Complete Sequence-Specific ^1H NMR Assignments for Human Insulin

Allen D. Kline* and Richard M. Justice, Jr.
Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285-1513
Received August 8, 1989; Revised Manuscript Received November 14, 1989

ABSTRACT: Solvent conditions where human insulin could be studied by high-resolution NMR were determined. Both low pH and addition of acetonitrile were required to overcome the protein's self-association and to obtain useful spectra. Two hundred eighty-six ^1H resonances were located and assigned to specific sites on the protein by using two-dimensional NMR methods. The presence and position of numerous d_{NN} sequential NOE's indicate that the insulin conformation seen in crystallographic studies is largely retained under these solution conditions. Slowly exchanging protons were observed for seven backbone amide protons and were assigned to positions A15 and A16 and to positions B15–B19. These amides all occur within helical regions of the protein [Chawdhury, S. A., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Tolley, S. P., Blundell, T. L., Cleasby, A., Pitts, J. E., Tickle, I. J., & Wood, S. P. (1983) *Diabetologia 25*, 460–464].

For the last 70 years insulin has been the subject of intense scientific research, often being used as the "standard protein" by which new chemical and physical techniques have been evaluated and applied. It was the first protein to be sequenced (Brown et al., 1955) and also the first to be chemically synthesized from its constituent amino acids (Meienhofer et al., 1963).

The most definitive source of information for understanding the insulin conformation has come from extensive X-ray crystallographic studies carried out in Oxford (Adams et al., 1969) and in Beijing (Wang et al., 1982). Crystallographic refinements of the porcine hexamer (Chang et al., 1986; Baker et al., 1988) have allowed a detailed examination of hydrogen bonds, of multiple side-chain conformations, and of numerous water molecules that occur within the complex. Other crystalline forms of insulin (Bentley et al., 1976; Dodson et al., 1978) and different insulin sequences (Chawdhury et al., 1983; Liang et al., 1985) have been examined. These studies have led to an understanding of conformation as a result of self-association and crystal packing. This understanding is critical, as the active form of the hormone is most certainly the monomer (Pekar & Frank, 1972).

In order to understand the solution conformation of monomeric insulin, several groups have undertaken NMR[1] studies of insulin and its related derivatives (Bradbury et al., 1981; Palmieri et al., 1988; Hua et al., 1989; Weiss et al., 1989). Yet the effects of self-association on the individual insulin molecule are still unknown. Application of multidimensional NMR techniques (Ernst et al., 1987; Wüthrich, 1986) along with isotopic labeling (LeMaster & Richards, 1988; Fesik 1988; Westler et al., 1988; Torchia et al., 1988; Lowry et al., 1988) should make it possible to study insulin in most of its associated forms. This paper describes our efforts to assign all 286 proton resonances for human insulin.

Experimental Procedures

Biosynthetically produced human insulin was obtained as a zinc-free lyophilized powder courtesy of Drs. Ron Chance and Bruce Frank of Lilly Research Laboratories. Water, acetonitrile, and trifluoroacetic acid with the highest possible levels of deuteration were purchased from MSD Isotopes, Montreal, Canada, and used without further purification. Deuterated insulin samples were prepared by exposing the dry powder to ^2H_2O for 30 min, followed by freeze-drying. This procedure was performed twice. The dried powder was placed into 0.325 mL of ^2H_2O, and small amounts of a ^2H_2O solution containing trifluoroacetic acid were added to dissolve the protein. A micro pH electrode was used to monitor the solution acidity, and the titration was halted when the meter read 3.6. Once the acidity was established, 0.175 mL of deuterated acetonitrile was added. In order to observe all amide protons, a separate sample was prepared by adding insulin to ^1H_2O, and the solution was then acidified with trifluoroacetic acid to a meter reading of 3.6. Finally, the protein solution received a similar volume of deuterated acetonitrile. All samples were centrifuged before being placed into a 5-mm NMR tube. The final concentration of insulin in all of the samples was calculated to be 2 mM.

All NMR experiments were carried out on a Bruker AM-500 spectrometer equipped with a process controller. The probe temperature was maintained at either 27 or 37 °C, and the sample tubes were not spun. Chemical shift values were referenced to deuterated TSP indirectly; that is, TSP was added to the sample only after the 2D spectra were collected. Referencing then occurred by measuring the shift of an isolated peak in the 1D spectrum and assigning its corresponding 2D peaks to that same value. 2QF-COSY (Rance et al., 1983), relay-COSY (Eich et al., 1982; Wagner, 1983a; Bax & Drobny, 1985), NOESY (Jeener et al., 1979), TOCSY (Braunschweiler & Ernst, 1983; Bax & Davis, 1985), and 2Q spectra (Wagner & Zuiderweg, 1983) were collected as previously described. ^1H_2O spectra were collected by presaturation of the water resonance by using continuous irradiation during the relaxation period (Wider et al., 1983). The repetition rate for all NMR experiments was 2 s. Frequency discrimination in ω_1 was accomplished by time-proportional phase incrementation (Marion & Wüthrich, 1983).

Data were processed on a VAX-8800 computer using software developed by Dennis Hare of Hare Research, Inc., Woodinville, WA. Phase-sensitive Fourier transformations were used in all cases except for some of the relay-COSY experiments. Spectral matrices were always 1024 complex points in ω_1 and 2048 complex points in ω_2. For high-resolution transforms ω_2 was zero-filled to 4096 complex points. The

[1] Abbreviations: NMR, nuclear magnetic resonance; 1D and 2D, one and two dimensional; ppm, parts per million; COSY, 2D correlated spectroscopy; 2QF-COSY, two-quantum-filtered COSY; relay-COSY, 2D relayed coherence transfer spectroscopy; 2Q spectroscopy, two-quantum spectroscopy; NOESY, 2D nuclear Overhauser spectroscopy; TOCSY, 2D total correlated spectroscopy; TSP, sodium 3-(trimethylsilyl)propionate-*2,2,3,3,d*$_4$; NOE, nuclear Overhauser enhancement; d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$ sequential NOE interactions between the amide on the following residue and the NH, C$^\alpha$H, and C$^\beta$H, respectively.

