Packed protein bilayers in the 0.90 Å resolution structure of a designed alpha helical bundle

GILBERT G. PRIVÉ,¹ DANIEL H. ANDERSON,² LAURA WESSON,² DUILIO CASCIO,²
AND DAVID EISENBERG²

¹Ontario Cancer Institute, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

²UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Molecular Biology Institute at UCLA and the Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095-1570

(RECEIVED December 30, 1998; ACCEPTED March 4, 1999)

Abstract

A 12-residue peptide designed to form an α -helix and self-associate into an antiparallel 4- α -helical bundle yields a 0.9 Å crystal structure revealing unanticipated features. The structure was determined by direct phasing with the “Shake-and-Bake” program, and contains four crystallographically distinct 12-mer peptide molecules plus solvent for a total of 479 atoms. The crystal is formed from nearly ideal α -helices hydrogen bonded head-to-tail into columns, which in turn pack side-by-side into sheets spanning the width of the crystal. Within each sheet, the α -helices run antiparallel and are closely spaced (9–10 Å center-to-center). The sheets are more loosely packed against each other (13–14 Å between helix centers). Each sheet is amphiphilic: apolar leucine side chains project from one face, charged lysine and glutamate side chains from the other face. The sheets are stacked with two polar faces opposing and two apolar faces opposing. The result is a periodic biomaterial composed of packed protein bilayers, with alternating polar and apolar interfaces. All of the 30 water molecules in the unit cell lie in the polar interface or between the stacked termini of helices. A section through the sheet reveals that the helices packed at the apolar interface resemble the four- α -helical bundle of the design, but the helices overhang parts of the adjacent bundles, and the helix crossing angles are less steep than intended (7–11° rather than 18°).

Keywords: atomic resolution; biomaterial; crystal engineering; peptide design; peptide physics

Our goal in protein design is to create sequences of amino acids that fold and aggregate into predetermined three-dimensional structures, and then to determine their crystal structures to better understand the design parameters. The design goal for “Alpha-1” was for the peptide to fold into alpha helices that aggregate into tetrameric bundles (Eisenberg et al., 1986).

The designed sequence of Alpha-1 incorporates the following features: (1) the residues are strong α -helix formers (Leu, Glu, Lys) (Chou & Fasman, 1978); (2) formation of α -helices would cluster the side chains into a solvated polar face, and a nonpolar face that would drive self-association into a hydrophobic-core tetrameric bundle; (3) intramolecular salt bridges between lysine and glutamate side chains at $i, i + 4$ intervals would stabilize a helix; (4) the same lysine and glutamate residues would further promote

antiparallel association of helices into a bundle; (5) to narrow the focus of the experiment to the interfaces rather than to the ends, acetylation of the N-terminus removes that positive charge, and the lysine near the C-terminus could cap the negative end of the helix dipole. The 12-mer peptide in this work was actually the shorter product of a designed 16-mer target molecule (Eisenberg et al., 1986). We also crystallized the 16-mer, but have never observed usable diffraction (Betz et al., 1996).

We have determined three crystal structures of the 12-mer peptide:

Acetyl-Glu-Leu-Leu-Lys-Lys-Leu-Leu-Glu-Glu-Leu-Lys-Gly.

The first structure (Hill et al., 1990) was of the peptide at low pH (about 3.2), which showed an α -helical fold, but aggregation more complex than the design: the crystal contained tetramers and hexamers, both packing leucine side chains into apolar interiors. The hexamers appeared to be stabilized by binding a sulfate ion between glutamatic acids. This is possible only at low pH. Our next goal for this cycle of design-structure-redesign was to determine the structure of Alpha-1 nearer to neutral pH, where stabilization of the protonated hexamer by sulfate would not obscure our view of interhelical interactions. Here we report the result. In the accom-

Reprint requests to: David Eisenberg, UCLA-DOE Molecular Biology Institute, UCLA, Box 951570, Los Angeles, California 90095-1570; e-mail: david@mbi.ucla.edu.

Abbreviations: F_{obs} , F_{calc} , observed and calculated structure factors; MPD, 2-methyl-2,4-pentandiol; ORTEP, Oak Ridge Thermal Ellipsoid Plot; SnB, “Shake-and-Bake” direct phasing program; U_{eq} , equivalent isotropic displacement parameter; U^{ij} , anisotropic displacement parameter; U^{ij} ellipsoid, ellipsoid drawn by ORTEP-like program.

panying publication, Patterson et al. (1999) report the crystal structure of a racemic mixture of the same sequence. The packing of these two crystals broadens the interest from a design project to an examination of what could be termed “biomaterial” formed from sheets of α -helical peptides. The high resolution diffraction from both crystals also allows a detailed view of amino acid behavior.

Discussion

Description of the structure

The Alpha-1 peptide molecules fold into near-ideal α -helices, as designed. The helices interact via parallel and antiparallel interfaces with a buried leucine core, but the assembly in the crystal is more complex than the simple four-helix bundle of the design. Because of the extended network of interhelix contacts in the crystal lattice, a unique tetrameric association unit (a “molecule”) cannot be assigned. We describe the structure in stages, beginning from the basic peptide elements, and ending with the full three-dimensional structure. We refer to the four independent peptide molecules in the unit cell with the chain identifiers of the Protein Data Bank (PDB) file: A (residues numbered 100–112), B (200–212), C (300–312), and D (400–412). In some contexts, the generic sequence number (Glu8, for example) will be used with the chain specified elsewhere. Residue “00” in each chain is the N-terminal acetyl group.

