Conformation of glucagon in a lipid-water interphase by 1H-NMR

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · NOVEMBER 1983

Impact Factor: 4.33 · DOI: 10.1016/S0022-2836(83)80143-0 · Source: PubMed

J. Mol. Biol. (1983) **169**, 921–948

Conformation of Glucagon in a Lipid–Water Interphase by ^{1}H Nuclear Magnetic Resonance

W. Braun†, G. Wider‡, K. H. Lee and K. Wüthrich

Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule
CH-8093 Zürich-Hönggerberg, Switzerland

(Received 30 March 1983)

A determination of the spatial structure of the polypeptide hormone glucagon bound to perdeuterated dodecylphosphocholine micelles is described. A map of distance constraints between individually assigned hydrogen atoms of the polypeptide chain was obtained from two-dimensional nuclear Overhauser enhancement spectroscopy. These data were used as the input for a distance geometry algorithm for computing conformations that would be compatible with the experiments. In the region from residues 5 to 29 the mobility of the polypeptide backbone and most of the amino acid side-chains was found to be essentially restricted to the overall rotational tumbling of the micelles. The secondary structure in this region includes three turns of irregular α-helix in the segment of residues 17 to 29 near the C terminus, a stretch of extended polypeptide chain from residues 14 to 17, an α-helix-like turn formed by the residues 10 to 14 and another extended region from residues 5 to 10. In the N-terminal tetrapeptide H-His-Ser-Gln-Gly- the two terminal residues are highly mobile, indicating that they extend into the aqueous phase, and the mobility of the residues Gln3 and Gly4 appears to be only partially restricted by the binding to the micelle. The absence of long range nuclear Overhauser effects between the peptide segments 5–9 and 11–29, and between 5–16 and 19–29 shows that the polypeptide chain does not fold back on itself and hence that micelle-bound glucagon does not adopt a globular tertiary structure. Previously it was shown that the polypeptide backbone of glucagon is located close to and runs roughly parallel to the micelle surface. Combination of these observations suggests that the overall spatial arrangement of the glucagon polypeptide chain in a lipid–water interphase is largely determined by the topology of the lipid support, in the present case the curvature of the dodecylphosphocholine micelles. The tertiary structure is further characterized by the formation of two hydrophobic patches by the side-chains of Phe6, Tyr10 and Leu14, and the side-chains of Ala19, Phe22, Val23, Trp25 and Leu26, respectively.

1. Introduction

Glucagon is a hormone which consists of a linear polypeptide chain of 29 amino acid residues and has a molecular weight of 3500. The primary target organ for

† Present address: Dept of Physics, Kyushu University 33, Fukuoka 812, Japan.
‡ Present address: Spectrospin AG, Industriestrasse 26, CH-8117 Fällanden, Switzerland.

0022-2836/83/280921-28 $03.00/0

glucagon is the plasma membrane of liver and other cells, where binding to a specific receptor site mediates activation of glycogenolysis (Pohl *et al.*, 1969). Evidence has been presented that recognition between glucagon and its receptor depends on the ordered lipid structures surrounding the receptor site in the membrane (Rodbell *et al.*, 1971; Rubalcava & Rodbell, 1973). Otherwise, structural data on the receptor system are scarce and therefore much effort has been invested to characterize the glucagon–receptor interactions through studies of the conformational properties of the hormone (Sasaki *et al.*, 1975; Blundell & Wood, 1982).

Early studies of glucagon conformation by circular dichroism and other physical-chemical techniques indicated a tendency of this polypeptide to adopt different spatial structures in different environments. For example, for monomeric glucagon in aqueous solution a flexible "random coil" structure was indicated (Panijpan & Gratzer, 1974). For self-aggregated glucagon in aqueous solution evidence was presented that it could adopt either an α-helical (Gratzer *et al.*, 1967; Srere & Brooks, 1969) or β-sheet (Epand, 1971; Moran *et al.*, 1977) secondary structure and that the species formed depended critically on the peptide concentration (Wagman *et al.*, 1980). Furthermore, interactions with lipids and detergents were found to induce changes of the glucagon conformation (Schneider & Edelhoch, 1972; Epand *et al.*, 1977; Bösch *et al.*, 1980). These observations indicated that other techniques, which would be capable of providing more detailed structural information based on measurements of a large number of parameters, should be applied to glucagon in different milieus. X-ray studies of glucagon trimers in single crystals showed that the individual peptide molecules had an α-helical conformation (Sasaki *et al.*, 1975). High resolution ^{1}H n.m.r.† studies confirmed the earlier observations that monomeric glucagon in aqueous solution adopts a predominantly flexible random coil form and further revealed a structured region involving the residues 22 to 25, with a conformation different from α-helix-type secondary structure (Bösch *et al.*, 1978). The present paper reports on the conformation of monomeric glucagon bound to dodecylphosphocholine micelles.

Previous studies using spectroscopy and other physical-chemical methods showed that with the experimental conditions used for the present experiments, glucagon-containing DPC micelles contain one molecule of glucagon and ~40 detergent molecules and have a molecular weight of ~17,000 (Bösch *et al.*, 1980; Wider *et al.*, 1982). When perdeuterated DPC is used (Brown, 1979) the ^{1}H n.m.r. lines of MB-glucagon are sufficiently well resolved for a detailed spectral analysis. Studies with conventional, one-dimensional n.m.r. techniques showed that the peptide has a non-globular, predominantly extended form and a low resolution conformation based on individual assignments for a limited number of amino acid side-chain protons was obtained for the segment of residues 19 to 27 (Braun *et al.*, 1981). In the meantime almost all resonances in the ^{1}H n.m.r. spectrum of MB-

† Abbreviations used: MB-glucagon, glucagon bound to dodecylphosphocholine micelles; DPC, [$^2H_{38}$]dodecylphosphocholine; n.m.r., nuclear magnetic resonance; p.p.m., parts per million; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; r.m.s.d., root-mean-square distance.

glucagon were assigned to specific residues in the amino acid sequence (Wider *et al.*, 1982). In this paper these resonance assignments provide the basis for structural interpretations of NOESY spectra.

2. Nuclear Magnetic Resonance Measurements

The structure determination for MB-glucagon relies on the evaluation of cross peak intensities in two-dimensional nuclear Overhauser enhancement spectra. For the NOESY experiments we used two samples, one in H_2O and one in ^2H_2O, which contained 0·015 M-glucagon, 0·7 M-[$^2H_{38}$]-dodecylphosphocholine and 0·05 M-phosphate buffer, with a pH meter reading of 6·0. The sample temperature during the n.m.r. measurements was 37°C. Under these conditions the solutions contain mixed micelles of molecular weight ~17,000, which consist of one molecule of glucagon and ~40 molecules of DPC (Bösch *et al.*, 1980; Wider *et al.*, 1982).

NOESY spectra were recorded with the pulse sequence (Jeener *et al.*, 1979; Anil Kumar *et al.*, 1980*a*):

$$(90° - t_1 - 90° - \tau_m - 90° - t_2)_n,$$

where t_1 is the evolution period, τ_m the mixing time and t_2 the observation period. To obtain a two-dimensional spectrum the measurement is repeated for a set of equidistant t_1 values. The signal-to-noise ratio is improved by accumulation of n transients for each value of t_1. After each observation the system is allowed to reach equilibrium during a fixed delay time. For all the experiments needed to record a complete NOESY spectrum, the same value for τ_m is used.

The NOESY spectra were recorded at 500 MHz on a Bruker WM 500 spectrometer. Quadrature detection was used, with the carrier frequency at the low field end of the spectrum. To eliminate experimental artifacts, groups of 16 recordings with different phases were added for each value of t_1 (Nagayama *et al.*, 1979,1980). To suppress contributions from coherent magnetization transfer to the cross peak intensities, mixing times shorter than 100 ms were stochastically modulated with a modulation amplitude of ±5 ms (Macura *et al.*, 1981). For measurements in H_2O the solvent resonance was suppressed by selective, continuous irradiation at all times except during data acquisition (Anil Kumar *et al.*, 1980*b*; Wider *et al.*, 1983). To end up with a 1024 × 1024 point frequency domain data matrix, which corresponds to the digital resolution given in the legends to Figures 1 and 2, the time domain matrix was expanded to 1024 points in t_1 and to 2048 points in t_2 by "zero-filling". Prior to Fourier transformation the time domain data matrix was multiplied in the t_1 direction with a phase-shifted sine bell, $\sin(\pi(t+t_o)/t_s)$, and in the t_2 direction with a phase-shifted sine squared bell, $\sin^2(\pi(t+t_o)/t_s)$. The length of the window functions, t_s, was adjusted for the bells to reach zero at the last experimental data point in the t_1 or t_2 direction, respectively. The phase shifts, t_o/t_s, were 1/64 and 1/128 in the t_1 and t_2 directions, respectively. The spectra were obtained in the absolute value mode.

Exploratory investigations of the dependence on τ_m of the intensities of several well separated NOESY cross peaks showed that after a rapid initial growth

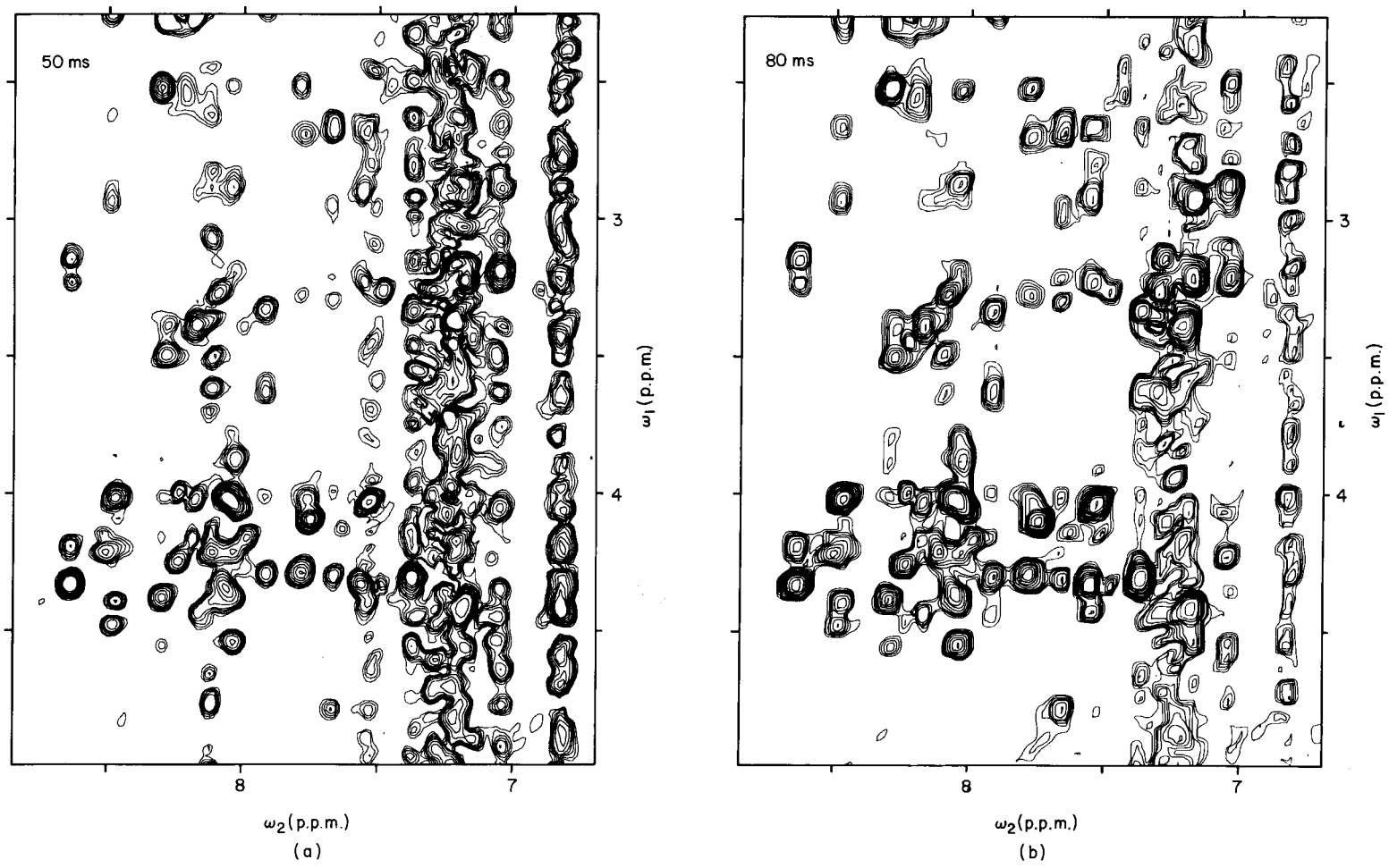

FIG. 1(a) and (b).

