Multi-Scaled Explorations of Binding-Induced Folding of Intrinsically Disordered Protein Inhibitor IA3 to its Target Enzyme

Jin Wang[1,2,3]*, Yong Wang[1], Xiakun Chu[2], Stephen J. Hagen[4], Wei Han[2], Erkang Wang[1]*

1 State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, People's Republic of China, 2 College of Physics, Jilin University, Changchun, Jilin, People's Republic of China, 3 Department of Chemistry, Physics and Applied Mathematics, State University of New York at Stony Brook, Stony Brook, New York, United States of America, 4 Department of Physics, University of Florida, Gainesville, Florida, United States of America

Abstract

Biomolecular function is realized by recognition, and increasing evidence shows that recognition is determined not only by structure but also by flexibility and dynamics. We explored a biomolecular recognition process that involves a major conformational change – protein folding. In particular, we explore the binding-induced folding of IA3, an intrinsically disordered protein that blocks the active site cleft of the yeast aspartic proteinase saccharopepsin (YPrA) by folding its own N-terminal residues into an amphipathic alpha helix. We developed a multi-scaled approach that explores the underlying mechanism by combining structure-based molecular dynamics simulations at the residue level with a stochastic path method at the atomic level. Both the free energy profile and the associated kinetic paths reveal a common scheme whereby IA3 binds to its target enzyme prior to folding itself into a helix. This theoretical result is consistent with recent time-resolved experiments. Furthermore, exploration of the detailed trajectories reveals the important roles of non-native interactions in the initial binding that occurs prior to IA3 folding. In contrast to the common view that non-native interactions contribute only to the roughness of landscapes and impede binding, the non-native interactions here facilitate binding by reducing significantly the entropic search space in the landscape. The information gained from multi-scaled simulations of the folding of this intrinsically disordered protein in the presence of its binding target may prove useful in the design of novel inhibitors of aspartic proteinases.

Introduction

"Intrinsically Disordered Proteins" (IDPs) are proteins that are disordered either in whole or in part. They play important roles in various cellular functions, including regulation, signaling and control processes [1]. Bioinformatic and statistical studies show that many proteins are intrinsically disordered: Of the crystal structures in the Protein Data Bank that contain no missing electron density, only about 30 percent show completely ordered structures [2,3]. From this perspective, biological function may not require ordered structure. A key question is then, how do intrinsically disordered proteins carry out biological function?

Experiment and theory are beginning to probe the relationship between the dynamics and function of highly flexible IDPs [1,4–12]. The intrinsically disordered proteinase inhibitor IA3, found in the cytoplasm of *Saccharomyces cerevisiae*, is an inhibitor of the protein vacuolar yeast peptidease A (YPrA). YPrA, which is also known as saccharopepsin [13], is a member of the aspartic proteinase family. The aspartic proteinases are present in many species, including vertebrates, fungi, plants and retroviruses [14], and they play a role in a range of pathologies that includes Alzheimers disease, hypertension, malaria and AIDS [15,16].

Until recently, few peptide inhibitors of aspartic proteinases were known [17]. Even fewer structures of inhibitor-enzyme complexes have been determined. One complex that has been studied is that of the yeast peptidase A with its naturally occurring peptide inhibitor, IA3 [18]. Free IA3 is a 68-residue peptide that lacks a stable structure in solution. Upon interaction with with YPrA, the N-terminal region of IA3 folds into an amphipathic helix that blocks the active site cleft of the enzyme. [19–21]. Therefore, IA3 undergoes a major disordered-to-ordered transition during binding to its target enzyme. Understanding this transition and the mechanism of IA3's interaction with YPrA may provide clues as to how IDPs regulate their function through dynamics.

Narayanan and coworkers recently used laser temperature-jump fluorescence spectroscopy and fluorescence resonance energy transfer (FRET) to investigate the kinetics of the binding-induced folding of IA3 with YPrA [22]. A rapid kinetic relaxation in IA3 was observed in the presence of YPrA, whereas this process was absent in free IA3. Modeling of the kinetic data for both free IA3 and the IA3/YPrA complex indicated that unfolded IA3 binds with YPrA prior to forming its N-terminal helix. The present work uses a multi-scaled simulation approach to explore the binding of N terminal IA3 to YPrA. (The structure of the C-terminus in the

Author Summary

The intrinsically disordered peptide IA3 is the endogenous inhibitor for the enzyme named yeast aspartic proteinase saccharopepsin (YPrA). In the presence of YPrA, IA3 folds itself into an amphipathic helix that blocks the active site cleft of the enzyme. We developed a multi-scaled approach to explore the underlying mechanism of this binding-induced ordering transition. Our approach combines a structure-based molecular dynamics model at the residue level with a stochastic path method at the atomic level. Our simulations suggest that IA3 inhibits YPrA through an induced-fit mechanism where the enzyme (YPrA) induces conformational change of its inhibitor (IA3). This expands the definition of an induced-fit model from its original meaning that the binding of substrate (IA3) drives conformational change in the protein (YPrA). Our result is consistent with recent kinetic experiments and provides a microscopic explanation for the underlying mechanism. We also discuss the important roles of non-native interactions and backtracking. These results enrich our understanding of the enzyme-inhibition mechanism and may have value in the design of drugs.

bound complex is unknown.) Although molecular dynamics (MD) simulation is a powerful tool for investigating biomolecules, the time scales for the IA3/YPrA folding and binding interaction are too long for simulation in atomic detail by MD, at least at the present time. In order to bridge the gap of time scales between experiment and computation, several approaches have been developed that reduce the number of degrees of freedom. One method is to construct a structure-based energy function at a coarse-grained residue level [23]. A second method is to identify and quantify the optimal kinetic paths between the initial disordered and final ordered native states [24,25]. The optimal paths are those paths that connect the reactant and product on the potential energy landscape surface with the largest statistical weight [26,27]. In this work, we first carry out a structure-based coarse-grained residue level study of IA3 binding and folding. This step uncovers the underlying thermodynamics of the binding-folding free energy landscape. We then identify several optimal paths of IA3 binding to YPrA, as initiated from different starting points, based on a fully atomistic description of the protein. We address the effect of non-native and native interactions on the binding-folding of IA3. We obtain results that are consistent with the experimental findings [22]. This multi-scaled approach provides a detailed dynamic picture of the folding of a natural peptide inhibitor in the presence of its target enzyme.

Results

Coarse Grained Free Energy Landscape

In order to understand the binding-folding process from a global thermodynamic perspective, we explored the free energy landscape with a coarse-grained structure-based model by MD simulation under constant temperature. In this work, the simulation temperature is chosen to be lower than the binding transition temperature so that binding is possible and the target enzyme is stable. Meanwhile a harmonic biasing potential is introduced to accelerate the sampling. The harmonic biasing potential serves two purposes: (a) It prevents the IA3 molecule from being too physically distant from YPrA. This approach prevents the molecule from consuming too much computational time wandering in free space and searching for its interaction

partner. It is analogous to simulating the system in a highly crowded cell-like enviroment where IA3 has higher chances of colliding with YPrA [5]. (b) The harmonic bias also facilitates crossing of the energy barrier by elevating the free energy basin of the complex. The biasing potential enhances sampling by minimizing trapping in less probable states. This idea is similar to the conformational flooding algorithm [28]. Finally, we can find the unbiased thermodynamic properties from our simulations by transforming back from the biased to the unbiased case, using Equation 2 of the *Methods*.

We take the normalized native contact fraction Q_f for folding of IA3 and the center of mass (R_{COM}) distance between IA3 and YPrA as the order parameters that quantify the progress of the folding and binding process towards the final conformation of the YPrA-IA3 complex. The free energy profile shown in Figure 1 suggests there are two stable configurations: one is the unfolded and unbound state of IA3 and the other is the native binding-folding complex. The transition state ensemble corresponds to the region where Q_f of IA3 is in the range [0.3–0.5] and R_{COM} is in the range [2.2–2.5] nm. The finding that the unbound state corresponds to a nonzero Q_f, so that IA3 is not entirely disordered in unbound state, is consistent with NMR and CD measurements [17,21,22], which indicate that the N-terminus of IA3 is approximately 15% folded when the peptide is free in solution. The fact that our result for the Q_f of the unbound state is larger than this value may reflect the fact that our chosen order parameter for the native contact fraction is not very sensitive to the fluctuations in the local contacts within the helical structure of IA3. We also measured the RMSD between the unbound and helical states of IA3. The average RMSD of $7.1\mathring{A}$ (from 31 C_α atoms) reflects the unstructured character of IA3 in unbound state. Overall, the coarse grained simulation reproduced the experimental properties of the system in a qualitative or semi-quantitative way. The free energy surface in Figure 1 indicates that binding and folding of IA3 are decoupled, with no folding occuring as the system approaches the transition state region. After the transition state however the binding and folding become strongly coupled. IA3 first approaches YPrA through binding from distant initial positions, then overcomes the transition state barrier, and finally folds itself into the structured conformation. Binding precedes folding.

Transition State and Key Residues Analysis

From the free energy profile in Figure 1 we can conclude that IA3 binds prior to folding. Here we address the question of which regions of YPrA interact with IA3 at the transition state.

We captured the contacts between IA3 and YPrA by using the cutoff algorithm instead of counting only the native contacts Q_b. Figure 2A shows that the interfacial contacts at the transition state are distributed widely with low populations. Many of these contacts do not coincide with the native contacts (labelled by red square points) in the PDB structure of the IA3/YPrA complex. This implies that the transition state may be characterized by many non-native contacts and only a few native contacts. The important role of non-native interactions in the early stages of IA3 binding to YPrA can not be captured quantitatively by the structure based residue-level model, but it is explored in our full-atomic model, which uses a physics-based force field whose energy function combines the AMBER and OPLS force fields. Figure 2B shows the distribution of interfacial contacts in the transition state. Contacts are mostly formed at the surface of active site groove of YPrA, which is shown in blue in the cartoon representation. This distribution shows unambiguously that the first stage of the interaction involves IA3 binding to the surface of the active site groove. The highest peak, colored in red for emphasis, corresponds to the "flap" region, a β hairpin loop formed by
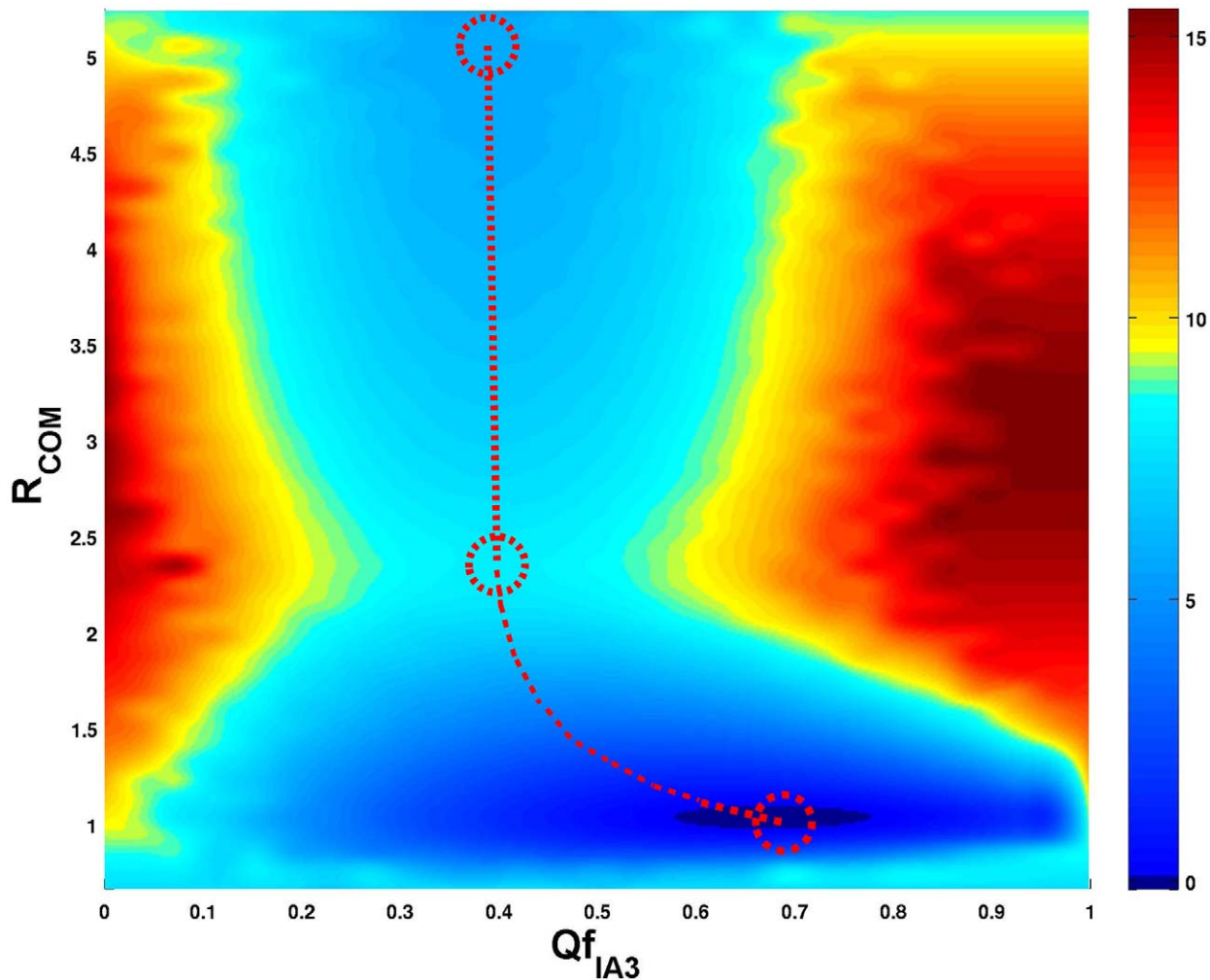
Figure 1. Unbiased free energy profile in terms of the IA3 folding coordinate (Qf_{IA3}) and the center of mass distance between YPrA and IA3 (R_{COM} in nm), as derived from the structure-based model at the residue level.
doi:10.1371/journal.pcbi.1001118.g001