Helix and side-chain geometry

There is no detectable bending or kinking of the α -helices. The ϕ and ψ backbone torsion angles (except for Gly12) cluster within 12 and 9° of (−60, −42) in each of the four α -helices. All four C-terminal glycine carboxylates protrude away from the helix axes and do not participate in the helical hydrogen bonding pattern. The acetyl groups help cap the N-termini of the helices by their hydrogen bonds to the amide N atom of residue Lys4 (Fig. 1). All but two of the 46 nonglycine side-chain χ_1 rotamers cluster within 10° from ideal trans or gauche values (180 or −60°), including the alternate conformations observed in the crystal (see Disorder in the peptide model below). At each residue, most of the χ_1 values are similar among the four molecules, with some differences. Leu6, for example, has χ_1 trans in molecules B and D, but gauche+ in A and C. Farther along the side chains, the dihedral angles become less correlated across the four molecules. In summary, the designed sequence specified helicity and mostly determined χ_1 in the four nonequivalent environments. As designed, the α -helical fold of Alpha-1 points glutamate and lysine side chains into a hydrophilic face opposite a hydrophobic stripe of leucine side chains (Eisenberg et al., 1986).

Columns

The unit cell *X* direction is nearly parallel to the axes of the four α -helices (Figs. 1, 2). Each of the four helices stacks head-to-tail with equivalent helices in adjacent unit cells, related by unit cell translations along *X*. Molecules A, B, C, and D stack analogously (Fig. 1). In each column interface, one direct hydrogen bond, a bridging water molecule, and the N-terminal acetyl space the interface and effectively continue the *i* to *i* + 4 helical hydrogen bonding pattern. Gly12 turns away from the helical path and is not directly involved in the helix stacking interface. Continuous

Fig. 1. Head-to-tail junction between two helices in a helical column and adjacent atoms of 2-methyl-2,4-pentandiol (MPD513). Balls and rods represent model atoms and their bonds: gray for carbon, cyan for nitrogen, and red for oxygen. For clarity, only the main-chain and CB peptide atoms are shown, and the view is restricted to the N-terminal half of the upper helix, the C-terminal half of the lower helix. These helices are related by a unit cell translation along the crystal *X* direction (vertical; see Figs. 2A, 2B). Broken green lines represent hydrogen bonds. Side-chain salt bridges with adjacent helix columns further stabilize this region (not shown). The A junction is the specific example in this figure, but the B, C, and D column interfaces form almost identical contacts. The Lys11 carbonyl oxygen (lower helix) directly hydrogen bonds to the Glu1 backbone nitrogen of the next molecule (upper helix). A bound water molecule links between Leu10 O of the lower helix and N of Leu2 in the upper helix (Water 619 in this A junction example). In all four junctions, at least one MPD hydroxyl is near the backbone carbonyls (MPD513 in two locations at right, in this example) and the hydrophobic atoms of the MPD molecules wedge between the terminal leucines (above and below the MPD, data not shown). When the four interfaces are oriented as in this figure, several more water molecules cluster around the right of the interface (data not shown). The four N-terminal acetyl groups (background of upper helix) cap the Lys4 N atoms in the first turn and form a hydrophobic contact to the lower helix. In all four helices, the C-terminal glycine carboxylates protrude away from the helix axes (top of lower helix, at left). All the figures were produced with XtalView (McRee, 1993); rendering was done with Raster3D (Merritt & Bacon, 1997).

helices would traverse the entire length of the crystals, if the C-terminal glycine were folded in, and another residue were inserted. As a result of this head-to-tail hydrogen-bonding pattern, the 12-residue peptide packs with a repeat length as though it were a 14-mer, incorporating the 7-mer repeat pattern necessary for extending long α -helices through space.

Sheets

The columns of helices along X closely pack against neighboring antiparallel columns to form sheets parallel to the X - Y plane. Columns of A helices, packed between columns of B helices, form distinct A,B and B,A side-to-side contacts in the crystal $\pm Y$ directions (Figs. 2A, 2B). The A and B helices do not register at the same X level, as they would to achieve the initial design (Eisenberg et al., 1986). Rather, each copy of molecule A contacts two copies of molecule B in each of the adjacent columns. The lattice repeat along Y generates two-dimensional sheets of tightly packed A and B helix columns throughout the crystal. Ridges-into-grooves interdigitation minimizes the interhelical distances along Y (Fig. 2C). Antiparallel columns of helices C and D follow a very similar arrangement. The close-packed C,D and D,C interfaces generate a second sheet of helices in a parallel X - Y plane. The packing of the amphiphilic helices in the A/B and C/D sheets results in one face of each sheet composed exclusively of Leu residues, while the opposite face is made up of charged side chains (Figs. 2C, 2D). The packing resembles a brick wall, strengthened by the offset between layers of bricks.

Multilayer packing

The alternate layering of A/B and C/D sheets along the Z direction forms the final three-dimensional network of helices (Figs. 2C, 2D). The leucine surface of an A/B sheet packs against the leucine surface of a C/D sheet, followed by a charged C/D surface against a charged A/B surface. The A/B and C/D sheets

stack in multilayers until cessation of crystal growth. The hydrophobic layer is about 13 Å thick (helix axis to helix axis distances) (Fig. 2C), thicker than packed helices that show side-chain interdigitation (9–11 Å). In the leucine interface there are contacts only between the CD1 and CD2 methyl groups in opposing sheets. The completely nonpolar leucine bilayers insulate the charged layers from each other.

The hydrophilic layer contains a densely charged zipper of direct and solvent mediated salt bridges. A surprising feature in the charge layer is the 2.5 Å contact between the carboxylates of chain A Glu108 and Gly412 in chain D in the next layer, stabilized by a ring of lysine charges. In each of the four peptide molecules, we find only one intrahelical salt bridge resembling the design. In chains A and D, Glutamates 8 are near Lysines 11; in chains B and C, Glutamates 1 are near Lysines 5. All the other salt bridges are intermolecular within and between sheets.

