Article

A Comparison of Three Dimensional Structures of Biosynthesized Preproinsulin and Insulin Using NMR

Mina Oh, K. Hun Mok[1] and Yoongho Lim*

Department of Applied Biology and Chemistry, Konkuk University, Seoul 133-701, Korea
[1]*Biomolecular Structure Research Unit, Korea Research Institute of Bioscience and Biotechnology*

The solution conformation of the human insulin precursor, preproinsulin, is described in terms of NMR spectral data. NMR experiments were performed on preproinsulin, whose structure was compared with the NMR structure of native human insulin. Despite the presence of the C-peptide and/or the signal peptide, secondary structure analyses indicate that the native structures of the A and B chains are well conserved even in preproinsulin. The observed relative robustness of the native structure in precursor forms permits further protein engineering experiments where the C-peptide or N-terminal signal sequence can be altered for the purpose of increasing expression or purification yields when producing recombinant human insulin.

Key words : *insulin, preproinsulin, NMR structure.*

Insulin is the major hormone responsible for regulating blood glucose levels in the human body. Since the pure isolation of insulin by Banting, Best, Collip, and Macleod in 1922, insulin has been used for the treatment of diabetes mellitus[1] and has become one of the most important therapeutic agents in medicine. Diabetes mellitus can be divided into two major clinical classes, insulin-dependent diabetes mellitus (IDDM, type I) and non-insulin-dependent diabetes mellitus (NIDDM, type II). Type I is caused by autoimmune destruction of the pancreatic β-cells leading to a deficiency in insulin secretion. Type II, which accounts for 95% of all cases, is characterized by insulin deficiency or insulin resistance, or both. Insulin is essential for both types I and II.[2,3]

Insulin is a small, globular protein of molecular weight 5700 Da with two polypeptide chains, the A chain (21 residues) and B chain (30 residues). These two are joined by two disulfide bonds at A7-B7 and A20-B19, with a third disulfide bond existing within the A chain, A6-A11. Insulin is synthesized in the pancreatic β-cells as an inactive single-chain precursor, preproinsulin, which has an N-terminal signal peptide that directs its passage into

secretory vesicles. Proteolytic removal of the signal peptide produces proinsulin, which is constituted of the A and B chains and the C peptide, a connecting peptide of 35 residues positioned between N-terminal of the A chain and C-terminal of the B chain. When the elevation of blood glucose triggers insulin secretion, proinsulin is converted into active insulin. Proinsulin is composed of 86 amino acids and preproinsulin, due to the additional 17-residue signal peptide, is composed of 103 amino acids.

Hua and Weiss suggested in 1991 that each structure of the A and B chains is conserved when preproinsulin turns into insulin.[4] This means that the existence of the flexible C chain positioned between the N-terminal of the A chain and the C-terminal of the B chain does not influence the structures of the A and B chains of preproinsulin. If this were to be the case, then the C peptide region of preproinsulin does not necessarily need to be biosynthesized as the intact sequence. In order to commercially produce insulin time- and cost-efficiently, a shorter C chain may be the better solution. To enlighten these possibilities, NMR experiments of native human insulin and biosynthetic preproinsulin were carried out. A comparison of the partial structures of preproinsulin with the three dimensional structure of insulin based on NMR is reported here.

*Corresponding author
Phone: 82-2-450-3760; Fax: 82-2-3436-5776
E-mail: yoongho@kkucc.konkuk.ac.kr

Abbreviations: COSY, correlated spectroscopy; CSI, chemical shift index; NOESY, nuclear Overhauser exchanged spectroscopy; PDB, protein data bank; TOCSY, total correlated spectroscopy; TPPI, time proportional phase incrementation.

Materials and Methods

Sample preparation. Biosynthetic human insulin was

purchased as a zinc-free lyophilized powder from Boehringer Mannheim. Biosynthetically produced preproinsulin was obtained as fusion protein expressed in *Escherichia coli* using recombinant DNA techniques[5,6] and purified. The deuterated solvents for the NMR experiments, D_2O and CD_3COOD, were purchased from Aldrich. These were also used without further purification.