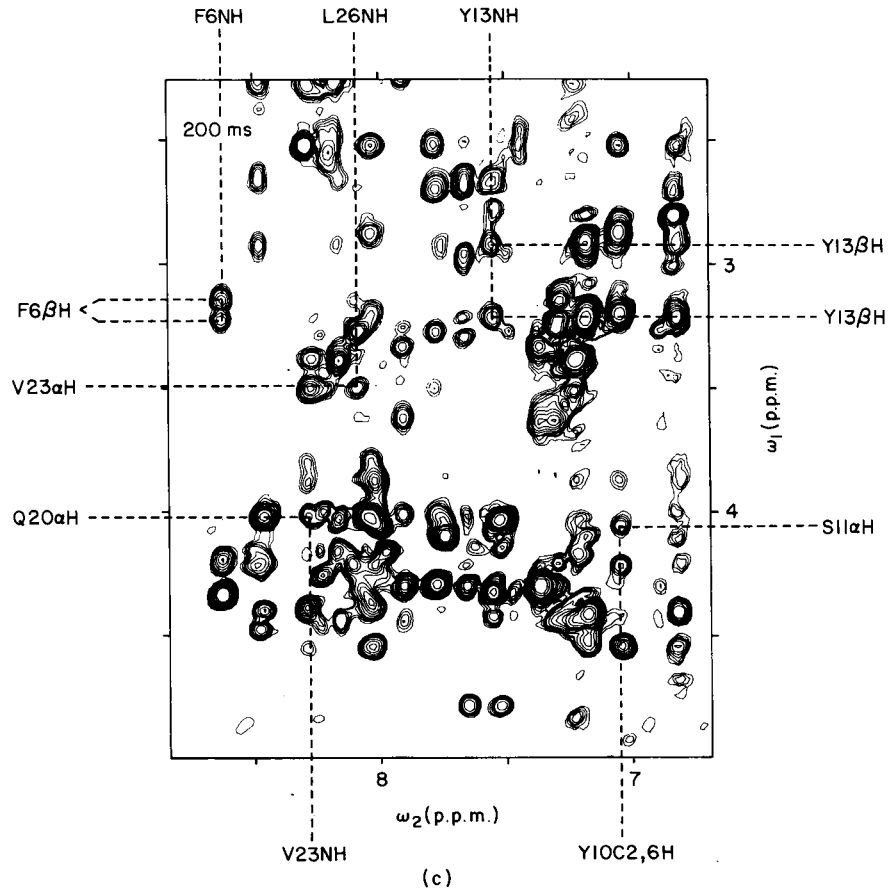

FIG. 1. (a) to (c) Contour plots of the spectral region ($\omega_1 = 2{\cdot}3$ to $4{\cdot}8$, $\omega_2 = 6{\cdot}7$ to $8{\cdot}8$) from 3 absolute value 500 MHz ^{1}H NOESY spectra of glucagon bound to perdeuterated dodecylphosphocholine micelles in H_2O solution. The 5 spectra were recorded under identical conditions (see the text), except for the different mixing times indicated in the upper left-hand corner. The digital resolution is 7·5 Hz/point. The contour levels are 2000, 2500, 3000, 3500, 4000, 5000, 6250, 7500, 8750, 10,000, 12,500, 15,000, 20,000, 25,000, 35,000, 50,000, 75,000, 100,000. As an illustration selected cross peaks are identified in spectrum (c), where the broken lines connect the cross peaks with the assignments of the 2 interacting resonances.

during the period $\tau_m = 0$ to $\sim$ 150 ms, the peak heights varied only little when τ_m was further increased. The intensity at $\tau_m = 1000$ ms was between 20 and 60% of the maximum intensity attained near $\tau_m = 200$ ms, depending on the cross peak considered. On the basis of these observations, mixing times of 30, 50, 80, 130 and 200 ms were selected for the NOESY spectra used for the present study (Figs 1 and 2; complete NOESY spectra of MB-glucagon in ^2H_2O (Bösch *et al.*, 1981) and in H_2O (Wider *et al.*, 1982) were previously presented). To ensure that the five spectra for each sample could be directly compared, they were recorded on five consecutive days without removing the sample from the spectrometer between measurements. Subsequently, identical data handling was used for each of the five

data sets and contour plots with identical contour levels were obtained (legends to Figs 1 and 2). The spectra in H_2O were recorded from 400 measurements, with t_1 values from 0·3 to 38 ms. 128 transients were accumulated for each value of t_1. The spectra in ^2H_2O were obtained from 420 measurements, with t_1 values from 0·3 to 50 ms. 144 transients were accumulated for each value of t_1.

Figure 1 illustrates the variation of the NOESY spectra when different mixing times were employed. The spectrum with $\tau_m = 50$ ms contains intense, vertical bands of noise at 6·8 p.p.m., from 7·0 to 7·6 p.p.m. and at 8·1 p.p.m. These artifacts originate from the presence of intense, sharp diagonal peaks of aromatic protons (Wider *et al.*, 1982) and represent "t_1 noise" as well as peak intensity smeared out along ω_1 by the random modulation of the mixing time τ_m (Macura *et al.*, 1981). In the spectra with longer mixing times these perturbations become less important, primarily because of the decrease of the diagonal peak intensities relative to the cross peak intensities (Anil Kumar *et al.*, 1981).

3. Regular Secondary Structures from Typical Patterns of NOESY Cross Peaks Between Backbone Hydrogen Atoms

Regular secondary structures in proteins contain characteristic patterns of short distances between amide, C^α and C^β protons (Billeter *et al.*, 1982; unpublished results). For example the distance d_1† is 2·2 Å in the extended polypeptide chain of a β-structure and 3·5 Å in the α-helix. In β-structures there are no other intrachain distances between backbone protons which would be shorter than 4·0 Å (Billeter *et al.*, 1982). In a regular α-helix one has further the short distances $d_2 = 2{\cdot}8$ Å and $d_1(i,i+3) = 3{\cdot}4$ Å, and $d_5(i,i+3)$ varies between 2·5 Å and 4·4 Å, depending on χ^1_{i+3}. For practical use we have, overall, that a succession of very short ($\lesssim$2·5 Å) distances d_1 in a peptide segment is indicative of an extended chain and a succession of short distances d_2 and/or $d_1(i,i+3)$ and/or $d_5(i,i+3)$ indicates that the polypeptide adopts a helical secondary structure.

The NOESY spectra of MB-glucagon in H_2O and in ^2H_2O were screened for cross peaks that correspond to d_1, d_2, $d_1(i,i+3)$ and $d_5(i,i+3)$. For this we used the previously published chemical shifts (Wider *et al.*, 1982) to determine the locations (ω_1,ω_2) where all the constraints d_1, d_2, $d_1(i,i+3)$ and $d_5(i,i+3)$ would be manifested in the NOESY spectra. From inspection of the experimental spectra (Figs 1 and 2) three situations were distinguished. (1) Connectivities

† The symbols used for characteristic $^1H-^1H$ distances in polypeptides were introduced elsewhere (Billeter *et al.*, 1982). i and j identify two residues in the same polypeptide chain. $d_1(i,j)$ is the distance between the C^α proton of residue i and the amide proton of residue j; for the sequential connectivity $d_1(i,i+1)$ the abbreviated symbol d_1 is used. $d_2(i,j)$ is the distance between the amide protons of residues i and j; for the sequential connectivity $d_2(i,i+1) \equiv d_2(i+1,i)$ the abbreviated symbol d_2 is used. $d_3(i,j)$ is the distance between the amide proton of residue j and the nearest C^β proton of residue i; for the sequential connectivity $d_3(i,i+1)$ the abbreviated symbol d_3 is used. $d_5(i,j)$ is the distance between the C^α proton of residue i and the nearest C^β proton of residue j.

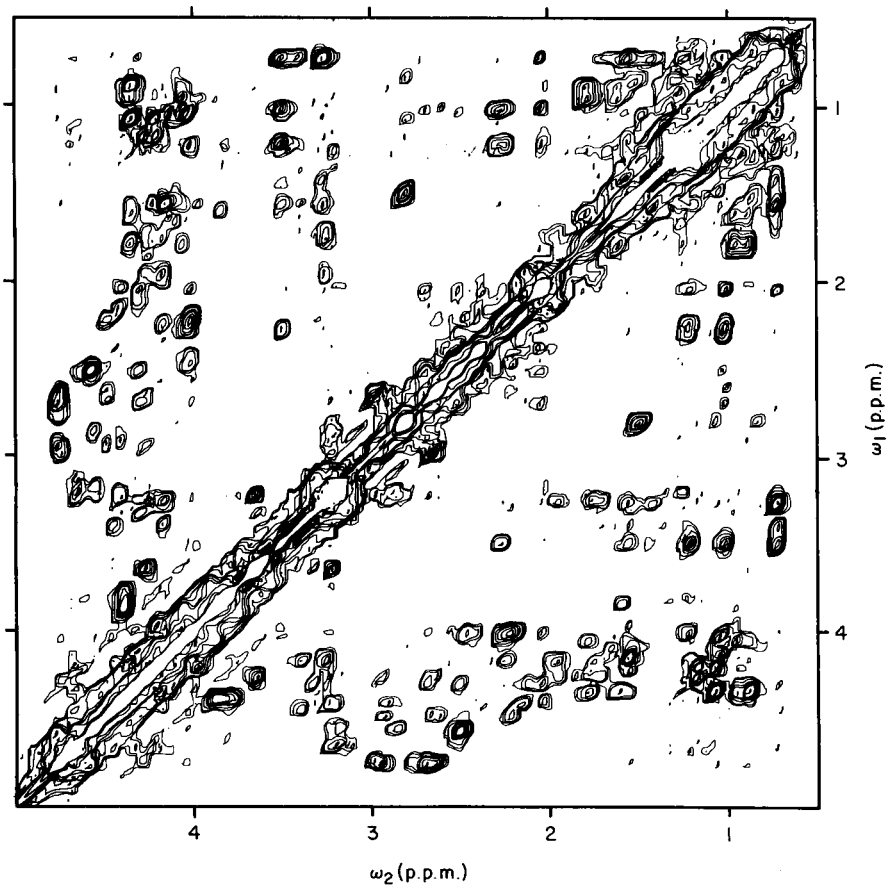

FIG. 2. Contour plot of the spectral region (ω_1 = 0·5 to 5·0 p.p.m., ω_2 = 0·5 to 5·0 p.p.m.) from a symmetrized, absolute value 500 MHz ^{1}H NOESY spectrum of glucagon bound to perdeuterated DPC micelles in ^2H_2O solution. The mixing time was 200 ms. Identical spectra were recorded with mixing times of 30, 50, 80 and 130 ms (see the text). The digital resolution is 6·4 Hz/point. The contour levels are 2000, 3000, 4000, 5000, 6500, 9000, 10,500, 12,000, 15,000 18,000, 23,000, 30,000, 40,000, 60,000, 80,000, 100,000.

manifested by resolved and unambiguously assigned cross peaks in the NOESY spectra (indicated in Fig. 3 by + or |—|, respectively). (2) Connectivities not manifested by cross peak intensity in the NOESY spectra. (3) Connectivities that would be overlapped with other peaks in a crowded region of the spectrum (indicated in Fig. 3 by ○ or | · · · · |, respectively). The connectivity patterns thus obtained (Fig. 3) indicate that the peptide segments from residues 3 to 9 and 14 to 17 adopt a predominantly extended secondary structure. For residues 12 to 14 formation of a loop or a turn is indicated and for the segment from residues 18 to 27 there is evidence for a helical structure.

