tively, could be established. It is clear, however, that the dichroic spectrum of ELG_{12} is characteristic of the "cross-β" conformation.

Conclusions

The experiments described here demonstrate that both oligopeptides and polypeptides are partially oriented upon incorporation in stretched films of polyoxyethylene. Furthermore, the infrared characteristics of the POE support allow determination of dichroic spectra for the $\nu(NH)$, amide I, and II transitions as well as for infrared bands below ~ 800 cm^{-1}. Since POE films rapidly absorb substantial amounts of water vapor, hydrogen to deuterium exchange reactions of suspended molecules can readily be studied. Important new conformational information is available from the combination of infrared dichroism and hydrogen exchange kinetics.

Incorporation of peptides has no measurable effect on the infrared transitions of the POE matrix. Apparently the peptides are confined to the amorphous phase of the POE film and do not alter its crystalline structure. Alignment of the long axes of suspended molecules parallel to the stretching directions appears to be the preferential mode or orientation of the peptides employed here. Only partial orientation is achieved; values for the minimum orientation factor f_m ranged from 0.4 for gramicidin S to 0.6 for PBLG. In spite of imperfect orientation, useful conformational information could usually be derived from the dichroic spectra.

Results presented here are consistent with the Hodgkin-Oughton–Schwyzer[23,24] conformation of gramicidin S. We speculate that its "cross-β" type dichroic spectrum results from extensive aggregation to form the ribbon-like β structures shown in Figure 6. Dichroic spectra of synthetic oligo- and polypeptides reveal that both α-helical and β conformations are oriented in POE films. The latter can be oriented either with the peptide chains parallel or perpendicular to the stretching direction of the POE film.

References and Notes

(1) Postdoctoral Fellow 1971; on leave from the Hebrew University, Jerusalem, Israel.
(2) (a) E. W. Thulstrup, J. Michl, and J. H. Eggers, *J. Phys. Chem.*, **74**, 3868 (1970); (b) A. Yogev, L. Margulies, D. Amar, and Y. Mazur, *J. Am. Chem. Soc.*, **91**, 4558 (1969).
(3) R. Schwyzer and P. Sieber, *Helv. Chim. Acta*, **40**, 625 (1957).
