Journal of Biomolecular NMR, 1 (1991) 457–471
ESCOM

J-Bio NMR 033

The value of chemical shift parameters in the description of protein solution structures

Yuan Gao, Nigel C. Veitch and Robert J.P. Williams*

Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, U.K.

Dedicated to the memory of Professor V.F. Bystrov

Received 27 May 1991
Accepted 31 July 1991

Keywords: Chemical shift; NOE; Protein structure; Mobility; Pseudocontact shift; Cytochrome b_5; Cytochrome *c*

SUMMARY

An increasing number of protein solution structures, calculated on the basis of nuclear Overhauser enhancement cross-peak intensities observed in two- or higher dimensional NOESY experiments, are becoming available. Among these structures regions of uncertainty are frequently observed particularly with respect to loops and surface side chains. These are commonly ascribed to either a lack of NOE constraints or to some intrinsic mobility within the protein. A powerful method of structural analysis which may resolve this problem is based on the information content of the chemical shift. The value of such an analysis is illustrated here with cytochromes b_5 and *c*, proteins for which high-quality crystallographic and NMR data are available. Comparison of these using a pseudocontact shift-based analysis indicates that NOE data should be combined with the chemical shift data in order to uncover fully the ensemble of protein states and their dynamics in solution.

INTRODUCTION

The analysis of protein structures in solution using nuclear magnetic resonance spectroscopy is based on the interpretation of relaxation effects, principally the nuclear Overhauser enhancement (NOE), and has led to excellent definition of the main-chain fold of many small proteins. The validity of the method at this level has been confirmed in a number of cases by detailed comparisons between X-ray diffraction and NMR-derived structures indicating that the latter is a well-authenticated method for determining the fold of a protein in solution. Recent increases in the sophistication of the experimental techniques has led to an ability to analyse protein structures of up to

* To whom correspondence should be addressed.

20 kDa. There are, however, regions of uncertainty in many of these structures and this applies to structures obtained by diffraction techniques to some degree but not so noticeably as to those obtained by NMR. The uncertainty in NMR-derived structures tends to be associated with side chains, both internal and external, though particularly the latter, and also with the definition of surface loops as the examples in Table I indicate. It may be that these regions are intrinsically mobile in solution, in which case it is to be expected that NMR studies in solution and diffraction studies in crystals will give a different impression of the structure and its mobility. There is good reason to suppose that energy minimization calculations applied to both types of structure will help to resolve some of these problems although they do not provide strict tests of the meaningfulness of the original observations, since it is known that despite very considerable effort the attempt to deduce protein structures directly from their primary sequence using energy calculations has so far failed. The central problem therefore lies in assessing experimentally the real intrinsic mobility within a given structure. Such knowledge is critical to an understanding of protein function as it is now well recognised that mobility within structure is the basis of activity (Williams, 1989).

Another approach to the definition of a protein structure in solution is to develop the analysis

TABLE I
REGIONS OF UNCERTAINTY IN SOME NMR SOLUTION STRUCTURES

Protein	Number of residues	Comments
Human TGF_{α}[a]	50	N- and C-termini vary considerably; loop from residues 8-15 shows great flexibility
BSPI-II[b]	64	Atom rms distribution 50% larger for side-chain than main-chain atoms; particularly marked for surface side chains especially in reactive site loop
Tendamistat[c] (*Streptomyces tendae*)	74	Residues 1-4 (N-terminus) virtually unrestrained by the NMR data
Plastocyanin[d] (*Scenedesmus obliquus*)	97	Connecting loops relatively poorly defined due partly to small number of constraints in these regions; Tyr^{68} not constrained by NOEs and adopts widely differing conformations
Acylphosphatase[e]	98	Surface loops and N-terminal segment less well defined due to few NOE constraints arising from either not enough contacts with the rest of the molecule or mobile regions fluctuating among several conformations
Interleukin 1β[f]	153	Some disorder noted in N- and C-termini; largest deviations occur in loop regions due to hinge motions involving one or two residues. Surface side chains partially disordered

[a] Kline et al., 1990.
[b] Clore et al., 1987.
[c] Kline et al., 1988.
[d] Moore et al., 1988.
[e] Saudek et al., 1989.
[f] Clore et al., 1991.

of the frequency dispersion of the chemical shift rather than only the relaxation rates from NMR measurements. A procedure for this type of analysis has long been used in the context of small molecule structures (Some of the very earliest structure methods used combinations of ring current shifts and NOE data). In essence a computer search programme can be devised to find a structure which takes into account shifts generated by diamagnetic centres e.g. aromatic rings or carbonyl groups, and/or by paramagnetic centres e.g. iron porphyrin systems or lanthanides. One such structure search programme was published under the name of Burlesque (Barry et al., 1971). Structures derived by this procedure included those of mononucleotides, penicillins and ribonucleotides (using lanthanide ions as probes) (Barry et al., 1971; Dobson et al., 1975,1978), porphyrin complexes with 2,4,7-trinitrofluorenone and caffeine (using the porphyrin as probe) (Barry et al., 1973a,b) and the indole–3-propionic acid complex with horseradish peroxidase (using heme iron as probe) (Burns et al., 1975). It should be also noted that it is always advantageous to include in shift data analysis coupling constants where these can be obtained and indeed they are usually included today in protein structure determinations (Smith et al., 1991) using relaxation (NOE) much as they were used earlier in small molecule studies. Professor Vladimir Bystrov, to whose memory this article is dedicated was an outstanding exponent of the use of such coupling constant methods.

A third method of structure determination which again uses paramagnetic probes is the use of a paramagnetic centre with a long relaxation time, Gd(III), Mn(II) and spin-labels for example, to give long-range relaxation. The method can be used with any of the other procedures. It is mentioned only in passing here, since it is not used in this paper.

There is a further interest in having these different methods for structure determination in solution. They are all based on different theoretical assumptions so that comparison between them may reveal difficulties peculiar to each. One such difficulty is associated with ensemble averaging which is differently linked to each method. In particular the degree of agreement between structural information derived from relaxation (NOE) and shift data will be analysed in this paper. Starting from certain protein crystal structures of high quality an assessment will be made as to how well NMR data, both NOE and shift, confirm the details of these structures (rather than their general fold) in solution. It is not our purpose therefore to show that the general fold in solution is correct in the cases chosen since this has been unequivocally established already, but rather to examine the degree of local agreement amongst methods, especially relating to the surfaces and side chains of the proteins. It would be extremely valuable for example if as a consequence of such analyses combinations of methods could be used to uncover mobility in solution with confidence.

