Crystal Structure of α_1: Implications for Protein Design

CHRISTOPHER P. HILL, DANIEL H. ANDERSON, LAURA WESSON, WILLIAM F. DEGRADO, DAVID EISENBERG*

X-ray diffraction shows the structure of a synthetic protein model, formed from noncovalent self-association of a 12-residue peptide and of sulfate ions at low pH. This peptide is a fragment of a 16-residue polypeptide that was designed to form an amphiphilic α helix with a ridge of Leu residues along one helical face. By interdigitation of the leucines of four such helices, the design called for self-association into a four–α-helical bundle. The crystal structure (2.7 angstrom resolution; *R* factor = 0.215) reveals a structure more complex than the design, with both a tetramer and a hexamer. The α-helical tetramer with leucine interior has more oblique crossing angles than most four–α-helical bundles; the hexamer has a globular hydrophobic core of 12 leucine residues and three associated sulfate ions. Computational analysis suggests that the hexameric association is tighter than the tetrameric one. The consistency of the structure with the design is discussed, as well as the divergence.

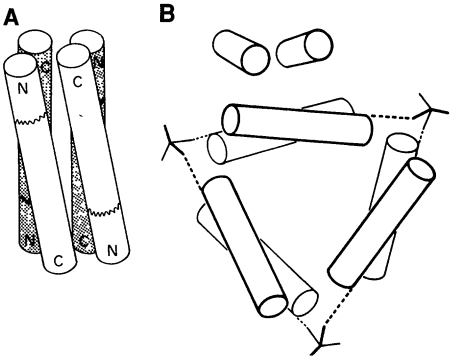

Fig. 1. (**A**) A schematic representation of a four–α-helix bundle, as in the original design of α_1A (*4*). The amino and carboxyl termini are designated by N and C, respectively. The serrated cuts show the helices shortened by 25%, as in α_1. (**B**) The packing of helices in α_1, shown schematically. All helices have identical structures and environments. An α_1 hexamer is shown below, built from two layers of trimers, hydrogen-bonded to SO_4^{2-} ions (represented as jacks) at the amino-termini of two helices. At the top is a tetramer, formed from two helices of the bottom hexamer and from two helices from another hexamer, the rest of which is not shown.

ONE APPROACH TO THE PROTEIN-folding problem (*1*) is protein design (*2*). An amino acid sequence can be designed to fold into a specific three-dimensional (3-D) structure; the protein is chemically synthesized or genetically expressed; and the purified protein is subjected to structural analysis to learn if the design has been achieved, or if not, to learn how the design principles need modification.

One aim of the present project is to design a four–α-helical bundle protein (Fig. 1A) (*3*). It started with the design and characterization of a 16-residue synthetic peptide engineered to self-associate into α-helical tetramers (*4–7*). Two 16-residue peptides, α_1A and α_1B, were designed based on geometric criteria and shown to form helical tetramers with considerable thermodynamical stability (*4, 5*). Ultimately, the helices were connected by three flexible loops (*5*) giving a single-chain protein (*6*). The 16-residue peptides failed to form x-ray quality crystals. However, in the course of this work, a 12-residue fragment, α_1 (Acetyl-Glu-Leu-Leu-Lys-Lys-Leu-Leu-Glu-Glu-Leu-Lys-Gly-COOH), was isolated as a by-product of the synthesis of α_1A, and this shortened peptide both crystallized and self-associated into helical aggregates in solution. At pH 7 in aqueous solution, the circular dichroism spectrum of α_1 shows a fourth-order dependence on peptide concentration (*4*), although the association was thermodynamically rather weak and showed a midpoint at ~5 mg/ml peptide concentration. This peptide was crystallized at pH 3.24 from 80% saturated ammonium sulfate. These conditions considerably enhance the self-association of α_1 (*8*), although the aggregation state has not yet been definitively determined. Herein we report the crystal structure of α_1 at 2.7 Å resolution.