(4) M. Goodman and J. Hutchison, *J. Am. Chem. Soc.*, **88**, 3677 (1966).
(5) M. Goodman and I. G. Rosen, *Biopolymers*, **2**, 503 (1964).
(6) F. E. M. O'Brien, *J. Sci. Instrum.*, **25**, 73 (1948).
(7) Y. A. Ovchinnikov, V. T. Ivanov, V. F. Bystrov, A. I. Miroshnikov, E. N. Shepel, N. D. Abdullaev, E. S. Efremov, and L. B. Senyavina, *Biochem. Biophys. Res. Commun.*, **39**, 217 (1970).
(8) E. M. Bradbury, L. Brown, A. R. Downie, A. Elliott, R. D. B. Fraser, W. E. Hanby, and T. R. R. McDonald, *J. Mol. Biol.*, **2**, 276 (1960).
(9) A. J. Geddes, K. D. Parker, E. D. T. Atkins, and E. Beighton, *J. Mol. Biol.*, **32**, 343 (1968).
(10) R. L. Davidson and M. Sittig, Ed., "Water-Soluble Resins", Reinhold Boak Corporation, New York, N.Y., 1968.
(11) A. Stern, W. A. Gibbons, and L. C. Craig, *Proc. Natl. Acad. Sci. U.S.*, **61**, 734 (1968).
(12) D. W. Urry and M. Ohnishi in "Spectroscopic Approaches to Biomolecular Conformation", D. W. Urry, Ed., American Medical Association, Chicago, Ill., 1970.
(13) Y. Takahashi and H. Tadokoro, *Macromolecules*, **6**, 672 (1973).
(14) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969, p 165.
(15) Y. Takahashi, I. Sumita, and H. Tadokoro, *J. Polym. Sci., Polym. Phys. Ed.*, **2**, 2113 (1973).
(16) H. Tadokoro, T. Yoshihara, Y. Chatani, and S. Murahashi, *Polym. Lett.*, **2**, 363 (1964).
(17) H. Tadokoro, *Macromol. Rev.*, **1**, 119 (1966).
(18) R. Iwamoto, Y. Saito, H. Ishihara, and H. Tadokoro, *J. Polym. Sci., Part A-2*, **6**, 1509 (1968).
(19) H. Tadokoro, Y. Chatani, T. Yoshihara, S. Tahara, and S. Murahashi, *Makromol. Chem.*, **73**, 109 (1964).
(20) M. Beer, *Proc. R. Soc. London, Ser. A*, **236**, 136 (1956).
(21) R. D. B. Fraser, *J. Chem. Phys.*, **21**, 1511 (1953).
(22) E. M. Bradbury and A. Elliott, *Spectrochim. Acta*, **19**, 995 (1963).
(23) D. C. Hodgkin and B. M. Oughton, *Biochem. J.*, **65**, 752 (1957).
(24) R. Schwyzer and V. Ludescher, *Biochemistry*, **7**, 2519 (1968).
(25) G. M. J. Schmidt, D. C. Hodgkin, and B. M. Oughton, *Biochem. J.*, **65**, 744 (1957).
(26) D. Balasubramanian, *J. Am. Chem. Soc.*, **89**, 5445 (1967).
(27) T. Miyazawa and E. R. Blout, *J. Am. Chem. Soc.*, **83**, 712 (1961).
(28) M. Tsuboi, *J. Polym. Sci.*, **59**, 139 (1962).
(29) M. Goodman, A. S. Verdini, C. Toniolo, W. D. Phillips, and F. A. Bovey, *Proc. Natl. Acad. Sci. U.S.*, **64**, 444 (1969).

