN, J. Mol. Biol, 2005, 352, 202–211. [PubMed: 16061250]

78. Sugase K, Dyson HJ and Wright PE, Nature, 2007, 447, 1021–1025. [PubMed: 17522630]

79. Dahal L, Kwan TO, Shammas SL and Clarke J, Biophysical journal, 2017, 113, 2713–2722. [PubMed: 29262364]

80. Tompa P and Fuxreiter M, Trends Biochem. Sci, 2008, 33, 2–8. [PubMed: 18054235]

81. Fuxreiter M, Simon I and Bondos S, Trends Biochem. Sci, 2011, 36, 415–423. [PubMed: 21620710]

82. Fuxreiter M, J. Mol. Biol, 2018.

83. Chu X and Muñoz V, Phys. Chem. Chem. Phys, 2017, 19, 28527–28539. [PubMed: 29044255]

84. Clementi C and Plotkin SS, Protein Sci, 2004, 13, 1750–1766. [PubMed: 15215519]

85. Levy Y, Onuchic JN and Wolynes PG, J. Am. Chem. Soc, 2007, 129, 738–739. [PubMed: 17243791]

86. Zhou H-X, Pang X and Lu C, Phys. Chem. Chem. Phys, 2012, 14, 10466–10476. [PubMed: 22744607]

87. Shammas SL, Travis AJ and Clarke J, J. Phys. Chem. B, 2013, 117, 13346–13356. [PubMed: 23875714]

88. Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, Tyagi S, Clarke J, Shammas SL, Blackledge M et al., Cell, 2015, 163, 734–745. [PubMed: 26456112]

89. Uversky VN, Gillespie JR and Fink AL, Proteins: Struct., Funct., Bioinf, 2000, 41, 415–427.

90. Ganguly D, Otieno S, Waddell B, Iconaru L, Kriwacki RW and Chen J, J. Mol. Biol, 2012, 422, 674–684. [PubMed: 22721951]

91. Rogers JM, Steward A and Clarke J, J. Am. Chem. Soc, 2013, 135, 1415–1422. [PubMed: 23301700]

92. Wang Y, Chu X, Longhi S, Roche P, Han W, Wang E and Wang J, Proc. Natl. Acad. Sci. U.S.A, 2013, 110, E3743–E3752. [PubMed: 24043820]

93. Shammas SL, Travis AJ and Clarke J, Proc. Natl. Acad. Sci. U.S.A, 2014, 111, 12055–12060. [PubMed: 25092343]

94. Iešmantavi ius V, Dogan J, Jemth P, Teilum K and Kjaergaard M, Angew. Chem. Int. Ed. Engl, 2014, 53, 1548–1551. [PubMed: 24449148]

95. Arai M, Sugase K, Dyson HJ and Wright PE, Proc. Natl. Acad. Sci. U.S.A, 2015, 112, 9614–9619. [PubMed: 26195786]

96. Ngo JCK, Giang K, Chakrabarti S, Ma C-T, Huynh N, Hagopian JC, Dorrestein PC, Fu X-D, Adams JA and Ghosh G, Molecular cell, 2008, 29, 563–576. [PubMed: 18342604]

97. Miller MA and Wales DJ, J. Chem. Phys, 1999, 111, 6610–6616.

98. Neelamraju S, Gosavi S and Wales DJ, J. Phys. Chem. B, 2018, 122, 12282–12291. [PubMed: 30495947]

99. Zor T, Mayr BM, Dyson HJ, Montminy MR and Wright PE, J. Biol. Chem, 2002, 277, 42241–42248. [PubMed: 12196545]

100. Wong ET, Na D and Gsponer J, PLOS Comput. Biol, 2013, 9, e1003192. [PubMed: 23990768]

101. Ganguly D, Zhang W and Chen J, PLoS Comput. Biol, 2013, 9, e1003363. [PubMed: 24278008]

102. Mao AH, Crick SL, Vitalis A, Chicoine CL and Pappu RV, Proc. Natl. Acad. Sci. U.S, 2010, 107, 8183–8188.

103. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, Nettels D and Schuler B, Proc. Natl. Acad. Sci. U.S.A, 2010, 107, 14609–14614. [PubMed: 20639465]