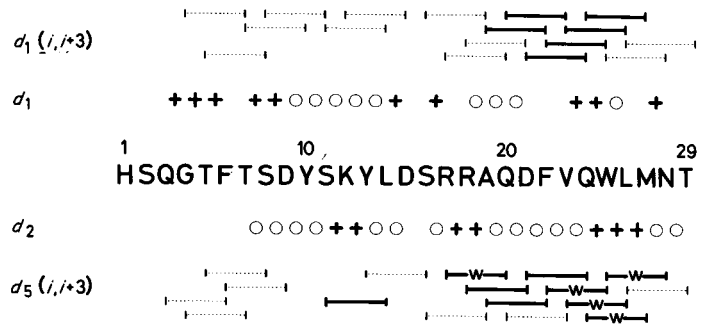

FIG. 3. Amino acid sequence of glucagon and NOESY connectivities d_1, d_2, $d_1(i, i+3)$ and $d_5(i, i+3)$ used for an initial, qualitative delineation of the secondary structure of MB-glucagon. The following symbols are used for d_1 and d_2, and for $d_1(i, i+3)$ and $d_5(i, i+3)$, respectively: (+) and (⊢—⊣) indicate that the connectivity was positively identified and the intensity of the cross peak was at least 2 contours in the NOESY spectrum recorded with a mixing time of 80 ms. Since the distance d_5 in an α-helix can vary appreciably because of the dependence on χ^1 (see the text) weak connectivities manifested by cross peaks with less than 2 contours after a mixing time of 80 ms but higher intensities at $\tau_m = 130$ and 200 ms are also indicated (⊢—w—⊣); these connectivities are not included in Table 2. (○) and (⊢ · · · ⊣) indicate that the presence or absence of the connectivity could not be determined unambiguously because of overlap in crowded spectral regions; absence of a symbol indicates that there is no connectivity manifested in the NOESY spectra.

4. Structural Interpretation of the NOESY Spectra With the Use of Distance Geometry Calculations

(a) *Correlation between intensity of the NOESY cross peaks and ^1H–^1H distance constraints*

When macromolecular species are studied at high magnetic fields, spin diffusion can become quite efficient (Hull & Sykes, 1975; Kalk & Berendsen, 1976) so that the information on proton–proton distances in a NOESY spectrum recorded with a fixed mixing time may be masked. On the other hand the initial build-up rates of NOEs in macromolecules are simply related to the inverse sixth power of the distance between the different groups of hydrogen atoms under consideration and can therefore provide the information needed for studies of the spatial molecular structure (Gordon & Wüthrich, 1978; Wagner & Wüthrich, 1979; Anil Kumar *et al.*, 1981; Bothner-By & Noggle, 1979; Braun *et al.*, 1981; Wüthrich *et al.*, 1982; Dobson *et al.*, 1982). When working with complex systems such as MB-glucagon, certain practical aspects must also be considered. For example, since the signal-to-noise ratio of NOESY spectra recorded with short mixing times is inherently poor (Fig. 1), it was not practical to measure initial NOE build-up rates by a series of experiments with different short τ_ms. Instead a NOESY spectrum with a long mixing time and correspondingly favourable signal-to-noise ratio was employed to properly define the location of the cross peaks in the two-dimensional spectrum (Figs 1(c) and 2). The peak intensities at these locations in spectra recorded with shorter mixing times but otherwise identical conditions (Fig. 1(a) and (b)) were then used as an approximate manifestation of the initial build-up rates.

When one correlates ^{1}H–^{1}H distances in proteins with NOEs the internal

motions of the macromolecular structure (Wüthrich & Wagner, 1978; Gurd & Rothgeb, 1979; Karplus & McCammon, 1981) must also be considered. These may affect both the effective rotational correlation time for dipole–dipole coupling between different protons and the distance between the protons (Braun *et al.*, 1981). For MB-glucagon negative NOEs could be observed between neighbouring protons within any of the residues 3 to 29 (Bösch *et al.*, 1980) and it was demonstrated that the negative NOEs seen by one-dimensional n.m.r. experiments correspond to the cross peaks in NOESY (Bösch *et al.*, 1981). This shows that with the exception of the N-terminal residues His1 and Ser2, the glucagon polypeptide chain is immobilized by the binding to the micelle and indicates that the rotational motions of the vectors joining different hydrogen atoms are effectively restricted to the overall tumbling motions of the micelles. In the semi-quantitative interpretation of the NOESY spectra used to prepare the input for the distance geometry calculations we therefore assumed a common correlation time for all the observed dipole–dipole interactions. It can be shown that in those locations of the molecular structure where the effective rotational correlation time is further affected by intramolecular motions, the proton–proton distances derived from the NOE data with the use of this assumption are upper limits to the actual distances (Braun *et al.*, 1981).

In a completely rigid structure the relative NOE build-up rates between different pairs of hydrogen atoms would correspond to $1/d^6$, where d is the distance between the interacting protons (Noggle & Schirmer, 1971). We have used this "rigid model" treatment for NOEs between hydrogen atoms that are separated by a sufficiently small number of bonds in the covalent polypeptide structure so that the relative spatial locations are determined by at most three torsion angles. For hydrogen atoms that are further separated in the covalent

TABLE 1

Correlations between intensity of the cross peaks in the NOESY spectra of MB-glucagon recorded in H_2O and constraints on ^1H–^1H distances, which were used as the input for the distance geometry calculations

NOESY		d (Å)† (rigid model)	d_{max} (Å)† (uniform averaging model)
τ_m (ms)	Intensity‡		
50	≥10	≤2·4	≤3·0
50	6–9	≤2·7	≤3·0
50	2–5	≤3·1	≤4·0
80	≥2	≤4·0	≤5·0

† The rigid model was applied for NOEs between hydrogen atoms that are separated by a sufficiently small number of bonds in the covalent structure so that the ^1H–^1H distance can be determined by 3 or less torsion angles about single bonds. In all other cases the uniform averaging model was applied (see the text). In the distance geometry calculations d or d_{max} were used as upper limits for the distance between the 2 groups of protons that are connected by the cross peak.

‡ The intensity of the NOESY cross peaks is given as the number of contour lines in the contour plots of the spectra. The contour levels are listed in the legend to Fig. 1.

TABLE 2

Distance constraints between groups of hydrogen atoms in the three-dimensional structure of glucagon bound to perdeuterated dodecylphosphocholine micelles obtained from NOESY spectra in H_2O and in ^2H_2O

	Distance constraints in Å†		
Sequential	Intra-residue	Long range backbone	Long range
His1			
Ser2			
Gln3			
HA G4 HN 2·7			
Gly4			
HN Q3 HA 2·7			
Thr5			
HN T5 HA 2·7	HN T5 MG 4·1 m (8·06, 1·07)		HB F6 QR 7·4 q (4·19, 7·27)
HA F6 HN 2·4	HA T5 MG 3·0 i (4·34, 1·07)		HB T7 MG 4·0 m (4·19, 7·17)
HB F6 HN 2·7			MG F6 HN 6·0 m (1·07, 8·63)
Phe6			
HN T5 HA 2·4	HN F6 PB 3·3 i (8·63, 3·13), (8·63, 3·22)		HN T5 MG 6·0 m (8·63, 1·07)
HN T5 HB 2·7	HA F6 QR 4·7 r (4·64, 7·27)		HA Y10 C1 7·0 r (4·64, 7·06)
HN F6 PB 3·3 i			QR T5 HB 7·4 q (7·27, 4·19)
			QR T7 HA 7·4 q (7·27, 4·23)
			QR T7 MG 8·4 qm (7·27, 1·17)
Thr7			
HN T7 HA 2·7	HN T7 HB 3·1 (8·11, 4·24)		HA F6 QR 7·4 q (4·23, 7·27)
HN T7 HB 3·1	HN T7 MG 3·4 m (8·11, 1·17)		HA Y10 C1 6·0 r (4·23, 7·06)
HA S8 HN 2·7	HA T7 MG 3·7 i (4·23, 1·17)		MG T5 HB 4·0 m (1·17, 4·19)
			MG F6 QR 8·4 qm (1·17, 7·27)
			MG S8 HN 5·0 m (1·17, 8·04)
			MG Y10 C1 8·0 mr (1·17, 7·06)
Ser8			
HN T7 HA 2·7	HN S8 PB 3·3 i (8·04, 3·86)		
HN S8 PB 3·3 i			HN T7 MG 5·0 m (8·04, 1·17)
HA D9 HN 2·7			
Asp9			
HN S8 HA 2·7	HN D9 PB 3·3 i (8·30, 2·52)		
HN D9 PB 3·3 i			PB Y10 C1 7·0 mr (2·52, 7·06)

Tyr10			
HN D9 PB 4·1 m	HN Y10 PB 3·3 i (8·05, 2·85), (8·05, 3·18)		HA Y13 C1 6·0 r (4·56, 7·19)
HN Y10 PB 3·3 i			C1 F6 HA 7·0 r (7·06, 4·64)
			C1 T7 HA 6·0 r (7·06, 4·23)
			C1 T7 MG 8·0 mr (7·06, 1·17)
			C1 D9 PB 7·0 mr (7·06, 2·52)
			C1 S11 HA 7·0 r (7·06, 4·06)
			C1 L14 QD 7·4 rq (7·06,0·88), (7·06, 0·96)
			C4 L14 QD 7·4 rq (6·82, 0·96)
Ser11			
HN S11 HA 2·4		HA L14 PB 6·0 m (4·06, 1·61)	HA Y10 C1 7·0 r (4·06, 7·06)
HN K12 HN 3·1			HA L14 HG 5·0 (4·06, 1·77)
			HA L14 QD 5·4 q (4·06, 0·88), (4·06, 0·96)
Lys12			
HN S11 HN 3·1	HN K12 PG 4·1 m (7·75, 1·10)		HA Y13 C1 7·0 r (4·09, 7·19)
HN K12 HA 2·4	HA K12 PD 5·0 m (4·09, 1·50)		PB Y13 C1 8·0 mr (1·59, 7·19)
HN Y13 HN 4·0			
PB Y13 HN 4·1 m			
Tyr13			
HN K12 HN 4·0	HN Y13 PB 3·7 i (7·55, 2·89), (7·55, 3·23)		C1 Y10 HA 6·0 r (7·19, 4·56)
HN K12 PB 4·1 m			C1 K12 HA 7·0 r (7·19, 4·09)
HN Y13 PB 3·7 i			C1 K12 PB 8·0 mr (7·19, 1·59)
			C1 L14 QD 9·4 rq (7·19, 0·88)
Leu14			
HA D15 HN 2·7	HN L14 MD15·0 m (7·52, 0·96)	PB S11 HA 6·0 m (1·61, 4·06)	HG S11 HA 5·0 (1·77, 4·06)
	HN L14 MD25·0 m (7·52, 0·88)		QD Y10 C4 7·4 rq (0·96, 6·82)
	HA L14 QD 4·1 s (4·33, 0·88), (4·33, 0·96)		QD Y10 C1 7·4 rq (0·88, 7·06), (0·96, 7·06)
			QD S11 HA 5·4 q (0·88, 4·06), (0·96, 4·06)
			QD Y13 C1 9·4 rq (0·88, 7·19)
			QD D15 HN 7·4 q (0·88, 7·55)
Asp15			
HN L14 HA 2·7	HN D15 PB 3·3 i (7·55, 2·76)		HN L14 QD 7·4 q (7·55, 0·88)
HN D15 PB 3·3 i			
Ser16			
HA R17 HN 2·7			
Arg17			
HN S16 HA 2·7	HN R17 PB 3·3 i (8·50, 1·96)	HA Q20 PB 5·0 m (4·17, 2·25)	
HN R17 PB 3·3 i			
HN R18 HN 3·1			

