Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins

Angela Lombardi*, Christopher M. Summa†, Silvano Geremia‡, Lucio Randaccio‡, Vincenzo Pavone*, and William F. DeGrado†§

*Department of Chemistry, University of Napoli ''Federico II,'' Via Mezzocannone, 4, I-80134 Napoli, Italy; †The Johnson Research Foundation, Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104-6059; and ‡Department of Chemical Science, University of Trieste, Via L. Giorgieri 1, I-34127 Trieste, Italy

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected April 27, 1999.

Contributed by William F. DeGrado, March 23, 2000

***De novo* protein design provides an attractive approach for the construction of models to probe the features required for function of complex metalloproteins. The metal-binding sites of many metalloproteins lie between multiple elements of secondary structure, inviting a retrostructural approach to constructing minimal models of their active sites. The backbone geometries comprising the metal-binding sites of zinc fingers, diiron proteins, and rubredoxins may be described to within approximately 1 Å rms deviation by using a simple geometric model with only six adjustable parameters. These geometric models provide excellent starting points for the design of metalloproteins, as illustrated in the construction of Due Ferro 1 (DF1), a minimal model for the Glu-Xxx-Xxx-His class of dinuclear metalloproteins. This protein was synthesized and structurally characterized as the di-Zn(II) complex by x-ray crystallography, by using data that extend to 2.5 Å. This four-helix bundle protein is comprised of two noncovalently associated helix-loop-helix motifs. The dinuclear center is formed by two bridging Glu and two chelating Glu side chains, as well as two monodentate His ligands. The primary ligands are mostly buried in the protein interior, and their geometries are stabilized by a network of hydrogen bonds to second-shell ligands. In particular, a Tyr residue forms a hydrogen bond to a chelating Glu ligand, similar to a motif found in the diiron-containing R2 subunit of *Escherichia coli* ribonucleotide reductase and the ferritins. DF1 also binds cobalt and iron ions and should provide an attractive model for a variety of diiron proteins that use oxygen for processes including iron storage, radical formation, and hydrocarbon oxidation.**

Proteins use a limited repertoire of metal ion cofactors to help catalyze a multitude of reactions. For example, diiron sites (1–4) mediate reversible oxygen binding in hemerythrins, whereas they function as hydrolytic centers in phosphatases. Structurally similar diiron sites also mediate a number of oxygen-dependent oxidative processes. Ferritins serve as ferroxidases, while other diiron proteins catalyze hydroxylation, epoxidation, and desaturation reactions. Further, a diiron site in *Escherichia coli* ribonucleotide reductase is responsible for the formation of a Tyr radical. How do the structures of these proteins tune the chemical properties of a common diiron center to obtain such a diversity of highly specific catalysts? This question is being addressed through the study of the natural proteins as well as the study of small-molecule diiron complexes (1–4). Although impressive progress has been made on both fronts, these approaches have inherent limitations. The study of large proteins is hampered by their extreme complexity, and it is difficult to synthesize small-molecule models capable of simultaneously binding diiron, oxygen, and various substrates. Recently, we and others have sought a molecular middle ground between these two extremes through the design of small proteins and peptides that self assemble to form complexes with hemes and metal ions (5–8).

Peptide models could have a number of distinct advantages relative to other synthetic models for metalloproteins. They allow the construction of hydrophobic pockets within water-soluble structures, and their structures may be synthesized easily and varied by using highly optimized methods of peptide synthesis. Further, peptide models address not only the issue of how the arrangement of atoms within an active site leads to function but also the very important question of how an amino acid sequence dictates the tertiary structure that supports this active site. Thus, several groups have described the design of minimal mimics of metalloproteins and heme-binding proteins that distill the quintessential elements believed to be responsible for the activities of metalloproteins into model proteins that are simpler and hence more easily understood than natural proteins (5–14). Through a careful characterization of the properties of such minimal metalloproteins, it should be possible to discern the features required for selective recognition of metal ions and for tuning their chemical properties.

The protein-folding problem is a primary challenge encountered in the design of complex metalloproteins. This problem may be circumvented by grafting inorganic cofactor-binding sites into the structures of natural proteins that normally do not bind metal ions. Automated methods have been developed recently for engineering such ion-binding sites (15, 16), and it has been possible to build a number of structural as well as redox-active metal ion-binding sites within several different proteins (17–20). Another successful approach has been to design small flexible peptides that are able to fold around metal sites such as Cu(II)-binding motifs (21–24) and small peptides that assemble into Fe_4S_4 clusters (25, 26).

