Kozak, M. (1986) *Cell 47*, 481-483.
Kren, B., Parsell, D., & Fuchs, J. A. (1988) *J. Bacteriol. 170*, 308-315.
Laemmli, U. K. (1970) *Nature 227*, 680-685.
LeMaster, D. (1986) *J. Virol. 59*, 759-760.
Lin, T.-Y., & Kim, P. S. (1989) *Biochemistry 28*, 5282-5287.
Manavalan, P., & Johnson, W. C., Jr. (1987) *Anal. Biochem. 167*, 76-85.
Mandel, M., & Higa, A. (1970) *J. Mol. Biol. 53*, 159-162.
Messing, J., & Vieira, J. (1982) *Gene 19*, 269-276.
Morel-Deville, F., Vachon, G., Sacerdot, C., Cozonne, A. J., Grunberg-Manago, M., & Cenatiempo, Y. (1990) *Eur. J. Biochem. 188*, 605-614.
Munson, J. M., Stormo, G. D., Niece, R. L., & Reznikoff, W. S. (1984) *J. Mol. Biol. 177*, 663-683.
Mutter, M. (1977) *J. Am. Chem. Soc. 99*, 8307-8314.
Pace, C. N. (1986) *Methods Enzymol. 131*, 266-280.
Pace, C. N., Grimsley, G. R., Thomson, J. A., & Barnett, B. J. (1988) *J. Biol. Chem. 263*, 11820-11825.
Pace, C. N., Shirley, B. A., & Thomson, J. A. (1989) in *Protein Structure: A Practical Approach*, (Creighton, T. E., Ed.) pp 311-330, IRL Press, New York.
Poland, D. C., & Scheraga, H. A. (1965) *Biopolymers 3*, 379-399.
Privalov, P., Tiktopulo, E., Venyaminov, S., Griko, Yu., Makhatadze, G., & Khechinashvili, N. (1989) *J. Mol. Biol. 205*, 737-750.
Read, S. M., & Northcote, D. H. (1981) *Anal. Biochem. 116*, 53-64.
Russel, M., & Model P. (1984) *J. Bacteriol. 159*, 1034-1039.
Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanyal, G., Kim, E., Thompson, F. M., & Brady, E. K. (1989) *Biochem. Biophys. Res. Commun. 165*, 772-781.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368-379.
Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 230-259.
Schmid, F. X. (1989) in *Protein Structure: A Practical Approach*, (Creighton, T. E., Ed.) pp 251-285, IRL Press, New York.
Smith, R. A., & Parkinson, J. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5370-5374.
Söderberg, B.-O., Sjöberg, B.-M., Sonnerstam, U., & Brändén, C.-I. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5827-5830.
Tabor, S., & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074-1078.
Tartof, K. D., & Hobbs, C. A. (1987) *Focus 9* (2), 12.
Teale, F. W. J. (1960) *Biochem. J. 76*, 381-388.
Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261-287.
Tsang, M. L.-S. (1981) *J. Bacteriol. 146*, 1059-1066.
Wiget, P., & Luisi, P. L. (1978) *Biopolymers 17*, 167-180.
Zuker, M. (1989) *Methods Enzymol. 180*, 262-288.

Solution Structure of Human Insulin-Like Growth Factor 1: A Nuclear Magnetic Resonance and Restrained Molecular Dynamics Study[†,‡]

Robert M. Cooke,[§] Timothy S. Harvey, and Iain D. Campbell*
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.
Received November 14, 1990; Revised Manuscript Received February 8, 1991

ABSTRACT: The solution structure of human insulin-like growth factor 1 has been investigated with a combination of nuclear magnetic resonance and restrained molecular dynamics methods. The results show that the solution structure is similar to that of insulin, but minor differences exist. The regions homologous to insulin are well-defined, while the remainder of the molecule exhibits greater disorder. The resultant structures have been used to visualize the sites for interaction with a number of physiologically important proteins.

Over recent years the requirement of protein growth factors for the control of cell growth and replication has become increasingly apparent, and the number of growth factors known to exist has multiplied dramatically. Most of these proteins are thought to act by binding to cell-surface receptors and activating intracellular tyrosine kinases [reviewed by Yarden and Ullrich (1988)]. Attempts have been made to subdivide the growth factors on the basis of sequence homologies (James & Bradshaw, 1984), and such consideration revealed that one group closely resembled insulin, prompting their renaming as insulin-like growth factors (IGFs)[1] (Rinderknecht & Humbel, 1976). Two distinct IGFs have been found in many mammalian species: IGF-1, the human form of which is also known as somatomedin C (Daughaday et al., 1972), and IGF-2, which was known as multiplication-stimulating activity (Dulak & Temin, 1973). The IGFs are also thought to be responsible for the nonsuppressible insulin-like activity observed in serum (Froesch et al., 1963).

[†] Supported by grants from the Monsanto Chemical Company, ICI Pharmaceuticals, and the Science and Engineering Research Council.
* Author to whom correspondence should be addressed.
[‡] The coordinates for the structures in this paper have been deposited with the Brookhaven Protein Database.
[§] Current address: Glaxo Group Research Greenford Road, Greenford, Middlesex UB6 0HE, U.K.

[1] Abbreviations: IGF-1, insulin-like growth factor 1; NMR, nuclear magnetic resonance; RMD, restrained molecular dynamics; IGF-2, insulin-like growth factor 2; TPPI, time-proportional phase incrementation; COSY, correlated spectroscopy; DQF-COSY, double-quantum filtered correlated spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy.

0006-2960/91/0430-5484$02.50/0

IGF-1 and IGF-2 are highly homologous single-chain proteins of 70 and 67 amino acids, respectively (Rinderknecht & Humbel, 1978a,b). Residues 3–29 of IGF-1 and residues 6-32 of IGF-2 are homologous to the B-chain of insulin and are thus known as the B-domains of the IGFs. Then follow the C-domains, analogous in location to the C-domain of proinsulin, but shorter and with no sequence homology either with each other or with proinsulin. These lead to the A-domains, residues 42–62 of IGF-1 and 41–61 of IGF-2, which are homologous to the A-chain of insulin. The short sequences at the carboxyl termini, although homologous to each other, have no insulin analogue and are known as the D-domains. Although it has so far proved impossible to obtain crystals of the IGFs suitable for diffraction studies, the high degree of sequence similarity prompted Blundell et al. (1978) to model the tertiary structure of IGF-1 on that of porcine insulin.

