A Heuristic Approach to Predicting the Tertiary Structure of Bovine Somatotropin‡

Louis Carlacci,§ Kuo-Chen Chou,* and Gerald M. Maggiora
Computational Chemistry, Upjohn Laboratories, Kalamazoo, Michigan 49001
Received October 9, 1990; Revised Manuscript Received February 5, 1991

ABSTRACT: A combination of a heuristic approach and energy minimization was used to predict the three-dimensional structure of bovine somatotropin (bSt), also known as bovine growth hormone, a protein of 191 amino acids. The starting points for energy minimizations were generated from the following two types of inputs: (a) the amino acid sequence and (b) the heuristic inputs, which were derived according to physical, chemical, and biological principles by piecing together all useful information available. The predicted 3-D structure of the bSt molecule has all the features observed in four-helix bundle proteins. The four α-helices in bSt are intimately packed to form an assembly with an approximately square cross section. All the adjacent α-helices are antiparallel, with a somewhat tilted angle between each of the adjacent pairs so that the assembly of the four helices looks like a left-handed twisted bundle. There are two disulfide bonds in the bSt structure: one "hooking" the middle of a long loop with helix 4 so as to pull the long loop onto the surface of the helix bundle and the other "hooking" the C-terminal segment with the same helix so as to force the C-terminal segment to bend toward the helix bundle. As a consequence, a considerable part of the surface of the four-helix bundle is closely packed or intimately embraced by the loop segments. The predicted bSt structure has a hydrophobic core and a hydrophilic exterior surface. The energetic analysis of the predicted bSt structure indicates that the interaction between helices and loops plays a dominant role in stabilizing the four-helix bundle structure from the viewpoint of both electrostatic and nonbonded interactions. A technique called FOLD was meanwhile developed, by which one can fold a polypeptide chain into any shape as desired. This tool proved to be very useful during the heuristic model-building process.

Pituitary somatotropins, or growth hormones, are responsible for normal growth and are involved in the regulation of various metabolic processes. Aside from their fundamental biological roles, somatotropins also have considerable commercial potential and have been used to treat dwarfism in humans, to enhance the growth rate of pigs, and to increase lactation in cows. Because of their importance in biology and their applications in medicine, agriculture, and related areas, there is increasing interest in exploring their molecular features (Brems et al., 1986; Violand et al., 1989; Brems & Havel, 1989; Mao, 1990). However, only the crystallographic study of a modified porcine somatotropin (pSt) by Abdel-Meguid et al. (1987) has provided 3-D structural information. In their study the N-terminal alanine residue was replaced by methionine to yield a biologically active Met-pSt analogue, which was shown to possess a four-α-helix bundle motif. The tertiary fold of the protein, adapted from the work of Abdel-Meguid et al. (1987), is depicted in Figure 1 and is seen to possess the antiparallel configuration of neighboring helices and the left-handed superhelical twist of the bundle, which are characteristic of essentially most four-α-helix bundle proteins (Weber & Salemme, 1980; Sheridan et al., 1982). In addition, the bundle is seen to possess a left-handed connective topology with "overhand" links between helices 1 and 2 (loop 1) and helices 3 and 4 (loop 3), as shown schematically in Figure 1. The kink in helix 2 is due to the presence of a proline residue that is conserved throughout the somatotropins. Due to their extremely high degree of sequence homology (Abdel-Meguid et al., 1987), it is very likely that all somatotropins would possess similar four-α-helix bundle motifs and similar connective topologies. The results of a recent study by Cohen and Kuntz (1987), in which a set of "heuristic" algorithms were used to predict the tertiary fold of human somatotropin (hSt), are consistent with this point of view, although they are not rigorous enough to independently confirm this assertion.

The goal of this study is to predict the 3-D structure of bovine somatotropin (bSt) based on the structure of Met-pSt. As shown in Figure 2, the sequence homology between pSt and bSt is extremely high (>90% based on exact matches, not just conserved amino acid residues). Thus, bSt is a suitable candidate for comparative or homology model-building based on the Met-pSt structure. The coordinates for Met-pSt have, however, not been made available by the authors, and thus the published Met-pSt structure can only serve as a qualitative, albeit quite helpful, guide in the modeling process. It is for this reason that a heuristic approach, which can effectively take advantage of limited data and fragmentary knowledge, was chosen in the present work: although some a priori folding algorithms exist, they currently are not robust or accurate enough to produce tertiary structures of sufficient accuracy.

On the basis of the high level of sequence homology between the two proteins, the locations of the helical segments in bSt can be determined with high probability from those in Met-pSt. This information can then be used in conjunction with energy-based helix-packing calculations to determine the structure of the bundle. Such calculations have been shown by Chou et al. (1988) and Carlacci and Chou (1990a) to predict the major structural features of four-α-helix bundle motifs in proteins successfully. Once the structure of the

‡ The coordinates of all atoms, including hydrogens, for the predicted three-dimensional bSt structure have been deposited with Protein Data Bank, Chemistry Department, Brookhaven National Laboratories, Upton, Long Island, NY 11973, from which copies are available.

* To whom correspondence should be addressed.

§ L.C. was a postdoctoral fellow in Upjohn Laboratories from July 1, 1988, to October 15, 1990. Present address: Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6059.

0006-2960/91/0430-4389$02.50/0

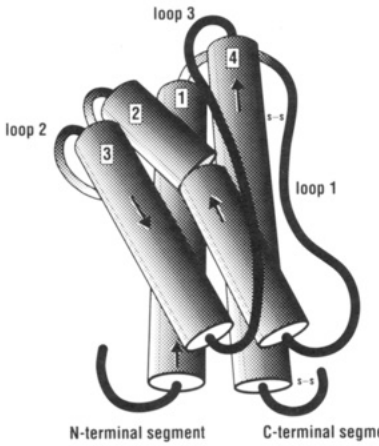

FIGURE 1: Schematic drawing to show the tertiary fold of the Met-pSt protein, adopted from the work of Abdel-Meguid et al. (1987). The cylinders, labeled 1–4, represent four α-helices. The tubes represent three loops, as well as C- and N-terminal segments. As shown in this picture, there are two S–S (disulfide) bonds: one is between helix 4 and loop 1, and the other is between the same helix and the C-terminal segment.

bundle has been determined, it remains to connect the helices with loops which ensure that the observed left-handed connective topology and other structural characteristics, such as the disulfide bond connecting helix 4 with loop 1 (Figure 1), are preserved. A heuristic approach can provide a practical means for accomplishing this, and this is the approach adopted here.

It is not expected that the current approach will yield a high-resolution structure comparable to what could be obtained by X-ray crystallographic methods. However, it is expected that the results obtained in the present work will be adequate to develop a useful structural model of bSt of suitable accuracy to answer questions concerning the relationship of bSt structure to its function.

The remainder of the paper is structured as follows. Methods and Formulation describes the methods employed in the energy calculations and energy-component analysis and the types of geometric parameters used to describe the key geometric features of four-α-helix bundles. The method for generating loops, however, involves more mathematical equations and is hence described in the Appendix. The following section describes the heuristic approach used in the construction of the tertiary structure of bSt, and the section after that analyzes the resulting structure in both geometric and energetic terms. There it is shown that the geometric features obtained for bSt correspond well with those generally observed for four-α-helix bundle proteins. More importantly, it is shown by energy component analysis that loop–helix interactions, especially those corresponding to van der Waals interactions, play a major role in stabilizing the overall helix bundle. The concluding section presents a summary of the work and also provides an appraisal of the usefulness of the heuristic approach employed here to specific types of protein structure prediction problems.

Methods and Formulation

(A) Energy Calculations and Energy Component Analysis. All energy calculations were carried out with ECEPP/2 (Némethy et al., 1983), which is an updated version of the original ECEPP program developed by Momany et al. (1975). As prescribed in ECEPP, only torsional degrees of freedom are allowed to vary; thus bond-stretching and angle-bending terms are omitted from the potential-energy function, which expresses the potential energy as a sum of electrostatic, nonbonded (van der Waals), hydrogen-bonding, and torsional terms. Energy minimizations were based on a general unconstrained optimization algorithm (Gay, 1983). All computations were carried out on an IBM 3090/400J computer at Upjohn Laboratories. The standard conventions for nomenclature of peptide conformations have been followed (IUPAC–IUB Commission on Biochemical Nomenclature, 1970).

