Chapter 9
Binding Induced Intrinsically Disordered Protein Folding with Molecular Dynamics Simulation

Haifeng Chen

Abstract Intrinsically disordered proteins lack stable tertiary and/or secondary structures under physiological conditions in vitro. Intrinsically disordered proteins undergo significant conformational transitions to well folded forms only on binding to partner. Molecular dynamics simulations are used to research the mechanism of folding for intrinsically disordered protein upon partner binding. Room-temperature MD simulations suggest that the intrinsically disordered proteins have non-specific and specific interactions with the partner. Kinetic analysis of high-temperature MD simulations shows that bound and apo-states unfold via a two-state process, respectively. Φ-value analysis can identify the key residues of intrinsically disordered proteins. Kolmogorov-Smirnov (KS) *P* test analysis illustrates that the specific recognition between intrinsically disordered protein and partner might follow induced-fit mechanism. Furthermore, these methods can be widely used for the research of the binding induced folding for intrinsically disordered proteins.

Keywords IDPs · Molecular dynamics simulations · Induced-fit mechanism

9.1 Introduction

Intrinsically disordered proteins lack stable tertiary and/or secondary structures under physiological conditions in vitro [1]. A large number of proteins (between 25 and 41 %) are intrinsically disordered. If the dogma dedicates that proteins need a structure to function, why do so many proteins live in the disorder state? [2] However, these intrinsically disordered proteins also play key function in regulation, signaling, and control upon binding with multiple interaction partners [3].

H. Chen (✉)
State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
e-mail: haifengchen@sjtu.edu.cn

D. Wei et al. (eds.), *Advance in Structural Bioinformatics*, Advances in Experimental Medicine and Biology 827, DOI 10.1007/978-94-017-9245-5_9

These proteins have many names, like rheomorphic, flexible or highly flexible, natively denatured, natively unfolded, intrinsically unstructured, intrinsically disordered. These proteins composed of an ensemble of highly heterogeneous conformations. After statistics of disordered protein database, IDPs include significantly higher levels of polar amino acids for Glu, Lys, Arg, Gln, Ser, Asp and Pro, and lower levels of hydrophobic residues for Ile, Leu, Val, Trp, Phe, Tyr, Thr, Met, Cys, His and Asn [4].

Furthermore, regions of disorder are found to be abundant in proteins associated with signaling, cancer, cardiovascular disease, amyloidoses, neurodegenerative diseases, and diabetes [5]. Different from structural protein as drug target, IDPs as drug target can bring low binding affinity and low side effect. There are two strategies for drug design targeting IDPs. Firstly, drug is binding to structured partner, thereby preventing the binding of the disordered partner. Secondly, drug is binding directly to the disordered partner, thereby preventing the association of two proteins. For this approach both partners were disordered, but small molecules bound to one of the two partners only. For example, c-Myc-Max inhibitors bind to distinct ID regions of c-Myc [6, 7]. These binding sites are composed of short contiguous stretches of amino acids that can selectively and independently bind small molecules. Inhibitor binding induces only local conformational changes, preserves the overall disorder of c-Myc, and inhibits dimerization with Max.

Furthermore, many intrinsically disordered proteins undergo significant conformational transitions to well folded forms only on binding to target ligands [8–11]. These experimental observations raise a set of interesting questions if these intrinsic disordered proteins obey an induced fit upon binding.

Coarse-grained modeling simulation [12] and all-atomic model with high temperature simulation [13] were used in intrinsically disordered protein folding coupled partner binding. So far the folding time scales of all atomic MD simulations are restricted to microsecond magnitude at room temperature (298 K), which is significant shorter than the folding half times of most proteins [14, 15]. In order to reveal the conformational changes within reasonable time, all MD simulations in explicit solvent at high temperature have been widely used to monitor the unfolding pathways of proteins. The unfolding timescales could be nanosecond at 498 K [14, 16]. Moreover, according to the principle of microscopic reversibility, experiments have demonstrated that the transition state for folding and unfolding is supposed to be same [14]. Therefore, MD simulations high temperatures are widely used in the folding of intrinsically disordered proteins coupled partner binding.