104. England JL and Haran G, Proc. Natl. Acad. Sci. U.S.A, 2010, 107, 14519–14520. [PubMed: 20682745]

105. Das RK and Pappu RV, Proc. Natl. Acad. Sci. U.S.A, 2013, 110, 13392–13397. [PubMed: 23901099]

106. Wicky BI, Shammas SL and Clarke J, Proc. Natl. Acad. Sci. U.S.A, 2017, 114, 9882–9887. [PubMed: 28847960]

107. Spolar RS and Record MT Jr, Science, 1994, 263, 11. [PubMed: 17748337]

108. Kriwacki RW, Hengst L, Tennant L, Reed SI and Wright PE, Proc. Natl. Acad. Sci. U.S.A, 1996, 93, 11504–11509. [PubMed: 8876165]

109. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C and Villafranca JE, Pac. Symp. Biocomput, 1998, 3, 473–484.

110. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST and Rose PW, Curr. Opin. Struct. Biol, 2002, 12, 197–203. [PubMed: 11959497]

111. Wang J, Zheng X, Yang Y, Drueckhammer D, Yang W, Verkhivker G and Wang E, Phys. Rev. Lett, 2007, 99, 198101. [PubMed: 18233118]
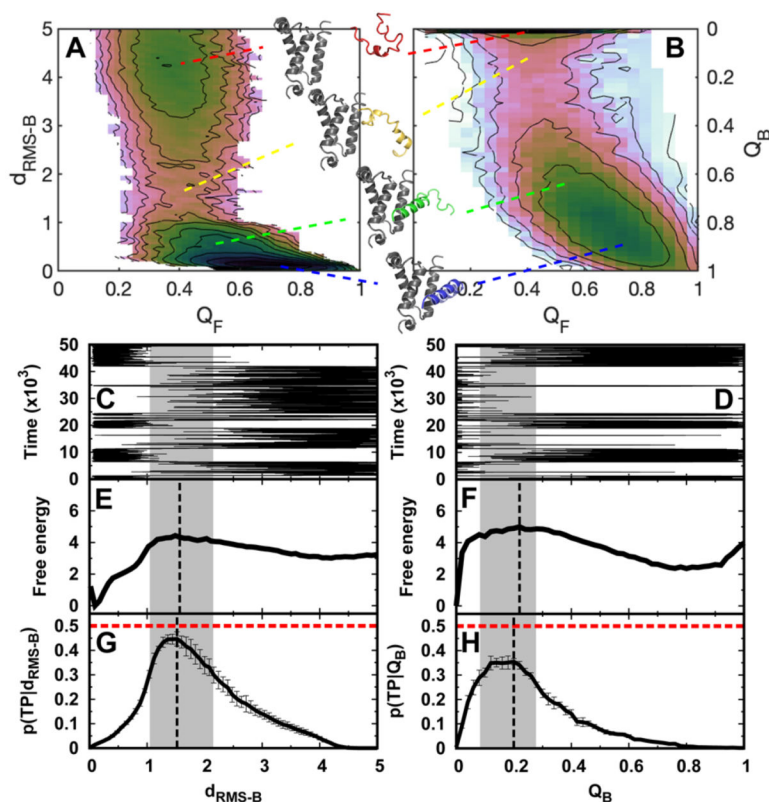
112. Yan Z, Zheng X, Wang E and Wang J, Chem. Sci, 2013, 4, 2387–2395.

**Fig. 1. Trajectories, transition path analyses, and free energy landscapes of binding-coupled-folding of pKID to KIX.**