The structure of α_1 is of interest as a model for the association of amphiphilic segments in proteins and also as a measure of the specificity of the design of the 16-residue α_1A and α_1B peptides. A priori, it appears unlikely that α_1 would self-associate into the same helical bundle structure designed for the 16-residue peptide (Fig. 1), since removal of the four residues at the amino terminus would result in staggered ends with a high solvent accessibility of apolar residues. However, a re-registration of the apolar Leu residues at the helical interfaces might result in a shortened bundle. Without structural studies it was impossible to know if α_1 forms such a shortened bundle, some other helical aggregate (*9*), or some unprecedented structure such as the hexamer of the α_1 crystals.

The structure was determined by isomorphous replacement to 3.0 Å resolution and refined to an *R* factor of 0.215 against 2 $\sigma(F)$ data of resolution 10 to 2.7 Å (Fig. 2). Because of the high symmetry of the space group, $I4_132$, the structure in the α_1 unit cell can be described as 48 monomeric α helices, 24 dimers (of several types), 16 trimers, 12 tetramers having 222 symmetry, or 8 hexamers having 32 symmetry. Which oligomeric state is the "true molecule" of α_1 depends on the relative binding energy of the various interhelical associations. We de-

Table 1. Energetics of dimer, tetramer, and hexamer formation of α_1. Calculations were performed with CHARMM (*27*) yielding the energy-minimized change ΔE and with atomic solvation parameters (*28*) yielding the solvation free energy ΔG_S, an estimate of hydrophobic interaction and other solvation terms. The values are expressed on a per monomer basis. The energy model for α_1 is based on the crystal structure (*29*). Formation of both hexamer and tetramer is ccompanied by small but significant decreases of both ΔE and ΔG_S. The decrease in solvent accessible surface area per helix (ΔArea) in formation of the hexamer and tetramer is small compared to most oligomeric proteins (*14*), where the minimum accessible area decrease is 670 Å^2.

Process	Energy changes (kcal/mol)		Area changes (Å^2)	
	ΔE	ΔG_S	ΔArea	$\Delta Area^*_{ap}$
Dimer from monomers $2\alpha \rightarrow \alpha_2$	−0.9	−2.4	140	140
Hexamer from monomers $6\alpha \rightarrow [\alpha_2]_2$	−28.6	−6.3	745	540
Tetramer from monomers $4\alpha \rightarrow [\alpha_2]_2$	−16.8	−4.3	546	417

*Change in apolar surface area (*28*). All areas computed with ACCESS (*30*).

C. P. Hill, D. H. Anderson, L. Wesson, D. Eisenberg, Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90024–1569.
W. F. DeGrado, E. I. du Pont de Nemours and Co., Central Research and Development, Experimental Station, Wilmington, DE 19880–0328.

*To whom correspondence should be addressed.

scribe first the structure of the monomer and the "antiparallel dimer"; then the hexamer, which is a protein molecule–like trimer of antiparallel dimers; and finally the tetramer, an association of two antiparallel dimers, each part of neighboring hexamers. The relation of these units (Fig. 1B), in comparison with Fig. 1A, shows that the tetramer more closely resembles the design than does any of the other oligomeric groupings of helices. Finally, we suggest on energetic grounds that the hexamer is probably a more stable aggregate of α_1 helices than is the tetramer.

The amino-terminal portion of the monomer is a standard α helix but the carboxyl-terminus is extended (Fig. 2). Residues 10 and 11 lie in the β region of the Ramachandran plot, and the last residue, Gly^{12}, is not well defined but best fits the density in the left-handed helical conformation. Two α_1 monomers forming a dimer lie approximately antiparallel to one another, with four Leu side chains (2, 3, 6, and 10) abutting the equivalent twofold related residues of the opposing helix (10, 3, 6, and 2) (Fig. 3, A and B), reminiscent of interactions in apolar peptides [(*10*), and references within]. The crossing angle for the two helices of −32° is just within the limits observed in helical bundles (*3*) but their separation of 12.9 Å is unusually large: antiparallel helical pairs generally show interhelical distances of 9 to 12 Å (*3, 11*). In brief, although the two monomers of this "antiparallel dimer" enjoy hydrophobic interaction with each other, the apolar groups are not interdigitated.