MATERIALS AND METHODS

The two proteins that will be considered here are bovine liver cytochrome b_5 and iso-1-cytochrome *c* from the yeast *Saccharomyces cerevisiae*. In the case of the latter both wild-type and variant forms will be discussed. The preparation and purification of these proteins has been described in detail elsewhere (Reid and Mauk, 1982; Pielak et al., 1985; Cutler et al., 1987). It is very necessary to have pure proteins since confusion readily arises where deamidated proteins are present as is frequently the case in commercial cytochrome *c* samples (Brautigan et al., 1978). The crystal structure of yeast iso-1-ferrocytochrome *c* has been refined to better than 1.5 Å resolution (Louie and Brayer, 1990) and that of ferricytochrome b_5 to 2.0 Å resolution (Mathews et al., 1979). In the

case of yeast iso-1-cytochromes *c* much additional data has been kindly provided by Dr. G.D. Brayer.

The procedures for the assignment of the NMR spectra of these proteins have already been described in detail and the data confirmed by the work of several independent groups (Feng et al., 1989; Gao et al., 1990; Guiles et al., 1990; Veitch et al., 1990). The analysis of the NOE relaxation data given below is based on the intensity of nuclear Overhauser enhancement cross peaks as recorded in two-dimensional NOESY experiments, checked by examination of cross sections in some cases. Typically the mixing times employed were between 100 and 150 ms. The NOE data collected here have not been used to deduce a solution structure for any of the proteins. The general agreement between the expectation of a measurable NOE (disregarding for the moment its magnitude) from the crystal coordinates and the observed NOE is so good that any alternative folding patterns for the main chain of the protein between solution and crystal states can quickly be discounted. Moreover, the overall agreement between the paramagnetic shift-based structure analyses and the structures of all the cytochromes under study is also so good that confidence in the use of the crystallographically determined fold is very high. The experimental data will be handled in the following manner.

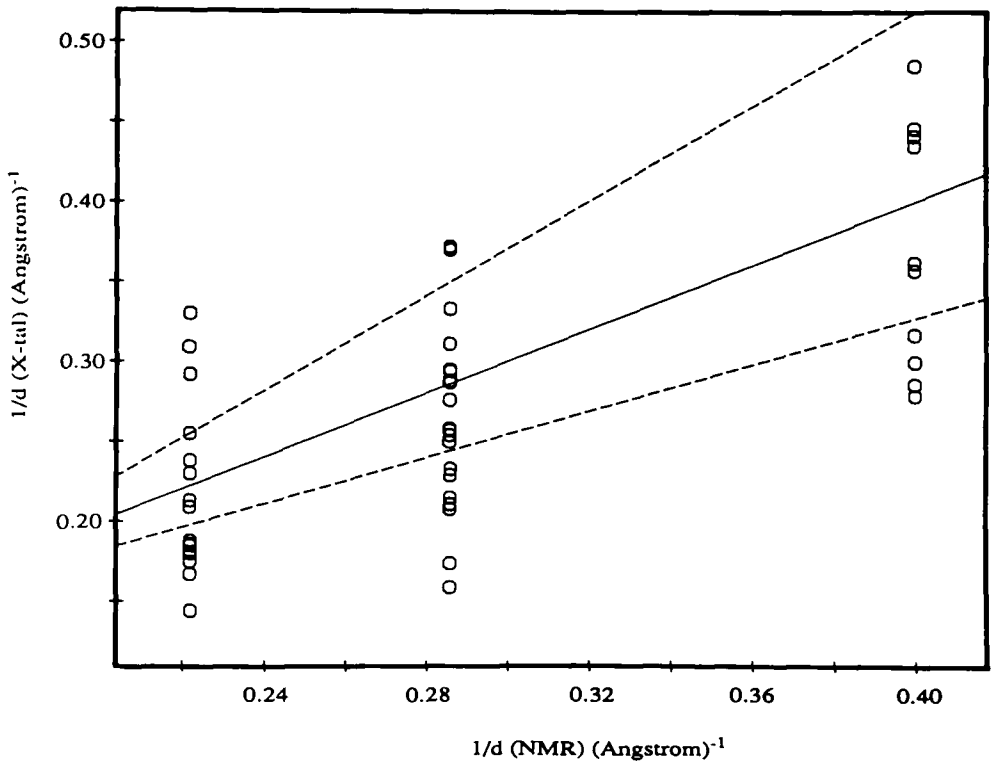

Fig. 1. Plot of proton-proton distances derived from crystal structure data of cytochrome b_5 against analogous distance constraints derived from NOE measurements for the protons of the residue Trp^{22}. The solid line is for reference only while the dashed lines give an approximate indication of the $\pm$ 0.5 Å error range of the NMR data.