On the hydrophilic side of each sheet, side-chain interdigitation between adjacent columns adds mechanical stabilization to the charge zipper (Fig. 2B). The charged surface of each sheet is relatively flat (Fig. 2D), and therefore alignment of charges closes the zipper between charge layers stacked along Z . Similarly, the leucines within each sheet interdigitate, but the hydrophobic surface of each sheet is flat. The rough and smooth interfaces result in large and small buried surface areas in the lattice (Table 1). The accumulation of all the interaction energies and tight packing results in the high resolution diffraction of the Alpha-1 crystals.

Distribution of solvent components

Solvent molecules pack essentially all the nonpeptide volume of the Alpha-1 crystal (Table 2). The hydrophilic charge layers are hydrated, as expected. A water molecule binds between each stacked pair of helices in the phantom residue space (see Columns section above) (Fig. 1). The solvent complexity and limitations in modeling it will be discussed below in the context of refinement.

Fig. 2 (*facing page*). Packing of the Alpha-1 crystal in its P1 space group. In panels **A** and **C**, the helices are represented by stick figures with CA atoms at the ends of each line segment. Near line segments are thicker than far segments. The C-terminal CA atom is represented by a + and labeled with a chain identifier: A (residues 100–112), B (200–212), C (300–312), and D (400–412). **A**: The stereo view is perpendicular to the charge layer, with an A/B sheet in the foreground and a C/D sheet in the background. The positive X direction is down the page. The helices align within a few degrees of the X axis. The A chain is antiparallel to the B chains on either side; the C and D chains are antiparallel. The left and right C chains are related by one unit cell translation along the Y axis, the two B chains by $(1+x, 1+y, 0)$ translation. The unit cell contains one each of A, B, C, and D chains. The helices in the A/B sheet straddle between helices in the C/D sheets. The helices pack into columns in the positive and negative X directions as shown in Fig. 1. Helix B overlaps much of helix C, but also contacts helix D. Helix A overlaps about half of each of two C and D helices in the underlying C and D columns. An MPD molecule partly covering the end of a helix in one sheet wedges between leucines of the first and second turns (or second and third turns) of a helix in the other sheet. There is no unique definition of a tetrameric “molecule” because no choice of tetramer fully buries the leucine surfaces. Crystallization into continuous peptide bilayer sheets (“biomaterial”) overlaps leucines of adjacent bundles (for example, by stacking more copies of this figure above and below along the $\pm X$ directions). **B**: Sheet of A and B helices in the X - Y plane, showing the close contacts between the side chains (all-atom representation; viewed perpendicular to the charge layer; positive X is down, as in Figs. 1, 2A). The A/B sheet contains antiparallel columns of A (yellow C atoms) and B (gray C atoms) molecules. **C**: Distances and thicknesses. The view is from the bottom of Fig. 2A. The thick line box shows the unit cell as we have defined it. The origin is marked “0,” and projections of the positive Y and Z directions (protruding behind the paper) are indicated. The space between the central A/B sheet and the lower C/D sheet is filled with leucine side chains. The polar layer between the central A/B sheet and the upper C/D sheet is full of lysine and glutamate side chains, plus all the bound water except the four in the helix stacking sites, and a bound chloride ion. The distances (shown in angstroms) label the thin lines that indicate the two helices involved in the measurement. The distances between pairs of helices are taken as the distances of closest approach of skew lines drawn along the helix axes (J. Bowie, pers. comm.). The perpendicular distances between helix sheets are about 12 Å across the leucine layers and about 12.8 Å across the polar layers. The leucine layer is relatively thick because the leucines contact through their CD1 and CD2 atoms, with little “knobs into holes” interdigitation. **D**: The crystal of L-Alpha-1 viewed down the X axis, with single-point perspective. For clarity, the figure shows only backbone CA atoms, leucines 6 and 7, Lys5, and Glu8 of each chain. The alternating polar and apolar interfaces are nearly flat when all atoms are displayed. Helix A is in the upper left corner.

We have modeled a large density peak as a chloride ion. It is bound between atom Lys311 NZ (3.11 Å), and atoms symmetry related to Lys104 NZ (3.12 Å), Lys204 NZ (3.20 Å), Lys405 CD, and CE (3.46 and 3.59 Å), Glu101 OE1 (3.75 Å). The volume of the chloride peak is larger than a well-ordered water.

Ethanolamine buffers the Alpha-1 crystals, and two bound ethanolamine molecules ("residues" 501 and 502) are found to form similar contacts to peptide atoms. Each ethanolamine molecule straddles both carboxylate oxygen atoms of a glutamate (Ethanolamine 501 on Glu301, 502 on Glu209). A lysine NZ atom is

Fig. 2. See caption on facing page.

Table 1. Buried solvent accessible surface areas in the Alpha-1 crystal lattice^a

Chain	End-to-end	Side-to-side	Leu layer	Charge layer	Total contacts	Isolated molecule	Buried (%)
A	73	866	178	346	1,463	1,518	96
B	72	830	215	316	1,433	1,479	97
C	72	857	202	300	1,431	1,531	93
D	77	833	174	365	1,449	1,480	98

^aSurface areas in lattice contacts were calculated with CCP4 (1994) programs AREAIMOL and RESAREA, using a probe radius of 1.4 Å. Only peptide atoms were included in the calculations, and the discretely disordered side chains were restricted to conformation A only. The areas shown (in Å²) are totals for each peptide chain. The “End-to-end” buried surface areas include both termini (as shown in Fig. 1). The “Side-to-side” areas include four contact regions per chain (Fig. 2A plus more symmetry mates). For example, the 866 Å² side-to-side surface for chain A is the sum of 322 and 190 Å² on the left (in Fig. 2A), and 312 and 42 Å² on the right. Similarly, the “Leu layer” and “Charge layer” columns list the surface areas on each chain covered by all the peptide molecules in the opposing leucine and charge layers (Figs. 2C, 2D). The “Total contacts” column contains sums of the contact areas. The “Isolated molecule” column shows the surface areas of the individual Alpha-1 molecules when excised from the crystal. The Alpha-1 molecules in this crystal are a “biomaterial” because of their high percentage burial. If a tetramer of Alpha-1 is dissected from this crystal, its buried accessible surface area is only 375 Å² per monomer, far less than the total area of about 1,450 Å² per monomer buried in the crystal (the two most compact choices of tetramer have the same total buried area, but distributed differently).