0006-2960/90/0429-2906$02.50/0

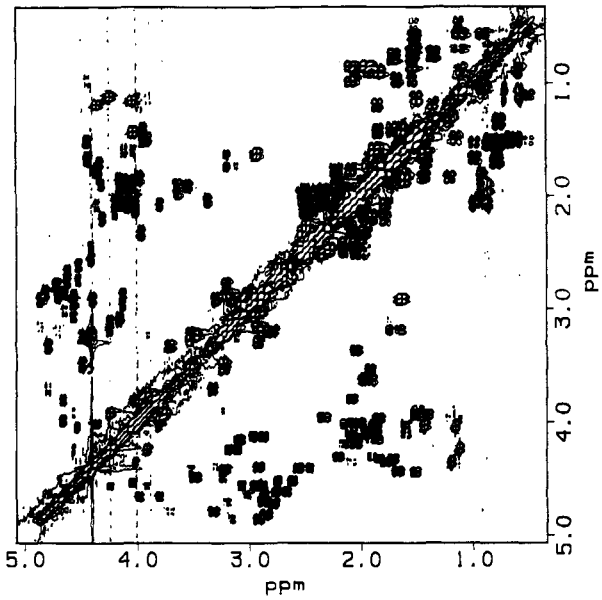

FIGURE 1: Aliphatic region of the 2QF-COSY spectrum of human insulin at 27 °C dissolved in completely deuterated solvents. The spectrum was plotted for both ω_1 (vertical axis) and ω_2 (horizontal axis) from 5.08 to 0.34 ppm. Both positive and negative contours are shown.

2D data files were transformed with several multiplication functions to allow a more complete analysis of the spectra. Third-order polynomial base-line corrections were used in both dimensions, particularly for in-phase NMR data.

Results

Selection of Solvent Conditions. The assigning of the insulin ^{1}H NMR spectrum by now classical 2D NMR methods (Wüthrich, 1986) required that the self-association properties of the protein (Jeffrey & Coates, 1966) be taken into account. Thus the initial task was to minimize protein aggregation without destroying the globular structure of the monomeric insulin.

The solvent conditions chosen for this study were low pH (3.6) and addition of 35% (volume) deuterated acetonitrile. This solvent reduced the self-association of the protein and allowed good-quality spectra to be obtained. The addition of the organic solvent did cause some shifting in positions of several signals in the 1D spectra (see supplementary material). However, the resulting NMR spectra (Figures 1 and 2) still show a considerable amount of chemical shift dispersion, indicative of a globular protein.

Identification of the Residue Spin Systems. Since the methods for obtaining sequential assignments are well established (Wüthrich, 1986) and many examples exist in the literature, only a few details illustrating the particular problems that arose during the assignment procedure will be presented.

After the initial analysis of all the ^2H_2O spectra, one glycine (B20) was still missing. This residue was later identified in the ^1H_2O spectra as having equivalent $C^\alpha H$ chemical shifts. The presence of both $C^\alpha H$ protons was subsequently confirmed by the observation of a glycine remote peak in the ^1H_2O 2Q spectrum.

The assignment of Thr B30 was complicated due to a near equivalency of its $C^\alpha H$ and $C^\beta H$ chemical shifts. The ^2H_2O spectra did not allow identification of the threonine $C^\beta H$–$C^\gamma H_3$ cross peak from an alanine $C^\alpha H$–$C^\beta H_3$ peak. Confirmation of its identity and its $C^\alpha H$ shift came later as a result of analyzing the ^1H_2O 2QF-COSY, NOESY, and TOCSY spectra.

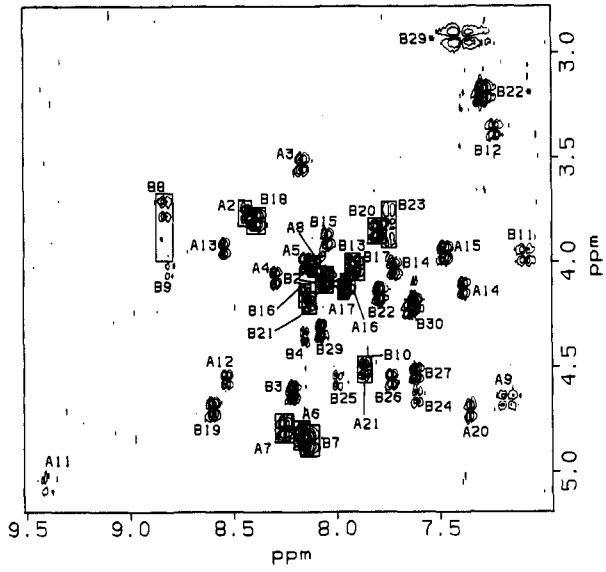

FIGURE 2: Fingerprint region of a 2QF-COSY spectrum of human insulin at 27 °C dissolved in ^1H_2O/acetonitrile. The spectrum was plotted for ω_1 (vertical axis) from 5.19 to 2.78 ppm and for ω_2 (horizontal axis) from 9.51 to 6.95 ppm. Both positive and negative contours have been plotted. Each NH–$C^\alpha H$ cross peak has been labeled with the assigned chain and sequence position. Boxes have been drawn to clearly indicate the peak positions in crowded regions. The two cross peaks marked with the asterisk are due to side-chain protons.

Serine B9 was represented by a single COSY cross peak (ω_2 = 3.80 ppm, ω_1 = 4.04 ppm) in a crowded portion of the spectrum. The lack of COSY information had to be supplemented with other cross peaks from the remaining spectra to fully assign this residue's protons. An NH–$C^\alpha H$ cross peak from the ^1H_2O 2QF-COSY spectrum identified the $C^\alpha H$ shift as 4.04 ppm. The 2Q spectrum in ^2H_2O gave a remote peak for the $C^\alpha H$ which confirmed that two protons were resident at 3.80 ppm.

Difficulties with the aromatic residues were limited to Phe B1 and Phe B24. For B1 an amide shift was never found. While the lack of an amide proton is consistent with its assignment to the N-terminal residue of the B chain, it does not substantiate its sequence position. Its position had to await the assignment of the sequential $d_{\alpha N}$ and $d_{\beta N}$ NOESY cross peaks.

Complications arose when trying to assign the ring protons of phenylalanine B24. The 2QF-COSY spectrum showed no obvious $C^\epsilon H$–$C^\zeta H$ cross peak, which would have permitted an assignment to phenylalanine. Only when the sequential assignment of B24 was certain was it known that a $C^\zeta H$ resonance should exist. The $C^\zeta H$ was tentatively assigned by default as the last remaining unassigned aromatic signal from the NOESY spectrum. This signal is near the $C^\epsilon H$ shift of Phe B24, which would make its corresponding COSY cross peak difficult to detect.

Cys A11's line widths were considerably broader than were most of the other insulin protons. The additional line width is responsible for the lack of $C^\alpha H$–$C^\beta H$ cross peaks in the 2QF-COSY. The only COSY data obtained on this spin system were weak $C^\beta H$–$C^\beta H$ and NH–$C^\alpha H$ cross peaks from the ^2H_2O and ^1H_2O spectra, respectively. These two spin system fragments were eventually connected by using a TOCSY experiment in ^2H_2O where the $C^\alpha H$–$C^\beta H$ cross peaks were manifest.