The IGFs are peculiar in their ability to bind to more than one receptor. Two distinct IGF receptors, type 1 and type 2, are found in many cell lines (Massague & Czech, 1982; Rechler & Nissley, 1985). IGF-1 binds to the type 1 receptor and, with a lower affinity, to the insulin receptor. It is also thought to possess a low affinity for the type 2 receptor, but this view has been challenged (Tally et al., 1987). In contrast, IGF-2 binds with high affinity to the type 2 receptor and with low affinity to the type 1 receptor but has no affinity for the insulin receptor. The homologous sequence relationships between the IGFs and insulin are paralleled by close similarities between the insulin and the type 1 IGF receptors (Massague & Czech, 1982; Yarden & Ullrich, 1988). The type 2 receptor is quite distinct (Massague & Czech, 1982) and appears to be identical with the mannose-6-phosphate receptor (Morgan et al., 1987). The ability of IGF-1 to bind to the insulin receptor may be partially responsible for the observed metabolic effects of this growth factor, although Rechler and Nissley (1985) have emphasized that this is unlikely to be the complete story.

Progress has been made in correlating the primary structures of IGFs with their relative affinities for different receptors. The affinities and/or activities of modified IGF-1 have been measured, and the results were used to predict which residues are important for receptor recognition [e.g., Cascieri et al. (1988, 1889); Bayne et al. (1988); Maly and Luthi (1988); Chen et al. (1988); and Cascieri and Bayne (1990)]. These hypotheses are, however, limited by the lack of detailed information concerning the tertiary structures of IGFs. As part of a project considering structure–function relationships in protein growth factors, we have examined the solution structure of human IGF-1, with nuclear magnetic resonance (NMR) and restrained molecular dynamics (RMD) methods. In this paper, we report the ^{1}H NMR assignments of IGF-1 and the utilization of RMD to generate a family of IGF-1 structures consistent with the NMR data. These structures are compared with that of insulin and are used to rationalize the observed biological activities of IGF-1.

Experimental Procedures

Human IGF-1 was a generous gift from the Mosanto Chemical Company. Samples for NMR studies were prepared by dissolving lyophilized protein in 100% deuterium oxide (D_2O) or 90% H_2O/10% D_2O and adding NaOD or DCl to yield a pH of 2.8–2.9 (uncorrected meter reading). The protein concentration ranged from 3–7 mM.

NMR spectra were acquired at 323 K with Bruker AM 500-MHz and AM 600-MHz spectrometers. Two-dimensional NMR spectra were recorded in the phase-sensitive mode with quadrature detection in F_1 provided by TPPI (Drobny et al., 1979; Bodenhausen et al., 1980). COSY (Bax & Freeman, 1981), DQF-COSY (Piantini et al., 1982), NOESY (Jeener et al., 1979), and TOCSY (Braunschweiler & Ernst, 1983) spectra were recorded with the standard Bruker microprograms. The mixing time in the NOESY experiments ranged from 80 to 300 ms, with no random variation. The spin-lock period in the TOCSY experiments was achieved with MLEV-17 (Bax & Davis, 1985) and ranged from 40 to 80 ms.

Restrained energy minimization and RMD simulations were performed with the GROMOS suite of programs (van Gunsteren & Berendsen, 1987). The initial structure of IGF-1 (IGF1.pdb from the Brookhaven Protein Data Bank) was first energy minimized by incorporating 344 upper and 23 lower distance restraints derived from NMR spectra. The structure was then equilibrated for 2 ps while coupled to a bath at 300 K with a coupling constant of $T = 0.01$ ps. The time step in this stage and in the remainder of the simulations was 0.002 ps, and all bond lengths were kept rigid by using the SHAKE algorithm (Ryckeart et al., 1977; van Gunsteren & Berendsen, 1977). NOE data were included with an initial force constant of 1000 kJ mol^{-1} nm^{-2}, which was increased to 8000 kJ mol^{-1} nm^{-2} over 15 ps of dynamics. Trajectories were calculated at 300 K for 40 ps, by using a different distribution of velocities at the equilibration stage. Most structures were obtained by averaging atomic positions over the final 5 ps of individual dynamics trajectories, followed by restrained energy minimization. Other structures were obtained by continuation of the above trajectories at 600 K for 5 ps, followed by cooling to 300 K, then averaging over a further 5 ps of dynamics, and minimizing as above. Each structure was then refined with 1000 steps of the discrete dynamic searching (SNIFR) method of Donnelly and Rogers (1988, 1989), again with a force constant of 8000 kJ mol^{-1} nm^{-2} for the NOE restraints.

Results

Assignment of Resonances. Initial experiments were aimed at determining a suitable temperature and pH for NMR studies. IGF-1 was observed to be insoluble (at millimolar concentrations) between pHs of ca. 5.5 and 10. In addition, spectra of samples above pH 3 were found to have poorer resolution than those of samples below pH 3; ultracentrifugation studies suggest that IGF-1 exists as a monomer below pH 3 and a dimer above this pH (R. M. Cooke et al., unpublished observations). Spectra recorded at room temperature had unacceptably poor resolution. At 323 K the resolution is improved, while the similar pattern of structurally shifted resonances suggests that the general tertiary structure is maintained at the higher temperature. Spectra of solutions of IGF-1 recorded before and after standing at 323 K for 48 h were not significantly different, suggesting that little degradation occurs under these conditions.

Resonance assignments were made by first identifying spin systems characteristic of particular amino acids. Analysis of spin-system patterns in various 2D spectra enabled 6 Ala, 3 Val, 6 Leu, 1 Ile, 3 Thr, and 3 Tyr spin systems to be identified, accounting for the total number of each of these residues in IGF-1. The DQF-COSY cross peaks involving the methyl resonances of the Val, Leu, and Ile residues are shown in Figure 1. The proximity of the Phe ring proton resonances to each other prevented the observation of complete spin systems. However, NOEs from the 2,6H resonances, allied with the absence of Trp and His from IGF-1, enabled identification of the aliphatic resonances of all four Phe residues. Only four of the seven Gly spin systems and two of the five Ser spin systems were unambiguously identified at this stage.