9.2 Materials and Method

The atomic coordinates of intrinsically disordered proteins were obtained from pdb data bank. Point mutants were modeled with SCWRL3 [17]. All hydrogen atoms were added using the LEAP module of AMBER [18]. Counter-ions were used to

maintain system neutrality. All systems were solvated in a truncated octahedron box of TIP3P waters with a buffer of 10 Å [19]. Particle Mesh Ewald (PME) [20] was applied to handle long-range electrostatic interactions with default setting in AMBER [18]. The parm99 force filed was used to compute the interactions within protein [21]. The SHAKE algorithm [22] was employed to constrain bonds including hydrogen atoms. All solvated systems were first minimized by steepest descent to remove any structural clash, followed by heating up and brief equilibration in the NPT ensembles at 298 K. The time step was 2 fs with a friction constant of 1 ps^{-1} using in Langevin dynamics. To study the folded state of each solvated system, multiple independent trajectories in the NPT ensemble at 298 K were simulated with PMEMD of AMBER. Then multiple independent unfolding trajectories were performed to investigate unfolding pathways for each solvated system in the NVT ensemble.

9.3 Results

TIS11d, KID, LEF, p53, CBP, and Brinker are partially or fully intrinsically disordered proteins. [13, 23–27] As transcription factor, they play key roles in signal transduction. Upon binding with DNA, RNA, or other transcription factors, they can well fold and will be introduced in this book. Their complex structures are illustrated in Fig. 9.1.

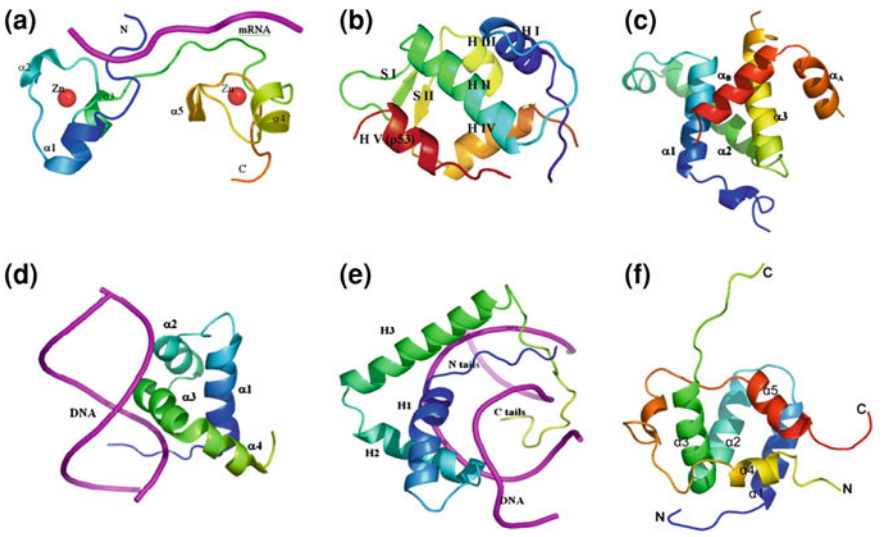

Fig. 9.1 The complex structure of intrinsically disordered proteins. **a** TIS11d/mRNA. **b** p53/MDM2. **c** pKID/KIX. **d** Brinker/DNA. **e** LEF/DNA. **f** p53/CBP

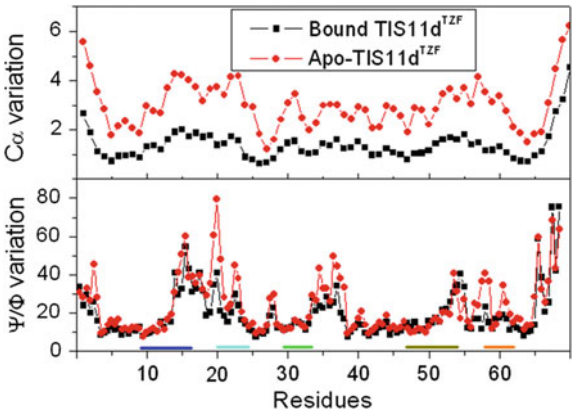

Fig. 9.2 Cα and Φ/Ψ variations for TIS11d

To capture the average properties of proteins, multiple trajectories for MD simulations (5–10) are necessary [28]. To study the recognition for intrinsically disordered proteins, multiple independent trajectories for apo-states and their complex were simulated at room temperature (298 K), respectively. Cα and Φ/ψ fluctuations for apo and bound states are researched. In general, the Cα variations of bound state are significant smaller than those of apo-state, especially in the region of the binding site. The results of apo and bound TIS11d are shown in Fig. 9.2 [26]. The Cα fluctuation of bound TIS11d^{TZF} is much smaller than that of apo-TIS11d^{TZF}, especially in the binding site of mRNA and zinc. This suggests that bound TIS11d^{TZF} become less flexible and more stable upon mRNA and zinc binding, which is consistent with the experiment. However, the Φ/ψ variation of bound TIS11d^{TZF} is similar to that of apo-TIS11d^{TZF}, suggesting that the secondary structure of bound TIS11d^{TZF} does not significantly change upon mRNA and zinc binding. Indeed, the helices of α1, α3 and α4 are already stable within apo-TIS11d^{TZF}.

