Volume 293, number 1,2, 72–80 FEBS 10388 November 1991

Simple techniques for the quantification of protein secondary structure by ^{1}H NMR spectroscopy

D.S. Wishart[1,2], B.D. Sykes[2] and F.M. Richards[1]

[1]*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA* and [2]*MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, AB, T6G 2H7, Canada*

Received 19 August 1991; revised version received 18 September 1991

Previous work by Wishart et al. (in press) and others [(1989) J. Magn. Reson. 83, 441–449; (1990) J. Magn. Reson. 90, 165–176] has shown a strong tendency for protein secondary structure to be manifested in ^{1}H NMR chemical shifts. Based on these earlier results, two techniques have been developed for the quantification of secondary structure in proteins. Both methods allow for the rapid and accurate determination of the percent content of helix, coil, and β-strand based on the integration (or peak enumeration) of selected portions of either 1-D or 2-D ^{1}H NMR spectra. These new and very simple procedures have been found to compare quite favorably to other well established techniques for secondary structure determination such as CD, Raman and IR spectroscopy.

NMR; Chemical shift; Secondary structure; Protein structure; Circular dichroism

1. INTRODUCTION

X-ray crystallography and, more recently, NMR spectroscopy have permitted biochemists to study the structure and conformation of peptides and proteins with unprecendented detail. However, structural determination and conformational analysis using these two techniques often requires many months, if not years, of painstaking work. Furthermore, the complexity of these procedures often limits the accessibility of such techniques to all but a few specialists. As a result, there is still a very strong demand among many protein chemists to seek more 'user-friendly' spectroscopy methods to conduct their own, lower resolution, conformational studies without the inconvenience of these more time-consuming techniques. Perhaps this, in part, explains the continued popularity and widespread use of CD, Raman and IR spectroscopy in protein chemistry to this day. Indeed, all three of these optical techniques permit quantitative, medium-resolution characterizations of protein structure within the time-frame of a single day.

In this paper we wish to demonstrate how NMR spectroscopy can provide the same quantitative assessment of protein structure within the same time-frame and with equal accuracy and ease as either CD, Raman or infrared spectroscopy. Indeed, a number of comparisons suggest that these new NMR methods may be more accurate and their implementation far easier than for CD, Raman or IR. Whereas these optical methods often require specialized software packages to conduct Fourier deconvolution [2], singular value decomposition [3,4] or multi-variate constrained regularization [5], this new NMR approach requires little more than an ability to count peaks or to integrate over peak areas.

The impetus to develop such an NMR approach began with attempts to find practical applications for the previously reported chemical shift tendencies found for amino acids in α-helices, β-strands and coils [1,6,65]. This work has already led to a number of other useful developments concerning chemical shift and protein structure but it is this particular application of secondary structure quantitation that we wish to report here.

Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; COSY, two-dimensional correlated spectroscopy; CD, circular dichroism spectroscopy; IR, infrared spectroscopy.

Correspondence address: D.S. Wishart, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA.

2. MATERIALS AND METHODS

Two methods are presented for the direct quantitation of protein secondary structure. Method 1 is for the quantitation of secondary structure through 1-D ^{1}H NMR while Method 2 is for the quantitation of secondary structure through 2-D ^{1}H NMR. Both techniques are essentially empirical in nature and both are based on the integration of peaks in selected regions of the α-proton and amide/aromatic regions of ^{1}H NMR spectra. In order to use either technique the protein of interest must meet several criteria. It must be: (i) water soluble; (ii) stable at pH 6.0 or less, and (iii) have its full sequence or amino acid content known.

Development and testing of both procedures have been based primarily on a database of protein chemical shift values and secondary structure assignments collected from a variety of published sources. These are listed in Table I. Secondary structure assignments for approximately half of the listed proteins were supplemented or corroborated by additional crystallographic data obtained from the Brook-

 Published by Elsevier Science Publishers B.V.

haven Protein Data Bank [66] as well as from several previously published structural databases [67,68]. A more detailed discussion of the secondary structural assignments is given in [1]. A database containing the ^1H chemical shift assignments and secondary structures of more than 70 proteins is available from the authors upon request.

2.1. *Method 1*

This protocol has been specifically designed for the determination of protein secondary structure content through simple integration of one-dimensional ^1H NMR spectra. Fig. 1 illustrates the central idea behind this particular approach. As can be seen from this figure, it is necessary to collect 2 identical spectra for the protein of interest, 1 in D_2O and another in H_2O. This may be done by dividing the original protein sample in two equal portions, lyophilizing each and then redissolving one aliquot in 90% H_2O/10% D_2O and the other in an equal volume of 99.9% D_2O. Alternatively it is possible to 'recycle' just 1 sample (as was done for this work) through each of the 2 solvent conditions. This can generally be done without any loss in accurary. In both cases it is necessary to collect the 2 spectra under identical conditions (temperature, pH, number of scans, resolution, spectral width, relaxation delays etc.).

The ^1H spectrum collected in H_2O is used for integrating over portions of the amide region while the ^1H spectrum collected in D_2O is used for integrating over portions of the α-^1H region. To be assured that all protons in the amide region are at their fullest intensity (in the H_2O spectrum) it is necessary to work under restricted pH conditions and to use either 'mild' saturation of the solvent resonance or tailored excitation of the protein resonances. Likewise, in order to shift the HDO resonance (in the D_2O spectrum) far enough away from the α-^1H envelope it is usually necessary to work under slightly elevated temperatures. Hence, by maintaining the pH between 3.0–6.0 and by keeping the temperature between 25–35°C, both requirements may be met. In general, these are not very restrictive conditions for most proteins.