TABLE 2 *(continued)*

Distance constraints in Å†			
Sequential	Intra-residue	Long range backbone	Long range
Arg18			
HN R17 HN 3·1	HN R18 PB 3·3 i (8·24, 1·86)		
HN R18 PB 3·3 i	HN R18 PG 4·1 m (8·24, 1·71)		
HN R18 HA 2·7			
HN A19 HN 3·1			
Ala19			
HN R18 HN 3·1	HN A19 MB 3·0 i (7·99, 1·56)	HA F22 HN 4·0 (4·17, 8·18)	HA F22 QR 7·4 q (4·17, 7·23)
HN A19 MB 3·0 i		HA F22 PB 6·0 m (4·17, 3·38)	MB F22 QR 8·4 mq(1·56, 7·23)
HN A19 HA 2·7		MB F22 PB 6·0 m (1·56, 3·38)	
HA Q20 HN 3·1		MB V23 HA 5·0 m (1·56, 3·49)	
MB Q20 HN 3·7 m			
Gln20			
HN A19 MB 3·7 m	HN Q20 PB 3·3 i (8·17, 2·25)	HA V23 HN 4·0 (4·02, 8·28)	HA V23 QG 6·4 q (4·02, 1·21)
HN A19 HA 3·1		PB R17 HA 5·0 m (2·25, 4·17)	
HN Q20 HA 2·7			
HN Q20 PB 3·3 i			
HA D21 HN 3·1			
PB D21 HN 4·1 m			
Asp21			
HN Q20 PB 4·1 m		HA Q24 HN 4·0 (4·47, 8·24)	
HN D21 HA 2·7		HA Q24 PB 5·0 m (4·47, 2·22)	
PB F22 HN 4·1 m			
Phe22			
HN D21 PB 4·1 m	HN F22 PB 3·3 i (8·18, 3·38)	HN A19 HA 4·0 (8·18, 4·17)	HA W25 H4 5·0 (4·44, 7·33)
HN F22 PB 3·3 i		HA W25 HN 5·0 (4·44, 7·92)	QR A19 HA 7·4 q (7·23, 4·17)
PB V23 HN 4·1 m		PB A19 HA 6·0 m (3·38, 4·17)	QR A19 MB 8·4 mq(7·23, 1·56)
		PB A19 MB 6·0 m (3·38, 1·56)	QR V23 QG 8·8 qq (7·23, 1·21)
			QR V23 HN 7·4 q (7·23, 8·28)
			QR V23 HA 7·4 q (7·23, 3·49)
			QR L26 QD 7·8 qq (7·23, 0·72)
			PB L26 QD 6·4 mq(3·38, 0·72)
Val23			
HN F22 PB 4·1 m	HN V23 HB 2·7 (8·28, 2·26)	HN Q20 HA 4·0 (8·28, 4·02)	HN F22 QR 7·4 q (8·28, 7·23)
HN V23 HB 2·7	HA V23 MG1 3·7 m (3·49, 1·21)	HA A19 MB 5·0 m (3·49, 1·56)	HA F22 QR 7·4 q (3·49, 7·23)
HN V23 HA 2·7	HA V23 MG2 3·7 m (3·49, 1·02)	HA L26 HN 3·0 (3·49, 8·09)	HA L26 QD 5·4 q (3·49, 0·72)
HA Q24 HN 3·1			QG Q20 HA 6·4 q (1·21, 4·02)

HB Q24 HN 3·1			QG F22 QR 8·8 qq (1·02, 7·23), (1·21, 7·23)
			QG Q24 HN 7·4 q (1·01, 8·24), (1·21, 8·24)
Gln24			
HN V23 HA 3·1	HN Q24 PB 3·3 m (8·24, 2·22)	HN D21 HA 4·0 (8·24, 4·47)	HN V23 QG 7·4 q (8·24, 1·02), (8·24, 1·21)
HN V23 HB 3·1		HA M27 HN 4·0 (4·00, 7·99)	
HN Q24 HA 2·7		HA N28 HN 4·0 (4·00, 7·66)	
HN Q24 PB 3·3 m		PB D21 HA 5·0 m (2·22, 4·47)	
HN W25 HN 3·1			
HA W25 HN 3·1			
PB W25 HN 4·1 m			
Trp25			
HN Q24 HN 3·1	HN W25 PB 3·3 i (7·92, 3·33), (7·92, 3·62)	HN F22 HA 5·0 (7·92, 4·44)	H2 L26 QD 6·4 q (7·37, 0·72)
HN Q24 HA 3·1	HN W25 H4 5·0 (7·92, 7·33)		H4 F22 HA 5·0 (7·33, 4·44)
HN Q24 PB 4·1 m	HA W25 H2 2·4 (4·30, 7·37)		H4 L26 HA 5·0 (7·33, 3·27)
HN W25 PB 3·3 i	PB W25 H2 3·7 m (3·33, 7·37), (3·67, 7·37)		H4 L26 PB 6·0 m (7·33, 1·57)
HN W25 HA 2·7	PB W25 H4 4·1 m (3·33, 7·33), (3·67, 7·33)		H4 L26 QD 5·4 q (7·37, 0·72)
HN L26 HN 4·0			H4 L26 HN 4·0 (7·33, 8·09)
PB L26 HN 4·1 m			H5 L26 PB 6·0 m (6·89, 1·57)
			H5 L26 QD 6·4 q (6·89, 0·72)
			H5 L26 HA 5·0 (6·89, 3·27)
			H6 L26 QD 7·4 q (7·11, 0·72)
Leu26			
HN W25 HN 4·0	HN L26 PB 3·7 m (8·09, 1·41)	HN V23 HA 3·0 (8·09, 3·47)	HN W25 H4 4·0 (8·09, 7·33)
HN W25 PB 4·1 m	HA L26 QD 4·1 s (3·27, 0·72)		HA W25 H4 5·0 (3·27, 7·33)
HN L26 HA 2·7			HA W25 H5 5·0 (3·27, 6·89)
HN L26 PB 3·7 m			QD F22 QR 7·8 qq (0·72, 7·23)
HN M27 HN 3·1			QD F22 PB 6·4 mq (0·72, 3·38)
PB M27 HN 4·1 m			QD V23 HA 5·4 q (0·72, 3·49)
			QD W25 H4 5·4 q (0·72, 7·33)
			QD W25 H2 6·4 q (0·72, 7·37)
			QD W25 H5 6·4 q (0·72, 6·89)
			QD W25 H6 7·4 q (0·72, 7·11)
			QD M27 HN 7·4 q (0·72, 7·79)
			QD M27 ME 8·4 qm (0·72, 2·04)
			PB W25 H4 6·0 m (1·57, 7·33)
			PB W25 H5 6·0 m (1·57, 6·89)
Met27			
HN L26 HN 3·1	HN M27 PB 3·7 i (7·79, 2·13)	HN Q24 HA 4·0 (7·79, 4·00)	HN L26 QD 7·4 q (7·79, 0·72)
HN L26 PB 4·1 m	HN M27 PG 4·1 m (7·79, 2·53), (7·79, 2·70)		ME L26 QD 8·4 qm (2·04, 0·72)
HN M27 HA 2·7			PG N28 HN 6·0 m (2·03, 7·66)
HN M27 PB 3·7 i			
HN N28 HN 3·1			
HA N28 HN 3·1			

TABLE 2 *(continued)*

	Distance constraints in Å†			
Sequential	Intra-residue	Long range backbone	Long range	
Asn28				
HN M27 HA 3·1	HN N28 PB 3·3 i (7·66, 2·68), (7·66, 2·97)	HN Q24 HA 4·0 (7·66, 4·00)	HN M27 PG 6·0 m (7·66, 2·03)	
HN N28 PB 3·3 i				
PB T29 HN 3·7 m				
Thr29				
HN N28 PB 3·7 m	HN T29 HB 4·0 (7·53, 4·11)			
HN T29 HB 4·0	HN T29 MG 4·1 m (7·53, 1·09)			
HN T29 HA 2·4	HA T29 MG 3·0 i (7·53, 1·07)			

† The amino acid residues are identified either by the 3-letter symbol or the 1-letter symbol followed by a number indicating the position in the amino acid sequence. HN stands for amide proton, HA for C^{α} proton, HB for C^{β} proton, HG for C^{γ} proton. C1 and C4 are the ring carbon atoms in positions 1 and 4 of the aromatic rings of Phe and Tyr, H2 and H4 to H7 are the protons attached to the indole ring carbons 2 and 4 to 7 of Trp. The following upper case letters indicate pseudo-atoms which substitute for a group of 2 or more protons and the lower-case letters indicate that the distance constraint indicated has been modified to account for the use of the pseudo-atom (see the text and Wüthrich *et al.*, 1983): M replaces the 3 protons of a methyl group and is located in the centre of the 3 proton positions, with the corrections factors $m = 1{\cdot}0$ Å (for long range constraints) and $i = 0{\cdot}6$ Å (for certain short range constraints). P replaces the 2 protons of a methylene group and is located in the centre between the 2 proton positions, m and i apply as for the methyl group. Q replaces the 2 methyl groups of Val or Leu and is located centrally with respect to the 2 methyl groups. The correction q is 2·4 Å, except that $s = 1{\cdot}7$ Å applies for the intraresidue constraint from $C^{\alpha}H$ to QD in Leu. QR is at the centre of the ring protons C2H, C3H, C5H and C6H in Phe or Tyr, the correction q us 2·4 Å. The correction $r = 2{\cdot}0$ Å applies for Tyr or Phe when one replaces either the 2,6 ring protons by C1 or the 3,5 ring protons by C4, and for the intraresidue constraint from $C^{\alpha}H$ to QR.

The rows following the 3-letter symbol for the amino acid list all the distance constraints which involve hydrogen atoms of this residue. For each residue the distance constraints are grouped into 4 classes, which are presented in the 4 columns. *Sequential* constraints are those between amide, C^{α} and C^{β} protons within each residue and in sequentially neighbouring residues. *Intraresidue* constraints are between side-chain hydrogen atoms and any other proton of the same residue. *Long range backbone* constraints are between amide, C^{α} and C^{β} protons in residues that are not nearest neighbours in the sequence. *Long range* constraints are all those that have not been listed in the first 3 columns. Note that the distance constraints between the backbone hydrogens and $C^{\beta}H$ of the same residue are listed in both the first and second column and that with the exception of the intra-residue constraints all entries appear twice, for 2 different residues. In each column the first entry identifies a hydrogen atom in the amino acid indicated by the 3-letter symbol. The second entry indicates a different residue and the position in this residue to which a short distance constraint exists. The third entry is the distance constraint in Å, possibly with symbols indicating that one or several correction factors were added to the NOESY distance constraint to allow for the use of pseudo-atoms. The fourth entry indicates the location of the NOESY cross peak in the ω_1–ω_2 plane (Figs 1 and 2), where (ω_1, ω_2) are given in p.p.m. This information has been omitted for the sequential constraints, since these were previously described in detail (Wider *et al.*, 1982).

structure the NOEs were interpreted with a previously described "uniform averaging model", which accounts for time variations of the proton–proton distances d in a flexible structure (Braun *et al.*, 1981). In this treatment the distance between two hydrogen atoms is allowed to vary uniformly between a minimum distance d_{min}, which is taken to be the sum of the van der Waals' radii, and a maximum distance d_{max}, which is determined by the NOE data. For a given NOE build-up rate d_{max} obtained with the uniform averaging model is always longer than the corresponding d obtained with the rigid model. Examples of corresponding distance constraints computed from the same NOE data for the rigid and the flexible model are 2·3 and 2·7 Å, 2·6 and 3·7 Å, and 2·9 and 4·5 Å (Braun *et al.*, 1981).