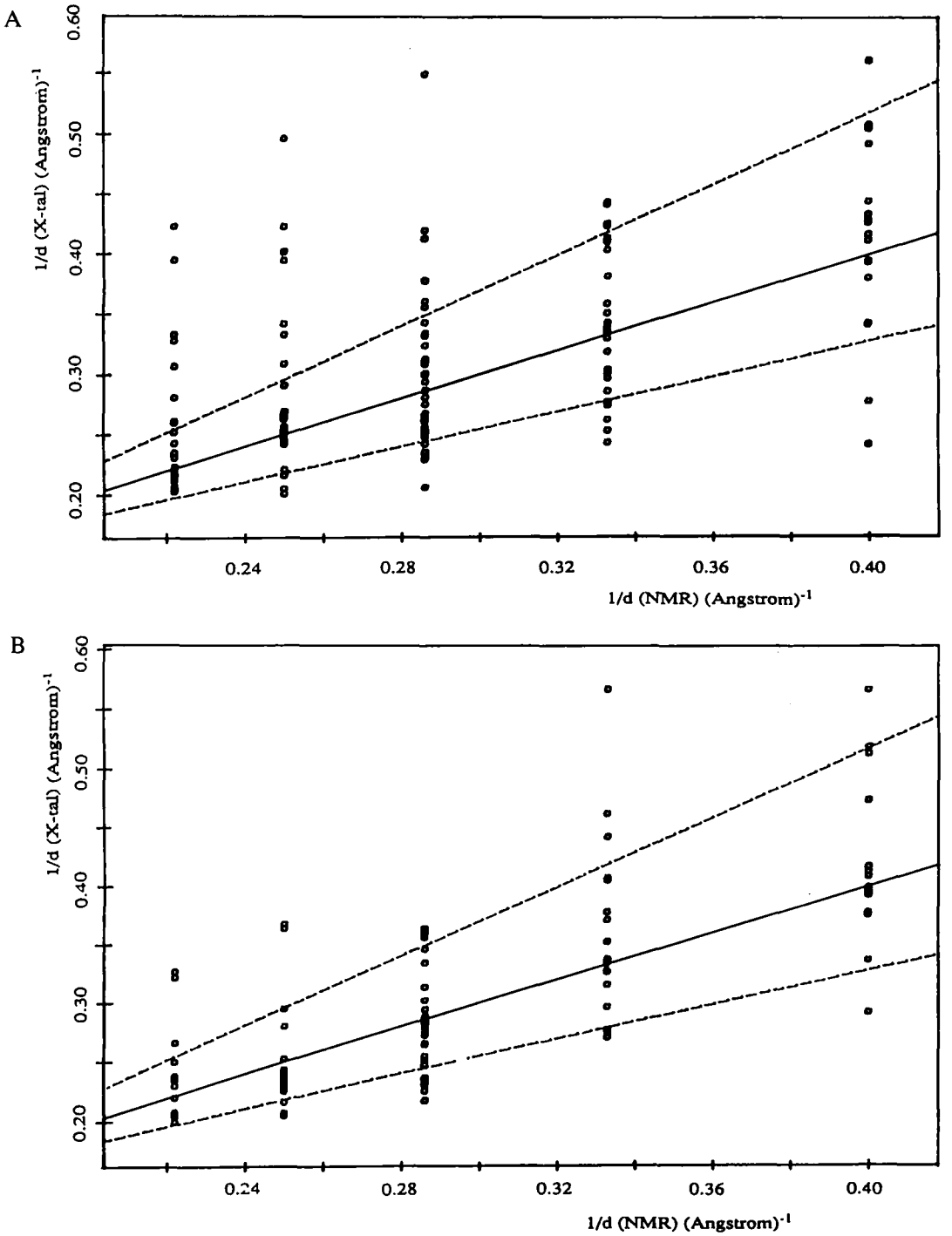

Fig. 2. Plot of proton-proton distances derived from crystal structure data of yeast iso-1-cytochrome *c* against analogous distance constraints derived from NOE measurements. (A) Heme protons, (B) Trp^{59} protons. The solid line is for reference only while the dashed lines give an approximate indication of the $\pm$ 0.5 Å error range of the NMR data.

RESULTS AND DISCUSSION

Nuclear Overhauser enhancements

The intensities of NOE connectivities recorded in two-dimensional experiments have been classified in various ways based on a system of strong, medium and weak cross peaks. These generally refer to distance constraints in structural work as follows: strong implies a distance of 2.5 ± 0.5 Å, medium 3.5 ± 0.5 Å and weak 4.5 ± 0.5 Å. This system is used here in the analysis of cytochrome b_5 spectra. Where the data are of sufficient quality, further distinctions of intensity can be made e.g. very weak, weak, medium, strong, very strong. This system is used here for cytochrome *c* NOE data where the corresponding distances are 2.5, 3.0, 3.5, 4.0 and 4.5 ± 0.25 Å. These NOE intensity data are plotted against observed distances between relevant protons taken from the crystal structures. Figures 1 and 2 show typical data for cytochromes b_5 and *c*, respectively. The scatter observed in these plots is somewhat surprising and it would be of interest to publish such analyses on other proteins for which NOE data and accurate crystal structures are available. The question of why such scatter is observed is of obvious importance in the context of NMR-derived protein structures.

Chemical Shift Analyses

(a) Diamagnetic shifts

The approach to a comparison of crystal structures and structures determined in solution can be illustrated first in a simple case, that of the analysis of the shift of α-CH protons due to neighbouring carbonyl groups. A quantitative analysis is possible in this case (Clayden and Williams, 1982; Dalgarno et al., 1983). The most thorough analysis of these shifts however has been recently carried out on the protein bovine pancreatic trypsin inhibitor. The agreement between theory and experiment for this system is excellent when both carbonyl shielding and ring current effects are taken into account (Asakura et al., 1991).

In Fig. 3 a plot given for cytochrome b_5 shows the strong opposite shifts of α-CH protons in β-sheets compared with those in α-helices. The shifts in this case are clearly diagnostic of secondary structure (Szilagyi and Jardetzky, 1989; Williamson, 1990). In contrast, the amide proton shifts are not as amenable to a similar analysis since the shift on the peptide NH depends on the hydrogen bond strength although there is a parallel shift due to the carbonyl. Note, however, that strong NH bonding shifts the NH downfield. Indeed the vast majority of the NH protons in the β-sheet of cytochrome b_5 are shifted markedly downfield relative to those NH resonances in α-helices as illustrated in Fig. 4.

The most conventional analysis of diamagnetic shifts is that due to ring currents from aromatic residues (Perkins, 1982). In particular it has been shown that the shift on aliphatic protons can be readily interpreted in a large number of cases and with good agreement with the expectation from crystal structures. However, neither ring current shifts nor carbonyl shifts give one the opportunity to compare protein surface structure in solution with that in crystals and the limitation of these methods to short-range effects of less than 5.0 Å makes detailed study difficult since the shifts can have multiple origins.

