

Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins

Jian Yao, H. Jane Dyson*, Peter E. Wright*

Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Received 12 August 1997; revised version received 10 October 1997

Abstract The intrinsic chemical shift dispersion for ^{15}N , ^1HN , $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, $^{13}\text{C}^\beta$ and ^{13}CO resonances has been evaluated utilizing complete resonance assignment data for unfolded apomyoglobin, together with two other unfolded and five folded proteins, obtained from the literature. The dispersion of $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, and $^{13}\text{C}^\beta$ resonances for the unfolded proteins is poor, whereas the dispersion of ^{15}N , ^1HN and ^{13}CO is much greater, reflecting the sensitivity of these nuclei to the nature of the neighboring amino acid in the primary sequence. By contrast, the dispersion of the $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, and $^{13}\text{C}^\beta$ nuclei are much greater in the folded proteins, reflecting the well-known dependence of the environments of these nuclei on secondary and tertiary structure. These differences in chemical shift dispersion dictate differences in strategies for resonance assignment in unfolded proteins compared with those most commonly used for folded proteins. Strategies utilizing the superior chemical shift dispersion of the ^{15}N , ^1HN and, in particular, the ^{13}CO nuclei, are indicated for use with unfolded or partially folded proteins.

© 1997 Federation of European Biochemical Societies.

Key words: NMR; Residual structure; Conformational preference

1. Introduction

Chemical shift dispersion describes the general spread of the chemical shifts for a particular nucleus in the NMR spectrum. The most noticeable difference in NMR spectra obtained for the same protein molecule in folded and unfolded states lies in the chemical shift dispersion observed [1–3]. In folded states, ring current and other dipolar interactions lead to increased chemical shift dispersion; these interactions are averaged out by conformational fluctuations in unfolded or partly folded states and chemical shifts are close to random coil values, i.e. there is little chemical shift dispersion for any given nucleus. Favorable chemical shift dispersion allows peaks to be separately identified in the NMR spectrum, thus facilitating resonance assignment. In larger biological molecules, the proton chemical shift dispersion becomes unfavorable for resonance assignment, due to the large number of protons in very similar chemical environments: overcoming the overlap problem in proton spectra of proteins was one of the main driving forces in the development and implementation of modern heteronuclear multi-dimensional techniques. Despite the importance of chemical shift dispersion in the assignment process, there has been little quantitative evaluation of the dispersion characteristic of certain chemically similar nuclei; the chemical shift

dispersion is discussed in the present paper as a useful guide to assignment strategies for difficult systems.

One of the major barriers to the structural characterization of unfolded or partly folded proteins has been the low chemical shift dispersion of key resonances in the unfolded state. For example, in a typical folded protein, the backbone amide protons are usually found over a range of at least 3 ppm, from 6.5 ppm to 9.5 ppm. This relatively well-dispersed region of the spectrum is routinely used as a basis for complete sequential resonance assignment by homonuclear or heteronuclear multi-dimensional methods. For denatured proteins, the assignment of chemical shifts suffers from very poor dispersion of chemical shifts and resonance overlap, problems that are not present to the same extent in folded proteins and which become increasingly troublesome as the size of the protein increases. Dispersing the spectrum into three or more dimensions only partially solves the problem. In order to make sequential assignments by 3D methods, it is still necessary for there to be sufficient dispersion in the chemical shifts of the nuclei detected in the three dimensions. Traditional sequential assignment strategies rely heavily on the dispersion of the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ or $^1\text{H}^\alpha$ and $^1\text{H}^\beta$ resonances, and these work well when there is good chemical shift dispersion for these nuclei, as occurs in a folded protein. However, such strategies are difficult to apply when the chemical shift dispersion is low, as in denatured states. This is illustrated in Fig. 1. When the $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^1\text{H}^\alpha$ and $^1\text{H}^\beta$ are badly dispersed, it becomes necessary to find alternative nuclei with better overall chemical shift dispersion to be used for sequential assignment purposes. The dispersion of the ^{15}N and ^1HN resonances has been successfully used for this purpose (see for example, [1,4–6]) but proved inadequate, in our hands, for unfolded apomyoglobin. A different strategy was therefore required, which utilized the high intrinsic dispersion of the ^{13}CO resonance.