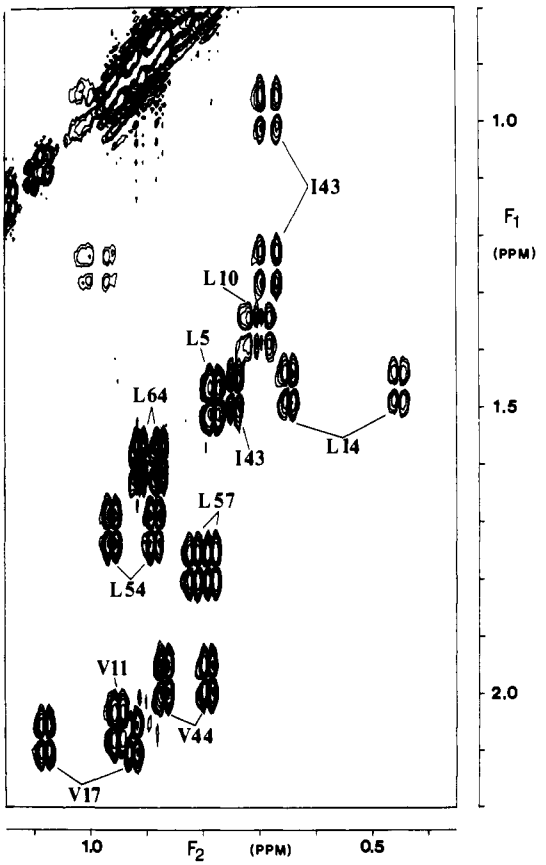

FIGURE 1: Region of a 500-MHz DQF-COSY spectrum of IGF-1 in D_2O. A total of 512 experiments, each of 4096 points, were acquired. The data were enhanced with shifted sine-bell window functions and zero filled twice in F_1 prior to Fourier transformation. The cross peaks involving the methyl resonances of the Ile, Leu, and Val residues are labeled.

Several other spin systems, which could not be unambiguously assigned to particular amino acids, were grouped in spectroscopically similar classes (e.g., Lys/Arg/Pro, Met/Gln/Glu, and Cys/Asn/Asp).

Spin systems known to arise from a particular amino acid were assigned to a specific location in the sequence by the standard procedures of sequential assignment (Billeter et al., 1982). This also enabled many spin systems, which previously could only be grouped into classes of amino acids, to be uniquely identified and sequentially assigned. The amide proton region of a NOESY spectrum of IGF-1 is shown in Figure 2, with several sequential NH–NH NOEs marked. Figure 3 shows the sequential connectivities observed for IGF-1. These allowed sequence-specific assignments of resonances of residues 1–33, 38–49, 51–63, and 65–67. The remaining Leu spin system was assigned to Leu 64, while the remaining Ala spin system was assigned to Ala 70, consistent with the pH dependence of its resonances. The resonance assignments are summarized in Table I.

Determination of Structure. The assignment of a large number of ^1H resonances and the identification of NOEs between resonances provided the basis for determination of the tertiary structure of IGF-1. A total of 344 NOEs were unambiguously identified. Of these, 147 occurred between resonances of the same residue, 123 occurred between resonances of sequentially adjacent residues, and 74 occurred between resonances of residues separated by at least one other amino acid in the sequence. These NOEs were grouped into classes according to intensity and the corresponding interproton distances, assigned a maximum value of 2.7, 3.0, 3.5, 4.0, 4.5,

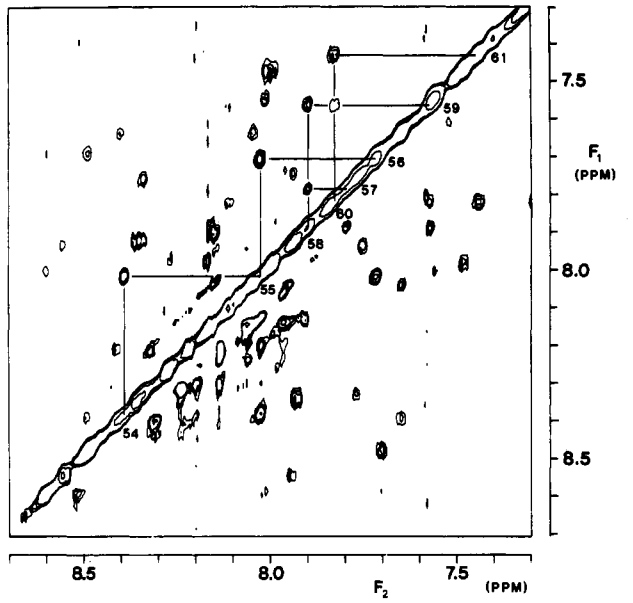

FIGURE 2: Downfield region of a NOESY spectrum of IGF-1 in H_2O. A total of 473 experiments, each of 4096 points, were acquired. The data were enhanced with a trapezoidal window function in F_2 and a shifted sine-bell window function in F_1 and were zero filled to 2048 points in F_1. The sequential cross peaks linking residues of the third helix are labeled. (The use of the trapezoidal window function has distorted the intensity of several cross peaks, particularly those close to the diagonal, causing variations in relative intensities of NOEs from those displayed in Figure 3 and used in the RMD simulations.)

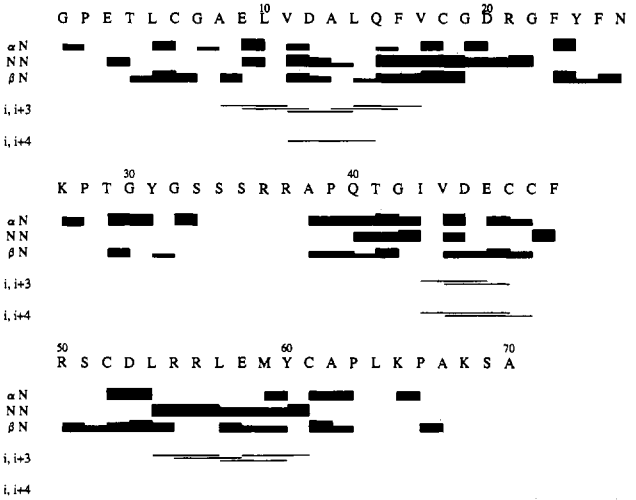

FIGURE 3: Pattern of sequential and "short-range" NOEs observed for IGF-1. The NOEs observed are indicated by bars linking the residues concerned. The first three rows represent the common classes of sequential NOEs (with the βN class including NOEs to other side chain resonances) and the height of each bar indicates the approximate NOE intensity. In the case of proline the δCH resonances were used in place of NH resonances. The next row indicates residues linked by $\alpha N(i,i+3)$ and $\alpha N(i,i+4)$ NOEs.

or 5.0 Å. In addition, several pairs of NH protons were assigned a minimum interatomic separation of 3 Å on the basis of the absence of NOEs. These cases were limited to sequentially adjacent residues and to situations where the NOEs concerned would be distinct from the diagonal resonances and from other NOESY cross peaks. A total of 23 such minimum distances were specified.

