Journal of Biomolecular NMR, **18:** 43–48, 2000.
KLUWER/ESCOM

Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView

Stephan Schwarzinger, Gerard J.A. Kroon, Ted R. Foss, Peter E. Wright* & H. Jane Dyson*
Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, MB-2, 10550 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

Received 19 May 2000; Accepted 29 June 2000

Key words: chemical shift, CSI, denaturant, NMRView, peptide, random coil

Abstract

Studies of proteins unfolded in acid or chemical denaturant can help in unraveling events during the earliest phases of protein folding. In order for meaningful comparisons to be made of residual structure in unfolded states, it is necessary to use random coil chemical shifts that are valid for the experimental system under study. We present a set of random coil chemical shifts obtained for model peptides under experimental conditions used in studies of denatured proteins. This new set, together with previously published data sets, has been incorporated into a software interface for NMRView, allowing selection of the random coil data set that fits the experimental conditions best.

Abbreviations: DMF, dimethylformamide; TFA, trifluoroacetic acid; Fmoc, 9-fluorenylmethoxycarbonyl; AcN, acetonitrile, TFE; 2,2,2-trifluoroethanol; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt; CSI, chemical shift index.

Introduction

Certain chemical shifts are sensitive indicators for secondary structure elements (Gross and Kalbitzer, 1988; Szilágyi and Jardetzky, 1989; Pastore and Saudek, 1990; Spera and Bax, 1991). Since chemical shifts are strongly dependent on the identity of the amino acid residue (for example, the $^{13}C\beta$ resonance of Thr is usually ∼70 ppm, while that of Ala is ∼17 ppm), the observed chemical shifts are usually normalized by subtraction of a standardized 'random coil' chemical shift that reflects the amino acid-specific component of the chemical shift. Over time a number of random coil data sets based on either a database (Wishart et al., 1991) or model peptides under a variety of experimental conditions (Richarz and Wüthrich, 1978; Bundi and Wüthrich, 1979; Jimenez et al., 1986; Braun et al., 1994; Thanabal et al., 1994; Merutka et al., 1995; Wishart et al., 1995a; Plaxco et al., 1997) have been published.

In addition to amino acid-specific effects, the chemical shifts of several backbone nuclei are strongly affected by the immediate local sequence and by conditions such as temperature and pH. These nuclei include ^{1}HN, ^{15}N and $^{13}C'$, and the superior chemical shift dispersion of these nuclei in unfolded states of proteins makes them of particular use in the assignment process for such states (Yao et al., 1997).

Secondary structures such as helix and β-sheet can be readily identified in folded proteins from Hα, Hβ, Cα and Cβ chemical shifts. These values are standardized by subtraction of the appropriate random coil shift, and the propensity for secondary structure at a given position in the amino acid sequence is frequently assessed by calculating the chemical shift index (CSI) which combines the data from these nuclei (Wishart et al., 1992). For unfolded proteins, the difference between the observed and random coil shifts is generally much smaller, and thus a meaningful assessment of the presence and extent of residual secondary structure in unfolded proteins requires accurate reference values obtained under similar experimental conditions. Con-

*To whom correspondence should be addressed.

centrated urea solutions at acidic pH have previously been used to investigate residual structure in unfolded proteins (Neri et al., 1992; Logan et al., 1994; Arcus et al., 1995; Frank et al., 1995; Schwalbe et al., 1997). At such extreme conditions, pH and hydrogen bonding from the co-solvent will contribute to changes of the chemical shifts observed. In order to evaluate residual secondary structure in urea-unfolded proteins we measured ^{13}C, ^{15}N and ^{1}H random coil chemical shift values for a series of model peptides (Ac-GGXGG-NH_2) in 8 M urea at a pH of 2.3.