In devising an assignment strategy for the acid-unfolded state of apomyoglobin, we have made a quantitative comparison of the chemical shift dispersion of various nuclei in different classes of chemical environment. Although unfolded apomyoglobin has some regions with residual propensities for helical conformations [7,8], this system provides a good paradigm for examining ranges of chemical shift dispersion, since it is of a reasonable size (153 residues) and contains a representative number of many amino acid residue types. These are present in a very similar solution environment in the unfolded protein, unlike a folded protein, where many nuclei are sequestered from solvent and are part of elements of secondary structure. Finally, a comprehensive calculation of chemical shift dispersion has been made, including data from two other unfolded proteins, barnase [5] and staphylococcal nuclease [6,9], and from five folded proteins, staphylococcal nuclease [10], human carbonic anhydrase [11], mutant dihydrofolate

*Corresponding authors. Fax: +1 (619) 784 9822.
E-mail: dyson@scripps.edu; wright@scripps.edu

reductase [12], recoverin [13] and ribonuclease A [14] for which full resonance assignments (including ^{13}C) were available in the BioMagResBank. The results underscore the superior dispersion characteristics of the ^{15}N , ^1H N and ^{13}C O resonances in unfolded states. Given the excellent dispersion of ^{13}C O [9], it is surprising that this nucleus has not been more widely used in NMR studies of unfolded proteins.

2. Materials and methods

Recombinant sperm whale apomyoglobin was expressed and labeled uniformly with ^{15}N and ^{13}C using previously described methods [15]. Unfolded apomyoglobin was prepared at pH 2.3 in 10 mM acetate- d_4 containing 5% $^2\text{H}_2\text{O}$. The pH was adjusted to 2.3 by the addition of dilute HCl. NMR spectra were recorded on a Bruker DMX750 spectrometer at 25°C and with a protein concentration of 200 μM ; the low protein concentration was required to keep the viscosity to a minimum for NMR. HSQC and triple resonance spectra were acquired according to standard methods [16,17]; the 3D (HCA)-CO(CA)NH experiment was acquired according to the method of Löhner and Rüterjans [18]. NMR spectra were Fourier transformed using the Felix software from MSI on a SGI Indy workstation. Full details of the assignments appear elsewhere [19].

The chemical shift dispersion was calculated for each nucleus within a given type of amino acid (e.g. alanine) by averaging the chemical shifts observed for all such amino acids in the protein and estimating the standard deviation. The calculation was performed separately for each amino acid type to reflect the fact that corresponding nuclei in different amino acids often have different intrinsic chemical shifts. The standard deviation reflects the general spread of the chemical shift values, but does not reflect the frequency range for a particular nucleus. For example, the spectral width for ^{15}N is much wider than for ^1H N, although the general spread of the two nuclei in the HSQC spectrum is similar. The chemical shift dispersion should reflect this difference in frequency range. The calculated standard deviation is therefore divided by the range of the chemical shifts of the particular nucleus to give a normalized measure of the intrinsic dispersion. The averages of the chemical shift ranges used for calculation of the dispersion for the unfolded and folded proteins are: (unfolded proteins) HN: 1.04, ^{15}N : 20.1, H^α : 0.96, ^{13}C O: 5.12, $^{13}\text{C}^\alpha$: 18.7, $^{13}\text{C}^\beta$: 51.5; (folded proteins) HN: 4.13, ^{15}N : 30.9, H^α : 3.14, ^{13}C O: 9.37, $^{13}\text{C}^\alpha$: 24.2, $^{13}\text{C}^\beta$: 54.6. The chemical shift dispersion in Fig. 4 was calculated independently for each of the proteins, and the results were then averaged separately for the three unfolded proteins and for the five folded proteins.