residues 72–82, which project out to cover the YPrA active site. This structural motif is commonly found in aspartic peptidases [15].

At our simulation temperature, YPrA is not a completely rigid partner in IA3 folding and binding. Figure 2C shows the effect of temperature in the RMS fluctuation in several local regions of YPrA. X-ray experiments [29] also show that the electron density is poor at the two loop regions marked with red squares in the figure. These two loop regions are the "flap" (or loop1) and a second region, named loop2. Comparing Figure 2B with Figure 2C shows a role for the "flap" region in controling IA3 binding to YPrA. The "flap" region forms the most contacts with IA3 although the RMS fluctuation data does not indicate a large capture radius. By contrast, loop2 has a largest capture radius as reflected by its structural fluctuation during binding, but it does not contribute to the interfacial contacts with IA3. Remarkably, at the tip of the flap, there is one absolutely conserved tyrosine (Tyr75) that is considered to play a crucial role in the capture and cleavage of substrates [30].

What Happens after Binding?

To gain further insight into the process that follows IA3 binding with the surface of active site groove, we investigated the distribution of the native interfacial contact fraction of individual IA3 residues

(Q_b) along the binding routes. In the crystal structure of the complex, the hydrophilic face of IA3 is oriented toward the solvent. The other face of IA3 is composed of the nine hydrophobic amino acid residues, V8, I11, F12, L19, A23, V25, V26, A29 and F30. This face is enveloped completely with the residues of the YPrA active site cleft and consists of three hydrophobic clusters: "cluster-1" (red) of V8-X-X-I11-F12 in the N-terminal, "cluster-2" (green) of L19-X-X-X-A23 in the mid region, and the C-terminal "cluster-3" (yellow) of V26-X-X-A29-F20 (see Figure 5 in Text S1). These clusters are indicated in Figure 3, which shows the evolution of the native interfacial contact fraction (Q_b) of individual IA3 residues. We find that Q_b is well-distributed and less than 0.2 at the transition state region. By following the evolution of distribution along the binding routes we see that the mid region of IA3 forms native contacts with YPrA first, followed by the C-terminal region, and finally the N-terminus. However, the distribution of IA3 intrachain contacts does not show a sequential order of IA3 folding. It seems that the folding of IA3 does not necessarily occur from a particular nucleation site.

Binding-Folding Path Revealed in Atomic Details

In studying protein folding and binding, the Q score (defined in the Text S1) for structural similarity has been extensively used as a
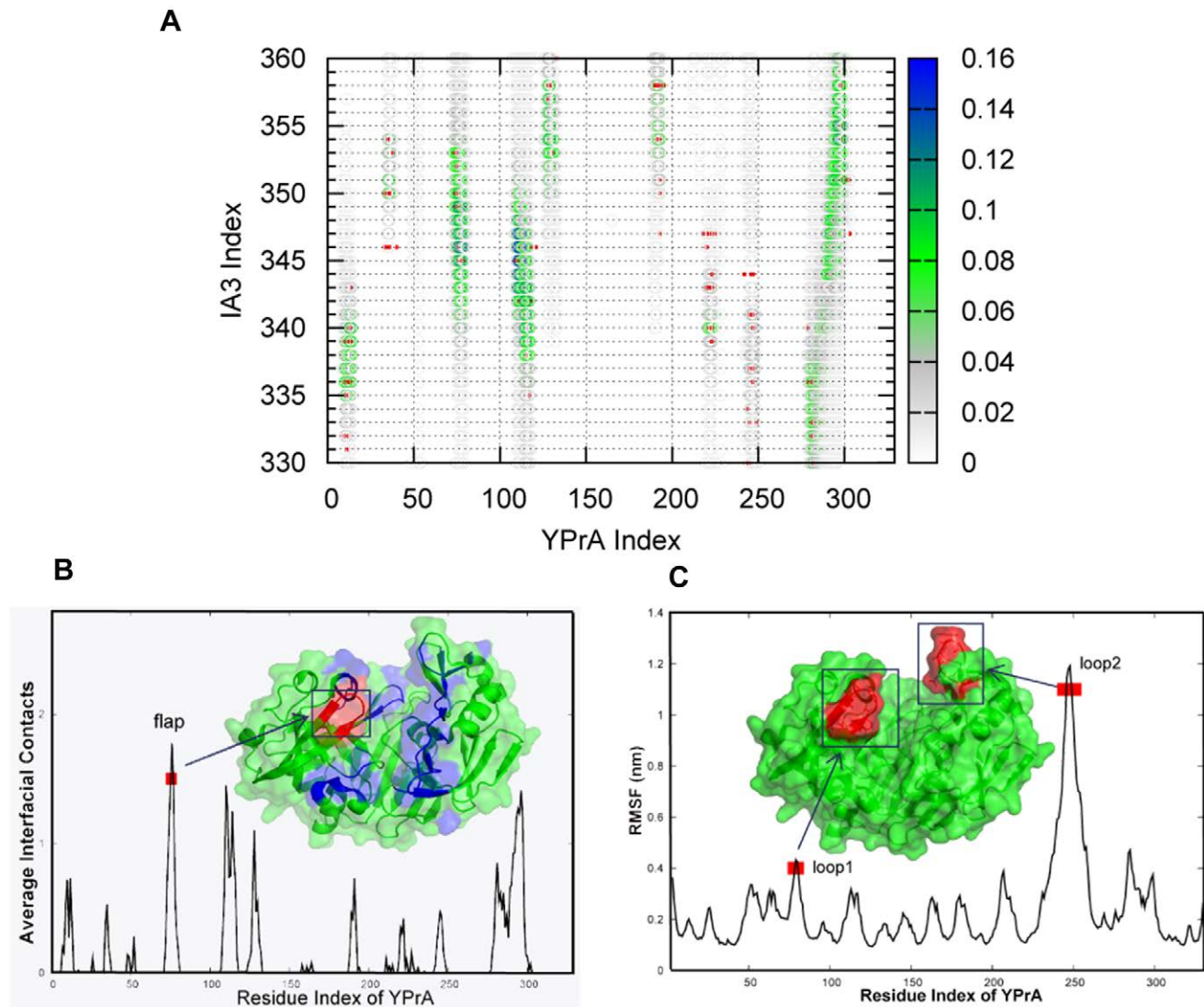
Figure 2. Transition state analysis obtained from coarse-grained structure-based model. (A) Binding contact map based on cut-off algorithm instead of Q_b. Native contacts are indicated by red squares. (B) Distribution of the average number of interfacial contacts. Contacts are primarily formed at the surface of active site groove of YPrA. The contacting residues are colored in blue in the cartoon. (C) RMS fluctuation of the C_α atoms as a function of residue number, as found by MD simulation. This provides a measure of the flexibility of local regions of YPrA.
doi:10.1371/journal.pcbi.1001118.g002

structural reaction coordinate [31–34]. Q represents the fraction of native contacts that have been formed and it characterizes the structure's similarity to a referenced structure. Here, the referenced structure is the crystal structure of the enzyme-inhibitor complex of YPrA-IA3 (PDB code: 1DP5). To monitor the folding and binding of IA3 interacting with YPrA in a fully atomistic description, optimal kinetic paths were calculated in order to determine the most probable pathways between the beginning and ending points. The optimal paths depend on the choice of initial and end points. For the end point we use the structure of the native IA3-YPrA complex, as resolved by xray crystallography. The initial point is disordered, unfolded IA3 and uninhibited, folded YPrA. Obviously, the initial point for IA3 should consist of an ensemble of conformations with a sufficient number of degrees of freedom. Unfolded conformations of IA3 were generated by molecular dynamics with explicit solvent at high temperature. Three paths were chosen to illustrate the folding and binding process in detail. We refer to them as *path*1, *path*2 and *path*3.

Figure 4 shows *path*1. Q_i and Q_y are the intrachain contact Q scores of IA3 and YPrA, respectively. The Q score of the interaction between them (Q_{inter}) represents the interfacial similarity relative to native binding complex of IA3 and YPrA. Figure 4A shows that the Q_y curve increases slowly close to 1, corresponding to the native inhibited structure. YPrA does not move much although it manifests some flexibility to accomodate the folding of IA3, mostly in the two loop regions located on the surface of the active site groove. Figure 4B shows the evolution of the folding score Q_i and binding score Q_{inter} along the path. Q_i does not vary much when Q_{inter} is less than 0.35. It even decreases slightly (IA3 unfolds) before grid 70 (where folding begins) due to backbone movements, not helix formation and breaking. This is consistent with the experimental indications that the pre-equilibrium of folding may not be helpful to IA3 folding through binding [22]. Q_{inter} increases more and more along the path, especially when it exceeds the Q_i of IA3 after grid 64 and reaches 0.5 at grid 80. It implies that IA3 does not fold itself before binding
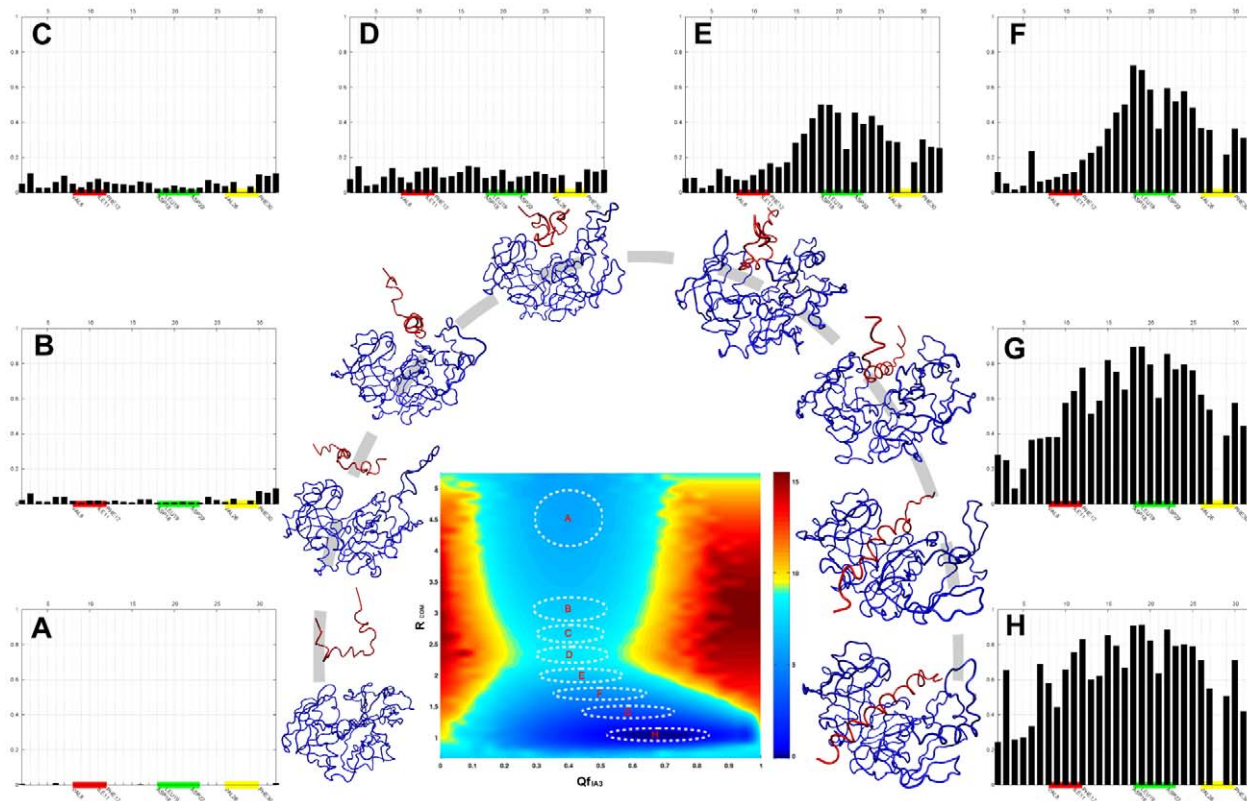
Figure 3. The evolution of the fraction of native interfacial contacts of N-terminal IA3 along the binding routes, as obtained by the coarse-grained structure-based model. Nine hydrophobic amino acid residues in IA3 (V8, I11, F12, L19, A23, V25, V26, A29 and F30) form three hydrophobic clusters. These are "cluster-1" (red), consisting of V8- X-X-I11-F12 in the N-terminus; "cluster-2" (green), consisting of L19-X-X-X-A23 in the mid region, and the C-terminal "cluster-3" (yellow), consisting of V26-X-X-A29-F20. These hydrophobic clusters are indicated by color along the abscissa. The plots showing evolution of Q_b are placed around the two-dimensional free energy landscape, from unbound state to the complex state, and are labelled A to H correspondingly. Typical conformations at those locations on the landscape are also shown.
doi:10.1371/journal.pcbi.1001118.g003