An Approach to the Tertiary Structure of Globular Proteins

I. D. Kuntz

Contribution from the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143. Received February 26, 1975

Abstract: The "distance plots" described by Rossman and Liljas are graphs of α-carbon distances derived from the known X-ray structures of globular proteins. These plots have been shown to be useful in locating "folding domains". The surprisingly regular patterns contain much additional information about protein tertiary structure. In this paper we show that repeating square and trapezoidal patterns can correspond to distorted three-dimensional "superhelical" structures that are principal constituents of folding domains.

Characterization of tertiary structure in globular proteins is not well developed. Kauzmann's suggestion that proteins contain a hydrophobic core remains the only major generalization.[1] With the availability of X-ray structural data for many globular proteins there have been some initial efforts to locate specific features such as calcium binding sites,[2] dinucleotide binding sites,[3] and hydrophobic regions.[4] A more general approach that has considerable promise has been the identification of "folding domains"

within the larger proteins.[5-7] One of the methods used for the location of domains employs "distance plots", described first by Phillips[8] and used in modified form by Rossman and Liljas.[6,7] "Distance plots" are graphs of C_α-C_α distances plotted against residue number, with contour lines drawn at fixed interatomic distances (see Figure 1 and Figures 1 and 2 of ref 7). These plots provide a concise summary of the structural information available from X-ray diffraction experiments. Rossman and Liljas have shown that
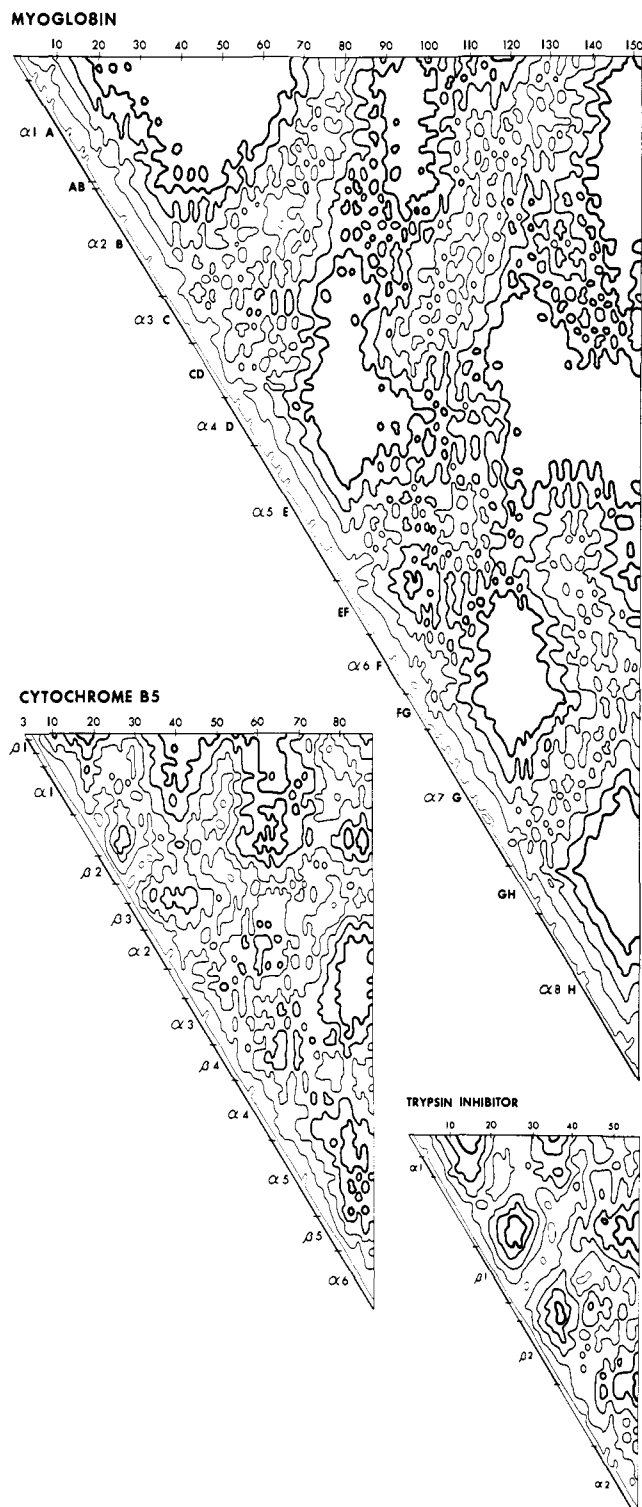
MYOGLOBIN

CYTOCHROME B5

TRYPSIN INHIBITOR

Figure 1. Representative distance plots. The data for pancreatic trypsin inhibitor (Huber and coworkers),[17] cytochrome b5 (Mathews and coworkers),[16] and for sperm whale myoglobin (Watson and J. C. Kendrew)[15] were made available by the Protein Data Bank, Brookhaven National Laboratory. Contours are drawn at 0, 15, 10, 15, 20, and 25 Å. The numerical results from which these diagrams were made were obtained upon a lineprinter and all slopes are too large by a factor of 1.6. This error does not alter the basic patterns of interest here. Conventional secondary structure is shown as assigned by the original workers.

such plots permit direct visualization of folding domains. We hope to show, here, that the distance plots also provide a very convenient and systematic approach to the study of other aspects of tertiary structure and permit recognition of some novel features of polypeptide chain folding.