The α_1 hexamer is formed by three antiparallel dimers around the threefold axis of symmetry (Fig. 3, C, D, E, and F). Each of the monomers contributes Leu^3 and Leu^7 to a hydrophobic interaction of 12 Leu side chains at the center of the hexamer. Two triplets of Leu^3 interdigitate around the threefold axis at the center of the hexamer. Above and below these triplets along the threefold are two triplets of Leu^7 side chains. A second type of interaction within the hexamer is that of each of the three sulfate ions with two α-helices. Each sulfate sits at the amino-terminus of one helix from the upper ring of three helices of the hexamer (Fig. 3, A, D, and F) and the amino-terminus of one helix from the lower ring. From both of these helices, each sulfate accepts the amide N–H hydrogen bond from residue 2. Thus the hexamer appears "protein molecule–like" in that it has a globular domain, with a 12-residue hydrophobic core surrounded by charged and polar groups in contact with solvent.

Some of the apolar functional groups of the antiparallel dimer are not wholly buried in the hexamer; groups remaining on the surface include portions of Leu^2, Leu^6, and Leu^{10} and some methylene groups of Lys^5, Glu^9, and Lys^{11}, all of which interact with groups from a symmetry-related antiparallel dimer of a neighboring hexamer. Two such interacting antiparallel dimers constitute the tetramer (Fig. 3G). The interhelical crossing angle between neighboring helices on adjacent antiparallel dimers is more open (41°) than in many four-helix bundles (often near 20°) and approaches the wide crossing angle (about 65°) in tetramers of the amphiphilic α-helical peptide, melittin (*12*). The helices are separated by 11 Å. Also linking dimers on adjacent hexamers are the sulfate ions: two sulfate oxygens each accept a hydrogen bond from the side chains of Glu^9 residues. Presumably the carboxylate function of Glu^9 is protonated at the low pH of the crystals.

Because both hexamer and tetramer appear to have some similarity to natural proteins, we need to ask which of these units is more tightly associated. This question is addressed in Table 1 by use of potential energy calculations, atomic solvation parameters, changes in total solvent accessible surface area, and change in apolar area. By all four criteria, the hexamer is the most stable aggregate within the crystal, as also indicated by the compactness (*13*). The tetramer is apparently less stable, but to the extent that both hexameric and tetrameric associations are of similar stability, the α_1 crystal is more akin to a molecular crystal such as ice or diamond than to a protein crystal, in which intersubunit associations are much tighter than intermolecular associations.

Despite its shorter chain length, α_1 shows several similarities to the original design (*4*): (i) the monomer is largely helical and amphiphilic; (ii) there is association of helices, mainly through hydrophobic interaction of Leu side chains; and (iii) there is an antiparallel helical pair. However, the structure also displays distinct differences from the design:

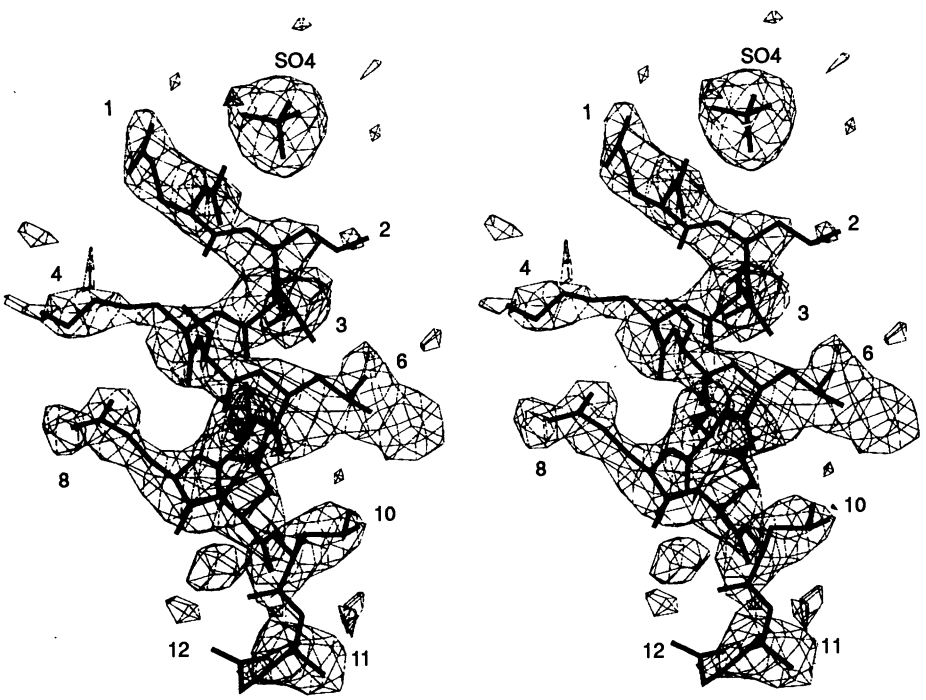

Fig. 2. A stereoview of the refined atomic model of α_1 shown in the multiple isomorphous replacement–solvent-flattened electron density. Crystals were grown in 20-μl hanging drops by a modification (*15*) of the procedure described earlier (*4*). Phases were determined by isomorphous replacement (*16–22*). The two-derivative electron-density map to 3.0 Å resolution (mean figure of merit of 0.64) showed a single, largely α-helical peptide in the asymmetric unit. Solvent flattening (*23*) marginally clarified the map, and a model was built (*24*) containing all α_1 atoms except the side chains of Lys^5 and Lys^{11}, which were not visible. This model explains all of the electron density except for a large spherical density on a crystallographic twofold axis near the amino termini of two helices. Such positions are often occupied by anions (*25*), and this density was interpreted as a sulfate (SO_4^{2-}) that forms hydrogen bonds with the main-chain amides of residue 2 from both helices. This interpretation was confirmed by 3.0 Å difference maps from crystals soaked for two days in a synthetic mother liquor in which the $(NH_4)_2SO_4$ is replaced by $(NH_4)_2SeO_4$. These maps had by far their largest peaks at the sulfate site when either MIR or model phases were used. Restrained least squares refinement (*26*) of positional parameters gave an *R*-factor of 0.21 for 2 σ (*F*) data of greater than 10.0 Å resolution, with the root-mean-square deviation of covalent bond lengths from their ideal target values being 0.013 Å. Temperature factors were fixed at 25 $Å^2$. The *R*-factor against all data of resolution 10 to 2.7 Å is 0.255, and the final $F_o - F_c$ map is reasonably flat.

(i) there is a hexamer, which has a 3-D hydrophobic core rather than the elongated surface of hydrophobic interactions expected for an α helical bundle; (ii) the Leu side chains of neighboring antiparallel helices abut rather than interdigitate; (iii) in the tetramer the dimer pairs cross at a wider angle than do the classic four–α-helical bundles (*3*); and (iv) sulfate ions are incorporated into the molecular structure.

What factors prevent α_1 from forming a classic antiparallel four–α-helical bundle in the crystal? The major factor is that the 12-residue helix of the structure is significantly shorter than the 16-residue helix of the design. The full 16-residue helix has a longer hydrophobic stripe of 6 Leu residues that apparently cannot be accommodated in the 3-D hydrophobic core of the α_1 hexamer. However, if this longer hydrophobic stripe were buried in a four–α-helical bundle, about 600 $Å^2$ of apolar surface area per helix (or about 670 $Å^2$ of total surface area) would be covered, perhaps yielding a stable tetramer (*14*). In contrast, an α_1 monomer has a total accessible surface area of only ~1650 $Å^2$, so that it has to cover between one third and one half of its total surface area to produce a stable aggregate, which it does by forming the novel 12 Leu hydrophobic core of the hexamer and tetramer.