Insulin exists either as a monomer, dimer, hexamer, and/or octamer depending on the solution pH. Insulin is known to exist stably as a monomer below pH 2 by the addition of acetic acid.[7-10] The sample should be kept below pH 2 in order to overcome the self-association of protein and to obtain good spectra.[4] The NMR sample of native human insulin was prepared in 80% D_2O/ 20% CD_3COOD (vol/vol) (pH 1.7). In order to observe all amide protons, a separate sample was prepared in 30% D_2O/20% CD_3COOD/50% H_2O (vol/vol). Sample concentrations were approximately 5 mM in the above solvent conditions. The NMR sample of preproinsulin was prepared in 100% D_2O or 50% D_2O/50% H_2O (vol/vol) (pH 1.75).

NMR experiments. All NMR experiments were carried out on a Bruker DPX 400 (9.4T) spectrometer equipped with a 5 mm dual probe. The probe temperature was maintained at 298 K. In all experiments, solvent suppression was achieved using selective saturation and a proton pulse length of 9.7 μsec. Presaturated one-dimensional NMR spectra[11] were measured using a spectral width of 4006.41 Hz and 32 K of data points. Two dimensional experiments were performed by the time proportional phase incrementation (TPPI) technique in all cases except for the COSY experiments. Two-dimensional COSY, TOCSY and NOESY spectra were acquired with 4096 data points in t2 and 512 t1 increments. 512 t1 FIDs were collected in each experiment with 256 scans per t1 value. Spectral width of 4006.41 Hz was used in all experiments of native human insulin, while 4166.67 Hz (100% D_2O) and 3881.99 Hz (50% H_2O) were used for preproinsulin. The pulse sequence for COSY was magnitude N-type with presaturation during the relaxation delay.[12] The pulse sequence used for TOCSY was an MLEV17 spin-lock where a 2.5 msec trim pulse was employed.[13] Mixing times of 100 msec and 50 msec were used in phase sensitive TOCSY experiments for measuring native human insulin and preproinsulin, respectively. Phase sensitive NOESY spectra were measured using a 200 msec mixing time in native human insulin and 100 msec in preproinsulin. Phase-shifted sine-bell squared functions were applied prior to Fourier transformation. All data were zero-filled to 4 K data points in the t1 dimension.