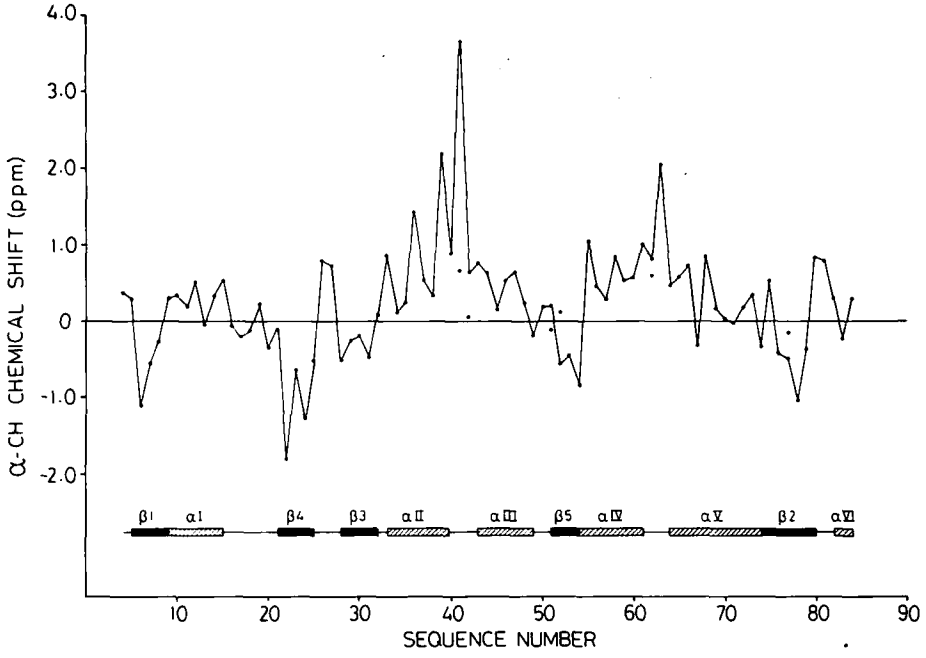

Fig. 3. The secondary shift of the α-CH protons of ferrocytochrome b_5 against the sequence number and the secondary structural elements as defined by Mathews et al. (1979). Where glycine residues are concerned the plot is drawn through the α-CH value of greatest magnitude for convenience.

Paramagnetic shift analysis

The paramagnetic shift analysis can be made from a single centre such as a metal atom. In this case the shift is measurable over distances of up to 15 Å. Today shifts between a diamagnetic blank and a paramagnetic system are measurable to a high degree of accuracy but are only fully interpretable where there is no change of structure of the protein between the diamagnetic and the paramagnetic states. The most favourable case is that of the heme proteins where the low-spin diamagnetic Fe(II) state can be used as the blank and the paramagnetic Fe(III) state as the probe (Williams et al., 1985).

Paramagnetic shifts based on the equation for the pseudocontact shift, δ_{pc}, where

$$\delta_{pc} = \delta\,(\text{diamagnetic blank}) - \delta\,(\text{paramagnetic system})$$

when the diamagnetic and paramagnetic molecules are known to be or are expected to be isomorphous can be interpreted through the pseudocontact shift equation:

$$\delta_{pc} = -\frac{N\beta^2 S(S+1)}{9kT}\left\{\left[g_z^2 - \tfrac{1}{2}(g_x^2+g_y^2)\frac{(3\cos^2\theta-1)}{r^3}\right] + \tfrac{3}{2}\left[(g_x^2-g_y^2)\frac{(\sin^2\theta\cos 2\varphi)}{r^3}\right]\right\}$$

where g_x, g_y and g_z are the principal values of the g-tensor describing the magnetic axis in a polar

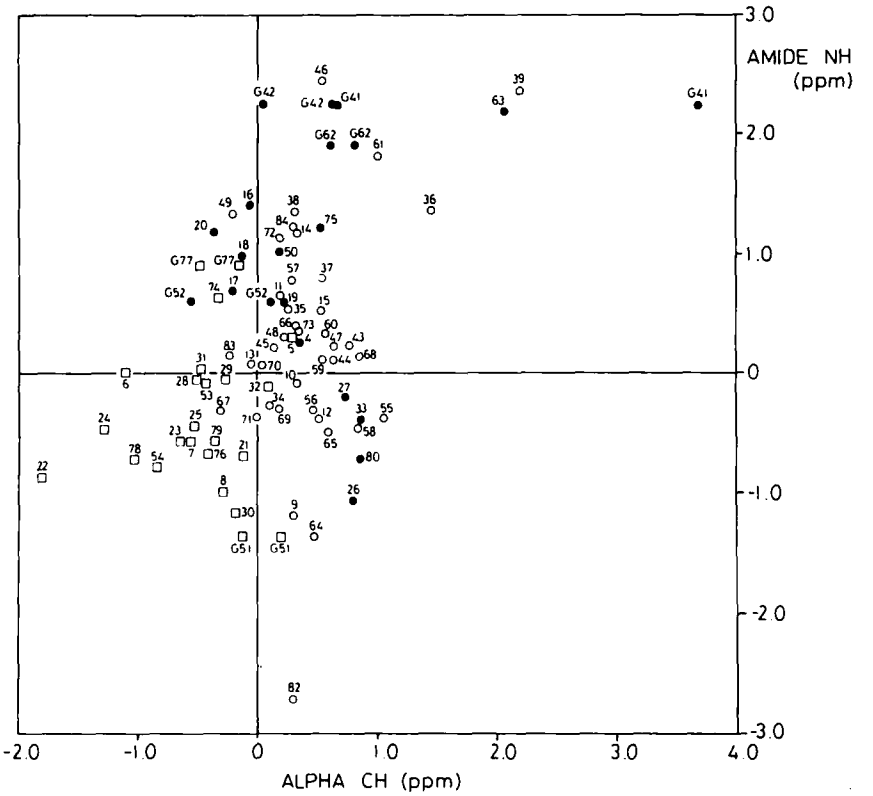

Fig. 4. Correlation of the main-chain NH and α-CH proton secondary chemical shifts for ferrocytochrome b_5. Residues are classified as follows: (□) β-sheet; (○) α-helix; (●)β-turn.

coordinate system (r,θ,φ) defined with respect to the heme iron atom (Kurland and McGarvey, 1970). The case of cytochrome b_5 will now be taken in detail and the paramagnetic shift data together with NOE information examined.