The structure of α_1 is consistent with the broad outlines of the design: α helices associate by hydrophobic interaction of Leu residues. However, the precise mode of association, involving unusual forms of both a hexamer and a tetramer, is unexpected. This surprise may mean that in de novo design it is easier to arrive at a sequence that folds in a protein-like manner than at a particular fold, especially in the multicomponent biological milieux, where ligands such as sulfate and protons can profoundly affect structure.

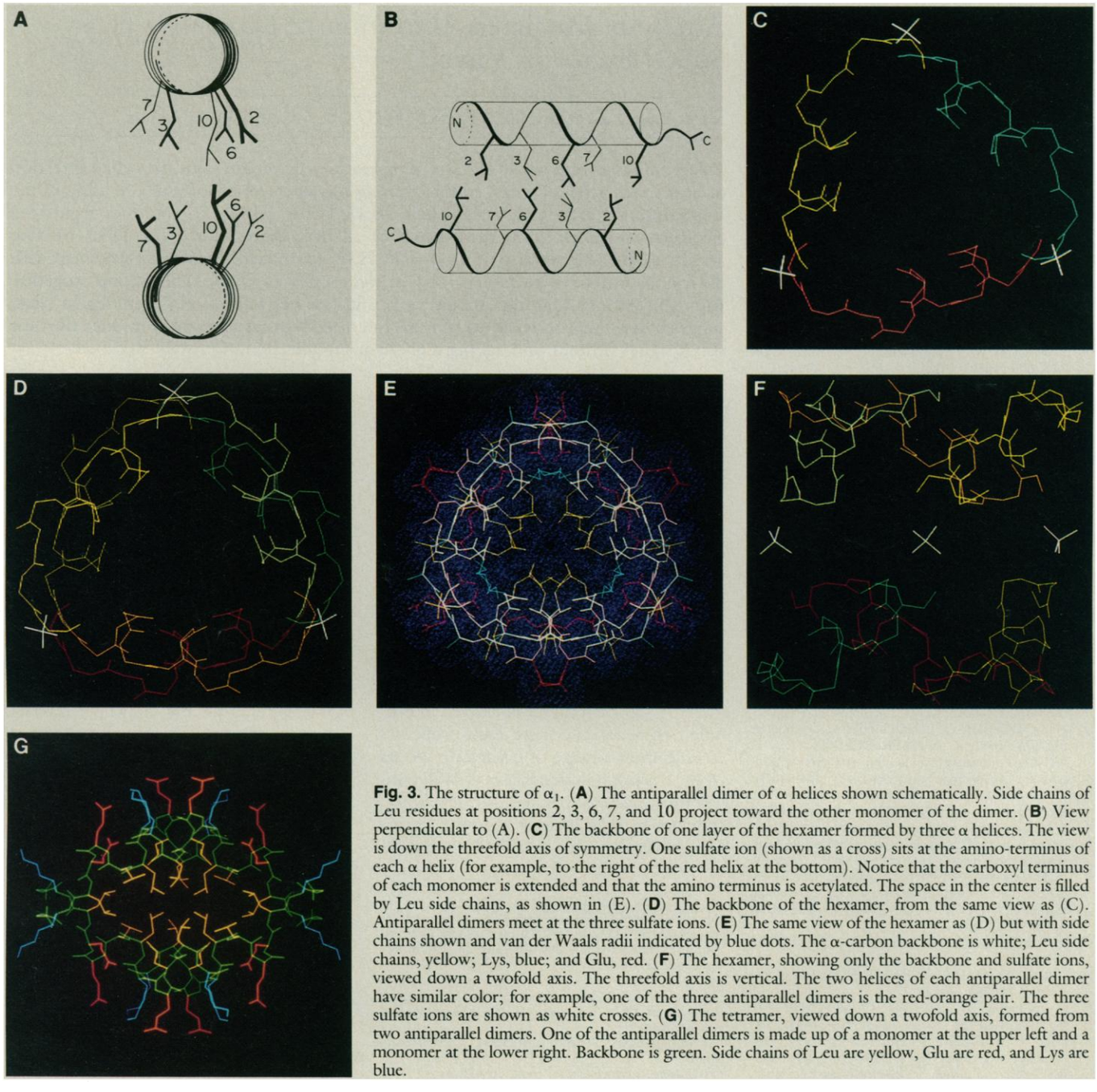

Fig. 3. The structure of α_1. (**A**) The antiparallel dimer of α helices shown schematically. Side chains of Leu residues at positions 2, 3, 6, 7, and 10 project toward the other monomer of the dimer. (**B**) View perpendicular to (A). (**C**) The backbone of one layer of the hexamer formed by three α helices. The view is down the threefold axis of symmetry. One sulfate ion (shown as a cross) sits at the amino-terminus of each α helix (for example, to the right of the red helix at the bottom). Notice that the carboxyl terminus of each monomer is extended and that the amino terminus is acetylated. The space in the center is filled by Leu side chains, as shown in (E). (**D**) The backbone of the hexamer, from the same view as (C). Antiparallel dimers meet at the three sulfate ions. (**E**) The same view of the hexamer as (D) but with side chains shown and van der Waals radii indicated by blue dots. The α-carbon backbone is white; Leu side chains, yellow; Lys, blue; and Glu, red. (**F**) The hexamer, showing only the backbone and sulfate ions, viewed down a twofold axis. The threefold axis is vertical. The two helices of each antiparallel dimer have similar color; for example, one of the three antiparallel dimers is the red-orange pair. The three sulfate ions are shown as white crosses. (**G**) The tetramer, viewed down a twofold axis, formed from two antiparallel dimers. One of the antiparallel dimers is made up of a monomer at the upper left and a monomer at the lower right. Backbone is green. Side chains of Leu are yellow, Glu are red, and Lys are blue.