3. Results and discussion

3.1. Evaluation of chemical shift dispersion

The intrinsic differences in chemical shift dispersion for various nuclei are noticeable in the spectra of unfolded apomyoglobin. Fig. 2 shows a comparison of the two-dimensional ^{15}N HSQC spectrum [20] of acid-unfolded apomyoglobin, which shows cross peaks between ^{15}N and ^1H , with the 2D HNCO spectrum [21], which shows cross peaks between the amide proton and the ^{13}C O of the previous residue in the polypeptide chain. The separation of the cross peaks, which is a function not only of the chemical shift dispersion, but also of the resonance linewidth and the spectrometer frequency, is quite similar in the two spectra, reflecting a similarity in the chemical shift dispersion of the ^{15}N and the ^{13}C O. The dispersion of the chemical shift can be estimated by assuming that the wider the dispersion, the greater the standard deviation from the mean value of a set of chemical shifts for equivalent nuclei. In this context, nominally equivalent nuclei may not be chemically equivalent, for example, the $^{13}\text{C}^\beta$ chemical shift for a threonine residue is generally close to 70 ppm, while that of an alanine is close to 20 ppm. The sets of nuclei for

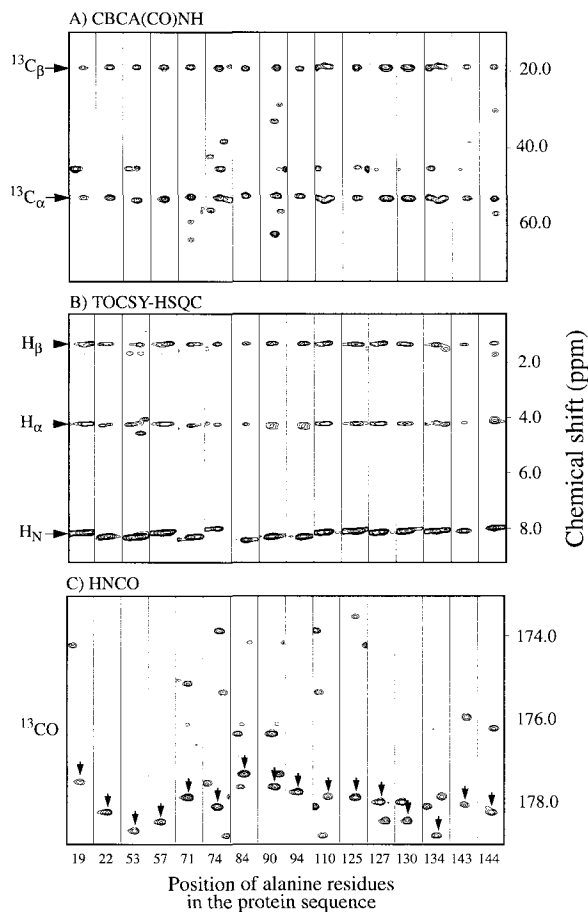


Fig. 1. Comparison of the resonances of the $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^1\text{H}^\alpha$, $^1\text{H}^\beta$, ^1H N, ^{15}N and ^{13}C O for the 15 alanine residues in apomyoglobin at pH 2.3, 10 mM acetate- d_6 and 5°C.

which the chemical shift dispersion, equated with the standard deviation, is calculated, are therefore averaged only over the particular nucleus in a given amino acid residue. The chemical shift dispersion calculated in this way is in general independent of the spectral resolution or spectrometer frequency used to obtain the chemical shift data. It can be seen from Fig. 3 that within a given residue type, the chemical shift dispersion varies greatly for different nuclei. For acid-unfolded apomyoglobin, the largest dispersion found is for the carbonyl carbon of histidine residues. On the other hand, the chemical shift dispersion for $^{13}\text{C}^\beta$ is small for all residue types.