However, complete control of a cofactor's environment might be best effected through the *de novo* design (27, 28) of proteins whose active sites are defined by the favorable free energy of folding of the polypeptide chain. Unfortunately, initial attempts to design proteins led to structures that formed molten globule-like states with dynamic behavior relative to natural proteins. More recently, it has been possible to design small uniquely folded proteins that incorporate all of the commonly occurring secondary structural and supersecondary structural motifs (29–36). These studies illustrated the delicate interplay of forces that define the uniquely folded structures of proteins; hydrophobicity provides a strong driving force for folding, but designs based on this consideration alone often adopt dynamically averaging

Abbreviations: DF1, Due Ferro 1; PDB, Protein Data Bank.

§To whom reprint requests should be addressed at: 1009B Stellar-Chance Building, University of Pennsylvania, Philadelphia, PA 19104-6059. E-mail: wdegrado@mail.med.upenn.edu.

structures. More subtle energetic features such as side-chain packing, hydrogen bonding, electrostatics, and conformational preferences are also important for the stability and structural specificity of proteins. Much progress has also been made in the *de novo* design of proteins that bind metalloporphyrins (10, 21–24, 37, 38) and Zn(II) (39, 40). The structures of some simple porphyrin peptide complexes have been solved by NMR (5, 41). However, the structures of larger designed proteins with bound cofactors have not been solved, possibly because they have dynamically averaging structures (37, 42). Therefore, methods for the design of uniquely folded cofactor-binding proteins are needed.

Recent computational approaches to *de novo* protein design have allowed the synthesis of novel proteins that fold into conformationally unique structures (36, 43–45). Side-chain repacking algorithms search for combinations of amino acid side chains capable of packing together in efficient low-energy combinations. However, with a few exceptions (36, 46), these methods have been used to redesign natural proteins rather than to engineer novel structures. Further, a major limitation of early side-chain repacking algorithms was their use of rigid backbone geometries, which placed somewhat artificial restrictions on the number of combinations of side chains that may be accommodated within a protein core. More recently, by parameterizing the main-chain conformation, it has become possible to vary systematically the structure of the backbone while simultaneously evaluating various side-chain combinations (45, 47). Also, simple analytical expressions for the main-chain conformation have been used to expedite the design of coiled-coil peptides, including a right-handed coiled coil whose three-dimensional structure had not previously been seen in nature (48). Thus, these methods have provided attractive tools for designing peptides and proteins that test the rules of protein folding. It is now important to extend these automated approaches to the design of functional proteins.

Materials and Methods

Protein Crystallization. The stock protein solution was prepared by dissolving 0.5 mg of protein in 5 μl of DMSO followed by 5 μl of 0.03 M aqueous zinc acetate. The solution was diluted with 65 μl of water and centrifuged. Small prismatic crystals, typically $100 \times 50 \times 50$ μm, were obtained from hanging drops at 4°C from 0.100 M sodium acetate (pH 4.6)/2.0 M ammonium sulfate solution. A single crystal was frozen in liquid nitrogen and used for data collection under a stream of cold nitrogen (at 100 K) by using glycerol as cryoprotectant. Data were collected by using a 345-mm Mar Imaging Plate at the Elettra Synchrotron (Trieste, Italy). Data were reduced by using MOSFLM in the orthorhombic space group $C222_1$ with cell dimensions of $a = 35.97$ Å, $b = 89.01$ Å, $c = 79.66$ Å. A summary of the data processing statistics is given in Table 2.

Protein Design. The initial backbone geometry of DF1 was generated as described previously (49), except that a D_2 symmetry operator was used to generate the structure. The side chains of the liganding Glu residues were placed in favorable rotamers (g^+, g^+) that would allow for formation of the site. Next, the interhelical distances were fine tuned to allow formation of the desired site while maintaining the D_2-symmetrical arrangement of the backbones of the helices, by using a macro written for PSSHOW. The remaining side chains were added as discussed previously (50) and the structure minimized by using the CVFF force field (Molecular Simulations, Waltham, MA). Figures were generated with the program INSIGHT II (Molecular Simulations).