tightly with YPrA. The binding score of IA3 to YPrA as a function of the folding score Q_i is shown in Figure 4B. Before entering into the hydrophobic cave of YPrA, IA3 searches the structural surface of YPrA with a continuous adjustment of its positioning, as reflected by the zigzag behavior of Q_i, until the binding score reaches 0.5. Folding proceeds once significant binding is realized on the interface, and binding and folding are subsequently coupled as the final native complex forms. The evolution of the structures and the related contact maps are shown in Figure 4c in Text S1.

Common Scheme Revealed by Multiple Pathways

The Q score measures only the residue-level similarity of the backbone to a reference structure. Therefore we introduce a Shadow Contact Map (SCP), an algorithm that calculates the interatomic contacts involving sidechain atoms [35]. The SCM algorithm describes the contact map excluding unphysical contacts. A cutoff distance is used to define the contacts, and this cutoff is set as $6.0 Å$ in our calculation. As in the definition of Q score, the total number of contacts can be divided into monomeric folding contacts and interfacial contacts. The contacts are grouped into two categories: native and non-native contacts, as determined by whether the contacts between residue pairs exist in final conformation. For final state, the atomic interfacial contact number and the intrachain contact number of IA3 are 604 and 103, respectively. This indicates that the interfacial interactions are far stronger than the intrachain interactions in IA3. This may explain why the kinetic process proceeds as binding followed by folding.

The evolution of the number of atomic contacts and the helix formation in IA3 along the folding pathway of *path*1, *path*2 and *path*3 are shown in Figure 5B–5D. For exploring the relationships between atom-atom contacts and IA3 folding, the contact number curves are overlaid with the evolution of α helix formation of IA3. Residues constituting α helix are assigned via analysis by the DSSP program [36], according to characteristic hydrogen-bond patterns, but other secondary structure elements such as coils and turns are excluded from this plot for the sake of clarity. Note that we have not detected beta sheet elements in our model, although there are experimental reports that IA3 may bind pepsin as a beta-strand and is therefore cut and digested as a substrate [19]. From the evolution of number of interfacial contacts (blue line) and the native contacts in IA3 (black line), we can see that IA3 binds with YPrA more and more tightly before it begins to form native contacts and helix structure. The native interfacial contacts then (green line) begin increasing until most of native contacts in IA3 are formed. For *path*2 and *path*3, the evolution of contact number is quite similar, but the corresponding processes of helix formation are significantly different. The long helix is formed from three nuclei located around the three hydrophobic clusters. Although there are significant differences between the three pathways, they reveal the common theme that IA3 binds to YPrA prior to folding. We also see clearly that non-native interactions are the dominant driving force in the initial stage of binding. Non-native contacts smoothly increase, while the native interface and native folding contacts only begin to appear at grid values near 70.
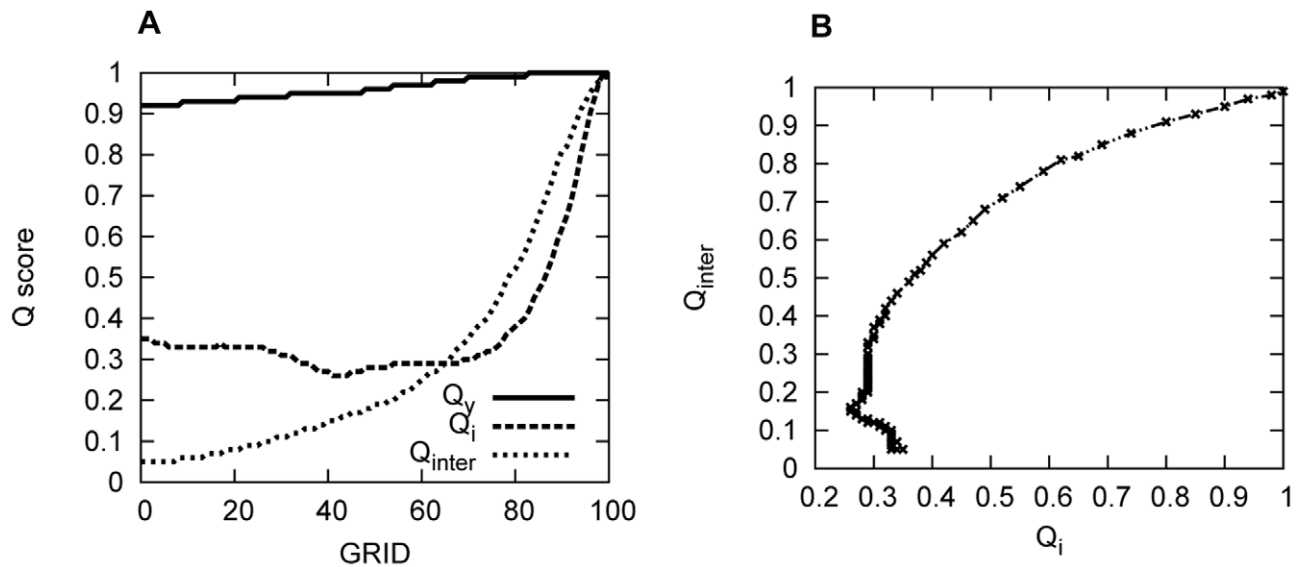
Figure 4. Folding and binding scores in _path_1, as obtained by the full atomic model. "GRID" refers to the discrete conformations along the kinetic pathway. The evolution of Q scores is shown in (A). The relationship between the IA3 folding score Q_i and the IA3/YPrA binding score Q_{inter} is plotted in (B).
doi:10.1371/journal.pcbi.1001118.g004

The Role of Non-Native Interactions Revealed by Pathway in Average

The average path is shown by the evolution of the average number of atomic contacts in Figure 6. A sharp increase in IA3 native contacts is observed in the black curve. This can be explained as the result of contact network forming in a highly synergistic way. Figure 6A shows that the native contacts of IA3 form together with non-native interfacial contacts while the number of interfacial native contacts remains nearly zero until grid 80. At the first stage of kinetic binding, the interactions between IA3 and YPrA are mostly contributed by non-native contacts. It is the non-native interactions between IA3 and the residues on the surface of YPrA that induce IA3 to bind with its target partner. Therefore, non-specific (non-native) interactions induce initial binding of IA3 to YPrA. After IA3 reaches YPrA, native interactions of binding set in by adjusting the conformation at the active site groove.IA3 folds into helical structure after binding with the active site groove of its target enzyme.

Why Backtracking?

It is noteworthy that in _path_1 (see Figure 5B) we observe a certain fraction of the helix content formed between grid 55 and 58 disappears and then later reappears. The formation and breakup of local secondary structure is observed not only in _path_1, but also in several other pathways that we calculated. In addition, we also found the formation and breaking of native contacts in _path_3 (see Figure 2c in the Text S1). Remarkably, an analogous process was observed in the investigation of the folding of Interleukin-1β (IL-1β) [37,38], knotted proteins [39], CheY-like family [40,41] and SAM-1 Riboswitch [42]. This behavior is known as "backtracking". Here, we define it as the interim formation of local secondary structures or native contacts along the reaction pathways. Experiments in _silico_ and in _vitro_ suggest that it is the result of topological frustration. Here, we propose that it results not only from topological factors but also from energetic contributions to the stability of IA3, as a partial compensation of entropy reduction during binding to YPrA.

During a biologically realistic interaction, binding to the special partner can help a protein to shrink the search space in the energy landscape. However, the associated free energy increases due to the rapid entropy reduction. As a partial compensation, occasional interactions may form only if they are energetically favored, irrespective of whether they are native contacts. These interactions do not form optimally but form easily in certain topologies. Therefore they are often unstable and fragile. During this process, both native interactions which are not stably formed and non-native interactions which are not included in the final state play roles in smoothing the free energy landscape. They stabilize the protein energetically, thus compensating the entropy reduction. As the molecule searches deeper in the free energy basin, the unstable native interactions will break down and reform stably in final structure. Hence backtracking is observed.

However we do not always observe non-native interactions or backtracking of native interactions along folding or binding pathways in nature. We explain this from three perspectives. First, the order parameters are usually coarse. They are not accurate enough to capture these details. Second, these interactions are transient and unstable. They may be difficult to measure. Third, not all the pathways are very rough. In this case of IA3-YPrA, it seems that non-native interactions play a more important role in IA3 binding while backtracking interactions are more significant in IA3 folding. From this, we believe non-native interactions and unstable native interactions can both play a role in protein folding and protein-protein recognition.