Based on the fundamental considerations outlined above and using NOESY cross peaks between neighbouring hydrogen atoms in the covalent structure for calibration purposes, the rules in Table 1 were used for the interpretation of the NOESY data recorded in H_2O solution of MB-glucagon. A similar set of rules (with different numbers to account for the different contour levels in the ^2H_2O and H_2O spectra, see legends to Figs 1 and 2) was used for the ^2H_2O spectra. The resulting distance constraints are listed in Table 2, where the numbers accompanied by one or two lower case letters include systematic corrections that account for the use of "pseudo-structures" for the amino acids (see the text below, footnotes to Table 2 and Wüthrich *et al.* (1983)).

(b) *Spatial structure determination with distance geometry calculations*

The problem to be solved by distance geometry calculations is the following (Crippen & Havel, 1978; Havel *et al.*, 1979; Braun *et al.*, 1981; Wako & Scheraga, 1982). Given upper limits and lower limits for the distances between the N atoms in a molecular structure, what are the conformations that are compatible with these distance constraints? In the present application to a polypeptide chain, the upper and lower limits for the distances between covalently linked atoms correspond to a set of standard geometries for the common amino acid residues (Momany *et al.*, 1975). For non-bonding interactions the lower limits correspond to the sum of the van der Waals' radii of the atoms considered and the upper limits on the interatomic distances are obtained from the analysis of the NOESY spectra (Table 2). Because of the limited resolution of n.m.r. spectra and the limitation of NOESY distance measurements to distances smaller than ~5·0 Å, n.m.r. provides an incomplete set of relatively inaccurate distance constraints. Therefore, distance geometry calculations will usually not provide a unique conformation, but repeated computations using the same set of n.m.r. distance constraints will provide a group of somewhat different molecular geometries that are all compatible with the experimental data (Fig. 4).

The distance geometry algorithm used was previously described in detail (Braun *et al.*, 1981). However, since a much more extensive set of distance constraints was available than for the earlier applications, new pseudo-structures for the amino acid residues were used for the input (Wüthrich *et al.*, 1983). In the pseudo-structures groups of hydrogen atoms for which no stereospecific

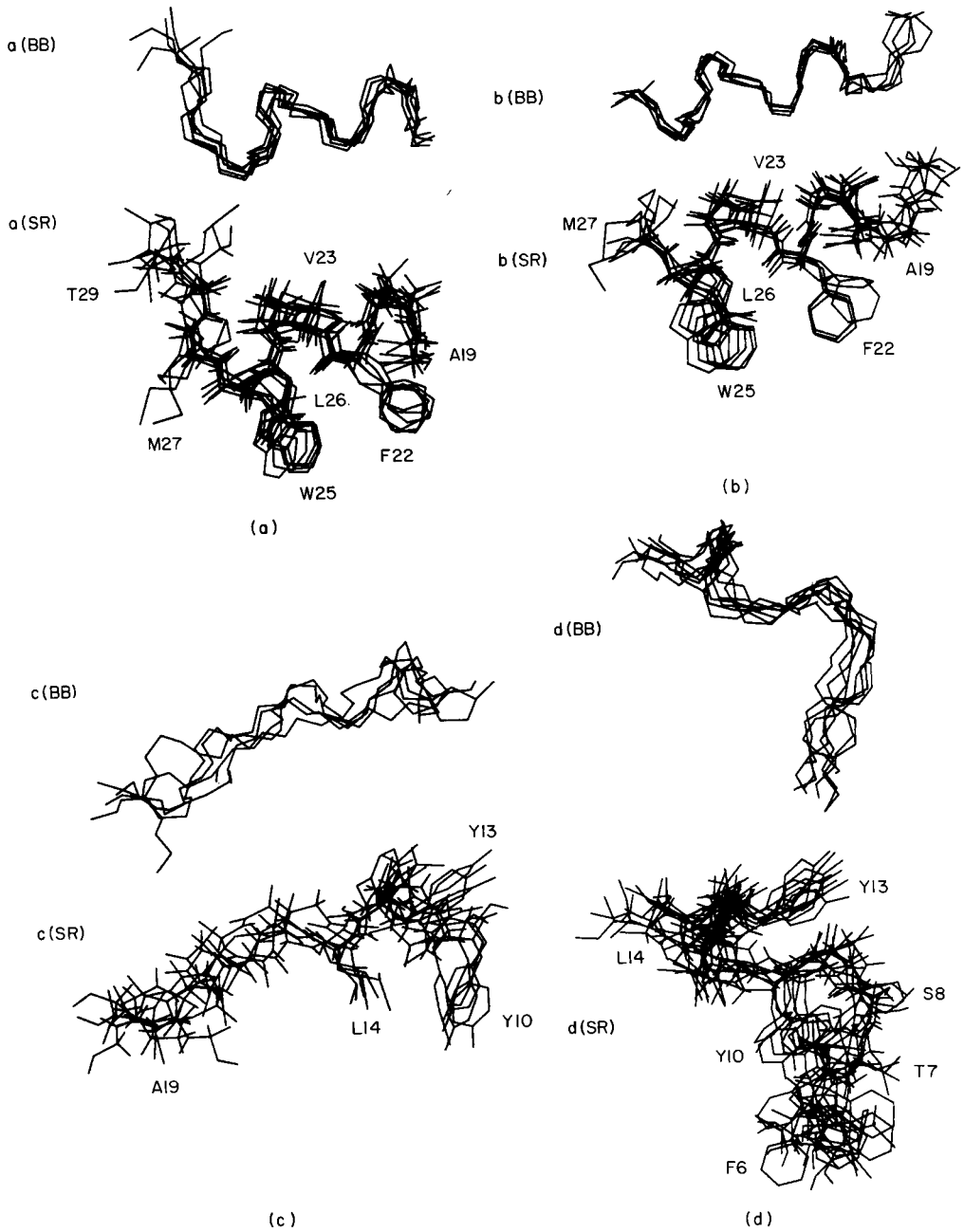

FIG. 4. Computer drawings of the spatial structure for 4 segments of MB-glucagon. From the final group of 10 computations for each segment all those structures that have satisfactory stereochemistry (see the text) are shown superimposed on each other. For each segment drawings of the backbone (BB) and of the "restrained side-chain" presentation (SR) are shown. SR includes the complete side-chains for the residues that are identified in the drawings and for which NOE constraints were observed for the peripheral protons (Table 2). For the other residues only the backbone atoms, including the C^{α} and amide protons and C^{β} are shown. CH_3 and CH_2 groups are represented by the spherical pseudo-atoms M and L, respectively, and the CH groups in the aromatic rings of Phe and Tyr by the spherical pseudo-atom K (see the text). (a) Segment 19–29, 6 structures are superimposed; (b) segment 17–27, 5 structures; (c) segment 10–20, 5 structures; (d) segment 5–15, 8 structures.

assignments of the n.m.r. lines were obtained (Wider *et al.*, 1982) are replaced by a pseudo-atom located in a central position relative to the protons for which it substitutes, and the pseudo-atom position is then used in the distance geometry calculations. The pseudo-atoms P, M, Q and QR are described in the footnotes to Table 2. Since the distance constraints manifested in the NOESY spectra are of course always between real hydrogen atoms, the use of pseudo-atoms must be accounted for by the introduction of corrections to the experimentally determined distance constraints. A list of the corrections used is given in the footnotes to Table 2 and a complete description of the dimensions of the pseudo-structures for amino acids to be used in n.m.r. structure determinations is contained in the following paper (Wüthrich *et al.*, 1983).

To reduce the computing time a different type of pseudo-atoms, M, L and K, was introduced to represent, respectively, the van der Waals' dimensions of the CH_3 and CH_2 groups and the CH fragments in positions 2, 3, 5 and 6 of the aromatic rings of Phe and Tyr. M is a sphere of radius 1·8 Å located in the centre of the three methyl protons. (Its centre coincides with the pseudo-atom position M used as the point of reference for NOESY distance constraints involving methyl groups.) L is a sphere of radius 1·6 Å located at the position of the methylene carbon. K is a sphere of radius 1·5 Å located at the position of the ring carbon atom. K, L and M are also used in the representations of the MB-glucagon structures in Figures 4 to 8.

Inspection of the distance constraints in Table 2 shows that no NOEs were observed between residues that would be further apart than five positions in the amino acid sequence. In the peptide segment 14–17 there are actually exclusively intra-residue and sequential NOEs. To save computer time the distance geometry calculations were therefore performed separately for four segments of the polypeptide chain, i.e. 5–15, 10–20, 17–27 and 19–29. This seemed justified, since in the absence of real "long range" distance constraints the overall shape of the molecule could anyway not be determined in a meaningful way. Approximate structures for MB-glucagon 5–29 were then obtained by a program that would minimize the root-mean-square-distance between the overlapping parts of the four segments.

For each peptide segment ten independent computer runs were made, with five G-cycles and 2000 refinement-cycles (Braun *et al.*, 1981). The weight of the NOE distance constraints relative to the distance constraints imposed by the covalent structure and the van der Waals' radii was taken to be 0·01. The criterion for proper convergence of each individual computer run was that none of the distance constraints by covalent bonds or by the van der Waals' radii was violated by more than 0·1 Å. Structures with larger violations were discarded. For those structures that had reasonable stereochemical properties (typically 5 to 8 out of a group of 10 computations) a record was made of all short distances between hydrogen atoms. Using the previously-determined (Wider *et al.*, 1982) chemical shifts the corresponding cross peak positions in a hypothetical NOESY spectrum of the computed conformation were calculated and compared with the input data (Table 2) and with the experimental spectra. This procedure was repeated three times to search for and subsequently eliminate inconsistencies between the

molecular geometry, the input data and the experimental spectra. Quite generally, most of the short distance constraints that were not contained in the input data were found to correspond to peaks in poorly resolved regions of the NOESY spectra, so that they could not be reliably assigned and evaluated in the initial spectral analysis. The checks on internal consistency between experiment and structures obtained included also a search for large violations of NOE distance constraints that would occur in the majority of the results. For the structures with satisfactory stereochemistry at least 90% of the input NOE constraints were typically fulfilled with violations of the order of 0·1 to 0·4 Å, which seems reasonable in view of the semi-quantitative data analysis (Tables 1 and 2). Consistently larger violations were found for the d_1-connectivities from Ala19 to Gln20, Val23 to Gln24 and Gln24 to Trp25, which were all estimated to be 3·1 Å (Table 2). These constraints were subsequently omitted in the last group of ten computations for the segments 17–27 and 19–29, which produced no significant changes in the molecular geometries. This showed quite convincingly that the structure determination in the region of residues 19 to 25 is dominated by the longer range constraints, such as $d_1(i,i+3)$ and $d_5(i,i+3)$ (Fig. 3).