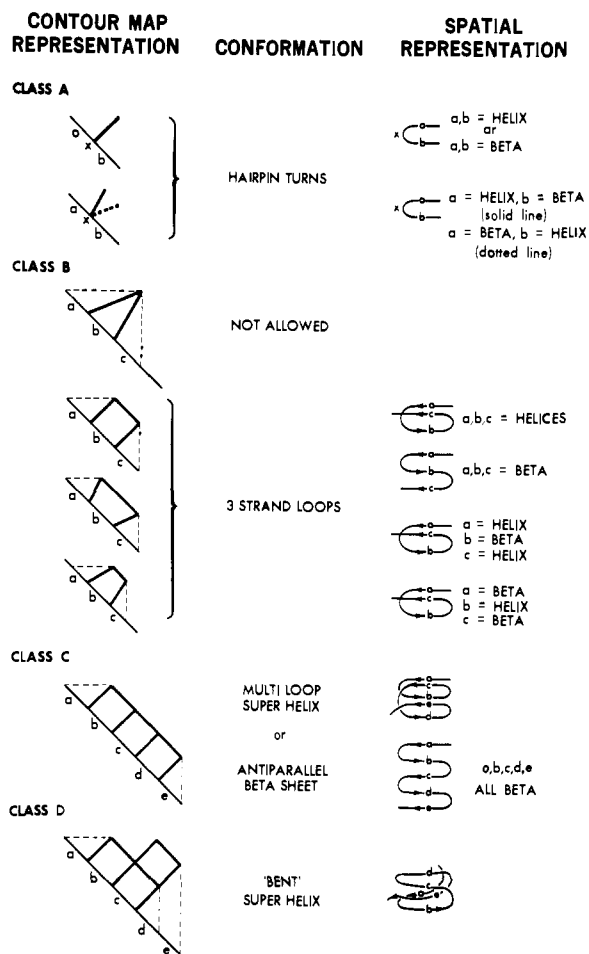CONTOUR MAP REPRESENTATION   CONFORMATION   SPATIAL REPRESENTATION

Figure 2. Relationships between idealized patterns on the distance plots and three-dimensional polypeptide conformations. The weak dotted lines (class B–D) are used to indicate the boundaries of the outer segments.

We begin by noting that the distance plots provide a "natural" organization of structural information. Secondary structure, that is, helical or extended chain conformations, involves residues close together along the peptide chain and is identified on the distance plots by inspection of the contours near the diagonal. Helical regions, for example, are portrayed by a low density of contour lines very near to, and parallel with, the diagonal, while β strands have a high density of contours along the diagonal. Local tertiary structures are explored by moving away from the diagonal. One first encounters representations of turns of various types. Further out from the diagonal, one finds patterns representing loops and other closed structures to be discussed below. On still larger excursions from the diagonal, there is information about the way in which these compact units interact with each other to form "domains". Finally, domain–domain interactions are depicted. Thus the distance plots provide an inherent structural hierarchy that allows us to focus our attention on different levels of structural complexity.

Our main concern in this paper is a qualitative discussion of the surprisingly regular features seen in the distance maps of Rossman and Liljas[7] and Figure 1. We shall proceed by considering various idealized patterns on the distance maps and asking to what backbone spatial conformations these patterns correspond. We make use of the structural hierarchy just described to discuss four classes of features (Figure 2). Class A patterns consist of simple line elements that stand for collections of contour lines that intersect the diagonal. Features of this type represent adjacent

segments[9] (a,b) of residues that come close together in space in an antiparallel manner. The common "β" turn linking two β strands is the most obvious example of this type. "Hairpin" turns can also link other secondary structural elements such as helix to helix or helix to β strand.[10,11] On the distance plot any of these combinations will be seen as a set of contours that we represent in idealized fashion as a single straight line that intersects the diagonal. The slope of this line will depend on the secondary structure because the slope is determined by the ratio of the number of residues per unit length for the segments a and b. This ratio is 2 if segment a is helical and segment b is a β strand. A β-helix arrangement would give a slope of 0.5. If both segments are of the same secondary structure, the slope is unity. These values represent limits for proteins since polypeptide structures more compact than helices or more extended than β strands are not normally expected.[12] Off-diagonal lines on the distance maps can have a wide range of slopes. Slopes of -0.5 to -2 correspond to parallel arrangements of the usual secondary structures. Very positive or negative slopes for contour lines off the diagonal represent segments that cross each other. Negative slopes are not possible at or very near the diagonal because two segments cannot be directly linked to each other so that the sequence proceeds in a parallel fashion while the residues of each segment maintain close contact.

We next consider idealized representations of distance plot contours that are combinations of the line elements forming simple closed figures such as triangles, squares, or trapezoids. These patterns come about when three segments (a,b,c) separated by two hairpin turns, are arranged in space so that each segment is close to the other two. Curiously, a triangular pattern of contours at the diagonal cannot occur for proteins. This comes about because the limited range of permissible slopes (0.5 to 2) generates a pattern (Figure 2) that would require that the three pairs of segments a,b and b,c and a,c be close to each other in space with antiparallel sequences. Three segments cannot be all antiparallel to each other. Hence, triangular patterns of this type should not occur on protein distance plots.