In order to compare and analyze the various energetic contributions that contribute to the overall stability of bSt, it is useful to separate the total potential energy, E_{tot}, into components (Carlacci & Chou, 1990a), which are defined in eq 1. In these expressions, the loop segments include the N- and C-terminal segments in addition to loops 1, 2, and 3 (see, e.g., Figure 1).

(B) Geometric Parameters of Four-Helix Bundles. In proteins, the relative orientation of two structures is usually expressed in terms of an orientation angle (Chothia et al., 1977, 1981; Chou et al., 1983, 1984, 1990b), denoted by Ω. For two helices, the orientation angle Ω_0 measures the tilting of the helix axes, with $\Omega_0 = 0°$ for parallel and $\Omega_0 = \pm 180°$ for antiparallel orientations, respectively. The orientation angle is positive if, starting from an initial parallel orientation of the helices (Ω_0

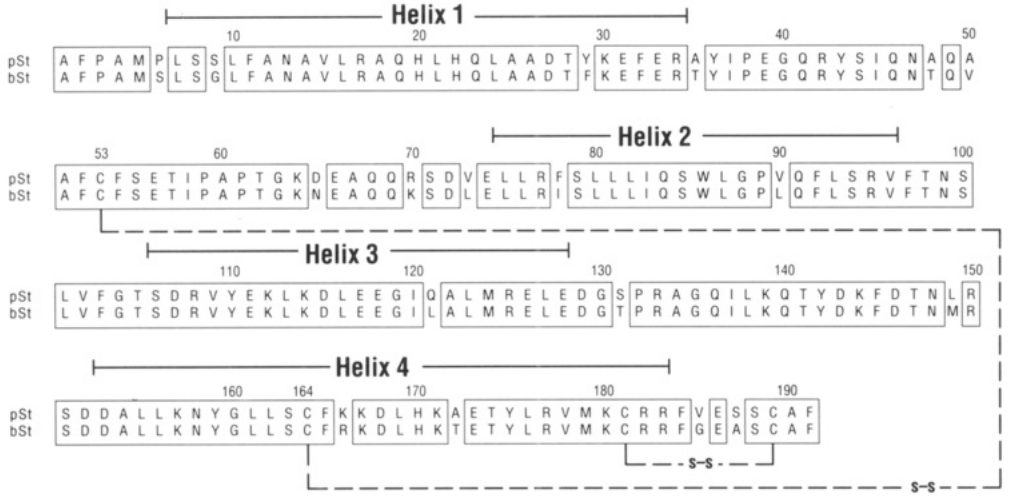

FIGURE 2: Amino acid sequence of pSt (Seeburg et al., 1983), and bSt (Graf & Li, 1974). Boxed amino acid residues are identical in these two sequences. The positions of the Met-pSt four helices are indicated above the amino acid sequences (Abdel-Meguid et al., 1987). Residue numbers are those used for the bSt molecule throughout this paper. The dotted lines connecting residue pairs (53, 164), and (181, 189) represent two disulfide bonds.

$$\begin{cases} E^{\alpha}_{\text{intra}} = \text{sum of the energies of the 4 individual constituent } \alpha\text{-helices} \\ E^{\alpha}_{\text{inter}} = \text{total intersegment energy among the 4 } \alpha\text{-helices} \\ E^{\alpha}_{\text{tot}} = E^{\alpha}_{\text{intra}} + E^{\alpha}_{\text{inter}} \\ E^{\text{loop}}_{\text{intra}} = \text{sum of the energies of individual loop segments} \\ E^{\text{loop}}_{\text{inter}} = \text{total intersegment energy among the loop segments} \\ E^{\text{loop}}_{\text{tot}} = E^{\text{loop}}_{\text{intra}} + E^{\text{loop}}_{\text{inter}} \\ E_{\text{tot}} = E^{\alpha}_{\text{tot}} + E^{\text{loop}}_{\text{tot}} + \varepsilon, \text{ where } \varepsilon \text{ is the interaction energy between the loop segments and the } \alpha\text{-helices of the molecule} \end{cases} \quad (1)$$

= 0°), the helix furthest from the viewer is rotated clockwise relative to the one nearest the viewer; it is negative if the furthest helix is rotated in the counterclockwise sense. Expressed in terms of unit vectors $\mathbf{e}_i$ and $\mathbf{e}_j$, which coincide with the helix axes (Chou et al., 1984) and point from the N- toward the C-terminus

$$\Omega_0 = \begin{cases} \cos^{-1}(\mathbf{e}_i \cdot \mathbf{e}_j) & \text{for clockwise rotation} \\ -\cos^{-1}(\mathbf{e}_i \cdot \mathbf{e}_j) & \text{for counterclockwise rotation} \end{cases} \quad (2)$$

The sign and magnitude of Ω_0 are independent of which helix is chosen as the nearest or furthest.

Since the left twisting is a typical feature in four-α-helix bundle proteins (Weber & Salemme, 1980), it is useful to develop a quantitative definition of this, which can be accomplished as follows. First, define the central axis of the helix bundle in an analogous manner to that used by Chou et al. (1990a) in their study of β-barrels; namely, that the central axis should pass through the mass center of all the C^{α} atoms of the four α-helices that compose the bundle, with the unit vector given by

$$\mathbf{e}_c = \tfrac{1}{4}\sum_{i=1}^{4} \tau_i \mathbf{e}_i \quad (3)$$

where $\mathbf{e}_i$ is the unit vector of the ith helix axis, as mentioned above, and

$$\tau_i = \begin{cases} 1 \text{ for } \mathbf{e}_1 \cdot \mathbf{e}_i \geq 0 \\ -1 \text{ for } \mathbf{e}_1 \cdot \mathbf{e}_i < 0 \end{cases} \quad (4)$$

This definition of τ_i ensures that the central axis is directed from the N- to the C-terminus of the first helix (Chou et al., 1985). Then the twist of a four-helix bundle can be defined as

$$\Theta = -\tfrac{1}{4}\sum_{i=1}^{4} \Omega^i \quad (5)$$

where Ω_i is the tilted angle of the ith helix with respect to the central axis and is defined by

$$\Omega_i = \begin{cases} \Omega_0^i & \text{for } -90° \leq \Omega_0 \leq 90° \\ \Omega_0^i - 180° & \text{for } 90° < \Omega_0 \leq 180° \\ \Omega_0^i + 180° & \text{for } -180° \leq \Omega_0 < -90° \end{cases} \quad (6)$$

where Ω_0^i is the orientation angle (cf. eq 2) between the ith helix axis and the central axis of the bundle. Note the difference between the orientation angle and the tilted angle as implied in eq 6: the former ranges from −180° to 180°, but the latter ranges from −90° to 90° so that a reversal of the axial direction of the ith helix is denoted by the same numerical value of Ω^i. In other words, when defining the orientation angle the directionality is assigned, but no such a directionality is assigned in defining the tilted angle. Obviously, in describing the twist of a four-helix bundle, the directionalities of axes are not of interest. Therefore, in eq 5 we should adopt tilted angles instead of orientation angles. According to the definition as given by eq 5, left-handed twisted, right-handed twisted, and nontwisted four-helix bundles are characterized by $\Theta < 0$, $\Theta > 0$, and $\Theta = 0$, respectively, and the larger the value of $|\Theta|$, the more the bundle is twisted.