5. The Spatial Structure of MB-Glucagon from n.m.r. Distance Constraints and Distance Geometry Calculations

The results of the above described structural interpretation of the NOESY distance constraints in MB-glucagon (Table 2) with the use of a distance geometry algorithm are presented in Figures 4 to 6 and Table 3. From the final ten computer runs for each of the segments 5–15, 10–20, 17–27 and 19–29, eight conformers for the segment 5–15, five conformers for each of the segments 10–20 and 17–29, and six conformers for 19–29 had stereochemically acceptable spatial structures (Fig. 4 and Table 3). Four different presentations of the molecular structures showing different amounts of detail are employed. In the presentation BB only the backbone atoms N, C^{α} and C′ are included. SB contains the backbone and all those atoms that are directly bonded to one of the backbone atoms, i.e. it includes the fragments N—H, $C^{\alpha}H$—C^{β} and C′=O. SR stands for "restricted side-chain representation", i.e. in addition to the atoms shown in SB the complete side-chains are shown for all those residues where the conformation is determined by NOEs involving the peripheral hydrogen atoms (Table 2). These residues are Thr5, Phe6, Thr7, Ser8, Tyr10, Tyr13, Leu14, Ala19, Phe22, Val23, Trp25, Leu26, Met27 and Thr29. In this presentation CH_3, CH_2 and the CH fragments of the rings of Tyr and Phe are replaced by the spherical pseudo-atoms M, L and K, respectively. In the HA (heavy atom) presentation all heavy atoms of all side-chains are included. This presentation is mainly added in Table 3 as a check for the sampling property of the algorithm. The structures in Figure 4 correspond to the presentations BB and SR. For each of the four segments in Figure 4 the conformer with the smallest deviations of the covalent structure from the standard ECEPP geometry (Momany *et al.*, 1975) was selected and mono- and stereo-drawings of these species are shown in Figure 5. In Figure 6 the four partial structures of MB-glucagon in Figure 5 were combined so as to minimize

TABLE 3

Comparison of all stereochemically acceptable conformers obtained from the final group of 10 computer runs for each of the 4 fragments of MB-glucagon

MB-glucagon segment	Number of conformers	Average r.m.s.d. (Å)†			
		BB‡	SB‡	SR‡	HA‡
5–15	8	1·52	2·03	2·63	2·79
10–20	5	2·08	2·67	3·06	4·09
17–27	5	0·85	1·18	1·59	1·94
19–29	6	1·15	1·47	1·77	2·07

The average of the root-mean-square distances between any 2 of the conformers is indicated.

† r.m.s.d. $= \left\{\frac{1}{N}\sum_{j=1}^{N}|Rx_j^u - x_j^v|^2\right\}^{\frac{1}{2}}$, where N is the number of atoms in the structure, x_j are the atomic co-ordinates, u and v indicate the 2 conformers which are compared, R is the rotation matrix which affords the best match in space between the 2 conformers. R was obtained using an algorithm proposed by McLachlan (1979).

‡ The following presentations of the molecular structures are used (see the text): BB, only the backbone atoms N, C^{α} and C′ are considered. SB, The backbone atoms with the directly bonded atoms are included in the calculation, i.e. N—H, $C^{\alpha}H—C^{\beta}$ and C′=O. SR, Restricted side-chain representation. Besides atoms of SB these structures contain all side-chain atoms (without hydrogens) for those residues, where NOE distance constraints were available up to the peripheral protons. These residues are T5, F6, T7, S8, Y10, Y13, L14, A19, F22, V23, W25, L26, M27, T29. HA, All heavy atoms are included.

the r.m.s.d. between the overlapping residues of any two segments. This Figure affords a survey of the secondary structure from residues 5 to 29 of MB-glucagon.

For the reasons discussed in section 4, above, the presently used procedures cannot be expected to provide a unique spatial polypeptide structure, but the result consists of a group of structures that are all compatible with the experimental data of Table 2. Each individual conformer is to be regarded as a typical member of the group of structures occupying the conformation space within the confines imposed by the NOE distance constraints, and not as an "average spatial structure" (Braun *et al.*, 1981). Therefore significant information on the quality of the results obtained comes primarily from comparison of the different structures in Figure 4 and Table 3.

Inspection of Table 2 shows that the number of NOE distance constraints used for the structure determination of the four segments 5–15, 10–20, 17–27 and 19–29 are, respectively, 51, 39, 77 and 79. There is a good correlation between the number of distance constraints used and the r.m.s.d. values in Table 3, i.e. smaller r.m.s.d. values prevail for the segments with a larger number of distance constraints per residue. The best constrained parts of the polypeptide chain are the backbone structures (BB) 17–27 and 19–29, which is also clearly seen in Figure 4. The r.m.s.d. values of 0·85 Å and 1·15 Å for BB 17–27 and BB 19–29, respectively, indicate that the atom positions in these fragments are determined nearly within the limiting uncertainty expected from the thermal fluctuations of

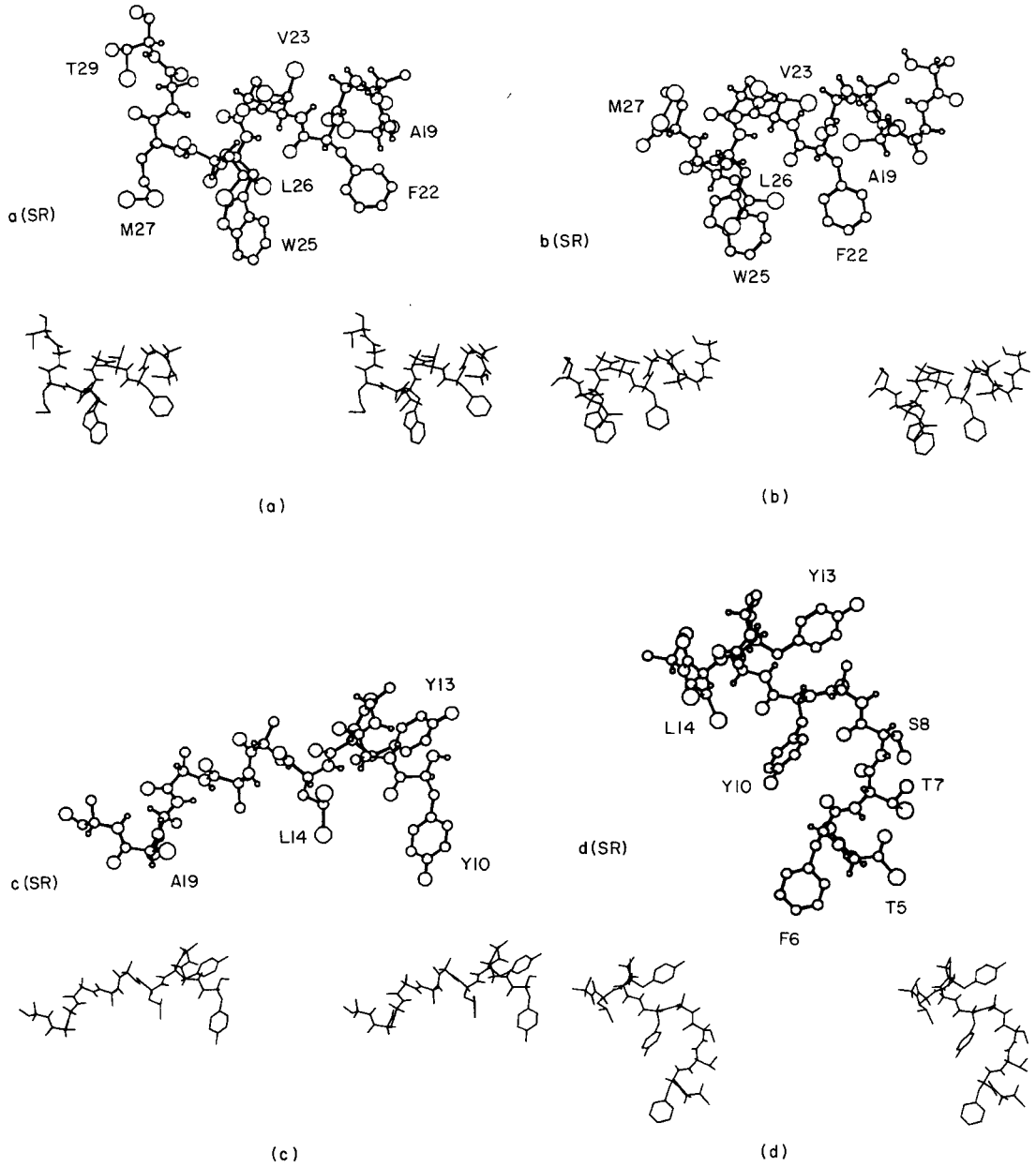

FIG. 5. Computer drawing of the stereochemically best structure (see the text) for each segment from the group of molecular geometries shown in Fig. 4. The SR presentation is shown and the same pseudo-atoms are used as in Fig. 4. For each segment a mono and a stereo drawing is shown. (a) Segment 19–29; (b) 17–27; (c) 10–20; (d) 5–15.

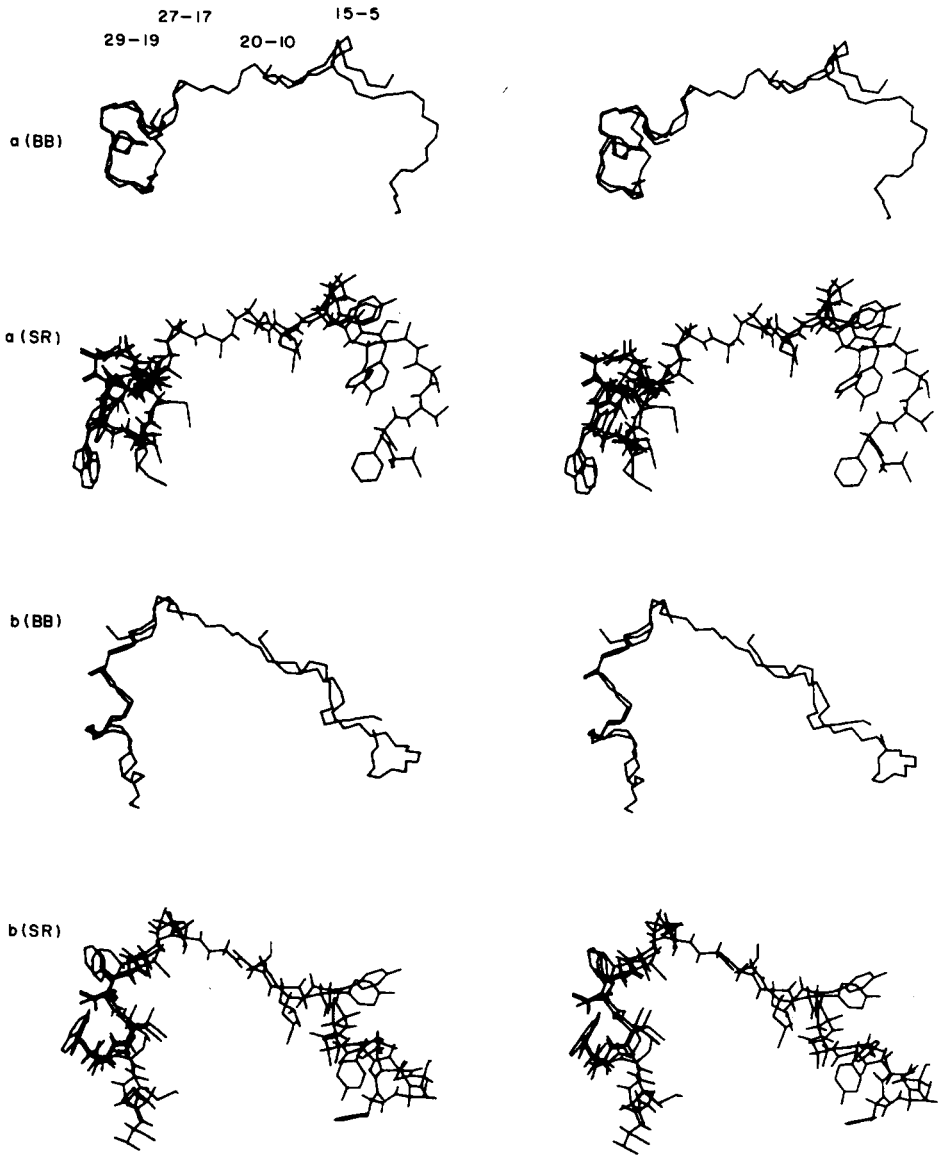

FIG. 6. Stereo drawings of the structure of MB-glucagon 5–29 obtained by combination of the structures for the 4 individually computed segments 5–15, 10–20, 17–27 and 19–29 in Fig. 5. The segments were fitted together so as to minimize the root-mean-square distance between the overlapping residues for each pair of segments. The r.m.s.d. values for the overlapping segments 11–14, 17–20 and 20–26 in the SR presentation are 1·7 Å, 1·7 Å and 1·1 Å, respectively. The BB and SR presentations of the structure are shown (see Fig. 4). The drawings (b) show the same structure as (a) after a 90° rotation about a horizontal axis aligned parallel to the projection plane.

the polypeptide chain. (For example, average root-mean-square fluctuations of 0·6 Å were observed for the C^{α} atom positions in simulations of the thermal motions in small globular proteins by molecular dynamics calculations (McCammon & Karplus, 1980).) The least restrained segment is 10–20. Here, the jump of the r.m.s.d. value from 3·06 Å to 4·09 Å between SR and HA indicates good sampling of the allowed conformation space, since half of the side-chains in this segment are only poorly constrained. Overall the numbers in Table 3 indicate and Figure 4 shows that there are many common features in all stereochemically acceptable conformers obtained from the same n.m.r. data in different computer runs, indicating that these are characteristic structural traits of MB-glucagon. These are discussed in the following in more detail.