Results and Discussion

Assignments of NMR data of human insulin. Even though the NMR experiments of native insulin had been previously carried out by Weiss *et al.*,[4] a reinvestigation of the assignments for native human insulin was carried out in order to better compare the structures of preproinsulin and insulin under identical conditions. Initially, 95 cross peaks out of a total of 102 proton resonances in the COSY spectrum were assigned. Based on the COSY assignments, 147 of the 161 cross peaks found in the TOCSY spectrum and 201 of the 223 of NOESY peaks were assigned, respectively. The data is summarized in Table 1. Most of the chemical shifts in Table 1 are in good agreement with the assignments of Hua and Weiss[4] except for B29 Lys. While the chemical shift differences between most of the residues ranged between 0.05~0.20 ppm, that of B29 Lys CεH2 showed a difference of about 0.70 ppm. In addition, it is interesting to note that a previously-overlapping chemical shift for B28 Pro CδH2 and B29 Lys CδH2[4] could be clearly resolved in the present series of experiments, as listed in Table 1.
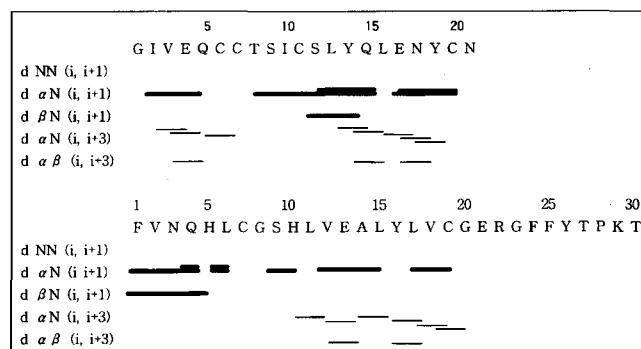
The secondary structure based on the sequence map. The secondary structure of human insulin revealed that there was α-helix in the B chain at B9-B20 and in the A chain at A1-A8 and A13-A20, with β-sheet in the B chain at B20-B23.[4] The presence and location of the α-helices in native human insulin could be established by observation of characteristic $d_{\alpha N(i,i+3)}$ and $d_{\alpha\beta(i,i+3)}$ sequential connectivities. Fig. 1 shows the summary of inter-residual NOE connectivities of native human insulin. As shown in Fig. 1, $d_{\alpha N(i,i+3)}$ NOEs were observed in roughly three regions composed of A2 Ile-A7 Cys, A13 Leu-A20 Cys, and B10 His-B19 Cys. These three regions agree well with the previously-observed helical regions. Helix region I was observed from $d_{\alpha N(i,i+3)}$ NOEs at the positions of A2 Ile-A4 Glu, A3 Val-A5 Gln, and A5 Gln-A7 Cys. Helix region II was observed at A13 Leu-A15 Gln, A14 Tyr-A16 Leu, A 16 Leu-A18 Asn, A17 Glu-A19 Tyr, and A18 Asn-A20 Cys. Helix region III was also observed at B10 His-B12 Val, B12 Val-B14 Ala, B14 Ala-B16 Tyr, B16 Tyr-B18 Val, and B17 Leu-B19 Cys. Interestingly, Cys residues were commonly found on the ends of all helices. Because the $d_{\alpha N(i,i+1)}$ NOEs were observed all over and conversely $d_{NN(i,i+1)}$ NOEs could not be observed at all, the region correspond- ing to the β-sheet conformation could not be determined from comparison of the strong and the weak $d_{\alpha N(i,i+1)}$ and $d_{NN(i,i+1)}$ NOEs. Despite the partially insufficient NOE data which resulted in an incomplete secondary structure assignment, all other secondary structure elements determined from NMR were found to be identical to the X-ray structure (Fig. 2) except for the β-conformation.

Secondary structure based on chemical shift index. The Wishart indices (Table 2) reported in 1992 were used as a guide for the determination of secondary structure.[14] The result of native human insulin analyzed with the Wishart indices is shown in Fig. 3, which is in accord

Table 1. Chemical shifts of the ¹H NMR resonances of native human insulin.

