Journal of Biomolecular NMR, **19:** 153–165, 2001.
KLUWER/ESCOM
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Chemical shifts in denatured proteins: Resonance assignments for denatured ubiquitin and comparisons with other denatured proteins

Wolfgang Peti[a], Lorna J. Smith[b], Christina Redfield[b] & Harald Schwalbe[c,*]
[a]*Institut für Organische Chemie, Universität Frankfurt, Marie-Curie-Strasse 11, D-60439 Frankfurt, Germany;* [b]*Oxford Centre of Molecular Sciences, New Chemistry Laboratory, South Parks Road, Oxford OX1 3QT, U.K.;* [c]*Massachusetts Institute of Technology, Department of Chemistry, Francis Bitter Magnet Laboratory, 170 Albany Street, Bldg. NW14, Cambridge, MA 02139, U.S.A.*

Received 8 August 2000; Accepted 6 November 2000

Key words: chemical shifts, denatured proteins, α-lactalbumin, lysozyme, random coil, resonance assignment, triple resonance NMR, ubiquitin

Abstract

Chemical shift assignment is reported for the protein ubiquitin denatured in 8M urea at pH 2. The variations in ^{15}N chemical shifts of three different proteins (ubiquitin, disulfide reduced, carboxymethylated lysozyme, all-Ala-α-lactalbumin), all without disulfides and denatured in 8M urea at pH 2 are compared to ‘random coil shifts’ of small model peptides (Braun et al., 1994) and to the averaged native chemical shifts taken from the BMRB database. Both parameterizations show a remarkable agreement with the averaged measured ^{15}N chemical shifts in the three denatured proteins. Detailed analysis of these experimental ^{15}N chemical shifts provides an estimate of the influence of nearest neighbors and conformational preferences on the chemical shift and provides a direct means to identify non-random structural preferences in denatured proteins.

Introduction

A large number of different techniques can be used to derive structural information for native states of proteins. This is different for the denatured state of a protein, for which only a few methods exist to obtain information at atomic resolution. It is now clear that NMR spectroscopy is the major technique to investigate the structural and dynamical characteristics of denatured proteins. ^{1}H,^{15}N,^{13}C resonance assignments for the denatured state have been reported for a number of proteins (Neri et al., 1992a, b; Arcus et al., 1994, 1995; Logan et al., 1993, 1994; Buck et al., 1995; Frank et al., 1995; Shortle, 1996; Wong et al., 1996; Yang and Kay, 1996; Schwalbe et al., 1997; Blanco et al., 1998; Dyson and Wright, 1998; Eliezer et al., 1998; Penkett et al., 1998; Hennig et al., 1999). The availability of methods to obtain site-specific resonance assignments forms the basis for a detailed analysis of NMR parameters such as NOE effects, J-couplings, and heteronuclear relaxation rates. The analysis of the NMR parameters of denatured proteins is challenging, since they reflect an average over the ensemble of conformers populated in the random coil state of a protein and methods have to be developed to describe the conformational averaging in denatured proteins. We have developed a statistical model for the random coil state of proteins (Fiebig et al., 1996; Smith et al., 1996a, b; Schwalbe et al., 1997; Hennig et al., 1999). Our random coil model assumes that all interactions in the unfolded polypeptide chain of a protein are local and that the conformational distribution can be described from the distribution of structure in native proteins. On this basis, NMR parameters such as J-couplings and chemical shifts should be predictable from analysis of their distribution observed in the native state of proteins.

The purpose of this study is to analyze the residual chemical shift dispersion observed in ^{1}H,^{15}N corre-

*To whom correspondence should be addressed. E-mail: schwalbe@mit.edu

lation experiments for denatured states of proteins. Here, we report the ^{1}H,^{15}N,^{13}C chemical shifts of ubiquitin denatured in 8M urea at pH 2. The chemical shift assignment of ubiquitin in its denatured state may serve as a reference for NMR studies in its native state. It contains 76 amino acids (8.6 kDa) without prosthetic groups or disulfide bridges. Ubiquitin is found in all eukaryotes and plays an important role as a control component of the ATP-dependent eukaryotic system for regulation of protein turnover (Rechsteiner, 1988; Hershko and Ciechanover, 1998). The structure of ubiquitin has been studied under a variety of different conditions. The native state of ubiquitin is well characterized by high-resolution X-ray crystallography (Vijay-Kumar et al., 1987) and NMR spectroscopy. NMR investigations of native ubiquitin have involved measurement of homonuclear NOEs, homo- and heteronuclear coupling constants (Di Stefano and Wand, 1987; Weber et al., 1987; Tjandra et al., 1995; Cavanagh et al., 1996) as well as residual dipolar couplings (Cornilescu et al., 1999). Auto- and cross-correlated relaxation rates have been measured to provide an insight into the backbone dynamics of ubiquitin (Schneider et al., 1992; Tjandra et al., 1995, 1996; Fushman et al., 1998; Fushman and Cowburn, 1998; Carlomagno et al., 2000).

Ubiquitin remains folded even at temperatures of 80 °C and for pH values ranging from 1.2 to 13 (Lenkinski et al., 1977; Jenson et al., 1980; Nash and Jonas, 1997). In 60% aqueous methanol at pH 2, ubiquitin forms a so-called A-state. In this state, the antiparallel β-sheet and the central α-helix are conserved in the N-terminal half of the protein (Brutscher et al. (1997) and references cited therein). The C-terminal half, which is rich in β-sheet character in the native state, undergoes a methanol-induced transition to a dynamic state that is purely α-helical. A wide range of dynamic properties were measured to confirm this model of a non-native state of ubiquitin. In 8M urea, ubiquitin is essentially completely denatured (Lenkinski et al., 1977). Under these conditions ubiquitin provides a good model for examining the ^{15}N chemical shift dispersion in the denatured state of a polypeptide chain without disulfides.

In order to analyze nearest neighbor effects on the ^{1}H^{N},^{15}N chemical shifts in denatured proteins, we compare ^{1}H^{N},^{15}N resonance assignments in human ubiquitin with chemical shift data for disulfide reduced, carboxymethylated hen egg white lysozyme (Schwalbe et al., 1997; Hennig et al., 1999), and a mutant form of human α-lactalbumin in which all cysteines have been replaced by alanines (Redfield et al., 1999). In addition, trends observed for chemical shift deviations are correlated with a large set of ^{3}J(H^{N},H$^{\alpha}$) coupling constants measured for the three different proteins. All proteins were studied in their unfolded state obtained by denaturation in 8M urea at pH 2.