Materials and methods

The peptides were synthesized on an automated synthesizer (Pioneer, PerSeptive) using standard Fmoc-chemistry on a 0.1 mM scale on Fmoc-[5-(4-Fmoc-aminomethyl-3,5-dimethoxyphenoxy)valeric acid]-polyethyleneglycol-polystyrene (PAL-PEG-PS) support. Piperidine (20% in DMF) was used for deblocking, benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphoniumhexafluorophosphate (PyBOP) (0.5 M in DMF) and diisopropylethylamine (DIPEA) (1 M in DMF) were used as activators, and a mixture of 2.8% acetic anhydride and 10% methylene chloride in DMF was used for N-terminal capping. After washing, the peptides were cleaved using a mixture of 2.5% of triisopropylsilane and 2.5% water in TFA and precipitated with chilled ether. The lyophilized peptides were found to be sufficiently pure by mass spectrometry and NMR analysis, and were used without further purification. Typically 5 to 8 mg of solid was dissolved in 0.5 ml 8 M urea, pH 2.3, containing 10% D_2O and 10 mM DSS as an internal reference, giving a peptide concentration of $\sim$20 mM. The pH of the sample was adjusted to 2.3 ($\pm$ 0.05) with concentrated hydrochloric acid. The final chloride content was about 150 mM.

NMR experiments were carried out on a Bruker DMX 750 Avance spectrometer equipped with a 5 mm TXI-gradient probe. The probe temperature was calibrated to 293 K using neat methanol (Van Geet, 1970). Proton shifts were measured using a 1D WATERGATE experiment (750.13 MHz, 14 ppm spectral width, with the carrier placed on the water signal at 5.08 ppm, td = 8K), carbon shifts were extracted from a 1D natural abundance ^{13}C experiment (188.62 MHz, 220 ppm spectral width, center frequency at 100 ppm, td = 16K, typically 8K scans), nitrogen chemical shifts were measured from a natural abundance ^{15}N-HSQC (Zhang et al., 1994) with water flip-back solvent suppression (Grzesiek and Bax, 1993) (SW = 8 ppm ^{1}H, 1923 Hz ^{15}N, center at 6.5 ppm ^{1}H, 117 ppm ^{15}N, 76.01 MHz, 1K times 64). DQF-COSY (Rance et al., 1983) (SW = 15 ppm, 8264 Hz, center at 5.08 ppm, 4K times 128), TOCSY (Braunschweiler and Ernst, 1983) (SW = 16 ppm, 8264 Hz, center at 5.08 ppm, 4K times 256) and natural abundance ^{13}C-HSQC (Vuister and Bax, 1992) (SW = 15 ppm in ^{1}H, 10204 Hz/13228 Hz in ^{13}C, center at 3 ppm ^{1}H, 43/40 ppm ^{13}C, 4K times 128) were used to resolve ambiguities. The total time required to collect these six experiments was typically less than 12 h per sample. Spectra were processed using the program NMRPipe (Delaglio et al., 1995) and analysis was carried out using the program NMRView (Johnson and Blevins, 1994).

Results and discussion

Examples of typical spectra are shown in Figure 1 (for ^{1}H and natural abundance ^{13}C) and Figure 2 (for natural abundance ^{15}N). Resonances were assigned by comparison to previously published data. Ambiguous assignments were resolved using the 2 dimensional NMR methods mentioned above. All proton chemical shifts were referenced to the DSS signal at 0 ppm, while the carbon shifts were referenced to the DSS signal at 22 ppm (calibrated to the DSS signal at 0 ppm in a solvent-only sample and by indirect referencing methods using a ratio of 0.25144953) (Wishart et al., 1995b). Neither the ^{13}C nor the ^{1}H chemical shift of DSS is influenced by the presence of 8 M urea, as demonstrated with a sample of DSS in 8 M urea containing a capillary with DSS in water. Nitrogen chemical shifts were referenced to DSS using a ratio of 0.10132912 (Wishart et al., 1995b).

The random coil chemical shifts for the 20 naturally occurring amino acids in 8 M urea at pH 2.3 and 20 °C are summarized in Table 1. A comparison of the values with those obtained by Wishart et al. (1995a) on Ac-GGXAGG-NH_2 hexapeptides in 1 M urea shows systematic deviations, which are attributed in part to the difference in the sequence (Ala or Pro instead of Gly following residue X) and in part to the difference in the solution conditions (concentration of cosolvent, pH). The biggest deviations are found for nuclei involved in hydrogen bonding (HN, C′). Only small deviations were found for those nuclei that are little affected by the sequence or by hydrogen bonding

