Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6434–6437, September 1987
Biochemistry

Three-dimensional structure of a genetically engineered variant of porcine growth hormone

(somatotropin/crystal structure/protein folding)

SHERIN S. ABDEL-MEGUID, HUEY-SHENG SHIEH, WARD W. SMITH, HENRY E. DAYRINGER, BERNARD N. VIOLAND, AND LARRY A. BENTLE

Monsanto, 700 Chesterfield Village Parkway, Chesterfield Village, MO 63198

Communicated by Howard A. Schneiderman, May 26, 1987 (received for review April 9, 1987)

ABSTRACT **The three-dimensional structure of a genetically engineered variant of porcine growth hormone, methionyl porcine somatotropin (MPS), has been determined at 2.8-Å resolution, using single crystal x-ray diffraction techniques. Phases were obtained by use of a single isomorphous K_2OsCl_6 derivative and were improved by use of the density modification procedure. The MPS structure is predominantly helical. It consists mainly of four antiparallel α-helices arranged in a left twisted helical bundle, a structural motif observed in a number of other unrelated proteins. However, the way the four helices are connected in the bundle is unusual and, to our knowledge, has never been reported before. Alignment of the amino acid sequence of MPS with that of other growth hormones reveals that residues within the α-helices are predominantly invariant and thus these invariant residues are necessary to maintain the structural integrity of these proteins.**

Growth hormone, also known as somatotropin, is a 22,000-dalton protein secreted by the anterior pituitary gland in mammals. It is required for normal development during childhood and is involved in the regulation of a variety of anabolic processes. The family of growth hormones is related to two other families of hormones, the prolactins, also secreted by the anterior pituitary, and the placental lactogens (1–4). The amino acid sequence homology among the members of these families is considerable, suggesting that their three-dimensional structures are basically the same. The elucidation of the three-dimensional structure of any member of these families should provide considerable insight into the molecular basis of action of these hormones.

Recently, interest in understanding the structure–function relationships of these hormones has increased. This is partially due to their important commercial applications for human health and agriculture. The ability to produce large quantities of these proteins using genetic engineering techniques has made these applications possible.

Procedures for the crystallization of two of these hormones were reported (5, 6) as early as 1948. The crystals produced by such procedures were very small and not suitable for crystallographic studies. Efforts to obtain more suitable crystals were hampered for years by the heterogeneity of these hormones isolated from natural sources. In recent years, improved methods for protein purification and the availability of homogeneous proteins through genetic engineering have made it possible to produce crystals of three of these hormones suitable for single crystal x-ray diffraction studies. These include a member of the placental lactogen family, human chorionic somatomammotropin (7), and two members of the growth hormone family, bovine growth hormone (8) and a genetically engineered variant of porcine growth hormone (9). We report here the three-dimensional structure of the latter protein, methionyl porcine somatotropin (MPS; ref. 10), in which the alanine residue at position one in porcine growth hormone has been replaced with a methionine. This structure was determined from a 2.8-Å electron density map. To our knowledge, this is the first structure determination reported for any of these hormones.

METHODS

The type II MPS crystals used in this study were grown as described (9). These crystals are trigonal and belong to the space group $P3_221$, with $a = b = 87.7$ Å and $c = 58.7$ Å. They contain one monomer of MPS in the crystallographic asymmetric unit. X-ray diffraction data were measured from a single native crystal using a two-dimensional position sensitive detector (11). Approximately 28,000 reflections were measured in 1 day to give 6300 unique reflections to 2.8-Å resolution. Heavy-atom derivative Friedel pairs of reflections were measured from seven different crystals using a Rigaku AFC-5 four-circle diffractometer on a Rigaku RU200 rotating anode x-ray generator using an ω scan technique and graphite-monochromated CuKα radiation. For the diffractometer data, crystal decay was monitored by 10 standard reflections, and an absorption correction was applied according to the semiempirical method using ψ scans (12) of reflections with χ values near 90°. Data from the seven derivative crystals were corrected for decay, which averaged 18.7%, and were scaled together giving an R factor of 7.2% for 14,051 reflections,

$$R = \frac{\Sigma|I_i - \langle I\rangle|}{\Sigma\langle I\rangle}$$

(where I_i is the intensity of reflection i and $\langle I\rangle$ is the mean intensity of that reflection).