residue	chemical shifts			
	NH	C$_\alpha$H	C$_\beta$H	others
A1 Gly				
A2 Ile	8.37	3.94		Cγ H₃ 0.74
A3 Val	8.13	3.65	1.94	Cγ H₃ 0.86/0.91
A4 Glu	8.14	4.19	2.06	Cγ H₂ 2.49
A5 Gln	8.20	4.04	2.05/2.12	Cγ H₂ 2.481/2.46
A6 Cys		4.86	2.84/3.21	
A7 Cys	8.20	4.80	3.74/3.27	
A8 Thr	8.16	4.04	4.35	Cγ H₃ 1.23
A9 Ser	8.57	4.73	4.00/3.85	
A10 Ile	7.81	4.33	1.52	Cγ H₂ 1.05/0.41
				Cγ H₃ 0.61; Cδ H3 0.48
A11 Cys	9.76	4.81		
A12 Ser	8.70	4.56	4.27/3.97	
A13 Leu	8.60	3.84	1.32/1.39	Cγ H 1.32
				Cδ H₃ 0.71/0.76
A14 Tyr	7.45	4.12	2.87/2.97	C2,6H 7.04; C3,5H 6.82
A15 Gln	7.52	3.95	1.99/2.35	Cγ H₂ 2.37
A16 Leu	8.00	4.14	1.51/1.88	Cγ H 1.70; Cδ H3 0.78
A17 Glu	8.00	4.14	2.04	Cγ H₂ 2.31/2.56
A18 Asn	7.38	4.45	2.52/2.60	
A19 Tyr	7.86	4.42	3.30/2.95	C2,6H 7.30; C3,5H 6.76
A20 Cys		4.70	2.84	
A21 Asn		4.70	2.70	
B1 Phe		4.22	3.13	C2,6H 7.17; C3,5H 7.30
				C4H 7.22
B2 Val	8.08	4.09	1.87	Cγ H₃ 0.83/0.83
B3 Asn	8.48	4.68	2.70	Nε H₂ 6.85/7.49
B4 Gln	8.39	4.45	1.88/2.05	Cγ H₂ 2.20
B5 His	8.66	4.42	3.53/3.22	C2H 8.65; C4H 7.38
B6 Leu	9.04	4.47	1.74	Cγ H₂ 1.57
B7 Cys		4.95	2.94/3.20	
B8 Gly		4.00/3.82		
B9 Ser	8.80	4.07	3.85	
B10 His	7.85	4.47	3.28/3.56	C2H 8.68; C4H 7.45
B11 Leu				
B12 Val	7.22	3.28	2.01	Cγ H₃ 0.89/0.89
B13 Glu				
B14 Ala	7.70	4.04	1.47	
B15 Leu		4.33		Cγ H₂ 1.45
B16 Tyr	8.05	4.20	3.11	C2,6H 7.07; C3,5H 6.72
B17 Leu	7.72	4.04	1.61	Cγ H₂ 1.80
				Cδ H₃ 0.89/0.91
B18 Val	8.30	3.82	2.07	Cγ H₃ 0.99/0.85
B19 Cys		4.74	2.91	
B20 Gly				
B21 Glu				
B22 Arg		4.17	1.94	Cγ H₂ 1.72; Cδ H₂ 3.21
				Nε H₂ 7.05
B23 Gly		3.90/3.78		
B24 Phe		4.20	3.09/3.09	C2,6H 6.90; C3,5H 7.03
B25 Phe	7.59	4.64	3.07/2.98	C2,6H 7.14; C3,5H 7.29
				C4H 7.25
B26 Tyr		4.57	2.87	C2,6H 6.94; C3,5H 6.69
B27 Thr		4.53	4.03	Cγ H₃ 1.15
B28 Pro		4.28	2.21	Cγ H₂ 1.88; Cδ H₂ 3.60
B29 Lys		4.35	1.85/1.75	Cγ H₂ 1.45; Cδ H₂ 1.65
				Cε H₂ 2.95; Nε H₂ 7.52
B30 Thr		4.44	4.35	Cγ H₃ 1.15

Fig. 1. Schematic representation of NOE sequential connectivities of native human insulin.
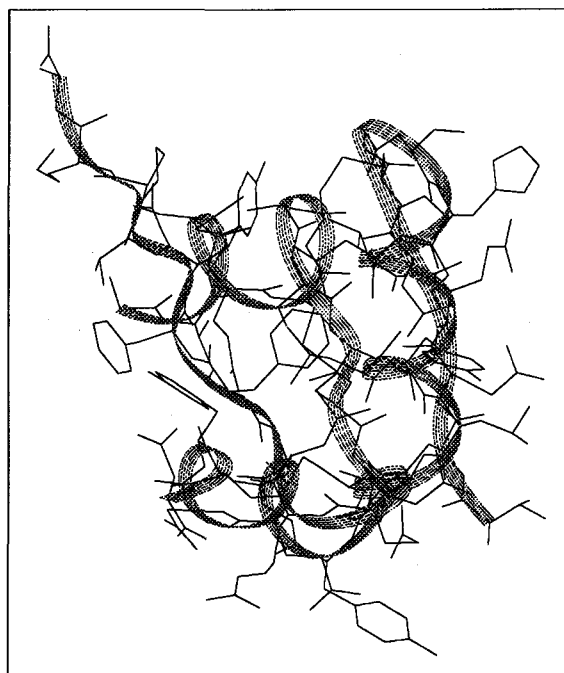
Fig. 2. X-ray crystallographic structure of insulin obtained from PDB (2INS).