To clearly illustrate the conformational difference, the landscapes of distance difference between the average pairwise intra-molecular distance of bound states and corresponding average pairwise intra-molecular distance of apo states for intrinsically disordered protein are shown in Fig. 9.3 [24]. The landscapes can reflect the relative conformational change of DNA and LEF backbone. The deep red area indicates that the distance difference for bases 5–8 and 23–26 is positive value. These bases are corresponding to the minor groove. This suggests that the minor groove is widened upon LEF-binding. Furthermore, disordered C-tail of LEF is located at the minor groove. This suggests that the disordered C-tail of LEF has interactions with DNA and open the minor groove of DNA. The deep blue area represents that the distance difference is negative value. It suggests that the major groove is contracted. That is consistent with the experimental observation that DNA is bended upon LEF-binding [29, 30]. For LEF, the deep red and blue areas are locked at disordered C-tail. This suggests that C-tail of LEF has significant conformational change.

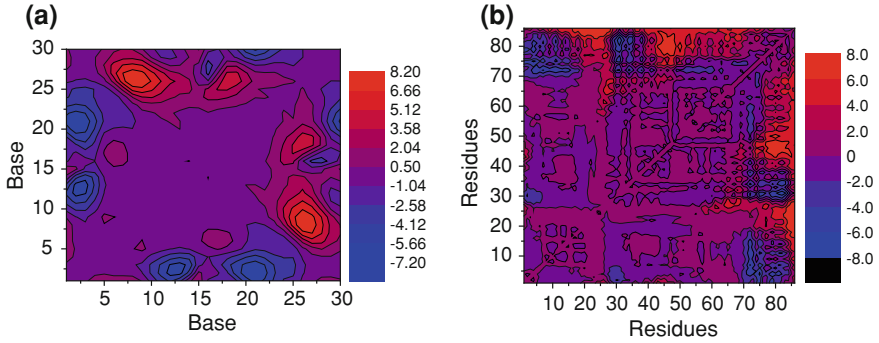

Fig. 9.3 Distance difference landscapes for DNA and LEF. **a** DNA. **b** LEF

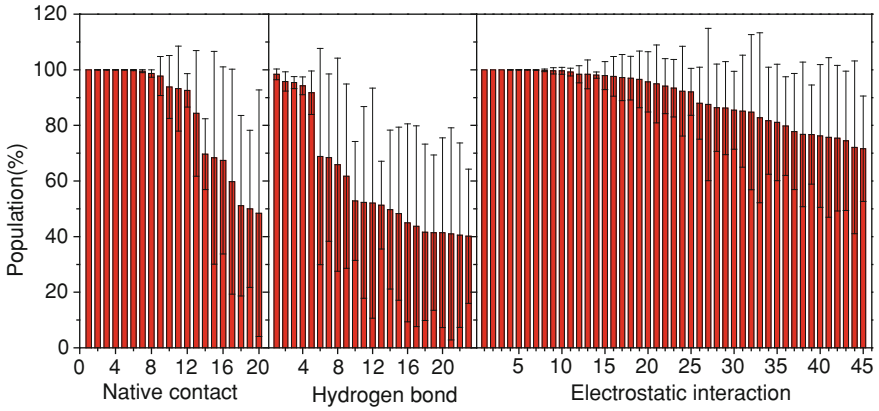

Fig. 9.4 Interactions between LEF and DNA

To study the driving force for these conformational adjustments, the electrostatic, hydrophobic, and hydrogen-binding interactions between intrinsically disordered protein and partner were analyzed and shown in Fig. 9.4. From this figure, stable electrostatic interactions, hydrogen bonds, and hydrophobic interactions can be calculated. In general, partner binding will introduce more electrostatic interactions, native contacts and hydrogen bonds at the interface which are responsible for the higher stability for intrinsically disordered proteins.