Material and methods

^{15}N labeled ubiquitin and ^{15}N,^{13}C labeled ubiquitin (both 5 mg; gifts from VLI Research Inc., Malvern, PA) were denatured in 8M urea at pH 2 in 95% H_2O/5% D_2O. 10 mg of ^{15}N,^{13}C labeled ubiquitin was dissolved at pH 4.7 in 95% H_2O/5% D_2O for the native state reference sample. Experiments for the assignment of backbone resonances are summarized in Table 1. Pulsed field gradient versions of the experiments, with water flip-back pulses (Grzesiek and Bax, 1993) and sensitivity enhancement (Sattler et al. (1999) and references cited therein) were recorded on a Bruker DRX 600 (TXI HCN z-grad) at 303 K. A recycle delay of 1.5 s was used for all experiments. The carrier positions were: 117.5 ppm for ^{15}N, 174 ppm for ^{13}C′, 38.2 ppm for aliphatic ^{13}C, and 4.8 ppm for ^{1}H, respectively. All ^{1}H pulses were centered around the water resonance. All carbon pulses were implemented as Gaussian cascades (Emsley and Bodenhausen, 1990); Q3 (Emsley and Bodenhausen, 1992) pulses for inversion and G4 and time reversed G4 pulses for excitation were used. Fourier transformation, mirror image linear prediction and data analysis were performed using the program Felix 98.0 (Molecular Simulation Inc., San Diego, CA). Resonance positions were referenced using internal DSS and TSP to obtain ^{1}H chemical shift referencing and for the ^{13}C and ^{15}N dimensions by calculation as described in Wishart et al. (1995). No difference was observed between DSS and TSP referencing. For ubiquitin, a total of 71 out of 73 possible ^{1}H^{N},^{15}N correlation peaks can be resolved at a proton resonance frequency of 600 MHz.

Coupling constant determination

The measurement of ^{3}J(H^{N},H$^{\alpha}$) coupling constants was performed as described for denatured human ubiquitin (Peti et al., 2000). ^{3}J(H^{N},H$^{\alpha}$) coupling constants for all-Ala α-lactalbumin were measured using a 0.4 mM sample of ^{15}N labeled protein dissolved in 8M urea (95% H_2O/5% D_2O) at pH 2 and 293 K (Redfield et al., 1999). A 2D HMQC-J (Kay and Bax,

Table 1. Summary of experiments carried out for the assignment of ubiquitin unfolded in 8M urea at pH 2

Experiment	No. of data points/sweep width			References
	t_3	t_2	t_1	
2D $^{1}H,^{15}N$-HSQC		4096(^{1}H); 5482.45 Hz	512(^{15}N); 1149.425 Hz	
3D NOESY-HSQC (200 ms)	1024(^{1}H); 5482.45 Hz	64(^{15}N); 1149.425 Hz	256(^{1}H); 5482.45 Hz	(Driscoll et al., 1990; Marion et al., 1989a)
3D TOCSY-HSQC (60 ms, 80 ms)	1024(^{1}H); 5482.45 Hz	64(^{15}N); 1149.425 Hz	256(^{1}H); 5482.45 Hz	(Marion et al., 1989b)
3D HSQC-NOESY-HSQC (150 ms)	1024(^{1}H); 5482.45 Hz	48(^{15}N); 1149.425 Hz	64(^{15}N); 1149.425 Hz	(Frenkiel et al., 1990)
3D HNCO	1024(^{1}H); 5482.45 Hz	64(^{13}C); 1666.67 Hz	64(^{15}N); 1149.425 Hz	(Kay et al., 1990)
3D CBCA(CO)NH	1024(^{1}H); 5482.45 Hz	64(^{15}N); 1149.425 Hz	128(^{13}C); 9615.385 Hz	(Grzesiek and Bax, 1992)
3D HNCACB	1024(^{1}H); 5482.45 Hz	128(^{13}C); 9615.385 Hz	128(^{15}N); 1149.425 Hz	(Wittekind and Mueller, 1993)
3D CC(CO)NH	1024(^{1}H); 5482.45 Hz	64(^{15}N); 1149.425 Hz	128(^{13}C); 9615.385 Hz	(Grzesiek et al., 1993)

Figure 1. Annotated $^{1}H,^{15}N$-HSQC spectrum of ubiquitin unfolded in 8M urea, pH 2, recorded at 303 K and a ^{1}H frequency of 600 MHz. Sequential assignments are indicated with the one-letter amino acid code and the residue number.

1990) dataset was collected at 600 MHz. The sweep widths were 6514.66 and 1152.07 Hz in ω_2 and ω_1, and 2048 and 300 complex points were collected in t_2 and t_1, respectively. After zero-filling, the digital resolution in ω_1 (^{15}N) was 0.56 Hz/point. The coupling constants were extracted from ω_1 columns using an in-house fitting program. Lysozyme chemical shift and coupling constant data were analyzed taking previously published data (Schwalbe et al., 1997; Hennig et al., 1999).

Results and discussion

Chemical shift assignments

The $^{1}H,^{15}N$-HSQC spectrum, with annotated assignments, of ubiquitin denatured in 8M urea at pH 2 is shown in Figure 1. The proton chemical shift dispersion in denatured ubiquitin is only 0.6 ppm (excluding the second amino acid at the N-terminus of the protein). The ^{15}N chemical shifts in ubiquitin show an amino acid type specific clustering: this is clearly visible for glycine, threonine and serine residues. The other amino acids cluster together with ^{15}N chemical shifts in the range 117–125 ppm (Table 2). For a given amino acid, the residual chemical shift dispersion of ^{15}N is large compared to the narrow chemical shift dispersion of $^{13}C^{\alpha}$,$^{13}C^{\beta}$ and $^{13}C'$.

A second characteristic feature observed in experiments on denatured proteins is the narrow line width. The long ^{1}H and ^{15}N transverse relaxation properties ($T_2(^{15}N) \approx 350$ ms for ubiquitin at 60 MHz ^{15}N frequency) are very uniform for denatured states of proteins lacking disulfide bridges. This enables experiments to be performed on denatured proteins that are only feasible in native proteins after deuterium incorporation. Increased mobility, as evidenced by a negative heteronuclear $\{^{1}H\}$-^{15}N NOE, might provide anNOE, might provide an additional hint to identify terminal amino acids in unfolded proteins.