It has been determined, through a series of tests and collected

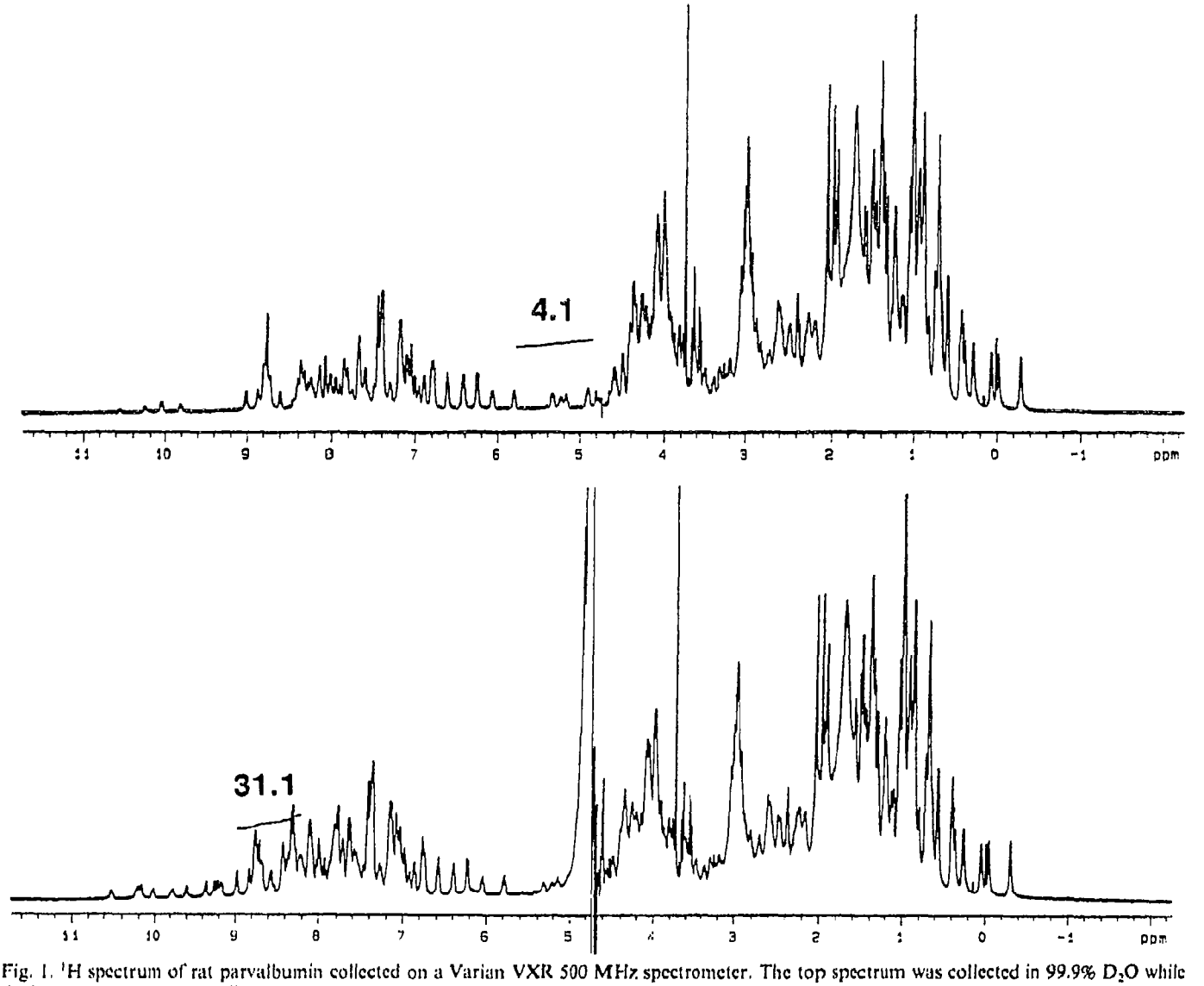

Fig. 1. ^1H spectrum of rat parvalbumin collected on a Varian VXR 500 MHz spectrometer. The top spectrum was collected in 99.9% D_2O while the bottom spectrum was collected in 90% H_2O/10% D_2O. The spectral sweepwidth for both was 7000 Hz and the relaxation delay was set to 4.0 s. The H_2O resonance was suppressed by simple presaturation. The pH for this sample was 4.6 and the protein concentration approximately 1.5 mM. Both spectra were collected at 30°C using a total of 32 scans each. Integration was performed after baseline correction and appropriate scaling. Values of the respective integrals (corresponding to the estimated number of resonance) are printed above each region (4.85–5.90 ppm and 8.20–9.00 ppm).

Table I

Summary of peptides and proteins used in development and testing of Methods 1 and 2

Protein (# of residues)	Source	Conditions		Reference[a]
		pH	°C	
Acyl Carrier Protein (77)	*E. coli*	6.1	25	[8]
α Bungarotoxin (74)	*B. multicinctus*	4.0	35	[9]
α Neurotoxin (60)	*D. polylepsis*	4.2	36	[10]
α Purothionin (45)	Wheat	4.0	25	[11]
Anaphylotoxin C3a (77)	Bovine	5.5	27	[12]
Anaphylotoxin C5a (75)	Bovine	2.3	10	[13]
Anenome Toxin BDS I (43)	*A. sulcata*	3.0	27	[14]
Antennapedia Homeo (68)	*D. melanogaster*	4.3	20	[15]
Anthopleurin A (49)	*A. xanthogrammica*	4.5	27	[16]
Apamin (18)	*A. mellifera*	2.0	25	[17]
Arc Repressor (53)	Phage P22	5.4	50	[18]
Bull Seminal Inhibitor (57)	Bovine	4.9	45	[19]
Calbindin (76)	Porcine	6.0	27	[20]
Calmodulin (148)	*D. melanogaster*	6.3	47	[21]
Cardiotoxin CTX II (60)	*N. mossambica*	3.6	45	[22]
Cecropin A (37)	*H. cecropia*	5.0	25	[23]
Cellobiohydrolase I (36)	*T. reesei*	3.9	27	[24]
Cro repressor (66)	Phage λ	6.8	35	[25]
Cytochrome B5 (82)	Bovine	4.0	40	[26]
Cytochrome C (104)	Horse	5.8	40	[27]
Cytochrome C2 (116)	*R. capsulatus*	6.0	30	[28]
Cytochrome C551 (82)	*R. aeruginosa*	3.5	32	[29]
Elgin C (69)	Leech	3.0	36	[30]
Flavodoxin (137)	*M. elsdenii*	8.3	43	[31]
Hirudin (65)	*H. medicinalis*	3.0	25	[32]
Histidine Cont. Protein (85)	*E. coli*	6.5	30	[33]
Histone H5 (77)	Chicken	3.7	25	[34]
Insulin (51)	Human	3.6	27	[35]
Interleukin 1-β (153)	Human	5.4	36	[36]
Interleukin 8 (72)	Human	5.2	40	[37]
Lac Repressor (51)	*E. coli*	6.9	18	[38]
Leucine Zipper GCN4 (33)	Yeast	5.0	20	[39]
Lysozyme (129)	Chicken	3.8	35	[40]
Lysozyme (130)	Human	3.8	35	[41]
Lysozyme (164)	Phage T4	5.6	20	[42]
Ner Protein (69)	Phage μ	7.0	27	[43]
Neutrophil Peptide 5 (33)	Rabbit	3.5	20	[44]
Ovomucoid 3rd Domain (56)	Turkey	4.2	25	[45]
Parvalbumin (108)	Pike	6.1	62	[46]
Plastocyanin (97)	*S. obliquus*	6.2	30	[47]
Procarboxypeptidase (81)	Bovine	6.5	15	[48]
Protein EL-30 (58)	*E. coli*	3.5	35	[49]
Ribonuclease A (124)	Bovine	3.2	30	[50]
Ribonuclease T1 (104)	*A. oryzae*	5.5	40	[51]
Staph. Nuclease (136)	*S. aureus*	7.4	37	[52]
Tendamistat (74)	*S. tendae*	3.2	50	[53]
Thioredoxin (105)	Human	5.5	40	[54]
Thioredoxin (108)	*E. coli*	5.7	35	[55]
Trypsin Inhibitor (62)	Ascaris	2.4	40	[56]
Trypsin Inhibitor (58)	Bovine	4.6	68	[57]
Trypsin Inh. Type E (59)	*D. polylepsis*	3.2	50	[58]
Trypsin Inhibitor (28)	Squash	2.0	32	[59]
Ubiquitin (76)	Human	4.7	50	[60]