Table 2. Chemical shifts of α-proton used for the determination of the secondary structure.

residue	α-¹H range (ppm)	residue	α-¹H range (ppm)
Ala	4.35±0.10	Met	4.52±0.10
Cys	4.65±0.10	Asn	4.75±0.10
Asp	4.76±0.10	Pro	4.44±0.10
Glu	4.29±0.10	Gln	4.37±0.10
Phe	4.66±0.10	Arg	4.38±0.10
Gly	3.97±0.10	Ser	4.50±0.10
His	4.63±0.10	Thr	4.35±0.10
Ile	3.95±0.10	Val	3.95±0.10
Lys	4.36±0.10	Trp	4.70±0.10
Leu	4.17±0.10	Tyr	4.60±0.10

with the known information of the secondary structure of native insulin as described above. Unlike the X-ray structure and NMR structure of Section A, the result obtained based on CSI did show only two helices situated at A13 Leu-A19 Cys and B9 Ser-B20 Cys. The β-sheet confor-
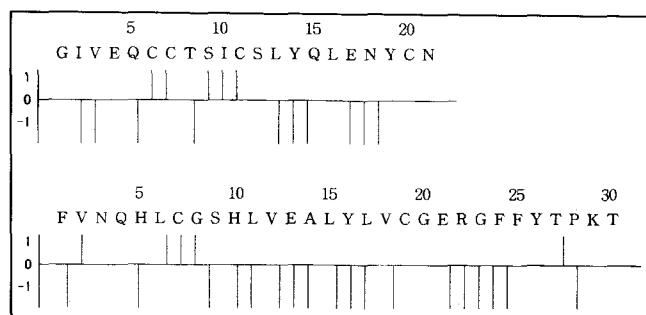
Fig. 3. CSI graph of native human insulin.

Table 3. Chemical shifts of the ¹H NMR resonances of signal peptide of preproinsulin.

residue	chemical shifts			
	NH	C$_\alpha$H	C$_\beta$H	others
S1 Met	7.91	4.14	2.10/2.14	Cγ H$_2$ 2.53
S2 Thr	8.21	4.28	4.11	Cγ H$_3$ 1.10
S3 Met	7.95	4.22	2.04/1.90	Cγ H$_2$ 2.44
S4 Ile	8.30	4.15	1.79	Cγ H$_2$ 1.40; Cγ H$_3$ 0.82
				Cδ H$_3$ 0.76
S5 Thr	8.27	4.29	4.03	Cγ H$_3$ 1.14
S6 Asn	8.40	4.65	2.73	
S7 Ser	8.23	4.23	3.65/3.72	
S8 His		4.55	2.99/2.99	C4H 7.18
S9 His		4.56	3.04/3.04	C4H 7.17
S10 His		4.59	3.07/3.13	C4H 7.10
S11 His		4.60	3.12/3.12	C4H 7.18
S12 His		4.58	3.14/3.12	C4H 7.19
S13 His		4.57	3.17/3.20	C4H 7.18
S14 Lys	8.15	4.17	1.76	Cγ H$_2$ 1.48; Cδ H$_2$ 1.66
				Cϵ H$_2$ 3.09
S15 Phe	8.58	4.46	2.48/1.96	
S16 Arg	8.40	4.16	1.23/1.61	Cγ H$_2$ 1.60; Cδ H$_2$ 2.90
S17 Arg	8.40	4.16	1.29/1.61	Cγ H$_2$ 1.60; Cδ H$_2$ 2.90

mation region of the B chain was nevertheless difficult to determine as in the case above. A possible explanation would be due to the extremely low pH (pH 1.7) used for the present NMR experiments. This pH range is considerably lower than the documented model peptide and protein NMR conditions used in determining the CSI (pH 2.3 to pH 7.3)[14] and hence may not be suitable for a simplistic numerical comparison.