The distance restraints were incorporated in RMD simulations of the structure of IGF-1, starting from the insulin-based structure. Prior to RMD, application of the NOE-derived distance restraints to this structure produced a large number of restraints violations, which suggested that this

Table I: Resonance Assignments[a]

residue	chemcal shift (ppm)[b] NH	αCH	βCH	other	residue	chemcal shift (ppm)[b] NH	αCH	βCH	other
Gly 1		4.02					3.88		
		4.02			Ser 33	8.11			
Pro 2				δCH, 3.61	Ala 38	8.04	4.59	1.35	
				δCH, 3.61	Pro 39		4.42	2.25	γCH, 1.99
Glu 3	8.55	4.45	2.19	γCH, 2.47				1.99	δCH, 3.74
			2.01						δCH, 3.61
Thr 4	7.94	4.44	4.27	γCH_3, 1.14	Gln 40	8.36	4.35	2.14	γCH, 2.37
Leu 5	8.18	4.40	1.60	γCH, 1.49				2.00	εNH, 7.36
				δCH_3, 0.78					εNH, 6.73
				δCH_3, 0.78	Thr 41	7.93	4.30	4.23	γCH_3, 1.21
Cys 6	8.09	4.59	3.05		Gly 42	8.34	4.04		
Gly 7	8.48	3.90			Ile 43	7.77	3.93	1.48	γCH, 1.26
		3.73							γCH, 0.99
Ala 8	8.35	4.04	1.44						γCH_3, 0.74
Glu 9	8.01	4.06	2.21	γCH, 2.58					δCH_3, 0.68
			2.09	2.49	Val 44	7.75	3.66	1.97	γCH_3, 0.86
Leu 10	7.48	3.88	1.69	γCH, 1.36					γCH_3, 0.78
				δCH_3, 0.71	Asp 45	7.94	4.50	2.86	
				δCH_3, 0.68				2.86	
Val 11	7.48	3.37	2.06	γCH_3, 0.95	Glu 46	7.91	4.16	2.12	γCH, 2.52
				γCH_3, 0.95				2.12	γCH, 2.52
Asp 12	7.98	4.38	2.95		Cys 47	8.14	4.80	3.22	
			2.81					3.03	
Ala 13	7.90	4.07	1.38		Cys 48	7.97		3.14	
Leu 14	8.16	3.83		γCH, 1.47				2.73	
				δCH_3, 0.64	Phe 49	7.75	4.68	3.37	2,6H, 7.28
				δCH_3, 0.45				3.07	
Gln 15	8.05	4.06	2.18	γCH, 2.46	Ser 51	7.85	4.50	3.80	
			2.05	γCH, 2.37	Cys 52	8.82	4.88	3.31	
Phe 16	7.65	4.42	3.20	2,6H, 7.21				3.19	
			3.20	3,5H, 7.24	Asp 53	8.04	4.76	2.96	
Val 17	8.40	3.72	2.08	γCH_3, 1.08				2.92	
				γCH_3, 0.92	Leu 54	8.39	3.99	1.65	γCH, 1.71
Cys 18	8.48	4.82	3.32						δCH_3, 0.96
			2.91						δCH_3, 0.88
Gly 19	7.70	3.97			Arg 55	8.03	4.00	1.87	γCH, 2.01
		3.97			Arg 56	7.72		2.01	
Asp 20	8.59	4.60	2.93					1.97	
			2.93		Leu 57	7.80	4.15	1.92	γCH, 1.78
Arg 21	8.01	4.21	1.92	γCH, 1.67				1.57	δCH_3, 0.81
			1.77						δCH_3, 0.78
Gly 22	7.56	3.97			Glu 58	7.90	4.18	2.11	γCH, 2.58
		3.76							γCH, 2.42
Phe 23	7.55	4.99	3.10	2,6H, 6.89	Met 59	7.57	4.22	2.02	γCH, 2.50
			2.85	3,5H, 7.14				1.93	
Tyr 24	8.34	4.64	3.07	2,6H, 7.04	Tyr 60	7.83		3.44	2,6H, 7.30
			2.94	3,5H, 6.78				2.91	3,5H, 6.76
Phe 25	8.07	4.65	2.94	2,6H, 7.14	Cys 61	7.44	5.00	3.18	
			3.10					2.87	
Asn 26	8.05	4.74	2.76	δNH, 7.41	Ala 62	8.18	4.44	1.26	
			2.66	δNH, 6.78	Leu 64	8.10	4.25	1.57	γCH, 1.58
Lys 27		4.48	1.68						δCH_3, 0.91
			1.64						δCH_3, 0.87
Thr 29	8.00	4.31	4.18	γCH_3, 1.19	Lys 65	8.11	4.61	1.79	
Gly 30	8.17	3.94						1.70	
		3.89			Pro 66		4.39	2.27	δCH, 3.80
Tyr 31	7.99	4.55	3.04	2,6H, 7.10				1.90	δCH, 3.62
			2.89	3,5H, 6.84	Ala 67	8.22	4.38	1.41	
Gly 32	8.27	3.98			Ala 70	8.12	4.30	1.40	

[a] At pH 2.9, 323 K. [b] Chemical shifts are accurate to ±0.02 ppm.

structure differed significantly from that observed in solution.