[a]References are primarily to sources containing resonance assignments and preliminary secondary structural information.

statistics from numerous proteins [1], that the number of resonances found between 8.20 and 9.00 ppm in the 'full intensity' amide envelope is approximately equal to 90% of the number of 'coil' residues in a protein. Note that this range of 8.20–9.00 ppm is almost identical to the range for 'random coil' amide protons quoted by Wüthrich [7]. Additional work has also revealed that the number of resonance found between 4.85–5.90 ppm is almost always equal to half the number of residues in β-strands. It therefore stands to reason that by integrating from 4.85–5.90 ppm and from 8.20–9.00 ppm and by scaling these integrals appropriately it is possible to accurately determine the β-

strand content, the coil content and, by default, the α-helix content of almost any protein.

Accurate integration and careful scaling are absolutely vital to the success of this method. It is therefore important to pay particular attention to baseline distortion and to correct for it when necessary. It is also important to collect a sufficient number of scans to get adequate (≈ 50:1) signal intensity so as to ensure accurate integration. Scaling may be done most conveniently by looking for a well-resolved single peak (such as isolated α-proton, a distinctive up-field methyl group, a single histidine resonance, or an isolated amide) and adjusting the integral scale accordingly. Alternatively, for very large proteins, (where single resonances are almost impossible to identify) it is possible to scale quite accurately by integrating over the whole amide region (6.0–11.0 ppm) and equating this measured value to the total number of amide and aromatic protons known to be in the protein. This 'total downfield ^{1}H content', N_{down}, is given by:

$$N_{down}=\#res + (2\times\#N) + (2\times\#Q) + (2\times\#H) + (4\times\#Y) + (5\times\#F) + (6\times\#W)$$

where #res is the total number of residues (excluding prolines), and #N, #Q, #H etc. represent the number of amide or aromatic ring-bearing amino acids found in the protein (using the standard single letter code for amino acids).

To summarize, the step-by-step protocol for quantifying secondary structure content via one-dimensional ^{1}H NMR is as follows.

1. Prepare two identical samples of the protein of interest. One sample should be dissolved in a defined volume of 90% H_2O/10% D_2O, the second should be dissolved in an identical volume of 99.9% D_2O. Both samples must be at equal concentrations. After adjusting the pH (which should be between pH 3.0–6.0), collect standard one-dimensional ^{1}H spectra of both samples using identical collection conditions.

2. After performing the necessary baseline corrections and the required scaling, use direct integration to determine the number of resonances between 8.20–9.00 ppm for the H_2O sample. Call this value <C>. (Note that since the spectra were collected in 90% H_2O, this integral will actually represent 90% of the total number of resonances between 8.20–9.00 ppm and should therefore be equal to the number of residues in the coil configuration.)

3. After performing the necessary baseline corrections and the required scaling, use direct integration to determine the number of resonances between 4.85–5.90 ppm for the D_2O sample. Call this value <B>.

4. The number of residues in the coil configuration is given by:

$$\#coil=<C>$$

The number of residues in the β-strand configuration is given by:

$$\#\beta=2.0\times<B>$$

The number of residues in the α-helix configuration is given by:

$$\#\alpha=\#residues-\#\beta-\#coil$$

The actual portion of secondary structure is given as:

$$\%coil = \frac{\#coil}{\#residues} \quad \%\beta = \frac{\#\beta}{\#residues} \quad \%\alpha = \frac{\#\alpha}{\#residues}$$

(Where #residues is the total number of residues in the protein or peptide.)

There are several caveats and cautions that bear mentioning, these are as follows: (i) If one finds that #α<0, then set #α=0. Calculate the percentage of β-strand and coil using the formule %β=#β/(#β+#coil) and %coil=#coil/(#β+#coil); (ii) If one finds there are no peaks between 4.85–5.90 ppm (i.e.<B>=0) then set #β=0; (iii) Due to the statistical and empirical nature of this method, peptides with fewer than 40 residues are likely to give spurious results.