Obtaining positive evidence for cysteine assignments without using sequential NOE's or isotopic labeling is difficult due to a lack of characteristic features demonstrated in the ^1H

spectrum. One possible characteristic of the disulfide-bridged residues is that NOE's could occur between protons of the linked residues. This is the case for four of the six remaining three-spin systems. Spin systems A6 and A11 have numerous NOE's between their protons (see supplementary material). The same is also seen between A7 and B7. For the two remaining three-spin systems, A20 and B19, the $C^{\alpha}H$ and $C^{\beta}H$ shifts were too similar to determine the existence of these interresidue NOE's. The pairing of the four spin systems was used to designate these as cysteine with a disulfide bridge to its NOE partner.

The assignment of Ile A2 required the use of numerous 2D spectra to establish its connectivities, as no $C^{\alpha}H-C^{\beta}H$ or $C^{\beta}H-C^{\gamma}H$ cross peaks could be found in the 27 °C 2QF-COSY. At that stage the residue consisted of three spin system fragments, a $C^{\beta}H-C^{\gamma}H_3$ piece, a $C^{\gamma}H_2-C^{\delta}H_3$ piece, and an unassigned $NH-C^{\alpha}H$ cross peak. The ^2H_2O TOCSY spectrum established a direct link between the $C^{\alpha}H$ shift (3.78 ppm) and the $C^{\beta}H$ (1.12 ppm) and $C^{\gamma}H_3$ (0.71 ppm) shifts. The amide at 8.43 ppm had NOE's to the $C^{\gamma}H_2-C^{\delta}H_3$ portion of the side chain, helping to tentatively assign the ethyl portion of the side chain to the rest of the residue. This charting of the protons within A2 was later confirmed by analysis of the 37 °C spectra where all the COSY cross peaks are visible.

For three of the leucine residues, A13, B15, and B17, no COSY cross peaks were found to connect the $C^{\alpha}H-C^{\beta}H_2$ portion of the spin system to the corresponding $C^{\gamma}H-(C^{\delta}H_3)_2$ portions, though sufficient numbers of each type of cross peak were found. Neither the relay-COSY nor the TOCSY provided any clear $C^{\alpha}H-C^{\gamma}H$ or $C^{\beta}H-C^{\delta}H_3$ connections. This occurs due to the unfortunate overlap of the $C^{\beta}H$ and $C^{\gamma}H$ chemical shifts within each residue. Despite the lack of single-relay cross peaks, double-relay cross peaks did occur in the relay-COSY experiment between $C^{\alpha}H$ and $C^{\delta}H_3$ shifts for all three of the leucine spin systems being discussed here. Such relays have been observed by others (Chazin et al., 1988) and are due to strong coupling effects between the $C^{\beta}H$ and $C^{\gamma}H$ resonances (Kay et al., 1987).

At this place in the assignment procedure, seven long-chain spin systems remained, which corresponded exactly to the number of glutamines and glutamic acids in human insulin. These $C^{\alpha}H-C^{\beta}H_2$ fragments were established through the 2QF-COSY and the 2Q spectra. The attachment of a backbone amide proton (Figure 2) was straightforward for all these spin systems, except for B13 and B21, which required analysis of the 37 °C spectra.

The classification of a long-chain spin system as a glutamine or a glutamic acid required the appending of $C^{\gamma}H$ protons which generally have a lower field position than their corresponding $C^{\beta}H$ signals. Due to the number of five-spin systems and the similarity of their $C^{\beta}H$ and $C^{\gamma}H$ shifts, it was not possible to reliably assign the complicated region where these COSY cross peaks are found. Alternatively, the relay-COSY and TOCSY turned out to be most useful, giving relay cross peaks between the $C^{\alpha}H$ and $C^{\gamma}H$ protons (see supplementary material) for five of the seven long-chain systems. For A5 and the A15 the presence of $C^{\alpha}H-C^{\gamma}H$ relay peaks was indefinite, so connections to the $C^{\gamma}H$ protons could be accomplished only through intraresidue and sequential NOE's.

The search for side-chain amides on the three insulin glutamines did not proceed until the sequential assignments were completed. The ^1H_2O and 2QF-COSY and NOESY revealed three remaining side-chain NH_2 groups which had not been assigned to asparagine residues. The attachment of amide signals proceeded by identifying NOE's between side-chain amide and $C^{\gamma}H$ signals. For A5 and B4 this was straightforward as $C^{\gamma}H$ positions were previously assigned. The final NH_2 pair was assigned to A15, partially by default, though it clearly had NOE's to its $C^{\beta}H$.

The pH meter reading for both ^2H_2O and ^1H_2O samples was 3.6, which gives a small difference in pH (0.4) between the two kinds of samples. Despite the potential for pH-dependent shift changes, a comparison of the $C^{\alpha}H$ and $C^{\beta}H$ resonances in both solvents showed no measurable differences.

A considerable effort was made to sort all resonances into residue spin systems and to identify the residue type. This information proved critical for determining that a one-to-one correspondence was occurring under these conditions between the 51 amino acid residues in the insulin sequence and the 51 spin systems found in the NMR spectra, i.e., that each residue in insulin gives only one set of NMR signals. In addition, such a complete spin system identification was extremely helpful in eliminating ambiguities during the sequential assignment process.

Sequence Assignment of the Spin Systems. The sequential NOE's reported here are primarily those observed in the 27 °C NOESY spectra. After these assignments were made, the 37 °C NOESY spectra were also analyzed for data which helped to remove ambiguities in the 27 °C spectra. So some of the NOE cross-peak assignments reflect the additional data found in the higher temperature NOESY.

As would be expected for the large amount of helix in this protein (Chawdhury et al., 1983; Baker et al., 1988), the insulin NOESY spectrum is heavily populated with cross peaks around the amide diagonal (Figure 3). Those readily visible d_{NN} NOE's, used to establish sequential connectivity, are shown with connecting lines. For the A chain (Figure 3A) d_{NN} connectivities are found between residues 2 and 3, from 6 to 9, from 13 to 16, and from 17 to 20. The large break in the NOE pattern between 3 and 6 results from the unfavorable overlap of the amide proton shifts in this region. The disruption of the pattern near residue A10 results more from a lack of an NOE than from an inability to assign the spectral region. The final break occurs at residue 16 and is easily seen from Figure 3A to be the overlap of the amide shifts for 16 and 17.

The d_{NN} cross peaks for the B chain are illustrated in both panels B and C of Figure 3. The NOE pattern can be followed by residue 4 to 8 (Figure 3B), from residue 9 to 20 (Figure 3C), from residue 21 to 27 (Figure 3B), and between residues 29 and 30 in Figure 3C. All of the breaks that occur in these d_{NN} assignments along the B chain are caused by overlap problems, except for proline 28 where an amide proton does not exist. The coincidence of cross peaks 23–24 and 26–27 was deduced by examination of the 37 °C NOESY.