Heuristic Approach to Tertiary Structural Prediction

The following two types of inputs are used to predict the 3-D structure of bSt: (a) the amino acid sequence as given in Figure 2 for the bSt molecule and (b) the heuristic inputs derived according to physical, chemical, and biological principles; e.g., the rotational constraint of a helix around its own axis due to optimal hydrophobic interaction, the distance constraint between some atoms due to disulfide bonds, and the similarity of loop shape due to homologous relationships. However, as noted in the introduction, the lack of coordinate data for pSt precludes a straightforward homology model building approach in the case of bSt. Nevertheless, data derived from the published structure of pSt, although limited, can serve as a practical guide in our modeling efforts. The heuristic approach employed here, which seeks to utilize this data to build a "working model" of bSt, consists of the following steps: (1) construct each individual α-helix by energy minimization, (2) generate a set of initial helix-bundle "core" structures, (3) "pack" the helices in each bundle by using energy minimization methods, (4) generate a set of appropriate loop structures, and (5) carry out full structural optimization on the resulting bSt structures.

(*1*) *Construction of Individual Helices.* The locations of residues in the bSt helices were taken from Figure 2 and are based on the close sequence homology between pSt and bSt. The values of the initial backbone dihedral angles for each residue were those in the computed minimum-energy conformation of an isolated regular poly(L-Ala) α-helix (Chou et al., 1988), namely, $(\phi, \psi, \omega) = (-68.0°, -38.0°, 180.0°)$. The initial side-chain dihedral angles were selected from the data (Vásquez et al., 1983) on the basis of energy minimization of the 20 amino acid residues (N-methylated and C-amidated) according to the following two criteria: (1) the corresponding ϕ and ψ must be in the helix region and (2) if more than two sets of side-chain dihedral angles are compatible with an α-helical backbone, only the one with relatively the lowest energy is selected. Each of the helices thus generated is subjected to energy minimization: initially only the side-chain dihedral angles were allowed to vary, followed by complete optimization of all dihedral angles. Interestingly, the bend in helix 2 (Figure 1) due to the presence of a conserved proline residue, i.e., Pro-89 (Figure 2), is well accounted for in the present calculations.

(*2*) *Generation of Initial Helix-Bundle Core Structures.* The four helices were initially positioned in an antiparallel arrangement with their axes perpendicular to a common plane such that the midpoint of each helix axis lies at an appropriate corner of a square lying in the plane. The length of each side of the square was taken to be 18 Å, the sum of the maximal radii of any two of the four helices (Chou et al., 1984), in order to remove repulsive atomic overlaps from the starting structure. As noted earlier, the antiparallel arrangement and relative placement of the helices are based on the results obtained by Abdel-Meguid et al. (1987) for Met-pSt as depicted in Figure 1.

In addition, since the helix bundle in Met-pSt possesses the usual left-handed twist observed for four-α-helix bundle pro-

teins (Weber & Salemme, 1980; Sheridam et al., 1982), a similar twist was introduced into the initial bSt helix-bundle structures. This reduced the number of structures (Chou et al., 1988) that must be minimized and aided in expediting the minimization process itself. To accomplish this, each helix was rotated by 20° around the connecting line between the midpoint of the helix axis and that of the central axis (Chou et al., 1990a). Although the initial value of the tilt angle is somewhat arbitrary, it is automatically adjusted by the energy minimization and ultimately reaches its optimal value for each of the four helices.

(*3*) *Packing of the Four Separate α-Helices.* The four α-helices in bSt were "packed" by means of energy minimization methods described earlier (Chou et al., 1984, 1985, 1986), although the implementation of analytical first derivatives (Carlacci & Chou, 1991) in the present work significantly reduced the overall computation time. Energy minimization of the helix bundle was carried out in two steps. First, only the side-chain dihedral angles were allowed to vary, which eliminates strong (or close) contacts between side-chain atoms. Second, the six rigid-body coordinates of each helix, three of which describe the midpoint of the helix axis and three of which are the Euler angles that describe its orientation (Chou et al., 1984), were also allowed to vary. Energy minimization with respect to the rigid-body variables provides a means for generating energetically favorable orientations and positions for each of the α-helices in the bundle (Chou et al., 1988).

In order to find the most favorable interactions between neighboring helices, a set of starting structures was generated by rotating each helix about its own axis such that the hydrophobic and hydrophilic surface features of each helix are appropriately positioned with hydrophobic residues toward the interior of the bundle and hydrophilic residue primarily on its "exposed" surface. In the case of helices 2 and 4, additional data were used to insure that the starting structures generated were consistent with other experimental results. On the basis of the work of Lehrman et al. (1990), Trp-86, which is located in the middle of helix 2, must face the inside core of the bundle. Furthermore, the bend in helix 2 due to Pro-89 must also be tilted toward the inside of the bundle (see Figure 1). Finally, Cys-164 and Cys-181 of helix 4 must face the outside of the bundle because they are involved in forming disulfide bonds with Cys-53 (loop 1) and Cys-189 (C-terminal segment), respectively (see Figure 1).

Thus, by means of MOSAIC facilities (Howe et al., 1991) which are based, in part, on MACROMODEL (Mohamadi et al., 1990), we can rotate each of the four helices on screen to the positions within the scope confined by the above constraints.

By using this procedure, 20 initial helix bundles were generated and subjected to energy minimization as described above. After the energy minimizations were done for all these 20 helix bundles, it was found that four of them converged to the same bundle, which is at least 50 kcal/mol lower in energy than the others. Such a bundle structure with the lowest packing energy was chosen as a "core" and used in all subsequent calculations.

(*4*) *Generating a Set of Initial Loop Structures.* The generation of loop structures is a nontrivial task even for relatively short loops, i.e., loops with 6–7 residues. A number of approaches have been developed to address this problem, and these include both knowledge-based and energy-based procedures (Bruccoleri & Karplus, 1987; Chou et al., 1989). Such procedures are not, however, applicable to loops of the size encountered in bSt: as indicated in Figures 1 and 2, loop 1 (40 residues) includes residues from Thr-35 to Glu-74, loop 2 (9 residues) from Phe-97 to Thr-105, and loop 3 (24 residues) from Asp-129 to Asp-152. Fortunately, the disulfide bond that covalently links Cys-51 in loop 1 to Cys-164 in helix 1 provides a tether at approximately the midpoint of loop 1, which serves to restrict the possible conformations of loop 1. Nevertheless, generation of appropriate loop 1 structures remains a very difficult problem.

In order to address the general problem of loop generation in bSt, a heuristic approach has been developed that uses templates based on structural information obtained from loops of Met-pSt (Figure 1) to reduce the "search space" of possible loop conformations. In order to realize this, a new procedure, described in the Appendix, was developed that is based on distance optimization and is functionally similar to the one described earlier for generating βαβ crossover structures (Chou et al., 1989). However, the new method differs in that it provides a means for controlling the "shape" of a loop during the connection process. Thus, initial loop structures, that is, structures prior to energy minimization, generated in this way will be "clustered" about the approximate structure of the loop in Met-pSt: there is no need to generate a wide diversity of loops that lie far from the basic structure of the loop as it appears in Met-pSt. As noted above, the existence of the disulfide bond connecting loop 1 with helix 4 further restricts the available geometries that loop 1 can attain. As described in the Appendix, the disulfide bond constraint can also be and is incorporated into the loop generation scheme. The short N- and C-terminal segments were also connected in a similar manner to mimic the corresponding segments observed in Met-pSt.

It is important to point out, however, that controlling the shape of a loop during loop generation is a "trial-and-error" process due to the limited "resolution" of the available Met-pSt structural data. Nevertheless, by requiring that the loops generated be smoothly connected to the appropriate helices and approximately resemble the corresponding loops in Met-pSt, one can always produce a reasonable set of initial loop structures by means of the technique given in the Appendix. To realize this, the following two criteria were used to decide whether to accept or reject the connected structures: (a) the distance-optimized function F_1 as defined by eq A1 in the Appendix must be smaller than 0.01 Å^2 (Chou et al., 1989) in order to guarantee a smooth connection and (b) the loop thus generated must be approximately similar in shape to the corresponding loop in Met-pSt (Figure 1), as detailed in the second section of the Appendix. More than 1000 loops were subjected to this process, resulting in six properly connected bSt structures for further minimization.