9.3.1 Unfolding Kinetics

High temperature simulation was used to research the unfolding kinetics of intrinsically disordered proteins with the parameters of the fraction of native tertiary contact (Qf) and the fraction of native binding contact (Qb). Time evolutions

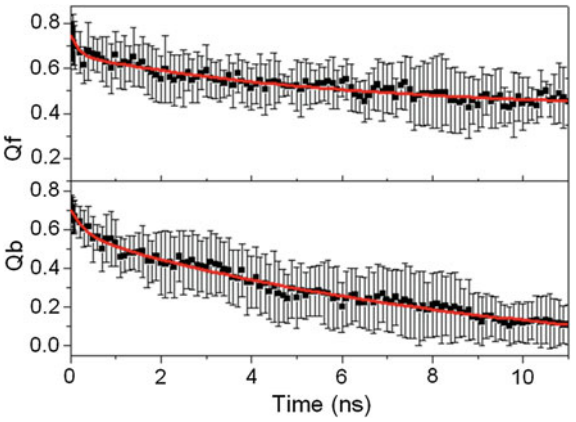

Fig. 9.5 Unfolding kinetics for bound pKID

of Qb and Qf for apo and bound states are shown in Fig. 9.5 [23]. The tertiary unfolding and unbinding can be fitted well by a single exponential function, indicating first order kinetics in the NVT ensemble at high temperature (498 K). This suggests that the binding of partner significantly postpones the tertiary unfolding of intrinsically disordered proteins. This is in agreement with the experimental observations [8, 31].

9.3.2 Φ-Value Prediction

Φ values have been widely used by theoretical and experimental works to identify the key residues for protein folding [32–34]. The Φ values of pKID were predicted and shown in Fig. 9.6. Note also that the highest Φ values are found for Asn139, Asp140 and Leu141, suggesting these residues play key role in the folding of

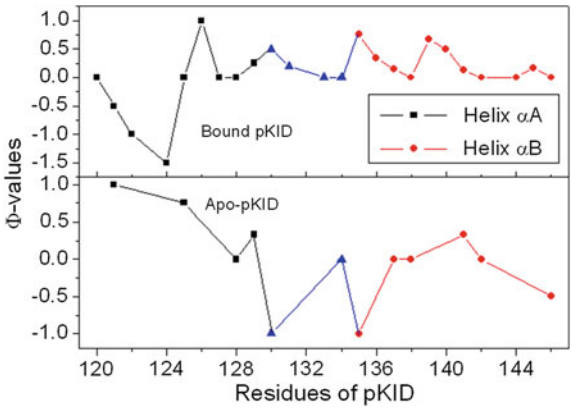

Fig. 9.6 Φ-values for bound and apo pKID

pKID [23]. A critical role of Leu141, which deeply extends into the hydrophobic groove of KIX, forms three hydrophobic contacts with KIX. All predicted Φ values can be confirmed by experiments.

9.3.3 Unfolding Pathway

According to the unfolding kinetics analysis, the unfolding orders of bound intrinsically disordered protein are shown in Fig. 9.7 [13]. If we assumed folding is reverse of unfolding, the proposed folding pathway of bound intrinsically disordered protein is from the unfolded state, then secondary structure folding, tertiary folding, partner binding, then to the folded state.

9.3.4 Recognition Mechanism

Conformational selection and induced fit are two widely used models to interpret the recognition between intrinsic disordered proteins [35]. According to the conformational selection paradigm, various conformational ensembles explore the free energy landscapes corresponding to diverse stable unbound states in equilibrium. During the binding process, the favorable conformation compatible with binding selectively stabilize, and the populations of conformational ensembles shift towards stabilizing state [36–39]. However, the induced fit scenario proposes that the favorable conformation results from significant changes of unbound ensembles upon allosteric binding [40–43]. It is worthy to point out that conformational selection and induced fit models cannot be distinguished absolutely [44]. Indeed, some systems involve kinetic elements of both mechanisms [45, 46].