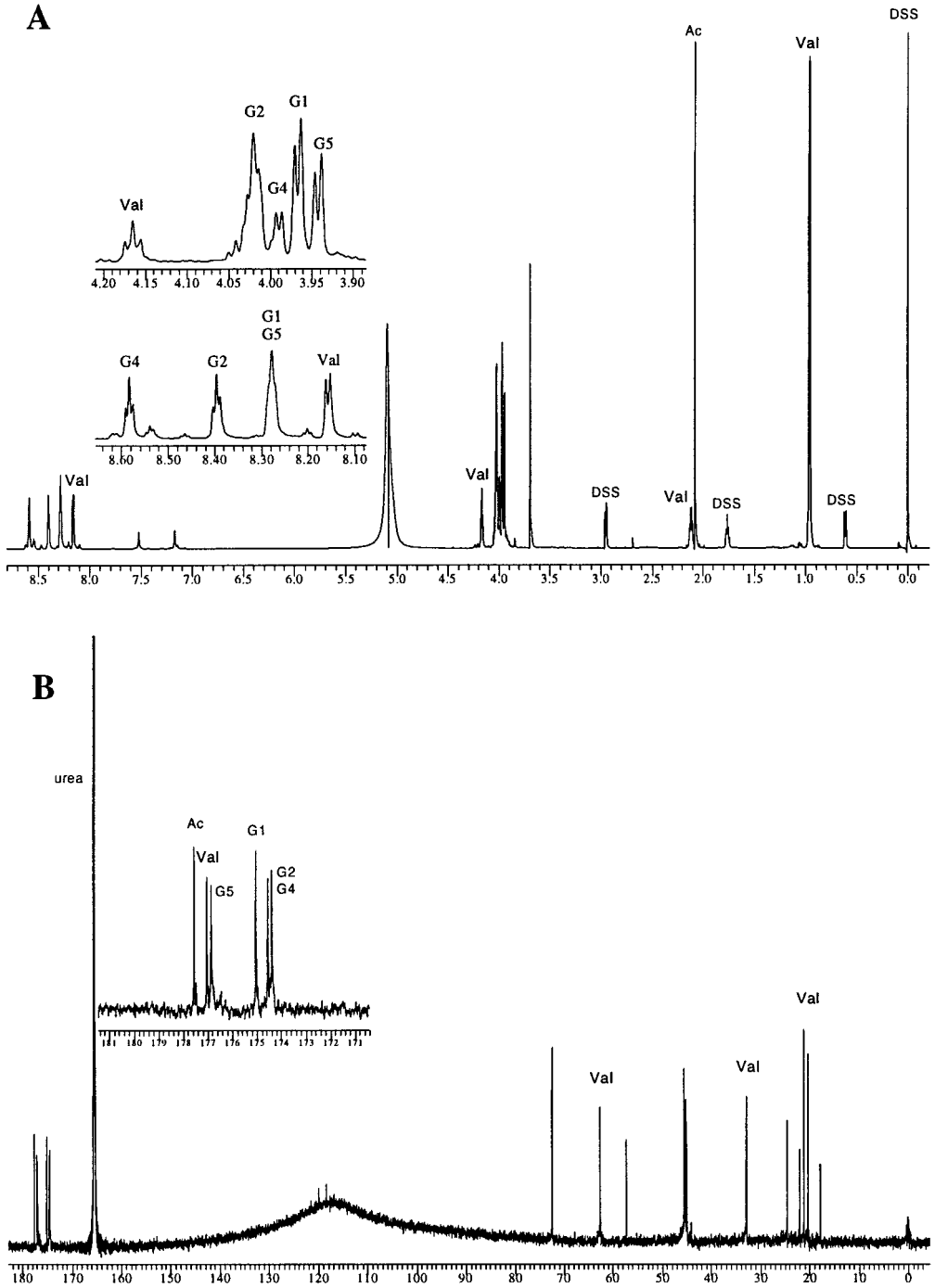

Figure 1. Spectra obtained for the peptide Ac-G-G-V-G-G-NH_2. (a) 1D ^{1}H spectrum with resonance assignments marked. Insets show the Hα and HN regions. (b) 1D ^{13}C spectrum with assignments marked. Individual glycine Cα resonances have not been assigned due to overlap. Unmarked resonances belong to DSS. The inset shows the ^{13}CO region of the spectrum. Figures were generated by the program MestRe-C (Cobas et al., 1999).

