Deciphering the mechanisms of binding induced folding at nearly atomic resolution: The Φ value analysis applied to IDPs

Stefano Gianni1,2,*, Jakob Dogan3, and Per Jemth3,*

1Dipartimento di Scienze Biochimiche "A. Rossi Fanelli"; Istituto di Biologia e Patologia Molecolari del CNR; Università di Roma "La Spaienza"; Rome, Italy; 2Department of Chemistry; University of Cambridge; Cambridge, UK; 3Department of Medical Biochemistry and Microbiology; Uppsala University; Uppsala, Sweden

Keywords: kinetics, mutagenesis, structure, thermodynamics, transition state

Introduction

The ultimate goal of the biophysicist is to provide experimental depictions of entire pathways. But, in practice, observed reactions often occur in a co-operative manner and only a very limited number of intermediates may be detected. Thus, the information that is accessible to the experimentalist is generally very limited and only few snapshots between reactants and products might be characterized. The binding induced folding reaction of intrinsically disordered proteins (IDPs) does not represent an exception to this statement. In fact, the recognition between IDPs and their partners is a complex reaction involving, in theory, at least a folding and a binding step.1,2 Yet, it is striking to observe that folding and binding often occur in a concerted manner, such that all the experimental probes may be consistent with an all-or-none reaction, where only the fully unstructured IDP in isolation and the fully folded bound state may be identified.3-6 In other cases, the experimental data are consistent with multiple binding steps, with different rate-limiting transition states under different conditions (e.g., ligand concentration).7,8 In all cases, the study of the transition state(s) of the reaction becomes critical to capture the residues driving the binding-induced-folding reaction of IDPs.

Since the transition state never accumulates, information about its structure can only be obtained indirectly.9 In organic chemistry, the fraction of bond formation in transition states is generally described using the so-called Leffler α- or Brønsted β-values.10 By following this technique, a chemical modification is made in a non-reacting part of the reagent, and the effect of the modification on the free energy of activation and the free energy of equilibrium is measured. The α- or β-value is the fraction between these 2 free energy terms and reflects the extent of covalent bond formation in the reaction.

In the case of proteins, the usage of site-directed mutagenesis allows performing a similar technique.11-14 In fact, in analogy to the α- or β-value analysis, by systematically mutating protein residues, while probing the effect of the mutation on the folding kinetics and native state stability, it is possible to map, one by one, interaction patterns in the transition states (Fig. 1). Mutations that destabilize the transition state (slowing down the folding reaction) target contacts that are formed in its structure. The relative formation of the contact is commonly called the Φ value. By producing and characterizing a large number of point mutants in a given protein it is therefore possible to draw a structural map of the transition state, with detection of native like (Φ values tending to 1) and denatured like (Φ values tending to 0) clusters. This is commonly called “the Φ value analysis.”11-14

Because of its power and very high resolution, the Φ value analysis has been widely employed in protein folding studies and, we argue, will represent a future direction for the analysis of the interactions between IDPs and their partners. In this review, we will briefly describe the basic equations of the Φ value analysis,
the possible complications arising when studying IDPs along with a few examples, which can be already found in the literature.

The Basic Principles of Φ Value Analysis: Equations

The Φ value analysis is based on measuring the effect of mutations on protein folding kinetics and equilibrium. It is therefore important to consider that, when mutating the native state \( N \) into \( N' \), the experimentally measurable quantities of the wild-type and mutant proteins are the \( \Delta G_{D-N} \), the \( \Delta G_{D'-N'} \), the \( \Delta G_{D-TS} \), the \( \Delta G_{D'-TS'} \), the \( \Delta G_{TS-N} \), and the \( \Delta G_{TS'-N'} \), where the symbol ‘\( \dagger \)’ denotes the mutant protein (Fig. 1). Thus a change in native stability upon mutation may be calculated as:

\[
\Delta \Delta G_{D-N} = \Delta G_{D'-N'} - \Delta G_{D-N}
\]

and the change in activation energy is:

\[
\Delta \Delta G_{D-TS} = \Delta G_{D'-TS'} - \Delta G_{D-TS}
\]

Quantitatively, the Φ value for folding is defined as:

\[
\Phi = \frac{\Delta \Delta G_{D-TS}}{\Delta \Delta G_{D-N}}
\]

Note that, in case of proteins displaying a residual structure in their denatured state, the mutation may also cause a change in its free energy, defined as \( \Delta G_{D-D} \), which, as discussed elsewhere, could potentially jeopardize a rigorous interpretation. These complications, however, tend to cancel out when the value of Φ tends to 0 or 1 and all mutations return an interpretable value.