REFERENCES AND NOTES

1. See *Prediction of Protein Structure and the Principles of Protein Conformation*, G. Fasman, Ed. (Plenum, New York, 1989); T. E. Creighton, *Proteins* (Freeman, New York, 1983).
2. W. F. DeGrado, *Adv. Prot. Chem.* **39**, 51 (1989); S. Y. M. Lau *et al.*, *J. Biol. Chem.* **259**, 13253; J. S. Richardson and D. C. Richardson, in *Protein Engineering*, D. L. Oxender and C. F. Fox, Eds. (Liss, New York, 1987), pp. 149–163; W. F. DeGrado, Z. R. Wasserman, J. D. Lear, *Science* **243**, 122 (1989); J. S. Richardson and D. C. Richardson, *Trends Biochem. Sci.* **14**, 2041 (1989).
3. P. C. Weber and F. R. Salemme, *Nature* **287**, 82 (1980); J. S. Richardson, *Adv. Protein. Chem.* **34**, 167 (1981); S. R. Presnell and F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6592 (1989).
4. D. Eisenberg *et al.*, *Proteins Struct. Funct. Genet.* **1**, 16 (1986).
5. S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751 (1987).
6. L. Regan and W. F. DeGrado, *Science* **241**, 976 (1988).
7. W. F. DeGrado and J. D. Lear, *J. Am. Chem. Soc.* **107**, 7684 (1985).
8. W. F. DeGrado, unpublished observations.
9. A. G. Murzin and A. V. Finkelstein, *J. Mol. Biol.* **204**, 749 (1988).
10. I. L. Karle, J. Flippen-Anderson, K. Uma, P. Balaram, *Biochemistry* **28**, 6696 (1989).
11. C. Chothia *et al.*, *J. Mol. Biol.* **145**, 215 (1981); A. M. Lesk and C. Chothia, *ibid.* **136**, 225 (1980).
12. T. C. Terwilliger and D. Eisenberg, *J. Biol. Chem.* **257**, 6010 (1982); *ibid.*, p. 6016.
13. The compactness Z, the ratio of the surface area of a protein to the surface area of a sphere of equal volume, was defined by M. H. Zehfus and G. D. Rose [*Biochemistry*, **25**, 5759 (1986)] and was found to be 1.64 ± 0.08 for 20 globular proteins. For the α_1 hexamer, $Z = 1.63$, and for the tetramer, $Z = 1.68$, as determined from the energy-minimized model.
14. S. Miller, A. M. Lesk, J. Janin, C. Chothia, *Nature* **328**, 834 (1987); J. Janin, S. Miller, C. Chothia, *J. Mol. Biol.* **204**, 155 (1988).
15. The reservoir solution was 80% saturated (at 23°C) $(NH_4)_2SO_4$, 0.05 M sodium citrate buffer, pH 3.24, 0.005 mg/ml NaN_3. Lyophilized peptide made up to a concentration of 10 mg/ml was mixed 1:1 with the reservoir solution to form the drop. Crystals grew to full size in 1 month at 30°C and were harvested in a synthetic mother liquor of 85% saturated $(NH_4)_2SO_4$, 75 mM sodium acetate at pH 3.9. The space group is $I4_132$ with $a = 62.3$ Å and one helix in the asymmetric unit.