Discussion

Like many IDPs, IA3 forms an ordered structure in the presence of its interaction partner. Its binding and folding dynamics play an essential role in the regulation of its target enzyme, YPrA. Molecular dynamics simulations can help us to explore the interaction at a level of detail that is difficult to obtain in laboratory experiments. However, standard MD is often limited by the temporal range it can probe at atomic detail. In this work, we developed a multi-scaled approach to provide a comprehensive
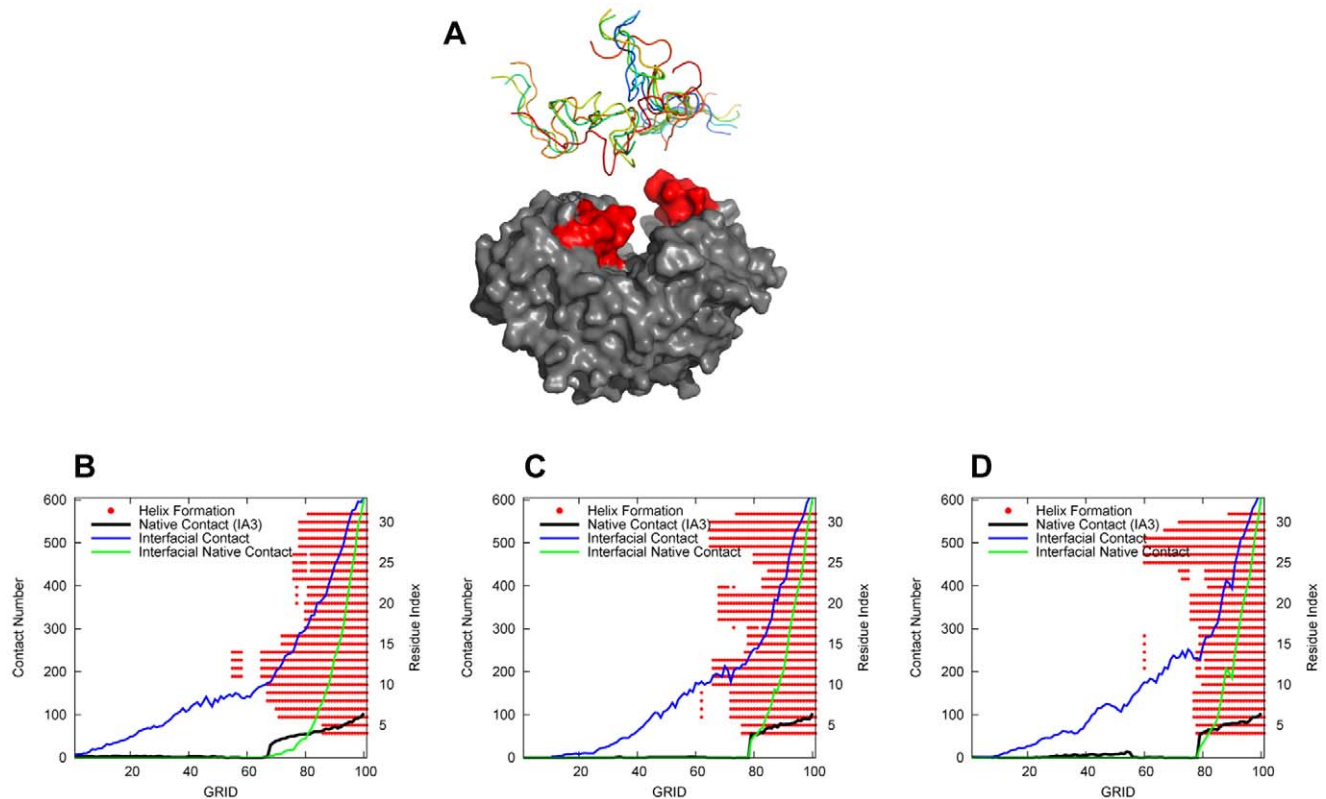
Figure 5. Kinetic pathways by full atomic model. (A) The initial structures for the binding pathways calculated. Different initial unfolded structures of IA3 are represented by coils with different colors. The two loop regions of YPrA are indicated in red. The evolution of atomic contact number and helix formation along the folding pathway are shown for *path*1 (B), *path*2 (C) and *path*3 (D). In these panels, the number of native contacts of IA3 is shown by the black line. The total interfacial contacts and the native interfacial contacts are shown in blue and green lines respectively. The red bars indicate IA3 α helix formation, with the residue index on the right axis. For clarity, other secondary structural elements are not shown.
doi:10.1371/journal.pcbi.1001118.g005

picture of the protein binding-folding dynamics, including both global thermodynamic landscape and atomic details of structural evolution paths.

Several reaction pathways were generated from different starting points to the final conformation of the protein-inhibitor complex. Although there are significant differences between the
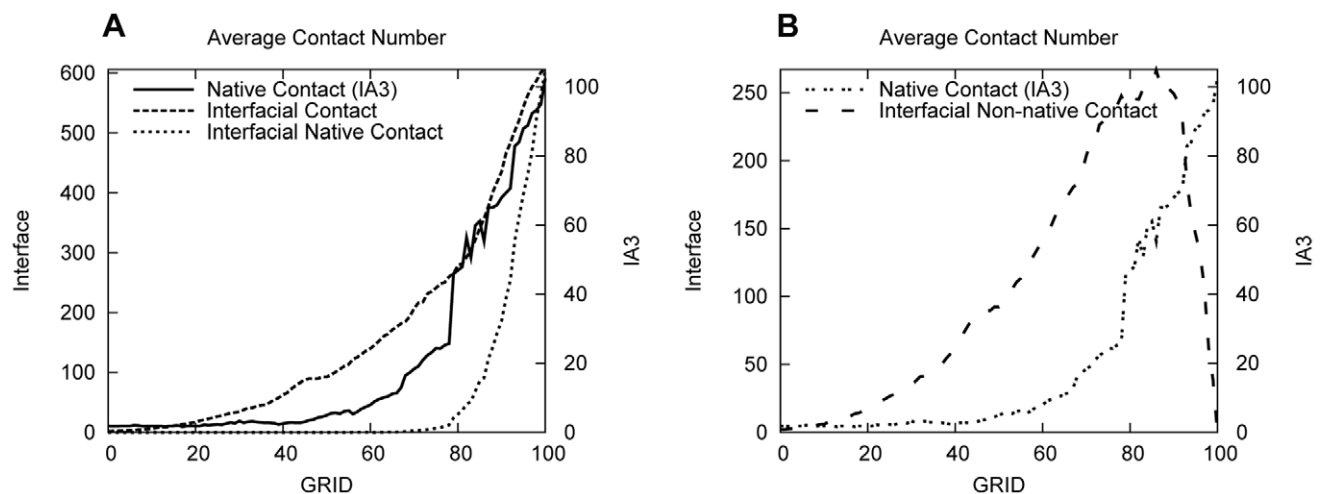

Figure 6. Evolution of the average number of atomic contacts along all the pathways we calculated. (A) The development of native contacts of IA3 is overlaid with curves showing the formation of interfacial contacts, both total and native. The native contacts of IA3 are shown by the black line. The total interfacial contacts and its native contacts are shown by the dashed line and dot line, respectively. (B) The development of IA3 non-native interfacial contacts, as well as native interfacial contacts, as a function of progress along the kinetic pathway.
doi:10.1371/journal.pcbi.1001118.g006

multiple pathways, which reflect the multidimensional nature of the underlying energy landscape [23,25,27,43], all reveal a common theme that IA3 binds to its target enzyme prior to folding itself into a helix. This finding is consistent with that of a coarse-grained free energy landscape from a structure-based MD simulation. In summary, the following folding and binding mechanism emerges. In the first step IA3 moves close to YPrA and binds to the surface of the active site groove via non-native interactions, through the long range electrostatic attraction. Before overcoming the free energy barrier, most of IA3 remains unstructured. Once IA3 enters into the cleft, its motion is greatly restrained, due to the lack of space for motion. In this highly hydrophobic environment, IA3 finally folds into an amphipathic α helix at the long cleft. In addition, we found that the mid region of the IA3 sequence, consisting of hydrophobic *cluster* 2, forms native interactions with YPrA earlier than the two terminal regions. This may be the result of stabilization by the interactions with the YPrA "flap". During binding, YPrA plays the role of a template to induce IA3 folding into the characteristic structure that blocks the active site of the enzyme. In other words, the mechanism of saccharopepsin inhibition by IA3 as revealed by our simulation is in favor of the "induced-fit" model [12]. In this context, an "induced-fit" mechanism refers to a target enzyme that induces in its inhibitor a significant conformational change.

We also examined the non-native interaction by classifying the atomic contacts, as calculated by a new algorithm (SCM). At the first stage, the interactions between IA3 and YPrA are under the control of non-native interfacial contacts. The recognition process of the inhibitor-enzyme complex is dominated by these non-native interactions, which have been reported to play a role in protein assembly [9,44–48]. The great success of simulating protein folding using structure-based models [49] which depend on the native topology suggests that the native contacts govern the folding of a protein that is well-designed by evolution. In the conventional view, non-native interactions are the major factor contributing roughness to the energy landscape [31,46]. Why does non-native interaction seem to play a facilitating role for binding in the IA3/YPrA system? It is easy to explain the results in the view of the structure of the enzyme-inhibitor complex. As the target binding site is located deep in the groove, IA3 has to search the molecular surface of YPrA to find an entropically favored and energetically optimized path to the hydrophobic cleft. Experimental studies have already hinted that non-native contacts from the helix-forming (enzyme-inhibiting) N terminus as well as the disordered C terminus (not included in this study) of IA3 assist the kinetics during early stages of the interaction without affecting the final stability of the complex [17,20,22]. The importance of nonnative interactions was also observed in pKID and KIX binding experiments [1,4,9] and DNA-binding proteins [48]. These earlier findings support our conclusion here as well as the fly-casting mechanism [50].

From the energy landscape perspective, the underlying landscape of the entire binding process must be funnel-like in order to guarantee biological recognition and native binding complex formation. There are several ways to guarantee the underlying landscape to be funnel-like [23,51]. One way is to enhance the native interactions or native bias. The other way is to reduce the non-native interactions or the roughness of the landscape. Those two ways are natural and conventionally emphasized. However there is another way to help the formation of the funneled landscape. A reduction in the entropy can significantly shrink the search space of the landscape. Here we see non-native interactions, even if not energetically favored, can contribute significantly to forming the binding funnel by reducing the entropy, bringing IA3 closer to the target YPrA interface). In this sense, non-native interactions can help the binding process.

Here we reveal the interaction mechanism of an aspartic proteinase and its endogenous inhibitor. Our studies provide a greater understanding of this unprecedented mode of enzyme inhibition. The results demonstrate the success of the multi-scaled approach for exploring the interaction of IA3 and YPrA, and they are consistent with the conclusions from time resolved experiments, which suggest non-specific binding followed by folding [14]. The combined method may be useful in understanding other enzyme-inhibitor systems. It also may offer valuable insights into the design of drugs inhibitors for the aspartic proteinases generated by pathogenic organisms.

Methods

Coarse-Grained Structure-Based Model

We performed the molecular dynamics simulations using a structure-based Hamiltonian to describe the energy of the protein in a given configuration. A structure-based Hamiltonian takes into account only native interactions, and each of these interactions enters into the energy balance with the same weighting. Therefore the model does not have heterogeneity in energy and it includes only topological frustration. Each amino acid is described by a single bead on a polymer chain located on the C_α position [52]. The C_α structure-based Hamiltonian is given by the expression:

$$H = \sum_{bonds} K_r(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedral} K_\phi^{(n)}[1 + cos(n \times (\phi - \phi_0))] +$$
$$\sum_{i<j-3}^{native} \epsilon(i,j)[5(\frac{\sigma_{ij}}{r_{ij}})^{12} - 6(\frac{\sigma_{ij}}{r_{ij}})^{10}] + \sum_{i<j-3}^{non-native} \epsilon_2(i,j)(\frac{\sigma_{NC}}{r_{ij}})^{12}$$

The total energy is divided into bond stretching, angle bending, torsion and nonbonded interactions. r, θ and ϕ are the virtual bond length, bond angle, and torsion angle defined by C_α position. r_0, θ_0 and ϕ_0 are the corresponding native values from the PDB structure. Nonbonded interactions are considered when two C_α atoms i and j are separated sequentially by at least three residues on a chain or when they come from different chains, are subdivided into native interactions and nonnative interactions. For native contacts, σ_{ij} is the distance between the C_α positions of contacting residues i and j. For non-native contacts, σ_{NC} provides excluded volume repulsion. We treat the nonlocal interactions within a chain and between the chains with the same strength. The native contact map is derived from a shadow contact map (SCM) [35]. Parameters K_r, K_θ, K_ϕ, ϵ, ϵ_2 weight the relative strength of each kind of interactions contributed to energy, and $K_r = 10000 kJ/(mol \cdot nm^2)$, $K_\theta = 20 kJ/mol$, $K_\phi^{(1)} = 1 kJ/mol$, $K_\phi^{(3)} = 0.5 kJ/mol$, $\epsilon = 1 kJ/mol$, $\epsilon_2 = 1 kJ/mol$. In order to sample more binding transitions we added a bias potential into the Hamiltonian. The bias potential is intended to make binding transitions more frequent by raising the free energy of the bound state. Here, we choose a harmonic form where the bias potential energy depends on the center of mass distance (R_{COM}) between YPrA and IA3.

$$V_{biased}(R_{COM}) = \frac{1}{2}k(R_{COM} - R_0)^2 \qquad (1)$$

Here k is the force constant, R_0 is the equilibrium position, R_{COM} is the COM distance between the two chains. Then the Hamiltonian has a new form $H_{biased} = H_{unbiased} + V_{biased}$. In the

native complex, where IA3 is bound to YPrA, R_{COM} (only C_α) is 1.063 nm, while for R_{COM} larger than 3.5 nm, we consider the system to be in the unbound state. We choose $R_0 = 3.0 nm$ to ensure that the bias potential lifts the free energy of bound states more than unbound states. Through many trials we found that an optimized value for the force constant is $k = 4.25 kJ/(mol \cdot nm^2)$.