In ubiquitin, sequential NOE cross peak patterns are uniform throughout the sequence and H^{N},H^{N} (i,i+1) and H^{α},H^{N}(i,i+1) cross peaks are observed with similar intensity. This reflects the fact that in a random coil both the conformational regions corresponding to the α and β main chain ϕ,ψ conformations are populated (Smith et al., 1996a, b). Provided that the primary sequence of the polypeptide chain

Table 2. Averaged chemical shifts for three different proteins denatured in 8M urea at pH 2. ^{13}C chemical shifts are reported for human ubiquitin and cysteine reduced, carboxymethylated hen eggwhite lysozyme. Std. is the standard deviation of the chemical shifts

	^{15}N (ppm)	Std. (ppm)	^{1}H^{N} (ppm)	Std. (ppm)	^{13}C$^{\alpha}$ (ppm)	Std. (ppm)	^{13}CO (ppm)	Std. (ppm)	Number of residues
Ala	125.11	1.27	8.14	0.15	52.26	0.25	177.10	0.37	25
Arg	122.41	1.42	8.25	0.19	55.96	0.38	175.80	0.46	15
Asn	120.43	1.05	8.31	0.14	52.79	0.19	174.71	0.40	17
Asp	120.09	1.36	8.40	0.09	52.66	0.16	174.91	0.49	23
Cys	119.49	0.64	8.26	0.19	55.05	0.02	173.82	0.22	3
Gln	122.13	1.17	8.37	0.18	55.46	0.25	175.67	0.42	16
Glu	121.63	1.79	8.33	0.13	55.26	0.12	175.74	0.56	17
Gly	109.56	0.91	8.19	0.20	44.93	0.34	173.58	0.37	23
His	119.36	1.46	8.48	0.16	54.64	0.56	174.15	0.06	4
Ile	121.66	1.31	8.09	0.17	60.47	0.26	175.68	0.33	23
Leu	123.55	1.61	8.18	0.20	54.79	0.32	177.01	0.34	29
Lys	122.49	1.57	8.30	0.16	55.87	0.49	175.97	1.35	23
Met	121.38	1.82	8.15	0.13	55.12	0.37	174.50	2.05	4
Phe	122.27	2.08	8.27	0.19	57.39	0.11	175.13	0.32	9
Pro					63.47	0.36	176.27	0.46	3
Ser	116.95	1.34	8.24	0.16	58.11	0.36	174.25	0.53	20
Thr	115.11	1.64	8.12	0.14	61.37	0.63	174.27	0.54	21
Trp	121.08	0.62	7.91	0.13	57.19	0.44	175.81	0.30	8
Tyr	120.61	0.74	8.02	0.09	57.87	0.12	175.38	0.25	8
Val	122.47	1.59	8.15	0.22	61.91	0.26	175.36	0.73	11

is not repetitive, ^{13}C labeling is not always required for the ^{1}H and ^{15}N resonance assignment (Schwalbe et al., 1997; Redfield et al., 1999). For the sequential (i±1) assignment, the NOESY-HSQC and HSQC-NOESY-HSQC spectra were found to provide the best resolution. Figure 2 shows strips taken from ^{1}H,^{15}N-NOESY-HSQC and ^{1}H,^{15}N-TOCSY-HSQC spectra of denatured ubiquitin with annotated (i,i+1) connectivities. In contrast to folded proteins, for which H^{N},H^{N}(i,i+1) cross peaks in ^{1}H,^{15}N-HSQC-NOESY-HSQC are observed only in α-helical secondary structure elements, these cross peaks are observed throughout the polypeptide chain in denatured proteins and therefore provide useful correlations for the sequential assignment in cases for which the N^{H},N^{H}(i,i+1) cross peaks are overlapped in the ^{1}H,^{15}N-NOESY-HSQC spectrum. Residues with neighboring glycine, threonine or serine residues serve as starting points for assignment: the ^{15}N chemical shift for these three amino acids differ from those of other amino acid type (see Figure 1) and differentiation between serine and threonine can readily be obtained from the ^{1}H,^{15}N-TOCSY-HSQC spectrum. While a majority of residues can be assigned in denatured proteins labeled with ^{15}N, incorporation of both ^{15}N and ^{13}C has often been reported to be vital for complete resonance assignment in denatured proteins. Standard 3D assignment pulse sequences like HNCACB, CBCA(CO)NH or HNCO resolving ^{13}C chemical shifts in the relatively well resolved ^{1}H,^{15}N plane can be applied as they are for native states of proteins. However, due to the observation that the variation in the carbon chemical shift (e.g. C$^{\alpha}$ and C$^{\beta}$) is very small, analysis of ^{1}H,^{15}N NOESY-HSQC spectra is often vital for the assignment of denatured proteins.

^{1}H^{N},^{15}N,^{13}C backbone resonances for 71 out of the 73 non-proline residues of ubiquitin in 8M urea, pH 2, could be assigned. The ^{1}H and ^{13}C resonances of two prolines were also assigned. Residues around Pro19 showed a doubling of the ^{1}H^{N},^{15}N correlation peaks. The side chain H$^{\alpha}$,H$^{\beta}$,C$^{\alpha}$,C$^{\beta}$ resonances could be assigned for 75 of the 76 residues (no assignment was possible for Pro37 because there are two neighboring prolines in the sequence). Additional side chain proton chemical shifts were identified in the ^{15}N edited TOCSY-HSQC and the NOESY-HSQC spectra. The chemical shift values of proton, nitrogen and carbon resonances have been deposited in the

Table 3. Comparison of averaged experimental chemical shifts for each amino acid compared with the averaged chemical shifts predicted using the empirical values by Braun et al. (column a) and the BMRB values (column b). Column c reports the correlation coefficient of the average ^{15}N chemical shift for each amino acid compared with the chemical shifts predicted by Braun et al. (1994), column d reports the χ^2 value between experiment and predictions and column e gives the number of residues in the three protein datasets

	^{15}N averaged measured values		(c) Correlation coefficient	(d) χ^2	(e) Number of residues
	(a) averaged Braun et al.	(b) averaged BMRB			
Ala	−0.59	2.14	0.88	0.32	25
Arg	−0.29	2.04	0.88	0.44	15
Asn	−0.07	1.63	0.79	0.41	17
Asp	−0.71	−0.22	0.83	0.58	23
Cys	−0.61	0.19	0.65	1.06	3
Gln	0.23	2.42	0.64	0.78	16
Glu	−0.17	0.95	0.93	0.65	17
Gly	−1.14	0.36	0.73	0.39	23
His	−0.74	0.01	0.96	0.16	4
Ile	−0.14	−0.09	0.88	0.41	23
Leu	−0.25	1.84	0.81	0.92	29
Lys	−0.61	1.48	0.92	0.36	23
Met	−0.52	1.34	0.94	0.4	4
Phe	0.37	1.54	0.87	1.08	9
Ser	−0.65	0.64	0.85	0.72	20
Thr	−0.79	0.01	0.89	0.56	21
Trp	−1.62	−0.18	0.61	1.02	8
Tyr	−1.39	−0.09	0.50	0.68	8
Val	1.37	1.42	0.92	0.42	11

BioMagResBank in Madison, WI, U.S.A. (accession number 4375, http://www.bmrb.wisc.edu).