Four-sided patterns such as squares or trapezoids do not, generally, deal with the interactions between *four* segments. Rather, they also represent the contacts among *three* segments that come close together in space. Such a pattern requires that segments a,b be antiparallel to each other as are segments b,c. However, segments a,c are now parallel to each other. Two related idealized structures can meet these requirements: (1) three segments arranged in a plane as in an antiparallel β sheet,[13] a structure we will call "S", and (2) a three-strand loop roughly assembling a figure "6". This latter structure can be thought of as a distorted "β sheet" in which the last strand is rotated significantly out of the plane of the first two strands. With sufficient out-of-plane rotation, this structure would permit the side chains of segments a and c to be in contact.

Trapezoidal forms on distance plots are found whenever the hairpin turns link helix to extended segments or vice versa. We have pointed out that such linkages will generate initial slopes that differ from unity and, hence, that will give rise to the trapezoidal patterns. Square and trapezoidal patterns represent the simplest and smallest units of tertiary structure that can form *parallel* spatial arrangements of the polypeptide chain. They also are indicative of quite compact packing of the chain. We shall call such patterns "contour cells". We have not considered higher level polygons of contour lines (pentagons, hexagons, etc.) although some of the patterns on protein maps might best be represented in this way.

We next consider patterns that are combinations of contour cells. First let us look at contour cells that are adjacent to each other and each lying along the diagonal (class C, Figure 2). Diagonal contour cell patterns represent extensions of the structures just described for single cells. One such structure is a multistranded antiparallel β sheet. If there are m cells in the pattern, there will be $m + 2$ strands of antiparallel β conformation in the structure. The other limiting structure represented by a class C pattern is a repeating "three-strand loop". This structure has a quasi-helical shape with $(m + 2)/3$ helical turns per pattern of m cells. One isolated contour cell would then represent a single superhelical turn. We use the name "superhelix" for such structures in analogy to the well-characterized superhelical DNA.[14] Note that there appear to be no restrictions on the types of secondary structures that can contribute to a "superhelix"; thus, both helical and extended conformations could be included. Formation of a β sheet within the superhelix would tend to flatten it, and a number of adjacent β strands would yield the conventional twisted pleated sheet.

Patterns in which some of the connected cells lie off the diagonal are grouped together in class D. Stacking of the cells on top of each other requires that some contacts be made between the polypeptide chain at both ends of a super helix. These patterns represent complicated three-dimensional structures that can be approximately characterized as multiconnected loops or "bent" superhelices. Further interpretation would require careful inspection of molecular models and a consideration of the actual size of the loops. We shall not attempt such a treatment here. Nor shall we discuss features of the distance plots that are more than two cells off the diagonal, even though such patterns contain information on the packing together of domains and are potentially of much interest for an understanding of the complete tertiary structure.

We have implicitly assumed throughout this discussion that all segments are linear. In fact, as noted earlier, nonlinear segments are often observed. Nonlinear segments make detailed analysis of structure more difficult since they permit considerable distortion in the three-dimensional conformations represented by the contour cells. Such effects will be pointed out in the next section.

To this point we have been concerned, primarily, with idealized features of the distance maps. We must now ask what features are actually found on the maps prepared from the X-ray data for some globular proteins. Inspection of maps for several proteins confirms the detailed analysis of the slopes of the contour lines for β-β, helix-helix, and β-helix linkages. One can also readily pick out secondary structure and many hairpin turns. As suggested above, only about one-half of the hairpin turns reported in earlier studies[10,11] produces major intersecting contours on distance plots. This comes about for two reasons. First, some of the hairpin turns overlap to form a composite turn. Second, some of the turns, while meeting the various criteria in the literature, are not sharp enough to bring many residues in adjacent segments into close contact.