2.2. *Method 2*

This protocol was specifically designed for the determination of protein secondary structure content through regional integration (peak counting) in the 'fingerprint' region of two-dimensional COSY or DQF-COSY ^{1}H NMR spectra. While generally limited to smaller proteins, Method 2 does offer several advantages over Method 1. For example, rather than having to collect spectra from two identical samples dissolved in different solvents, it is possible to get the same kind of information using just a single spectrum collected with the sample dissolved in a single solvent (H_2O). Furthermore, rather than having to conduct very careful integrations to get the necessary peak intensities it is possible to simply count the visible COSY peaks to get the same kind of information. Additionally, because the fingerprint region actually separates α-protons from β-protons (which is not done in one-dimensional spectra) it is possible to make use of a much better method to calculate the α-helix content. This helix content measurement, which is also empirically derived, is based on the observation that the number of helical residues in a protein is proportional to the number of α-protons (excluding glycines) between 3.40–4.10 ppm [1].

Fig. 2 illustrates the central idea behind Method 2. Each shaded region represents the portion of the spectrum which has been found to correlate with the content of α-helices, β-strands and coils. Counting peaks in each of the differently shaded regions (note that some peaks are actually counted twice) will give one a rather accurate assessment of the secondary structure content of the protein.

For a more specific outline of how the procedure should be conducted, we present the following itemized protocol.

1. With the protein of interest dissolved in H_2O (ideally at pH <6.0 and a temperature >25°C), collect a 2-D COSY or DQF-COSY ^{1}H NMR spectrum.

2. Through direct peak counting in the fingerprint region determine the number of resonances found between 8.20–9.00 ppm. The number of peaks in this spectral region is proportional to the number of residues found in the coil conformation. Call this value <C>.

3. Through direct peak counting in the fingerprint region determine the number of resonances found between 4.85–5.90 ppm. The number of peaks in this spectral region is proportional to the number of residues found in the β-strand conformation. Call this value <B>.

4. Through direct peak counting in the fingerprint region determine the number of resonances found between 3.40–4.10 ppm. The number of peaks in this spectral region (after subtracting out the expected number of glycine peaks) is proportional to the number of residues found in the α-helix conformation. Call this value <A>.

5. An estimate of the proportion of residues in the coil configuration is given by:

$$<\#coil> = 0.9\ <C>$$

An estimate of the proportion of residues in β-strands is given by:

$$<\#\beta> = 2.0\ \{<B>\}$$

An estimate of the proportion of residues in the α helix configuration is given by:

$$<\#\alpha> = 2.0\ \{<A>-2.0\ (\#glycine)\}$$

An estimate of the total number of residues is given by:

$$TOTAL = <\#coil> + <\#\beta> + <\#\alpha>$$

The actual portion of secondary structure is given as:

$$\%coil = \frac{\#coil}{TOTAL} \quad \%\beta = \frac{\#\beta}{TOTAL} \quad \%\alpha = \frac{\#\alpha}{TOTAL}$$

(Where #residues is the total number of residues and #glycine is the number of glycines known to be in the peptide or protein.)

As before, there are a few other caveats and cautions that bear further mention: (i) It has been found that the results could be slightly