An important step in the refolding of ubiquitin is thought to involve a cis-trans proline isomerization as described by Briggs and Roder (1992) using hydrogen–deuterium exchange labeling in conjunction with rapid mixing methods and two-dimensional NMR analysis. Ubiquitin has three proline residues. In our studies, a doubling of resonances with a ratio of 12.5% to 87.5% for the cis and trans conformations of Pro19, respectively, was observed for resonances of residue Ser20, indicating slow cis-trans proline isomerization in denatured ubiquitin (cis-Pro19: C^{α}: 61.396 ppm, C^{β}: 33.507 ppm; trans-Pro19: C^{α}: 61.98 ppm, C^{β}: 31.13 ppm). These findings are in good agreement with the earlier study using pulsed hydrogen exchange experiments in which cis/trans proportions of 9% to 91%, 14% to 86% and 19% to 81% were found for Pro19, Pro37 and Pro38, respectively.

Analysis of chemical shifts

Residual dispersion of chemical shifts in denatured proteins is mainly observed for backbone amide nitrogen and carbonyl atoms, while deviation of individual ^{13}C and other ^{1}H chemical shifts from their mean is small and decreases for side chain atoms (Table 2). Interestingly, the standard deviation of the carbonyl shifts is small and comparable to the other carbon chemical shift deviations. The mean chemical shifts found here and in other proteins are close to the averaged values commonly referred to as random coil chemical shifts that have been measured in a large number of small peptide constructs and for which solvent effects have been characterized (Wishart et al., 1991a, b; Thanabal et al., 1994; Plaxco et al., 1997). For example, the peptide constructs GGXGG are assumed to adopt no preferred conformation and therefore to represent the random coil in solution (Merutka et al., 1995). An extensive characterization of random coil chemical shifts was carried out by Wishart et al. (1995) using the peptides GGXAGG and GGXPGG to quantify the influence of a proline residue on the

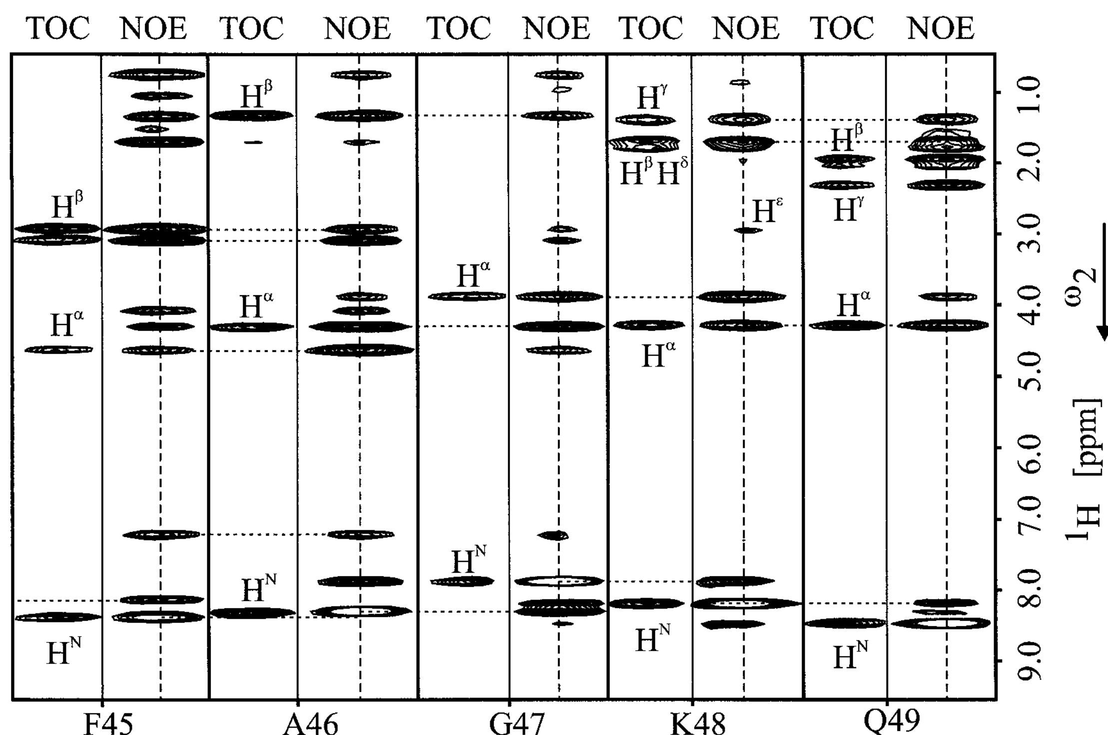

Figure 2. Strips taken from the ^{1}H,^{15}N-NOESY-HSQC and the ^{1}H,^{15}N-TOCSY-HSQC spectra of unfolded ubiquitin. The strips of residues F45 to Q49 show features typical of NOESY spectra for unfolded proteins. A uniform pattern of sequential (i,i+1) NOE cross peaks for H^{N},H^{N} (i,i+1) and H^{α},H^{N} (i,i+1) is observed with similar intensity throughout the entire sequence, in contrast to native folded sequences. The intraresidual correlation peaks are annotated.

chemical shifts of the preceding residue. The reported chemical shift values are routinely used to interpret deviations as being indicative for specific ranges of values of the backbone angles ϕ and ψ. Interestingly, C′, C^{α} and C^{β}, but not ^{15}N chemical shifts are used to predict the secondary structure of proteins (Glushka et al., 1989; Spera and Bax, 1991; Wishart et al., 1992, 1995, 1997; Braun et al., 1994; Wishart and Sykes, 1994a, b; Wishart and Nip, 1998) Recently, Cornilescu et al. (1999) proposed an elegant way to extract the angle information from these chemical shifts to restrain ϕ, ψ torsion angles in the structure calculation of native proteins using a database approach.

Here, we concentrated on the ^{15}N chemical shifts since these have a significant dispersion in the denatured state, and we analyzed data for three proteins denatured in 8M urea at pH 2: human ubiquitin, reduced carboxymethylated hen egg white lysozyme and all-Ala α-lactalbumin. We compared the residue-specific averaged chemical shift of these denatured proteins with averaged chemical shifts of native proteins as deposited in the BMRB chemical shift database. The average chemical shift statistics of all native folded diamagnetic proteins was used as its last update in February 1999. Aromatic amino acids, cysteine and methionine residues were excluded from the statistical analysis because of the lack of experimental data in our dataset of denatured proteins (less than 10 chemical shifts) and because aromatic residues have been found to be involved in non-random structure at least in lysozyme (Schwalbe et al., 1997; Hennig et al., 1999). The analysis reveals that the mean observed chemical shifts in native proteins are very close to the mean observed values values in denatured proteins (shown in Figure 3); the correlation coefficient is 0.98. In native states of proteins a large spread of chemical shifts around the mean is observed. For the unfolded proteins studied here, the distribution is narrow.