(*5*) *Energy Minimization of the Entire Protein Molecule.* The six conformations obtained in the previous step were subjected to further energy minimizations, which were carried out in three steps. First, all side-chain dihedral angles were allowed to vary; second, only the backbone dihedral angles in the loop regions were allowed to vary; and third, all the dihedral angles were freely varied. After energy minimizations, the one with the lowest total energy was chosen as the predicted bSt structure. Such a choice is based on the following grounds: (a) the predicted structure is at least 100 kcal/mol lower in energy than the other five structures and (b) the shape of its loops is more similar to that of Met-pSt. Not unexpectedly, the process of energy minimization not only relaxed the strain of the loops but also improved the packing, as will be discussed further in the next section.

The flow chart given in Figure 3 provides an overview of the entire heuristic model-building process used here.

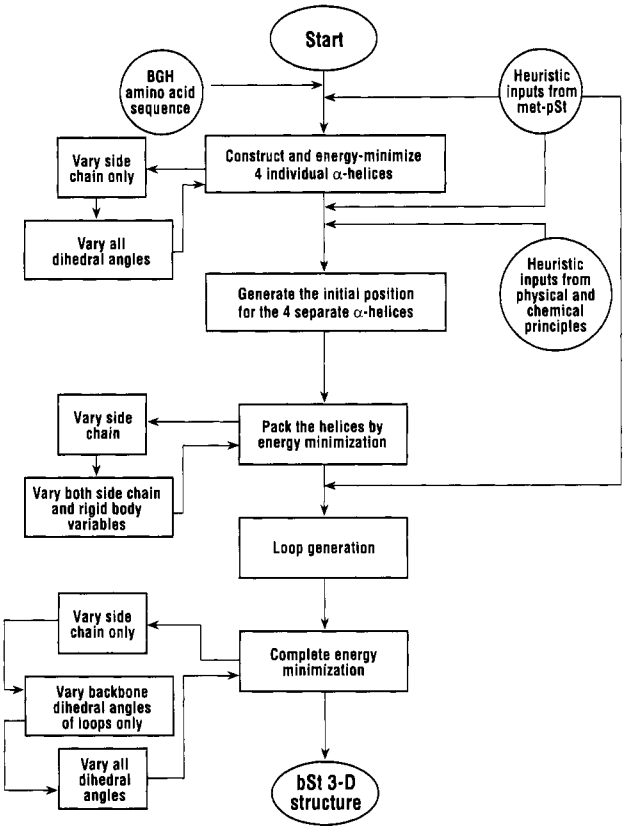

FIGURE 3: Flow diagram showing how to combine the heuristic approach and energy minimization to predict the 3-D structure of bSt protein.

Evaluation and Analysis of the Predicted Structure

The lowest energy structure found through the procedures described in the above section represents the predicted structure for the bSt molecule. Below, let us further analyze such a structure from both geometry and energetics aspects.

Geometrical Features. The geometrical parameters characterizing the four-helix bundle in the predicted bSt molecule are listed in Table I, from which we can see that all the orientation angles between two adjacent helices are $\Omega_0 = -152°$, within ±7°. This indicates that the four α-helices in bSt form a left-handed twisted bundle (Chou et al., 1988). The sense of twisting of the four-helix bundle in bSt agrees with that in many observed helix bundles of proteins (Argos et al., 1988; Weber & Salemme, 1980; Sheridan et al., 1982; Kaumaya et al., 1990) as well as its homologue Met-pSt (Abdel-Meguid et al., 1987).

In order to provide a quantitative description of such a bundle twisting feature in the bSt structure, in Table I are also listed the values of the tilted angles of helices 1–4 to their central axis and the distances of closest approach of the axes of helices 1–4 to the central axis. Substituting the data of Ω_i into eq 5 yields that the twisted angle of the four-helix bundle in bSt is $\Theta = -20°$. According to the definition as given in eq 5, the negative value of Θ indicates that the bundle is left-handed twisted.

Such a left-handed twisted feature of the four-helix bundle is also clearly shown by the stereo drawings in Figures 4 and 5.

Figure 4a is the ribbon drawing of the predicted structure, viewed from the flank of the four-helix bundle. Figure 4b is the corresponding drawing viewed looking down the bundle axis, with the N- and C-terminal ends close to the reader. The top and bottom panels of Figure 5 are space-filling drawings with only backbone atoms and all heavy atoms shown, respectively. In Figure 5, the helix regions are drawn in red and the loop regions as well as N- and C-terminal segments are drawn in yellow.

Table I: Geometric Parameters Characterizing the Four-Helix Bundle in the Predicted bSt Protein Structure

relationship between helices[a]				relationship between helix and central axis[b]		
adjacent pair	diagonal pair	Ω_0[c] (deg)	D[d] (Å)	helix[a]	Ω_i[e] (deg)	R_i[f] (Å)
1–3		−149	10.4	1	24	9.2
3–2		−159	12.2	2	23	8.6
2–4		−145	10.7	3	15	7.4
4–1		−152	11.2	4	19	6.9
	1–2	47	17.8			
	3–4	35	14.5	Θ[g]	−20	

[a] See Figures 1 and 4. [b] The central axis of a four-helix bundle is defined by eq 3. [c] Ω_0 is the orientation angle between two helices as defined by Chou et al. (1983, 1984). An explanation of such a definition can also be found in Chou et al. (1988). [d] D is the distance of closest approach between two helix axes as defined by Chou et al. (1983). [e] Ω_i is the tilted angle of the ith helix to the central axis of the four-helix bundle. [f] R_i is the closest approach (Chou et al., 1983) of the ith helix axis to the central axis of the four-helix bundle. [g] The twisted angle of the four-helix bundle is defined in eq 5.

It is seen from Figures 4 and 5 that helices 1 and 2 are connected by a long loop and helices 3 and 4 are connected by another long loop. Actually, as shown in the bottom panel of Figure 5, a considerable part of the flank surface of the four-helix bundle is closely embraced and packed by the two long loops. Residues 53, 164, 181, and 189 of the bSt molecule are cysteines (Figure 2). In the predicted structure, the distances between the S atom of residue 53 and that of residue 164 and between the S atom of residue 181 and that of residue 189 are 2.1 and 1.7 Å, respectively. Since such a distance in an idealized disulfide bond is 2.040 Å (Momany et al., 1975), it is reasonable to think that loop 1 (i.e. the loop connecting helices 1 and 2) is "hooked" with helix 4 (Figure 4a) through the disulfide bond between residues 53 and 164 and that the C-terminal attachment is forced to bend to helix 4 due to the existence of the disulfide bond between residues 181 and 189. Overall, the geometrical and conformational features of the loops in the predicted bSt structure are quite consistent with the observations by Tramontano et al. (1989) that large loops in proteins usually involve hydrogen bonding, disulfide bonding, and packing interactions within the loop and with other parts of the protein.

It is seen from Figure 4b that there is a remarkable bend in helix 2. This is because of the existence of the proline in helix 2 at the position of residue 89 (Figure 2). According to our calculation, the bend angle is 26.2°, very close to the angle of ≈30° observed by Abdel-Meguid et al. (1988) for the corresponding helix in the Met-pSt molecule.

The helices in the predicted structure are closely packed as shown in the bottom panel of Figure 5, even though all hydrogen atoms are not included. This can also be seen from the low adjacent interhelix axis distances of closest approach. These distances are 10.4, 12.2, 10.7, and 11.2 Å, respectively (Table I). According to the analysis by Weber and Salemme (1980) based on the available coordinate or literature data, the mean adjacent interhelix axis distance of closest approach in four-helix proteins is 9.6 Å, with a standard deviation of 1.4 Å, indicating that the interhelix distances in the predicted bSt structure fell within the general observed region.