Assignments of NMR data of preproinsulin and its comparison with native insulin using NMR. Assignments of NMR data of preproinsulin were performed with reference to assignments of native insulin and the signal peptide. The assignments of the signal peptide of preproinsulin are listed in Table 3. The primary structure of preproinsulin is shown in Fig. 4. 122 of 139 cross peaks of the COSY spectrum (Fig. 5) were assigned. Based on the COSY assignments, 169 of 184 cross peaks of the TOCSY spectrum were assigned. Its assignments and figures are provided as supplementary materials. The assignments of preproinsulin were performed only in the A and B chains. The chemical shifts and assignments obtained from the COSY and TOCSY spectra are listed in Table 4. The flexible C chain region of preproinsulin could not be observed even in the COSY spectrum. On

the other hand, the chemical shifts of the α-protons of preproinsulin were nearly in accord with those of native insulin. Hence, despite the lack of assignments for the C-peptide of preproinsulin, the resulting spectra of the A and B chains of preproinsulin appears to be extremely similar to that of native insulin. The secondary structure of preproinsulin was determined based on the Wishart indices (Table 2). The CSI of preproinsulin is shown in Fig. 6 which nearly coincides with that of native insulin as shown in Fig. 3. Thus, the patterns of the secondary structure for chains A and B appear well conserved. Fig. 6 shows that the A and B chains of preproinsulin are conserved partly in comparison with X-ray structure and NMR structure of native insulin.
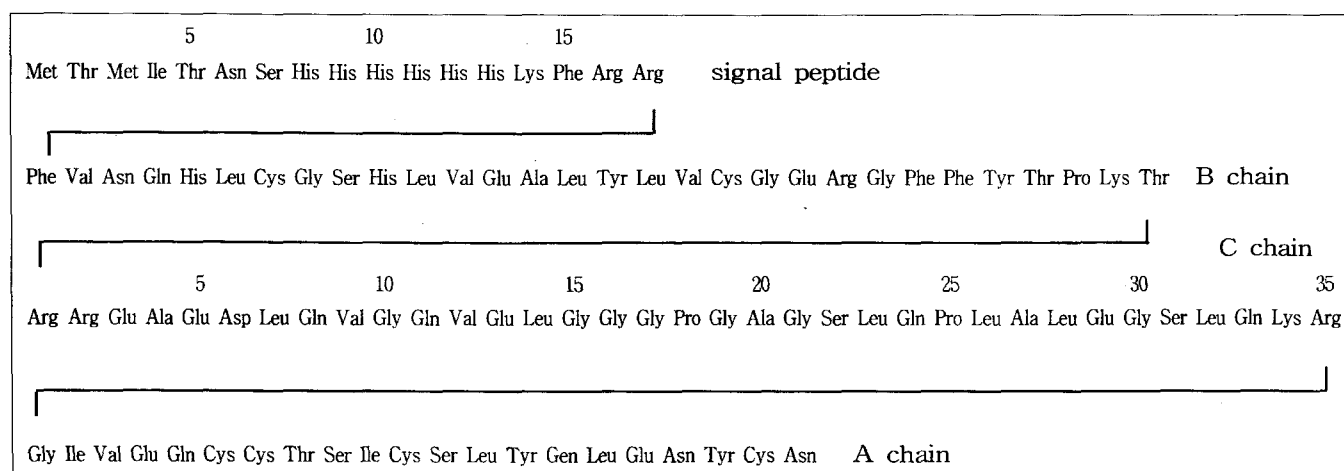

Fig. 4. The primary structure of preproinsulin.