In 1994, Braun et al. investigated the ^{15}N chemical shifts in random coil peptides. The chemical shift for a given amino acid averaged over all 20 possible dipeptide pairs as given by Braun et al. correlates well ($R = 0.99$, shown in Figure 4) with the averaged ^{15}N chemical shift for a given amino acid derived from the proteins studied here (Figure 4). Braun et al. (1994) re-

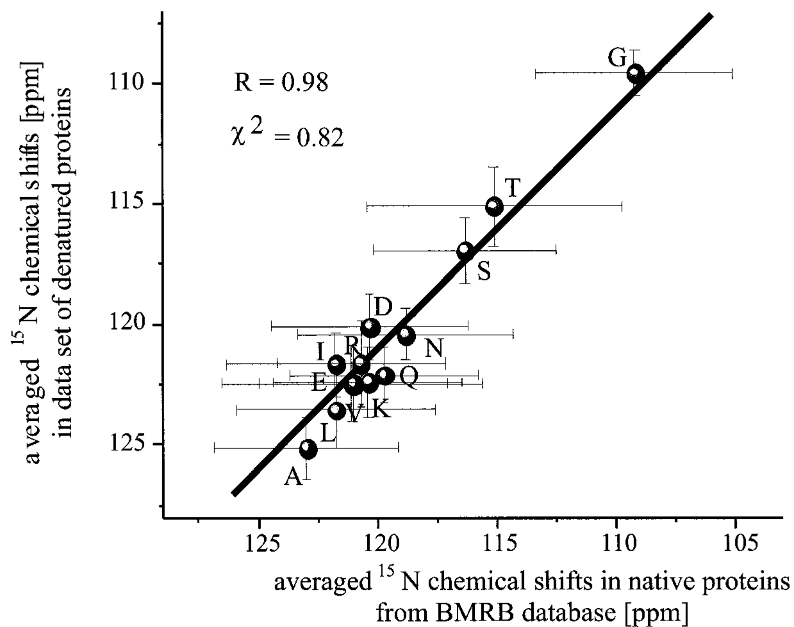

Figure 3. Comparison of the averaged ^{15}N chemical shifts of three denatured proteins with the averaged ^{15}N chemical shifts of folded proteins in the BMRB database. The average was taken for all amino acids of a given kind. The balls represent the mean values and the error bars are standard deviations. Aromatic amino acids, cysteine and methionine were excluded from this comparison as discussed in the text. Large variations in ^{15}N chemical shifts are observed for native proteins. A much smaller variation can be seen for the ^{15}N chemical shift of the three denatured proteins studied.

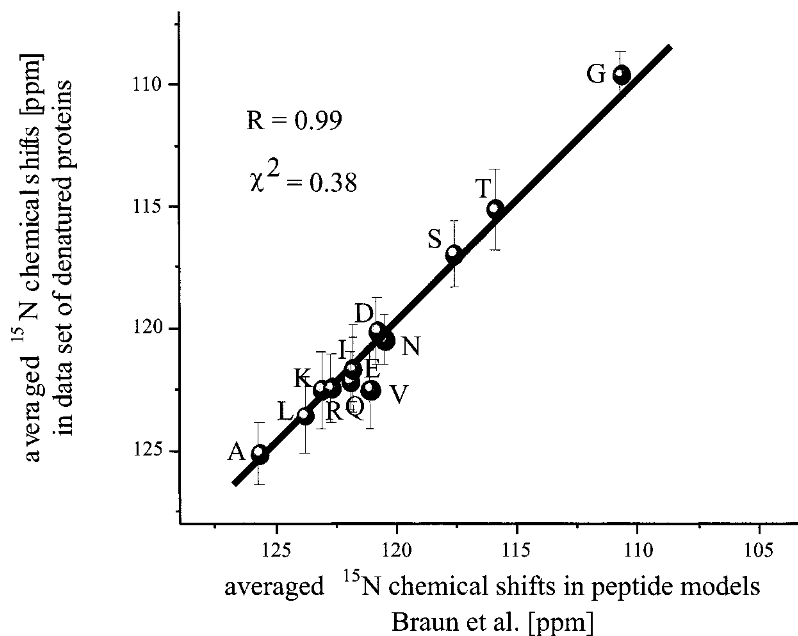

Figure 4. Comparison of the averaged ^{15}N chemical shifts of three denatured proteins with the ^{15}N chemical shift calculated by Braun et al. (1994).

ported a correction factor to account for the influence of the i−1 amino acid on the ^{15}N chemical shift of amino acid i. A comparison with the averaged experimental data provided here is shown for all residues in Figure 4 and is summarized in Table 3. The data show that the chemical shift for a specific amino acid corrected for the i−1 amino acid and averaged over the three datasets from ubiquitin, lysozyme and all-Ala α-lactalbumin correlates well with the random coil data from the peptide work.

Figure 5 shows the correlation of experimental ^{3}J(H^N,H$^\alpha$) coupling constants in the three proteins individually and averaged over the three different proteins and predictions from the random coil model. The coupling constants for a random coil were predicted (Smith et al., 1996) using an enlarged database of 402 high resolution protein crystal structures (Hennig et al., 1999). Predictions were made using only those residues that are not in regions of recognized secondary structure (COIL parameter set) using three different Karplus parameterizations (Pardi et al., 1984; Ludvigsen et al., 1991; Vuister and Bax, 1993).

The correlations reveal the following: using all amino acids, the averaged correlation coefficient R is found to be 0.85. Excluding Asp and Glu as residues that are protonated at low pH, the averaged correlation coefficient increases to 0.89. This can also be seen for the individual proteins. Aromatic residues are often involved in non-random conformations in the denatured states of proteins. This is also reflected here. Reduction of the dataset (−aromatic residues and −D, −E) increases the correlation coefficient (ubiquitin: $R = 0.96$; lysozyme: $R = 0.94$; all-Ala-α-lactalbumin: $R = 0.90$; and $R = 0.95$ for all three proteins). The increase in correlation between prediction and experiment is independent of the protein and seems therefore independent of both the primary sequence and the native structure of the protein. Lysozyme and α-lactalbumin contain a significant portion of α-helices in their native state, while ubiquitin is mainly a β-sheet protein. This result is in agreement with the database approach to describe the denatured state of proteins; no bias of the experimental coupling constants depending on the secondary structure in the native protein is observed. This comparison is also in agreement with the observation that aromatic residues are involved in non-random conformations in the denatured states of proteins. Analysis of coupling constants and chemical shifts shows that aromatic residues are found to deviate from the model predictions. Therefore, measurement of ^{15}N chemical shifts appears to be a very sensitive method to measure this deviation from random conformational sampling.