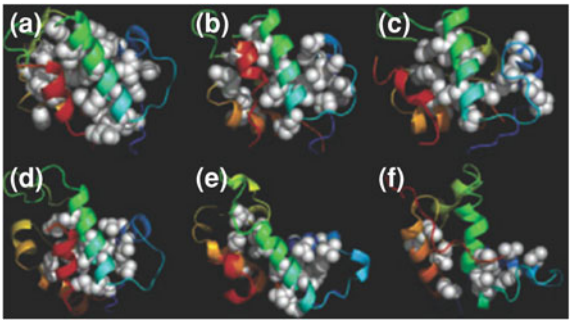

Fig. 9.7 Unfolding pathway for bound p53. **a** fold state. **b** unbinding. **c** tertiary unfolding. **d** helix 3/5 unfolding. **e** helix 1/2/4 unfolding. **f** unfolded state

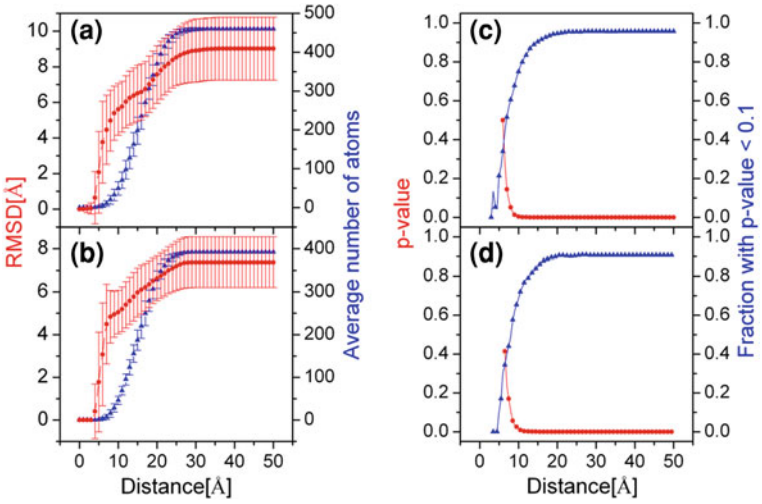

Fig. 9.8 Local conformational RMSD differences between bound and apo conformations as a function of distance from the centroid of binding partner and statistical significance of conformational selection in p53 and CBP binding. Average local RMSD for 10 pairs of bound conformations and the most similar apo conformation and for 90 pairs of bound NCBD and the other apo conformations, as a function of distance from the centroid of binding partner. **a** CBP. **b** p53. **c** CBP. **d** p53

The possible magnitudes of conformational selection and induced fit [47] are calculated to reveal the recognition mechanism. To explore the recognition mechanism, the average RMSD deviations of bound conformation and apo conformations are analyzed as a function of distance from the centroid of binding partner and shown in Fig. 9.8 [27]. This figure illustrates that the RMSD variation gradually increases until to the global level. This suggests that there is an induced fit far away for the binding site.

To address the statistical significance for differences of deviations between these two systems, two sample Kolmogorov-Smirnov test [48] is used to calculate the P value for each distance group. Figure 9.8c illustrates the median of P values and the fraction with $P < 0.1$ for all 100 pairs of CBP conformations in each distance group. It is found that the median P values are typically smaller than 0.1 in most distance group, especially in some larger distance group with median P values approximates 0. The conformations with $P < 0.1$ exceed 50 % in most distance groups. This suggests that the bound CBP is significant different from the apo conformation far away from the binding site and the differences are statistically significant. In summary, the recognition between intrinsic disordered CBP and p53 might obey an induced fit based on the RMSD and P-value analysis.

9.4 Conclusion and Remark

Intrinsically disordered proteins lack stable tertiary and/or secondary structures under physiological conditions in vitro. Intrinsically disordered proteins undergo significant conformational transitions to well folded forms only on binding to partner. Molecular dynamics simulations are used to research the mechanism of folding for intrinsically disordered protein upon partner binding. Room-temperature MD simulations suggest that the intrinsically disordered proteins have non-specific and specific interactions with the partner. Kinetic analysis of high-temperature MD simulations shows that bound and apo-states unfold via a two-state process, respectively. Φ-value analysis can identify the key residues of intrinsically disordered proteins. Kolmogorov-Smirnov (KS) *P* test analysis illustrates that the specific recognition between intrinsically disordered protein and partner might follow induced-fit mechanism. Furthermore, these methods can be widely used for the research of the binding induced folding for intrinsically disordered proteins.