The unbiased thermodynamic average of a function A can be calculated as follows:

$$<A>_{unbiased} = \frac{<Ae^{V_{biased}(\eta)/k_BT}>_{biased}}{<e^{V_{biased}(\eta)/k_BT}>_{biased}}$$

where η is the reaction coordinate. The free energy of the system at η is given by $F(\eta) = -k_BT \ln P(\eta)$, where $P(\eta)$ is the equilibrium probability.

$$P_{unbiased}(\eta) = P_{biased}(\eta) \cdot e^{V_{biased}(\eta)/k_BT} \cdot \frac{1}{<e^{V_{biased}(\eta)/k_BT}>_{biased}}$$

The unbiased free energy can be calculated as

$$F_{unbiased}(\eta) = -k_BT \ln(P_{biased}(\eta)) - V_{biased}(\eta) + C \qquad (2)$$

The constant C does not change in constant temperature simulations, so in our simulations we can select the value that sets the lowest $F(\eta)$ to zero.

The simulations were performed using the Gromacs software package [53]. We put the protein system in a 50 nm cubic box corresponding to a low protein concentration. In fact, the effective box length is about 8.4 nm (the largest R_{COM} in coarse grained MD simulation). Nonbonded interactions are cut off at 3 nm. The time step was 0.5 fs. Stochastic dynamics were used with a drag coefficient $\gamma = 1.0$. We started our trajectories with 9 different configurations in either native or nonnative state. The actual total constant temperature simulation time is $1.35 \mu s$. The total data include 214 binding and dissociation transitions, allowing us to observe how the dynamics change during the folding and binding. We calculated the free energy from the trajectories using WHAM (Weighted Histogram Analysis Method) [54], and using the formula 2 to get unbiased free energy. The simulation temperature is set at 176 K. Plotting the 2-D free energy surface for the binding/folding behavior requires two independent reaction coordinates, representing binding and folding respectively. From the transformation equation 2 we know that $V_{biased}(\eta)$ has an explicit expression only when η contains R_{COM}. R_{COM} can describe the binding behavior. For folding, we choose Qf_{IA3}, which is defined as the fraction of native spatial tertiary contacts. A native contact is formed if the distance between the two C_α atoms is shorter than 1.2 times their native distance σ_{ij}.

$$F_{unbiased}(Qf_{IA3}, R_{COM}) = -k_BT \ln(P_{unbiased}(Qf_{IA3}, R_{COM})) - V_{biased}(R_{COM})$$

Simulation Details of Optimal Paths Calculation at Atomic Level

The free and inhibited states of YPrA were generated from crystal structures taken from the Protein Data Bank (1FMU and 1DP5, respectively). The unfolding of IA3 was generated by Langevin dynamics by NAMD with the Charmm32 force field. Then, the initial and final structures of the complex were modelled. Crystallographic water molecules and carbohydrate moieties were removed. After modeling of the reactant and product, the paths connecting these states were calculated with the MOIL package [55]. The MOIL energy function combines the AMBER and OPLS force fields [56,57]. We can solve the minimum energy path if the pre-specified initial and final states are known. Given the minimized endpoint structures, the initial guesses for the trajectory are determined by the minimum-energy-path self-penalty walk (SPW) [58] functional embedded in the CHMIN module. Then these paths were optimized in the SDP module with steepest descent. The solvation effects are described by the Generalized Born model [59,60]. The high-frequency modes from the trajectories are filtered and modeled as Gaussian white noise. The cut-off distance for van der Waals interactions is $9.0 \mathring{A}$.

Steepest Descent Path Algorithm

The steepest descent path is widely used in qualitative interpretation of chemical reactions [61,62]. In analogy to the classical action, an action S as a function of length in a discrete representation is defined to represent a most probable Brownian trajectory as follows:

$$S = \sum_{j=1}^{N-1} \sqrt{H_s + (\frac{\partial U}{\partial x_j})^2} |x_{j+1} - x_j| \qquad (3)$$

We split the path by N grids to approximate the path by a set of discrete conformations. x_j is the entire vector of conformational coordinates at grid j. The initial conformation is x_1 and the final conformation is x_N. The potential energy U is a function of the mass-weighted coordinate vector. The constant H_s is an arbitrary positive value that mimics the energy in classical mechanics. Optimal paths with different thermal energies are generated by tuning this parameter. The steepest descent path is the limiting path that optimizes the action for $H_s \to 0$. The shortest path between x_i and x_f as generated by linear interpolation is the optimal path for $H_s \to \infty$. Here we used $H_s = 10^{-5}$.

Given the two end structures, the SDP module will minimize the target function:

$$T = \sum_{j=1}^{N-1} \sqrt{H_s + (\frac{\partial U}{\partial x_j})^2} |x_{j+1} - x_j| + C \qquad (4)$$

$$C = \lambda \sum_j (\Delta l_{j,j+1} - <\Delta l>)^2 \qquad (5)$$

$$<\Delta l> = \frac{1}{N-1} \sum_{j=1}^{N-1} \Delta l_{j,j+1} \qquad (6)$$

where C is a restraint to ensure that configurations x_j's are distributed approximately uniformly along the pathway. The target function T is minimized by conjugate gradient local minimization. $\Delta l_{j,j+1} = M^{1/2} |x_j - x_{j+1}|$ is the arc-length of the path in mass weighted coordinates between conformation x_j and x_{j+1}. λ is the strength of a penalty function that restrains the step length to the average length $<\Delta l>$. For further details see Ref. [63].

SDP is a continuous curve with a low-energy barrier that connects the reactants and the products. An important advantage of an SDP is
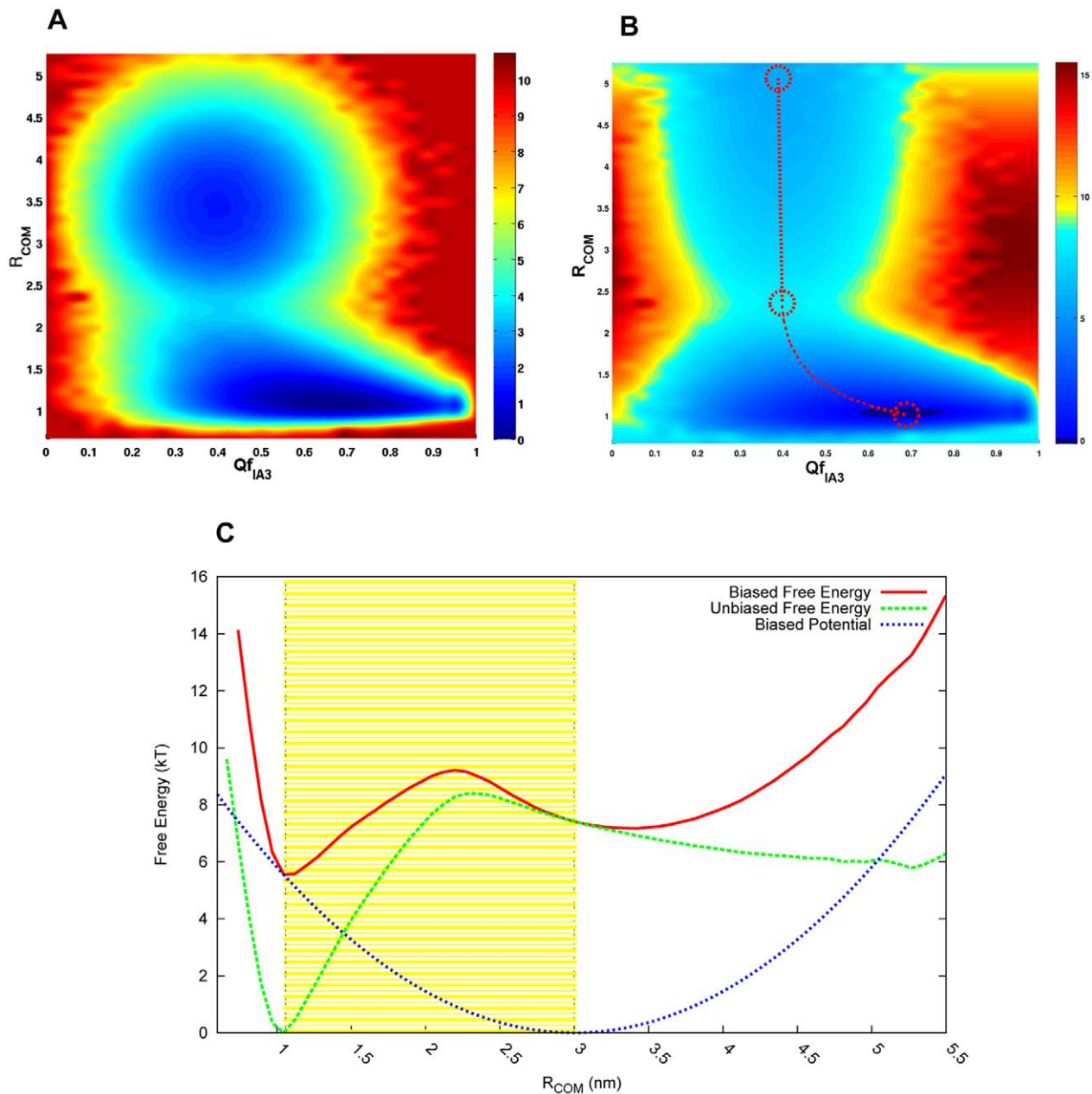
Figure 7. One- and two-dimensional free energy profiles (in units of $k_B T$) obtained by the coarse grained structure based model at $T = 176K$. (A) The free energy surface before unbiasing is shown as a function of R_{COM} and Qf_{IA3}. (B) The unbiased free energy surface is shown as a function of R_{COM} and Qf_{IA3}. The binding route is marked by the red dotted line. The initial state, transition state and final state are labelled by red dashed circles. The binding route shows that IA3 interacts with YPrA prior to folding into a helical structure. (C) The biased and unbiased free energy profiles and the harmonic biasing potential are shown along the coordinate representing the center of mass distance between YPrA and IA3. The biasing potential is introduced to accelerate the sampling around the important transitions (yellow region), by raising both the native binding state ($R_{COM} = 1.06nm$) and nonnative unbinding state ($R_{COM} \geq 3.5nm$) with only a small perturbation at the transition state.
doi:10.1371/journal.pcbi.1001118.g007

that it allows testing of a concrete mechanism. The disadvantage is that it gives no information about the properties of the system far from the steepest descent path. Other non-native interactions which are away from the binding groove cannot be sampled on the steepest descent path. As we know, the Milestoning method [64,65] has been developed by Ron Elber and coworkers to solve this problem. However, it is still an open question how to calculate kinetics and thermodynamics of long-time biological processes, which are typically not accessible by straightforward MD simulation.

Boundary Structures Preparation

Based on boundary conditions, the initial and final coordinates must be specified. We take the crystalline structure of YPrA complexed with IA3 mutant inhibitor (PDB code: 1DP5) as the endpoint of the transitional trajectories. We assume that the trajectories start with the uninhibited enzyme and unfolded IA3 that is far from the binding site of proteinase A. For free YPrA the coordinates in trigonal and monoclinic crystal forms are accessible from Protein Data Bank under accession codes 1FMU and 1FMX,

respectively. Here, we adopt the coordinates of the trigonal crystal form, not only for the clarity of the electron density in the "flap" consisting of a β hairpin loop extending over the active site, but also because of the possible presence of some hydrolysis products in the monoclinic crystal [29]. However, in the crystal structure of the trigonal form, there are two disordered regions in which the electron density is relatively poor. The two highly flexible regions, located at the peptide segments 162–165 and 243–245, are considered to make less contribution to the inhibitor binding as their locations on the molecular surface are far from the active site. The missing segments were modelled by structure prediction. Although YPrA is glycosylated its covalently binding carbohydrate moieties are not considered in the simulation. The structural difference between the initial and final states of YPrA is shown in Supplemental Figure 1A in Text S1. Obviously, the unfolded state of the inhibitor can not be represented by a single structure. It should be an ensemble of conformations having a sufficient number of degrees of freedom. These were generated by Langevin dynamics using NAMD [66] with Charmm22 force field. The initial conformation of the system was constructed in stages, starting with the complex of folded IA3 and uninhibited YPrA with a center of mass distance of 3 nm from each other, followed by packing the complex with a 4 nm thick water sphere. We then carried out minimization using the conjugate gradient algorithm with 1000 steps. The initial distance between IA3 and YPrA was set to about 3 nm as hinted by the coarse grained model. The region of most interest is the range R_{COM} from 1.06 nm to 3.0 nm, where folding and binding occur, which is indicated by the yellow region in Figure 7. In order to generate the conformation of the complex, the minimized system was heated to 500 K in the canonical ensemble. The procedure employed a Langevin thermostat with a $5ps^{-1}$ damping parameter. Constraints were applied to the lengths of all bonds involving hydrogen atoms, thus allowing a 2 fs time step. A spherical boundary condition was used to control the 4.5 nm thick water sphere from the center of mass of the complex. YPrA was fixed during the high temperature dynamics. After heating IA3 for 1 ns, we extracted intermediate structures (without water) whose RMSDs from the helical structure were larger than $7.5Å$ and that were separated by interval steps larger than 1 ps.