Modification of this structure in RMD simulations resulted in large decreases in both the potential energy and the average restraints violation. The results are summarized in Table II. When the backbone atoms (N, Cα, C, O) of these structures are overlaid, the average RMSD is 5.1 Å. With an iterative fitting procedure similar to that described by Nilges et al. (1987), the most variable regions of the structure were identified and ignored in subsequent fitting procedures. When the backbone atoms of residues 3–19, 22–26, and 43–61 only are used to fit structures, the average RMSD for these residues is 1.7 Å. The Cα atoms of these residues are overlaid in Figure 4.

Description of the IGF-1 Structure. From the family of RMD structures, well-defined regions of the molecule can be identified. Residues 3–6 adopt a β-strand-like conformation. The first helix starts at residue 8 and continues until residue 17. A poorly defined turn leads to residues 23–26, which are in an extended conformation. The end of the B-domain and all of the C-domain are ill-defined in the RMD structures, as evidenced by the RMSD of >4 Å for the backbone atoms of residues 30–39 and the large variation in dihedral angles. The

Table II: Energies of the Calculated IGF-1 Structures

structure	total potential energy (kJ mol^{-1})	sum of restraint violations (nm)
initial	2.976×10^{11}	5.06
I	−4164	0.84
II	−4230	0.89
III	−4138	1.05
IV	−4604	0.85
V	−4626	0.97
VI	−4644	0.83
VII	−4040	0.79
VIII	−4665	0.86
IX	−4365	0.81
X	−4142	0.99

FIGURE 4: The 10 RMD structures of IGF-1 overlaid by using the backbone atoms of residues 3–19, 22–26, and 43–61 (with an average RMSD of 1.7 Å for these atoms). The Cα atoms are linked by heavy lines for the residues just described and lighter lines for the remainder of the molecule. The location of the first and second helices (residues 8–17 and 43–49, respectively) are indicated. Note the lack of definition for the residues denoted by the light lines.

second helix encompasses residues 44–49. A disulfide bridge between residues 47 and 52 imposes a turn on the structure leading to the third helix. This runs from residue 54 to 59, although it appears irregular (particularly at Arg 55) and less well-defined than the other helices. Following residue 60 the D-domain is ill-defined.

The protein is cross linked by three disulfide bridges. The Cys 6–Cys 48 bridge is one the surface of the molecule, while those between Cys 18 and Cys 61 and between Cys 47 and Cys 52 are completely buried in the protein core. Other residues whose side chains are mostly buried are Leu 10, Ala 13, Leu 14, Val 17, Ile 43, Leu 57, and Tyr 60. This "hydrophobic core" is surrounded by the three helices and is, to a large extent, shielded from the solvent by the polypeptide backbone. The side chains of several residues fill most of the gaps in this shield: Leu 5, Phe 16, Arg 21, and Glu 58 cover the core between the first and third helices; Val 11, Phe 23, Phe 25, and Val 44 fill the gap between the first and second helices and the backbone of residues 22–26; and Gln 40 and Glu 46 cover the core in the region of Tyr 60. The backbone of IGF-1 is mostly arranged around the side chains packed in the hydrophobic core, although the N terminus crosses the A-domain, bringing Glu 3 close to Ser 51; and, after the final cysteine, the backbone of Ala 62 crosses that of Phe 23.

Discussion

The "Quality" of the IGF-1 Structure. The quality of NMR-derived structures should be assessed on two grounds: the degree to which the structure fits the experimental data and the variability of the calculated structures. The former can be assessed from consideration of the violations of the NOE-derived distance restraints. As shown in Table II, the sum of the restraints violations in the NMR/RMD structures ranges from 0.84 to 1.05 nm. This is a significant improvement on the initial structure in which the sum of the violations is 5.06 nm. The variability of structures is usually measured by the average RMSD for the atomic positions of overlaid structures. In the case of IGF-1, the average RMSD for the backbone atoms of the entire structure of 5.1 Å is quite high, suggesting poor structural definition. The immediate cause for this is the limited amount of NMR data. The NMR-derived restraints consisted exclusively of NOEs; large line widths prevented the use of NH–αCH coupling constants, while the temperature at which experiments were performed was too high for slowly exchanging amide protons to be identified and thus for hydrogen bonds to be inferred.

It is, however, apparent that some regions of the structure are far better defined than others. When only residues 3–19, 22–26, and 43–61 were used to fit structures, the average RMSD for their backbone atoms is reduced to 1.7 Å. The structures of these regions, which include the three helices and the hydrophobic core, are thus well-defined. The variation in structural definition results primarily from the fact that the NOEs are not distributed evenly throughout the sequence but are concentrated in the regions corresponding to the well-defined residues. The lack of NOEs for the poorly defined regions is to some degree inevitable due to limited assignments, yet even for assigned resonances from these regions few NOEs are observed. The absence of NOEs may be caused by a number of factors: rapid motion could decrease the NOE intensity, while the existence of several conformational states could diminish the relative intensities of NOEs resulting from only some states. Conformational flexibility is also suggested by the relatively narrow resonances observed from the ill-defined regions; the absence of a single conformation also correlates with the observation that few resonances from these regions are significantly shifted from their "random-coil" positions. The lack of definition apparent from the RMD simulations of the C- and D-domains may thus reflect conformational disorder that occurs in solution. The absence of a defined structure for the C-domain is consistent with the observation by Bayne et al. (1989) that replacement of residues 28–37 with four glycine residues has little effect on the remainder of the protein. It is also possible that disorder within the C- and D-domain regions is hindering attempts to crystallize IGF-1.

The Structural Relatonship with Insulin. The results presented in this paper support the suggestion that the solution structure of IGF-1 is similar to that of insulin. Patterns of NOEs are diagnostic of the secondary structural elements within a protein (Wuthrich et al., 1984). In the case of IGF-1, the pattern of NN(i,i+1), αN(i,i+3), and αN(i,i+4) NOEs (Figure 3) indicates that three helices are present, corresponding in approximate position to the three helices of insulin.