Table 1. Random Coil Chemical shifts (in ppm) for the 20 common amino acids in acidic 8 M urea

Residue	H^N	H^α	$H^\beta 1$	$H^\beta 2$	N^H	C^α	C^β	C'
Ala	8.35	4.35	1.42		125.0	52.8	19.3	178.5
Cys-ox	8.54	4.76	3.29	3.02	118.7	55.6	41.2	175.5
Cys-red	8.44	4.59	2.98	2.98	118.8	58.6	28.3	175.3
Asp	8.56	4.82	2.98	2.91	119.1	53.0	38.3	175.9
Glu	8.40	4.42	2.18	2.01	120.2	56.1	29.9	176.8
Phe	8.31	4.65	3.19	3.04	120.7	58.1	39.8	176.6
Gly	8.41	4.02			107.5	45.4		174.9
His	8.56	4.79	3.35	3.19	118.1	55.4	29.1	175.1
Ile	8.17	4.21	1.89		120.4	61.6	38.9	177.1
Lys	8.36	4.36	1.89	1.77	121.6	56.7	33.2	177.4
Leu	8.28	4.38	1.67	1.62	122.4	55.5	42.5	178.2
Met	8.42	4.52	2.15	2.03	120.3	55.8	32.9	177.1
Asn	8.51	4.79	2.88	2.81	119.0	53.3	39.1	176.1
Pro-trans		4.45	2.29	1.99		63.7	32.2	177.8
Pro-cis		4.60	2.39	2.18		63.0	34.8	n.d.
Gln	8.44	4.38	2.17	2.01	120.5	56.2	29.5	176.8
Arg	8.39	4.38	1.91	1.79	121.2	56.5	30.9	177.1
Ser	8.43	4.51	3.95	3.90	115.5	58.7	64.1	175.4
Thr	8.25	4.43	4.33		112.0	62.0	70.0	175.6
Val	8.16	4.16	2.11		119.3	62.6	31.8	177.0
Trp	8.22	4.70	3.34	3.25	122.1	57.6	29.8	177.1
Tyr	8.26	4.58	3.09	2.97	120.9	58.3	38.9	176.7

(Hα, Hβ, Cα, Cβ). However, because of the significantly smaller differences between the chemical shifts in unfolded proteins, compared with those of folded proteins, these small variations have to be considered in order to avoid systematic errors in chemical shift calculations of unfolded proteins, leading to the over- or under-estimation of the population of residual secondary structure.

In addition, an estimate of the random-coil $^3J_{HN,H\alpha}$ coupling constant was made for a total of 13 amino acids for which the NH region of the 1D spectrum was sufficiently well-resolved for the measurement to be made by inspection. A comparison of these values with those previously obtained in GdmHCl (Plaxco et al., 1997) is shown in Table 2. The differences in $^3J_{HN,H\alpha}$ are negligible between the two denaturing solvents.

The random coil chemical shift data set for the amino acids in 8 M urea at pH 2.3 was incorporated, together with previously published data sets, into a new tool for NMRView (Johnson and Blevins, 1994). The purpose of this is to allow the user to choose the random coil data set closest to the experimental condi-

Table 2. Comparison of $^3J(H^NH^\alpha)$ coupling constants measured in 8 M urea at pH 2.3 and in 6 M GdmHCl at pH 5.0 (Plaxco et al., 1997)

Residue	$^3J_{HNH\alpha}$ (Hz)	
	Urea	GdmHCl
Ala	5.8	6.1
Cys-red	7.4	7.3
Asp	7.5	7.8
Phe	7.2	7.3
His	7.7	7.8
Ile	7.2	7.1
Lys	6.8	7.0
Leu	6.5	6.8
Asn	7.6	7.7
Gln	7.4	7.1
Arg	6.7	6.9
Thr	7.8	7.9
Val	7.3	7.2

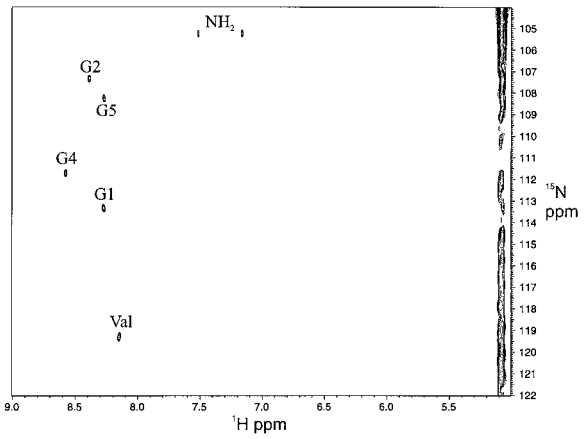

Figure 2. Natural abundance ^{15}N-^{1}H HSQC spectrum obtained for the peptide Ac-G-G-V-G-G-NH_2. Assignments are indicated. The residual water resonance is visible on the right of the panel.

tions or to compare with results previously obtained with different data sets. To match the experimental conditions as closely as possible the user is able to correct for temperature, pH and cosolvent effects, as well as for different standard compounds used for referencing. Options have also been included to correct for isotope effects in deuterated proteins and for small chemical shift deviations arising, for example, in TROSY-type experiments (Pervushin et al., 1997).