A closer look at the distance plots for myoglobin and cytochrome b$_5$ permits an evaluation of the utility of idealized patterns. Myoglobin contains eight α helices linked together by "corners" of varying complexity.[15] The folds in the chain occur after each helix at or near residues 18–20, 35–36, 43–50, 57–58, 77–85, 95–99, and 119–124. The molecule contains no β structure. The distance map indicates major, high-angle folds occur at residues 40–50 (this loop includes the junction of the B and C helices, the C helix itself, and the CD corner), as well as at the EF, FG, and GH corners (residues 77–85, 95–99, 119–124, respectively). The junction of the A and B helices and the junction of the D and E

helices are not major features on the contour map because they generate only "weak" turns ca. 90°. The five segments of the polypeptide chain, as marked out by the "strong" turns, are shown in Table I. The first two segments are quite nonlinear, while the last three are each composed of a single helix and an adjacent corner and are thus approximately linear. There are three complete contour cells and parts of two others lying along the diagonal (Table I and Figure 1). Generally, we find that the peptide chains that go with each contour cell have a more complicated folding pattern than the idealized shapes we have discussed. This is primarily due to the nonlinear segments. Thus, for example, the peptide chain for cell A (an incomplete cell that spans helices A-E) is a bent hairpin shape and cell B (helices C-F) is also quite distorted. Cells C and D, however, are composed of nearly linear segments and the corresponding peptide chains are, in fact, in an approximate right-handed superhelical conformation. The last contour cell is incomplete but serves to extend the superhelix by one third of a turn. Thus, in very rough terms, we can represent the myoglobin molecule as being folded in two overlapping parts. The first part, resides 1-77, has a compact but irregular tertiary structure. The second section, residues 58-153, forms about one and one-half turns of a superhelix. The two sections are joined by the common residues 58-77 (D and E helices), and their mutual interactions are represented on the distance map by the collection of close contacts between residues 5-58 and 99-120 (the large off-diagonal set of contours in the upper right part of the myoglobin map).

The X-ray data for cytochrome b5 are incomplete in that several terminal residues have not been located. The protein contains an assortment of secondary features including six helices and five short pieces of β structure.[16] The distance map for cytochrome b5 is less well marked than that for myoglobin. This seems to be part of a general tendency for larger proteins to produce more distinct patterns on distance maps than do smaller proteins, perhaps because of the longer linear segments in the former molecules. There are five segments in cytochrome b5 and four reasonably distinct diagonal contour cells (Table I), with two partial cells also present. Some of the segments are nonlinear, giving rise to distorted cells, two of which actually incorporate other cells. The first cell is a partial one that represents a closed loop spatial conformation that is almost a single turn of superhelix. The second cell incorporates two strands of β structure and the corresponding peptide chain is a nearly planar "S" shape. The third and fourth cells include nonlinear segments that lead to looped structures, but of complex shape. The last two cells (one complete and one partial) again represent a roughly superhelical chain conformation. It is interesting that both myoglobin and cytochrome b5 have adopted a similar tertiary feature to form most of the heme pocket.

In summary, we agree with Rossman and Liljas that the distance plots are a very useful tool for locating folding domains. Further, the distance plots suggest a general method for representing many levels of tertiary structure in proteins. Our preliminary analysis of the experimental maps leads to an ordering of tertiary structural features. The lowest level of folding is, of course, the "hairpin" turn. Such turns bring sets of residues that we have termed "segments" into close contact. These segments typically contain 10-15 residues and can be made of helical, β, or other secondary conformations. The next level of tertiary structure deals with the arrangement of three segments into compact units such as β sheet or superhelical loops. Distance maps provide a straightforward way to recognize such units because they generate the square or trapezoidal patterns we have called contour cells. Such cells represent, typically, 30-50 residues

Table I. Tertiary Features in Myoglobin and Cytochrome b₅

A. Myoglobin

Segment	Residues	Secondary structure	Description
I	1–35	α1, AB corner, α2	Nonlinear
II	36–77	α3, CD corner, α4, α5	Nonlinear
III	78–99	EF corner, α6	Linear
IV	100–119	FG corner, α7	Linear
V	120–153	GH corner, α8	Linear

Contour cells	Residues	Segments	Description
A	1–77	I, II	Partial cell, bent hairpin
B	36–99	II, III	Distorted
C	58–120	½ II, III, IV	3-Strand loop ⎫ right-hand
D	78–153	III, IV, V	3-Strand loop ⎬ superhelix
E	100–153	IV, V	Partial cell ⎭

B. Cytochrome b₅

Segment	Residues	Secondary structure	Description
I	3–16	β1, α1	Nonlinear
II	17–25	β2	Linear
III	26–39	β3, α2	Nonlinear
IV	40–49	α3	Linear
V	50–63	β4, α4	Nonlinear
VI	64–87	α5, β5, α6	Nonlinear