Although there is fairly good agreement with the model of Blundell et al. (1978), several of the NOEs observed in our NMR studies are incompatible with the model structure of IGF-1. This was confirmed by the RMD simulations: for most of the protein, the range of RMD structures encompasses the model-built IGF-1 structure, but there are regions in which the RMD structures differ significantly from the model-built structure. These residues are assigned positive deviations in Figure 5; residues whose standard deviations encompass the

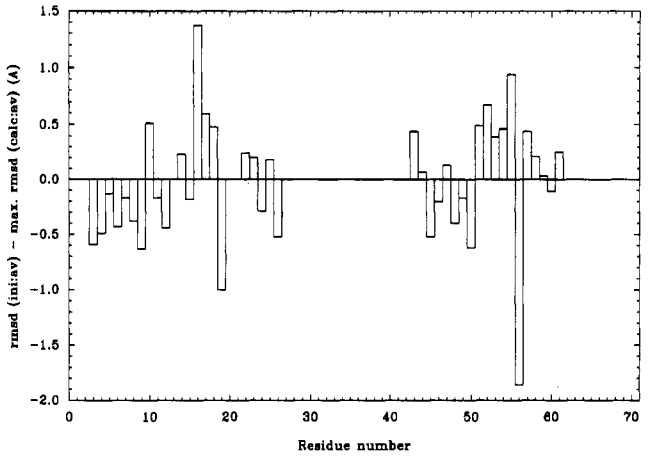

FIGURE 5: Differences between the calculated and starting structures for the well-defined regions of the IGF-1: residues 3–19, 22–26, 43–61. Each of the 10 calculated structures was fitted to their average and the maximum RMSD for each residue calculated [max (calc:av)]. This was then subtracted from the RMSD between the initial structure and the average structure (ini:av). Thus for any residue, a positive value indicates that its position in the initial structure lies outside the spread of those in the final structures, while a negative value indicates that it lies within the spread of calculated structures.

model-built structure are represented by negative values. Within the B-domain, the most significant changes in backbone coordinates occur at the end of the helix (residues 16–18), as the side chain of Phe 16 is drawn to a less solvent-exposed position by NOE restraints to Leu 54 and Leu 57. It is not possible to say whether the differences observed in the C-domain are significant as we have little information concerning the position of this region. There is, however, no evidence that the two β-turns predicted by Blundell et al. (1978) occur at the junction between the B- and C-domains. Several differences are observed in the A-domain. The location of the first helix in the A-domain (residues 44–49) differs from that in the predicted structure (residues 43–48). There is also a shift in the position of the backbone within this helix and in the link between the two A-domain helices. The structure of the second A-domain helix (residues 54–59) is also significantly different from that in the modeled structure. Although deviations from ideal helical conformations were expected for this region, the deformations appear to be larger than forecast. The final residues in each of the three helices display larger values for the ϕ dihedral angle than is expected for an ideal α-helix, suggesting that in solution the ends of the helices are not very stable. The D-domain was predicted by Blundell et al. (1978) to bend back on itself. This seems to be the case although the position of the bend is variable and this domain appears to be located closer to the rest of the molecule than predicted. However, in this case the lack of long-range NOE restraints limits the significance of such differences.

As in insulin, two dusulfide bridges of IGF-1 are buried, while the third is on the surface, and homologous residues constitute the hydrophobic core of each protein. Although there are small shifts in the relative orientation of the helices between the predicted and experimental structures (Table III), movement of the side chains relative to their backbones allow the packing of the core to be preserved. There are, however, changes in the positions of the side chains of Ile 43, Tyr 60, and the four buried cysteines. Differences are also observed for some of the residues surrounding the core. Phe 16 and Arg 21 cover the core more than in the predicted structure. A movement of the backbone of residues 23–26 away from the B-domain helix means that the side chain of Phe 23 moves with respect to its backbone to maintain the same position over the core. In the case of Phe 25, the movement of the backbone is larger than can be compensated for, and the ring adopts an altered location with respect to the core.

Table III: Interhelix Angles in IGF-1

	angle (deg)		
structure	8–18:44–49	44–49:54–59	8–18:54–59
I	69	143	82
II	82	148	74
III	91	140	52
IV	91	128	56
V	66	125	64
VI	75	120	55
VII	66	143	85
VII	76	123	56
IX	69	138	83
X	61	140	92
average	75 ± 15	135 ± 12	70 ± 22
initial	63	144	90

The roles of particular glycine residues in the structures of insulins and IGFs have previously attracted comment. In all the RMD structures of IGF-1, Gly 7 and Gly 42 possess positive ϕ angles. They are thus required for the backbone of the protein to adopt unusual conformations. On the other hand, the dihedral angles of Gly 19 and Gly 22 vary considerably between structures, and their role is apparently to allow flexibility in the backbone. It is likely that the corresponding glycines in insulin perform the same roles. The remaining glycines in IGF-1, at positions 1, 30, and 32, also appear to be quite flexible.

Interactions with Other Proteins. Using modified IGFs, several researchers have attempted to locate the residues required for interaction with their various biological partners: the insulin and type 1 and type 2 receptors and the serum binding proteins [reviewed by Cascieri and Bayne (1990)]. Using the NMR/RMD structures of IGF-1, we can briefly examine the disposition of "important" residues to help explain their significance for activity.

It is clear that the N-terminal tripeptide of IGF-1 is required for interaction with the serum binding proteins (Szabo et al., 1988), and the involvement of residues 49–51 has also been inferred (Cascieri et al., 1989). These regions form a patch at one end of the molecule, as shown in Figure 6A.

The studies of Bayne et al. (1988) and Cascieri et al. (1989) have implicated residues 1, 2, 8, 9, 12, 49, 50, 51, 55, and 56 interactions with the type 2 receptor. These residues form two patches on the surface of IGF-1 (Figure 6B). It is possible that the receptor-binding residues of IGF-1 form noncontiguous patches, interacting with a relatively large receptor. Alternatively, the importance of other residues lying between the patches has yet to be demonstrated. For example, Asp 53, which is conserved in IGFs but not in insulins and lies on the surface between the patches marked in Figure 6B, may also contribute to binding.