adjacent to each ethanolamine oxygen atom (Lys304 for Ethanolamine 501, Lys105 for 502). The hydroxyl hydrogen of each ethanolamine must orient toward the glu side chains. An additional carboxylate oxygen atom contacts each ethanolamine N atom (Glu308 OE2 for 501, Gly112 OXT for 502). The ethanolamine C1 and C2 atoms have long contacts to hydrophobic peptide atoms. The displacement parameters for Ethanolamine 502 are about triple those of 501. The differences seem to be that Ethanolamine 502 is in a more open binding site, and the symmetry packs a destabilizing Gly312 O 3.57 Å away from the Ethanolamine 502 oxygen atom, but a stabilizing Glu109 OE2 3.44 Å from the nitrogen atom of Ethanolamine 501. Both ethanolamines bind between peptide columns, not obviously influenced by any residual helix dipole.

The precipitant used to crystallize Alpha-1 is 2-methyl-2,4-pentanediol (MPD), and four MPD molecules occupy five sites near helix termini and the exposed surfaces of the leucine layers. The MPD hydroxyls all hydrogen bond to peptide backbone atoms (Fig. 1), except that the surrounding leucines bury the hydrogen bond between MPD511 O2 and MPD512 O4. MPD510 is the R isomer, while the other bound MPD molecules are S isomer (atom assignments are discussed under Refinement below). The hydrophobic atoms of MPD510, 511, and 512 wedge between terminal leucines in the stacking regions between pairs of C helices and pairs of D helices. The offset between helix termini in the A/B and C/D layers (Fig. 2A) also inserts these MPD molecules between leucines in the first and second turns of helix A, and the second and third turns of helix B. MPD511 also extends into the terminal junction between B helices. The disordered MPD513 covers the exposed leucine edges at the helix A junctions (Fig. 1) and extends between the leucines in the first and second turns of helix D. Thus the structure reveals how the MPD molecules participate in hydrophobic burial during aggregation of Alpha-1 peptide into crystals.

Comparison of the Alpha-1 structure with the design, and conclusions

Our goal in determining the Alpha-1 structures was to learn more about the parameters of peptide design. Our results have taught us what not to design next time.

In this paper, we have described our second crystal structure of the designed peptide Alpha-1. The peptide folds into helices as designed, but the precipitant drives it into amphiphilic continuous sheets, rather than into the intended discrete four-helix bundles. As the crystal grows, the aggregating Alpha-1 does not fully bury its leucine surface by bundle formation. Hydrophobic burial continues by binding MPD and by addition of more peptide until crystal growth ceases. Similarly, in the previous low-pH Alpha-1 crystal (Hill et al., 1990) the hexameric aggregate leaves an exposed hydrophobic edge that is buried only by continuation of crystal growth. The repetitive sequence of Alpha-1 does not specify a unique registration of adjacent helices in the hydrophobic core.

A “biomaterial” is an aggregate of biological molecules having intermolecular binding energy comparable to typical intramolecular noncovalent binding energies. We distinguish between the Alpha-1 “biomaterial” and more typical protein crystals by the almost total surface involvement of Alpha-1 in intermolecular contacts. Table 1 shows solvent accessible surface areas buried by crystallization. In solution, Alpha-1 aggregates into a tetramer (Eisenberg et al., 1986; Betz et al., 1996), but the concentration dependence suggests that the binding energy is not strong. If the solution tetramer contact areas resemble those found in the tetrameric units of the present crystal, the contact areas (375 Å² per monomer) (Table 1) would be at the small extreme of the observed range for dimerization (368 to 4,746 Å²) (Jones & Thornton, 1996). Taylor et al. (1996) compare the packing of Alpha-1 to “peptide F” and discuss some examples of biological functions of aggregated peptides. The extremely hard and dense crystals of pheromone Er-1 (Anderson et al., 1996) can be viewed as a “biomaterial.” Weiss et al. (1995) propose that disruption of an Er-1 lattice across the cell surface triggers the switch to mating behavior in that organism. Antifreeze proteins (Wen & Laursen, 1992) and “peptitogens” (Schafmeister et al., 1993) could similarly aggregate onto surfaces. We speculate that a periodic self-assembled peptide-based coating for a biological-contact surface could be derived from available designed peptide crystal structures and could achieve very dense surface coverage. For many applications, Alpha-1 is too soluble to be used without modification.

Head-to-tail stacking of α -helices has been previously described. “Peptide F” of Taylor et al. (1996) packs into parallel

Table 2. Summary of the refined atomic model

Statistics of the atomic model refined against X-ray data ^a					
Resolution (Å)		Model r26		Model r26a	
Lower limit	Upper limit	R_{cryst}	GooF	R_{cryst}	GooF
25.7	2.04	0.085	1.535	0.086	1.525
2.04	1.62	0.074	1.062	0.075	1.068
1.62	1.42	0.071	0.983	0.072	0.984
1.42	1.29	0.074	0.927	0.075	0.922
1.29	1.20	0.075	0.876	0.076	0.877
1.20	1.13	0.080	0.896	0.082	0.905
1.13	1.07	0.083	0.883	0.082	0.878
1.07	1.01	0.088	0.848	0.088	0.844
1.01	0.96	0.135	1.153	0.140	1.166
0.96	0.90	0.163	1.030	0.164	1.031
25.7	0.90	0.085	1.038	0.086	1.038

Content of the final model (r26a) ^b			
Molecule	Number	Number of nonhydrogen atoms	"Residue" numbers
Alpha-1 peptide	4	408	<i>n</i> 00– <i>n</i> 12
Ethanolamine	2	8	501, 502
2-Methyl-2,4-pentanediol	4	32	510–513
Chloride	1	1	520
Water	30	30	601–634