The structure was determined at 2.8-Å resolution using single isomorphous phases from a K_2OsCl_6 derivative. The position of the single osmium atom bound to the protein molecule was identified using Patterson techniques and refined by conventional methods (13). A summary of heavy-atom derivative preparation and refinement statistics is shown in Table 1. Although the calculated single isomorphous replacement (SIR) map at 2.8 Å clearly showed 80% of the MPS molecule, it was difficult to trace the backbone completely. Further improvement of this SIR map using the density modification procedure described by Wang (14) produced a significantly improved map. In the latter map it was much easier to follow the polypeptide chain, to identify the locations of the amino and carboxyl termini, and to identify the locations of the two disulfide bonds. The side chain density was clear for most residues, particularly the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MPS, methionyl porcine somatotropin; SIR, single isomorphous replacement.

Table 1. K_2OsCl_6 heavy-atom derivative preparation and refinement

	Resolution limit, Å								
	10.7	7.8	5.9	4.9	4.1	3.6	3.1	2.8	Total
f-rms/E-rms	2.10	2.05	2.64	3.97	2.78	2.49	1.97	1.54	2.44
R_c	0.49	0.43	0.44	0.38	0.55	0.60	0.68	0.66	0.53
Number of reflections	89	212	367	563	801	1049	1270	1446	5797
Mean figure of merit	0.67	0.70	0.61	0.76	0.61	0.52	0.39	0.28	0.48

f-rms, the rms calculated for heavy-atom structure factor; E-rms, the rms lack of closure determined from centric reflections only;

$$R_c = \frac{\Sigma \| F_{PH} - F_P | - F_H |}{\Sigma | F_{PH} - F_P |},$$

where F_{PH} and F_P are the observed structure factor amplitudes for the derivative and parent crystals, respectively, and F_H is the calculated heavy-atom structure amplitude. Soaking solutions included 35% saturated ammonium sulfate in 100 mM Tris·HCl (pH 7.0) for 12 days.

aromatic residues. This map allowed an unambiguous determination of the protein fold. A model containing all of the 191 amino acid residues of the MPS molecule was built using an Evans and Sutherland PS300 picture system and the FRODO package of programs (15). The protein coordinates have been partially refined at 2.8-Å resolution to an R factor of 0.33 using Hendrickson–Konnert (16) restrained least-squares refinement.

RESULTS AND DISCUSSION

MPS is an elongated molecule with approximate dimensions of 55 Å by 35 Å by 35 Å (Fig. 1). Four long helices are clearly evident in the map, two of these helices contain ≈20 amino acid residues each, the other two contain ≈30 amino acid residues each. Thus, 54% of the amino acid residues of the protein can be accounted for by these four helices. The helices are arranged in a tightly packed antiparallel helix bundle (17) similar to that found in cytochrome *b*-562 (18), cytochrome *c'* (19), and myohemerythrin (20). Superposition of the MPS α-carbon atoms with the α-carbon atoms of each of the other three molecules shows a remarkable degree of similarity in their four helix bundles. The MPS helix bundle superimposes well with that of each of the other three proteins, with cytochrome *b*-562 giving the smallest rms deviation. Table 2 shows the α-carbon atoms rms deviation in angstroms between MPS and some of the other proteins containing antiparallel four helix bundles. Detailed description of the subtle differences between the MPS bundle and those of the other three proteins must await the completion of the MPS structure refinement.

Although the structure analogy between the MPS four helix bundle and that of other proteins is striking, the MPS connectivity is different. In cytochrome *b*-562, cytochrome *c'*, and myohemerythrin, helix A is connected to B, B to C, and C to D, each with a short distance connection (Fig. 2*a*). This can be referred to as an up-down-up-down connectivity. The MPS molecule connectivity, however, is such that helix A is connected to C, C to B, and B to D, and it is made up of one short distance connection and two long distance