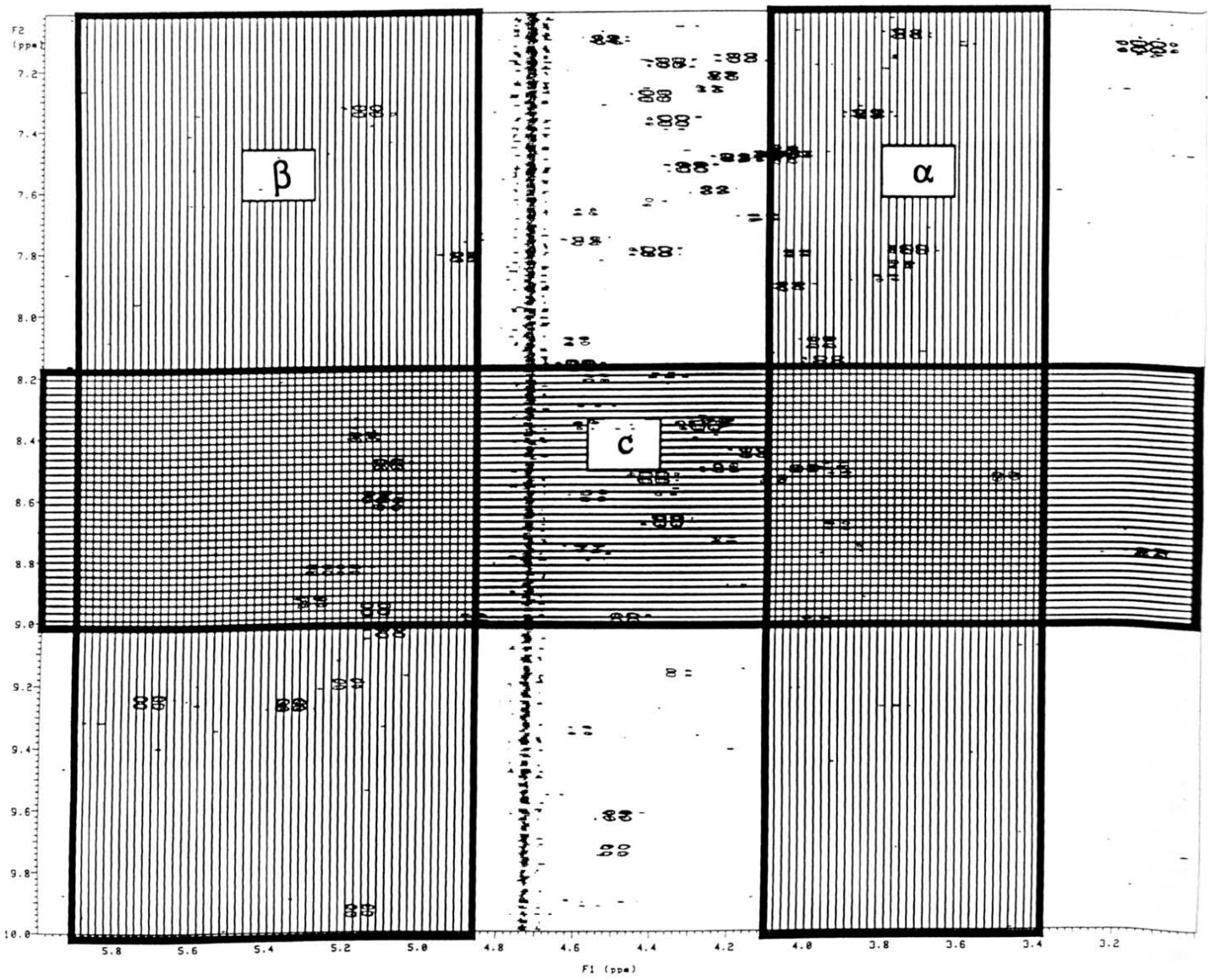

Fig. 2. Fingerprint region of a DQF-COSY spectrum of *E. coli* thioredoxin collected at 30°C in 90% H_2O/10% D_2O. The shaded areas correspond to the regions of the spectrum which contain important structural information. The two hatched, vertical blocks marked by **β** and **α** correspond to the areas used to estimate the β-strand and α-helix content, respectively. The hatched transverse region marked by **c** is used to estimate the coil content. Note that some peaks appear in overlapping regions and hence are counted twice when making secondary structure estimates.

improved if the estimate of the number of β-strand residues is scaled according to the peptide size using the following equation:

$$<\#\beta> = 2.0\ \{<B>-0.05\ (\#\text{residues} - 60)\}$$

This seems to correct for the preponderance of cysteine and glycine residues found in very small proteins (which tend to have low-valued, upfield chemical shifts). It is not clear if this correction generally holds for much larger proteins (> 15 kDa). (ii) If, in using Method 2, one finds there are no peaks between 4.85-5.90 ppm (i.e. <B>=0) then set #β=0. (iii) The procedure works best for proteins where assignments have been carried out between pH 3.0-7.0 and temperatures between 20-50°C. (iv) Method 2 appears to be generally applicable to peptides and proteins with molecular weights <20 kDa, however, due to the statistical basis and empirical nature of the technique, peptides of fewer than 40 residues sometimes give spurious results.

3. RESULTS

3.1. *Method 1*

As a general check of the utility and applicability of Method 1, the above procedure was applied to 6 proteins of widely differing sizes and folding classifications: rat parvalbumin (109 residues), ribonuclease A (124 residues), HEW lysozyme (129 residues), troponin C (162 residues), soybean trypsin inhibitor (181 residues) and carbonic anhydrase (256 residues). It is interesting to note that under the conditions used to collect the troponin C spectrum, this protein was almost completely dimerized so the 'effective' number of residues was probably closer to 324. Fig. 1 illustrates a typical result, with

scaled integrals included, as obtained for rat parvalbumin.

Table II compares the estimates of secondary structure content obtained by Method 1 with those obtained by crystallography for these 6 proteins. With the exception of soybean trypsin inhibitor, the agreement is rather good. There is a slight tendency to underestimate the coil content in larger proteins and this may be due to the level of saturation transfer that takes place during the presaturation of the H_2O resonance. This presaturation leads to generally reduced amide intensities and hence to a smaller value for the integrated area under the 8.20–9.00 ppm envelope. Perhaps a more refined solvent suppression technique or, alternatively, a more refined selective excitation protocol would help in these mtters.

The success of this technique for proteins ranging in size up to 324 residues indicates that the method has the potential to be applicable to proteins of almost all sizes. However, given the small size of the experimental sample, it was felt that additional data would have to be accumulated before we could be confident of the method's general applicability.

As a further check of this method, idealized data of integrated peak areas were generated using the chemical shift assignments of 'solved' proteins which had been previously published in the literature. Note, however, that no attempt was made to correct for possible linewidth differences within this 'simulated' data set. These values were then used to calculate the percent content of secondary structure employing the same methods described above. This permitted an effective comparison to be conducted on more than 50 different proteins. The results of this tabulation are presented in Table III (column 2). It can be seen that the agreement to previous structural assignments is quite good for nearly all of the listed proteins. This result, in combination with the actual experimental findings given in Table II suggests that this technique is, indeed, generally applicable. Further refinements in the methodology and additional studies on other proteins should help to improve its accuracy in the future.

3.2. *Method 2*

In Table III we also list the secondary structure estimates made using Method 2. These estimates may be compared with those obtained from X-ray crystallography which are presented in column 1. With the exception of few very small proteins, the agreement is rather striking. It is important to note, however, that most of the NMR data used in preparing this table was taken from published sequential assignments and, as a result, the data are somewhat idealized. Application of these procedures to actual COSY spectra from published sources, as well as those obtained locally, was also done in order to check the general validity of the method under simulated 'experimental' circumstances. Very little disagreement was found. Clearly, however, it is important to collect high quality COSY spectra with good signal to noise ratio so as to be as certain as possible that nearly all expected resonances are visible in the fingerprint region.

3.3. *Comparison to CD, IR and Raman*

It was of some interest to compare the methods just described to other techniques which have traditionally been used in secondary structure quantification, notably CD, IR and infrared Raman spectroscopy. Table IV presents a comparison between the secondary structure content as determined by Method 2 and the secondary structure content as determined by CD. The CD values were obtained from a variety of published sources while the X-ray values were obtained principally from [66,67,69]. If one compares the two sets of values with those obtained crystallographically, it is clear that Method 2 is significantly more accurate. A more limited comparison with IR and Raman data yielded similar results.

While the success and simplicity of Method 2 is obvious, it is important to note that this procedure typically requires an order of magnitude more material than does CD or other comparable forms of spectroscopy. This should not be overlooked when sample quantity or sample concentration is an issue.

4. DISCUSSION

The procedures outlined above provide two simple techniques for the rapid determination of secondary structure content in proteins. All that is required is either a regular 1-D or 2-D COSY spectrum of a protein dissolved in H_2O (or D_2O). Using simple peak counting (2-D COSY) or peak integration (1-D spectra) over a few selected regions of the spectrum it is quite possible to determine secondary structure content with an accuracy that equals or exceeds that of such well established techniques as CD, IR and Raman spectroscopy.