Supporting Information

Text S1 Supporting information of flexible binding-folding of IA3 to YPrA.
Found at: doi:10.1371/journal.pcbi.1001118.s001 (5.21 MB PDF)

References

1. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6: 197–208.
2. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, et al. (2003) Predicting intrinsic disorder from amino acid sequence. Proteins 53: 566–572.
3. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your id: intrinsic disorder as an id for recognition, regulation and cell signaling. J Mol Recognit 18: 343–84.
4. Wright PE, Dyson HJ (2009) Linking folding and binding. Curr Opin Struct Biol 19: 31–38.
5. Uversky VN (2009) Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. Protein J 28: 305–325.
6. Ganguly D, Chen J (2009) Atomistic Details of the Disordered States of KID and pKID. Implications in Coupled Binding and Folding. J Am Chem Soc 131: 5214–5223.
7. Chen J (2009) Intrinsically Disordered p53 Extreme C-Terminus Binds to S100B (beta beta) through \Fly-Casting". J Am Chem Soc 131: 2088–2089.
8. Huang YQ, Liu ZR (2009) Kinetic advantage of intrinsically disordered proteins in coupled foldingbinding process: A critical assessment of the "y-casting" mechanism. J Mol Biol 393: 1143–1159.
9. Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-induced folding of a natively unstructured transcription factor. PLoS Comput Biol 4: e1000060.
10. Levy Y, Onuchic JN, Wolynes PG (2007) Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. J Am Chem Soc 129: 738–739.
11. Tsai C, Ma B, Sham Y, Kumar S, Nussinov R (2001) Structured disorder and conformational selection. Proteins 44: 418–427.
12. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 35: 539–46.
13. Dreyer T, Valler MJ, Kay J, Charlton P, Dunn BM (1985) The selectivity of action of the asparticproteinase inhibitor ia3 from yeast (saccharomyces cerevisiae). Biochem J 231: 777–9.
14. Parr CL, Keates RA, Bryksa BC, Ogawa M, Yada RY (2007) The structure and function of saccharomyces cerevisiae proteinase a. Yeast 24: 467–80.
15. Davies DR (1990) The structure and function of the aspartic proteinases. Annu Rev Biophys Biophys Chem 19: 189–215.
16. Winterburn TJ, Wyatt DM, Phylip LH, Berry C, Bur D, et al. (2006) Adaptation of the behaviour of an aspartic proteinase inhibitor by relocation of a lysine residue by one helical turn. Biol Chem 387: 1139–42.
17. Green TB, Ganesh O, Perry K, Smith L, Phylip LH, et al. (2004) Ia3, an aspartic proteinase inhibitor from saccharomyces cerevisiae, is intrinsically unstructured in solution. Biochemistry 43: 4071–81.
18. Padron-Garcia JA, Alonso-Tarajano M, Alonso-Becerra E, Winterburn TJ, Ruiz Y, et al. (2009) Quantitative structure activity relationship of ia(3)-like peptides as aspartic proteinase inhibitors. Proteins 75: 859–869.
19. Li M, Phylip LH, Lees WE, Winther JR, Dunn BM, et al. (2000) The aspartic proteinase from saccharomyces cerevisiae folds its own inhibitor into a helix. Nat Struct Biol 7: 113–7.
20. Phylip LH, Lees WE, Brownsey BG, Bur D, Dunn BM, et al. (2001) The potency and specificity of the interaction between the ia3 inhibitor and its target aspartic proteinase from saccharomyces cerevisiae. J Biol Chem 276: 2023–30.
21. Ganesh OK, Green TB, Edison AS, Hagen SJ (2006) Characterizing the residue level folding of the intrinsically unstructured ia3. Biochemistry 45: 13585–96.
22. Narayanan R, Ganesh OK, Edison AS, Hagen SJ (2008) Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase ypra. J Am Chem Soc 130: 11477–85.
23. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. Proc Natl Acad Sci U S A 101: 511–516.
24. Cardenas AE, Elber R (2003) Kinetics of cytochrome c folding: atomically detailed simulations. Proteins 51: 245–257.
25. Wang J, Zhang K, Lu H, Wang E (2005) Quantifying kinetic paths of protein folding. Biophys J 89: 1612–1620.
26. Henkelman G, Uberuaga BP, Jonsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113: 9901–9904.
27. Wang J, Zhang K, Lu H, Wang E (2006) Dominant kinetic paths on biomolecular binding-folding energy landscape. Phys Rev Lett 96: 168101.
28. Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: Conformational ooding. Phys Rev E 52: 2893–2906.
29. Gustchina A, Li M, Phylip LH, Lees WE, Kay J, et al. (2002) An unusual orientation for tyr75 in the active site of the aspartic proteinase from saccharomyces cerevisiae. Biochem Biophys Res Commun 295: 1020–6.
30. Tang J, Koelsch G (1995) A possible function of the aps of aspartic proteases: The capture of substrate side chains determines the specificity of cleavage positions. Protein Pept Lett 2: 257–266.
31. Hardin C, Luthey-Schulten Z, Wolynes PG (1999) Backbone dynamics, fast folding, and secondary structure formation in helical proteins and peptides. Proteins 34: 281–94.
32. Zong C, Papoian GA, Ulander J, Wolynes PG (2006) Role of topology, nonadditivity, and watermediated interactions in predicting the structures of alpha/beta proteins. J Am Chem Soc 128: 5168–76.
33. Sutto L, Latzer J, Hegler JA, Ferreiro DU, Wolynes PG (2007) Consequences of localized frustration for the folding mechanism of the im7 protein. Proc Natl Acad Sci U S A 104: 19825–30.

34. Weinkam P, Pletneva EV, Gray HB, Winkler JR, Wolynes PG (2009) Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. Proc Natl Acad Sci U S A 106: 1796–801.

35. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN (2010) Smog@ctbp: simplified deployment of structure-based models in gromacs. Nucleic Acids Res 38 Suppl: W657–61.

36. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–637.

37. Gosavi S, Chavez LL, Jennings PA, Onuchic JN (2006) Topological frustration and the folding of interleukin-1 beta. J Mol Biol 357: 986–96.

38. Capraro DT, Roy M, Onuchic JN, Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? Proc Natl Acad Sci U S A 105: 14844–8.

39. Sulkowska JI, Sulkowski P, Szymczak P, Cieplak M (2008) Stabilizing effect of knots on proteins. Proc Natl Acad Sci U S A 105: 19714–9.

40. Hills RD, Jr., Brooks CL, III (2008) Subdomain competition, cooperativity, and topological frustration in the folding of chey. J Mol Biol 382: 485–495.

41. Hills RD, Jr., Kathuria SV, Wallace LA, Day IJ, Brooks CL, III, et al. (2010) Topological Frustration in beta alpha-Repeat Proteins: Sequence Diversity Modulates the Conserved Folding Mechanisms of alpha/beta/alpha Sandwich Proteins. J Mol Biol 398: 332–350.

42. Whitford PC, Schug A, Saunders J, Hennelly SP, Onuchic JN, et al. (2009) Nonlocal Helix Formation Is Key to Understanding S-Adenosylmethionine-1 Riboswitch Function. Biophys J 96: L7–L9.

43. Rylance GJ, Johnston RL, Matsunaga Y, Li CB, Baba A, et al. (2006) Topographical complexity of multidimensional energy landscapes. Proc Natl Acad Sci U S A 103: 18551–5.

44. Karanicolas J, Brooks rCL (2003) Improved go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. J Mol Biol 334: 309–325.

45. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: Theory and simulation. Prot Sci 13: 1750–1766.

46. Cho SS, Levy Y, Wolynes PG (2009) Quantitative criteria for native energetic heterogeneity inuences in the prediction of protein folding kinetics. Proc Natl Acad Sci U S A 106: 434–9.

47. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, et al. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. Proc Natl Acad Sci U S A 105: 9999–10004.

48. Toth-Petroczy A, Simon I, Fuxreiter M, Levy Y (2009) Disordered Tails of Homeodomains Facilitate DNA Recognition by Providing a Trade-Off between Folding and Specific Binding. J Am Chem Soc 131: 15084+.

49. Go N (1983) Theoretical studies of protein folding. Annu Rev Biophys Bioeng 12: 183–210.

50. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the y-casting mechanism. Proc Natl Acad Sci U S A 97: 8868–8873.

51. Wang J, Verkhivker GM (2003) Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. Phys Rev Lett 90: 188101.

52. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. J Mol Biol 298: 937–953.

53. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) Gromacs: fast, exible, and free. J Comput Chem 26: 1701–18.

54. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules .1. the method. J Comput Chem 13: 1011–1021.

55. Elber R, Roitberg A, Simmerling C, Goldstein R, Li HY, et al. (1995) Moil - a program for simulations of macromolecules. Comput Phys Commun 91: 159–189.

56. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, et al. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106: 765–784.

57. Jorgensen WL, JTirado-Rives (1988) The opls potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110: 1657–1666.

58. Czerminski R, Elber R (1990) Self-avoiding walk between 2 fixed-points as a tool to calculate reaction paths in large molecular-systems. Int J Quantum Chem 38: 167–186.

59. Hawkins GD, Cramer CJ, G TD (1995) Pairwise solute screening of solute charges from a dielectric medium. Chem Phys Lett 246: 122.

60. Tsui V, Case DA (2000) Theory and applications of the generalized born solvation model in macromolecular simulations. Biopolymers 56: 275–91.

61. Olender R, Elber R (1996) Calculation of classical trajectories with a very large time step: Formalism and numerical examples. J Chem Phys 105: 9299–9315.

62. Yang Z, Majek P, Bahar I (2009) Allosteric transitions of supramolecular systems explored by network models: Application to chaperonin groel. PLoS Comput Biol 5: e1000360.

63. Majek P, Elber R, Weinstein H (2009) Pathways of conformational transitions in protein. In: Voth GA, ed. Coarse-Graining of Condensed Phase and Biomolecular Systems. Florida: CRC Press. pp 185–203.

64. West AMA, Elber R, Shalloway D (2007) Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. J Chem Phys 126: 145104.

65. Kuczera K, Jas GS, Elber R (2009) Kinetics of helix unfolding: Molecular dynamics simulations with milestoning. J Phys Chem A 113: 7461–7473.

66. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with namd. J Comput Chem 26: 1781–802.

Text S1

1 Microscopic Perspective at Atomic Level

Two hundred paths were calculated and nine were optimized successfully. It is a non-trival task to tune the parameters in the path model for optimizing the stochastic kinetic paths with full-atomic details. The difficuties lie in 1) the side chain-dependent steric clash which may cause high local energy and blow up the system; 2) inreasonable chain crossing during the process due to the protein topological complexity. It is the major reason that we failed to optimize most of the paths. We expect that this will be improved in the future version of Moil. In our calculation, all the nine paths share the similar global mechanism. Their differences lie in the detailed routes and the nucleation position in IA3. In this paper, we choose the three paths (labelled path1, path2, path3 in main text) as concrete examples in an arbitrary way. At first, we count the native contacts in IA3 and interfacial contacts along these paths. It is shown in Fig. S2. We found the formation and breaking of native contacts along the path to the final state. This is particularly clear in path3.