Contour cells	Residues	Segments	Description
A	3–25	I, II	Partial cell, loop
B	7–39	½ I, II, III	"S"
C	17–54	II, III, IV, ½ VI	Distorted
D	26–73	III, IV, V, ½ VI	Distorted
E	40–80	IV, V, ½ VI	3-Strand loop ⎱ right-hand
F	50–87	V, VI	Partial cell ⎰ superhelix

and have spatial dimensions of 10-20 Å. A concatenation of loops results in a few turns of superhelical structure that is often found folded back on itself. Since the superhelical dimensions are 10-20 Å across and require 30-50 residues per turn, it is unlikely that more than a few turns will be present per helix. Most of the folding domains we have examined appear to consist, at least in part, of this "superhelical" structure.

References and Notes

(1) W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959).
(2) R. M. Tufty and R. H. Kretsinger, *Science*, **187**, 167 (1975).
(3) S. T. Rao and M. G. Rossman, *J. Mol. Biol.*, **76**, 241 (1973).
(4) I. D. Kuntz, *J. Am. Chem. Soc.*, **94**, 8568 (1972).
(5) D. B. Wetlaufer, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 697 (1973).
(6) A. Liljas and M. G. Rossman, *Annu. Rev. Biochem.*, **43**, 475 (1974).
(7) M. G. Rossman and A. Liljas, *J. Mol. Biol.*, **85**, 177 (1974).
(8) D. C. Phillips in "British Biochemistry, Past and Present", T. W. Goodwin, Ed., Academic Press, London, 1970, pp 11–28.
(9) A *segment* of polypeptide chain is taken as the continuous string of residues that fall between line elements that intersect the diagonal (Figures 1 and 2). As we shall see, segments are often in helical or extended chain conformations. However, many segments include nonlinear spatial conformations, the most common being a weak (θ <90°) turn. These weak turns do not show pronounced features on the distance plots, being represented by small protuberances near the diagonal (see, for example, the junction of the A,B helices on the myoglobin map, Figure 1).
(10) (a) P. N. Lewis, F. A. Momany, and J. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 2293 (1971); (b) I. D. Kuntz, *J. Am. Chem. Soc.*, **94**, 4009 (1972).
(11) J. L. Crawford, W. N. Lipscomb, and C. G. Schlman, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 538 (1973).

(12) R. E. Dickerson and I. Geis, "The Structure and Action of Proteins", Harper and Row, New York, N.Y., 1969.
(13) This possibility arises because the backbone carbon atoms of the first and third strands of β structure actually come quite close together (i.e., 8–10 Å), even though the side-chain atoms of these strands are not in contact.
(14) J. C. Wang, J. Mol. Biol., 43, 263 (1969).
(15) H. C. Watson, Prog. Stereochem., 4, 299 (1968).
(16) F. S. Mathews, P. Argos, and M. Levine, Cold Spring Harbor Symp. Quant. Biol., 36, 387 (1972).
(17) R. Huber, D. Kukla, A. Ruhlmann, and W. Steiglmann, Cold Spring Harbor Symp. Quant. Biol., 36, 141 (1972).

[Sar1-A]Insulin, a Biologically Active Analog[1,2]

Yoshio Okada and Panayotis G. Katsoyannis*

Contribution from the Department of Biochemistry, Mount Sinai School of Medicine of the City University of New York, New York, New York 10029. Received February 19, 1975

Abstract: The synthesis and isolation in purified form of [Sar1-A]insulin, a biologically active analog of sheep insulin, are described. This analog differs from the parent molecule in that the N-terminal amino acid residue of the A chain, glycine, has been replaced with sarcosine. This change which in essence is the replacement of one hydrogen of the amino group of the A^1 residue with a methyl group, with a concomitant increase of the basic character of that residue, results in a small decrease of the biological activity and a more pronounced decrease of the immunoreactivity of insulin (20 and 9 IU/mg, respectively, vs. 23–25 IU/mg for the natural hormone). For the synthesis of this analog, the [Sar1]A chain was synthesized by the fragment condensation method and isolated as the S-sulfonated derivative. Conversion of the latter compound to the sulfhydryl form and combination of the reduced product with the S-sulfonated derivative of the B chain of sheep (bovine) insulin afforded the [Sar1-A]insulin. Isolation and purification of the insulin analog was accomplished by chromatography on a carboxymethylcellulose column with an exponential sodium chloride gradient.