In order to obtain a general estimate of the chemical shift dispersion for different nucleus types, the chemical shift dispersion can be averaged over the different residue types. Fig. 4 shows a comparison of the chemical shift dispersion of different nuclei, averaged over three unfolded proteins, apomyoglobin [19], barnase [5] and staphylococcal nuclease [6,9]. It is immediately obvious that the chemical shift dispersions for the backbone amide proton and backbone ^{15}N are large. This calculation confirms the utility of the accepted strategy using these two nuclei as a basis for current ^{15}N HSQC-based 3D spectra of proteins. However, the dispersions of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ are quite small: clearly an alternative to assignment strategies that utilize these nuclei must be found in the case of unfolded proteins. The data in Fig. 4 suggest such an alternative. The chemical shift dispersion for the carbonyl carbon

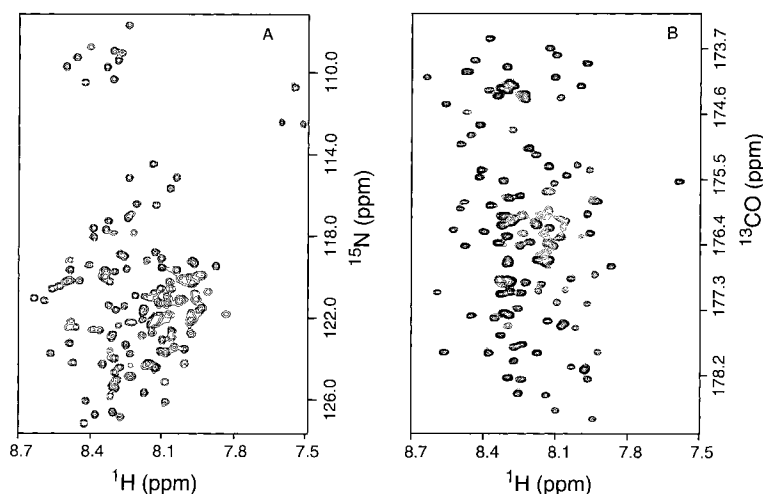


Fig. 2. A: 750 MHz ^{15}N - ^1H HSQC spectrum and B: 750 MHz HNCOC spectrum of apomyoglobin at pH 2.3, 10 mM acetate- d_6 and 5°C .

is large, comparable to those of the backbone ^{15}N and amide proton, consistent with the qualitative picture provided in Figs. 1 and 2.

Fig. 4 also shows the chemical shift dispersion for the same set of nuclei averaged over five folded proteins, staphylococcal nuclease [10], human carbonic anhydrase [11], mutant dihydrofolate reductase [12], recoverin [13] and ribonuclease A [14]. All nuclei in folded proteins have larger chemical shift dispersions than the corresponding nuclei in unfolded proteins: dispersion is not so acute a problem in folded proteins as far as assignment strategy is concerned. It is interesting to note the superior chemical shift dispersion of ^{13}CO even in the folded proteins, a circumstance which suggests that this nucleus may be of great utility in the sequential assignment of large or difficult proteins.

These findings have been of great utility in the development of strategies for the sequential resonance assignment of unfolded and partly folded states of moderately sized proteins such as apomyoglobin. To overcome the inherent degeneracy of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ or $^1\text{H}^\alpha$ and $^1\text{H}^\beta$, as well as the degeneracy in the dipeptide sequences in apomyoglobin, we utilized the large chemical shift dispersion of the backbone ^{13}CO , illustrated in Figs. 1 and 2. Our assignment strategy relies mostly on 3D spectra such as HNCOC [21] and (HCA)CO(CA)NH [18] to make sequential correlations via the well-dispersed ^{13}CO resonances. In these experiments, the chemical shift of the carbonyl carbon is encoded in ω_1 , allowing sequential assignment through the correlation with the amide nitrogen (ω_2) and amide proton (ω_3) of the following residue in the amino acid sequence. Application of this methodology, together with the more conventional C^α -based strategy, allowed us to obtain complete backbone sequential assignments for acid-unfolded apomyoglobin [19]. Conventional methods of combining HNCOC and HCACO are not effective in this case due to lack of dispersion in H^α and C^α .

