Deciphering the mechanisms of binding induced folding at nearly atomic resolution – the Φ value analysis applied to IDP’s

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

¹Dipartimento di Scienze Biochimiche “A. Rossi Fanelli” and Istituto di Biologia e Patologia Molecolari del CNR, Università di Roma “La Sapienza”, P.le A. Moro 5, 00185, Rome, Italy.
²Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
³Department of Medical Biochemistry and Microbiology, Uppsala University, BMC Box 582, SE-75123 Uppsala, Sweden.
SUMMARY

The $\Phi$ value analysis is a method to analyze the structure of metastable states in reaction pathways. Such a methodology is based on the quantitative analysis of the effect of point mutations on the kinetics and thermodynamics of the probed reaction. The $\Phi$ value analysis is routinely used in protein folding studies and is potentially an extremely powerful tool to analyze the mechanism of binding induced folding of intrinsically disordered proteins. In this review we recapitulate the key equations and experimental advices to perform the $\Phi$ value analysis in the perspective of the possible caveats arising in intrinsically disordered systems. Finally, we briefly discuss some few examples already available in the literature.
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) \(^7\). 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 \(^8\). In organic chemistry, the fraction of bond formation in transition states is generally described using the so called Leffler \(\alpha\)- or Brønsted \(\beta\)-values \(^9\). 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 \(\alpha\)- or \(\beta\)-value is the fraction between these two 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 \(^10-13\). In fact, in analogy to the \(\alpha\)- or \(\beta\)-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 (Figure 1). Mutations that
destabilise 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 $\Phi$ 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 ($\Phi$ values tending to 1) and
denatured like ($\Phi$ values tending to 0) clusters. This is commonly called “the $\Phi$ value analysis” 10-
13.

Because of its power and very high resolution, the $\Phi$ 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 $\Phi$ 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 $\Phi$ value analysis – equations.**

The $\Phi$ value analysis is based on measuring the effect of mutations on protein folding kinetics and
equilibrium 10. 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 ’ denotes the
mutant protein (Figure 1). Thus a change in native stability upon mutation may be calculated as:

$$
\Delta \Delta G_{D-N} = \Delta G_{D'-N'} - \Delta G_{D-N}
$$

Equation 1.

and the change in activation energy is:
\[ \Delta \Delta G_{D-TS} = \Delta G_{D-TS} - \Delta G_{D-TS} \]  
Equation 2.

Quantitatively, the \( \Phi \) value for folding is defined as:

\[ \Phi = \frac{\Delta \Delta G_{D-TS}}{\Delta \Delta G_{D-N}} \]  
Equation 3.

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 \( \Phi \) 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 \( \Phi \) value analyses can only be applied to two-state systems. Neither a folding nor binding needs to be two state to be amenable to \( \Phi \) 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-chain is mutated, the free energy of each state may be affected as a result of essentially three components: $\Delta G_{\text{noncov}}$, due to the changes in non-covalent interactions; $\Delta G_{\text{reorg}}$, the change in free energy arising from the reorganization of structure upon mutagenesis; $\Delta G_{\text{solv}}$, the change in solvation energy upon mutagenesis. Because the aim of Φ value analysis is indeed to probe the fractional formation of non-covalent interactions in metastable states, the two terms $\Delta G_{\text{reorg}}$ and $\Delta G_{\text{solv}}$ should be minimized. Thus, the ideal mutations for Φ values are aliphatic to aliphatic deletion mutations. Mutations of a large hydrophobic side chain to a hydrophilic side chain and vice-versa lead to very large $\Delta G_{\text{reorg}}$ and $\Delta G_{\text{solv}}$ and are not well suited for Φ value analysis. Finally, an important assumption of the Φ value analysis is that the folding pathway is not substantially modified by the mutation *per se* \(^{10}\). Thus, conservative deletion mutations are generally to be favored over non-conservative mutations to minimize any possible re-routing induced by the change in sequence. Accordingly, the recommended strategy is to mutate buried hydrophobic side chains, possibly by introducing minor truncations, without altering the stereochemistry (i.e. Ile→Val→Ala→Gly; Leu→Ala→Gly; Thr→Ser; Phe→Ala→Gly) \(^{13}\). Because of their characteristic sequence compositions, in the case of IDPs \(^2\), it might also be required to mutate charged residues. In these cases we recommend to perform the highest possible number of site-directed mutants, rather than rely on a few values, to further substantiate any experimental observation arising from mutations in charged positions. Additionally, the extent of formation of $\alpha$-
helices may be probed by performing an Ala→Gly scanning approach on surface exposed positions

**Application of Φ value analysis to IDPs**

The Φ value analysis is an extremely powerful technique to analyze the structure of the rate-determining step of a given reaction. This methodology can, in theory be applied to any type of reactions, including bi-molecular binding. However, even when the correct mutations are designed and the analysis is carried out rigorously, a major caveat in applying the Φ value analysis to IDPs lies in the assumption that the introduced mutation does not imply a change in rate limiting step, further, for the interpretation of measured Φ values, it is important that the overall reaction is limited by the folding or the binding steps. Thus, as described in this Section, application of the Φ value analysis to IDPs demands additional caution.

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.