Figure 6 shows the correlation of the experimental chemical shifts of the 29 assigned leucine residues in the three datasets with predictions using the correction factors of Braun et al. (1994). It is apparent that the size of the neighboring amino acid side chain influences both the chemical shifts as well as the coupling constants (summarized in Table 4). In principle, both electronic and steric factors exerted by amino acid i−1 will influence the ^{15}N chemical shift of amino acid i and will therefore account for the variation between *different* dipeptides. For a given dipeptide, the variations are likely to reflect differences in the conformational sampling of the specific subunit. Therefore, it is interesting to investigate the relationship between the variation of chemical shifts and experimental parameters like ^{3}J(H^N,H$^\alpha$) coupling constants that depend on the angle ϕ. An extensive investigation is beyond the scope of this paper, mainly because of the limited dataset of only 230 measured coupling constants, which are summarized in Table 4. A previous analysis of ^{3}J(H^N,H$^\alpha$) coupling constant data for unfolded fibronectin binding protein (Penkett et al., 1997) showed that there is an increase in coupling constant values when residue (i−1) has a β branched or aromatic side chain. The increased coupling constants reflect a greater population of β ϕ,ψ conformers in the ensemble of rotamers. Overall, a similar trend is seen here. For example, for asparagine the average ^{3}J(H^N,H$^\alpha$) coupling constant is 7.3 Hz when residue (i−1) has a β branched or aromatic side chain compared to 6.7 Hz for other residue types at position (i−1), while for isoleucine the average ^{3}J(H^N,H$^\alpha$) coupling constant is 7.6 Hz when residue (i−1) has a β branched or aromatic side chain compared to 6.7 Hz when another residue is in the preceding position. Additional factors to these are also important, however. It is interesting to see that there is a considerable variation in both the chemical shift and the ^{3}J(H^N,H$^\alpha$) coupling constant *within* the dipeptide pair Thr/Leu (Thr7/Leu8: δ^{15}N = 124.21 ppm, ^{3}J(H^N,H$^\alpha$) = 6.9 Hz; Thr14/Leu15: δ^{15}N = 125.14 ppm, ^{3}J(H^N,H$^\alpha$) = 7.4 Hz; Thr55/Leu56: δ^{15}N = 124.45 ppm, ^{3}J(H^N,H$^\alpha$) = 7.0 Hz; Thr66/Leu67: δ^{15}N = 123.76 ppm, ^{3}J(H^N,H$^\alpha$) = 7.6 Hz). The most low-field ^{15}N chemical shift of Leu15 corresponds with the largest ^{3}J(H^N,H$^\alpha$) coupling constant. For Thr14/Leu15, Thr55/Leu56, and Thr66/Leu67, there is a correlation between downfield ^{1}H^N and ^{15}N chemical shift and increase in ^{3}J(H^N,H$^\alpha$) coupling constant,

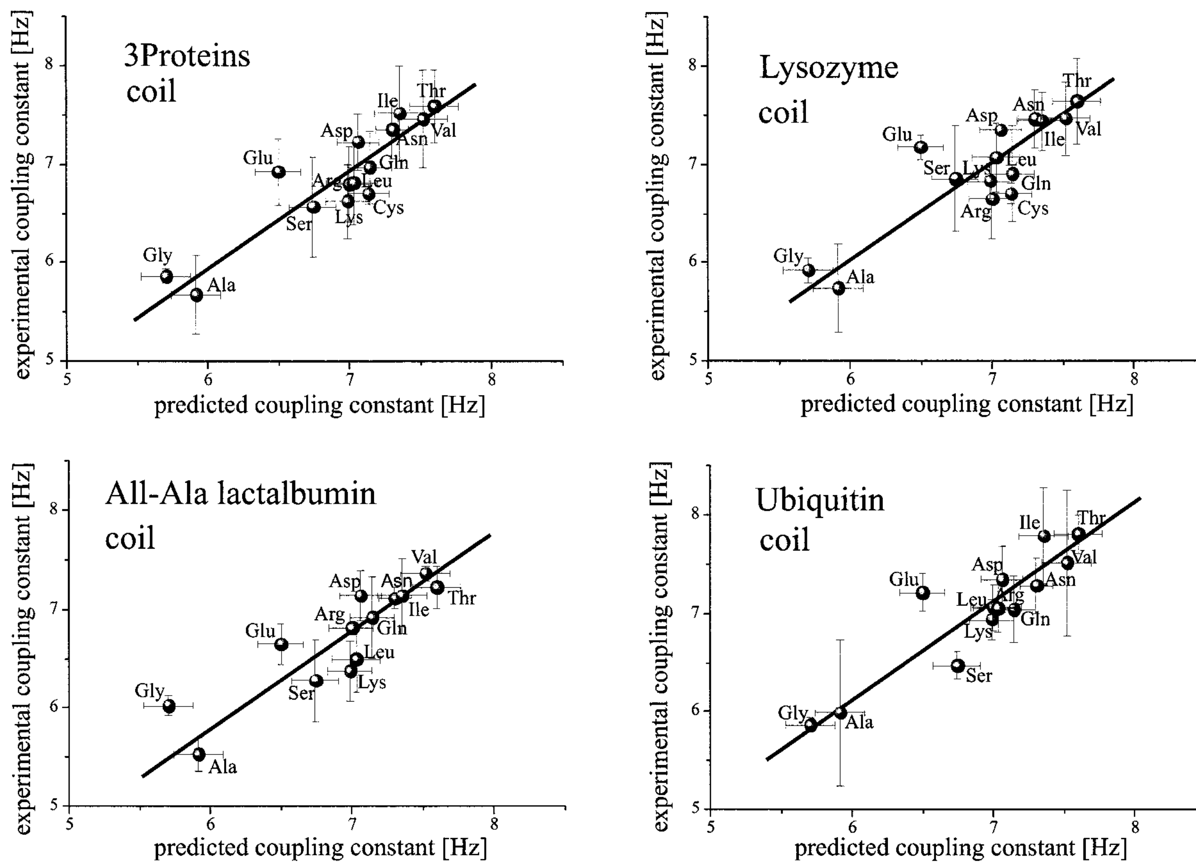

Figure 5. Comparison of experimental $^3J(H^N,H^\alpha)$ coupling constants (aromatic residues are excluded) with predicted values from the statistical model for a random coil. The values were predicted as described by Smith et al. (1996a, b) using an enlarged database of 402 native protein structures (Hennig et al., 1999). Values for correlation including aromatic residues are given in Table 4.

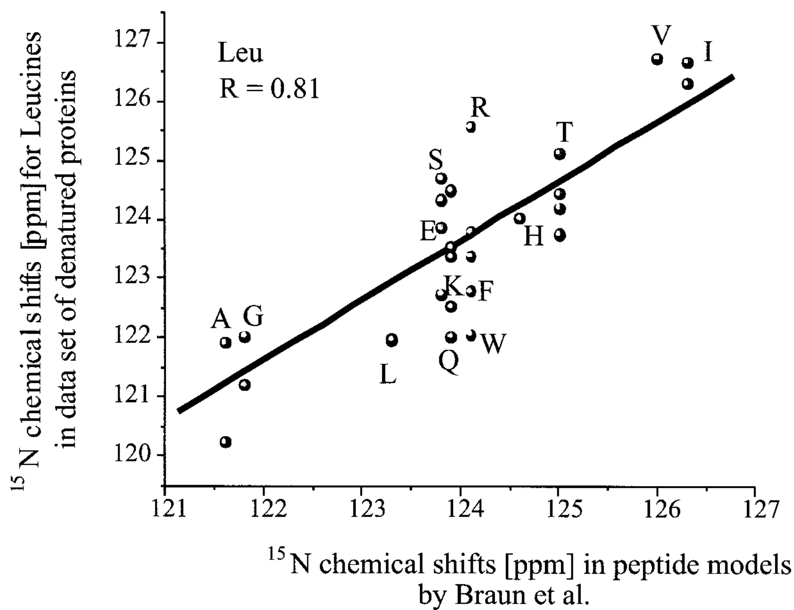

Figure 6. Comparison of experimental ^{15}N chemical shifts of 29 leucine residues with predicted values of Braun et al. (1994).