In pure protein folding studies, the measured parameters are the folding and unfolding rate constants \( k_f, k_u \) and the equilibrium constant \( K \) (either from the kinetic constants or independently from for example urea or heat denaturations at equilibrium). In the case of IDPs and folding-induced binding we need to measure the association and dissociation rate constants \( k_{on}, k_{off} \) and the equilibrium constant \( K_d \), either from the rate constants or independently by for example isothermal titration calorimetry or fluorescence-monitored titrations. However, \( \Delta \Delta G \) values should preferably be obtained from kinetic rate constants since these are very precise and accurate. Thus, \( K_d \) values are obtained from the ratio \( k_{off}/k_{on} \). It is a common misconception that Φ value analyses can only be applied to two-state systems. Neither a folding nor binding needs to be two state to be amenable to Φ value analysis. Even overall \( K_d \) values can usually be correctly calculated from apparent \( k_{on} \) and \( k_{off} \) values, despite a multi-step binding mechanism, since the influence of each step is incorporated into the observed rate constants. It is however important to remember that what we characterize is the structure of the rate-limiting transition state for the overall binding reaction in the concentration range where we collect data. Thus, careful interpretation of data is imperative.

The Basic Principles of Φ Value Analysis: Choice of Mutations

It is of critical importance to define how to make mutations that are suitable for the Φ value analysis. In fact, when a side-
A comprehensive kinetic mechanism describing a monomeric intrinsically disordered protein undergoing a ligand induced conformational change is described by a square scheme, as exemplified in Scheme 1.

Within this context a binding event progressing through pathways 1 and 2 is representative of an induced-fit model, whereby ligand binding induces a conformational change. Alternatively, a binding event progressing through pathways 3 and 4 assumes that two alternative conformations are in pre-equilibrium in the absence of the ligand, formally similar to a concerted Monod-Wyman-Changeux (MWC) model. However, it should be noticed that the different order of events implied by the MWC and induced fit mechanisms do not unambiguously assign the rate-determining step and different scenarios are possible. A plausible bi-dimensional free energy diagram of the induced-fit and MWC scenarios involving folding or binding as different rate limiting steps is reported in Figure 2.

We will now briefly analyze the different kinetic behavior expected from the different scenarios depicted in Figure 2 and its implication for the \( \Phi \) value analysis to an IDP system.

**Folding before binding**

The folding before binding scenario implies the IDP to explore the folded conformation in the absence of ligand. Rapid mixing of the ligand shifts the equilibrium toward the complex, thereby promoting folding. Under such conditions, 2 possible scenarios may occur: (1) folding might be a slow step prior to fast binding (Fig. 2A) or (2) fast folding might precede slow binding (Fig. 2B).

If folding were slow, when a binding experiment is performed by challenging the protein with increasing concentrations of ligand, the observed rate constant would be equal to

\[
\text{k}_{\text{obs}} = \text{k}_F
\]

(4)

Observed kinetics would therefore appear independent on ligand concentration and the analysis of rate constants would not allow inferring information about the stability of the complex. In these cases, because slow folding is rate limiting, we predict the \( \Phi \) values to resemble what is classically observed in protein folding studies, with clusters of native-like structure forming in the transition state, that is expected to resemble a distorted version of the native structure. Finally, it should be noticed that, because the folding rate constant correlate with topological complexity of the native state, the presence of a slow folding step prior to binding is expected to be very rare in the case of small IDPs that display a simple topology in their folded state.

An alternative possibility of the folding before binding scenario implies the folding reaction to be faster than binding (\( k_F \), \( k_D \gg k_{\text{on}} \)), where \( k_{\text{on}} = k_{\text{off}} \times [L] \) (Fig. 2B). Under such conditions, the apparent bi-molecular rate constant would be a combination of microscopic rate constants as formalized below:

\[
k^{\text{app}}_{\text{on}} = \frac{k_{\text{on}}}{1 + K_D - N}
\]