Sequential $d_{\alpha N}$ cross peaks do occur for insulin; however, these cross peaks are generally weaker than the d_{NN} peaks. A few notable exceptions are documented in Figure 4. Five isolated $d_{\alpha N}$ peaks found for the A chain are shown in Figure 4A along with a series from residue 13 to 16. The cross peak between 13 and 14 is overlapped with another cross peak and had to be verified by using the 37 °C NOESY spectrum.

Consistent with the depiction of extended structure from the X-ray model, the B chain has more $d_{\alpha N}$ cross peaks. A relatively long series of NOE's are shown in Figure 4B for residues 16–24. Figure 4C contains three shorter strings, 1–4, 7–10, and 11–13. The close proximity of the amide signals for Gly B8 and Ser B9 produces fused cross peaks which were best analyzed by using a resolution-enhanced spectrum.

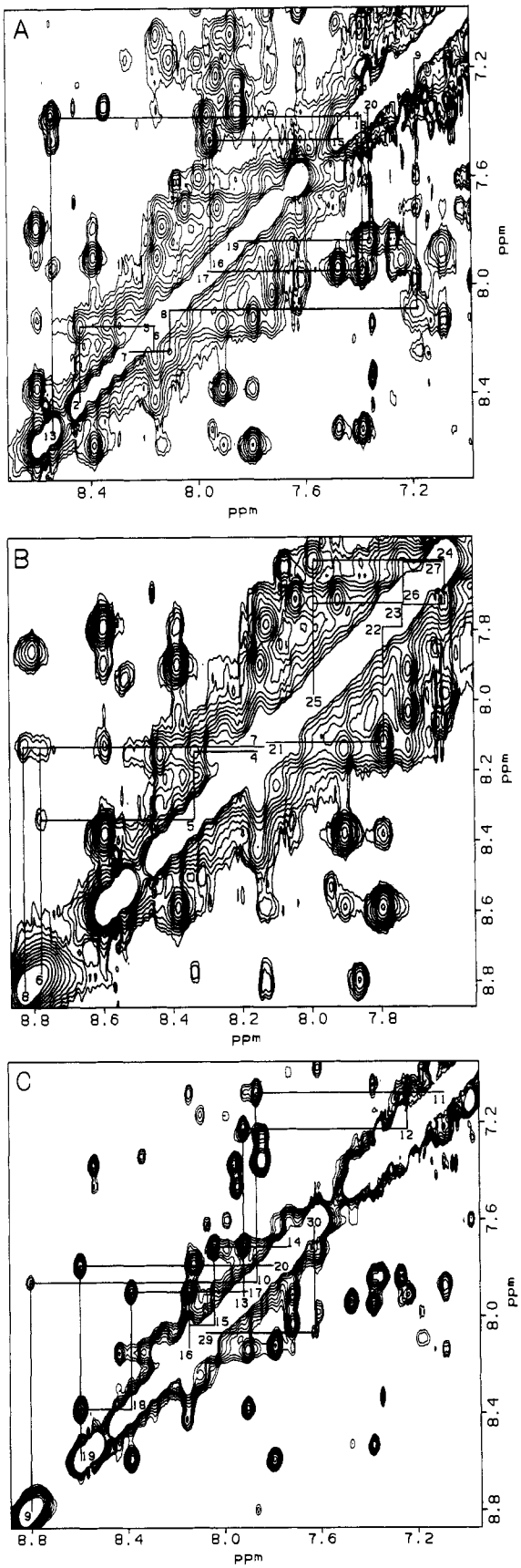

FIGURE 3: Amide diagonal region of a 200-ms NOESY spectrum of human insulin at 27 °C dissolved in ^1H_2O/acetonitrile. Region A shows the sequential d_{NN} cross peaks for the A chain, while regions B and C show the sequential cross peaks for the B chain. The spectra were plotted for both ω_1 (vertical axis) and ω_2 (horizontal axis) from (A) 8.71 to 6.98 ppm, (B) 8.87 to 7.53 ppm, and (C) 8.88 to 6.93 ppm. Only positive contour levels have been drawn. Sequence positions are shown along the diagonal with lines connecting the appropriate cross peaks.