Structure Determination. The V_M values suggested the presence of 24 monomer molecules per unit cell, implying three monomers per asymmetric unit. If the crystal were built up by dimers forming the four-α-helix-bundle motif, the independent unit should contain one dimer and one monomer. The monomer should also be located near a crystallographic 2-fold axis to generate the four-α-helix-bundle motif. The structure was solved by molecular replacement by using the program AMORE (51) using the coordinates of the designed protein as the starting model. The rototranslation solutions were calculated by using the four α-helix bundle model omitting the side-chain atoms and was visually examined with the program O (52). A macro was used for the automatic generation of the solutions and visualization of the relative crystal packing. The solution for the monomer location was found visually when the dimer lying on a crystallographic 2-fold axis generated a superimposed object by space group symmetry. The 161st rototranslation (R_{conv} and correlation coefficient, Cc = 0.454) was selected as a fixed solution for the monomer. The rototranslational solution for the independent dimer, compatible with the complete crystal packing, gave an R_{conv} of 0.515 and a Cc of 0.511. The first rigid body refinement of the three monomers gave an R factor of 0.488. The structure was refined by using REFMAC (53) following a process of manual model building by using O. Throughout the initial refinement during model building, 5% of data set was flagged for calculation of R_{free}. The final refinement gave $R_{conv} = 0.237$ and $R_{free} = 0.304$. The program PROCHECK (54) was used to assess the geometry of the final structure [Protein Data Bank (PDB) no. 1EC5].

Analytical Ultracentrifugation. Equilibrium sedimentation analysis was performed at 25°C with a Beckman XLA ultracentrifuge (Beckman Coulter) at 25,000 and 35,000 rpm using an AN-60TI rotor. The total protein monomeric concentration was 60 μM in 10 mM Mops, pH 6.5, in the presence of 2.0 equivalents $CoCl_2$. The protein concentration for all measurements was determined by the absorption at 280 nm ($\varepsilon = 9{,}530\ M^{-1}\ cm^{-1}$). Equilibrium was assumed when successive radial scans, taken at 2-h intervals at the same velocity, were identical. Data were analyzed as described previously (49).

CD Measurements. CD measurements were obtained by using a Jasco J-700 spectropolarimeter (Jasco, Easton, MD). Urea and Gdn·HCl denaturation studies were carried out manually at 25°C by adding aliquots of the denaturant solution to the peptide solution, and the curve obtained was fit to an equation for dimerization-linked folding (55) by using KALEIDAGRAPH (Synergy Software, Reading, PA).

Metal Complex Preparation. Cobalt was inserted into DF1 as described for the reconstitution of hemerythrin (56, 57). To a solution of *apo*-DF1 (10 mg, $1.49 \cdot 10^{-3}$ mmol) in 1.5 ml of 6 M Gdn·HCl in 50 mM Mops buffer (pH 7.0) was added dropwise 15 μl of a $CoCl_2$ 5 M solution in 50 mM Mops buffer (pH 7.0, 50-fold excess cobalt) over 10 min. The solution was stirred at room temperature for 2 h and then diluted by slow addition (50 μl/min) of 50 mM Mops buffer (pH 7.0) to a final Gdn·HCl concentration was 0.6 M. Excess reagents were removed by filtration through an Amicon membrane (molecular weight cutoff, 3,000 Da). The complex was then diluted with H_2O/10% DMSO, lyophilized, and redissolved in 50 mM Mops buffer, pH 7.0 ($1.42\ 10^{-5}$ M concentration). The UV-visible spectrum was obtained by using a Perkin–Elmer Lambda 7 UV spectrophotometer ($\lambda = 524$ nm, $\varepsilon = 143\ M^{-1}\ cm^{-1}$; $\lambda = 574$ nm, $\varepsilon = 190\ M^{-1}\ cm^{-1}$; $\lambda = 594$ nm, $\varepsilon = 165\ M^{-1}\ cm^{-1}$; $\lambda = 625$ nm, $\varepsilon = 80\ M^{-1}\ cm^{-1}$; extinction coefficients, calculated per DF1 monomer), are typical of five-coordinate Co(II), and show a very close correspondence with the extinction coefficients of bacterioferritin ($\lambda = 520$ nm, $\varepsilon = 126\ M^{-1}\ cm^{-1}$; $\lambda = 555$ nm, $\varepsilon = 155\ M^{-1}$