(5)
In a disordered conformation, and its partner. Under such conditions, if folding \( k_{f} \) were slow in relation to dissociation \( (k_{off}) \) (Fig. 2C), the apparent association rate constant would be equal to

\[
k_{on}^{app} = \frac{k_{on}}{1 + \frac{k_{off}}{k_{f}}}
\]

where \( k_{on} \) and \( k_{off} \) represent respectively the rate constants of association and dissociation between the disordered state and the ligand, and \( k_{f} \) the folding rate constant. On the other hand, if the folding rate constant were larger than the microscopic \( k_{off} \) of binding (Fig. 2D), the apparent association rate constant will be equal to the microscopic \( k_{on} \). We note that these 2 different scenarios are expected to return very different clusters of \( \Phi \) values. In the latter case the protein is expected to be largely unfolded in the transition state and \( \Phi \) values calculated from mutations reporting on the folding of the IDP are therefore expected to return low values. In the former case the main rate limiting barrier is associated to a folding step and \( \Phi \) values may be interpreted as genuine folding probes. Accordingly, observation of relatively high values of \( \Phi \), as observed for example in the binding between KIX and the transactivation domain of KIX,\(^{26}\) may suggest that \( k_{f} < k_{off} \) and allows to exclude the presence of a fast folding step occurring after slow binding.

### Examples in Literature

Our laboratories have recently contributed 2 of the first examples of \( \Phi \) value analyses on IDPs. In this section we will briefly describe the major conclusions of these studies.

The CREB-binding protein (CBP) is a co-activator that modulates the interaction between DNA-bound activator proteins and the components of the basal transcription complex.\(^{27}\) A globular domain of CBP, namely the KIX domain, is one mediator of such interactions.\(^{28}\) The KIX domain binds different IDP systems via two distinct, but energetically connected, binding sites, called “c-Myb” and “MLL” sites.\(^{29}\) The interaction between KIX and the transactivation domain of c-Myb, which folds into a helical structure upon binding following a folding-after-binding scenario,\(^{3} \) has been studied by \( \Phi \) value analysis.\(^{26} \) It is of interest to discuss both the structural distribution and the magnitude of the observed \( \Phi \) values measured for this system. In fact, mapping the measured \( \Phi \) values on the structure between KIX and c-Myb suggests that c-Myb folds via 2 distinct nuclei displaying medium or high \( \Phi \)-values located at the N- and C-terminal ends of the helix, with a region with lower values of \( \Phi \) was located at the center of the helix. Importantly, some of the residues displaying high \( \Phi \) values did not make direct contact with KIX in the complex, providing additional support to the proposal that the \( \Phi \) values for the interaction between KIX and c-Myb are dominated by the intra- rather than intermolecular contacts and confirming the reaction to be rate limited by the folding of c-Myb rather than by recognition of KIX (Fig. 2C). Furthermore, an analysis using Ala to Gly scanning of the helical structure of c-Myb revealed the...
central part of the IDP to display $\Phi > 1$, which is classically interpreted as a signature of misfolding in the transition state. Of additional interest, it was noted that the average $\Phi$ value for the recognition between KIX and c-Myb was in the order of 0.89. This number is in stark contrast with the value of ca. 0.3, which is typically observed in the folding of single domain globular proteins, and suggests that the transition state contains a very high degree of native-like structure. Because of these findings, it was concluded that c-Myb recognizes KIX with a high degree of geometrical precision, which appears incompatible with models suggesting protein disorder to be a mechanism to speed up partner recognition.

Another system that has been the subject of numerous studies in the last couple of years is the interaction between the nuclear coactivator binding domain (NCBD) of CBP, and the activation domain from the p160 transcriptional co-activator for thyroid hormone and retinoid receptors (ACTR).$^{31,32}$ It has been shown that NCBD binds a diverse set of proteins, including IDPs and folded proteins, and that it also adopts different bound conformations depending on interacting ligand.$^{33}$ NCBD has molten globular properties, whereas ACTR is completely disordered, and they synergistically fold upon interacting with each other.$^{34}$ A $\Phi$-value binding analysis was performed,$^{16}$ with hydrophobic mutations made at positions that are involved in intermolecular interactions on both NCBD and ACTR. With a few exceptions, the $\Phi$-values displayed low values with an average value of 0.14, suggesting that native hydrophobic contacts form late, at the downhill side of the rate-limiting barrier for association. This scenario, in contrast to that of KIX/c-Myb, is consistent with a rate limiting TS1 as depicted in Figure 2D. The highest $\Phi$-values were situated at the N-terminal helix of ACTR. Indeed, in a subsequent study,$^{35}$ where the secondary structure content of the N-terminal helix of ACTR was modulated by mutations at positions where no tertiary contacts are made, the $\Phi$-values were generally higher than those previously determined, and the authors were able to show that preformed secondary structure accelerates the binding to NCBD. NMR experiments revealed that changes in compaction due to the mutationally increased helix content was very small, thus the so-called fly casting effect$^{36}$ could not explain the observation that the association rate constant increased with increasing helicity. The high helix formation in the transition state is similar to the c-Myb/KIX system, but in contrast to the low $\Phi$-values for formation of hydrophobic contacts discussed above. Thus, folding of the N-terminal helix of ACTR is partially rate-limiting and possibly described by a different energy landscape than the intermolecular hydrophobic interactions.

