1

Mapping Protein Folding Landscapes by NMR Relaxation

P.E. Wright, D.J. Felitsky, K. Sugase, and H.J. Dyson

Abstract. The process of protein folding provides an excellent example of the interactions of water with biomolecules. The changes in the water–protein interactions along the protein folding pathway provide an important impetus for the formation of the final natively folded structure of the protein. NMR spectroscopy provides unique insights into the dynamic protein folding process, and during the past 20 years we have seen the development of a wide range of NMR techniques to probe the kinetic and thermodynamic aspects of protein folding. In particular, with the advent of high-field spectrometers and stable isotope labeling techniques, the structure and dynamics of a wide range of disordered and partly ordered proteins at equilibrium have been characterized by NMR. Efforts in our laboratory over a number of years have allowed the sequence-specific identification of sites of local hydrophobic collapse, as well as secondary structure formation and transient long-range interactions in several protein systems, most notably for apomyoglobin, which will be highlighted in this article.

1.1 NMR Techniques for Studying Protein Folding

Kinetic folding pathways for proteins that fold on a millisecond timescale can be probed using hydrogen exchange pulse labeling [1,2], where differential protection of amide protons at various points during folding is detected by NMR. More recently, with the advent of high-field spectrometers and ^{13}C, ^{15}N, and ^{2}H labeling techniques, the structure and dynamics of disordered and partly ordered proteins at equilibrium have been characterized by NMR. The upper reaches of the protein folding landscape can be mapped using chemical shift, nuclear Overhauser effect (NOE), spin labeling, relaxation data, and residual dipolar coupling measurements (reviewed in [3]). Efforts in our laboratory over a number of years have allowed the sequence-specific identification of sites of local hydrophobic collapse, secondary structure formation, and transient long-range interactions in several protein systems, most notably in apomyoglobin.

1.2 The Apomyoglobin Folding Landscape

Apomyoglobin, the heme-free version of the muscle protein myoglobin, contains eight helices folded into the canonical globin fold. The kinetic folding pathway, elucidated by hydrogen exchange pulse labeling [4,5], shows the rapid formation of an intermediate species containing the A, B, G, and H helices, which is followed by the slower (∼ms) folding of the remainder of the protein. The equilibrium folding landscape for apomyoglobin is typical for a single-domain protein. In the presence of high concentrations of urea, the protein is completely unfolded [6], and populates an ensemble of structures with little detectable propensity for structure formation. In the acid-unfolded state at pH 2, the protein is largely unfolded, but samples transient secondary structure and hydrophobic clusters in certain parts of the protein but not in others [7]. Equilibrium intermediates corresponding to the ABGH kinetic intermediate are formed at intermediate pHs in the absence of urea. These species, termed *molten globules*, contain relatively stable helical secondary structure, but fluid tertiary structure. Resonances of the F helix of folded apomyoglobin at pH 6 are invisible because of an exchange on an intermediate timescale between two or more structures with different chemical shifts [8].

1.3 Structure of the Kinetic Molten Globule State

All globins so far studied pass through a kinetic molten globule intermediate that contains some but not all of the helices. The particular helices that are present in the kinetic intermediate vary according to the amino acid sequence; for example, the intermediate in the folding of the monomeric plant hemoglobin apoleghemoglobin contains the E, G, and H helices instead of the A, B, G, and H helices of apomyoglobin. An extensive series of kinetic and equilibrium folding studies on mutants of apomyoglobin [9–11] have identified a non-native structure that slows down folding and allows the intermediate to be detected. This is illustrated in Fig. 1.1, which shows the proton occupancies in the molten globule intermediate of apomyoglobin mapped onto the structure of the native, fully folded protein. The most highly protected areas in the intermediate, which likely correspond to the coalesced portion of the polypeptide, do not correspond to contiguous regions in the fully folded protein. Instead, the H helix appears to be translocated in the intermediate by about one helical turn. We conclude that the transition state for folding thus involves resolution of this small area of non-native structure before the final native contacts can be made.