Deviations from restraint values ^c		
Restraint	RMS deviation	Sigma
Bond lengths	0.025 (0.018) Å	0.03 Å
Bond angles	(1.9°)	0.06 Å
Chiral volumes	0.065 Å ³	0.2 Å ³
Flat groups	0.241 Å ³	0.3 Å ³
Anti-bump	0.192 Å	0.1 Å
DELU	0.006 Å ²	0.01 Å ²
SIMU	0.026 Å ²	0.073 Å ²
ISOR	0.108 Å ²	0.1 Å ²

^a R_{cryst} is the conventional crystallographic R -factor. Goodness of fit (GooF) far from 1.0 indicates systematic error in the model. The SHELXL X-ray weighting function adjusted GooF to be near 1.0 as a function of intensity, but not as a function of resolution as shown in this table. R_{free} is an indicator of bias (Brünger, 1997), evaluated for 2,368 excluded reflections. We refined against 21,313 reflections up to refinement r26 (R_{free} was 0.105), then repeated refinement r26 using all 23,681 data. In this paper we describe the all-data model r26a.

^bAtoms in multiple locations are counted once. Seven peptide side chains are modeled in two conformations each. Each water molecule is modeled in a single location, despite the need for increased complexity discussed in the text. SHELXL applied a diffuse solvent correction and generated a riding hydrogen model (Sheldrick & Schneider, 1997). The value of n in the last column can be 1, 2, 3, or 4.

^cThe sigma values control the rigidities of the geometric and other restraints imposed on the model by SHELXL (Engh & Huber, 1991; Sheldrick & Schneider, 1997). Most of the numbers in this table were generated by SHELXL; figures in parentheses were determined by WHAT_CHECK (Hooft et al., 1996). SHELXL implicitly restrains the bond angle around an atom by restraining the distance between the two atoms around it (with double the bond-length sigma value). The chiral volume restraint sets the volumes of the tetrahedra around CA atoms. The flat groups in this structure are the peptide bonds and the carboxylates. DELU is a rigid-bond restraint. SIMU restrains toward gradual changes of the U^{ij} parameters for adjacent atoms. ISOR restrains water U^{ij} 's toward isotropy. The RMS deviations are mostly less than 1σ. Correlated displacements of adjacent atoms would result in average coordinates that appear to violate the anti-bump restraint. The U^{ij} ellipsoids for water molecules are far from round because of unmodeled discrete disorder.

sheets, antiparallel to the other sheet of the bilayer. Several helix termini mutually cap, with some mediation by bound water. Karle et al. (1990) describe their stacked helices as parallel zippers because of the extensive interdigitation of side-chain groups. The only water in that crystal binds to a modified lysine side chain. Three direct hydrogen bonds exclude water from the helix stacking region. Similar "natural" extensions are often seen in crystals of decameric B-DNA oligonucleotide duplexes, with "normal" helix parameters for the step spanning molecules of adjacent unit cells, even though there is a break in the backbone of both strands at the step (Dickerson, 1990). During crystallization of B-DNA and these peptide helices, hydrogen bonds and apolar surfaces provide a strong driving force to stack and densely pack these molecules, even without continuous covalent linkage of backbone atoms.

The Alpha-1 stacking interaction described here illustrates a simple way to generate long columns from short α -helical segments in an engineered crystal. Ogihara et al. (1997) proposed a scheme for design of coiled-coils that can then crystallize head-to-tail, by defining "residue-equivalents" in the end-to-end packing. Here we find that the "residue-equivalent" concept can also be applied to the head-to-tail packing of straight α -helices. The repeat unit in the Alpha-1 crystal X direction contains 12 amino acid residues (Gly12 turns away from the helical path). The N-terminal acetyl group and the water molecule coordinated between helices each contribute approximately the rise of another residue. The total repeat length is therefore 14 "residue-equivalents" (the a cell dimension is 20.85 Å for a rise of 1.5 Å per "residue-equivalent"). It would have been more difficult to predict the cell dimension of the 18 residue "peptide F" molecule (Taylor et al., 1996) because the geometry of the C-terminal portion is 3₁₀. It seems to pack with a 19 "residue-equivalent" repeat along Z (the c cell dimension is 29.24 Å). We speculate that the originally designed 16 residue Alpha-1 (Eisenberg et al., 1986) would not densely pack head-to-tail because of its intermediate length. A less dense packing that permits variegated states would result in the poor diffraction of the 16-mer Alpha-1 crystals (Betz et al., 1996). We also speculate that the hydrophobic capping function of the MPD molecules would be less essential if some of the terminal leucines were changed to valines, thus closing the interhelical gaps. Several stacked coiled-coils contain alternating leu and val residues, for example, Ogihara et al. (1997).

The high resolution of the Alpha-1 structure provides some amino acid physics lessons. Similar peptide and solvent disorder is probably present in every protein crystal, whether or not the resolution allows its visualization (Smith et al., 1986). Although still largely unexplored, the availability of "alternate" conformations is just as important to protein function as are the "main" rotamers. In the Alpha-1 structure, the modeled and unmodeled leucine disorder is mostly in the terminal leucines. The central leucines are buried by other leucines, but the terminal leucines are partway covered by extrudible MPD molecules and are more free to torsion.