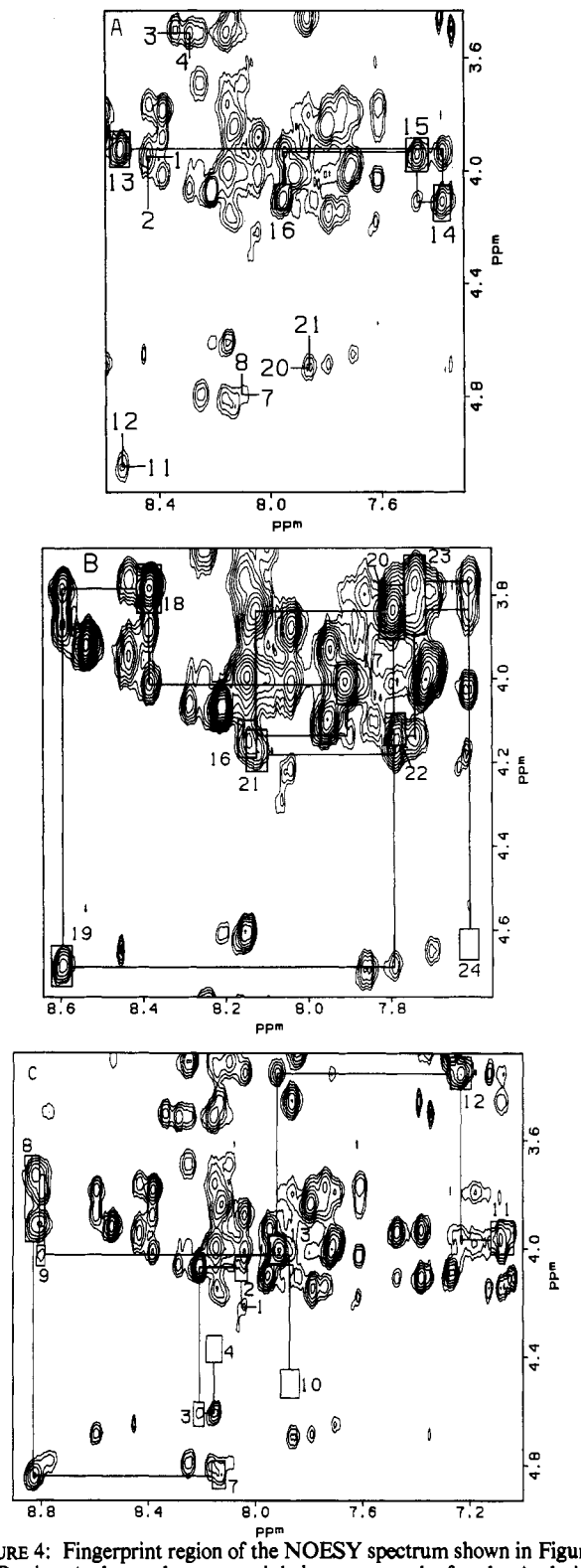

FIGURE 4: Fingerprint region of the NOESY spectrum shown in Figure 3. Region A shows the sequential $d_{\alpha N}$ cross peaks for the A chain, while regions B and C show the sequential cross peaks for the B chain. The spectra were plotted for ω_1 (vertical axis) from (A) 5.13 to 3.43 ppm, (B) 4.76 to 3.69 ppm, and (C) 4.93 to 3.27 ppm and for ω_2 (horizontal axis) from (A) 8.59 to 7.30 ppm, (B) 8.64 to 7.55 ppm, and (C) 8.90 to 7.00 ppm. Only positive contour levels have been drawn. Where a series of sequential $d_{\alpha N}$ cross peaks occur, the COSY cross-peak position is indicated by a box and is labeled with the sequence number. Lines connect the boxes with the appropriate NOESY cross peaks. For those connections not a part of a series, the sequence position is given directly at the cross peak.

A summary of all the d_{NN} and $d_{\alpha N}$ sequential connectivities for both chains in human insulin is shown in Figure 5. Also shown are the $d_{\beta N}$ connectivities which helped in establishing

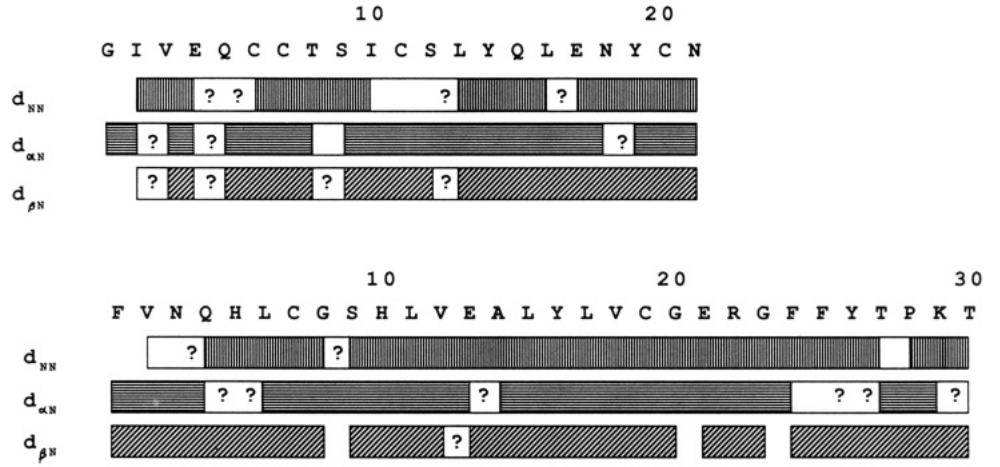

FIGURE 5: Amino acid sequence of human insulin and a survey of the sequential connectivities used in establishing the ^{1}H NMR assignments. The vertically hatched blocks represent sequential d_{NN} connections; horizontally hatched blocks represent sequential $d_{\alpha N}$ connections; diagonally hatched blocks represent $d_{\beta N}$ connections. A blank block indicates that no NOE was detected, while a question mark denotes a position where the existence of an NOE is uncertain. For proline no such amide protons exist, so the $C^{\delta}H$ protons were used instead.

Table I: ^{1}H Chemical Shifts (±0.01) in ppm for Human Insulin at pH = 3.6, 27 °C, Dissolved in 65%/35% Water/Acetonitrile

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	others		residue	NH	$C^{\alpha}H$	$C^{\beta}H$	others	
Gly A1	n.o.[c]	3.99,[a] 3.92[a]								$C^{\delta}H_3$	0.83, 0.72
Ile A2	8.43	3.78	1.12	$C^{\gamma}H$	1.09, 0.90	Cys B7	8.14	4.86	3.16, 2.91		
				$C^{\gamma}H_3$	0.71	Gly B8	8.82	3.93, 3.74			
				$C^{\delta}H_3$	0.57	Ser B9	8.80	4.04	3.80,[a] 3.80[a]		
Val A3	8.16	3.53	1.91	$C^{\gamma}H_3$	0.92, 0.85	His B10	7.86	4.49	3.49, 3.23	$C^{\delta}H$	7.40
Glu A4	8.29	4.09	2.08, 1.95	$C^{\gamma}H$	2.49, 2.27					$C^{\epsilon}H$	8.59
Gln A5	8.15	4.01	2.04,[a] 2.00[a]	$C^{\gamma}H$	2.44,[b] 2.33[b]	Leu B11	7.08	3.97	1.84, 1.20	$C^{\gamma}H$	1.35
				$N^{\epsilon}H$	7.44, 6.73					$C^{\delta}H_3$	0.78, 0.73
Cys A6	8.17	4.82	3.19, 2.86			Val B12	7.23	3.37	2.04	$C^{\gamma}H_3$	0.95, 0.91
Cys A7	8.24	4.79	3.71, 3.32			Glu B13	7.91	4.02	2.13, 2.04	$C^{\gamma}H$	2.48,[a] 2.44[a]
Thr A8	8.10	4.02	4.36	$C^{\gamma}H_3$	1.19	Ala B14	7.72	4.03	1.43		
Ser A9	7.17	4.66	3.98, 3.80			Leu B15	8.04	3.90	1.48, 1.14	$C^{\gamma}H$	1.51
Ile A10	7.70	4.45	1.51	$C^{\gamma}H$	1.12, 0.50					$C^{\delta}H_3$	0.68, 0.57
				$C^{\gamma}H_3$	0.64	Tyr B16	8.15	4.16	3.09,[a] 3.09[a]	$C^{\delta}H_2$	7.08
				$C^{\delta}H_3$	0.57					$C^{\epsilon}H_2$	6.72
Cys A11	9.39	5.07	3.32, 2.99			Leu B17	7.90	4.02	1.88, 1.60	$C^{\gamma}H$	1.83
Ser A12	8.53	4.57	4.23, 3.91							$C^{\delta}H_3$	0.92, 0.90
Leu A13	8.54	3.94	1.51, 1.40	$C^{\gamma}H$	1.53	Val B18	8.38	3.80	2.08	$C^{\gamma}H_3$	0.99, 0.86
				$C^{\delta}H_3$	0.86, 0.78	Cys B19	8.60	4.70	3.19, 2.87		
Tyr A14	7.38	4.14	2.97, 2.87	$C^{\delta}H_2$	7.05	Gly B20	7.79	3.85,[a] 3.85[a]			
				$C^{\epsilon}H_2$	6.80	Glu B21	8.13	4.20	2.14, 2.05	$C^{\gamma}H$	2.48, 2.43
Gln A15	7.47	3.96	2.33, 1.96	$C^{\gamma}H$	2.42,[a] 2.33[b]	Arg B22	7.79	4.16	1.93, 1.85	$C^{\gamma}H$	1.71,[a] 1.66[a]
				$N^{\epsilon}H$	7.37, 6.85					$C^{\delta}H$	3.19, 3.19[b]
Leu A16	7.94	4.11	1.86, 1.59	$C^{\gamma}H$	1.70					$N^{\epsilon}H$	7.29
				$C^{\delta}H_3$	0.80, 0.75	Gly B23	7.74	3.87, 3.77			
Glu A17	7.96	4.13	2.06, 1.96	$C^{\gamma}H$	2.52, 2.33	Phe B24	7.60	4.64	3.02, 2.86	$C^{\delta}H_2$	6.98
Asn A18	7.38	4.41	2.56, 2.45	$N^{\delta}H$	7.09, 6.34					$C^{\epsilon}H_2$	7.13
Tyr A19	7.84	4.40	3.29, 2.91	$C^{\delta}H_2$	7.28					$C^{\zeta}H$	7.13[b]
				$C^{\epsilon}H_2$	6.71	Phe B25	7.99	4.56	3.07, 2.94	$C^{\delta}H_2$	7.19
Cys A20	7.35	4.72	3.20, 2.78							$C^{\epsilon}H_2$	7.28
Asn A21	7.86	4.52	2.75, 2.61	$N^{\delta}H$	7.35, 6.50					$C^{\zeta}H$	7.22[a]
Phe B1	n.o.	4.24	3.19, 3.10	$C^{\delta}H_2$	7.24	Tyr B26	7.73	4.56	2.96,[a] 2.87[a]	$C^{\delta}H_2$	7.02
				$C^{\epsilon}H_2$	7.37					$C^{\epsilon}H_2$	6.73
				$C^{\zeta}H$	7.32[a]	Thr B27	7.61	4.54	4.04	$C^{\gamma}H_3$	1.15
Val B2	8.05	4.08	1.94	$C^{\gamma}H_3$	0.86,[a] 0.86[a]	Pro B28		4.31	2.20, 1.89	$C^{\gamma}H$	1.97, 1.90
Asn B3	8.21	4.62	2.76, 2.66	$N^{\delta}H$	7.42, 6.73					$C^{\delta}H$	3.63, 3.63[b]
Gln B4	8.15	4.36	2.08, 1.78	$C^{\gamma}H$	2.21, 2.21[b]	Lys B29	8.07	4.33	1.84, 1.72	$C^{\gamma}H$	1.42,[a] 1.41[a]
				$N^{\epsilon}H$	7.29, 6.59					$C^{\delta}H$	1.63, 1.63[b]
His B5	8.35	4.42	3.53, 3.22	$C^{\delta}H$	7.36					$C^{\epsilon}H$	2.92, 2.92[b]
				$C^{\epsilon}H$	8.48					$N^{\zeta}H_3$	7.39
Leu B6	8.77	4.44	1.67, 0.98	$C^{\gamma}H$	1.53	Thr B30	7.63	4.20	4.24	$C^{\gamma}H_3$	1.12