Table II

Comparison of secondary structure content predictions from Method 1 with those obtained by X-ray crystallography

Protein name (# of residues)	X-ray[a]			Method 1		
	%α	%β	%c	%α	%β	%c
Rat parvalbumin (109)	64	6	30	64	7	28
Ribonuclease A (124)	22	48	30	16	44	39
HEW Lysozyme (129)	45	19	36	49	23	28
Troponin C (162)	69	7	24	65	9	26
Soybean Trypsin Inhibitor (181)	53	3	44	38	20	43
Carbonic Anhydrase (256)	41	13	46	40	23	37

%α=% α-helix; %β=% β-strand; %c=% 'coil'

[a]X-ray values were obtained from the PDB [66] and other published sources [67,68]

Table III

Comparison of predictions from Method 1 and Method 2 to previously reported X-ray crystallographic and/or NMR measurements of secondary structure

Protein name (#residues)	X-ray/NMR[a]			Method 1[b]			Method 2		
	%α	%β	%c	%α	%β	%c	%α	%β	%c
Acyl Carrier Protein (77)	61	0	39	67	5	28	68	5	27
Algal Plastocyanin (97)	8	59	33	5	54	41	6	53	41
α Bungarotoxin (74)	0	51	49	2	59	39	0	61	39
α Neurotoxin (62)	0	58	42	0	62	38	0	62	38
α Purothionin (45)	35	22	43	40	22	38	31	26	43
Anaphylotoxin C3a (77)	69	0	31	66	3	31	66	3	31
Anaphylotoxin C5a (75)	69	0	31	55	0	45	56	0	44
BDS I (43)	0	56	44	12	42	46	0	47	53
Antennapedia (68)	66	0	34	62	0	38	71	0	29
Anthopleurin A (49)	0	49	51	18	49	33	0	60	40
Apamin (18)	50	0	50	72	0	28	62	0	38
Arc Repressor (53)	53	17	30	45	17	38	55	14	31
Ascaris Trypsin Inh. (62)	0	55	45	16	48	36	7	54	39
Bovine Trypsin Inh. (58)	26	36	38	34	38	28	24	44	32
Bull Seminal Inh. (57)	19	42	39	37	25	38*	18	32	50
Calbindin (76)	61	8	31	55	16	29	63	13	24
Calmodulin (148)	62	8	30	61	9	30	58	10	32
CTX II (60)	0	62	38	0	63	37	0	63	37
Cecropin A (37)	84	0	16	73	11	16	81	8	11
Cellobiohydrolase I (36)	0	45	55	25	33	42*	0	44	56
Cro Repressor (66)	41	30	29	42	27	31	44	26	30
Cytochrome B5 (82)	46	27	27	37	30	33	40	29	31
Cytochrome C (104)	49	6	45	61	6	33	57	7	36
Cytochrome C551 (82)	68	0	32	73	0	27	70	0	30
Cytochrome C2 (116)	51	5	44	61	3	36	56	4	40
Ner Protein (69)	71	0	29	66	11	23	57	14	29*
E. coli Thioredoxin (108)	40	31	29	44	26	30	38	29	33
Elgin C (69)	22	42	36	23	42	35	7	51	42*
Flavodoxin (137)	42	35	23	50	28	22	37	35	28
HEW Lysozyme (129)	45	19	36	47	23	30	35	29	36
Human Lysozyme (130)	43	17	40	45	18	37	54	15	31
Hirudin (65)	0	38	62	34	22	44*	0	32	68
Histidine Protein (85)	41	36	23	28	42	30	37	37	26
Histone H5 (77)	48	10	42	38	16	46	50	13	37
Human Thioredoxin (105)	42	34	24	33	36	31	41	32	27
Insulin (51)	58	10	32	67	12	21	67	12	21
Interleukin 1β (153)	0	60	40	0	58	42	14	50	36
Interleukin 8 (72)	21	42	27	18	47	35	25	43	32
Lac Repressor (51)	64	0	36	61	0	39	64	0	36
Leucine Zipper GCN4 (33)	96	0	4	61	0	39	71	0	29
T4 Lysozyme (164)	64	13	23	64	6	30	60	6	34
NP-5 Peptide (33)	0	76	24	0	64	36	0	64	36
Ovomucoid (56)	23	36	41	34	21	45	5	31	64*
Parvalbumin (108)	64	6	30	75	4	21	67	5	28
Procarboxypeptidase (81)	27	35	38	30	27	43	26	29	45
Ribosomal EL-30 Prot. (53)	31	33	36	31	29	30	23	39	38
PTI Type E (59)	22	36	42	20	20	60*	33	33	34
Ribonuclease A (124)	22	48	30	15	52	33	19	51	31
Ribonuclease T1 (104)	14	44	42	20	44	36	25	41	34
Squash Trypsin Inh. (28)	0	64	36	0	56	44	0	56	44
Staph. Nuclease (136)	30	38	32	33	38	29	36	36	28
Tendamistat (74)	0	69	31	7	49	44	0	52	48
Ubiquitin (76)	26	41	33	16	45	39	16	45	39

[a]Values obtained from X-ray crystallographic measurements, well-defined NMR structures or from previously published reports on secondary structure content measured by standard 2-D NMR methods

[b]Data for this column was obtained using 'simulated' 1-D NMR data prepared from assignments previously reported in the literature

*Indicates an unusually poor prediction

%α=% α-helix; %β=% β-strand; %c=% 'coil'

Table IV

Comparison of predictions from Method 2 to CD results

Protein name (#residues)	Actual[a]			Method 2			CD			Reference[b]
	%α	%β	%c	%α	%β	%c	%α	%β	%c	
Acyl Carrier Protein (77)	61	0	39	68	5	27	33	0	67	[61]
α Purothionin (45)	35	22	43	31	26	43	54	20	26	[62]
Anaphylotoxin C3a (77)	69	0	31	66	3	31	43	0	57	[63]
Anaphylotoxin C5a (75)	69	0	31	56	0	44	43	0	57	[63]
Bovine Trypsin Inh. (58)	26	36	38	24	44	32	21	28	51	[5]
Calmodulin (148)	62	8	30	58	10	32	61	4	35	[4]
Cytochrome C (104)	49	6	45	57	7	36	33	9	58	[5]
E. coli Thioredoxin (108)	40	31	29	38	29	33	37	19	44	[64]
HEW Lysozyme (129)	45	19	36	35	29	36	45	21	34	[5]
Human Thioredoxin (105)	42	34	24	41	32	27	49	35	16	[54]
Insulin (51)	58	10	32	67	12	21	49	23	27	[5]
Parvalbumin (108)	64	6	30	67	5	28	58	0	42	[5]
Ribonuclease A (124)	22	48	30	19	51	31	25	37	38	[5]
Staph. Nuclease (136)	30	38	32	36	36	28	32	25	43	[5]