Materials and methods

Peptide and crystals

Synthesis and purification of Alpha-1 are described in the accompanying paper (Patterson et al., 1999). In summary, Alpha-1 was synthesized on a solid support and purified to homogeneity by reverse phase chromatography. Crystals were grown by the sitting-drop method: 5 μ L of a 20 mg/mL peptide solution in water were

Table 3. Statistics of the X-ray diffraction data^a

Resolution (Å)		Unique reflections	Completeness (%)	Redundancy	R_{sym}	$\langle I/\sigma \rangle$
Lower limit	Upper limit					
25.7	1.94	2,793	100.0	4.9	0.066	16.2
1.94	1.54	2,811	100.0	5.1	0.077	24.3
1.54	1.34	2,793	99.9	4.7	0.090	21.7
1.34	1.22	2,765	99.5	3.6	0.101	16.7
1.22	1.13	2,766	98.4	3.2	0.123	14.0
1.13	1.07	2,584	92.2	2.2	0.115	12.7
1.07	1.01	2,359	84.8	1.7	0.078	11.2
1.01	0.97	2,059	73.5	1.6	0.105	8.5
0.97	0.93	1,740	62.3	1.5	0.151	5.7
0.93	0.90	1,011	36.0	1.3	0.141	4.3
25.7	0.90	23,681	85.9	3.3	0.076	16.3

^aData were processed with HKL programs (Otwinowski & Minor, 1997), with $I/\sigma > 1$. The completeness falls to 50% at about 0.92 Å.

mixed with 5 μL of a reservoir solution (78 mM triethanolamine-HCl pH 8.0, 52 mM ethanolamine-HCl pH 9.75, 91% MPD), then equilibrated with 1 mL of the reservoir solution. Plate crystals grew in about one month.

Structure determination

A single crystal of dimensions $0.5 \times 0.2 \times 0.1$ mm was flash frozen to 100 K. A data set to 0.90 Å resolution (Table 3) was collected on a 30 cm diameter MAR Research detector at beamline X12C at the National Synchrotron Light Source at Brookhaven. The wavelength was set to 0.92 Å. The crystal is triclinic P1, with unit cell dimensions $a = 20.85$ Å, $b = 20.91$ Å, $c = 27.06$ Å, $\alpha = 102.4^\circ$, $\beta = 95.3^\circ$, and $\gamma = 119.6^\circ$. Several orientations of the crystal and the detector were required to cover the triclinic reciprocal lattice. Low resolution reflections were measured with reduced exposure times. A total of 133 images were processed with the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997), resulting in 77,116 observations of 23,681 unique reflections.

Based on molecular replacement calculations with ideal α -helices as search fragments, we expected that our eventual crystal structure would reveal four peptide molecules in the triclinic cell. The

crystal would therefore have a Matthews number of $1.67 \text{ Å}^3/\text{Da}$ and a solvent content of about 26% of volume or about 10% of the number of peptide atoms (Matthews, 1968). Since each dodecapeptide is made up of 102 nonhydrogen atoms, we expected a total of 448 atoms in the model (10% more than 4×102 because of the solvent content).

The structure was solved by the Shake-and-Bake algorithm revised to incorporate E-Fourier refinement (SnB, version 1.5) (Miller et al., 1994). E values were generated from the intensities using the programs BAYES, LEVY, and EVAL (Blessing et al., 1996, 1998). Input parameters for SnB were the default values calculated from the expected 448 atoms (Table 2 in Miller et al., 1994). SnB used the 4,480 largest E values, picked 448 peaks on every cycle, and refined the phases to find the models with the lowest R_{min} values (Miller et al., 1993). More recently than the Alpha-1 structure solution, the authors of SnB have recommended searching for fewer peaks than the expected number of peptide atoms in the final model. The presence of a bound chloride ion was not discovered until later; we did not exploit its contrasting density to increase the phasing power of SnB. The SnB refinement results resemble Figure 3 of Miller et al. (1993): Of the 100 random-atom starting models, 73 gave nonsolutions with R_{min} in the range of 0.427–

Fig. 3. Examples of refined electron density. In each example, the blue wire frame represents the $2F_{\text{obs}} - F_{\text{calc}}$ map (with model r26a phases) (Table 2) contoured at 1σ . The $F_{\text{obs}} - F_{\text{calc}}$ map is contoured at $\pm 0.24 \text{ e/Å}^3$, green for positive, red for negative. The atom color scheme is the same as in Figure 1. **A:** Lys105 (chain A) is well ordered, as is most of the structure. The high resolution breaks much of the density into discrete atoms. **B:** Lys305 (chain C) modeled in two rotamers. The discrete disorder of the peptide is well modeled. The geometry around the CA atom would improve if the backbone atoms were also split. **C:** Leu102 (chain A) provides an example of why the peptide model is not at maximal complexity. The difference density indicates a second low-occupancy conformation. The transformation from the present model to the alternate rotamer would require χ_1 and χ_2 rotations, and a twist of the backbone to move the CB atom into the paper. The increased complexity at this site and elsewhere would minimally improve the model, while worsening the data-to-parameter ratio. **D:** Leu302 (chain C). Two rotamers with nearly opposite χ_2 angles fill the bulged density. The leucines near the termini are partway buried by MPD, not fully buried by other leucines, and alternate conformations are thus possible. **E:** Glu408 (chain D). We built the side chain in two rotamers because a distended U^{ij} ellipsoid for atom OE2 could not mimic its lobed density. The OE1 atom of conformation B (one of the overlapped atoms in this model) is nonpositive definite. **F:** MPD510. Four molecules of 2-methyl-2,4-pentanediol in five sites cover exposed leucine surfaces. MPD510 is the best ordered of the four, but still shows residual density that could indicate a second rotamer of the dimethyl portion (at the top). The electron density around 10 of the 30 bound water molecules also suggests alternate locations, but the data-to-parameter ratio limits the complexity of the model.

Fig. 3. See caption on facing page.

0.450, 22 models made transitions to pseudo-solutions whose final R_{min} values ranged from 0.383–0.410, and five gave unambiguous solutions (deeper transitions) with R_{min} of 0.375. The peaks of the pseudo-solutions cluster into four cylinders, but their positions do not make chemical sense.