[a] Uncertainty in the chemical shift value is ±0.03 ppm. [b] Assignment is not definitive; that is, it relied on NOE data or the lack of cross peaks to other shift values. [c] n.o. = not observed.

sequential neighbors. Considering only the NOESY spectra collected at 27 °C, there were numerous ambiguities present, particularly in the first part of the A chain. The problem was critical for Gln A5, which had no clearly assignable sequential NOE to either Glu A4 or Cys A6. Even with these ambiguities only one arrangement of spin systems to amino acid sequence fits completely with the NOE's identified. That arrangement was used to produce a complete listing of proton resonances and their assignments for human insulin at 27 °C (Table I).

By assigning the 37 °C NOESY spectrum as well, it was possible to gain additional sequential connections due to shifting in the amide resonances. These new connections did

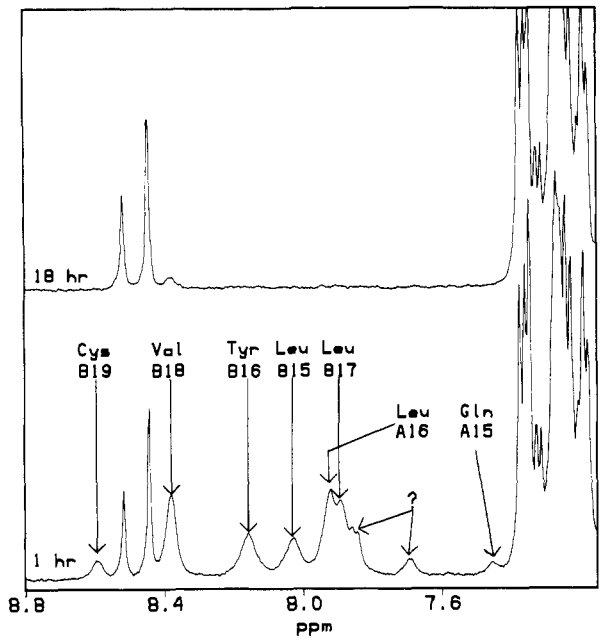

FIGURE 6: 1D NMR spectra of human insulin at 27 °C in completely deuterated solvents showing the region 8.80–7.16 ppm. The length of time the protein spent exposed to solvent is shown for both spectra. Each spectrum was collected, processed, and plotted in an identical manner. Amide proton assignments are shown on the lower spectrum.

not contradict those obtained previously and so were included in Figure 5. With the use of longer NOE mixing times and multiple sample temperatures, it was possible to obtain 82% of all possible d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$ connections. Most of the missing connections are due to cross-peak overlap.

One COSY cross peak was consistently found in all spectra but was not included in Table I. It has a similar appearance to and turns up in close proximity with the Thr B27 $C^{\beta}H$–$C^{\gamma}H_3$ cross peak. The intensity of the shadow peak is only about 5% of the main peak at 27 °C but increases to 10% when the temperature is raised to 37 °C. No other NMR information was available on these shift positions. Though a definitive assignment was not possible, it was noted that the shadow peak moved in concert with Thr B27 $C^{\beta}H$–$C^{\gamma}H_3$ as the sample conditions were altered.

Identification of Slowly Exchanging Amide Protons. In ^2H_2O samples, a number of amide signals were detectable within 12 h of the sample preparation, though after 24 h no such signals were ever encountered. This is illustrated in Figure 6 on a ^2H_2O sample, 1 and 18 h after preparation. The identity of these slowly exchanging amide signals was determined by analysis of a NOESY spectrum collected after the 1-h spectrum had been taken. Amide resonances were determined by reference to the assigned chemical shifts or by the presence of assigned NOESY cross peaks. Seven amide protons were assigned in this fashion and are identified in the figure. From the 1-hour spectrum (Figure 6), it appears that two or three additional signals occur. These signals did not have unique amide proton shifts, nor did they give identifiable cross peaks, so assignments were not possible.