References

1. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246
2. Chouard T (2011) Structural biology: breaking the protein rules. Nature 471:151–153
3. Liu J, Faeder JR, Camacho CJ (2009) Toward a quantitative theory of intrinsically disordered proteins and their function. Proc Natl Acad Sci USA 106:19819–19823
4. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59
5. Metallo SJ (2010) Intrinsically disordered proteins are potential drug targets. Curr Opin Chem Biol 14:481–488
6. Wang H, Hammoudeh DI, Follis AV, Reese BE, Lazo JS, Metallo SJ, Prochownik EV (2007) Improved low molecular weight Myc-Max inhibitors. Mol Cancer Ther 6:2399–2408
7. Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. J Am Chem Soc 131:7390–7401
8. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021–1025
9. Canon F, Ballivian R, Chirot F, Antoine R, Sarni-Manchado P, Lemoine J, Dugourd P (2011) Folding of a salivary intrinsically disordered protein upon binding to tannins. J Am Chem Soc 133:7847–7852
10. Wang J, Wang Y, Chu X, Hagen SJ, Han W, Wang E (2011) Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target enzyme. PLoS Comput Biol 7:e1001118
11. Zhang W, Ganguly D, Chen J (2012) Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins. PLoS Comput Biol 8:e1002353

12. Chen J (2009) Intrinsically disordered p53 extreme C-terminus binds to S100B(betabeta) through "fly-casting". J Am Chem Soc 131:2088–2089
13. Chen HF, Luo R (2007) Binding induced folding in p53-MDM2 complex. J Am Chem Soc 129:2930–2937
14. Fersht AR, Daggett V (2002) Protein folding and unfolding at atomic resolution. Cell 108:573–582
15. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE (2009) Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 19:120–127
16. Shea JE, Brooks CL 3rd (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 52:499–535
17. Canutescu AA, Shelenkov AA, Dunbrack RL Jr (2003) A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 12:2001–2014
18. Case DA, Darden TA, Cheatham TE, Simmerling ICL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2010) Amber 11, University of California, San Francisco
19. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926
20. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089
21. Wang JM, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. J Comput Chem 21:1049–1074
22. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
23. Chen HF (2009) Molecular dynamics simulation of phosphorylated KID post-translational modification. PLoS ONE 4:e6516
24. Qin F, Ye W, Chen Y, Chen X, Li Y, Zhang J, Chen HF (2012) Specific recognition between intrinsically disordered LEF and DNA. Phys Chem Chem Phys 14:538–545
25. Qin F, Jiang Y, Chen Y, Wu M, Yan G, Ye W, Li Y, Zhang J, Chen HF (2011) Conformational selection or induced fit for Brinker and DNA recognition. Phys Chem Chem Phys 13:1407–1412
26. Qin F, Chen Y, Li YX, Chen HF (2009) Induced fit for mRNA/TIS11d complex. J Chem Phys 131:115103
27. Yu Q, Ye W, Wang W, Chen HF (2013) Global conformational selection and local induced fit for the recognition between intrinsic disordered p53 and CBP. PLoS ONE 8:e59627
28. Day R, Daggett V (2005) Ensemble versus single-molecule protein unfolding. Proc Natl Acad Sci USA 102:13445–13450
29. Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature 376:791–795
30. Love JJ, Li X, Chung J, Dyson HJ, Wright PE (2004) The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. Biochemistry 43:8725–8734
31. Radhakrishnan I, Perez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator: coactivator interactions. Cell 91:741–752
32. Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. Proc Natl Acad Sci USA 97:1525–1529

33. Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L (2004) Solvation in protein folding analysis: combination of theoretical and experimental approaches. Proc Natl Acad Sci USA 101:2834–2839
34. Fersht AR, Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci USA 101:7976–7981
35. Boehr DD, Wright PE (2008) Biochemistry. How do proteins interact? Science 320:1429–1430
36. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254:1598–1603
37. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci USA 96:9970–9972
38. Boehr DD, McElheny D, Dyson HJ, Wright PE (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. Science 313:1638–1642
39. Weikl TR, von Deuster C (2009) Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. Proteins 75:104–110
40. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA 44:98–104
41. Rini JM, Schulze-Gahmen U, Wilson IA (1992) Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 255:959–965
42. Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-induced folding of a natively unstructured transcription factor. PLoS Comput Biol 4:e1000060
43. Schrank TP, Bolen DW, Hilser VJ (2009) Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. Proc Natl Acad Sci USA 106:16984–16989
44. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 35:539–546
45. James LC, Tawfik DS (2003) Conformational diversity and protein evolution–a 60-year-old hypothesis revisited. Trends Biochem Sci 28:361–368
46. Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. Proc Natl Acad Sci USA 105:11182–11187
47. Wlodarski T, Zagrovic B (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. Proc Natl Acad Sci USA 106:19346–19351
48. Massey Jr, FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. J Am stat Assoc, 68–78