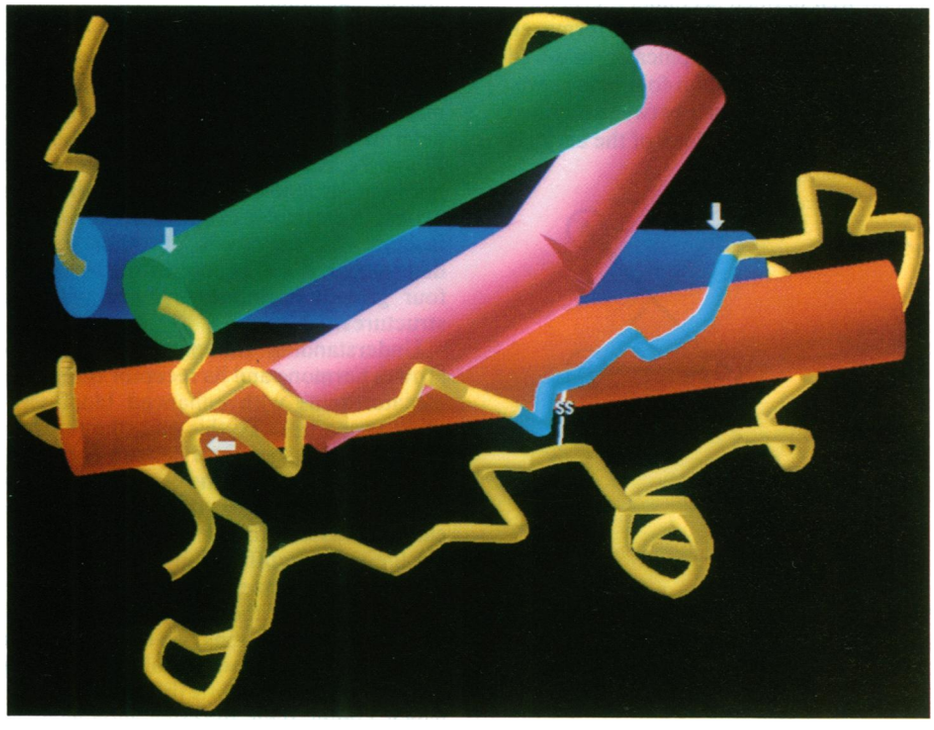

FIG. 1. Computer representation of the MPS polypeptide fold as seen in the 2.8-Å electron density map. Each of the four α-helices is represented by a cylindrical rod, whereas regions of essentially nonhelical polypeptide are shown as a thin tube. The part of the tube involved in the ribbon structure around the cystallographic two-fold axis (see text) is shown in cyan. One of the two disulfide bonds found in the molecule is shown, but the other is hidden in this view. The amino terminus is located at the upper left-hand corner, whereas the carboxyl terminus is at the lower left-hand corner. The arrows point toward the position where introns have been found within the coding sequence.

Table 2. Similarities among MPS, cytochrome *b*-562, cytochrome *c'*, and myohemerythrin

Comparison		Number of residues		Number of equivalenced residues	% of equivalenced residues		rms deviation, Å
Molecule 1	Molecule 2	Molecule 1	Molecule 2		Molecule 1	Molecule 2	
MPS	Cytochrome *b*-562	191	103	69	36.1	67.0	2.8
MPS	Cytochrome *c'*	191	127	66	34.6	52.0	3.2
MPS	Myohemerythrin	191	118	69	36.1	58.5	3.0
Cytochrome *b*-562	Cytochrome *c'*	103	127	63	61.2	49.6	3.1
Cytochrome *b*-562	Myohemerythrin	103	118	76	73.8	64.4	2.5
Cytochrome *c'*	Myohemerythrin	127	118	76	59.8	64.4	2.5

α-Carbon coordinates for cytochrome *b*-562, cytochrome *c'*, and myohemerythrin were obtained from the Protein Data Bank (21). Almost all equivalenced MPS residues are within the MPS four-helix bundle. Structural similarities were calculated using the Rossmann and Argos algorithm (22).

connections (Fig. 2*b*). We believe this represents a subgroup within the antiparallel helix bundle group of proteins that, to our knowledge, has never before been reported and that might be called an up-up-down-down connectivity. There is only one other way to connect α-helices in an antiparallel left twisted four-helix bundle—that is to connect A to B, B to D, and D to C with two short distance connections and one long distance connection (Fig. 2*c*). This connectivity has been found in apoferritin (23). This can be referred to as up-down-down-up connectivity. Thus, with the determination of the MPS structure, at least one example of each subgroup of antiparallel four helix bundles has been reported.