To shed light on the structural evolution from microscopic perspectives, we examine the formation of hydrogen bonds and salt bridges during the process. In Fig. S3 (a), the number of hydrogen bonds of interface (first line), intrachain of IA3 (mid line) and the complex (last line) are plotted along these pathways. Several non-native hydrogen bonds and salt bridges are found both of intrachain of IA3 and interchain between IA3 and YPrA during the folding and binding process. For instance, in $path1$, there are intrachain salt bridges: L7-E10, L16-E17; in $path2$, intrachain salt bridge D4-L7 and interchain hydrogen bond T79A-E17B; in $path3$, interaction network formed by salt bridge L16-E17 and hydrogen bonds Q13-E17 and Q13-L16, interchain hydrogen bond T113A-L18B. They are shown in Fig. S3 (b). It is worthwhile noticing that we didn't find special specific non-native hydrogen bonds and salt bridges on these pathways. The results are consistent with that of coarse grained model.

2 Reaction Coordinate: Q score

It is the total fraction of the target segment of polypeptide to characterize the structural similarity to the referenced structure. Here, the referenced structure is the crystal structure of the enzyme-inhibitor complex (PDB code: 1DP5). Q value is computed by the formula as follow:

$$Q = \frac{2}{(N-1)(N-2)} \sum_{i<j-1} exp\Big[- \frac{(r_{ij} - r_{ij}^N)^2}{2\sigma_{ij}^2}\Big].$$ (1)

Where, N represents the total residue number of target polypeptide, i and j are the residue sequence number. r_{ij} and r_{ij}^N refer to the C_α distances between residue i and j of intermediate conformations and the native complex structure, respectively. σ_{ij} is the width of the function and equal to $|i-j|^{0.15}$ and the normalization $(N-1)(N-2)/2$ is the number of non-nearest neighbor pairs given the total residue number N. The Q value fluctuates between 0 (completely different with non-overlap with the native complex structure) and 1 (perfectly correspond with the final structure).

3 Estimation of Binding Affinity

To estimate the equilibrium disassociation constant K_d reflected binding affinity between IA3 and YPrA, we consider two-state kinetics for the formation of the enzyme-inhibitor complex C from free IA3 and

the protease at equilibrium, which may be expressed as,

$$[I] + [Y] \underset{K_{off}}{\overset{K_{on}}{\rightleftharpoons}} [C] \tag{2}$$

where K_{on} is the second-order rate constant as the association rate, while K_{off} is the first order rate constant for the dissociation. The ratio between the two rate constants yields the equilibrium disassociation constant K_d (has units of concentration) given by:

$$K_d = \frac{K_{off}}{K_{on}} = \frac{[I] * [Y]}{[C]}$$

We define the concentration of free enzyme $[E_f]$ and free inhibitor $[I_f]$, so total enzyme concentrantion $[Y_0] = [Y_f] + [C]$, and total IA3 concentration $[I_0] = [I_f] + [C]$. Given that the effective simulation box is spherial of radius 8.337 nm (the largest R_{COM} in our simulation), the IA3 concentration is inversely proportional to the volume of the sphere, $4\pi R^3/3$. Then $[I_0] = 0.684 * 10^3 \mu M$. Likewise the total enzyme concentration is $[Y_0] = 0.684 * 10^3 \mu M$. ΔG is is about 6 kT at this concentration which is estimated from the free energy difference between free IA3 and the complex state (from figure 7 in main text). Then $\frac{[C]}{[I_f]} = exp^{(\Delta G/KT)}$. So that $K_d = \frac{[I_f][Y_f]}{[C]} = 4.192 nM$.

4 Cut-off Algorithm to Count Native and Non-native Contacts in the Coarse-grained Structure-based Model

We used a cut-off algorithm to describe the interactions between the two chains, which can describe both native and non-native contacts formed in the trajectory. We find that in the native complex the average distance between two C_α atoms that form a contact between the two chains is 8.831 $\mathring{A}$. We therefore used two radius cut-offs to define a contact value. If the distance between two C_α atoms is shorter than 6 $\mathring{A}$, we define the contact value as 1, while if the distance is between 6 $\mathring{A}$ and 10.60 $\mathring{A}$ (which is 1.2 times 8.831 $\mathring{A}$) the contact value is 0.5. Using this algorithm to describe the native complex, we find that the total inter-chain contact value is 145.5, or a little higher than the 134 that is calculated by SCM.

In order to measure which part of YPrA interacts with IA3 at the transition state, it is easy to use the cut-off algorithm which has been used to count the interfacial contacts in a simple and direct way. However, the cut-off distance should be chosen carefully. If it is too large, non-native contacts are overcounted. If the cut-off is too small, native contacts will be undercounted. In our simplified coarse-grained structure-based model, the non-native contacts are treated as repulsive interactions, and only the native contacts are energetically favorable. Non-native contacts may be overestimated because a distance cutoff algorithm is coarse and cannot avoid finding non-native contacts near a native contact. We separated the non-native contacts into two groups: near-native and away-native. If in a non-native contact pair, both of the C_α atoms are more than 5.0 $\mathring{A}$ from any atoms that participate in a native contact pair, this non-native contact is considered as away-native. Otherwise, it is near-native. This is illustrated schematically in Fig. S6, where residues in IA3 and YPrA are represented by magenta and yellow spheres, respectively.

We found that native contacts, near-native non-native contacts, and away-native non-native contacts form 13.4%, 25.8%, and 60.8%, respectively. Therefore, there are more away-native non-native contacts than near-native non-native contacts at the transition state. The non-native contacts are primarily away-native non-native contacts. Overall, the cutoff algorithm can provide useful information about native and non-native contacts in the transition state topology, although it is coarse grained. We have explored more quantitatively the role of non-native interactions of IA3 binding to YPrA by our full-atomic model, using the combination of AMBER and OPLS force field (not the structure-based model).

5   Contact Maps based on SCM

The atomic contact maps are also based on the algorithm of SCM. If there is at least an atom-atom contact between a pair of residues, a contact between this pair of residues is considered to be formed which is represented by a hollow triangle in the plot. In order to describe the degree of strength of interaction between a pair of residues, the corresponding number of atomic contacts existing in the residue pair is shown by the color bar. To make the contacts map more informative, the native contacts are shown by squared points colored according to the number of atomic contacts in the residue pair. In this way, we can estimate that the native contact forms by the triangle with a squared point situated at the center, and the non-native contact forms by a single triangle. Two grey lines are used to divide the regions of intra-chain contacts of IA3 and interfacial contacts. In addition, the contacts of residue pair of which the index separation is less than four are not considered, because in any case these contacts are required to maintain the backbone not being broken. The residue sequence number of aspartic proteinase A is rearranged from 2 to 330, and that of IA3 is from 2 to 32. In the contact map, the residue index of YPrA is from 1 to 329 and that of IA3 is 330 to 360. Noted that, as a concrete example of pathway, only the $path1$ is shown in this supporting information. The contact maps and their corresponding structures on the folding path are shown in Fig. S4 (c). To view the structural evolution clearly, the relation between contact maps and structure is shown in Fig. S4 (a-b). The surface areas in YPrA are labelled A-J. They are colored blue exclude that the two loop regions are red. C region is the "flap". H region is loop2. Along $path1$, the initial and final states are placed at top left corner and left bottom. In the lower right corner, it corresponds to the grid 55. The interfacial contacts are labelled by red rectangles.

It is clear that the interfacial contacts are mostly contributed by non-native interactions before grid 66. The region between grid 55 to 66 may correspond to the transition state region predicted by our coarse grained model. We can also see, after this region, native contacts increase dramatically. From the evolution of contact map, we can also see that IA3 forms partially ordered structure (labelled by blue triangles) only after TS region. Considering together with the formation of native interfacial contacts also after this region, this may support a coupled folding and binding process after transition state consistent with the result by our coarse grained model. In addition, we can see the sequential order of IA3 binding to YPrA. It seems that the first regions which IA3 contacts are loop1 (labelled by C) and partial D region. The second region is loop2 (labelled by H). Then it is the region J. The four regions (C,D,H,J) all locate on the outer surface of the active site groove. After IA3 entering the groove, it contacts the surface at the bottom of the groove (labelled by A,B,E,F,G,I). In addition, we found that several non-native hydrogen bonds and salt bridges formed during the process. For example, intrachain salt bridges: L7-E10, L16-E17 and interchain hydrogen bonds: T79A-E17B, T246A-Q5B. It is worthwhile pointing out that, among 8 pathways we calculated, in four paths of them there is a nonnative intrachain salt bridge: L16-E17. This special nonnative salt bridge can neutralize partial charges in acidic and basic sidechains and contribute the stabilization of IA3. The reason of this salt bridge forming more frequently may be that sidechains of the two neighboring residues are able to contact easily in topology. we also found that hydrophobic cluster2 forms at first at grid 57. It contributes most part of non-native contacts at the first stage that IA3 bind to the outer surface of YPrA (C,D regions). Note that, this case may not be inevitable along other paths in the multi-dimensional energy landscape.
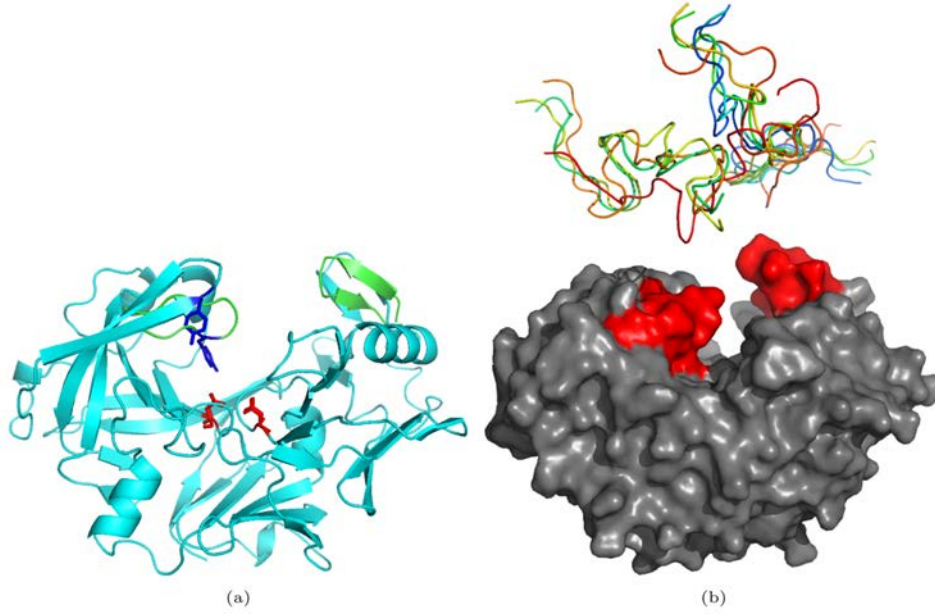
Figure 1. (a) Structural alignment between the initial state and final state of YPrA during this simulation. Their conformations are modelled based on the structures of proteinase A alone in trigonal form (1FMU) and its inhibited form by N terminal IA3 mutant (1DP5). Superimposition of the conformation of initial point onto that of the final point reveals an overall RMSD of 0.59 $\mathring{A}$. The structural distinction is reflected in two loops located on the surface of the two domains. One is the flap and the other one is the segment 243-249 in the C-terminal domain whose coordinates can not be captured by X-ray crystal diffraction in trigonal crystal of free PrA. The two loops in free form are colored in cyan to emphasize their structural diversity from the complex form. The other parts of the proteinase are quite stable upon inhibitor binding including the two catalytic aspartic acid residues even more immobile which are colored red. (b) The initial structures of all the pathways we have calculated. The different initial structures of IA3 are represented by coils with different colors. The two loop regions are colored in red.
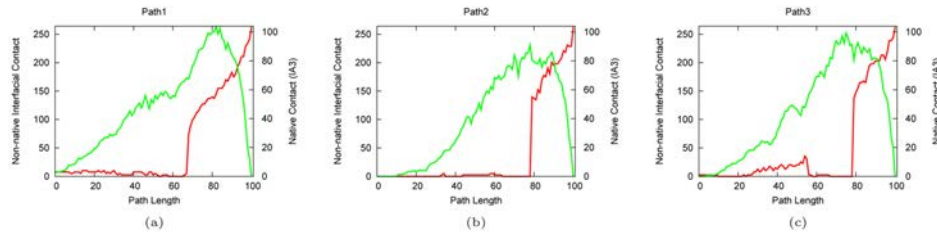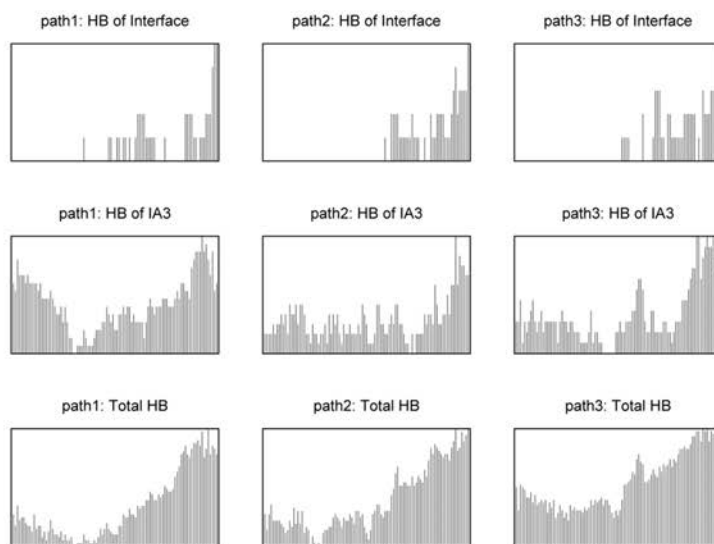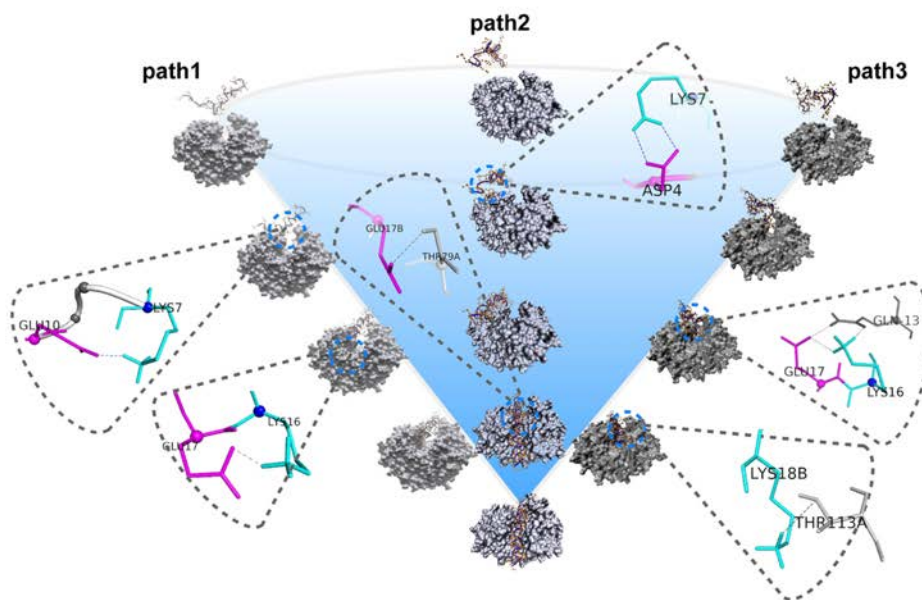