On the helix regions there are 47 hydrophobic side chains, of which 32 are buried or partly buried inside the helix bundle, 11 are embraced by the loops, and only 4 are exposed. This

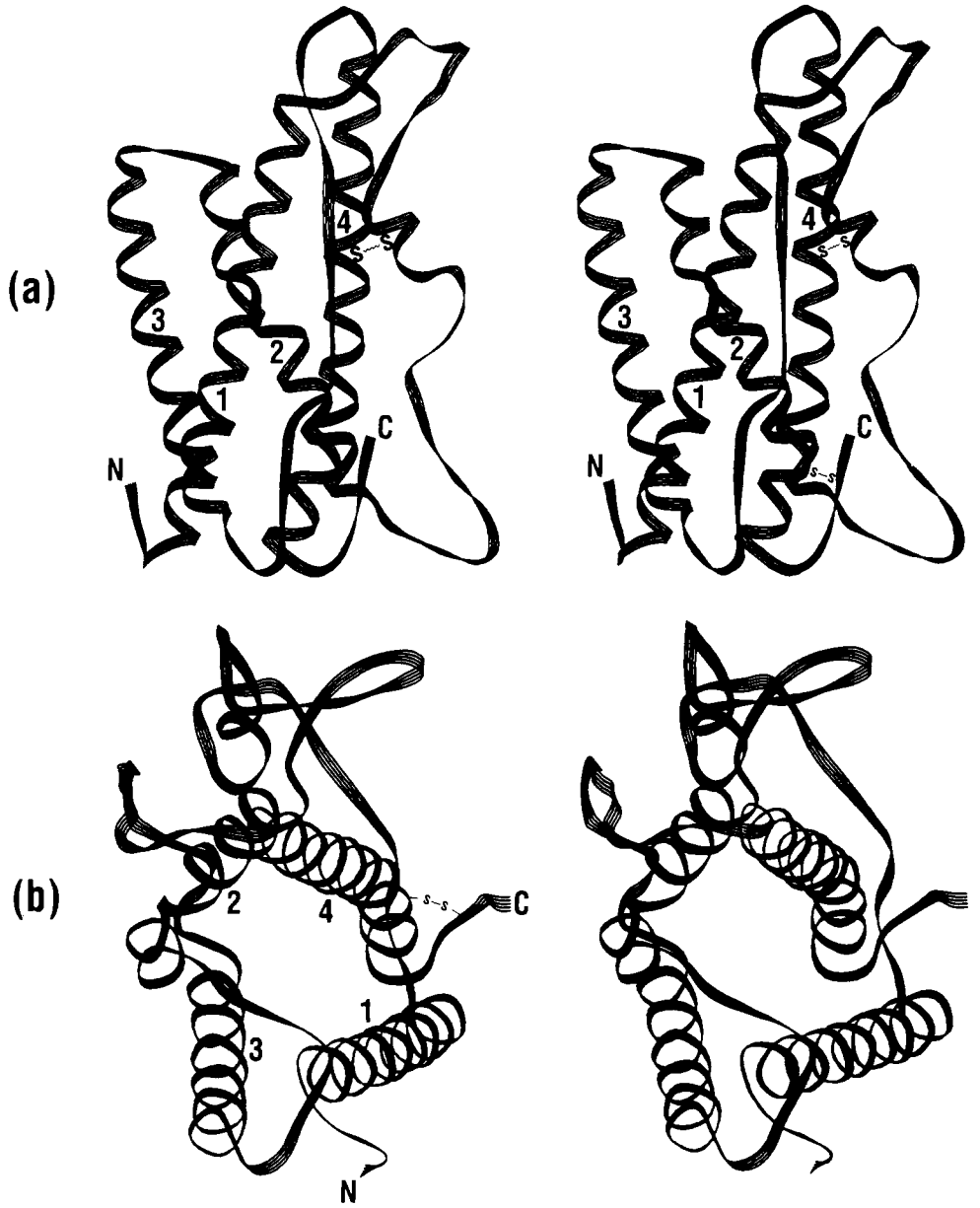

FIGURE 4: Stereo ribbon drawings for the predicted bSt structure viewed (a) from the flank of the four-helix bundle therein and (b) from the top of the bundle, with the N- and C-terminal ends close to the reader. The short zigzag line represents a disulfide bond. The four α-helices are labeled 1–4.

indicates that the core of the four-helix bundle is actually a hydrophobic one. Furthermore, the loop segments contain 32 hydrophobic side chains, of which 17 are buried or partly buried either between a helix and a loop or between loops. Therefore, in general 77% of hydrophobic side chains of the protein are buried or partly buried either inside the helix bundle or between a helix and a loop or between loops. This feature is definitely in favor of the structural stability when the solvation effect is taken into account.

Energetic Features. Listed in Table II are the values of the energetic terms as defined in eq 1 for the predicted bSt structure. We can see from Table II that the total conformational energy for the bSt molecule is $E_{tot} = -1040.7$ kcal/mol, of which −915.7 kcal/mol is from E_{tot}^{α}, the total energy of the α-helix set, 118.5 kcal/mol from E_{tot}^{loop}, the total energy of the loop segment set, and −243.5 kcal/mol from ϵ, the interaction energy between helix set and loop segment set.

It is interesting to note that the total energy of the loop segment set, E_{tot}^{loop}, is a positive term. It seems at first glance that the existence of loops would not help in stabilizing the structure. However, the existence of loops has generated an additional interaction energy ϵ, which is negative and very large in magnitude, overriding the positive energetic term of the loop segment set alone. Therefore, the loops did play an important role, although indirectly, in stabilizing the four-helix-bundle protein.

Although there have been many discussions about the dipolar or, generally speaking, electrostatic interactions among α-helices (Wada, 1976; Hol et al., 1978; 1981; Sheridan & Allen, 1980; Sheridan et al., 1982; Weber & Salamme, 1980; Chou et al., 1983, 1984, 1988; Gilson & Honig, 1989) and their role in stabilizing the packing of α-helices, the role of dipole interactions in stabilizing (Sheridan et al., 1982) or destabilizing (Gilson & Honig, 1989) the α-helix bundle is still a controversial issue (Chou et al., 1990b). Therefore, it will certainly provide us with useful insights into this problem if each of the interaction energy terms as defined in eq 1 are further separated into electrostatic energy and nonbonded energy, respectively. Such itemized results are also listed in Table II, from which we can see that the electrostatic energy