1.4 The Upper Reaches of the Folding Landscape

One of the strengths of NMR is that it can give per-residue structural information on ensembles of molecules that may contain different local structures. An example of this is the acid-unfolded state of apomyoglobin. Chemical shift

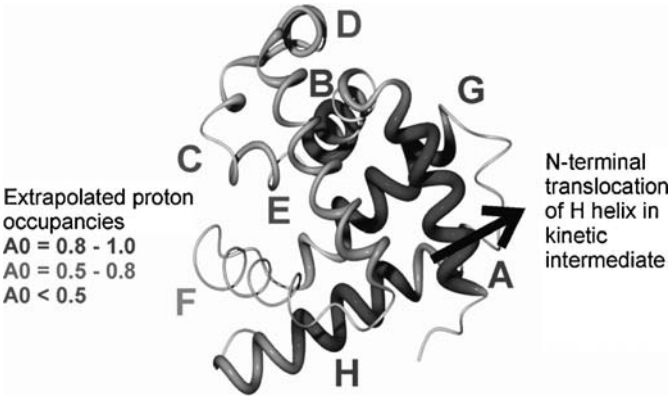

Fig. 1.1. Model of the apomyoglobin kinetic folding intermediate based on hydrogen exchange pulse labeling and mutagenesis data. The proton occupancies are mapped onto the structure of the holomyoglobin [12]. The degree of amide proton exchange protection is indicated by the intensity of the gray shading and the thickness of the backbone. The most protected regions are indicated by the darkest shade and the thickest backbone. The figure was prepared using MolMol [13]

data show that there is a detectable propensity for helical backbone dihedral angles in the regions of the protein that correspond to the H and A helices in the native folded state. Relaxation data [7] and spin-labeling studies [14] show the presence of transient native-like long-range interactions between the A and G helix regions in acid-unfolded apoMb. That these transient interactions are native-like and nonrandom must be a consequence of the amino acid sequence alone, and a series of mutant studies of apomyoglobin [9–11] showed that the propensity for local and transient long-range ordering in acid-unfolded apomyoglobin could be correlated with the property "average area buried upon folding" (AABUF) [15] or the modified hydrophobic effect [16]. In addition, the proton occupancy in the kinetic intermediate also correlates with the AABUF, and changing the local AABUF by designed point mutations also changes the pattern of proton occupancy in the kinetic intermediate [10] (Fig. 1.2). These experiments showed conclusively that the local regions with high AABUF adopt stable structure early in the protein folding process. We next turned to the question of the means whereby the hydrophobic clusters, sometimes separated by long intermediate stretches of the unfolded polypeptide, can interact, and to the hierarchy of folding events. These questions are addressed by using paramagnetic relaxation enhancement (PRE) (spin labels) and ^{15}N R_2 relaxation dispersion.

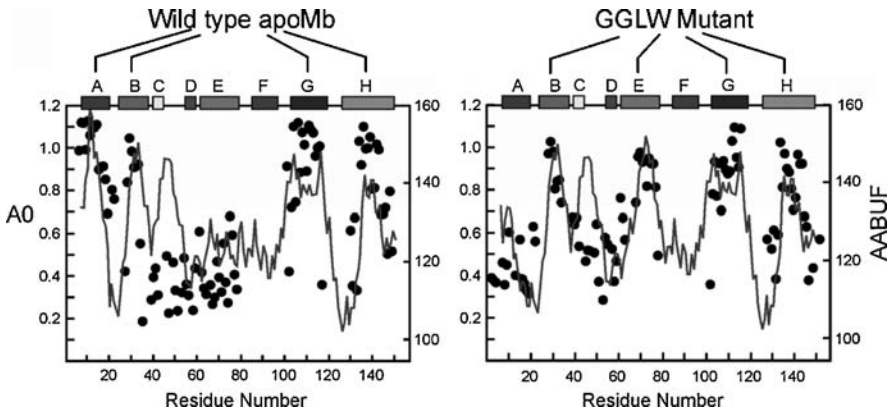

Fig. 1.2. Correlation between proton occupancies in the kinetic burst phase intermediate (*black circles*) and average area buried upon folding (AABUF, *gray lines*) for wild-type apomyoglobin and for a quadruple mutant (Leu11Gly, Trp14Gly, Ala71Leu, Gly73Trp – termed the GGLW mutant). Reproduced with permission from [10]

1.5 Paramagnetic Relaxation Probes: Spin Labeling of Apomyoglobin

The incorporation of a paramagnetic spin label results in broadening of the NMR resonances of nuclei within 15–20 Å from the site of spin labeling. This makes spin labels powerful probes of conformational ensembles. A preliminary spin-label study of apomyoglobin [14] showed that the transient contacts that occur at equilibrium in acid-unfolded apoMb are sequence specific and region specific. Resonances are broadened in the immediate vicinity of the spin label, but for some spin label sites, such as E18 (Fig. 1.3), broadening is observed at long range in the G and H helix regions, while for a spin-label site in the E helix, no such long-range broadening is observed. We have recently undertaken a comprehensive spin-label study of apomyoglobin using the data to derive a model that gives rise to a quantitative evaluation of the population of various transient collapsed states [17].