Discussion

A noticeable feature of insulin NMR spectra is the presence of a few extremely broad signals. This may be attributed to the occurrence of multiple states, which are exchanging on an intermediate NMR time scale (Sandström, 1982). This type of behavior is not unusual for proteins and has been observed in NMR spectra before. At the assignment stage, however, the concern is whether each amino acid residue in the protein is represented by a single set of signals. The assumption of a one-to-one correspondence between residues and NMR spin systems makes the sequential assignment task much easier. With the possibility of multiple states existing for insulin, however, such an assumption was avoided. This resulted in additional efforts being made to identify all the spin systems' members and ultimately to establish their residue type before proceeding to the sequential assignment stage. It was found that a good correspondence existed between the 51 residues in human insulin and the 51-residue spin systems from the NMR data.

With such complete knowledge of the spin systems' amino acid type and with full confidence in the amino acid sequence for human insulin, making assignments using sequential NOE's was greatly expedited. The high percentage of sequential connections observed (Figure 5) attests to this. The ambiguities that arose due to chemical shift overlap were usually resolved by reference to the sequence or by examining the similar peaks at the other temperature.

The unassigned shadow peak of Thr B27 has been tentatively assigned as a second $C^{\beta}H$–$C^{\gamma}H_3$ peak of Thr B27, though no further evidence has been identified. This assignment is based solely on the minor peak moving in concert with Thr B27 $C^{\beta}H$–$C^{\gamma}H_3$ peak as the solvent conditions were changed. The small percentage of the minor component suggests that prospect of a proline cis–trans isomerization (Evans et al., 1987; Chazin et al., 1989). This explanation is particularly appealing because of the presence of a neighboring proline residue at position B28.

The necessity of adding organic solvent to remove or reduce the self-association of insulin has been noted by others (Fredericq, 1957; Weiss et al., 1989). A direct study of the insulin monomer under aqueous conditions is not presently possible with NMR spectroscopy as the dissociation constant is near micromolar concentrations (Jeffrey & Coates, 1966; Paselk & Levy, 1974). Despite the presence of the organic solvent, the spectra suggest that the protein has retained its folded conformation.

The secondary structure that the protein adopts has a major influence on the types of sequential NOE's found (Wüthrich et al., 1982; Wüthrich, 1986). These NOE patterns are often used to identify the type of secondary structure the protein has in solution (Wüthrich et al., 1984; Wüthrich, 1986). The primary purpose of this study was to establish the chemical shift assignments for the 286 proton resonances in human insulin. The NOESY spectra collected to accomplish that goal had relatively long mixing times (200 ms). These long mixing times compromise their use for carefully determining the secondary structures in a protein as the peak intensities are complicated by spin diffusion. Despite this, the NOE patterns do reveal some general features about insulin under these conditions.

The type and size of sequential NOE cross peaks predicted from the X-ray model compare well with those observed. The only major discrepancy occurs in the C-terminal region of the B chain. The various X-ray models all display an extended structure for this region, which is involved in intermolecular interactions across the dimer interface. Such a conformation in solution would be expected to give a series of strong $d_{\alpha N}$ connectivities and weak or nonexistent d_{NN}. This is not what is observed in the NMR spectra. It is the $d_{\alpha N}$ that are weak or nonexistent (no $d_{\alpha N}$ cross peak is detected between B24 and B25). In contrast, the d_{NN} peaks in this region are clearly visible, though not as intense as the NOE cross peaks of the B chain central helix. What can be concluded from these data is that the extended structure seen in the X-ray models is not

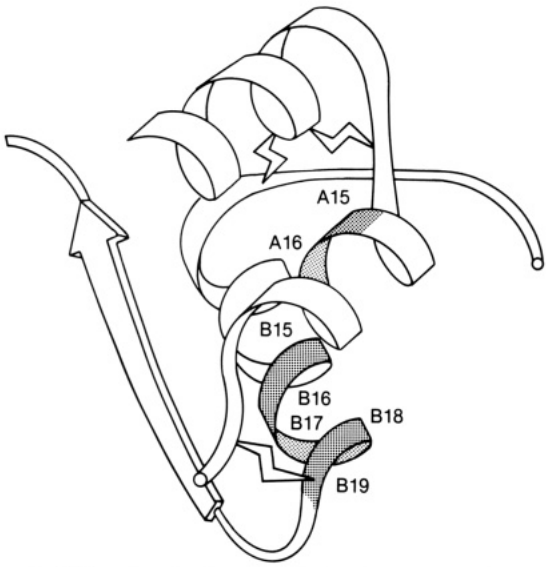

FIGURE 7: Ribbon drawing, based on the insulin X-ray model, showing the residues that have slowly exchanging amide proton signals. These residues are shaded and labeled with chain and sequence number.

the predominant conformation under these solution conditions. X-ray studies indicate that this region is strongly involved in interactions forming the dimer. Our choice of solvent may have reduced these interactions and may account for the conformational differences seen.

Hydrogen exchange studies utilizing assigned NMR signals of amide protons provide site-specific identification of the extent of solvent exposure within proteins (Wagner, 1983b; Englander & Kallenbach, 1984). A slowed exchange rate for amide proton is almost always associated with hydrogen bonding of that proton and often occurs as a result of secondary structure formation (Wagner & Wüthrich, 1982; Wang et al., 1987; Otting et al., 1987; Redfield & Dobson, 1988). When examined over an entire region, the exchange rates give an impression of the stability of that portion of secondary structure (Wagner, 1983b). For this study no exchange rates were measured, but the amide protons were distinguished by whether they had exchanged within 1 h. This course classification allows only the most stable parts of the protein to be identified.

The identification of a series of five amide protons from the C-terminal of the central B-chain helix (Figure 6) indicates that this portion of the molecule (Figure 7) has the most stable intramolecular hydrogen bonds. These data agree remarkably well with a recent neutron diffraction study of hydrogen exchange from porcine crystals (Wlodawer et al., 1989). Such agreement over widely varying conditions strongly suggests the importance of this particular secondary structure to the overall stability of the protein. It also implies that the protein conformation in hexamer crystals and mixed aqueous solutions must be very similar, at least in this area.

A smaller region on the second helix in the A chain is also implicated by the NMR amide proton exchange as having hydrogen bonds. But for A15 and A16, which shows slow exchange in solution, the neutron diffraction study indicated that these sites were not protected from the solvent in crystalline insulin (Wlodawer et al., 1989). Differences must occur on the dynamic, if not on the static, conformation in this region for the less aggregated solution state of the protein to protect these amide sites better than in the hexameric crystalline state. The reason for this additional protection is not obvious.