Low $\Phi$-values for helix formation were also observed in a third $\Phi$-value analysis of an IDP binding reaction, that between the S-peptide interacting with the S-protein.$^{37}$ Furthermore, most native intermolecular contacts in the S-peptide/S-protein system are not present at the transition state, which is similar to NCBD/ACTR but distinct from c-Myb/KIX.

Overall, these studies show how carefully designed $\Phi$-value analyses can be used to understand the binding reactions of IDPs. The conclusion thus far is that binding precedes folding but there does not seem to exist any general trend on mechanisms for IDP/target interactions, in terms of formation of intra- and intermolecular interactions.$^{25}$

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

Work partly supported by grants from the Italian Ministero dell’Istruzione dell’Universita’ e della Ricerca (PNR-CNRS Aging Program 2012–2014 to S.G.), the Sapienza University of Rome (C26A13T9NB to S.G.) and EMBO (to S.G.), the Swedish Research Council (to P.J.), the O. E. and Edla Scientific Foundation (to J.D.), and the Magnus Bergwall Foundation (to J.D.).

References

1. Dunker AK, Silman I, Uversky VN, Sauman JL. Function and structure of inherently disordered proteins. Curr Opin Struct Biol 2008; 18:756-64; PMID:18952168; http://dx.doi.org/10.1016/j.cub.2008.10.002 2. Uversky VN, Dunker AK. Understanding protein non-folding. Biochim Biophys Acta 2010; 1804:1231-64; PMID:20117254; http://dx.doi.org/10.1016/j.bbadis.2010.01.017 3. Gianni S, Morrone A, Giri R, Brunori M. A folding-after-binding mechanism describes the recognition between the transactivation domain of c-Myb and the KIX domain of the CREB-binding protein. Biochim Biophys Acta 2012; 1823:85-94; PMID:225437; http://dx.doi.org/10.1016/j.bbamem.2012.08.002 4. Kiefhaber T, Bachmann A, Jensen KS. Dynamics and mechanisms of coupled protein folding and binding reactions. Curr Opin Struct Biol 2012; 22:21-9; PMID:22129832; http://dx.doi.org/10.1016/j.sbi.2011.09.010 5. Narayanan R, Ganesh OK, Edison AS, Hagen SJ. Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase YPA1. J Am Chem Soc 2008; 130:11477-85; PMID:18681437; http://dx.doi.org/10.1021/ja803221c 6. Rogers JM, Steward A, Clarke J. Folding and binding of an intrinsically disordered protein: fast, but not ‘diffusion-limited’. J Am Chem Soc 2013; 135:1415-22; PMID:23301700; http://dx.doi.org/10.1021/ja309527b 7. Dogan J, Schmidt T, Mu X, Engström A, Jensen P. Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. J Biol Chem 2012; 287:34316-24; PMID:22915388; http://dx.doi.org/10.1074/jbc.M111.399436 8. Chemes LB, Sánchez JE, de Prat-Gay G. Kinetic recognition of the retinoblastoma tumor suppressor by a specific protein target. J Mol Biol 2011; 412:267-84; PMID:21787785; http://dx.doi.org/10.1016/j.jmb.2011.07.015 9. Eyring H. The Activated Complex in Chemical Reactions. J Chem Phys 1935; 3:107-15; http://dx.doi.org/10.1063/1.1749604 10. Leffler JE. Parameters for the Description of Transition States. Science 1953; 117:340-1; PMID:17741025; http://dx.doi.org/10.1126/science.117.3039.340 11. Fersht AR, Matsosiek A, Sato S. Phi-value analysis and the nature of protein-protein folding transitions. Proc Natl Acad Sci USA 2004; 101:7976-81; PMID:15150406; http://dx.doi.org/10.1073/pnas.0402684101 12. Cho JH, Raleigh DP. Denatured state effects and the origin of nonclassical phi values in protein folding. J Am Chem Soc 2006; 128:16492-3; PMID:17177385; http://dx.doi.org/10.1021/ja0608578 13. Matsosiek A, Kells JTF Jr., Serrano I, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. Nature 1989; 340:122-6; PMID:2739734; http://dx.doi.org/10.1038/340122a0 14. Fersht AR. Sato S. Phi-value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci USA 2004; 101:7976-81; PMID:15150406; http://dx.doi.org/10.1073/pnas.0402684101 15. Cho JH, Raleigh DP. Denatured state effects and the origin of nonclassical phi values in protein folding. J Am Chem Soc 2006; 128:16492-3; PMID:17177385; http://dx.doi.org/10.1021/ja0608578 16. Dogan J, Mu X, Engström A, Jensen P. The transition state structure for coupled binding and folding of disordered protein domains. Sci Rep 2013; 3:2876; PMID:23799450; http://dx.doi.org/10.1038/srep02076 17. Zhou HX, Pang X, Lu C. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. Phys Chem Chem Phys 2012; 14:4066-76; PMID:22744607; http://dx.doi.org/10.1039/c2cp41196b
20. Kiel C, Selzer T, Shaul Y, Schreiber G, Herrmann C. Electrostatically optimized Ras-binding Ral guanine nucleotide dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proc Natl Acad Sci U S A 2004; 101:9223-8; PMID:15197281; http://dx.doi.org/10.1073/pnas.0401160101