The four helices of MPS are made up of residues 7–34, 75–87, 106–127, and 152–183 (in this paper amino acid residues are numbered starting with methionine at position 1; see Fig. 3 for amino acid sequence). The region from 89 to 96 is also in an α-helical conformation. It can be thought of as an extension of the second helix, except that it is ≈30° away from being collinear with it. The region from 35 to 74 is predominantly loops and turns. One of these turns is a 1.5-turn α-helix (residues 53–58). The region from 128 to 151, on the other hand, has no well-defined secondary structure. Part of this region, residues 137–143 (Fig. 1), forms an antiparallel ribbon with the crystallographically two-fold related molecule. The area around the two-fold axis represents the most significant region of contact between any two molecules in the unit cell. Thus, one can think of the crystals as composed of MPS dimers.

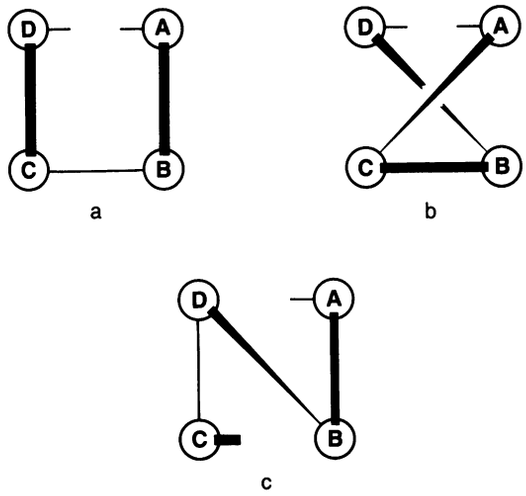

FIG. 2. Schematic diagram of the three possible ways to connect the α-helices in a left-handed antiparallel four-helix bundle. The helices, labeled A–D, are shown end on as circles. Thin lines represent connections below the page, thick lines represent connections above the page, and tapered lines represent connections that go through the page.

Although MPS exists as monomer in solution, it has been reported (24) that human growth hormone can exist as a dimer in solution. It is reasonable to speculate that the human growth hormone dimer is held together in a similar fashion to that seen in the MPS crystal.

The amino acid sequence of MPS contains four cysteine residues at positions 53, 164, 181, and 189. Residues 53 and 164 form one disulfide bridge, whereas residues 181 and 189 form the other. Both Cys-164 and Cys-181 are within the carboxyl-terminal α-helix. The presence of Cys-164 midway up the fourth helix does not cause any distortion of this helix. It is interesting that there are no disulfide bonds in any of the other known antiparallel helix-bundle structures (17).

Fig. 3 shows the alignment of the amino acid sequence of growth hormones from different species and the position of the four α-helices relative to these sequences. Although there is extensive homology among all of these sequences; it is clear that the α-helical regions of these molecules show the most extensive sequence homology. Two-thirds of the identical residues lie within the four α-helices. This suggests that these invariant residues are necessary to maintain the structural integrity of these hormones and are not required for species specificity.

It has been suggested (32) that exons frequently represent modules—polypeptides that fold in a compact fashion. The position of the three introns within the coding sequences of mature human (29) and rat (28) growth hormones has been determined. They are located between amino acids 32 and 33, 71 and 72, and 125 and 126 (Fig. 3). Inspection of the growth hormone structure (Fig. 1) clearly shows that each of the four exons, in particular exon 3 and exon 4, is folded in a compact fashion. As structures of other cell growth factors are determined, it will be interesting to see whether any of these four growth hormone modules will appear in the other structures.

Understanding the nature of the interaction between growth hormone and its receptor is important to the design of somatotropins with enhanced bioeffectiveness as well as other properties. To this end, the information derived from this structural study and genetic engineering techniques can be used to identify the amino acid residues involved in binding of porcine growth hormone to its receptor. Also, data can be measured to extend the resolution of this structure to the limit of the diffraction (2.1 Å).

Due to the extensive degree of amino acid sequence homology (33) of the growth hormones, the placental lactogens, and the prolactins, we believe that we have determined the general three-dimensional fold of these proteins. Yet to be determined are the subtle differences between these proteins, which confer their differing biological activity. When the three-dimensional structures of other members of this hormone family have been determined and the differences between them have been analyzed, considerable insight into understanding the molecular basis of the action of these hormones will be provided.