Some discrepancies are seen between the exchange data from NMR and early exchange studies monitored by infrared spectroscopy (Coleman & Willumsen, 1969). This paper reported that one-fourth of all the amide protons were still nonexchanged after 24 h at 21 °C and pH 3.7. The NMR conditions have reached this level of exchange after only 1 h. This time frame discrepancy may be accounted for by the differences between the experimental conditions. The NMR conditions were developed to minimize the aggregation effects while the infrared study was run at conditions where significant self-association occurs. Additional experiments are needed to understand the importance of solvent conditions on amide exchange rates.

This report represents our initial efforts to understand the solution behavior of insulin, which necessarily must start with the acquisition of reliable NMR assignments. These efforts have already provided initial glimpses at the secondary structures in solution and their stability. In a further publication we will explore the dynamic conditions which are the source of the broad lines observed in the NMR spectra.

Acknowledgments

We thank Drs. Ron Chance and Bruce Frank of Lilly Research Laboratories for providing purified human insulin and especially Dr. Frank for his constructive evaluation of the manuscript. We acknowledge helpful discussions with Dr. Jim Shields, Lilly Research Laboratories, and Dr. Mike Weiss, Harvard Medical School. We also thank Prof. Jane Richardson of Duke University for providing ribbon drawings of the porcine insulin.

Supplementary Material Available

A series of 1D spectra revealing the effects of increasing amounts of acetonitrile (Figure S1), an expanded portion of a ^2H_2O NOESY spectrum with cysteine interresidue NOE's labeled (Figure S2), and a section of a relay-COSY demonstrating $C^{\alpha}H$ to $C^{\delta}H$ cross peaks (Figure S3) (3 pages). Ordering information is given on any current masthead page.

Registry No. Human insulin, 11061-68-0; insulin, 9004-10-8; acetonitrile, 75-05-8.

References

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Timmer, R., & Sheet, S. (1969) *Nature 224*, 491–496.

Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., & Vijayan, N. M. (1988) *Philos. Trans. R. Soc. London 319*, 369–456.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bax, A., & Drobny, G. (1985) *J. Magn. Reson. 61*, 306–320.

Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., & Mercola, D. (1976) *Nature 261*, 166–168.

Bradbury, J. H., Ramesh, V., & Dodson, G. (1981) *J. Mol. Biol. 150*, 609–613.

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

Brown, H., Sanger, F., & Kitai, R. (1955) *Biochem. J. 60*, 556–565.

Chang, W., Stuart, D., Dai, J., Todd, R., Zhang, J., Xie, D., Kuang, B., & Liang, D. (1986) *Sci. Sin. B29*, 1273–1284.

Chawdhury, S. A., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Tolley, S. P., Blundell, T. L., Cleasby, A., Pitts, J. E., Tickle, I. J., & Wood, S. P. (1983) *Diabetologia 25*, 460–464.

Chazin, W. J., Hugli, T. E., & Wright, P. E. (1988) *Biochemistry 27*, 9139–9148.

Chazin, W. J., Kordel, J. Drakenberg, T., Thulin, E., Brodin, P., Grundström, T., & Forsen, S. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2195–2198.
Coleman, D. L., & Willumsen, L. (1969) *C. R. Trav. Lab. Carlsberg 37*, 1–20.
Dodson, E. J., Dodson, G. G., Lewitova, A., & Sabesan, M. (1978) *J. Mol. Biol. 125*, 387–396.
Eich, G., Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 3731–3732.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521–655.
Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford.
Evans, P. A., Dobson, C. M., Kautz, R. A., Hatfull, G., & Fox, R. O. (1987) *Nature 329*, 266–268.
Fesik, S. W. (1988) *Nature 332*, 865–866.
Fredericq, E. (1957) *J. Am. Chem. Soc. 79*, 599–601.
Hau, Q. X., Chen, Y. J., Wang, C. C., Wang, D. C., & Roberts, G. C. K. (1989) *Biochim. Biophys. Acta 994*, 114–120.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Jeffrey, P. D., & Coates, J. H. (1966) *Biochemistry 5*, 3820–3824.
Kay, L. E., Jones, P. J., & Prestegard, J. H. (1987) *J. Magn. Reson. 72*, 392–396.
LeMaster, D. M., & Richards, F. M. (1988) *Biochemistry 27*, 142–150.
Liang, D., Stuart, D., Dai, J., Todd, R., You, J., & Lou, M. (1985) *Sci. Sin. B28*, 472–483.
Lowry, D. F., Redfield, A. G., McIntosh, L. P., & Dahlquis, F. W. (1988) *J. Am. Chem. Soc. 110*, 6885–6886.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Meienhofer, J. E., Schnabel, E., Bremer, H., Brinkhoff, O., Zabel, R., Sroka, W., Klostermeyer, H., Brandenburg, D., Okuda, T., & Zahn, H. (1963) *Z. Naturforsch. 18*, 1120–1121.
Otting, G., Steinmetz, W. E., & Wüthrich, K. (1987) *Eur. J. Biochem. 168*, 609–620.
Palmieri, R., Lee, R. W. K., & Dunn, M. F. (1988) *Biochemistry 27*, 3387–3397.
Paselk, R. A., & Levy, D. (1974) *Biochemistry 13*, 3340–3346.
Pekar, A. H., & Frank, B. H. (1972) *Biochemistry 11*, 4013–4016.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Redfield, C., & Dobson, C. M. (1988) *Biochemistry 27*, 122–136.
Torchia, D. A., Sparks, S. W., & Bax, A. (1988) *Biochemistry 27*, 5135–5141.
Sandström, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, New York.
Wagner, G. (1983a) *J. Magn. Reson. 55*, 151–156.
Wagner, G. (1983b) *Q. Rev. Biophys. 16*, 1–87.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 160*, 343–361.
Wagner G., & Zuiderweg, E. R. P. (1983) *Biochem. Biophys. Res. Commun. 113*, 854–860.
Wang, D., Chang, W., & Dai, J. (1982) *Sci. Sin. B25*, 603–610.
Wang, Q., Kline, A. D., & Wüthrich, K. (1987) *Biochemistry 26*, 6488–6493.
Weiss, M. A., Nguyen, D. T., Khait, I., Inouye, K., Frank, B. H., Beckage, M., O'Shea, E., Shoelson, S. E., Karplus, M., & Neuringer, L. J. (1989) *Biochemistry 28*, 9855–9873.
Westler, W. M., Kainosho, M., Nagao, H., Tomonaga, N., & Markley, J. L. (1988) *J. Am. Chem. Soc. 110*, 4093–4095.
Wider, G., Hosur, R. V., & Wüthrich, K. (1983) *J. Magn. Reson. 52*, 130–135.
Wlodawer, A., Savage, H., & Dodson, G. (1989) *Acta Crystallogr. B45*, 99–107.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.