The initial model of the peptide molecules was generated by assigning names to the SnB peaks of one R_{min} 0.375 solution. For example, peak 80 became Leu3 CA of molecule B. The 448 atom SnB solution contained 202 of the 208 expected peptide main chain atoms, 110 of the 200 expected side-chain atoms, and some solvent structure. Further atoms were located through rounds of E-Fourier cycling with SnB, leading to the identification of 399 of the 408 expected peptide atoms prior to any crystallographic model refinement.

Refinement

We performed intensity-based refinement of the model with SHELXL-93 and SHELXL-97 (Sheldrick & Schneider, 1997). To monitor bias through R_{free} (Brünger, 1997), 10% of the data was isolated from the “working set” of reflections prior to the start of refinement. The same “test set” reflections were used in SHELXL-93 and SHELXL-97. The initial Alpha-1 model coordinates were unmodified from the SnB results (except for symmetry operations). Atoms not found by SnB were positioned in electron density maps (unweighted Fourier coefficients $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$; refined model phases). Model examination and building were performed with O (Jones & Kjeldgaard, 1993) for the first 18 models, and XtalView (McRee, 1993) for the remainder. We summarize the content of the “final” refined model in Table 2. All the non-hydrogen atoms including solvent components are modeled with anisotropic displacement parameters. Isotropic hydrogen atoms “ride” on their peptide, ethanolamine, and MPD atoms (Sheldrick & Schneider, 1997). The final difference Fourier map has a maximum of $0.84 \text{ e}/\text{\AA}^3$, a minimum of $-0.50 \text{ e}/\text{\AA}^3$, and an RMS level of $0.08 \text{ e}/\text{\AA}^3$, indicating little unmodeled structure. Figure 3 shows examples of electron density. The coordinates and structure factors are available from the Brookhaven PDB (codes 1BYZ and R1BYZSF).

Disorder in the peptide model

All the peptide atoms were initially modeled in single conformations, then alternate rotamers were added starting with the most obvious (Lys211). As the refinement progressed and the R and R_{free} factors decreased, more detail emerged from the noise of the maps, and the model gained complexity until reaching the limitation of the data-to-parameter ratio (the ratio is 4.5 with evaluation of R_{free} , 5 without). In our “final” model, 7 of the 44 side chains are built in two conformations each: Leu203, Lys205, Lys211, Leu302 (Fig. 3D), Lys305 (Fig. 3B), Leu307 and Glu408 (Fig. 3E). The refined occupancies for the second conformations range from 0.455 for Leu302 to 0.200 for Lys205. The $2F_{obs} - F_{calc}$ density is weak for Lys205 conformation B (final model phases; contoured at 1σ), but convincing for Leu307 conformation B (occupancy 0.253). In addition, there are residual peaks indicating very low occupancy second conformations for Leu102 (Fig. 3C), Leu202, Gly312, and Lys411, and second and third conformations of Leu403. The large swing to the alternate position of Lys411 would require a propagating shift of backbone atoms at least as far as Leu410 C and O, and Gly412 N. Addition of the unmodeled very low occupancy

disorder would minimally improve the model and R_{free} , while adding many more refineable parameters.

The solvent model

As the refinement progressed, we filled the solvent volume with model atoms, but the data-to-parameter ratio finally also halted the increase of complexity of the solvent model. We summarize the content of the solvent model in Table 2. We added water molecules to explain density peaks, subject to chemical sensibility. The anisotropic displacement parameters (restrained by SHELXL to be approximately isotropic) model in part the shapes of the water density. The “final” water model does not fully silence the difference map (contoured at $\pm 3\sigma$ or $\pm 0.24 \text{ e}/\text{\AA}^3$). The electron density suggests coordinate splits for 10 of the 30 water sites (Waters 616, 617, 618, 619, 621, 623, 624, 632, 633, and 634). Unmodeled positive residual density between water sites could represent a solvent continuum or a second network (the largest is about 0.3 to $0.4 \text{ e}/\text{\AA}^3$, near Wat629 and Wat631). A complete description of the solvent behavior would be more complex than in the present model. The low contour level needed to see evidence of missing details discouraged us from further complicating the water model.

The Alpha-1 solvent model includes bound ethanolamine and 2-methyl-2,4-pentanediol molecules (Table 2). Their polar atoms were assigned to optimize salt bridges and hydrogen bonding. Also, the N and O atoms of Ethanolamine 502 generated difference density when assigned differently than in the “final” model. The geometries of the ethanolamine and three of the MPD molecules were stable to refinement despite the omission of many restraints during most of the process. Addition of a second location for MPD513 (Fig. 1) left little residual at that site. We were unable to flatten the residual density around MPD512; a full set of restraints stabilizes its geometry. The disorder of MPD512 and 513 is likely correlated to the nearby disordered Leu203 and Leu302. The residual density around MPD510 (Fig. 3F) did not justify further complication of the model.

Error analysis

Error estimates for each refined parameter were calculated by a full matrix refinement using SHELXL with enlarged arrays. The coordinate error (and therefore bond length error) correlates to the displacement parameters. The positional error estimates for most of the peptide atoms cluster between 0.01 and $0.025 \text{ \AA}$ for atoms near U_{eq} of $0.06 \text{ \AA}^2$. The few peptide atoms with U_{eq} around 0.25 – $0.38 \text{ \AA}^2$ are much more poorly determined, with coordinate errors around 0.06 to $0.1 \text{ \AA}$. The estimated errors in the bond lengths derived from the positions cluster around 0.08 – $0.015 \text{ \AA}$, rising to more than $0.02 \text{ \AA}$ for highly displacing atoms.

Six peptide atoms of disordered residues are refined to negative eigenvalues of their U^{ij} matrices. The six atoms are: Lys205, conformation B, atoms CB, CE, and NZ; Lys305, both conformations, atom NZ; Glu408, conformation B, atom OE1. These six nonpositive definite atoms overlap other atoms in their own disordered side chains, but other overlapped atoms in the model are positive definite. The statistical measures of agreement (R , R_{free} , goodness of fit, and residual density) are unchanged when the six nonpositive definite atoms are made isotropic. The U^{ij} matrices are well behaved for all the solvent atoms, although a few MPD513 atoms are nearly flat in U^{ij} ellipsoid representation.