The reason for the greater chemical shift dispersion of the ^{13}CO , compared with $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ is at least in part due to a greater dependence of the resonance frequency on the identity of the two adjacent amino acids side chains, i.e. both the residue to which the CO group is nominally attached (residue i) and the following residue (residue $i+1$), whereas for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ the chemical shift is dependent almost completely (in the absence of a fixed 3D structure) on the amino acid spin

system itself. Thus, the dispersion of the ^{13}CO chemical shift is dependent primarily on the local amino acid sequence, while the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ resonance dispersion is determined primarily by local secondary and tertiary structure. This dependence on the local sequence is also true for the ^{15}N and amide protons, and can account for the high dispersion in these cases also [22]. Thus, the most effective strategy for resonance assignment in unfolded proteins makes use of correlations to the relatively well-dispersed ^{15}N , ^1HN and ^{13}CO resonances. For folded proteins, correlations to these nuclei can be supplemented by information from other backbone and side chain nuclei, the dispersion of which is enhanced by the differences in local environment within the 3D structure of the protein. While the dispersion in the chemical shift of ^{15}N has frequently been utilized to facilitate sequential assignment of

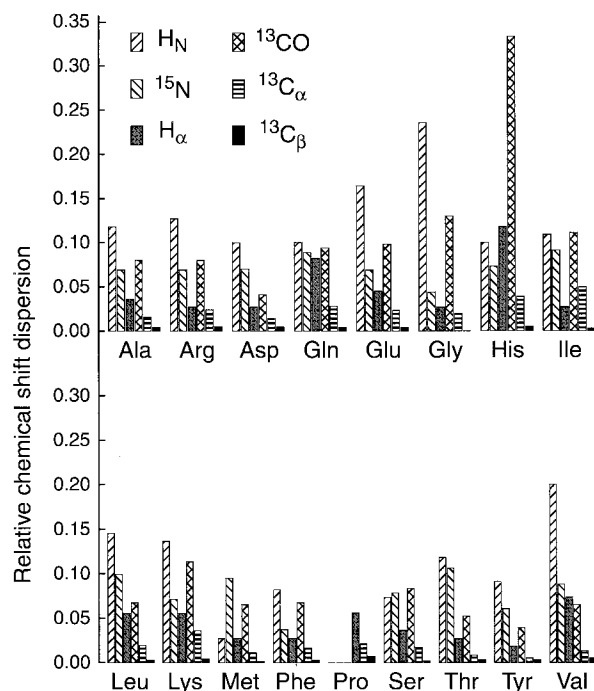


Fig. 3. Chemical shift dispersion calculated for each residue type from assigned chemical shifts of apomyoglobin unfolded at pH 2.3 [19].

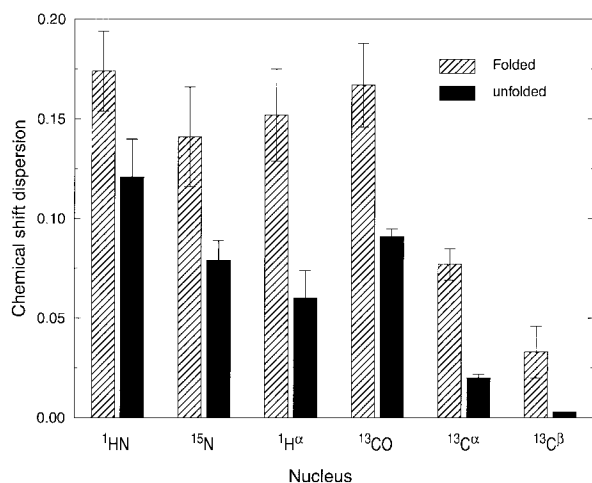


Fig. 4. Chemical shift dispersion of various nucleus types averaged over all residue types for folded and unfolded proteins. The assigned chemical shifts are averaged for unfolded proteins apomyoglobin [19], barnase [5] and staphylococcal nuclease [6,9]. The assigned chemical shifts are averaged for five folded proteins, staphylococcal nuclease [10], human carbonic anhydrase [11], mutant dihydrofolate reductase [12], recoverin [13] and ribonuclease A [14]. The chemical shift dispersion is calculated as described in the text.