The importance of the amino acid residues located at the amino terminal regions of the A and B chains of insulin to the biological activity has become evident by studies carried out in several laboratories.[3] Our own investigations have shown that removal of the N-terminal tetrapeptide sequence from the A chain results in complete inactivation of the molecule.[4] Of particular interest was our finding that replacement of the α-amino group of the N-terminal glycine by hydrogen results in a substantial decrease of the biological activity of the hormone[5] (7–10 IU/mg vs. 23–25 IU/mg for the natural compound). Similarly, removal of the N terminus of the A chain, glycine, leads to a 90% loss of the biological activity of insulin.[6] On the other hand, elimination of the N-terminal amino acid residue of the B chain does not cause any appreciable change in the biological activity of the hormone.[7] These data therefore clearly demonstrate that the A^1 residue is critically involved in the maintenance of high biological activity of insulin. It appears, however, that the amino group at A^1 is not per se involved with the manifestation of high biological activity of insulin. Indeed, modification of the amino group of glycine A^1 with a variety of acyl groups does not strikingly affect the biological potency and immunoreactivity of the hormone.[3,8,9] The preparation of numerous analogs of insulin with the A^1 amino group modified indicates that the relative size of the modifying acyl group affects the biological profile of the hormone and that large acyl groups decrease the biological activity of insulin.[9] These results led to the speculation[9] that the N-terminal residue of the A chain is involved in the stabilization of the "active site" of insulin and that this site is affected either by conformational changes and/or a steric effect of the large acyl groups at A^1.

Characteristic of almost all insulin analogs involving modifications at A^1 that have been prepared thus far is that they no longer possess a basic group at A^1. It was therefore of interest to investigate the effect on the biological activity and immunogenicity of insulin of modifications at A^1 that increase the basic character of that residue. This investiga-

tion describes the synthesis, isolation in pure form, and biological evaluation of such an analog, [Sar1-A] sheep insulin, which differs from the parent molecule in that the N terminus of the A chain, glycine ($pK_2 = 9.6$), has been replaced with sarcosine ($pK_2 = 10.00$). By the mouse convulsion assay method this analog was found to possess a potency of 20 IU/mg. By the radioimmunoassay method [Sar1-A]insulin possesses a potency of 9 IU/mg (vs. 23–25 IU/mg for the natural hormone). It thus becomes apparent that replacement of one hydrogen of the amino group of the A^1 residue with a methyl group with a concomitant increase of the basic character of that residue results in but a modest decrease of the biological activity and a more pronounced decrease of the immunogenicity of insulin. It is interesting to note that a parallel situation exists with another insulin analog, the [Arg-Gly$^{A^1}$]insulin, where one hydrogen of the amino group at the A^1 residue is replaced with the basic arginine residue.[10] In this case, too, there is a more pronounced decrease of the immunogenicity than of the biological activity of the molecule (60 vs. 40%). The synthesis of several insulin analogs modified at the A^1 residue is, of course, required before a firm conclusion regarding such a relationship between chemical structure, biological activity, and immunogenicity is established.

General Aspects of the Synthesis. [Sar1-A] sheep insulin was prepared by combination of the S-sulfonated form of the B chain of bovine (sheep) insulin with the sulfhydryl form of the [Sar1]A chain of sheep insulin. The S-sulfonated bovine B chain, which is identical with the respective chain of sheep insulin,[11,12] was prepared by oxidative sulfitolysis of bovine insulin, followed by separation of the resulting S-sulfonated derivatives of the A and B chains by continuous flow electrophoresis. The sulfitolysis of insulin and the separation of the resulting A and B chain derivatives have been described in previous communications from this laboratory.[13] The synthesis of the [Sar1]A chain with its functional groups protected (XII) was accomplished by the classical methods of peptide synthesis, namely, a combination of the "stepwise elongation" and "fragment conden-