Table 4. $^{3}J(H^{N},H^{\alpha})$ coupling constants for all pairs of amino acids observed in ubiquitin, lysozyme and all-Ala α-lactalbumin. The standard deviation and the number of measured pairs are given in brackets. Rows give the amino acid i, colums give the amino acid i−1

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	5.72(0.66/4)		5.75		5.74(0.06/2)	5.63(0.12/2)				5.67(0.13/3)	5.39(0.09/3)	5.4(0.08/2)		6.51		5.95(0.37/2)	6.01			5.11
Arg			6.53(0.08/2)			6.86		7.29(0.27/2)			7.14(0.33/2)	6.42				6.7(0.21/3)		6.28		
Asn		7.3(0.26/3)	7.18	7.64	7.53(0.57/2)		7.06(0.04/2)	7.75(0.03/2)		7.1			7.35	7.1		7.22			7.48	
Asp	6.87(0.27/2)			7.21(0.08/2)		7.36	7.39			7.27	7.34	6.89			6.77	7.52(0.15/S)	7.45			
Cys		6.64(0.04/2)									6.85							6.69		
Gln			6.63	7.35		6.58	6.3			7.04(0.31/2)	7.47	6.96(0.16/2)	7.39	7.18		7.11				6.61
Glu	6.5(0.01/2)		6.78		7.09					7.28	7.04(0.54/3)	7.02(0.25/2)		7.26	6.42		6.81(0.02/2)			7.11(0.22/2)
Gly	5.96	5.84		5.81			5.86	5.77									5.93			
His	**7.25**										**7.95**			**7.34**						
Ile			**7.42(0.24/3)**	**7.09(0.04/2)**		**7.87(0.86/2)**		**7.19(0.48/3)**			**8.21**	**7.41**					**7.87(0.42/2)**	**7.72**		
Leu	6.38	7.46(0.35/2)	7.24			6.42(0.57/2)	6.55(0.51/3)	6.74(0.35/2)	7.1	7.03(0.08/2)	6.52(0.39/2)	6.45(0.42/2)		6.58		6.63	7(0.27/4)	6.9(0.01/2)		7.31
Lys	6.42(0.37/4)		6.67	6.63(0.49/2)	6.7		6.72(0.4/3)	6.94(0.24/2)	6.28	6.72	6.71	6.15	7.03				5.91			7.01(0.25/2)
Met																				
Phe			**7.15**			**7.21**				**8.02(0.43/2)**	**7.08**	**7.14**								
Pro																				
Ser	8		6.5		6.33	6.15	6.6(0.26/3)	6.49		6.24	6.17(0.59/3)	6.25			6.45	6.81	6.65		7.2	6.83
Thr	**7.44(0.32/3)**	**7.64**	**7.62(0.44/2)**	**7.26(0.56/2)**				**7.94**	**7.43**	**7.44(0.61/3)**	**8.02**	**7.67(0.01/2)**				**7.93(0.06/2)**				
Trp	**6.26(0.23/2)**	**6.05**	**6.56**								**6.87**							**6.52**	**6.34**	
Tyr				**6.23(1.07/2)**			**7.08**													
Val			**7.36**	**7.09**		**7.31**	**7.93**			**7.41**	**8.21**			**6.54**				**7.66(0.26/2)**		

suggesting that both chemical shifts and coupling constants reflect on the tendency towards more extended structure. For Thr7/Leu8, for which this correlation is not observed, we propose a non-random structure in ubiquitin as evidenced by the differences in chemical shifts and by increased heteronuclear relaxation rates (Peti and Schwalbe, in preparation).

Conclusions

The chemical shift assignments for denatured ubiquitin have been reported. A good correlation between the mean chemical shifts in the database of native folded proteins and the random coil values of the three proteins investigated here has been found. Taking nearest neighbor effects into account, the averaged experimental chemical shifts can be correlated with the prediction given by Braun et al. (1994). Further analysis, however, also reveals pronounced effects that cannot be predicted only by analysis of dipeptide pairs in the unfolded states of proteins. This implies that, in order to describe with accuracy the averaged conformation in a denatured protein, additional nearest neighbor effects need to be taken into account.

Acknowledgements

5 mg ^{15}N labeled ubiquitin and 5 mg ^{15}N,^{13}C labeled ubiquitin were awarded from VLI Research Inc., Malvern, PA. Spectra were recorded at the Large Scale Facility for Biomolecular NMR at the University of Frankfurt, at Bruker Karlsruhe and at the Francis Bitter Magnet Laboratory at the Massachusetts Institute of Technology. We thank Dr. W. Bermel for measurement time and support. We thank Prof. C. Griesinger and Prof. C.M. Dobson for encouragement, and Steffen Maurer for help with the data analysis. The work was supported by the Fonds der Chemischen Industrie (Kekulé stipend for W.P.), by the DFG (Schw 701/3-1) and by the Massachusetts Institute of Technology. L.J.S. is a Royal Society University Research Fellow. C.R. acknowledges the BBSRC for an Advanced Research Fellowship. C.R. and L.J.S. are members of the Oxford Centre for Molecular Science, which is founded by BBSRC, EPSRC and MRC. A list of ^{1}H,^{15}N, ^{13}C chemical shifts for denatured human ubiquitin in 8M urea, pH 2, can be obtained upon request (schwalbe@mit.edu).