unfolded states of proteins, it has not been generally appreciated that the ^{13}CO chemical shift is also well-dispersed and provides a powerful pathway to sequence-specific assignments. It is only very recently that the ^{13}CO resonance has been utilized for this purpose [23].

3.2. Use of chemical shift as a measure of local conformational preference

Generally speaking, the chemical shift dispersion becomes smaller as a protein is denatured. The chemical shift in the absence of a defined conformation is approximated by the so-called 'random coil' chemical shift. Random coil values for various nuclei of the different amino acids have been obtained from peptides in which the backbone dihedral angles are unrestricted. This has been achieved by incorporating each of the 20 amino acids in peptides containing only glycine or alanine [22,24–26]. Conformational analysis can be performed based on the deviation of these chemical shifts from random coil values. Influences from factors other than local conformation result in finite chemical shift dispersion in the random coil state. For example, the chemical shifts of backbone amide protons, ^{15}N and ^{13}CO are significantly influenced by local sequence, while the $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^1\text{H}\alpha$ and $^1\text{H}\beta$ show a much smaller effect. For a nucleus to serve as a sensitive probe for local conformation, as measured by deviations from random coil chemical shifts, two conditions need to be met: it is necessary for the nucleus to have a wide chemical shift dispersion when the residue is involved in secondary (or tertiary) structure. At the same time, the nucleus should ideally have a small chemical shift dispersion in the random coil state, otherwise corrections for the sequentially adjacent amino acid must be made [22]. The calculated chemical shift dispersions for unfolded proteins shown in Fig. 4 largely reflect the random coil state; these values can be used to provide a quantitative reference for conformational analysis by chemical shift deviation from random coil values. This type of analysis is particularly

powerful when a protein is partially unfolded or for regions of a protein that are less structured, where it can be very difficult to obtain medium- or long-range NOEs that would normally be diagnostic for the presence of secondary structure.

The ^{13}C chemical shifts of CO , $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ are useful indicators of secondary structure in folded proteins [27–30]. Other nuclei such as $^1\text{H}\alpha$ are also used, although proton chemical shifts are influenced to a greater degree by local dipolar effects, making them somewhat less useful for the detection of secondary structure. According to Fig. 4, the nucleus with the greatest increase in chemical shift dispersion as the proteins are folded is $^{13}\text{C}\alpha$, indicating that this nucleus ought to have the greatest sensitivity for the prediction of secondary structure. This is entirely consistent with the prevailing opinion in the literature, derived from empirical observations of differences in chemical shift between different elements of secondary structure [29]. We therefore suggest that the $^{13}\text{C}\alpha$ chemical shift in unfolded proteins provides the most useful and sensitive measure of the presence of residual secondary structure; although the ^{13}CO resonance is also very sensitive to secondary structure, its dependence on the local amino acid sequence makes it less reliable for diagnosis of backbone conformational propensities in unfolded states.

The chemical shift is a parameter that is potentially a rich source of structural information, and is the first sequence-specific datum obtained by NMR. In addition to its use in the structure determination of well-folded biomolecules, the present demonstration of the quantitation of the chemical shift dispersion shows that it is also a powerful tool in studies of unfolded or partly folded proteins. Such studies are becoming of increasing importance in the understanding of protein folding, ligand-protein interactions, and structural characteristics of the increasing number of proteins that are unfolded or only partly folded in their physiologically active states [31].