[a]Values in the first column represent those taken from X-ray crystal structures or published NMR structures. See [1,66–68] for more details or specific assignments

[b]References to CD measurements

%α=% α-helix; % β-strand: %c=% 'coil'

Furthermore, unlike the complex manipulation and data 'massaging' that is generally required for the interpretation of IR, Raman and CD data, this new technique requires little more than simple peak counting: no curve fitting, second derivitization, deconvolution or transformation is needed. Furthermore, the procedure is not significantly affected by the 'basis set' or parameterization problems that plague such techniques as CD or IR. Instead, this NMR-based procedure is simple, relatively unparameterized and fast. In fact, with a bit of practice it is quite possible (once the spectra are collected) to determine the secondary structure content of a protein in less than a few minutes.

Although high-resolution NMR has reached a stage such that is now possible to determine the atomic structure of a small protein to a great degree of accuracy, the procedure often requires many weeks, if not months, to do so. The possibility that NMR may now be used for a very quick determination or verification of secondary structure opens up a number of potential and highly useful applications. For instance, it is not hard to see how this procedure could be used in preliminary characterization of proteins for which detailed structural studies are about to begin. It is also not difficult to see how this technique could be adapted or extended to studies of titrated conformational changes in proteins (folding, unfolding etc.). Such work could be done either in the fast or slow exchange limit, although the slow exchange studies may prove to be more technically demanding.

It is also worth noting that the same integration or peak-picking procedure for secondary structure determination can be applied to other nuclei as well. Preliminary work with ^{13}C carbonyl resonance assignments and/or one-dimensional ^{13}C spectra for BPTI [69], calmodulin [21] and *E. coli* thioredoxin has yielded some promising results. In particular, it has been found that the number of ^{13}C carbonyl resonances found below 172.0 ppm is proportional to the number of residues in β-strand configurations. In a like manner, the number of resonances found above 176.0 ppm is proportional to the number of residues in helical configurations. If a proper accounting of side chain carbonyl or carboxyl resonances is done, this procedure appears to be at least as accurate in secondary structure determination as the methods previously described for protons. It will be of some interest to explore this approach in greater detail as more ^{13}C spectroscopy is done in the future.

To conclude, we have demonstrated that it is possible to use NMR spectroscopy to quantify protein secondary structure content with the same ease and swiftness as can be done with more popular or more conventional techniques such as CD, IR or Raman spectroscopy. It is hoped that this application of NMR spectroscopy might be adopted by others, both specialist and non-specialist alike, as a useful and very simple method for characterizing protein structures and structural transitions in solution.

Acknowledgements: The authors wish to thank Colleen Wilson for her help in preparation of the manuscript. Financial support by the MRC Group in Protein Structure and Function and by the National Institutes of Health (NIH Grant GM-22778) is gratefully acknowledged.