1.6 Model for Transient Interactions

For unfolded and partly folded states, the spin label reports on parts of the polypeptide chain that are in transient contact with the segment bearing the spin label. The extent of relaxation enhancement (line broadening) depends on both the distance to the paramagnetic spin label and the lifetime of the interaction. When the chain conformers rapidly interconvert, as is the case

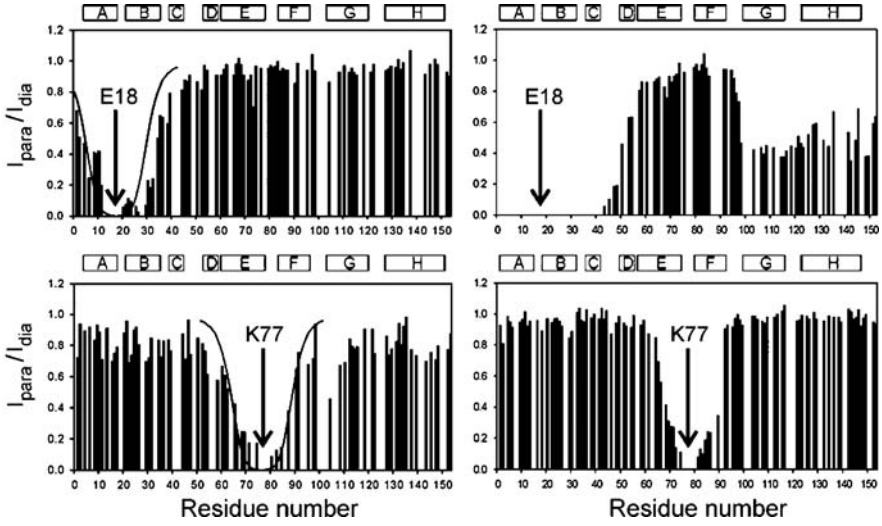

Fig. 1.3. Paramagnetic relaxation enhancement profiles for apomyoglobin unfolded at pH 2.3 in the presence (*left panels*) and absence (*right panels*) of 8 M urea. Data for spin labels attached at residues 18 and 77 is shown. The plots show the ratio of HSQC cross-peak intensity with the spin label oxidized (paramagnetic) and reduced (diamagnetic) as a function of residue number. The *solid lines* in the *left panels* represent the broadening profile expected for a random coil polypeptide. The figure is adapted from data reported in [14]. The positions of the helices in holomyoglobin are shown by the bars at the top of the figure

in unfolded apomyoglobin, the relaxation enhancement becomes a weighted average over all members of the ensemble:

$$R_{2\mathrm{P}} = \Sigma_{\mathrm{i}}\, K_{\mathrm{i}} p_{\mathrm{i}} / r_{\mathrm{i}}^{6},$$

where p_{i} is the fractional population of state i, r_i is the distance between the backbone amide proton which gives rise to the NMR cross-peak and the spin label, and K_i is a proportionality constant which depends on both the gyromagnetic ratio of the nucleus under investigation and the correlation time for the electron–nuclear dipole–dipole interaction. The magnitude of K_i is such that even very small populations ($<\sim$1%) can contribute measurably to the overall relaxation rate.

Detailed PRE measurements on a total of 14 spin-label sites distributed throughout the molecule confirm that transient long-range interactions in the acid-unfolded apomyoglobin chain are restricted to the five regions corresponding to the AABUF maxima in the primary sequence; these are designated as regions A, B, C, G, and H (corresponding to the high-AABUF sequences in the A and B helices, the CD loop, and the G and H helices, respectively). The localization of interaction sites to distinct segments of the chain suggests that the paramagnetic relaxation may be modeled via a chemical

kinetics approach whereby these five sites transiently associate in various combinations. The unfolded ensemble can thereby be divided into 52 macrostates, which correspond to the 51 different possible combinatorial arrangements of the five interaction sites and the completely dissociated substrate. For example, A can combine with G and H in one cluster, while B and C are interacting in a second cluster to form the macrostate AGH–BC.