Acknowledgements: We thank John Chung, David Eliezer, Patricia Jennings, Stefan Prytulla and Yawen Bai for helpful discussions, and Linda Tennant for technical assistance with protein preparation. This work was supported by grant GM38794 from the National Institutes of Health.

References

- [1] Neri, D., Wider, G. and Wüthrich, K. (1992) Proc. Natl. Acad. Sci. USA 89, 4397–4401.
- [2] Zhang, O. and Forman-Kay, J.D. (1995) Biochemistry 34, 6784–6794.
- [3] Dyson, H.J. and Wright, P.E. (1996) Annu. Rev. Phys. Chem. 47, 369–395.
- [4] Logan, T.M., Thériault, Y. and Fesik, S.W. (1994) J. Mol. Biol. 236, 637–648.
- [5] Arcus, V.L., Vuilleumier, S., Freund, S.M.V., Bycroft, M. and Fersht, A.R. (1994) Proc. Natl. Acad. Sci. USA 91, 9412–9416.
- [6] Alexandrescu, A.T., Abeygunawardana, C. and Shortle, D. (1994) Biochemistry 33, 1063–1072.
- [7] Griko, Y.V., Privalov, P.L., Venyaminov, S.Y. and Kutysenko, V.P. (1988) J. Mol. Biol. 202, 127–138.
- [8] Eliezer, D., Yao, J., Dyson, H.J. and Wright, P.E. (1997) Nature, to appear.
- [9] Zhang, O., Kay, L.E., Shortle, D. and Forman-Kay, J.D. (1997) J. Mol. Biol. 272, 9–12.
- [10] Torchia, D.A., Sparks, S.W. and Bax, A. (1989) Biochemistry 28, 5509–5524.
- [11] Sethson, I., Edlund, U., Holak, T.A., Ross, A. and Johnson, B.-H. (1996) J. Biomol. NMR 8, 417–428.

- [12] Altmann, S., Labhardt, A.M., Senn, H. and Wüthrich, K. (1997) *J. Biomol. NMR* 9, 445–446.
- [13] Ames, J.B., Tanaka, T., Stryer, L. and Ikura, M. (1994) *Biochemistry* 33, 10743–10753.
- [14] Shimotakahara, S., Rios, C.B., Laity, J.H., Zimmerman, D.E., Scheraga, H.A. and Montelione, G.T. (1997) *Biochemistry* 36, 6915–6929.
- [15] Jennings, P.A., Stone, M.J. and Wright, P.E. (1995) *J. Biomol. NMR* 6, 271–276.
- [16] Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.* 26, 131–138.
- [17] Clore, G.M. and Gronenborn, A.M. (1994) *Methods Enzymol.* 239, 349–363.
- [18] Löhr, F. and Rüterjans, H. (1995) *J. Biomol. NMR* 6, 189–197.
- [19] Yao, J., Wright, P.E. and Dyson, H.J. (1997) *J. Mol. Biol.*, to appear.
- [20] Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990) *J. Magn. Reson.* 86, 304–318.
- [21] Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.* 89, 496–514.
- [22] Braun, D., Wider, G. and Wüthrich, K. (1994) *J. Am. Chem. Soc.* 116, 8466–8469.
- [23] Zhang, O., Forman-Kay, J.D., Shortle, D. and Kay, L.E. (1997) *J. Biomol. NMR* 9, 181–200.
- [24] Bundi, A. and Wüthrich, K. (1979) *Biopolymers* 18, 285–297.
- [25] Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biomol. NMR* 5, 14–24.
- [26] Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR* 5, 67–81.
- [27] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.* 222, 311–333.
- [28] Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.* 113, 5490–5492.
- [29] Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.* 239, 363–392.
- [30] Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR* 4, 171–180.
- [31] Kriwacki, R.W., Hengst, L., Tennant, L.L., Reed, S.I. and Wright, P.E. (1996) *Proc. Natl. Acad. Sci. USA* 93, 11504–11509.