Comparison between NOE and chemical shift data

(a) Cytochrome b_5

The confidence to be placed in the overall analysis of the structure by the paramagnetic shift method can be made quantitative using the parameter $\Sigma(\delta_{calc}-\delta_{obs})^2$ for all resonances where δ_{obs} and δ_{calc} are the observed and calculated pseudocontact shifts, respectively. The value for cytochrome b_5 is 6.0 compared with the sum, $\Sigma\delta^2_{obs}$ of 200 (Veitch et al., 1990). The difference between calculated and observed structural features for individual protein groups is more clearly seen by examining the function $\frac{\delta_{obs}-\delta_{calc}}{\delta}$ where δ is the larger of the two, δ_{obs} and δ_{calc}. This second analysis is not weighted by the magnitude of δ. Examination of the data for cytochrome b_5 shows that there are discrepancies only for a very few protons. These include:

(1) Aromatic ring protons. However, if the flipping of rings, i.e. fast averaging of δ- and ε-pro-

tons, respectively, for all the aromatic rings is taken into account then the discrepancy is removed.

(2) Aromatic protons of a surface histidine, His^{26}, whose chemical shift values are affected by the pH, thus making direct comparison with diffraction data difficult.

(3) The shifts observed on Val^{45} protons which can only be accounted for if this surface residue shows fast rotation around the β-CH-α-CH bond.

(4) The residue Gly^{42} has two conformations indicating mobility in the sequence Gly-Gly(42)-Glu-Glu. This is not altogether surprising in the light of similar observations for yeast iso-1-cytochrome *c* (Gao et al., 1990).

No attempt has been made in this analysis to treat data for surface side chains such as those of lysine.

The examination of surface loops either through examination of $(\delta_{obs} - \delta_{calc})^2$ or by $\frac{\delta_{obs} - \delta_{calc}}{\delta}$ indicates that they are as well defined as the helices and the β-sheet. The loops or turns are in almost all cases very short. The conclusion in the case of cytochrome b_5 is that the structure of all the backbone and of internal side chains is of low mobility in virtually all respects. With one or two exceptions mentioned above this is also true of the small and/or more hydrophobic surface residues. The stability in solution is such that this protein can withstand a change of unit charge at its centre, Fe(II) to Fe(III), with virtually no influence on the structure. The examination of crystals of the protein in its two oxidation states also shows that the structure is fixed, excluding from this analysis the side chains of lysines and other long polar surface residues (Mathews et al., 1979).

Against this background the data on the NOE connectivities between groups in the protein can be examined. In general there is no difference between the NOE data for the two oxidation states. In Figs. 1 and 2 the plots show only wide bands of 'agreement' which implies that at the local level of structure there is no simple way of matching NOE-structures with crystallographic structures to $\sim \pm 1.0$ Å or with shift data (In no way does this reflect on the ability to deduce an overall fold for a protein). The suggestion is that the definition of structure by NOE procedures is imprecise locally but that here this does not reflect any considerable mobility since the shift data indicate that solution and crystal structures are very similar. To take a detailed example it is observed that agreement is relatively poor for the spatial relationship of Trp^{22}, which is chosen as it is likely to be rigid, and Glu^{11}, Ile^{12} and Val^{29} side chains. Of these, Glu^{11} is a surface residue while Ile^{12} and Val^{29} form part of the second of two hydrophobic cores.

Previously the NOE data for another protein, acylphosphatase have been examined (Saudek et al., 1989). In this case the short loops of structure were as well defined as the secondary structure of β-strands and α-helices. However, the larger loops were less well defined and reasons were proposed for supposing that the definition was poor through increased mobility. In both studies it could well be that the NOE method exaggerates the disorder or mobility in a structure locally.

(b) Cytochromes c

The analysis of the structure of cytochromes *c* by paramagnetic shift data using the two oxidation states of the iron was carried out in the above manner. In the case of this protein it was necessary to determine the crystal structures of the cytochromes in both the reduced and oxidised states and it was shown that there was a noticeable structural difference between the redox states in certain regions on the methionine side of the heme (Louie and Brayer, 1990). In these circumstances the paramagnetic shift analysis has been used somewhat cautiously (Gao et al., 1991). It

TABLE 2
PROTONS FOR WHICH THE DIFFERENCE SQUARED BETWEEN THE OBSERVED (δ_{obs}) AND CALCULATED (δ_{calc}) PSEUDOCONTACT SHIFTS OF C102T YEAST ISO-1-CYTOCHROME *c* IS GREATER THAN 0.02 AND THE PERCENTAGE ERROR IS GREATER THAN 50%

Residue	Proton	Shift		Error		Assignment	
		Calc	Obs	(Obs − Calc)²	%	Red	Oxid
Ser[2]	HB1	−0.113	−0.330	0.047	66	3.95	3.62
Ser[2]	HB2	−0.140	−0.360	0.048	61	3.76	3.40
Phe[10]	HE1	1.127	0.390	0.543	65	6.85	7.22
Phe[10]	HD2	−0.519	0.010	0.280	102	7.15	7.16
Cys[14]	HN	−0.481	−0.160	0.103	67	8.08	7.92
Leu[15]	HB2	0.450	0.150	0.090	67	1.38	2.53
Gln[16]	HB1	0.196	0.450	0.065	56	2.02	2.47
Gln[16]	HG1	0.113	0.330	0.047	66	2.53	2.86
Val[28]	HN	0.067	0.490	0.179	86	7.05	7.54
Gly[29]	HN	−0.069	−0.240	0.029	71	7.71	7.47
Pro[30]	HA	0.422	0.190	0.054	55	3.61	3.80
Pro[30]	HB1	0.230	0.090	0.020	61	1.28	1.37
Ile[35]	HN	0.145	−0.050	0.038	135	7.22	7.17
Phe[36]	HA	−0.127	−0.290	0.027	56	4.35	4.06
Arg[38]	HA	−0.037	−0.200	0.026	81	4.87	4.67
Arg[38]	HB2	−0.028	−0.180	0.023	84	2.02	1.84
Arg[38]	HE	−0.023	−0.400	0.142	94	7.68	7.28
His[39]	HN	−0.053	0.230	0.080	123	7.97	8.20
His[39]	HA	−0.055	−0.410	0.126	87	5.66	5.25
Ser[40]	HA	−0.164	−0.460	0.088	64	4.98	4.52
Gly[41]	HN	−0.332	0.720	1.106	146	8.23	8.95
Gln[42]	HB1	−0.087	−0.260	0.030	67	1.94	1.68
Gln[42]	HG2	−0.073	−0.240	0.028	70	2.28	2.13
Gln[42]	HE22	−0.069	−0.230	0.026	70	7.54	7.31
Ala[43]	HN	−0.200	−0.600	0.160	67	8.72	8.12
Ala[51]	HN	−0.007	0.290	0.088	102	7.74	8.03
Asn[52]	HA	0.210	0.470	0.068	55	4.15	4.62
Asn[52]	HB2	−0.266	−0.040	0.051	85	2.96	2.92
Ile[53]	HA	−0.122	0.250	0.139	149	3.40	3.65
Lys[55]	HN	0.075	−0.070	0.021	194	7.46	7.39
Lys[55]	HB1	0.230	0.040	0.036	83	1.90	1.94
Asn[56]	HN	0.038	0.260	0.049	86	7.04	7.30
Asn[56]	HD22	−0.030	0.150	0.032	120	6.35	6.50
Val[57]	HN	0.067	−0.120	0.035	156	7.40	7.28
Val[57]	HA	0.051	0.390	0.115	87	3.70	4.09
Val[57]	HB	0.066	0.330	0.070	80	1.02	1.35
Trp[59]	HZ2	0.193	0.410	0.047	53	7.14	7.55
Trp[59]	HE3	−0.229	−0.070	0.025	69	7.59	7.52
Asp[60]	HN	−0.154	0.450	0.365	134	9.47	9.92
Glu[61]	HB2	−0.218	0.030	0.062	114	1.12	1.15
Asn[63]	HB1	−0.104	0.130	0.055	180	3.13	3.26
Asn[63]	HB2	−0.110	0.050	0.026	145	2.85	2.90
Asn[63]	HD21	−0.039	−0.210	0.029	81	7.58	7.37