Mina Oh *et al.*

Table 4. Chemical shifts of the 1H NMR resonances of pre-proinsulin.

residue	chemical shifts			
	NH	CαH	CβH	others
A1 Gly				
A2 Ile	8.04	4.04	2.05	Cγ H₃ 0.73
A3 Val		3.60	1.98	
A4 Glu	8.47	4.20	1.98	Cγ H₂ 2.43
A5 Gln		4.12	2.12/2.14	Cγ H₂ 2.50/2.53
A6 Cys				
A7 Cys		4.81		
A8 Thr	8.06	3.84	4.10	Cγ H₃ 1.25
A9 Ser		4.74	3.84/4.03	
A10 Ile		4.26	1.48	Cγ H₃ 0.65; Cγ H₂ 0.81
A11 Cys				
A12 Ser	8.47	4.57	4.29/3.98	
A13 Leu				Cγ H₂ 1.58; Cγ H₃ 0.80
A14 Tyr	8.29	4.11	2.88	C2,6H 7.02; C3,5H 6.78
A15 Gln			1.98/2.08	Cγ H2 2.30
A16 Leu				
A17 Glu				
A18 Asn	8.30	4.42	2.52/2.68	Nε H₂ 6.82/7.52
A19 Tyr		4.44	3.25/290	C2,6H 7.20; C3,5H 6.70
A20 Cys				
A21 Asn				Nε H₂ 6.90/7.58
B1 Phe				
B2 Val	8.04	4.04	2.05	Cγ H₃ 0.91/0.86
B3 Asn				Nε H₂ 6.80/7.46
B4 Gln	8.42	4.52	1.98/2.01	CγH₂ 2.48/2.52 Nε H₂ 6.78/7.46
B5 His		4.42	3.52	C2H 8.42; C4H 7.34
B6 Leu				
B7 Cys				
B8 Gly				
B9 Ser		4.02	3.82	
B10 His		4.44	3.55	C2H 8.42; C4H 7.43
B11 Leu				
B12 Val	8.16	3.20	1.98	Cγ H₃ 0.88
B13 Glu				
B14 Ala		4.00	1.40	
B15 Leu				
B16 Tyr		4.25	3.10	C2,6H 7.03; C3,5H 6.79
B17 Leu				
B18 Val	8.27	3.97	1.94	Cγ H3 0.82
B19 Cys				
B20 Gly				
B21 Glu				
B22 Arg	7.96	4.19	1.62	Cγ H₂ 1.47; Cδ H₂ 3.07 Nε H₂ 7.13
B23 Gly				
B24 Phe		4.13	2.85/2.99	C2,6H 7.08; C3,5H 7.21 C4H 7.23
B25 Phe		4.13	2.98	C2,6H 7.18; C3,5H 7.27 C4H 7.30
B26 Tyr	8.19	4.58	2.85/2.85	C2,6H 6.97; C3,5H 6.64
B27 Thr	8.38	3.92	4.39	
B28 Pro		4.35	2.20	Cγ H₂ 2.02; Cδ H₂ 3.60
B29 Lys	8.13	4.32	2.28	Cγ H₂ 1.43; Cδ H₂ 1.80 Cε H₂ 2.92 ; Nε H₂ 7.55
B30 Thr	8.29	4.24	4.24	Cγ H₃ 1.16
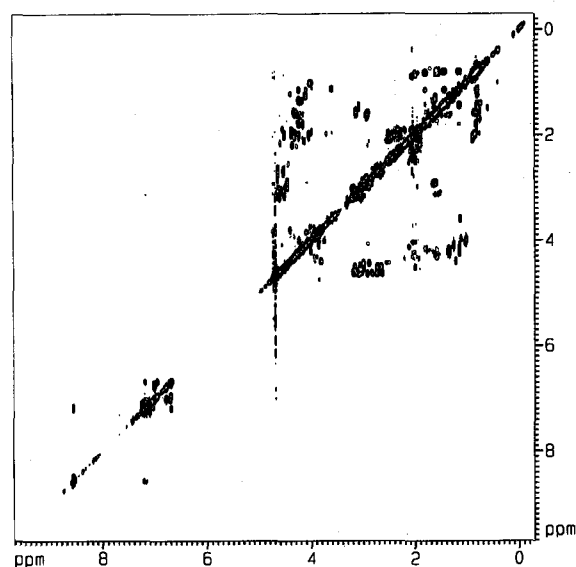
Fig. 5. The COSY spectrum of a 5 mM solution of preproinsulin in D₂O at 300K.
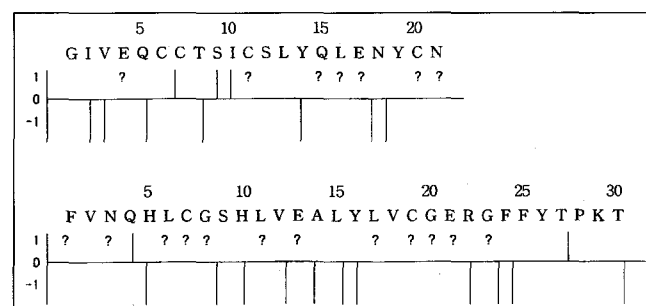
Fig. 6. CSI graph of preproinsulin.