An extrapolation from structure/function relationships in insulin (Pullen et al., 1976; Baker et al., 1988) shows that residues 21, 23, 24, 25, 42, 43, 44, 46, 60, and 62 of IGF-1 appear to be important for binding to the insulin receptor. Arg 21, Phe 23, Tyr 24, Phe 25, and Val 44 lie along a cleft on one face of IGF-1, separating the A- and B-domains from the C-domain (Figure 6C). The side chains of the three aromatic residues lie flat across the surface with the Phe rings directed toward the A- and B-domains and the Tyr ring directed toward the C-domain. The other five "important" residues do not contribute directly to the surface of the cleft, although Gly 42 and Ile 43 are buried beneath Phe 25 and Val 44 and may

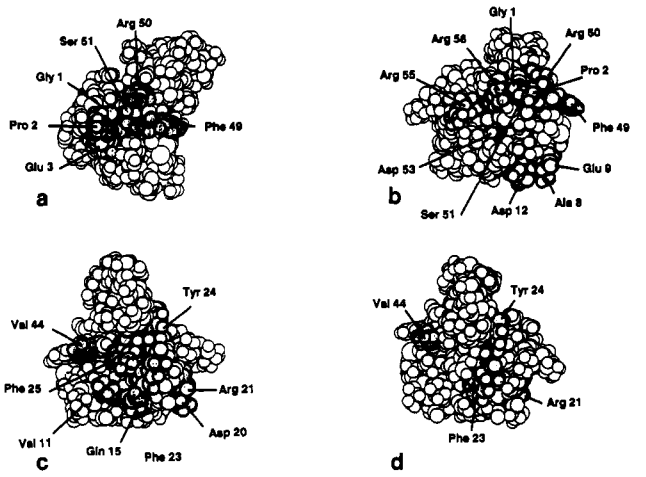

FIGURE 6: The average minimized RMD structure of IGF-1 with residues highlighted corresponding to the sites of interaction with the following species: (a) The serum binding protein, residues 1–3 and 49–51 of IGF-1. (b) The type 2 receptor, residues 1, 2, 8, 9, 12, 49–51, 55, and 56 of IGF-1. Ser 51 is partially obscured by the N terminus in this orientation. The position of Asp 53 is also indicated, which may be important in this particular interaction (see text for details). (c) The insulin receptor, residues 21, 23, 24, 25, 42, 43, 44, 46, 60, and 62 of IGF-1. Glu 46 is located on the opposite side of the molecule, approximately over Gly 42, Ile 43, and Tyr 60, buried in the core. (d) The type-1 receptor, residues 21, 23, 24, and 44 of IGF-1. It is interesting to note the total overlap of this surface with that of the insulin receptor site in panel c.

serve to determine their location. Glu 46 located on the surface of the molecule, on the opposite face to the cleft, while Tyr 60 and Ala 62 may be important for stabilizing the tertiary structure. The D-domain, which is known to inhibit binding to the insulin receptor (Cascieri et al., 1988; Bayne et al., 1989), is located at one end of the cleft, near Arg 21. Three residues earlier in the B-domain, Val 11, Gln 15, Asp 20, border the cleft and may contribute to receptor binding.

Fewer residues have been shown to interact with the type 1 receptor. Modification of Phe 23 and Tyr 24 altered the affinity for this receptor (Cascieri et al., 1988), and the effect of modifications of Tyr 24 were more pronounced for type 1 receptor affinity than for insulin receptor affinity. The use of these residues for binding to both the insulin and the type 1 receptors and the proposed homology between these receptors suggest that other residues involved in binding to the insulin receptor may be used in type 1 receptor binding. Arg 21 and Val 44, which are conserved in higher insulins and IGFs, are the most likely candidates. The locations of residues 21, 23, 24, and 44 are marked in Figure 6D. Some or all of the residues within the C-domain, which adjoins the cleft in IGFs but not in insulins, appear to be required for distinguishing between the type 1 and insulin receptors (Bayne et al., 1989; Cascieri & Bayne, 1990). Our studies, however, show the C-domain to be particularly mobile in the uncomplexed gowth factor. Maly and Luthi (1988) found that binding to the receptor protected Tyr 24, Tyr 31, and Tyr 60 from iodination, indicating that these residues formed part of or were close to the binding site.

In summary, the residues that bind to the type 1 receptor appear to overlap those that bind to the insulin receptor, whereas those that bind to the type 2 receptor overlap those that interact with the serum binding proteins. The former pairing correlates with the likely structural relationship between those receptors. Morgan et al. (1987) suggested that the type 2 receptor may be related to the serum binding proteins. Although the sequences of two binding proteins have since been shown to have no homology with that of the type 2 receptor (Brewer et al., 1988, Binkert et al., 1989), they do possess cysteine-rich regions that form the ligand-binding sites, akin to membrane-bound growth factor receptors.

Conclusions. This paper presents the first detailed description of the solution structure of IGF-1. The assignment of the vast majority of the ^{1}H NMR resonances and a qualitative assessment of the observed NOEs show the tertiary structure of IGF-1 to be similar to that of insulin. Incorporating the NMR information in molecular dynamics simulations confirm this while enabling differences between the structures of insulin and IGF-1 to be observed. The molecular dynamics simulations also show variations in structural definition throughout the molecule, with flexible regions connected to a well-defined core, which corresponds to the region of insulin homology. Consideration of the results of mutagenesis studies in terms of our NMR/RMD structures has enabled the regions of IGF-1 involved in interactions with its various biological partners to be visualized.

ACKNOWLEDGMENTS

This is a contribution from the OCMS. We thank S. S. Wang for assistance in the early stages of this project, W. F. van Gunsteren for the gift of GROMOS, and R. A. Donnelly for the gift and help in the use of SNIFR.

Registry No. IGF-1, 67763-96-6; IGF-1 (human), 68562-41-4; insulin, 9004-10-8.

REFERENCES

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

Bax, A., & Freeman, R. (1981) *J. Magn. Reson. 44*, 542–561.

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

Bayne, M. L., Applebaum, J., Chicchi, G. G., Hayes, N. S., Green, B. G., & Cascieri, M. A. (1988) *J. Biol. Chem. 263*, 6233–6239.

Bayne, M. L., Applebaum, J., Underwood, D., Chicchi, G. G., Green, B. G., Hayes, N. S., & Cascieri, M. A. (1989) *J. Biol. Chem. 264*, 11004–11008.

Billeter, M., Braun, W., & Wuthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Binkert, C., Landwehr, J., Mary, J.-L., Schwander, J., & Heinrich, G. (1989) *EMBO J. 8*, 2497–2502.

Blundell, T. L., Bedarkar, S., Rinderknecht, E., & Humbel, R. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 180–184.

Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.