TABLE 2 (continued)

Residue	Proton	Shift		Error		Assignment	
		Calc	Obs	(Obs − Calc)²	%	Red	Oxid
Ser^{65}	HB2	−0.208	0.100	0.095	148	3.73	3.83
Glu^{66}	HB1	0.025	0.280	0.065	91	1.69	1.97
Glu^{66}	HB2	0.114	0.310	0.038	63	1.81	2.12
Leu^{68}	HA	−0.505	−0.170	0.112	66	3.07	2.90
Asn^{70}	HB1	0.349	0.170	0.032	51	3.09	3.26
Asn^{70}	HD21	0.309	0.640	0.109	52	7.29	7.93
Pro^{71}	HD2	0.976	0.440	0.287	55	3.09	3.53
Tyr^{74}	HA	0.475	0.950	0.225	50	4.23	5.18
Thr^{78}	HG1	1.114	0.140	0.948	87	8.78	8.92
Lys^{79}	HN	−0.092	−0.700	0.370	87	8.77	8.07
Lys^{79}	HA	0.092	0.430	0.115	79	4.44	4.87
Ala^{81}	HN	0.290	0.110	0.032	62	8.15	8.26
Ala^{81}	HB3	0.197	−0.050	0.061	125	1.38	1.33
Phe^{82}	HA	0.774	0.220	0.307	72	4.45	4.67
Phe^{82}	HD1	0.649	−0.550	1.439	185	6.73	6.18

has been concluded that the somewhat poorer fit of the NMR shift data to the calculated data for these cytochromes is due to:

(1) A rearrangement of the hydrogen bond network close to the heme on the Met^{80} side including several changes in association with the heme propionates.

(2) Changes in the NH bond network on the Met^{80} side of the heme as evidenced by some very considerable changes in chemical shifts which are not predicted by the paramagnetic shift analyses (Table 2).

(3) Small changes in coordinates associated particularly with helix 50-56 but general to the structural region of sequence 50–85, and perhaps a small movement of the heme within the protein. There can be no doubt, and here the agreement with very high resolution crystallographic data is excellent, that there are multiple NH-bond distance and non-bond distance changes of the order of 0.5–1.0 Å but never larger than this throughout this zone of the protein. The NOE data for the oxidised and reduced proteins have been examined to see if there is any reflection of these changes but little evidence can be found. As Fig. 2 shows, comparison of expected NOE intensity with known crystallographic data leaves the impression that the structural interpretation of the NOE is not likely to be accurate in this case to better than about ± 1.0 Å when the above changes can not be seen. The analysis can be repeated with modified cytochrome *c* where the strength of the fold under the heme, i.e. that part connected to the Met^{80} and thence to the iron atom, is deliberately and progressively weakened. This can be achieved by making mutations at Phe^{82}, e.g. Ser and Gly for Phe, or by breaking the Met^{80} bond by reacting cytochrome *c* with cyanide.

(c) Yeast iso-1-cytochrome c C102T:F82S variant

In the case of the C102T:F82S variant the same degree of assignment of protons has been achieved as for the native protein. The NMR observations show that in solution and in the re-

duced state there are small changes relative to the native protein, as indicated by chemical shift differences (Gao et al., 1991) in agreement with the crystallographic data (Louie et al., 1988). In the oxidised proteins the differences are in the same regions but magnified. The changes in the reduced protein are in the same regions as the changes seen between oxidised and reduced *native* proteins but with additional changes around the region 28-35. In addition the oxidised form of C102T:F82S shows a considerable change in internal mobility as measured by increased rates of ring flipping or increased rates of hydrogen exchange (Table 3).

We next look for NOE evidence concerning these changes. There is very little that agrees with the above due to a major problem within the NOE data set. The difficulties are illustrated in Fig. 5 where an NOE pattern for the 5-methyl group of the heme is shown in both the native and the C102T:F82S oxidised variant. There is a very large difference between the two and this NOE result could indicate a considerable structural change. In fact it is clear from the shift calculations and the crystal structure that the change is nowhere more than 0.5 Å. It must therefore be concluded that due to the internal mobility the NOE is reflecting perturbations in relaxation times rather than significant changes in distances.

(d) The cyanide complex of horse cytochrome c

The assignment of the NMR spectrum of the cyanide complex of horse cytochrome *c* has been completed to the same level as that of the oxidised native protein. The chemical shifts indicate that the structures are very similar over large parts of the protein including the two terminal helices 1-14 and 85-102 (Gao, 1991). The general fold has also the same structural features as illustrated in Fig. 6. The region from 78-85 is modified as would be expected from the breaking of the Met^{80} to Fe(III) bond. In fact the use of paramagnetic shift analysis, with the native reduced protein as the diamagnetic reference state, shows the following changes in more detail.