2D binding–folding free energy landscapes are calculated from REMD simulations and projected along (A) $d_{RMS-B}$, $Q_F$ and (B) $Q_B$, $Q_F$. $d_{RMS-B}$ is the root-meansquare deviation of the distances of binding native contact pairs to those in the native structure, while $Q_B$ and $Q_F$ are the fractions of binding and folding native contacts for pKID, respectively. The typical structures of pKID-KIX are shown corresponding to the indicated binding stages. A sample binding trajectory (C, D), 1D free energy landscape (E, F), and transition path analyses (G, H) are shown along $d_{RMS-B}$ and $Q_B$, respectively. $p(TP|x)$ is the conditional probability of being on a transition path, where $x$ is the reaction coordinate. The dashed vertical lines in free energy landscapes and $p(TP|x)$ are plotted according to the maximum values as an identification of the transition state. The values at the peak of $p(TP|x)$ for $d_{RMS-B}$ and $Q_B$ are 0.45±0.03 and 0.35±0.01, respectively. The transition state locations are very similar, according to different identifications of the barrier of free energy landscapes and the peak of $p(TP|x)$. In detail, for $d_{RMS-B}$, the transition states' $d_{RMS-B}$ are 1.49 and 1.52 from the free energy landscape and $p(TP|x)$, respectively, while for $Q_B$, the transition states' $Q_B$ are 0.22 and 0.20 from the free energy landscape and $p(TP|x)$, respectively. We denote the transition state ensembles that have $p(TP|x)$ higher than 0.30 by gray shadow regions. The errors of $p(TP|x)$ are calculated by analyzing different trajectories. A total of 1108 and 1099 (un)binding transition paths are observed along $d_{RMS-B}$ and $Q$, respectively. Time is in reduced units, $d_{RMS-B}$ is in units of $nm$, and free energy is in units of $kT_B$, where $T_B$ is the binding temperature.
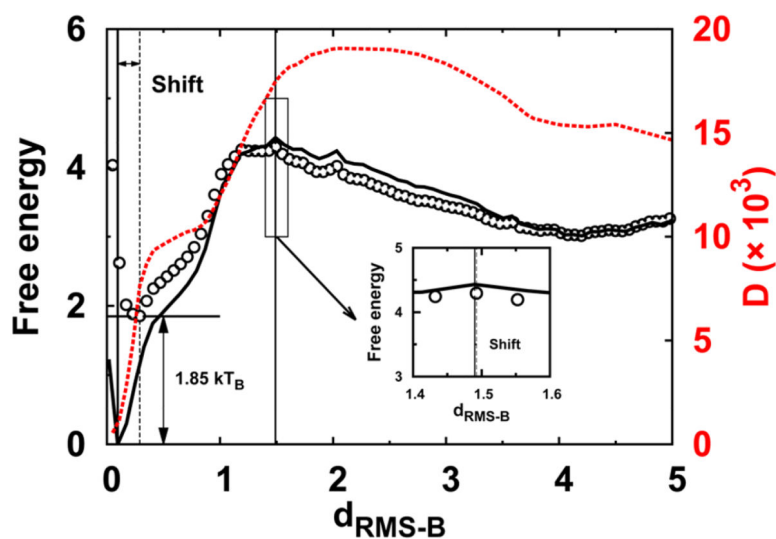