The backbone conformation includes a predominantly extended polypeptide segment from residues 5 to 9, one helix-like turn formed by residues 10 to 14, another stretch of extended chain between residues 14 and 17 and three turns of a distorted α-helix from residues 17 to 29 (Fig. 6). In the segment 5–15 (Figs 4(d) and 5(d)) there is a clear spatial separation of the hydrophobic residues Phe6, Tyr10 and Leu14 from the charged or polar residues. If one connects the C^{β} atoms (which are indicated for all residues in the SR representations of Figs 4 and 5) separately for polar and non-polar residues, one finds two lines that are always on opposite sides of the polypeptide backbone and never cross each other. Tyr13 seems not to be included in the hydrophobic patch formed by residues 6, 10 and 14. In the C-terminal dodecapeptide a hydrophobic patch is formed by the side-chains of Ala19, Phe22, Val23, Trp25 and Leu26 (Figs 4(a) and (b) and 5(a) and (b)).

The overall shape of the molecule in Figure 6 is not reliably characterized by the NOE data, since no long range distance constraints extending over more than five residues were observed. However, the backbone outlines approximately the curvature of the dodecylphosphocholine micelles (Bösch *et al.*, 1980), and since MB-glucagon was found to be located near the micelle surface (Brown *et al.*, 1981), the overall shape of the molecule in Figure 6 might nonetheless coincide rather closely with the micelle-bound polypeptide.

In the structures of Figure 4, which were computed with relative weights of 0·01 for distance constraints by NOEs and 1 for distance constraints by the covalent structure, the three peptide bonds between residues 20 and 21, 23 and 24, and 26 and 27 deviate quite markedly from planarity, with ω-angles between 160 and 165°. To improve the planarity of the peptide bonds we have applied further refinement cycles to the stereochemically most satisfactory conformers in Figure 5, with a relative weight of 0·001 of the NOE distance constraints to the stereochemical distance constraints. After this regularization no ω-angle deviates more than 15° from planarity, which corresponds to an energy of about 2RT. This regularization involved only small changes in the structures (Fig. 7).

A plot of the ϕ–ψ torsion angles in the regularized structures of Figure 7 is presented in Figure 8. The sterically allowed regions in the ϕ–ψ plane (Ramachandran & Sasisekharan, 1968) are also indicated. The data points for MB-glucagon fall within or near to the allowed areas, as is also generally observed in X-ray structures of proteins (Richardson, 1981). Compared to typical

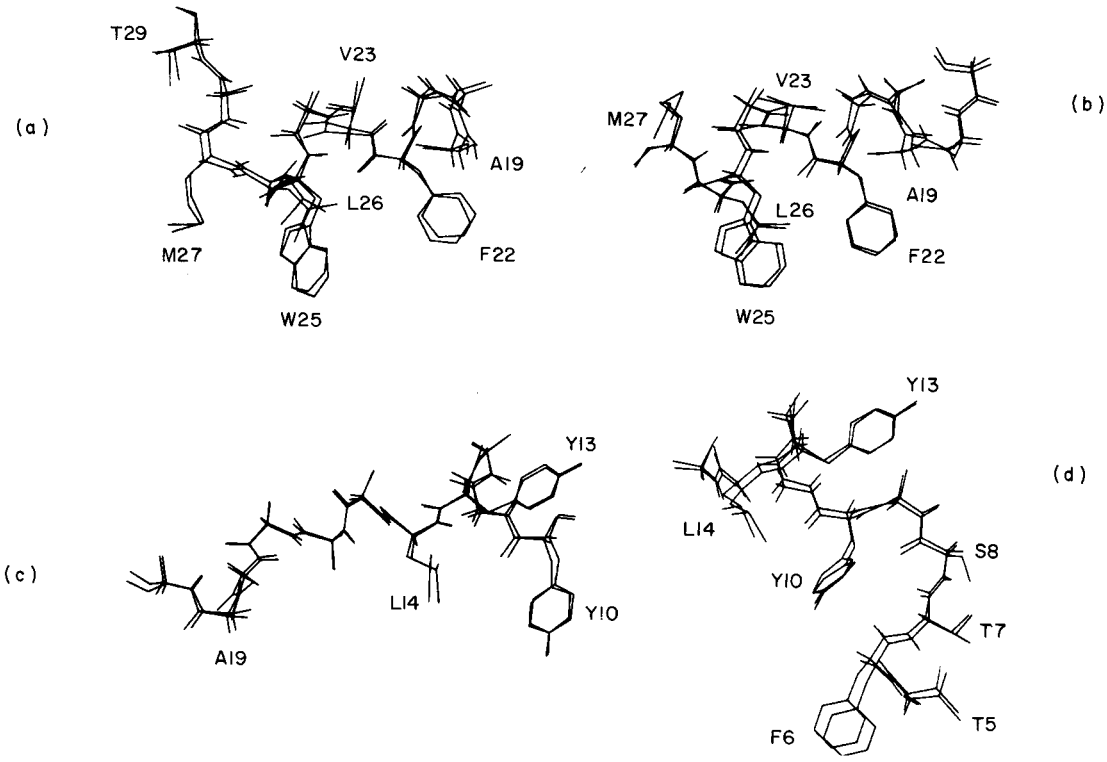

FIG. 7. Structural changes obtained when the stereochemically best structures of the 4 segments of MB-glucagon shown in Fig. 5 (see the text) were subjected to the process of "regularization" to get energetically acceptable peptide bonds (see the text). After the regularization the dihedral angles ω for all the peptide bonds deviate at most by 15° from planarity. The root-mean-square distances between the 2 structures before and after regularization are (d) 0·5 Å for segment 5–15, (c) 0·3 Å for segment 10–20, (b) 0·5 Å for segment 17–27 and (a) 0·6 Å for segment 19–29. The SR presentation of the molecular structures with pseudo-atoms K, L and M described in Fig. 4 is used.

distributions of ϕ–ψ values in globular proteins, MB-glucagon contains relatively many data points near $\phi = +60°$. The ϕ–ψ values for the residues 18 to 29 emphasize that the helical structure formed by this region of the polypeptide chain is pronouncedly irregular.

Using the criteria that the carbonyl oxygen-to-amide proton distance in a hydrogen bond should be between 1·8 and 2·5 Å, and the distance from the carbonyl oxygen to the peptide nitrogen between 2·8 and 3·5 Å, evidence was obtained that C=O . . . H—N hydrogen bonds could be formed between residues 11 and 13, 11 and 14, 17 and 19, 17 and 20, 22 and 25, 23 and 26, 23 and 27, 25 and 27. The irregular helix from residues 17 to 29 thus contains relatively few intramolecular hydrogen bonds. Since the polypeptide is located near the dodecylphosphocholine head groups in the micelles, it is tempting to speculate that the lipid might compete with the polar groups of the polypeptide for the proton acceptor and donor sites and that many of the backbone groups of MB-glucagon are actually hydrogen bonded with lipid head groups.

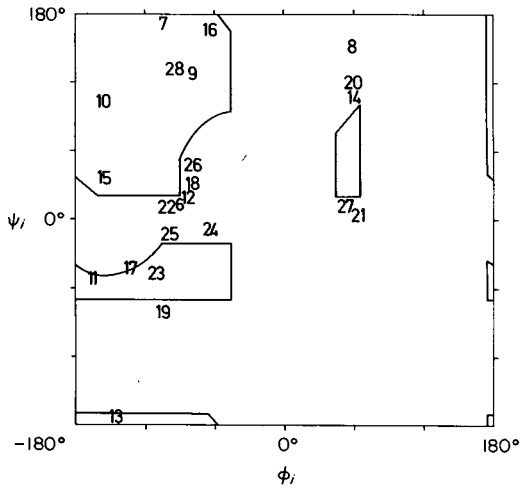

FIG. 8. Plot of the backbone torsion angles ϕ and ψ (Ramachandran & Sasisekharan, 1968) for the regularized (see the text) structure of MB-glucagon in Fig. 7. The data for residues 5 to 12, 13 to 18, 19 to 24 and 25 to 29 are, respectively, from the calculations for the segments 5–15, 10–20, 17–27 and 19–29.

6. Discussion

(a) *Limiting factors in the n.m.r. structure determination of MB-glucagon*

When the present study of MB-glucagon is compared with conceptually similar studies reported previously for MB-glucagon (Braun *et al.*, 1981) and micelle-bound melittin (Brown *et al.*, 1982) it is clear that a much more thorough and extensive structure determination was possible because of the availability of nearly complete resonance assignments to specific residues in the primary structure. However, the text of sections 2 to 5 should also have shown that there is room for further improvements in the future. Thus, while the present structure determination relied on the availability of a large number of relatively inaccurate ^1H–^1H distance constraints, improved accuracy of the measurements of the individual distances will be of interest, e.g. by improvement of the signal-to-noise ratio in NOESY spectra recorded with short mixing times (Fig. 1(a)) and replacement of the uniform averaging model (Braun *et al.*, 1981) with a more realistic treatment for NOE distance constraints in flexible structures. A more quantitative treatment of the NOE distance measurements would probably also have to include a more thorough investigation of the possibility that different parts of the molecular structure might have different effective motional correlation times.

When compared with similar studies of small globular proteins in aqueous solution (unpublished results) certain limitations of the n.m.r. structure determination resulted from the inherent complexity of the MB-glucagon system. For example, the structure determination had to rely entirely on the NOE distance constraints in Table 2 because it was for practical reasons not possible to

obtain complementary information either from spin–spin coupling constants or from studies of the amide proton exchange with the solvent (Wüthrich, 1976) and exploitation of the chemical shifts was limited because of the scarcity of reference data on polypeptide ^{1}H n.m.r. shifts in a lipid–water interphase. The polypeptide backbone is more precisely constrained than most of the side-chains (Table 3), since in most regions there is a dense network of distance constraints between backbone hydrogen atoms (Table 2) and none of the backbone–backbone distance constraints had to be referred to pseudo-atoms (Wüthrich *et al.*, 1983). It is also worth noting that the determination of regular secondary structure elements based on recognition of characteristic NOESY cross peak patterns (Fig. 3) coincides well with the structure obtained from the distance geometry calculations. On the other hand, because of the absence of a globular structure NOE distance constraints between side-chain hydrogens are relatively scarce.