21. Koshland DEJ Jr., Némethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 1966; 5:365-85; PMID:5938952; http://dx.doi.org/10.1021/bi00865a027

22. Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: a plausible model. J Mol Biol 1965; 12:88-118; PMID:14343300; http://dx.doi.org/10.1016/S0022-2836(65)80285-6

23. Shammas SL, Travis AJ, Clarke J. Remarkably fast coupling folding and binding of the intrinsically disordered protein accelerates ligand binding. Angew Chem Int Ed Engl 2014; 53:1548-51; PMID:24449148; http://dx.doi.org/10.1002/anie.201307712

24. Schreiber G, Hanan G, Zhou HX. Fundamental aspects of protein-protein association kinetics. Chem Rev 2009; 109:839-60; PMID:19156002; http://dx.doi.org/10.1021/cr800373w

25. Dogan J, Gianni S, Jerm P. The binding mechanisms of intrinsically disordered proteins. Phys Chem Chem Phys 2014; 16:6323-31; PMID:24317797; http://dx.doi.org/10.1039/c3cp54226b

26. Giri R, Morotne A, Toto A, Brunori M, Gianni S. Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. Proc Natl Acad Sci U S A 2013; 110:14942-7; PMID:23980173; http://dx.doi.org/10.1073/pnas.1307371110

27. Goto NK, Zor T, Martinez-Yamout M, Dyson HJ, Wright PE. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. J Biol Chem 2002; 277:43168-74; PMID:12205094; http://dx.doi.org/10.1074/jbc.M207660200

28. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ; Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell 1997; 91:741-52; PMID:9413984; http://dx.doi.org/10.1016/S0092-8674(00)80463-8

29. Zor T, De Guzman RN, Dyson HJ, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. J Mol Biol 2004; 357:521-34; PMID:15019774; http://dx.doi.org/10.1016/j.jmb.2004.01.058

30. Naganathan AN, Muoz V. Insights into protein folding mechanisms from large scale analysis of mutational effects. Proc Natl Acad Sci U S A 2010; 107:8611-6; PMID:20418505; http://dx.doi.org/10.1073/pnas.100088107

31. Chen H, Lin RJ, Schulz RL, Chakravarti D, Nash A, Nagy L, Privalsky ML, Nakarai Y, Evans RM. Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. Cell 1997; 90:569-80; PMID:9267036; http://dx.doi.org/10.1016/S0002-8674(00)80156-4

32. Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. Genes Dev 2000; 14:1553-77; PMID:10887150

33. Qin BY, Liu C, Simha H, Lam SS, Correia JJ, Derynck R, Lin K. Crystal structure of IRF-3 in complex with CBP. Structure 2005; 13:1269-77; PMID:16154084; http://dx.doi.org/10.1016/j.str.2005.06.011

34. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, Wright PE. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature 2002; 415:549-53; PMID:11823864; http://dx.doi.org/10.1038/415549a

35. Ieiamantavičius V, Dogan J, Jerm P, Teišum K, Kjaergaard M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. Angew Chem Int Ed Engl 2014; 53:1548-51; PMID:24449148; http://dx.doi.org/10.1002/anie.201307712

36. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A 2000; 97:8868-73; PMID:10908673; http://dx.doi.org/10.1073/pnas.160259600

37. Bachmann A, Wildemann D, Praetorius F, Fischer G, Kiefhaber T. Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. Proc Natl Acad Sci U S A 2011; 108:3952-7; PMID:21325613; http://dx.doi.org/10.1073/pnas.1012668108