**Fig. 2. Position-dependent diffusion coefficient and free energy landscapes.**
1D binding thermodynamic (solid line), effective (circles) free energy landscapes, and the diffusion coefficient $D$ are plotted along $d_{RMS-B}$. The insert plot indicates the slight height and position shift of the transition state between the thermodynamic and effective free energy landscapes. The free energy landscape is obtained from thermodynamic REMD simulations, and $D$ is calculated from restraining simulations, at constant binding temperature. $d_{RMS-B}$ is in units of $nm$, and free energy is in units of $kT_B$, where $T_B$ is the binding temperature. $D$ is in units of $nm^2/time$
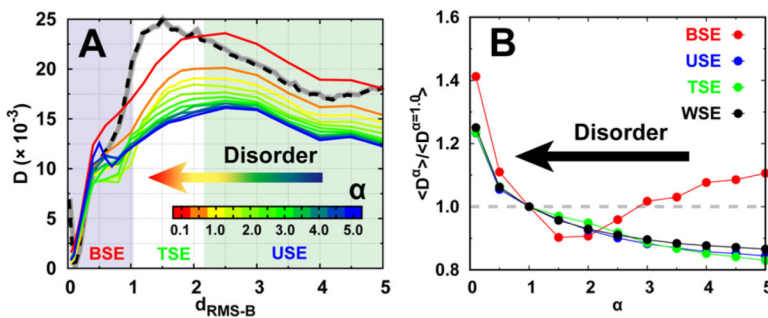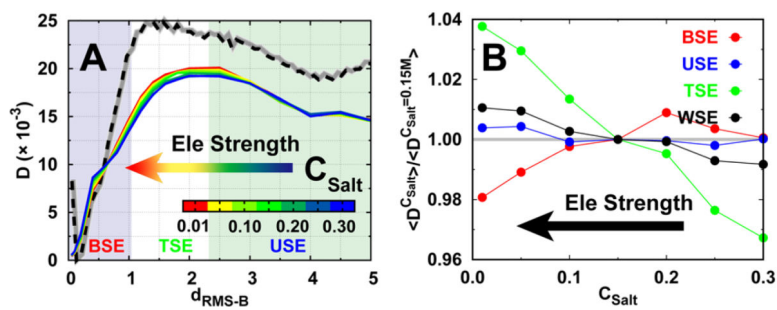
**Fig. 3. Diffusion coefficient at different degrees of conformational disorder.**
(A) The position-dependent $D$ along $d_{RMS-B}$. Free energy landscape with default disorder parameter $\alpha = 1.0$ at binding temperature is shown with dashed line as a guidance of the binding process, which can be further divided into three stages: the binding state ensemble (BSE), transition state ensemble (TSE), and unbinding state ensemble (USE), based on the transition path analysis shown in Fig. 1E. (B) The ratio between the mean $D(d_{RMS-B})$ of different degrees and default ($\alpha = 1.0$) parameter of conformational disorder for different binding stages. "WSE" is an acronym for the "whole state ensemble".
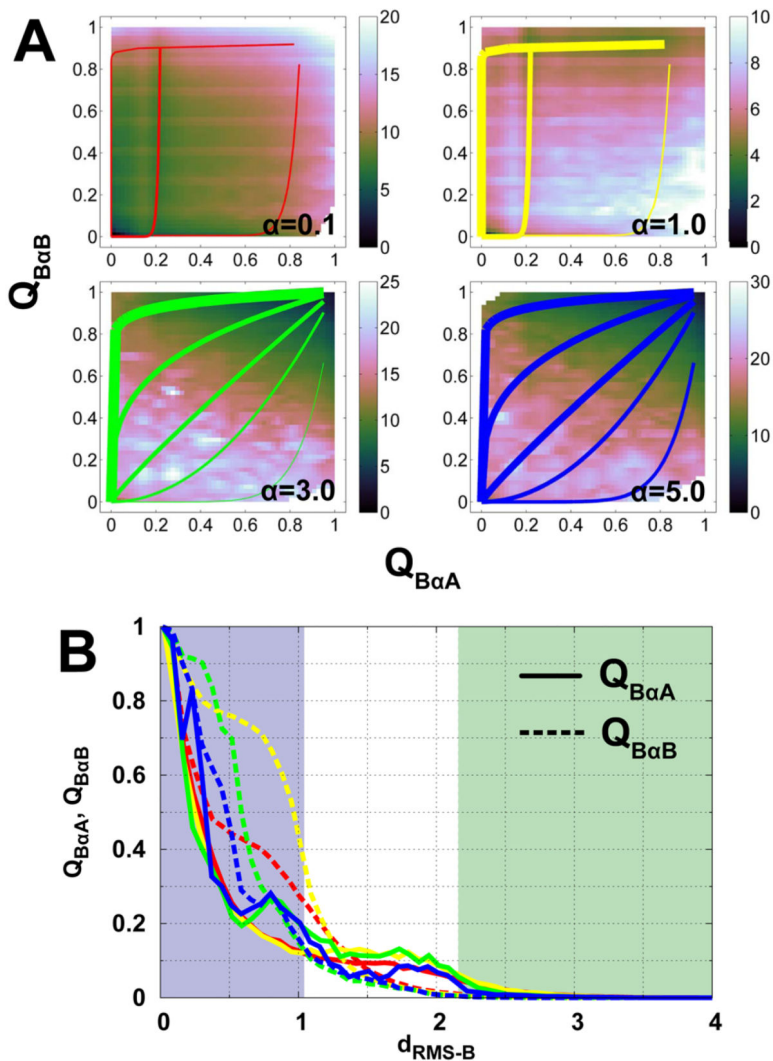
**Fig. 4. Binding mechanisms at different degrees of conformational disorder.**
(A) 2D binding free energy landscapes projecting onto $Q_{B\alpha A}$ and $Q_{B\alpha B}$. $Q_{B\alpha A}$ and $Q_{B\alpha B}$ are the fractions of native binding contacts of helices $\alpha A$ and $\alpha B$ of pKID to KIX, respectively. The free energy landscapes of $\alpha$=0.1, 1.0, 3.0, and 5.0 are plotted. The lines in each panel illustrate pathways, with thick ones indicating large flux and vice versa. (B) Evolutions of binding contacts of helices $\alpha A$ and $\alpha B$ of pKID along $d_{RMS-B}$. Solid and dashed lines are $Q_{B\alpha A}$ and $Q_{B\alpha B}$, Respectively, and different colored lines correspond to different degrees of conformational disorder. The color of shadows and lines follow the same scheme used in Fig. 3A.