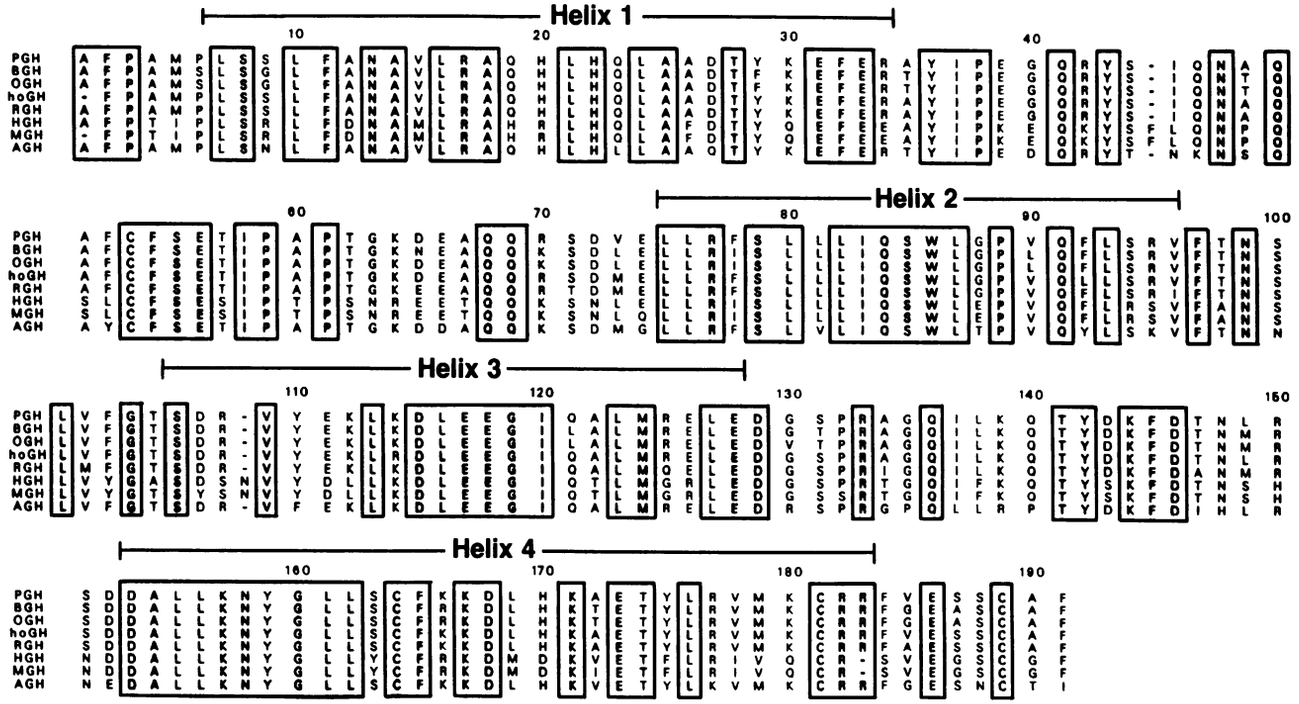

FIG. 3. Amino acid sequences of porcine growth hormone (PGH; ref. 10), bovine growth hormone (BGH; ref. 25), ovine growth hormone (OGH, ref. 26), horse growth hormone (hoGH; ref. 27), rat growth hormone (RGH, ref. 28), human growth hormone (HGH; ref. 29), monkey growth hormone (MGH; ref. 30), and avian growth hormone (AGH; ref. 31). Boxed amino acid residues are identical in all eight sequences. The positions of the MPS four helices are indicated above the amino acid sequences. Residue numbers are those used for the MPS molecule throughout this paper.

We are grateful to Dr. David Eisenberg for the use of his x-ray facility and to Ms. Nancy Leimgruber for valuable assistance.