The long-range intramolecular contacts in a given macrostate act as a set of topological restraints which reduce the chain's configurational entropy. For nonspecific, transient interactions, the relative entropy loss for the formation of different contacts (loops) may well be expected to dominate the thermodynamics and thus determine the relative populations of the different macrostates. This entropy loss depends directly on the length of the intervening chain segment(s) and relates to the distance distribution function $P(r)$ between two *noninteracting* sites separated by the same linker length. More specifically, the entropy loss is defined by the fraction of the distribution in which the interaction site centers are close enough for the two regions to coalesce; this distance is typically estimated from the sum of the radii of two spheres with volumes equivalent to the total van der Waals volumes of all residues within each interaction site. The required distance distributions can be extracted from the paramagnetic relaxation enhancement of spin labels attached centrally in the chain (such as at positions 57 or 77) where no long-range interactions occur, utilizing radius of gyration information to help determine the long-range tails of the distributions.

A model in which the relative stabilities of various clusters (and macrostates) are determined solely by the entropic barriers to loop closure discussed above can be fitted to the experimental paramagnetic relaxation data, with a single parameter reflecting the mean favorable free energy of interaction required to overcome the entropic penalty for contact formation. The model fits the experimental data surprisingly well (pale gray lines in Fig. 1.4), but not perfectly, as it cannot explain the experimentally observed preference for the C-terminus (G/H regions) to interact with A/B over C. (This latter region should interact more strongly solely on the basis of loop entropy considerations.) When this fact is accounted for (by modeling in weaker pairwise interaction free energies of the C region with other interaction sites), the loop entropy model gives an excellent fit to the PRE data for the 12 spin-label sites that show long-range interactions, as shown via the fit in Fig. 1.4 (dark gray lines). Spin-label data obtained in the presence of 8 M urea at pH 2.3 could also be well fitted by this model; the only contacts persisting under these more destabilizing conditions are relatively short range (AB and GH).

The most highly populated macrostates in the pH 2.3 ensemble in the absence of urea include, as expected, the species with no association (A–B–C–G–H). What is perhaps more surprising is that this completely unfolded substate is a minority (30%); most ensemble members have one or more long-range interaction. The most populated (17–40% population) interactions involve

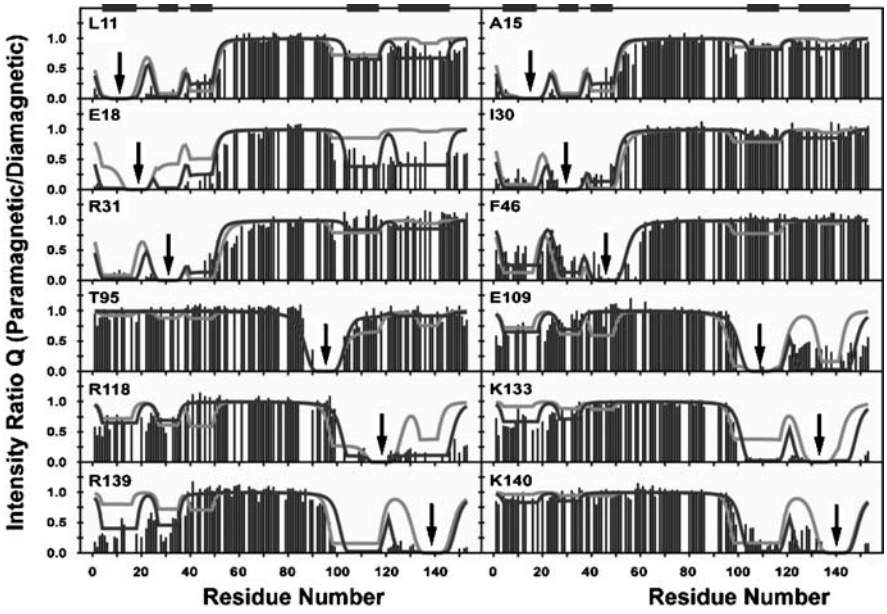

Fig. 1.4. Paramagnetic relaxation enhancement profiles for apomyoglobin unfolded at pH 2.3 with spin labels at the positions indicated by the *arrows*. The *fitted curves* show the initial (*pale gray*) fits and the fits obtained after correction for differences in pairwise cluster interaction free energies (*dark gray lines*).The location of hydrophobic clusters defined from regions of high AABUF are indicated by *bars* at the *top* of the figure. Reproduced with permission from [17]

some of the smallest loop closure events (e.g., AB, GH, and BC) and are independent of whether the interaction is native (GH, BC) or non-native (AB). The former of these observations rationalizes the relatively modest reduction ($\sim$15%) in radius of gyration relative to more completely (denaturant) unfolded ensembles despite the presence of long-range interactions in a majority of the ensembles.