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

Brewer, M. T., Stetler, G. L., Squires, C. H., Thompson, R. C., Busby, W. H., & Clemmons, D. R. (1988) *Biochem. Biophys. Res. Commun. 152*, 1289–1297.

Cascieri, M. A., & Bayne, M. L. (1990) in *Molecular and Cellular Biology of IGFs and their Receptors* (LeRoth, D., & Raizada, M. K., Eds.) Plenum Press, London.

Cascieri, M. A., Chicchi, G. C., Applebaum, J., Hayes, N. S., Green, B. C., & Bayne, M. L. (1988) *Biochemistry 27*, 3229–3233.

Cascieri, M. A., Chicchi, G. C., Applebaum, J., Green, B. G., Hayes, N. S., & Bayne, M. L. (1989) *J. Biol. Chem. 264*, 2199–2202.

Chen, Z. Z., Schwartz, G. P., Zong, L., Burke, G. T., Chanley, J. D., & Katsoyannis, P. G. (1988) *Biochemistry 27*, 6105–6111.

Daughaday, W. H., Hall, K., Raben, M. S., Salmon, W. D., Van den Brande, J. L., & Van Wyck, J. J. (1972) *Nature 235*, 107.
Donnelly, R. A., & Rogers, J. W. (1988) *Int. J. Quantum Chem., Quantum Chem. Symp. 22*, 507–513.
Donnelly, R. A., & Rogers, J. W. (1989) *J. Opt. Theory Appl. 61*, 111–121.
Drobny, G., Pines, A., Sinton, S., Weitkamp, D. P., & Wemmer, D. (1979) *Faraday Symp. Chem. Soc. 13*, 49–55.
Dulak, N., & Temin, H. J. (1973) *J. Cell. Physiol. 81*, 153–160.
Froesch, E. R., Burgi, H., Ramseier, E. B., Bally, P., & Labhart, A. (1963) *J. Clin. Invest. 42*, 1816–1834.
Froesch, E. R., Schmid, C., Schwander, J., & Zapf, J. (1985) *Annu. Rev. Physiol. 47*, 443–467.
James, R., & Bradshaw, R. A. (1984) *Annu. Rev. Biochem. 53*, 259–292.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Maly, P., & Luthi, C. (1988) *J. Biol. Chem. 263*, 7068–7072.
Massague, J., & Czech, M. P. (1982) *J. Biol. Chem. 257*, 5038–5045.
Morgan, D. O., Edman, J. C., Standring, D. N., Fried, V. A., Smith, M. S., Roth, R. A., & Rutter, W. J. (1987) *Nature 329*, 301–307.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1987) *FEBS Lett. 219*, 11–16.
Piantini, U., Sorensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J., & Gammeltoft, S. (1976) *Nature 259*, 369–373.
Rechler, M. M., & Nissley, S. P. (1985) *Annu. Rev. Physiol. 47*, 425–442.
Rinderknecht, E., & Humbel, R. E. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 4379–4381.
Rinderknecht, E., & Humbel, R. E. (1978a) *J. Biol. Chem. 253*, 2769–2776.
Rinderknecht, E., & Humbel, R. E. (1978b) *FEBS Lett. 89*, 283–286.
Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Physics 23*, 327–341.
Szabo, L., Mottershead, D. G., Ballard, F. J., & Wallace, J. C. (1988) *Biochem. Biophys. Res. Commun. 151*, 207–214.
Tally, M., Enberg, G., Li, C. H., & Hall, K. (1987) *Biochem. Biophys. Res. Commun. 147*, 1206–1212.
van Gunsteren, W. F., & Berendsen, H. J. C. (1977) *Mol. Phys. 34*, 1311–1327.
van Gunsteren, W. F., & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Nijenborgh 16, Groningen, The Netherlands.
Wuthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.
Yarden, Y., & Ullrich, A. (1988) *Annu. Rev. Biochem. 57*, 443–478.

Interaction of Influenza Hemagglutinin Amino-Terminal Peptide with Phospholipid Vesicles: A Fluorescence Study

Michael J. Clague,*,‡ Jay R. Knutson,§ Robert Blumenthal,‡ and Andreas Herrmann‖
Section of Membrane Structure and Function, National Cancer Institute, and Laboratory of Cell Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, and Humboldt-Universität, Bereich Biophysik, der Sektion Biologie, Invalidenstrasse 42, 1040 Berlin, Germany

Received October 30, 1990; Revised Manuscript Received March 4, 1991

ABSTRACT: We have studied tryptophan fluorescence from a 20-residue synthetic peptide corresponding to the amino terminal of the HA2 subunit of the influenza virus hemagglutinin protein, a putative "fusion" peptide. Decay-associated spectra have been obtained at pH 7.4 and at pH 5 (the optimal pH for influenza virus fusion) in the presence and absence of liposomes. We demonstrate that a blue shift in the total steady-state fluorescence spectrum upon binding to liposomes is due to a movement in characteristic emission wavelength and increased lifetime of one of the resolved spectral components. In contrast, a further shift after lowering the pH is the product of a redistribution in the relative amplitudes of spectral components. Also, each decay component is quenched by spin-labels or anthroxyl groups normally located within the hydrocarbon interior of the membranes. Calculations are presented leading to an estimate of the distance of the tryptophan residue from the bilayer center, suggesting that the tryptophan residues are at or near the hydrocarbon–polar interface. No gross positional change was detected between pH values. Rotational depolarization is shown to be retarded by liposome binding, more so at low pH.

The induction of fusion between influenza virus and target cell membranes upon lowering of the pH has been ascribed to a conformational change of the hemagglutinin (HA)[1] spike glycoprotein, the predominant protein on the viral surface (Skehel et al., 1982). One consequence of this structural rearrangement is the exposure of the amino-terminal segment of the HA-2 subunit (White & Wilson, 1987), which, because of its hydrophobic character, is an attractive candidate for involvement in the interaction with target membranes. In order

* Address correspondence to this author at the European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany.
‡ National Cancer Institute.
§ National Heart, Lung, and Blood Institute.
‖ Humboldt-Universität.

[1] Abbreviations: BHA, hemagglutinin, bromelain fragment; DAS, decay-associated spectra; DMPC, dimyristoylphosphatidylcholine; DPPC, dipalmitoylphosphatidylcholine; HA, hemagglutinin.