References

1. Born, G. V. R., Cuatrecasas, P. and Herken, H. (1990) The biosynthesis of insulin. In *Handbook of experimental pharmacology*, vol. **92**, pp. 67-92, Springer-Verlag, New York.

2. Taira, M., Hashimoto, N., Shimada, F., Suzuki, Kanatsuka, Y., Nakamura, A. F., Ebina, Y., Tatibana, M., Makino, H. and Yoshida, S. (1989) Human diabetes associated with a deletion of the tyrosine kinase domain of the insulin receptor. *Science* **245**, 63-66.

3. Odawara, M., Kadowaki, T., Yamamoto, R., Shibasaki, Y., Tobe, K., Accili, D., Bevins, C., Mikami, Y., Matsuura, N., Akanuma, Y., Takaku, F., Taylor, S. I. and Kasuga, M. (1989) Human diabetes associated with a mutation in the tyrosine kinase domain of the insulin receptor, *Science* **245**, 66-68.

4. Hua, Q. and Weiss, M. A. (1991) Comparative 2D NMR studies of human insulin and des-pentapeptide insulin:

Sequential resonance assignment and implications for protein dynamics and receptor recognition, *Biochemistry* **30**, 5505-5515.

5. Sung, W. L., Yao, F. L., Zahab, D. M. and Narang, S. A. (1986) Short synthesis oligodeoxyribonucleotide leader sequences enhance accumulation of human proinsulin synthesized in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **83**, 561-565.

6. Kang, Y. and Yoon, J. W. (1991) Development of a high-expression vector (PYK10-9) of human proinsulin gene, *Biotechnol. Lett.* **13**, 755-760.

7. Blundell T. L., Dodson, G. G., Hodgkin, D. C. and Mercola, D. A. (1972) Insulin: The structure in the crystal and its reflection in chemistry and biology. *Adv. Protein Chem.* **26**, 379-402.

8. Blundell T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Mercola, D. A. and Vijayan, M. (1971) Atomic positions in rhombohedral 2-zinc insulin crystals. *Nature* **231**, 506-511.

9. Milthorpe, B. K., Nicol, L. W. and Jeffrey, P. D. (1977) The polymerization pattern of zinc (II) insulin at pH 7. *Biochem. Biophys. Acta* **495**, 195-202.

10. Goldman, J. and Carpenter, F. (1974) Zinc binding insulin dichroism and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. *Biochemistry* **13**, 4566-4574.

11. Bax, A. (1985) A spatially selective composite 90° radiofrequency pulse. *J. Magn. Reson.* **65**, 142-145.

12. Nagayama, K., Kumar, A., Wüthrich, K. and Ernst, R. R. (1980) Experimental techniques of two-dimensional correlated spectroscopy. *J. Magn. Reson.* **40**, 321-334.

13. Bax, A. and Davis, D. G. (1985) MLEV-17-Based two-dimensional homonulear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.

14. Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) The chemical shifts index: A fast and simple methods for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651.