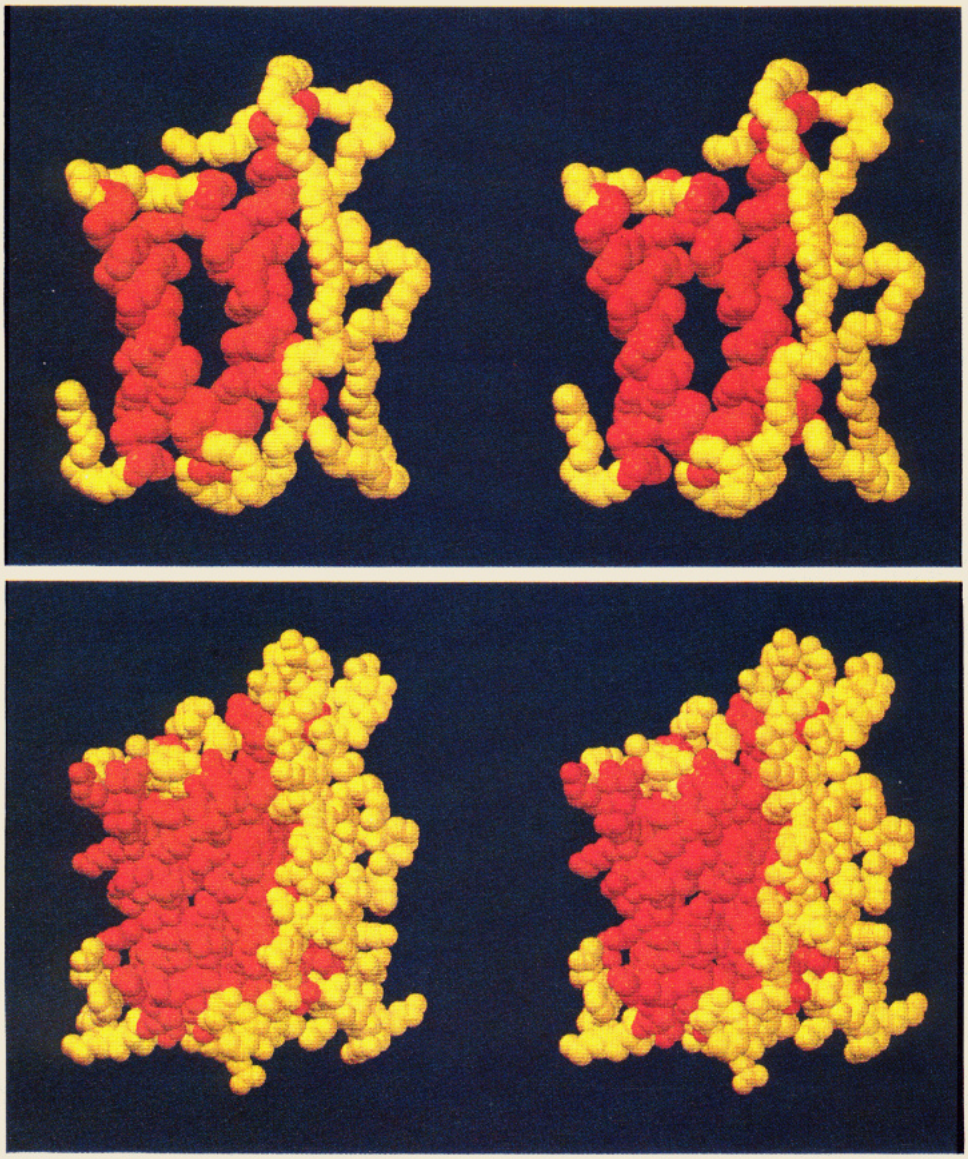

FIGURE 5: Space-filling stereo drawings for the predicted bSt structure: (top) only backbone atoms are included and (bottom) all atoms, except hydrogen atoms, are included. Atoms in the four α-helix regions are drawn in red, and all the remaining atoms are drawn in yellow.

among the four α-helices is $\epsilon_{ES}^{\alpha} = -5.5$ kcal/mol, the electrostatic energy among the loop segments is $\epsilon_{ES}^{loop} = -2.7$ kcal/mol, but the electrostatic energy between the α-helices and loop segments is $\epsilon_{ES} = -33.1$ kcal/mol. This means that, in the four-helix bundle, the electrostatic interaction between helices and loop segments is about 6 times the size of that among the four α-helices and about 12 times that among the loop segments themselves. Furthermore, it is shown in Table II that $\epsilon_{NB}^{\alpha} = -115.7$ kcal/mol, $\epsilon_{NB}^{loop} = -35.4$ kcal/mol, and $\epsilon_{NB} = -218.3$ kcal/mol, indicating that even for nonbonded energy the contribution from the interaction between helices and loop segments is much larger than those among the four α-helices and among the loop segments themselves.

Consequently, according to our calculation here, it is the interaction between helices and loop segments that plays a dominant role in stabilizing the four-helix bundle protein from the viewpoint of both electrostatic and nonbonded interactions, in agreement with the conclusion drawn earlier by analyzing the energetics of a theoretical four-helix bundle model (Carlacci & Chou, 1990a,b). It should be pointed out, however, that all these calculations were based on ECEPP (Momany et al., 1975). As pointed out by one of the reviewers, it would be interesting to see whether the conclusion will still hold true if the calculations are carried out in terms of other empirical potentials, such as those in CHARMM (Brooks et al., 1983) and AMBER (Weiner et al., 1984), where the electrostatic energy is much more heavily emphasized than in ECEPP.

Finally, it is interesting to make a comparison between the energy-minimized four-helix bundle obtained by packing four separate bSt α-helices as described under Heuristic Approach to Tertiary Structural Prediction (in section 3) and the four-helix bundle obtained by minimizing the energy of the entire bSt molecule as described (in section 5). For the sake of distinction, the former is called the *isolated* bSt helix bundle, representing that no loop effect is involved in its formation, and the latter is called the *nonisolated* bSt helix bundle, meaning that its formation is affected by the existence of loops. Listed in Table III are the geometric and energetic parameters for the energy-minimized *isolated* bSt helix bundle, i.e., the structure obtained by packing four separate bSt α-helices without the presence of loops. However, the geometric parameters listed in Table I and the corresponding energetic parameters in Table II were obtained in the presence of loops. Comparing Tables III and I, we find that, owing to the loop effect, the average closest approach between adjacent pair of helices was reduced from 11.6 to 11.1 Å, i.e., by 0.5 Å, and

Table II: Various Energetic Terms[a] Characterizing the Predicted bSt Structure

Energy of α-Helix Set (kcal/mol)				
intrahelix	interhelix			total
E^{α}_{intra}	ϵ^{α}_{ES} [b]	ϵ^{α}_{NB} [c]	E^{α}_{inter} [d]	E^{α}_{tot}
−794.6	−5.5	−115.7	−121.1	−915.7
Energy of Loop Set (kcal/mol)				
intraloop	interloop			total
E^{loop}_{intra}	ϵ^{loop}_{ES} [e]	ϵ^{loop}_{NB} [f]	E^{loop}_{inter} [g]	E^{loop}_{tot}
156.6	−2.7	−35.4	−38.1	118.5
Loop–Helix Interaction Energy (kcal/mol)				
ϵ_{ES} [h]	ϵ_{NB} [i]	ϵ^{S-S}_{tor} [j]	ϵ [k]	
−33.1	−218.3	7.9	−243.5	
Total Energy of bSt Molecule (kcal/mol)				
$E_{tot} = E^{\alpha}_{tot} + E^{loop}_{tot} + \epsilon = -1040.7$				

[a] See eq 1 and the relevant footnote below for the definition of each of the energetic terms listed in this table. [b] Electrostatic interhelix energy. [c] Nonbonded interhelix energy. [d] $E^{\alpha}_{inter} = \epsilon^{\alpha}_{ES} + \epsilon^{\alpha}_{NB}$. [e] Electrostatic interloop energy. [f] Nonbonded interloop energy. [g] $E^{loop}_{inter} = \epsilon^{loop}_{ES} + \epsilon^{loop}_{NB}$. [h] Electrostatic loop–helix interaction energy. [i] Nonbonded loop–helix interaction energy. [j] Torsional energy due to the disulfide bonds between helices and loops. [k] $\epsilon = \epsilon_{ES} + \epsilon_{NB} + \epsilon^{S-S}_{tor}$.

Table III: Geometric and Energetic Parameters[a] Characterizing the Energy-Minimized Isolated bSt Helix-Bundle Structure

Geometric Parameters						
adjacent pair	diagonal pair	Ω_0 (deg)	D (Å)	helix	Ω_i (deg)	R_i (Å)
1–3		−145	10.5	1	27	9.4
3–2		−160	12.0	2	25	9.2
2–4		−142	12.5	3	15	7.4
4–1		−147	11.5	4	22	8.0
	1–2	51	18.7			
	3–4	37	15.7	Θ	−22	

Energetic Parameters (kcal/mol)				
intrahelix energy	interhelix energy			total energy
E^{α}_{intra}	ϵ^{α}_{ES}	ϵ^{α}_{NB}	E^{α}_{inter}	E^{α}_{tot}
−782.4	−4.1	−99.6	−103.7	−886.1

[a] See the corresponding footnotes to Tables I and II for definitions of the symbols in this table.

that between diagonal pair of helices was reduced from 17.2 to 16.1 Å, i.e., by 1.1 Å, while the average closest approach of the four helices to the central axis was reduced from 8.5 to 8.0 Å, i.e., by 0.5 Å. Comparing Tables III and II, we further find that the total energy of the four-helix bundle was lowered from −886.1 to −915.7 kcal/mol, i.e., by 29.6 kcal/mol. All of these indicate that, driven by the interaction between loops and helices, the nonisolated four-helix bundle has a better pack than the isolated four-helix bundle, further supporting the conclusion drawn from our previous study on a theoretical model that the interaction between loops and the helices plays a significant role in stabilizing four-helix bundle structures (Carlacci & Chou, 1990a,b).