Once chemical shifts have been assigned in NMRView, the new tool is used as follows: after a random coil database has been selected, the program calculates the difference (Δ-ppm) between observed and random shifts for a given nucleus in a given residue, and evaluates the CSI. For each nucleus type, two output files are generated, one with the residue number and CSI and the other with residue number, observed chemical shift and Δ-ppm. In addition, there is an assignment flag to verify assignments for unfolded proteins or flexible parts of a protein. This flag is 0 for residues whose chemical shift is within a user selected boundary (a multiple/fraction of the cutoff values employed in CSI calculations) and is 1 for possible mis-assigned residues. This flag will give very clear results for most residues, but it has to be used with care for certain amino acid types with similar random coil shifts (such as Phe/Tyr or Glu/Gln/Arg). The program also calculates a consensus CSI value (Wishart and Sykes, 1994) for the selected nuclei.

The random coil chemical shift values in 8 M urea at pH 2.3, 20 °C have been submitted to the BMRB (accession number 4747). The extended CSI interface for NMRView is available for download as part of the TSRI contribution for NMRView (http://garbanzo.scripps.edu/nmrgrp/wisdom/pipe/tips_scripts.html).

Acknowledgements

We thank Dr. John Chung for assistance with the NMR experiments. S.S. acknowledges an Erwin Schrödinger Fellowship of the Austrian Science Funding Agency (FWF J1736-CHE). This project was supported by grant GM57374 from the National Institutes of Health.

References

Arcus, V.L., Vuilleumier, S., Freund, S.M.V., Bycroft, M. and Fersht, A.R. (1995) *J. Mol. Biol.*, **254**, 305–321.

Braun, D., Wider, G. and Wüthrich, K. (1994) *J. Am. Chem. Soc.*, **116**, 8466–8469.

Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.*, **53**, 521–528.

Bundi, A. and Wüthrich, K. (1979) *Biopolymers*, **18**, 285–297.

Cobas, C., Cruces, J. and Sardina, F.J. (1999) http://qobrue.usc.es/jsgroup/MestRe-C/MestRe-C.html

Delaglio, F., Grzesiek, S., Vuister, G.W., Huang, Z., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Frank, M.K., Clore, G.M. and Gronenborn, A.M. (1995) *Protein Sci.*, **4**, 2605–2615.

Gross, K.-H. and Kalbitzer, H.R. (1988) *J. Magn. Reson.*, **76**, 87–99.

Grzesiek, S. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 12593–12594.

Jimenez, M.A., Nieto, J.L., Rico, M., Santoro, J., Herranz, J. and Bermejo, F.J. (1986) *J. Mol. Struct.*, **143**, 435–438.

Johnson, B.A. and Blevins, R.A. (1994) *J. Chem. Phys.*, **29**, 1012–1014.

Logan, T.M., Thériault, Y. and Fesik, S.W. (1994) *J. Mol. Biol.*, **236**, 637–648.

Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biomol. NMR*, **5**, 14–24.

Neri, D., Billeter, M., Wider, G. and Wüthrich, K. (1992) *Science*, **257**, 1559–1563.

Pastore, A. and Saudek, V. (1990) *J. Magn. Reson.*, **90**, 165–176.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.

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.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Richarz, R. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2133–2141.

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.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Szilágyi, L. and Jardetzky, O. (1989) *J. Magn. Reson.*, **83**, 441–449.

Thanabal, V., Omecinsky, D.O., Reily, M.D. and Cody, W.L. (1994) *J. Biomol. NMR*, **4**, 47–59.

Van Geet, A.L. (1970) *Anal. Chem.*, **42**, 679–680.

Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.* , **98**, 428–435.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.
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 Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Yao, J., Dyson, H.J. and Wright, P.E. (1997) *FEBS Lett.*, **419**, 285–289.
Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR*, **4**, 845–858.