REFERENCES

[1] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) J. Mol. Biol. (in press).
[2] Byler, D.M. and Susi, H. (1986) Biopolymers 25, 469–487.
[3] Hennessey, Jr. J.P. and Johnson, Jr. W.C. (1981) Biochemistry 20, 1085–1094.
[4] Johnson, C.W. (1990) Proteins: Struct. Funct. Gen. 7, 205–215.
[5] Provencher, S.W. and Glockner, J. (1981) Biochemistry 20, 33–37.
[6] Szilagyi, L. and Jardetzky, O. (1989) J. Magn. Reson. 83, 441–449.
[7] Wüthrich, K. (1986) NMR of Protein and Nucleic Acids, Wiley, New York.
[8] Holak, T.A. and Prestegard, J.H. (1986) Biochemistry 25, 5766–5774.
[9] Basus, V.J., Billeter, M., Love, R.A., Stroud, R.M. and Kuntz, I.D. (1988) Biochemistry 27, 2763–2768.
[10] Labhardt, A.M., Hunziker-Kwik, E.H. and Wuthrich, K. (1988) Eur. J. Biochem. 177, 295–305.
[11] Clore, G.M., Sukumaran, D.K., Gronenborn, Teeter, M.M., Whitlow, M. and Jones, B.L. (1987) J. Mol. Biol. 193, 571–578.
[12] Chazin, W.J., Hugli, T.E. and Wright, P.E. (1988) Biochemistry 27, 9139–9148.
[13] Zuiderwig, E.R.P., Mollison, K.W., Henkin, J. and Carter, G.W. (1988) Biochemistry 27, 3568–3580.
[14] Driscoll, P.C., Clore, G.M., Beress, L. and Gronenborn, A.M. (1989) Biochemistry 28, 2178–2187.
[15] Billeter, M., Qian, Y.O., Otting, G., Muller, M., Gehring, W.J. and Wuthrich, K. (1990) J. Mol. Biol. 214, 183–197.
[16] Mabbutt, B.C. and Norton, R.S. (1990) Eur. J. Biochem. 187, 555–563.
[17] Wemmer, D. and Kallenbach, N.R. (1983) Biochemistry 22, 1901–1906.
[18] Breg, J.N., Boelens, R., George, A.V.E. and Kaptein, R. (1989) Biochemistry 28, 9826–9833.
[19] Strop, P., Wider, G. and Wuthrich, K. (1983) J. Mol. Biol. 166, 641–651.
[20] Drakenberg, T., Hofmann, T. and Chazin, W.J. (1989) Biochemistry 28, 5946–5954.
[21] Ikura, M., Kay, L.E. and Bax, A. (1990) Biochemistry 29, 4659–4667.
[22] Otting, G., Steinmetz, W.E., Bougis, P.E., Rochat, H. and Wuthrich, K. (1987) Eur. J. Biochem. 168, 609–620.
[23] Holak, T.A., Engstrom, A., Kraulis, P.J., Lindeberg, G., Bennich, H., Jones, T.A., Gronenborn, A.M. and Clore, G.M. (1988) Biochemistry 27, 7620–7629.
[24] Kraulis, P.J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J. and Gronenborn, A.M. (1989) Biochemistry 28, 7241–7257.
[25] Weber, P.L., Wemmer, D.E. and Reid, B.R. (1985) Biochemistry 24, 4553–4562.
[26] Guiles, R.D., Altman, J., Kuntz, I.D. and Waskell, L. (1990) Biochemistry 29, 1276–1289.
[27] Wand, A.J., DiStefano, D.L., Feng, Y., Roder, H. and Englander, S.W. (1989) Biochemistry 28, 186–196.
[28] Gooley, P.R. and Mackenzie, N.E. (1988) Biochemistry 27, 4032–4040.
[29] Chau, M.H., Cai, M.L. and Timkovich, R. (1990) Biochemistry 29, 5076–5087.
[30] Hyberts, S.G. and Wagner, G. (1990) Biochemistry 29, 1465–1474.
[31] Van Mierlo, C.P.M., Vervoort, J., Muller, F. and Bacher, A. (1990) Eur. J. Biochem. 187, 521–541.
[32] Sukumaran, D.K., Clore, G.M., Preus, A., Zarbock, J. and Gronenborn, A.M. (1987) Biochemistry 26, 333–337.
[33] Klevit, R.E., Drobny, G.P. and Waygood, E.B. (1986) Biochemistry 25, 7760–7769.
[34] Zarbock, J., Clore, G.M. and Gronenborn, A.M. (1986) Proc. Natl. Acad. Sci. USA 83, 7628–7632.
[35] Kline, A.D. and Justice Jr., R.M. (1990) Biochemistry 29, 2906–2913.
[36] Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T. and Gronenborn, A.M. (1990) Biochemistry 29, 3542–3556.
[37] Clore, G.M., Appella, E., Yamada, M., Matsushima, J. and Gronenborn, A.M. (1989) J. Biol. Chem. 264, 18907–18911.
[38] Zuiderweg, E.R.P., Kaptein, R. and Wuthrick, K. (1983) Eur. J. Biochem. 137, 279–292.
[39] Oas, J.G., McIntosh, L.P., O'Shea, E.K., Dahlquist, F.W. and Kim, P.S. (1990) Biochemistry 29, 2891–2894.
[40] Redfield, C. and Dobson, C.M. (1988) Biochemistry 27, 122–136.
[41] Redfield, C. and Dobson, C.M. (1990) Biochemistry 29, 7201–7214.
[42] McIntosh, L.P., Wand, A.J., Lowry, D.E., Redfield, A.G. and Dahlquist, F.W. (1990) Biochemistry 29, 6341–6362.
[43] Gronenborn, A.M., Wingfield, P.T. and Clore, G.M. (1989) Biochemistry 28, 5081–5089.
[44] Bach, A.C., Selsted, M.E. and Pardi, A. (1987) Biochemistry 26, 4389–4397.
[45] Robertson, A.D., Westler, W.M. and Markely, J.L. (1988) Biochemistry 27, 2519–2529.
[46] Padilla, A., Cave, A. and Parello, J. (1988) J. Mol. Biol. 204, 995–1017.
[47] Moore, J.M., Chazin, W.J., Powls, R. and Wright, P.E. (1988) Biochemistry 27, 7806–7816.
[48] Vendrell, J., Wider, G., Aviles, F.X. and Wuthrich, K. (1990) Biochemistry 29, 7515–7522.
[49] Van de Ven, F.J.M. and Hilberts, C.W. (1986) J. Mol. Biol. 192, 389–440.
[50] Robertson, A.D., Purisima, E.O., Eastman, M.A. and Scheraga, H.A. (1989) Biochemistry 28, 5930–5938.
[51] Hoffman, E. and Ruterjans, H. (1988) Eur. J. Biochem. 177, 539 560.
[52] Torchia, D.A., Sparks, S.W. and Bax, A. (1989) Biochemistry 28, 5509–5524.
[53] Kline, A.D. and Wuthrich, K. (1986) J. Mol. Biol. 192, 869–890.
[54] Forman-Kay, J.D., Clore, G.M., Driscoll, P.C., Wingfield, P.T., Richards, F.M. and Gronenborn, A.M. (1989) Biochemistry 28, 7088–7097.
[55] Dyson, H.J., Holmgren, A. and Wright, P.E. (1989) Biochemistry 28, 7074–7087.
[56] Gronenborn, A.M., Nilges, M., Peanasky, R.J. and Clore, G.M. (1990) Biochemistry 29, 183–189.
[57] Wagner, G. and Wuthrich, K. (1982) J. Mol. Biol. 155, 347–366.
[58] Arseniev, A.S., Wider, G., Joubert, F.J. and Wuthrich, K. (1982) J. Mol. Biol. 159, 323–335.
[59] Heitz, A., Chiche, L., Le-Nguyen, D. and Castro, B. (1989) Biochemistry 28, 2392–2398.
[60] Weber, P.L., Brown, S.C. and Mueller, L. (1987) Biochemistry 26, 7282–7290.
[61] Schulz, H. (1985) J. Biol. Chem. 250, 2293–2298.
[62] Teeter, M.M. and Whitlow, M. (1988) Proteins: Struct. Funct. Gen. 4, 261–271.
[63] Hugli, T.E., Morgan, W.T. and Muller-Eberhard, H.J. (1975) J. Biol. Chem. 250, 1479–1485.
[64] Reutimann, H., Straub, B., Luisi, P.L. and Holmgren, A. (1981) J. Biol. Chem. 256, 6796–6802.
[65] Pastore, A. and Saudek, V. (1990) J. Magn. Reson. 90, 165–176.
[66] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M.J. (1977) J. Mol. Biol. 112, 535–542.
[67] Levit, M. and Greer, J. (1977) J. Mol. Biol. 114, 181–293.
[68] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[69] Tuchsen, E. and Hanson, P.E. (1988) Biochemistry 27, 8568–8576.