References

Arcus, V.L., Vuilleumier, S., Freund, S.M., Bycroft, M. and Fersht, A.R. (1995) *J. Mol. Biol.*, **254**, 305–321.
Arcus, V.L., Vuilleumier, S., Freund, S.M., Bycroft, M. and Fersht, A.R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 9412–9416.
Blanco, F.J., Serrano, L. and Forman-Kay, J.D. (1998) *J. Mol. Biol.*, **284**, 1153–1164.
Braun, D., Wider, G. and Wüthrich, K. (1994) *J. Am. Chem. Soc.*, **116**, 8466–8469.
Briggs, M.S. and Roder, H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2017–2021.
Brutscher, B., Brüschweiler, R. and Ernst, R.R. (1997) *Biochemistry*, **36**, 13043–13053.
Buck, M., Schwalbe, H. and Dobson, C.M. (1995) *Biochemistry*, **34**, 13219–13232.
Carlomagno, T., Maurer, M., Hennig, M. and Griesinger, C. (2000) *J. Am. Chem. Soc.*, **122**, 5105–5113.
Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.
Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1999) *J. Am. Chem. Soc.*, **120**, 6836–6837.
Di Stefano, D.L. and Wand, A.L. (1987) *Biochemistry*, **26**, 7272–7281.
Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T. and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 3542–3556.
Dyson, J. and Wright, P.E. (1998) *Nat. Struct. Biol.*, **5**, 499–503.
Eliezer, D., Yao, J., Dyson, H.J. and Wright, P.E. (1998) *Nat. Struct. Biol.*, **5**, 148–155.
Emsley, L. and Bodenhausen, G. (1990) *Chem. Phys. Lett.*, **165**, 469–476.
Emsley, L. and Bodenhausen, G. (1992) *J. Magn. Reson.*, **97**, 135–148.
Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J. and Dobson, C.M. (1996) *J. Phys. Chem.*, **100**, 2661–2666.
Frank, M.K., Clore, G.M. and Gronenborn, A.M. (1995) *Protein Sci.*, **4**, 2605–2610.
Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. and Feeney, J. (1990) *J. Magn. Reson.*, **90**, 420–425.
Fushman, D. and Cowburn, D. (1998) *J. Am. Chem. Soc.*, **120**, 7109–7110.
Fushman, D., Tjandra, N. and Cowburn, D. (1998) *J. Am. Chem. Soc.*, **120**, 10947–10952.
Glushka, J., Lee, M., Coffin, S. and Cowburn, D. (1989) *J. Am. Chem. Soc.*, **111**, 7716–7722.
Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson.*, **B101**, 114–119.
Grzesiek, S. and Bax, A. (1992) *J. Am. Chem. Soc.*, **114**, 6291–6293.
Grzesiek, S. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 12593–12594.
Hennig, M., Bermel, W., Spencer, A., Dobson, C.M., Smith, L.J. and Schwalbe, H. (1999) *J. Mol. Biol.*, **288**, 705–723.
Hershko, A. and Ciechanover, A. (1998) *Annu. Rev. Biochem.*, **67**, 425–479.
Jenson, J., Goldstein, G. and Breslow, E. (1980) *Biochim. Biophys. Acta*, **624**, 378–385.
Kay, L.E. and Bax, A. (1990) *J. Magn. Reson.*, **86**, 110–126.
Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.
Lenkinski, R.E., Douglas, C.M., Glickson, J.D. and Goldstein, G. (1977) *Biochim. Biophys. Acta*, **494**, 126–130.
Logan, T.M., Olejniczak, E.T., Xu, R.X. and Fesik, S.W. (1993) *J. Biomol. NMR*, **2**, 225–231.

Logan, T.M., Theriault, Y. and Fesik, S.W. (1994) *J. Mol. Biol.*, **236**, 637–648.
Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) *J. Mol. Biol.*, **217**, 731–736.
Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989a) *Biochemistry*, **28**, 6150–6156.
Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989b) *J. Am. Chem. Soc.*, **111**, 1515–1517.
Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biomol. NMR*, **5**, 14–24.
Nash, D.P. and Jonas, J. (1997) *Biochem. Biophys. Res. Commun.*, **238**, 289–291.
Neri, D., Billeter, M., Wider, G. and Wüthrich, K. (1992a) *Science*, **257**, 1559–1563.
Neri, D., Wider, G. and Wüthrich, K. (1992b) *Proc. Natl. Acad. Sci. USA*, **89**, 4397–4401.
Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **180**, 741–751.
Penkett, C.J., Redfield, C., Dodd, I., Hubbard, J., McBay, D.L., Mossakowska, D.E., Smith, R.A., Dobson, C.M. and Smith, L.J. (1997) *J. Mol. Biol.*, **274**, 152–159.
Penkett, C.J., Redfield, C., Jones, J.A., Dodd, I., Hubbard, J., Smith, R.A., Smith, L.J. and Dobson, C.M. (1998) *Biochemistry*, **37**, 17054–17067.
Peti, W., Hennig, M., Smith, L.J. and Schwalbe, H. (2000) *J. Am. Chem. Soc.*, **122**, 12017–12018.
Plaxco, K.W., Morton, C.J., Grimshaw, S.B., Jones, J.A., Pitkeathly, M., Campbell, I.D. and Dobson, C.M. (1997) *J. Biomol. NMR*, **10**, 221–230.
Redfield, C., Schulman, B.A., Milhollen, M.A., Kim, P.S. and Dobson, C.M. (1999) *Nat. Struct. Biol.*, **6**, 948–952.
Sattler, M., Schleucher, J. and Griesinger, C. (1999) *Prog. NMR Spectrosc.*, **34**, 93–158.
Schneider, D.M., Dellwo, M.J. and Wand, A.J. (1992) *Biochemistry*, **31**, 3645–3652.
Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J. and Dobson, C.M. (1997) *Biochemistry*, **36**, 8977–8991.
Shortle, D. (1996) *Curr. Opin. Struct. Biol.*, **6**, 24–30.
Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M. and Dobson, C.M. (1996a) *J. Mol. Biol.*, **255**, 494–506.
Smith, L.J., Fiebig, K.M., Schwalbe, H. and Dobson, C.M. (1996b) *Fold. Des.*, **1**, R95–106.
Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
Thanabal, V., Omecinsky, D.O., Reily, M.D. and Cody, W.L. (1994) *J. Biomol. NMR*, **4**, 47–59.
Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 12562–12566.
Tjandra, N., Szabo, A. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 6986–6991.
Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.
Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 7772–7777.
Weber, P.L., Brown, S.C. and Mueller, L. (1987) *Biochemistry*, **26**, 7282–7290.
Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995a) *J. Biomol. NMR*, **5**, 67–81.
Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995b) *J. Biomol. NMR*, **6**, 135–140.
Wishart, D.S. and Nip, A.M. (1998) *Biochem. Cell Biol.*, **76**, 153–163.
Wishart, D.S. and Sykes, B.D. (1994a) *J. Biomol. NMR*, **4**, 171–180.
Wishart, D.S. and Sykes, B.D. (1994b) *Methods Enzymol.*, **239**, 363–392.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991a) *J. Mol. Biol.*, **222**, 311–333.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991b) *FEBS Lett.*, **293**, 72–80.
Wishart, D.S., Watson, M.S., Boyko, R.F. and Sykes, B.D. (1997) *J. Biomol. NMR*, **10**, 329–336.
Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.*, **B101**, 201–205.
Wong, K.B., Freund, S.M. and Fersht, A.R. (1996) *J. Mol. Biol.*, **259**, 805–818.
Yang, D. and Kay, L.E. (1996) *J. Mol. Biol.*, **263**, 369–382.