Summary and Conclusions

The predicted 3-D structure of bovine somatotropin, composed of 191 amino acids, is a four-helix bundle protein, with two long loops (loop 1 and loop 3, whose residue numbers are 40 and 24) between helices 1 and 2 and between helices 3 and 4, respectively, and a short loop (loop 2, whose residue number is 9) between helices 2 and 3. Two short segments with 6 and 8 residues are attached to the N-terminus of helix 1 and the C-terminus of helix 4, respectively. There are two disulfide bonds. One is formed between residue 53 in loop 1 and residue 164 in helix 4, and the other is formed between residue 181

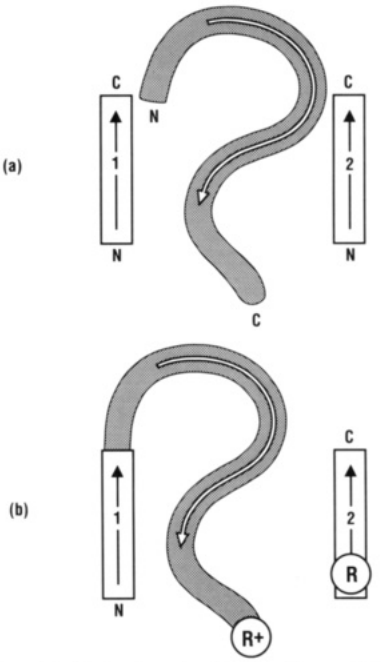

FIGURE 6: Schematic illustration showing how to use FOLD to connect two rigidly fixed structures with a flexible polypeptide chain: (a) The ribbons marked 1 and 2 represent two rigidly fixed structures, and the ribbon shaded with dots represents the flexible chain. (b) The N-terminus of the flexible chain is attached to the C-terminus of the first fixed chain, and then a dummy residue R^+, which is identical with the first residue, R, of the second fixed structure, is added to the end of the combined chain. See text for further explanation.

in helix 4 and residue 189 in the C-terminal attachment. The four α-helices are closely packed to form a left-twisted bundle, as observed in most four-helix bundle proteins (Argos et al., 1977; Weber & Salemme, 1980; Sheridan et al., 1982; Chou et al., 1988, 1990b; Kaumaya et al., 1990). A considerable part of the surface of the 4-helix bundle is embraced by the two long loops. More than 90% of hydrophobic chains on the helix regions are buried inside due to either the packing of the helices or the embrace of the loops. Particularly, there are 32 hydrophobic side chains that are buried or partly buried inside the helix bundle, indicating that the core of the 4-helix bundle is a hydrophobic one.

It is found through energetic analysis that, in forming a stable four-helix bundle structure, an overwhelming majority of favorable energy is from the interaction between α-helices and loops. This is particularly true for the electrostatic energy, in which the contribution from the interaction between α-helices and loops segments is about 6 times as large as that among the four α-helices themselves. Similar results have also been obtained recently from an energetic analysis of an α/β barrel, in which it was found that the interaction between the loops and all the other part of the molecule contributes a significant favorable energy in stabilizing the α/β barrel (Chou & Carlacci, 1991).

The heuristic approach developed here is an effective and feasible avenue to predict the three-dimensional structure of a protein on the basis of limited data and fragmentary knowledge, as indicated by its practical implementation for bovine somatotropin.

Acknowledgments

Illuminating discussions with Dr. R. Lehrman are gratefully acknowledged. We are also very much indebted to the reviewers, whose comments were very helpful in improving this paper.

APPENDIX

FOLD: Generation of a Loop between Two Rigidly Fixed Structures. The technique FOLD developed here can be used to connect two rigid polypeptide segments, placed in fixed relative positions, with a flexible peptide segment following a desired trend; i.e. to determine the dihedral angles in the latter that allow the smooth joining of the segments according a designated topological path. This is achieved by minimizing an object function F with respect to the dihedral angles of the flexible chain. Therefore, in principle, such a new technique can be used to fold a polypeptide chain into any pattern as desired. The principle of this procedure is similar to the one formulated earlier for the generation of $\beta\alpha\beta$ crossover fold structure (Chou et al., 1989). However, the earlier technique (Chou et al., 1989) can only generate a smooth connection, but the power in forcing the connecting chain to follow a designated path is relatively much weaker. Here let us give a detailed description of the FOLD technique.

Suppose two rigidly fixed structures are to be connected by a flexible polypeptide chain (Figure 6a). As a first step, the N-terminus of the flexible chain is attached to the C-terminus of the first fixed chain, and then a dummy residue R^+ is added to the end of the combined chain (Figure 6b). The dummy residue is identical with residue R, the first residue of the second fixed structure. The dihedral angles of these two identical residues, R and R^+, are also the same. This can be easily fulfilled by the routine chain extension procedure as described in ECEPP (Momany et al., 1975). The new chain thus obtained is called the "torso–arm" structure, in which the "torso" belongs to the fixed part of the structure and the "arm" belongs to the flexible part. Now let us introduce three subobject functions.

(*1*) *Subobject Function* $F_1 = F(dummy)$. In order to generate a smooth connection of the torso–arm structure with the second fixed structure, the dummy residue R^+ in the flexible arm must be superposed with the residue R in the second fixed structure (Figure 6b). This can be realized by minimizing the following distance function with respect to the dihedral angles of the flexible arm (Chou et al., 1989):

$$F_1 = F_1(\text{dummy}) = [\mathbf{r}(N^+) - \mathbf{r}(N)]^2 + [\mathbf{r}(C^{\alpha +}) - \mathbf{r}(C^\alpha)]^2 + [\mathbf{r}(C'^+) - \mathbf{r}(C')]^2 + [\mathbf{r}(O^+) - \mathbf{r}(O)]^2 \quad \text{(A1)}$$

where $\mathbf{r}(X^+)$ and $\mathbf{r}(X)$ are position vectors for atom X in the dummy residue R^+ and residue R, respectively (Figure 6b). In the course of the fitting procedure, the dihedral angles (ϕ,ψ) of the specified residues in the flexible arm are allowed to vary (with ω fixed at 180°, until a minimum in F_1 is reached, i.e., until the dummy residue R^+ is superposed on the terminal residue R of the second fixed structure. At this point a continuous polypeptide chain that incorporates the three segments has been generated.

(*2*) *Subobject Function* $F_2 = F(C^\alpha)$. During the connection process via distance minimization, in order to force the flexible arm to follow a designated path, a second object function F_2 is imposed, which can be expressed as

$$F_2 = F_2(C^\alpha) = \sum_i [\mathbf{r}(C_i^\alpha) - \mathbf{r}_i^\circ]^2 \quad \text{(A2)}$$

where $\mathbf{r}(C_i^\alpha)$ is the position vector of the ith of the C^α atoms selected from the flexible arm and $\mathbf{r}_i^\circ$ is the desired position of the atom. For the case studied here, $\mathbf{r}_i^\circ$ (i = 1, 2, ...) values were defined in the UVW system (Chou et al., 1985), based on the axes of the two helices to be connected, and their quantities were selected according to the loop shape of Met-pSt as illustrated in Figure 1. Thus, the minimization of F_2 will hold the selected C^α atoms of the flexible arm at a series of designated positions so as to force the loop to follow a desired path during connection. However, since no atomic coordinates of Met-pSt are available, the selection of these quantities can only be approximated on the basis of its outline as shown in Figure 1. Therefore, the selection of these data is actually a "trial-and-error" process, during which we have to often show the loops thus generated on a screen for a comparison so that these data can be gradually adjusted to improve the similarity.

(*3*) *Subobject Function* $F_3 = F_3(S\text{–}S)$. In order to bring two designated cysteines together to form a disulfide bond, a third object function F_3 is introduced:

$$F_3 = F_3(\text{S–S}) = \sum_j [(\mathbf{r}(S_1)_j - \mathbf{r}(S_2)_j) - 2.040]^2 + [(\mathbf{r}(S_1)_j - \mathbf{r}(C_2^\beta)_j) - 3.052]^2 + [(\mathbf{r}(C_1^\beta)_j - \mathbf{r}(S_2)_j) - 3.052]^2 + [(\mathbf{r}(C_1^\beta)_j - \mathbf{r}(C_2^\beta)_j) - 3.855]^2 \quad \text{(A3)}$$

where $\mathbf{r}(X_1)_j$ and $\mathbf{r}(X_2)_j$ are position vectors for atom X in the jth pair of cysteines, respectively. The index j in eq A3 runs over all cysteine pairs forming disulfide bonds in the segments concerned. The minimization of F_3 will assure that the relevant atomic distances between the two cysteines are as close to those in an idealized disulfide bond.