1. Catt, K. J., Moffat, B. & Naill, H. D. (1967) *Science* **157,** 321.
2. Sherwood, L. M. (1967) *Proc. Natl. Acad. Sci. USA* **58,** 2307–2314.
3. Bewley, T. A. & Li, C. H. (1971) *Experientia* **27,** 1368–1371.
4. Bewley, T. A. & Li, C. H. (1984) *Arch. Biochem. Biophys.* **233,** 219–227.
5. Wilhelmi, A. E., Fishman, J. B. & Russell, J. A. (1948) *J. Biol. Chem.* **176,** 735–745.
6. Lewis, U. J. & Brink, N. G. (1958) *J. Am. Chem. Soc.* **80,** 4429–4430.
7. Hunt, R. E., Moffat, K. & Golde, D. W. (1981) *J. Biol. Chem.* **256,** 7042–7045.
8. Bell, J. A., Moffat, K., Vonderhaar, B. K. & Golde, D. W. (1985) *J. Biol. Chem.* **260,** 8520–8525.
9. Abdel-Meguid, S. S., Smith, W. W., Violand, B. N. & Bentle, L. A. (1986) *J. Mol. Biol.* **192,** 159–160.
10. Seeburg, P. H., Sias, S., Adelman, J., de Boer, H. A., Hayflick, J., Jhurani, P., Goeddel, D. V. & Heyneker, H. L. (1983) *DNA* **2,** 37–45.
11. Xuong, N. H., Freer, S. T., Hamlin, R., Neilsen, C. & Vernon, W. (1978) *Acta Crystallogr. Sect. A* **34,** 289–296.
12. North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968) *Acta Crystallogr. Sect. A* **24,** 351–359.
13. Rossmann, M. G. (1976) *Acta Crystallogr. Sect. A* **32,** 774–777.
14. Wang, B. C. (1985) *Methods Enzymol.* **115,** 90–112.
15. Jones, A. T. (1978) *J. Appl. Crystallogr.* **11,** 268–272.
16. Hendrickson, W. A. & Konnert, J. H. (1980) in *Computing in Crystallography*, eds. Diamond, R., Ramasehan, S. & Venkatesan, K. (Indian Acad. Sci. for Internatl. Union of Crystallography), pp. 1301–1323.
17. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
18. Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D. & Mathews, F. S. (1981) *J. Mol. Biol.* **148,** 427–448.
19. Weber, P. C., Bartsch, R. G., Cusanovich, M. A., Hamlin, R. C., Howard, A., Jordan, S. R., Kamen, M. D., Meyer, T. E., Weatherford, D. W., Xuong, N. H. & Salemme, F. R. (1980) *Nature (London)* **286,** 302–304.
20. Hendrickson, W. A. & Ward, K. B. (1977) *J. Biol. Chem.* **252,** 3012–3018.
21. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
22. Rossmann, M. G. & Argos, P. (1976) *J. Mol. Biol.* **105,** 75–95.
23. Ford, G. C., Harrison, P. M., Rice, D. W., Smith, J. M. A., Treffry, A., White, J. L. & Yariv, J. (1984) *Philos. Trans. R. Soc. London* **B304,** 551–565.
24. Baxter, R. C. & Turtle, J. R. (1985) in *The Biochemistry of the Polypeptide Hormones*, eds. Wallis, M., Howell, S. L. & Taylor, K. W. (Wiley, Chichester, U.K.), pp. 410–433.
25. Graf, L. & Li, C. H. (1974) *Biochem. Biophys. Res. Commun.* **56,** 168–176.
26. Li, C. H., Dixon, J. S., Gordon, D. & Knorr, J. (1972) *Int. J. Peptide Protein Res.* **4,** 151–153.
27. Zakin, M. M., Poskus, E., Langton, A. A., Ferrara, P., Santome, J. A., Dellacha, J. M. & Paladini, A. C. (1976) *Int. J. Pept. Protein Res.* **8,** 435–444.
28. Page, G. S., Smith, S. & Goodman, H. M. (1981) *Nucleic Acids Res.* **9,** 2087–2104.
29. DeNoto, F. M., Moore, D. D. & Goodman, H. M. (1981) *Nucleic Acids Res.* **9,** 3719–3730.
30. Li, C. H., Chung, D., Lahm, H.-W. & Stein, S. (1986) *Arch. Biochem. Biophys.* **245,** 287–291.
31. Souza, L. M., Boone, T. C., Murdock, D., Langley, K., Wypych, J., Fenton, D., Johnson, S., Lai, P. H. & Everett, R. (1984) *J. Exp. Zool.* **232,** 465–473.
32. Go, M. (1981) *Nature (London)* **291,** 90–92.
33. Dayhoff, M. O., ed. (1976) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 2, pp. 120–121.