**Fig. 5. Diffusion coefficient at different strengths of electrostatic interactions.**
(A) The position-dependent $D$ along $d_{RMS-B}$. Free energy landscape with moderate salt concentration $C_{Salt}$=0.15 M at binding temperature is shown with dashed line as a guidance of the binding process (B) The ratio between the mean $D(d_{RMS-B})$ of different salt concentrations and moderate salt concentration ($C_{Salt}$=0.15$M$) for different binding stages.
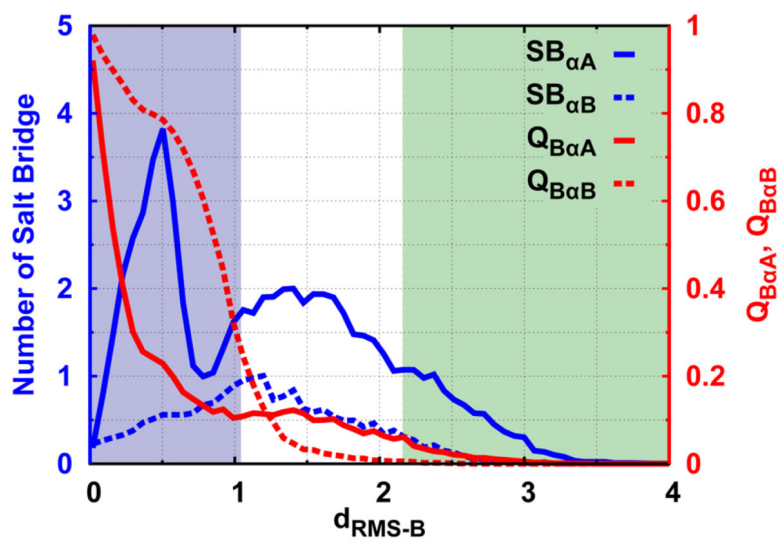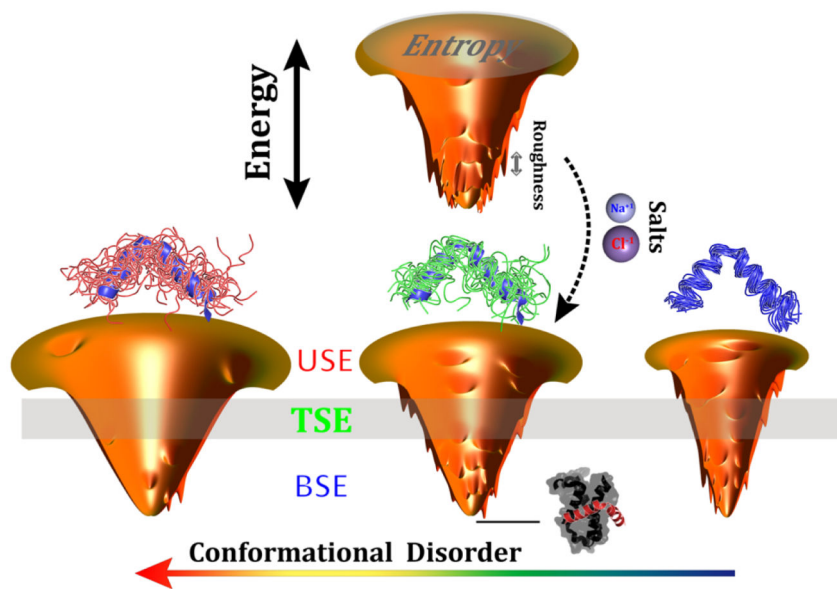
**Fig. 6. Non-native salt bridges and native contact formations of the two helices during binding at moderate salt concentration**

$C_{salt} = 0.15 M$. SB is an acronym for "salt bridge".

**Fig. 7. Scheme illustrating the effects of conformational disorder and salt concentrations on the topography of binding energy landscapes.**

The deepness, the size, and the roughness of the funnels represent the energy gap, entropy, and energy roughness of the energy landscapes, respectively. The single funnel at the top shows binding with moderate conformational disorder and low salt concentration (strong electrostatic interactions). The three funnels at the bottom show the change according to the conformational disorder. Conformational ensembles of pKID are shown on the top of each funnel, with the native structure colored blue. The bound complex of pKID-KIX is shown at the basin of the funnel, indicating the destination of binding.