16. From 75 heavy-atom–containing solutions, heavy-atom derivatives were screened by the sinking crystal method (*17*). Data were collected from some 18 crystals on a MSC Rigaku AFC 5R. Corrections were made for decay and absorption (*18*). A $K_2Pt(NO_2)_4$ derivative was located on the threefold symmetry axis by matching its isomorphous and anomalous difference Patterson functions with peaks from an E map, which was generated from the isomorphous differences by the direct methods program MITHRIL (*19*). This model was refined by the Fhle method (*20*) with REFINE2 (*21*). A $Bi(NO_3)_3$ derivative was located in a general position from a difference Fourier map with phases from the Pt derivative. The Bi derivative was refined by the origin-removed Patterson correlation method (*22*), and phases based on it returned the correct position of the Pt atom in a difference Fourier map. Both of these heavy atoms are located in chemically reasonable places on the protein surface: the Bi site is near the carboxyl terminus and Glu^8 side chain, and the Pt ion occupies a crevice near the side chains of Lys^4 and Glu^8.
17. P. B. Sigler and D. M. Blow, *J. Mol. Biol.* **14**, 640 (1965).
18. A. C. T. North, D. C. Phillips, F. S. Mathews, *Acta Cryst.* **A24**, 351 (1968).
19. C. J. Gilmore, *J. Appl. Cryst.* **17**, 42 (1984).
20. E. J. Dodson, in *Crystallographic Computing Techniques*, F. R. Ahmed, K. Hüml, B. Sedlacek, Eds. (Munksgaard, Copenhagen, 1976), pp. 259–268.
21. REFINE2 and programs used for phase calculation and atomic refinement were from the CCP4 program suite, Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, England.
22. T. C. Terwilliger and D. Eisenberg, *Acta Cryst.* **A39**, 813 (1983).
23. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
24. T. A. Jones, *ibid.*, p. 157.
25. W. G. J. Hol, *Adv. Biophys.* **19**, 133 (1985).
26. W. A. Hendrickson, *Methods Enzymol.* **115**, 252 (1985).
27. B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983). Version 19 of CHARMM used with parameter file PARMALLH3 and topology file TOPALLH6.
28. D. Eisenberg and A. D. McLachlan, *Nature* **319**, 199 (1986).
29. The energetically minimized model includes explicit hydrogen atoms whose positions were inferred from the x-ray model as well as the sulfate ions and reasonable conformations for the side chains of Lys^5 and Lys^{11}, which could not be located during the crystallographic analysis. Harmonic constraints were applied and the model energy minimized to convergence, which led to a root-mean-square change in atomic positions of about 0.2 Å. To simulate the pH of the crystals, carboxylates are protonated. The dielectric constant is taken as 40.
30. T. J. Richmond and F. M. Richards, *J. Mol. Biol.* **119**, 537 (1978).