(b) *Comparison of glucagon conformations in different environments*

Three structures are available for comparative studies, i.e. the glucagon trimers studied in single crystals (Sasaki *et al.*, 1975), a conformation of monomeric glucagon in aqueous solution (Bösch *et al.*, 1978) and the present structure of MB-glucagon, where the molecule is in a lipid–water interphase (Brown *et al.*, 1981). Figure 9 shows a comparison of the polypeptide backbone structures of the three segments 5–15, 10–20 and 19–29 of micelle-bound glucagon with the glucagon single crystal X-ray structure (Sasaki *et al.*, 1975). The backbone of each

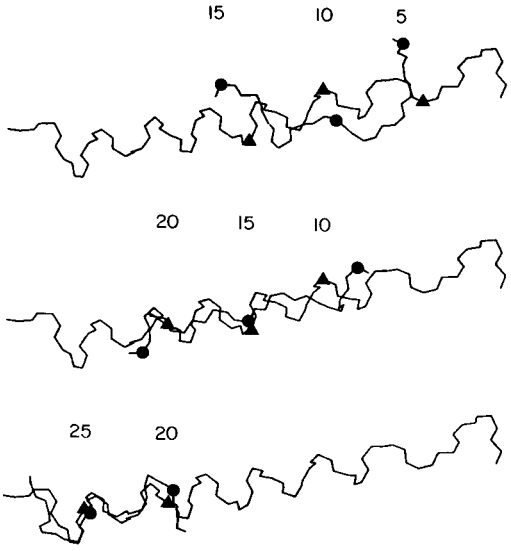

FIG. 9. Comparison of the polypeptide backbone of the structures for the 4 segments of MB-glucagon shown in Fig. 5 with the X-ray structure of glucagon in single crystals (Sasaki *et al.*, 1975). The symbols (●) for MB-glucagon and (▲) for the crystal structure of glucagon mark the C^{α} positions indicated by the numbers. Root-mean-square distances are 3·5 Å for segment 5–15, 3·0 Å for segment 10–20, 2·1 Å for segment 17–27 and 1·6 Å for segment 19–29.

segment was individually translated and rotated to give a best fit with the X-ray structure. Corresponding C^{α} positions are marked with the symbol (●) for MB-glucagon and with the symbol (▲) for the X-ray structure. There is obviously a trend to better coincidence towards the C-terminal region. The length of the segments 5–10 and 10–20 and the overall length of the sum of the different segments of MB-glucagon exceed the length of the helical crystal structure.

In monomeric glucagon in aqueous solution the residues 22 to 25 form a stable, non-random conformation, which differs from a helical structure by major rearrangements in the backbone (Bösch *et al.*, 1978). For the other regions of the glucagon polypeptide chain the data would be compatible with an extended, flexible conformation.

Earlier studies (Bösch *et al.*, 1980) showed that the conformation of glucagon bound to DPC micelles is representative for glucagon bound to various different micellar lipids and appears to be similar to the structure of glucagon bound to lipid bilayers. It may also be added that a previously published "low resolution" n.m.r. structure for the segment 19–27 of MB-glucagon (Braun *et al.*, 1981) indicated the formation of a hydrophobic cluster with the side-chains of Ala19, Phe22, Val23, Trp25 and Leu26 and that this structural feature is confirmed by the present study.

(c) *Glucagon conformation in a lipid–water interphase and biological function*

The following considerations rely on the hypothesis that glucagon adopts similar conformations in the lipid–water interphase near the surface of dodecylphosphocholine micelles and on the surface of the target organ. Fundamental considerations make it appear rather unlikely that the initial contact of glucagon with its target cell would be the formation of the specific complex with the receptor site. Rather, one could expect non-specific binding to the cell surface, followed by diffusion to the specific receptor site in the two-dimensional space provided by the lipid–water interphase on the surface of the cell membrane. With regard to this hypothetical scheme for the initial phase of the glucagon action on the surface of the target organ, two features in the above comparison of the glucagon conformations in different environments are particularly intriguing. First, a major structural rearrangement in the polypeptide region 22–24, which is primarily responsible for the binding of the hormone to the receptor site (Wright & Rodbell, 1979), is implicated for the transfer from dilute aqueous solution to the lipid–water interphase of the micellar surface. If specific binding of glucagon is determined by the conformation of the binding region, the structural features responsible for this specificity would thus be formed only after the hormone has been incorporated into the lipid–water interphase of the cell surface. Second, with regard to the observation that the activity of glucagon depends on the presence of the N-terminal histidine residue while the region near residues 20 to 26 is responsible for specific binding of the hormone to the receptor site (Rodbell *et al.*, 1971; Wright *et al.*, 1978; Wright & Rodbell, 1979), it is interesting that the polypeptide chain between these two locations is more extended in the MB-glucagon conformation than in the crystal structure (Fig. 9).

Use of the facilities at the Zentrum für Interaktives Rechnen (ZIR) of the ETH Zürich is gratefully acknowledged. Financial support was obtained from the Schweizerischer Nationalfonds (project 3.528.79) and through a special grant of the ETH Zürich. We thank Dr T. Blundell and Mrs J. Richardson for helpful discussions and correspondence on this work, and Mrs E. Huber and Mrs E. H. Hunziker for the careful preparation of the manuscript and the illustrations.

REFERENCES

Anil Kumar, Ernst, R. R. & Wüthrich, K. (1980*a*). *Biochem. Biophys. Res. Commun.* **95**, 1–6.

Anil Kumar, Wagner, G., Ernst, R. R. & Wüthrich, K. (1980*b*). *Biochem. Biophys. Res. Commun.* **96**, 1156–1163.

Anil Kumar, Wagner, G., Ernst, R. R. & Wüthrich, K. (1981). *J. Amer. Chem. Soc.* **103**, 3654–3658.

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

Blundell, T. L. & Wood, S. (1982). *Annu. Rev. Biochem.* **51**, 123–154.

Bösch, C., Bundi, A., Oppliger, M. & Wüthrich, K. (1978). *Eur. J. Biochem.* **91**, 204–214.

Bösch, C., Brown, L. R. & Wüthrich, K. (1980). *Biochim. Biophys. Acta,* **603**, 298–312.

Bösch, C., Anil Kumar, Baumann, R., Ernst, R. R. & Wüthrich, K. (1981). *J. Magn. Reson.* **42**, 159–163.

Bothner-By, A. A. & Noggle, J. H. (1979). *J. Amer. Chem. Soc.* **101**, 5152–5153.

Braun, W., Bösch, C., Brown, L. R., Gō, N. & Wüthrich, K. (1981). *Biochim. Biophys. Acta,* **667**, 377–396.

Brown, L. R. (1979). *Biochim. Biophys. Acta,* **557**, 135–148.

Brown, L. R., Bösch, C. & Wüthrich, K. (1981). *Biochim. Biophys. Acta,* **642**, 296–312.

Brown, L. R., Braun, W., Anil Kumar & Wüthrich, K. (1982). *Biophys. J.* **37**, 319–328.

Crippen, G. M. & Havel, T. F. (1978). *Acta Crystallogr.* **34**, 282–284.

Dobson, C. M., Olejniczak, E. T., Poulsen, F. M. & Ratcliffe, R. G. (1982). *J. Magn. Reson.* **48**, 97–110.

Epand, R. M. (1971). *Can. J. Biochem.* **49**, 166–169.

Epand, R. M., Jones, A. J. S. & Schreier, S. (1977). *Biochim. Biophys. Acta.* **491**, 296–304.

Gordon, S. L. & Wüthrich, K. (1978). *J. Amer. Chem., Soc.* **100**, 7094–7096.

Gratzer, W. B., Bailey, E. & Beaver, G. H. (1967). *Biochem. Biophys. Res. Commun.* **28**, 914–919.

Gurd, F. R. N. & Rothgeb, T. M. (1979). *Advan. Protein Chem.* **33**, 74–165.

Havel, T. F., Crippen, G. M. & Kuntz, I. D. (1979). *Biopolymers,* **18**, 73–81.

Hull, W. A. & Sykes, B. D. (1975). *J. Chem. Phys.* **65**, 867–880.

Jeener, J., Meier, B. H., Bachman, P. & Ernst, R. R. (1979). *J. Chem. Phys.* **71**, 4546–4553.

Kalk, A. & Berendsen, H. J. C. (1976). *J. Magn. Reson.* **24**, 343–366.

Karplus, M. & McCammon, J. A. (1981). *Crit. Rev. Biochem.* **9**, 293–349.

Macura, S., Huang, Y., Suter, D. & Ernst, R. R. (1981). *J. Magn. Reson.* **43**, 259–281.

McCammon, J. A. & Karplus, M. (1980). *Annu. Rev. Phys. Chem.* **31**, 29–45.

McLachlan, A. D. (1979). *J. Mol. Biol.* **128**, 49–79.

Momany, F. A., McGuire, R. F., Burgess; A. W. & Scheraga, H. A. (1975). *J. Phys. Chem.* **79**, 2361–2381.

Moran, E. C., Chou, P. Y. & Fassman, G. D. (1977). *Biochem. Biophys. Res. Commun.* **77**, 1300–1306.

Nagayama, K., Wüthrich, K. & Ernst, R. R. (1979). *Biochem. Biophys. Res. Commun.* **90**, 305–311.

Nagayama, K., Anil Kumar, Wüthrich, K. & Ernst, R. R. (1980). *J. Magn. Reson.* **40**, 321–334.

Noggle, J. H. & Schirmer, R. E. (1971). *The Nuclear Overhauser Effect,* Academic Press, New York.

Panijpan, B. & Gratzer, W. B. (1974). *Eur. J. Biochem.* **45**, 547–553.

Pohl, S. L., Birnbaumer, L. & Rodbell, M. (1969). *Science,* **164**, 566–569.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Advan. Protein Chem.* **25**, 283–437.
Richardson, J. (1981). *Advan. Protein Chem.* **34**, 167–339.
Rodbell, M., Birnbaumer, L., Pohl, S. L. & Sundby, F. (1971). *Proc. Nat. Acad. Sci., U.S.A.* **68**, 909–913.
Rubalcava, B. & Rodbell, M. (1973). *J. Biol. Chem.* **248**, 3831–3837.
Sasaki, K., Dockevill, S., Ackmiak, D. A., Tickle, I. J. & Blundell, T. L. (1975). *Nature (London)*, **257**, 751–757.
Schneider, A. B. & Edelhoch, H. (1972). *J. Biol. Chem.* **247**, 4986–4991.
Srere, P. A. & Brooks, G. C. (1969). *Arch. Biochem. Biophys.* **129**, 708–710.
Wagman, M. E., Dobson, C. M. & Karplus, M. (1980). *FEBS Letters,* **119**, 256–270.
Wagner, G. & Wüthrich, K. (1979). *J. Magn. Reson.* **33**, 675–680.
Wako, H. & Scheraga, H. A. (1982). *J. Protein Chem.* **1**, 5–45.
Wider, G., Lee, K. H. & Wüthrich, K. (1982). *J. Mol. Biol.* **155**, 367–388.
Wider, G., Hosur, R. V. & Wüthrich, K. (1983). *J. Magn. Reson.* **52**, 130–135.
Wright, D. E. & Rodbell, M. (1979). *J. Biol. Chem.* **254**, 268–269.
Wright, D. E., Hruby, V. J. & Rodbell, M. (1978). *J. Biol. Chem.* **253**, 6338–6340.
Wüthrich, K. (1976). *NMR in Biological Research: Peptides and Proteins,* North-Holland Publishing Company, Amsterdam.
Wüthrich, K. & Wagner, G. (1978). *TIBS,* **3**, 227–230.
Wüthrich, K., Bösch, C. & Brown, L. R. (1980). *Biochem. Biophys. Res. Commun.* **95**, 1504–1509.
Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). *J. Mol. Biol.* **155**, 311–319.
Wüthrich, K., Billeter, M. & Braun, W. (1983). *J. Mol. Biol.* **169**, 949–961.

Edited by J. C. Kendrew