Therefore, in order to include all the above functions, the objective function F should be composed of F_1, F_2, and F_3:

$$F = F_1(\text{dummy}) + F_2(C^\alpha) + F_3(\text{S–S}) \quad \text{(A4)}$$

By minimizing function F as defined above, the flexible chain can be manipulated according to a designated topological path to connect any two fixed structures. Therefore, using the technique developed here, one can in principle fold a polypeptide chain into any pattern as desired.

Registry No. bSt, 66419-50-9; growth hormone, 9002-72-6.

REFERENCES

Abdel-Meguid, S. S., Shieh, H. S., Smith, W. W., Dayringer, H. E., Violand, B. N., & Bentle, L. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6434–6437.

Argos, P., Rossmann, M. G., & Johnson, J. E. (1977) *Biochem. Biophys. Res. Commun. 75*, 83–86.

Brems, D. N., & Havel, H. A. (1989) *Proteins: Struct., Funct., Genet. 5*, 93–95.

Brems, D. N., Plaisted, S. M., Kauffman, E. W. & Havel, H. A. (1986) *Biochemistry 25*, 6539–6543.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Bruccoleri, R. E., & Karplus, M. (1987) *Biopolymers 26*, 137–168.

Carlacci, L., & Chou, K. C. (1990a) *Protein Eng. 3*, 509–514.

Carlacci, L., & Chou, K. C. (1990b) *Protein Eng. 4*, 225–227.

Carlacci, L., & Chou, K. C. (1991) *J. Comput. Chem. 12*, 410–415.

Chothia, C., Levitt, M., & Richardson, D. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4130–4134.

Chothia, C., Levitt, M., & Richardson, D. (1981) *J. Mol. Biol. 145*, 215–250.

Chou, K. C., & Carlacci, L. (1991) *Proteins: Struct., Funct., Genet. 9*, 280–295.

Chou, K. C., Némethy, G., & Scheraga, H. A. (1983) *J. Phys. Chem. 87*, 2869–2881.

Chou, K. C., Némethy, G., & Scheraga, H. A. (1984) *J. Am. Chem. Soc. 106*, 3161–3170.

Chou, K. C., Némethy, G., Rumsey, S., Tuttle, R. W., & Scheraga, H. A. (1985) *J. Mol. Biol. 186*, 591–609.

Chou, K. C., Némethy, G., Rumsey, S., Tuttle, R. W., & Scheraga, H. A. (1986) *J. Mol. Biol. 188*, 641-649.
Chou, K. C., Maggiora, G. M., Némethy, G., & Scheraga, H. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4295-4299.
Chou, K. C., Némethy, G., Pottle, M., & Scheraga, H. A. (1989) *J. Mol. Biol. 205*, 241-249.
Chou, K. C., Carlacci, L., & Maggiora, G. M. (1990a) *J. Mol. Biol. 213*, 315-326.
Chou, K. C., Némethy, G., & Scheraga, H. A. (1990b) *Acc. Chem. Res. 23*, 134-141.
Cohen, F. E., & Kuntz, I. D. (1987) *Proteins: Struct., Funct., Genet. 5*, 162-166.
Gay, D. M. (1983) *Assoc. Comput. Mach. Trans. Math. Software 9*, 503-524.
Gilson, M. K., & Honig, B. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1524-1528.
Hol, W. G. J., van Duijnen, P. T., & Berendsen, H. J. C. (1978) *Nature 273*, 443-446.
Hol, W. G. J., Halie, L. M., & Sander, C. (1981) *Nature 294*, 532-536.
Howe, W. J., Blinn, J. R., Moon, J. B., Hagadone, T. R., White, G. J., & Schultz, M. W. (1991) *J. Comput. Chem.* (submitted for publication).
IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry 9*, 3471-3479.
Kaumaya, P. T. P., Berndt, K. D., Heidorn, D. B., Trewhella, J., Kezdy, F. J., & Goldberg, E. (1990) *Biochemistry 29*, 13-23.
Lehrman, R., Tuls, J., Lund, M., & Havel, H. (1990) Model Studies of Bovine Growth Hormone Aggregation, presented at the Royal Society of Chemistry Conference on Protein Stability, Cambridge, U.K., March 26-29, 1990.
Mao, B. (1990) *Biopolymers 30*, 645-647.
Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., & Still, W. C. (1990) *J. Comput. Chem. 11*, 440-466.
Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361-2381.
Némethy, G., Pottle, M. S., & Scheraga, H. A. (1983) *J. Phys. Chem. 87*, 1883-1887.
Regan, L., & DeGrado, W. F. (1988) *Science 241*, 976-978.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167-339.
Sheridan, R. P., & Allen, L. C. (1980) *Biophys. Chem. 11*, 133-136.
Sheridan, R. P., Levy, R. M., & Salemme, F. R. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4545-4549.
Tramontano, A., Chothia, C., & Lesk, A. M. (1989) *Proteins: Struct., Funct., Genet. 6*, 382-394.
Vásquez, M., Némethy, G., & Scheraga, H. A. (1983) *Macromolecules 16*, 1043-1049.
Wada, A. (1976) *Adv. Biophys. 19*, 1-63.
Weber, P. C., & Salemme, F. R. (1980) *Nature 287*, 82-84.
Weiner, S. J., Kollman, P. A., Case, D. A., Chandra Singh, U., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765-784.

Proton NMR Assignments of Heme Contacts and Catalytically Implicated Amino Acids in Cyanide-Ligated Cytochrome *c* Peroxidase Determined from One- and Two-Dimensional Nuclear Overhauser Effects†

James D. Satterlee*
Department of Chemistry, Washington State University, Pullman, Washington 99164-4630
James E. Erman
Department of Chemistry, Northern Illinois University, DeKalb, Illinois 60115
Received December 12, 1990; Revised Manuscript Received February 13, 1991

ABSTRACT: Proton NMR assignments of the heme pocket and catalytically relevant amino acid protons have been accomplished for cyanide-ligated yeast cytochrome *c* peroxidase. This form of the protein, while not enzymatically active itself, is the best model available (that displays a resolvable proton NMR spectrum) for the six-coordinate low-spin active intermediates, compounds I and II. The assignments were made with a combination of one- and two-dimensional nuclear Overhauser effect methods and demonstrate the utility of NOESY experiments for paramagnetic proteins of relatively large size (M_r 34 000). Assignments of both isotope exchangeable and nonexchangeable proton resonances were obtained by using enzyme preparations in both 90% H_2O/10% D_2O and, separately, in 99.9% D_2O solvent systems. Complete resonance assignments have been achieved for the proximal histidine, His-175, and His-52, which is a member of the catalytic triad on the distal side of the heme. In addition, partial assignments are reported for Trp-51 and Arg-48, catalytically important residues, both on the distal side. Aside from His-175, partial assignments for amino acids on the proximal side of the heme are proposed for the alanines at primary sequence positions 174 and 176 and for Thr-180 and Leu-232.

Cytochrome *c* peroxidase (EC 1.11.1.5, CcP) is a 34-kDa ferriheme enzyme that reduces hydrogen peroxide using cytochrome *c* (Kraut, 1981; Poulos & Finzel, 1984). Mechanisms for its function have been proposed that involve acid-base catalysis and charge stabilization by a catalytically active triad of amino acids: Trp-51, His-52, and Arg-48 (Poulos & Kraut, 1980; Kraut, 1981). These amino acids lie close to the heme and the ligand-binding site on the distal side of the heme (Poulos & Kraut, 1980; Finzel et al., 1984; Edwards et al.,

† This work was supported by grants from the NIH (DK 30912, HL 01758 and RR 0631401 to J.D.S.) and the NSF (DMB 8716459 to J.E.E.). We are grateful for this support.
* To whom correspondence should be addressed.

0006-2960/91/0430-4398$02.50/0