Table 2 summarizes the deviations from the restraint target values (Engh & Huber, 1991). Although the restraints were more loosely enforced in this refinement than the recommended defaults for SHELXL, the average deviations are much less than the σ values, except for the antibump restraint.

Acknowledgments

We thank Russ Miller and Charles Weeks for their SnB help, Robert Sweet for assistance in data collection at Brookhaven beamline X12C, and Drs. William F. DeGrado and Emily Maverick for discussions. This work was supported by the National Science Foundation and the U.S. Department of Energy.

References

- Anderson DH, Weiss MS, Eisenberg D. 1996. A challenging case for protein crystal structure determination: The mating pheromone Er-1 from *Euplotes raikovi*. *Acta Cryst D52*(3):469–480.
- Betz S, Raleigh D, DeGrado W, Lovejoy B, Anderson DH, Ogihara N, Eisenberg D. 1996. Crystallization of a unique state from a molten globular ensemble. *Folding Design* 1(1):57–64.
- Blessing RH, Guo DY, Langs DA. 1996. Statistical expectation value of the Debye-Waller Factor and E(hkl) values for macromolecular crystals. *Acta Cryst D52*:257–266.
- Blessing RH, Guo DY, Langs DA. 1998. Intensity statistics and normalization. In: Fortier S, ed. *Direct methods for solving macromolecular structures*, NATO ASI Series Volume, Ser. C: Mathematical and Physical Sciences, Vol. 507. Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 47–71.
- Brünger AT. 1997. Free R value: Cross-validation in crystallography. *Methods Enzymol* 277:366–396.
- CCP4 (Collaborative Computing Project #4). 1994. The CCP4 suite: Programs for protein crystallography. *Acta Cryst D50*:760–763.
- Chou PY, Fasman GD. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol* 47:45–148.
- Dickerson RE. 1990. What do we really know about B-DNA. In: Sarma RH, Sarma MH, eds. *Structure and methods*, vol. 3: *DNA and RNA*. New York: Adenine Press.
- Eisenberg D, Wilcox W, Eshita SM, Pryciak PM, Ho SP, DeGrado WF. 1986. The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins Struct Funct Genet* 1:16–22.
- Engh RA, Huber R. 1991. Accurate bond length and angle parameters for X-ray protein-structure refinement. *Acta Cryst A47*:392–400.
- Hill CP, Anderson DH, Wesson L, DeGrado WF, Eisenberg D. 1990. Crystal structure of α -1: Implications for protein design. *Science* 249:543–546.
- Hoof RWW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature* 381:272.
- Jones S, Thornton JM. 1996. Principles of protein-protein interactions. *Proc Natl Acad Sci USA* 93:13–20.
- Jones TA, Kjeldgaard M. 1993. *The O computer graphics program*. Sweden: Uppsala University.
- Karle IL, Flippen-Anderson JL, Uma K, Balam P. 1990. Parallel zippers formed by α -helical peptide columns in crystals of Boc-Aib-Glu(OBzl)-Leu-Aib-Ala-Leu-Aib-Ala-Lys(Z)-Aib-Ome. *Proc Natl Acad Sci USA* 87(20):7921–7925.
- Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol* 33:491–497.
- McRee DE. 1993. *Practical protein crystallography*. San Diego, California: Academic Press.
- Merritt E, Bacon DJ. 1997. Raster3D: Photorealistic molecular graphics. *Methods Enzymol* 277:505–524.
- Miller R, DeTitta GT, Jones R, Langs DA, Weeks CM, Hauptman HA. 1993. On the application of the minimal principle to solve unknown structures. *Science* 259:1430–1433.
- Miller R, Gallo SM, Khalak HG, Weeks CM. 1994. SnB: Crystal structure determination via Shake-and-Bake. *J Appl Cryst* 27:613–621.
- Ogihara NL, Weiss MS, DeGrado WF, Eisenberg D. 1997. The crystal structure of the designed trimeric coiled coil coil-V₆L₄: Implications for engineering crystals and supramolecular assemblies. *Protein Sci* 6:80–88.
- Otwinowski Z, Minor W. 1997. Processing of diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Patterson WR, Anderson DH, DeGrado WF, Cascio D, Eisenberg D. 1999. Centrosymmetric bilayers in the 0.75 Å resolution structure of a designed alpha-helical peptide, D,L-Alpha-1. *Protein Sci* 8:1410–1422.
- Schafmeister CE, Miercke LJW, Stroud RM. 1993. Structure at 2.5 Å of a designed peptide that maintains solubility of membrane proteins. *Science* 262:734–738.
- Sheldrick GM, Schneider TR. 1997. SHELXL: High-resolution refinement. *Methods Enzymol* 277:319–343.
- Smith JL, Hendrickson WA, Honzatko RB, Sheriff S. 1986. Structural heterogeneity in protein crystals. *Biochemistry* 25:5018–5027.
- Taylor KS, Lou M-Z, Chin T-M, Yang NC, Garavito RM. 1996. A novel, multilayer structure of a helical peptide. *Protein Sci* 5:414–421.
- Weiss MS, Anderson DH, Raffioni S, Bradshaw RA, Ortenzi C, Luporini P, Eisenberg D. 1995. A cooperative model for receptor recognition and cell adhesion: Evidence from the molecular packing in the 1.6 Å crystal structure of the pheromone Er-1 from the ciliated protozoan *Euplotes raikovi*. *Proc Natl Acad Sci USA* 92:10172–10176.
- Wen D, Laursen RA. 1992. Structure-function relationships in an antifreeze polypeptide. The role of neutral, polar amino acids. *J Biol Chem* 267(20):14102–14108.