Transient interactions also occur between the N- and C-termini of the protein. The populations of these contacts are quite small (less than 4%) consistent with the greater reduction in entropy required to close loops involving the extended intervening linkers. The model predicts multiple species involving different combinations of A, B, G, and H, all of similar stability. Because of its stochastic basis, however, the model is likely to underestimate cooperativity; the results are not incompatible with a single ABGH cluster. Strikingly, spin labels that probe this interaction induce quite different extents of nonlocal line broadening. A spin label attached at position R139 (in the H region) induces much more N-terminal (A/B) line broadening than one attached at position K140. Similarly, spin labels at positions 11, 15, and 18 in the A region all induce different extents of line broadening in G and H. This

differential line broadening must reflect the relative orientations/positions of the side chains within the cluster, indicating a significant degree of specificity of interaction. In contrast to this observation, spin labels that probe interactions restricted to a single chain terminus all enhance relaxation to similar degrees, suggesting that more localized interactions are significantly more heterogeneous and less specific.

1.7 Information from Relaxation Dispersion Measurements

Relaxation dispersion measurements are applied to systems that are undergoing an exchange process on the microsecond to millisecond timescale. Measurement of the R_2 relaxation rate in a series of experiments where the pulsing frequency is varied results in additional intensity for resonances of nuclei involved in the exchange process [18,19]. The resulting dispersion curve, showing R_2^{eff} as a function of $1/\tau_{\text{CP}}$ can be fitted to functions such as [19, 20]

$$
\begin{aligned}
R_2^{\text{eff}} &= R_2^0 + \frac{1}{2}\left\{k_{\text{AB}} + k_{\text{BA}} - \frac{1}{\tau_{\text{CP}}}\cosh^{-1}\left[D_+\cosh\left(\eta_+\right) - D_-\cos\left(\eta_-\right)\right]\right\} \\
D_\pm &= \frac{1}{2}\left[\pm 1 + \frac{\psi + 2\Delta\omega^2}{\sqrt{\psi^2+\xi^2}}\right] \\
\eta_\pm &= \tau_{\text{CP}}\sqrt{\frac{1}{2}\left(\pm\psi + \sqrt{\psi^2+\xi^2}\right)} \\
\psi &= \left(k_{\text{AB}} + k_{\text{BA}}\right)^2 - \Delta\omega^2 \\
\xi &= 2\Delta\omega\left(k_{\text{AB}} - k_{\text{BA}}\right)
\end{aligned}
$$

The parameters derived from these fits give information on the relative population of the two states p_{A} and p_{B}, on the rate of exchange $k_{\text{ex}}(= k_{\text{AB}} + k_{\text{BA}})$ between the two states, and on the structure of the excited state, which is given by the chemical shift difference $\Delta\omega$.

1.8 Folding of an Intrinsically Disordered Protein Upon Binding to a Target

Coupled folding and binding is a frequent theme in the field of intrinsically disordered proteins (see Chap. 6). One of the earliest examples of this phenomenon was the interaction of the phosphorylated kinase-inducible domain (pKID) of the transcription factor CREB with the KIX domain of the transcriptional coactivator CBP. Free pKID is unfolded in solution [21], but folds into an orthogonal pair of helices, αA and αB, upon binding to the folded KIX domain (Fig. 1.5) [23]. We have recently posed the question, what is the

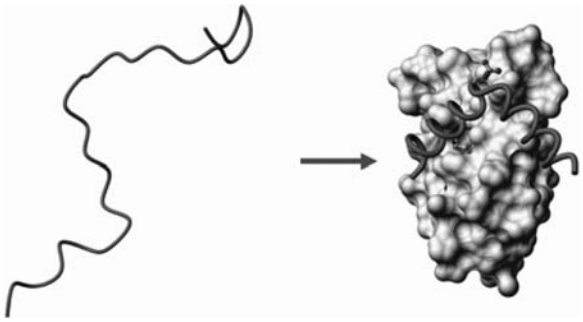

Fig. 1.5. Coupled folding and binding during the interaction of the phosphorylated kinase inducible activation domain of the transcription factor CREB (termed pKID) with the globular KIX domain of the CREB binding protein (CBP). The free pKID domain is intrinsically disordered (represented as the unfolded chain on the *left*), and in the absence of the binding partner it populates an ensemble of conformations. Upon binding to the globular KIX domain (*shown as gray surface*), it folds into a pair of orthogonal helices (*dark gray backbone trace*). Reproduced with permission from [22]