![Scheme 1](image-url)
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 \(^{20}\). 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 \(^{21}\). 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 behaviour 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 towards the complex, thereby promoting folding. Under such conditions, two possible scenarios may occur – i) folding might be a slow step prior to fast binding (Figure 2a) or ii) fast folding might precede slow binding (Figure 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

\[
k_{\text{obs}} = k_F
\]

Equation 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 ($k_F$, $k_U>>k_{on'}$, where $k_{on'}=k_{on}[L]$) (Figure 2b). Under such conditions, the apparent bi-molecular rate constant would be a combination of microscopic rate constants as formalized below:

$$k_{on}^{app} = k_{on} \frac{1}{1+K_{D-N}}$$  \hspace{1cm} \text{Equation 5}

where $k_{on}$ represents the bimolecular rate constant between the folded state and the ligand and $K_{D-N}$ the unfolding equilibrium constant. Importantly, because the apparent rate constant $k_{on}^{app}$ would be equal to the bimolecular rate constant slowed down by a factor equal to $(1+K_{D-N})$ and because a limit of about $10^{7-8}$ M$^{-1}$s$^{-1}$ can be set for diffusion limited reactions involving large macro-molecules, we notice that this scenario is very unlikely to any intrinsically disordered system (i.e. with $K_{D,N}>>100$) displaying an apparent association rate of more than $10^{5-6}$ M$^{-1}$s$^{-1}$. In these cases, because the speed of the reaction would essentially be limited by the association between two fully folded entities, we predict the $\Phi$ value analysis to return generally high values of $\Phi$ when using mutations probing folding of the IDP, with fractional values located at the interface between the IDP and its partner.

*Folding after binding*
The binding mechanisms of most IDPs has been reported to follow a folding after binding scenario, whereby the locking of the folded state follow the initial recognition between the IDP, in a disordered conformation, and its partner. Under such conditions, if folding ($k_F$) were slow in relation to dissociation ($k_{off}$) (Figure 2c), the apparent association rate constant would be equal to

$$k_{on\text{app}} = k_{on} \frac{1}{1 + \frac{k_{off}}{k_F}}$$

Equation 6

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 (Figure 2d), the apparent association rate constant will be equal to the microscopic $k_{on}$. We note that these two 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, 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 two 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. A globular domain of CBP, namely the KIX domain, is one mediator of such interactions. The KIX domain binds different IDP systems via two distinct, but energetically connected, binding sites, called “c-Myb” and “MLL” sites. 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, has been studied by Φ value analysis. It is of interest to discuss both the structural distribution and the magnitude of the observed Φ values measured for this system. In fact, mapping the measured Φ values on the structure between KIX and c-Myb suggests that c-Myb folds via two distinct nuclei displaying medium or high Φ-values located at the N- and C-terminal ends of the helix, with a region with lower values of Φ was located at the center of the helix. Importantly, some of the residues displaying high Φ values did not make direct contact with KIX in the complex, providing additional support to the proposal that the Φ values for the interaction between KIX and c-Myb are dominated by the intra- rather than inter-molecular contacts and confirming the reaction to be rate limited by the folding of c-Myb rather than by recognition of KIX (Figure 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 Φ > 1, which is classically interpreted as a signature of misfolding in the transition state. Of additional interest it was noted that the average Φ 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 circa 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 the last couple of years is the interaction between the nuclear co-activator binding domain (NCBD) of CBP, and the activation domain from the p160 transcriptional co-activator for thyroid hormone and retinoid receptors (ACTR) \(^{29,30}\). 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 \(^{31}\). NCBD has molten globular properties, whereas ACTR is completely disordered, and synergistically fold upon interacting with each other \(^{32}\). A $\Phi$-value binding analysis was performed \(^{15}\), with hydrophobic mutations made at positions that are involved in intermolecular interactions on both NCBD and ACTR. With a few exceptions, all $\Phi$-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 Fig. 2d. The highest $\Phi$-values were situated at the N-terminal helix of ACTR. Indeed, in a subsequent study \(^{33}\) 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 \(^{34}\) 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 \(^{35}\). 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 Φ-value analyses can be used to understand the binding reactions of IDPs. The conclusion thus far is that binding preceeds 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.

ACKNOWLEDGEMENTS

Work partly supported by grants from the Italian Ministero dell’Istruzione dell’Università e della Ricerca (PNR-CNR 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 Bergvall Foundation (to J.D.).

REFERENCES


LEGENDS TO FIGURES

Figure 1. The $\Phi$ value analysis. Reaction profiles are presented for a protein mutated in the region highlighted with a blue star. (Left) If the transition state is as unstructured as the denatured state in the site of mutation, the free energy will be unaffected by the mutation. Accordingly, $\Delta\Delta G_{TS-D} = 0$ and $\Phi = 0$. (Right) If the site of mutation is native-like in the folding transition state, $\Delta\Delta G_{TS-D}$ tends to $\Delta\Delta G_{D-N}$ and $\Phi = 1$. By introducing several conservative site-directed mutants it is possible to determine the structure of the folding transition state.

Figure 2. Schematic free energy diagrams for the folding before binding and folding after binding scenarios. The microscopic bi-molecular binding step is highlighted in grey in each diagram. The height of this barrier is dependent on the ligand (L) concentration. As described in the text, the signatures of the different $\Phi$ values distribution may be affected if the folding, rather than the binding, step is rate limiting for the overall reaction.