(1) General small changes especially in peptide NH resonances in the region on the Met^{80} side of the heme.

(2) Noticeable changes in the region 28-35 but there is not a change in the hydrogen bond pattern at Gly^{41}, i.e. relative to the reduced native protein.

(3) There is a greater change in mobility than was found for the C102T:F82S variant.

(4) Apart from the loss of some NH resonances of side-chain amides and increased ring-flipping

TABLE 3
RESONANCES OBSERVABLE IN C102T YEAST ISO-1-CYTOCHROME *c* SPECTRA AND ABSENT IN THOSE OF THE C102T:F82S VARIANT

Resonance	Comment
Asn^{31}, Asn^{52}, Asn^{70}	Side-chain amide resonances absent Fast exchange with H_2O
Phe^{10}, Tyr^{46}, Tyr^{48}, Tyr^{97}	Aromatic side-chain resonances absent Intermediate ring flip rate
$Gly^{29}(\alpha_1,\alpha_2)$, $Ser^{82}(\alpha,\beta_1,\beta_2)$	Backbone resonances absent Intermediate rate of internal motion

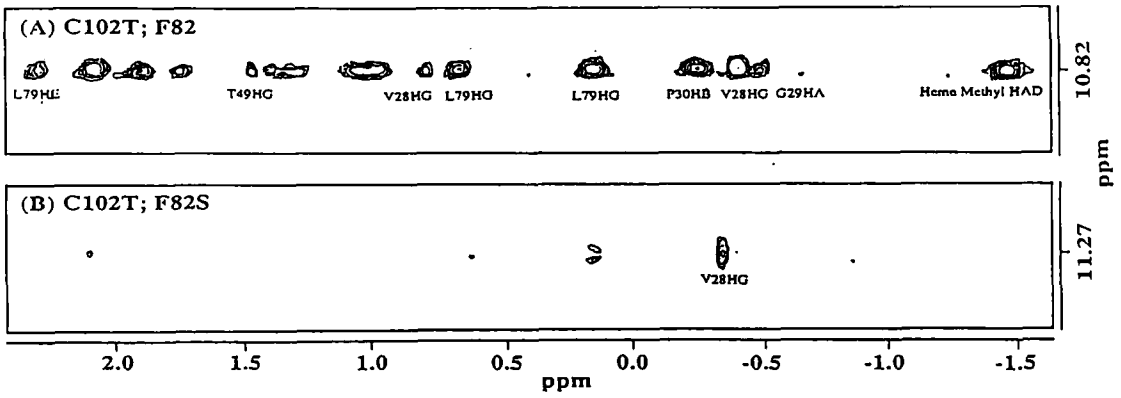

Fig. 5. NOE connectivities from the heme methyl-5 protons of (A) the C102T and (B) the C102T:F82S variants of yeast-iso-1-ferricytochrome *c*. The spectra were recorded in D_2O, pH 6.0, 27°C with a mixing time $\tau_m = 130$ ms.

rates it is also apparent that NMR signals from amino acids 78-84 were absent, with the exception of the side chain of Phe^{82} which could be located from NOE and shift data as being close to its position in the native state.

The loss of signals in the two-dimensional experiment is interpreted as indicative of a greatly increased mobility in the region from 78-84. Two sets of experiments confirm this impression: (1) In 1D spectra the Met^{80} ε-CH_3 resonance could be found at a position close to its random coil position; (2) By measuring 2D NOE spectra at four temperatures 7°, 17°, 27° and 37°C, considerable changes in NOE intensity could be followed with little change in chemical shift except that expected by comparison with the native protein. It is undoubtedly the case that use of NOE data could not give a correct structure in this region of increased mobility.

CONCLUSION

There is no doubt that the use of NOE data can give good outline structures of proteins. There is also no doubt that these structures can be refined in a methodically internally consistent manner such that pictures can be given of apparently very well defined structures. Claims are made that these structures are an accurate representation of the real state of the protein to better than 1.0 Å even for side chains. Now internal consistency is not a thorough check of such a claim to precision. In this paper a deliberate comparison between three methods of structure analysis has been made in order to test the validity of the NOE method.

The methods are: refined X-ray diffraction in crystals, chemical shift analyses and the NOE itself. There is excellent agreement between the crystal structure data and shift analysis but only general agreement concerning NOE data at the $\pm$ 1.0 Å level of resolution. Even in the case of a well-defined structure with almost no loose loops or turns, i.e. cytochrome b_5, the NOE method would not necessarily produce a rigid impression of a single structure. Rather it must produce an ensemble of related structures as can be seen at a local level given the data in Fig. 1. In the case of cytochrome *c* where internal mobility of the structure can be increased locally and in two steps from the native protein, i.e. by the C102T:F82S mutation and by forming a cyanide complex, the

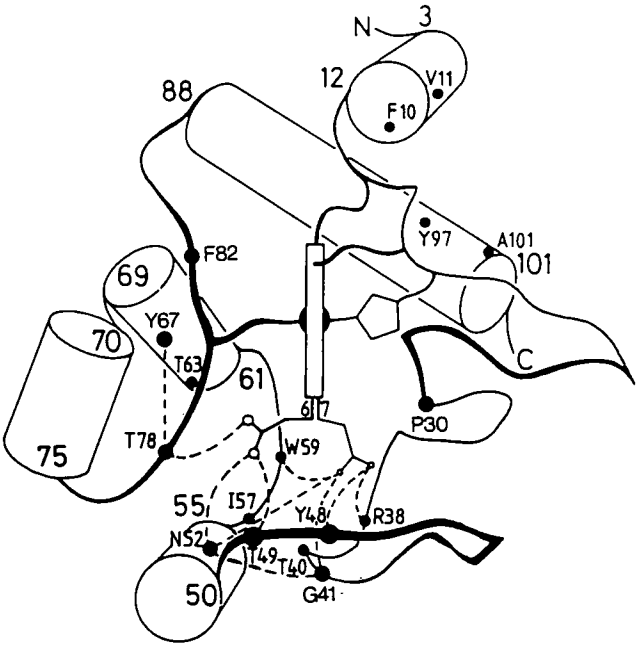

Fig. 6. Outline structure of cytochrome *c* with an indication of the internal hydrogen-bond network (dashed lines). The axial ligand Met^{80} appears on the left of the heme group, here viewed from the side.

evidence is clear that variations of internal mobility cause grave problems for the NOE method. It was impossible to deduce any feature of the structure changes seen by either crystal structure or NMR shift data. A similar impression was obtained in a study of acylphosphatase where tight and loose loops were compared. Similar data for other proteins are obviously important. For instance, in the case of hen lysozyme the loop of structure around residues 61-78 as found by X-ray structure analysis apparently does not conform to this structure when the molecule is in solution as judged by NOE methods (Smith, L.J., Redfield, C. and Dobson, C.M., personal communication).