31. We thank W. Wilcox, S. M. Eshita, J. Talafous, and D. C. Rees for discussions, J. Young for peptide synthesis, R. Houghton for technical assistance, and NSF and NIH for support. Coordinates and structure factor amplitudes have been deposited in the Brookhaven Data Bank.

19 January 1990; accepted 31 May 1990

Antibody-Mediated Activation of *Drosophila* Heat Shock Factor in Vitro

VINCENZO ZIMARINO,* SUSAN WILSON,† CARL WU‡

Eukaryotic cells respond to elevated temperatures by rapidly activating the expression of heat shock genes. Central to this activation is heat shock–inducible binding of the transcriptional activator, termed heat shock factor (HSF), to common regulatory elements, which are located upstream of all heat shock genes. The DNA binding activity of the inactive form of *Drosophila* HSF was induced in vitro by treatment with polyclonal antibodies to the purified, in vivo–activated factor. This finding, together with observations that high temperature and low pH activate HSF binding in vitro, suggests that the inactive form of HSF can directly recognize and transduce the heat shock signal without undergoing a covalent modification of protein structure.

ORGANISMS RESPOND TO AN ELEVATION in the ambient temperature by the rapid transcription and translation of RNAs encoding heat shock proteins (*1*). The transcriptional response to heat shock in eukaryotes is mediated by a positive control element, the heat shock element (HSE), which is present in multiple copies upstream of all heat shock genes (*2*). A transcriptional activator protein, termed heat shock factor (HSF), binds to HSEs (*3–8*) and activates transcription of heat shock genes in vitro in a binding site–dependent manner (*9–11*). Although the sequence of the HSE has been highly conserved in evolution, there are differences in the properties of HSF from a range of eukaryotic species (*6, 8*). Electrophoretic analyses of HSF purified from *Saccharomyces cerevisiae*, *Drosophila*, and human cells indicate that the factors have molecular sizes of 150, 110, and 83 kD, respectively (*11–13*). In addition, HSF is bound constitutively to the HSE in yeast (*6, 8, 14*), whereas HSF in *Drosophila* and vertebrates is unable to bind to the HSE unless the cells are heat shocked (*3, 5–8*). The relative increase in HSF binding is correlated with the severity of the heat shock stimulus, suggesting that the temperature-dependent induction of HSF binding activity is a critical regulatory switch in the activation of heat shock gene transcription in higher eukaryotes.

Experiments with protein synthesis inhibitors indicate that de novo protein synthesis is not required for the induction, reversal, and reinduction of HSF binding activity in *Drosophila*, *Xenopus*, and human cells (*5, 8, 15*). These studies indicate that HSF binding is likely to be regulated by a posttranslational mechanism in higher eukaryotes. Moreover, the binding activity of human HSF in HeLa cell cytoplasmic extracts is induced in vitro simply by heat treatment, suggesting that the inactive form of human HSF is able to sense the temperature elevation (*16*). Furthermore, low pH (5.8 to 6.4) activates human HSF in vitro (*17*). Using the gel mobility shift assay, we have found that heat (35°C) and low pH (KH_2PO_4 buffer, pH 6.5) also stimulate the DNA binding activity of HSF in unshocked cyto-

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 4C-09, Bethesda, MD 20892.

*Present address: Mikrobiologisk Institut, University of Copenhagen, Oster Farimagsgade 2A, 1353 Copenhagen K, Denmark.

†Present address: Genex Corporation, Gaithersburg, MD 20877.

‡To whom correspondence should be addressed.