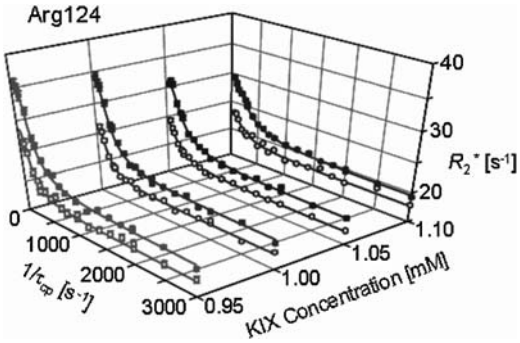

Fig. 1.6. ^{15}N R_2 relaxation dispersion profile for Arg124 of pKID recorded at 800 (*filled circles*) and 500 MHz (*open circles*). Dispersion curves for 1 mM [^{15}N]-pKID in the presence of 0.95, 1.00, 1.05, and 1.10 mM KIX are shown

mechanism by which the folding of the disordered pKID is coupled with binding to the KIX domain? Recent NMR studies including relaxation dispersion have provided intriguing insights into this question.

Coupled folding and binding of pKID to KIX was studied by recording a series of HSQC titrations and by ^{15}N R_2 dispersion measurements performed using ^{15}N-labeled pKID at two magnetic fields and over a range of pKID:KIX concentration ratios [24] (Fig. 1.6). The HSQC titrations show that at least two processes occur as KIX is titrated into pKID. Both fast (low-affinity) and slow (high-affinity) exchange processes are observed. The NMR data can

be fitted to a pseudo four-site exchange model, which gives important new insights into the mechanism of coupled folding and binding in this system:

$$\underset{\text{Free}}{\text{pKID}} + \text{KIX} \leftrightarrow \underset{\substack{\text{Encounter}\\ \text{complex}}}{\text{pKID}\ldots\text{KIX}} \leftrightarrow \underset{\text{Intermediate}}{\text{pKID} : \text{KIX}^*} \leftrightarrow \underset{\text{Bound}}{\text{pKID} : \text{KIX}}\,.$$

The encounter complex represents an ensemble in which nonspecific hydrophobic interactions occur at a number of sites. The primary interactions in the encounter complex involve a hydrophobic cluster (Y134, I137, L138, and L141) in the unfolded αB region of pKID contacting hydrophobic patches on KIX. The encounter complex was invoked to reconcile the behavior of the cross-peaks in the HSQC titrations with the $\Delta\omega$ values obtained from the relaxation dispersion measurements: a better correlation is observed between the $\Delta\omega$ values and equilibrium chemical shift differences $\Delta\delta$ which utilize the encounter complex (Fig. 1.7). The structure of the binding intermediate can also be inferred from the chemical shift and relaxation data. The αA helix is nearly fully folded in the intermediate, whereas the αB helix is only partially folded.

In summary, our NMR measurements show that the coupled folding and binding landscape of pKID is complex: Disordered pKID first makes transient hydrophobic contacts with KIX, forming an ensemble of encounter complexes that evolve to folded states without dissociation of pKID from the KIX surface. As for the folding of apomyoglobin, the most important interaction in the initiation of coupled folding and binding of pKID is the formation of hydrophobic interactions, which can then play a key role in directing the folding process towards the final folded state.