The intention of this paper is to indicate the strengths and limitations of the NOE method of structure determination. There is no doubt that well-defined secondary structure features are readily and accurately uncovered by this NMR method. The paper highlights the problems associated with an accurate description of individual side chains and with structure determination where there is variation in internal mobility. In fact it seems that a combination of observations of NOE and shift data will define structure more precisely and also give information about mobility. This is essential to an understanding of protein function and generally does not follow from crystallographic analyses.

ACKNOWLEDGEMENTS

We thank Jonathan Boyd and Nick Soffe for their generous assistance with the NMR experiments. R.J.P.W. is a member of the Oxford Centre for Molecular Sciences. N.C.V. thanks Merton College, Oxford, for the provision of a Harmsworth Senior Scholarship.

REFERENCES

Asakura, T., Nakamura, E., Asakawa, H. and Demura, M. (1991) *J. Magn. Res.*, **93**, 355–360.

Barry, C.D., North, A.C.T., Glasel, J.A., Williams, R.J.P. and Xavier, A.V. (1971) *Nature*, **232**, 236–245.

Barry, C.D., Hill, H.A.O., Mann, B.E., Sadler, P.J. and Williams, R.J.P. (1973a) *J. Am. Chem. Soc.*, **95**, 4545–4551.

Barry, C.D., Hill, H.A.O., Sadler, P.J. and Williams, R.J.P. (1973b) *Proc. Roy. Soc. A*, **334**, 493–504.

Brautigan, D.L., Ferguson-Miller, S. and Margoliash, E. (1978) *J. Biol. Chem.*, **253**, 130–139.

Burns, P.S., Williams, R.J.P. and Wright, P.E. (1975) *J. Chem. Soc. Chem. Commun.*, 795–796.

Clayden, N.J. and Williams, R.J.P. (1982) *J. Magn. Res.*, **49**, 383–396.

Clore, G.M., Gronenborn, A.M., Kjær, M. and Poulsen, F. (1987) *Prot. Eng.*, **1**, 305–311.

Clore, G.M., Wingfield, P.T. and Gronenborn, A.M. (1991) *Biochemistry*, **30**, 2315–2323.

Cutler, R.L., Pielak, G.J., Mauk, A.G. and Smith, M. (1987) *Prot. Eng.*, **1**, 95–99.

Dalgarno, D.C., Levine, B.A. and Williams, R.J.P. (1983) *Bioscience Rep.*, **3**, 443–452.

Dobson, C.M., Ford, L.O., Summers, S.E. and Williams, R.J.P. (1975) *J. Chem. Soc. Farad. Trans. II*, **71**, 1145–1153.

Dobson, C.M., Geraldes, C.F.G.C., Ratcliffe, R.G. and Williams, R.J.P. (1978) *Eur. J. Biochem.*, **88**, 259–266.

Feng, Y., Roder, H., Englander, S.W., Wand, A.J. and Di Stefano, D.L. (1989) *Biochemistry*, **28**, 195–203.

Gao, Y. (1991) D. Phil. Thesis, University of Oxford.

Gao, Y., Boyd, J., Williams, R.J.P. and Pielak, G.J. (1990) *Biochemistry*, **29**, 6994–7003.

Gao, Y., Boyd, J., Pielak, G.J. and Williams, R.J.P. (1991) *Biochemistry*, **30**, 1928–1934.

Guiles, R.D., Altman, J., Kuntz, I.D. and Waskell, L. (1990) *Biochemistry*, **29**, 1276–1289.

Kline, A.D., Braun, W. and Wüthrich, K. (1988) *J. Mol. Biol.*, **204**, 675–724.

Kline, T.P., Brown, F.K., Jeffs, P.W., Kopple, K.D. and Mueller, L. (1990) *Biochemistry*, **29**, 7805–7813.

Kurland, R.J. and McGarvey, B.R.J. (1970) *J. Magn. Res.*, **2**, 289–301.

Louie, G.V. and Brayer, G.D. (1990) *J. Mol. Biol.*, **214**, 527–555.

Louie, G.V., Pielak, G.J., Smith, M. and Brayer, G.D. (1988) *Biochemistry*, **27**, 7870–7876.

Mathews, F.S., Czerwinski, E.W. and Argos, P. (1979) In *The Porphyrins, Vol. VIIB* (Ed, Dolphin, D.) Academic Press, New York, pp. 107–147.

Moore, J.M., Case, D.A., Chazin, W.J., Gippert, G.P., Havel, T.F., Powls, R. and Wright, P.E. (1988) *Science*, **240**, 314–317.

Perkins, S.J. (1982) In *Biological Magnetic Resonance, Vol. 4* (Eds, Berliher, L. and Reuben, J.) Plenum Press, New York, pp. 193–336.

Pielak, G.J., Mauk, A.G. and Smith, M. (1985) *Nature*, **313**, 152–154.

Reid, L.S. and Mauk, A.G. (1982) *J. Am. Chem. Soc.*, **104**, 841–845.

Saudek, V., Williams, R.J.P., Stefani, M. and Ramponi, G. (1989) *Eur. J. Biochem.*, **185**, 99–103.

Smith, L.J., Sutcliffe, M.J., Redfield, C. and Dobson, C.M. (1991) *Biochemistry*, **30**, 986–996.

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

Veitch, N.C., Whitford, D. and Williams, R.J.P. (1990) *FEBS Lett.*, **269**, 297–304.

Williams, G., Clayden, N.J., Moore, G.R. and Williams, R.J.P. (1985) *J. Mol. Biol.*, **183**, 447–460.

Williams, R.J.P. (1989) *Eur. J. Biochem.*, **183**, 479–497.

Williamson, M.P. (1990) *Biopolymers*, **29**, 1423–1431.