Figure 2. The non-native interfacial contact curves (green line) overlaid with native contact curve of N terminal of IA3 (red line) along the folding pathway of *path*1 (a), *path*2 (b) and *path*3 (c). We found the formation and breaking of native contacts along the path to the final state. This is particularly clear in *path*3.

(a)

(b)

Figure 3. (a) Evolution of number of hydrogen bonds on *path*1, *paht*2 and *path*3. In first row, it is the hydrogen bonds formed in interface. Hydrogen bonds formed in N terminal of IA3 and the complex are shown in second row and third row. The three columns correspond to *path*1, *paht*2 and *path*3, respectively. (b) Schematic structural evolution along these three paths. On the route, we found several non-native hydrogen bonds and salt bridges formed during the process. They are emphasized and shown at the side of structural graphs. The hydrogen bonds and salt bridges are illustrated in dashed lines. Basic residues are colored in cyan sticks and blue spheres and acidic residues are represented by magenta sticks and spheres. For sake of clarity, we only show part of typical non-native hydrogen bonds and salt bridges. On *path*1, there are intrachain salt bridges: L7-E10, L16-E17; on *path*2, intrachain salt bridge D4-L7 and interchain hydrogen bond T79A-E17B; on *path*3, interaction network formed by salt bridge L16-E17 and hydrogen bonds Q13-E17 and Q13-L16, interchain hydrogen bond T113A-L18B.
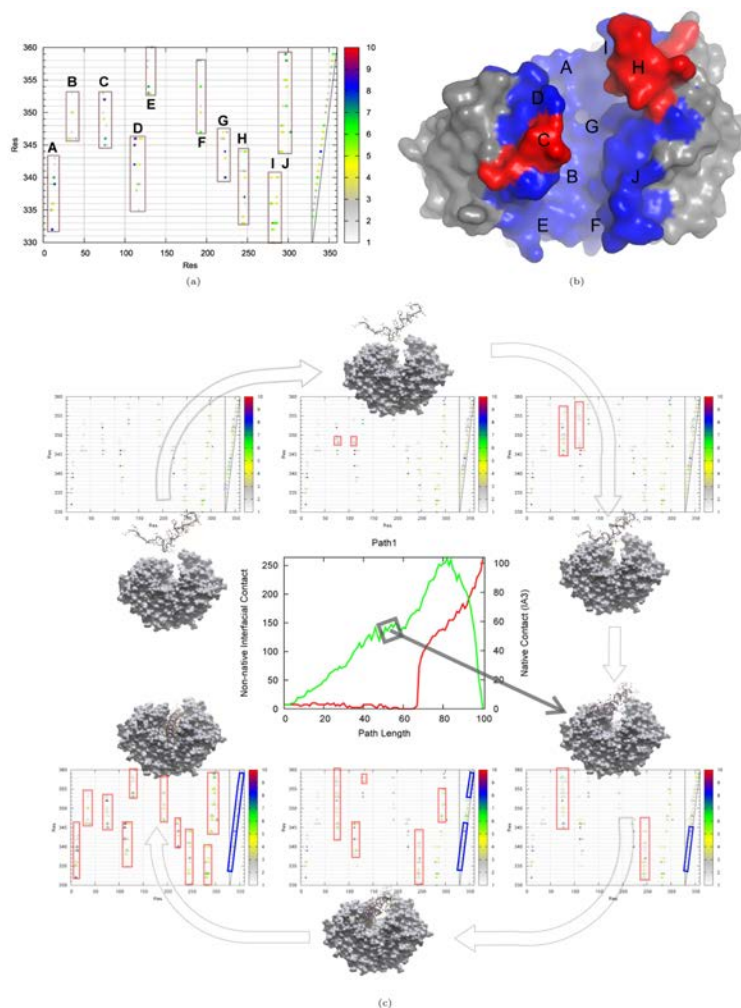
Figure 4. Native contact map of interface between N terminal of IA3 and YPrA and its corresponding regions in YPrA and the contact map evolution along the pathway: *path*1. (a) Native contact map in which the interfacial contacts are located on the surface of active site groove in YPrA. (b) Corresponding regions labelled A to J according to the interfacial contact map. They are colored blue exclude that the two loop regions are red. C region is the "flap". H region is loop2. (c) Contact map evolution along *path*1 and represented structures.
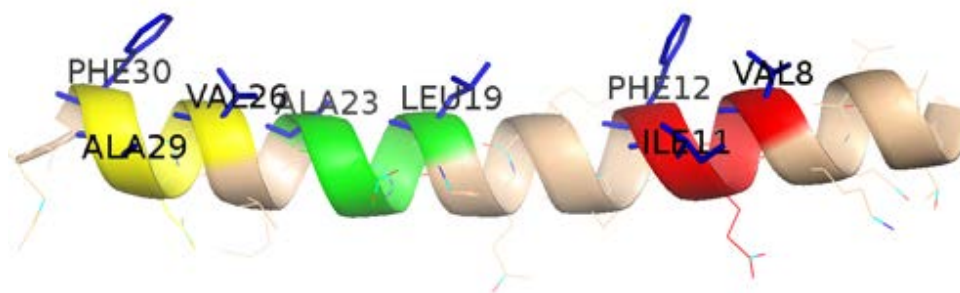
Figure 5. The helix structure of N terminal IA3 (residues 1-31) in cartoon. In the crystal structure of the complex, the hydrophilic face of IA3 is oriented toward the solvent, the other face enveloped completely with the residues of the active site cleft, is composed of nine hydrophobic amino acid residues, V8, I11, F12, L19, A23, V25, V26, A29 and F30. They constitute three hydrophobic clusters, "cluster-1" (red) of V8-X-X-I11-F12 in the N-terminal, "cluster-2"(green) of L19-X-X-X-A23 in the mid, and the C-terminal "cluster-3"(yellow) of V26-X-X-A29-F20.
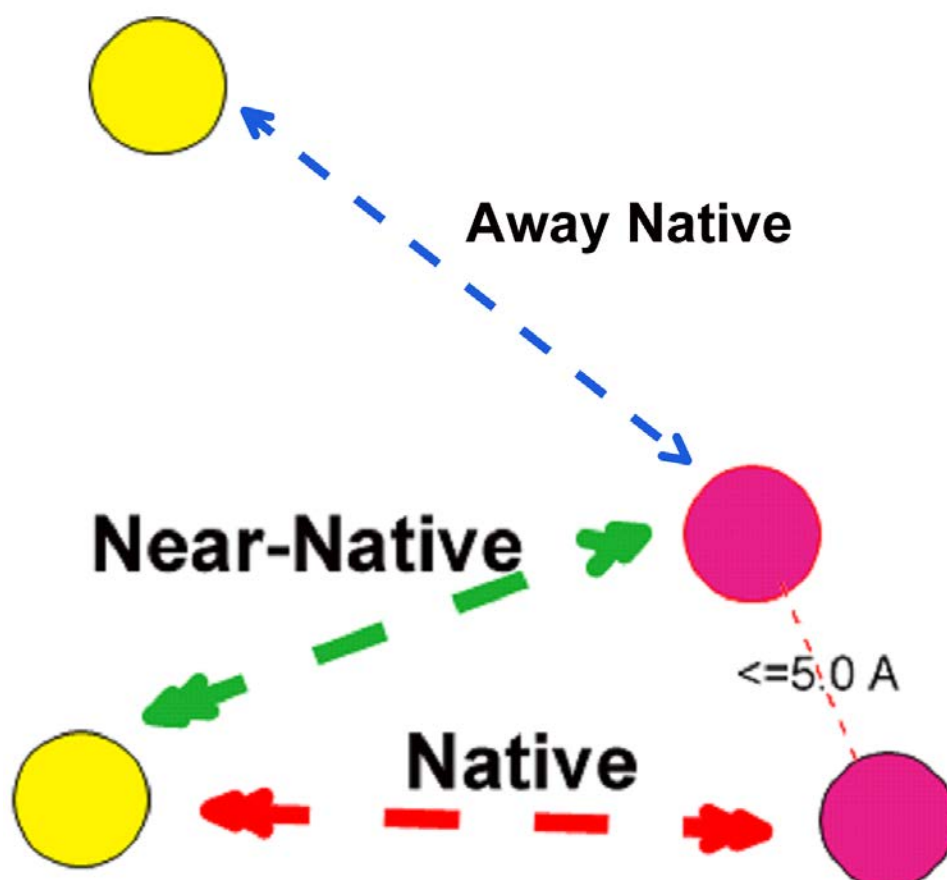
Figure 6. Schematic diagram of contacts calculation by cut-off algorithm. We used two radii to define the contact. If the distance of two C_α atoms is shorter than 6 $\mathring{A}$, we set the contact value 1, while the distance is between 6 $\mathring{A}$ and 10.60 $\mathring{A}$, which is 1.2 times average distance of two C_α atoms at native state, the contact value is 0.5. We separated the non-native contacts into two parts: near-native and away-native non-native contacts. If a non-native contact pair whose Ca atoms are not both within 5.0 Angstrom of any atoms in a native contact pair, this non-native contact is considered as away-native. Otherwise, it is near-native. Residues in IA3 and YPrA are represented by magenta and yellow spheres, respectively.