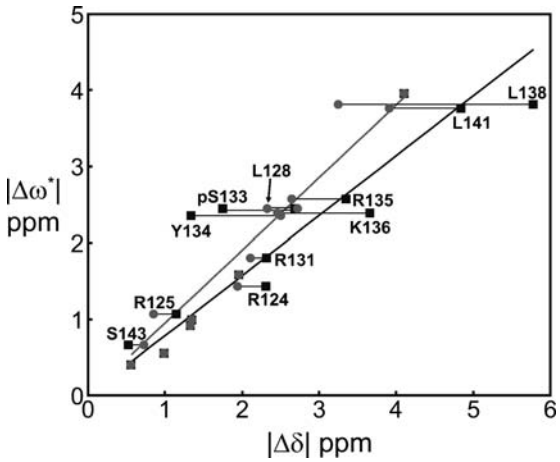

Fig. 1.7. Correlation of ^{15}N chemical shift differences ($\Delta\omega^*$) determined from the R_2 dispersion measurements with equilibrium shift differences. Chemical shift differences between free pKID and the fully bound state ($\Delta\delta_{\text{FB}}$) are shown as *black squares*, and between the encounter complex and fully bound state ($\Delta\delta_{\text{EB}}$) are shown as *gray circles*, with matching shades for the lines of best fit. Reproduced with permission from [24]

References

1. J.B. Udgaonkar, R.L. Baldwin, Nature **335**, 694–699 (1988)
2. H. Roder, G.A. Elöve, S.W. Englander, Nature **335**, 700–704 (1988)
3. H.J. Dyson, P.E. Wright, Chem. Rev. **104**, 3607–3622 (2004)
4. P.A. Jennings, P.E. Wright, Science **262**, 892–896 (1993)
5. C. Nishimura, H.J. Dyson, P.E. Wright, J. Mol. Biol. **322**, 483–489 (2002)
6. S. Schwarzinger, P.E. Wright, H.J. Dyson, Biochemistry **41**, 12681–12686 (2002)
7. J. Yao, J. Chung, D. Eliezer, P.E. Wright, H.J. Dyson, Biochemistry **40**, 3561–3571 (2001)
8. D. Eliezer, P.E. Wright, J. Mol. Biol. **263**, 531–538 (1996)
9. C. Nishimura, P.E. Wright, H.J. Dyson, J. Mol. Biol. **334**, 293–307 (2003)
10. C. Nishimura, M.A. Lietzow, H.J Dyson, P.E. Wright, J. Mol. Biol. **351**, 383–392 (2005)
11. C. Nishimura, H.J. Dyson, P.E. Wright, J. Mol. Biol. **355**, 139–156 (2006)
12. J. Kuriyan, S. Wilz, M. Karplus, G.A. Petsko, J. Mol. Biol. **192**, 133–154 (1986)
13. R. Koradi, M. Billeter, K. Wüthrich, J. Mol. Graphics **14**, 51–55 (1996)
14. M.A. Lietzow, M. Jamin, H.J. Dyson, P.E. Wright, J. Mol. Biol. **322**, 655–662 (2002)
15. G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee, M.H. Zehfus, Science **229**, 834–838 (1985)
16. H.J. Dyson, P.E. Wright, H.A. Scheraga, Proc. Natl. Acad. Sci. U. S. A. **103**, 13057–13061 (2006)
17. D.J. Felitsky, M.A. Lietzow, H.J. Dyson, P.E. Wright, Proc. Natl. Acad. Sci. U. S. A. **105**, 6278–6283 (2008)
18. J.P. Loria, M. Rance, A.G. Palmer, J. Am. Chem. Soc. **121**, 2331–2332 (1999)
19. M. Tollinger, N.R. Skrynnikov, F.A. Mulder, J.D. Forman-Kay, L.E. Kay, J. Am. Chem. Soc. **123**, 11341–11352 (2001)
20. D.G. Davis, M.E. Perlman, R.E. London, J. Magn Reson. B **104**, 266–275 (1994)
21. I. Radhakrishnan, G.C. Pérez-Alvarado, H.J. Dyson, P.E. Wright, FEBS Lett. **430**, 317–322 (1998)
22. H.J. Dyson, P.E. Wright, Nat. Rev. Mol. Cell Biol. **6**, 197–208 (2005)
23. I. Radhakrishnan, G.C. Pérez-Alvarado, D. Parker, H.J. Dyson, M.R. Montminy, P.E. Wright, Cell **91**, 741–752 (1997)
24. K. Sugase, H.J. Dyson, P.E. Wright, Nature **447**, 1021–1025 (2007)

http://www.springer.com/978-3-540-88786-7

Water and Biomolecules
Physical Chemistry of Life Phenomena
Kuwajima, K.; Goto, Y.; Hirata, F.; Terazima, M.; Kataoka, M. (Eds.)
2009, XVII, 307 p., Hardcover
ISBN: 978-3-540-88786-7