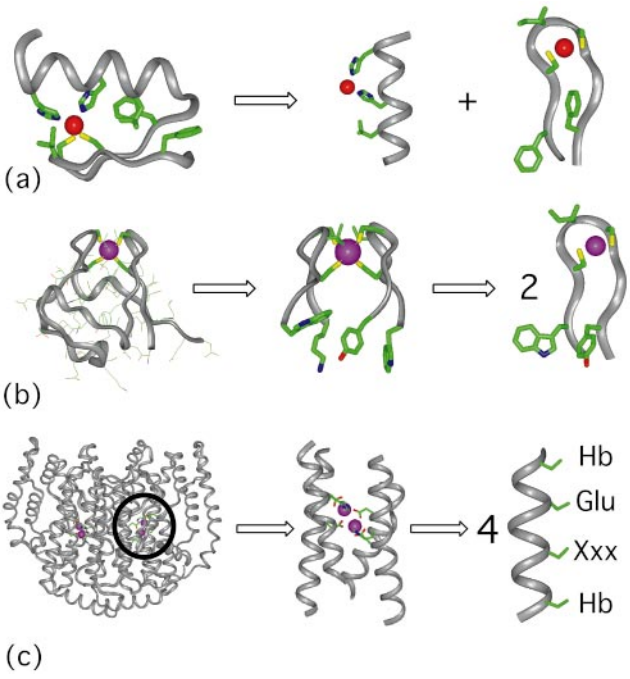

Fig. 1. Retrostructural analysis of metalloproteins. (*a*) A zinc finger from Zif268 residues 33 to 60 [PDB no. 1ZAA (61)]. The structure can be dissected in two separate secondary structure motifs including an idealized α-helix and β-hairpin. Shown are the two His and two Cys side chains critical for metal binding. Also shown are those hydrophobic resides that stabilize the fold as well as define the environment of the metal ion. (*b Left*) The full structure of rubredoxin [PDB no. 1BRF (84)], whose active site can be dissected in two symmetry-related idealized β-hairpins, residues 3–12, and 36–45 (*Center*). The β-hairpin from rubredoxin (*Right*) is remarkably similar to that found in Zif268 (*a Right*) in both its geometry and the placement of the aromatic and aliphatic side chains. (*c Left*) The full structure of Δ^9 ACP desaturase [PDB no. 1AFR, (71)], which is a dimer of two identical catalytic subunits. In contrast to the complex nature of the overall structure, the architecture of the diiron-binding site can be described as a D_2 symmetric four-helix bundle (*Center*). The idealized helix contains hydrophobic residues (Hb) that encapsulate and structurally stabilize the diiron site. Each helix contains a Glu side-chain ligand (shown). The position labeled ''Xxx'' is a His ligand in two of the helices. In the other two helices, this position helps define the environment and access surrounding the diiron center.

cm^{-1}; $\lambda = 600$ nm, $\varepsilon = 107\ M^{-1}\ cm^{-1}$; $\lambda = 625$ nm, $\varepsilon = 75\ M^{-1}\ cm^{-1}$) (58).

Results

Retrostructural Analysis of Metalloproteins. Often the metal ion-binding sites of metalloproteins are formed between two or more elements of secondary structure, a fact that allowed the correct prediction of the fold of the zinc finger before its experimental determination (59). A retrostructural approach can be used to identify geometric relationships between these secondary structural elements. The following section illustrates retrostructural analyses of several proteins with varying degrees of symmetry.

His_2Cys_2 zinc fingers bind Zn(II) at an asymmetric site between an α-helix and an antiparallel β-hairpin (Fig. 1*a*). The long axis of the helix and the average plane of the β-hairpin are nearly parallel to one another, with a spacing of approximately 8.8 Å. The helix donates two His ligands to the site, and the antiparallel β-hairpin donates two Cys ligands. Thus the overall structure of the major elements of secondary structure of the second domain of the zinc finger protein, Zif268 (60, 61), may be described to 0.66 Å rms deviation (Table 1) by considering only an idealized helix and an idealized β-hairpin with an α_r-α_r-α_r-α_l turn. The six adjustable parameters defining the docking of the helix onto the sheet include three translations and three Eulerian rotations.

In other metalloproteins, one often observes symmetric arrangements of these same secondary structure motifs. For example, the 2-Cys antiparallel β-hairpin motif that is found in zinc fingers (Fig. 1*b*) occurs in a tandem 2-fold repeat in the iron-binding sites of the rubredoxins. Indeed, the rubredoxin active site may be described to 0.78 Å rms deviation by using a C_2-symmetric arrangement of the identical idealized β-hairpins (Table 1). This finding has formed the basis for the design of a series of minimal dimeric models for rubredoxin (A.L., F. Nastri, V.P., unpublished work).

The α-helical portions of zinc finger proteins often include one of two metal-binding sequences: His-Xxx-Xxx-His or His-Xxx-Xxx-Xxx-His (62). These helical motifs—or variations in which one of the His residues is converted to Glu or Asp—are frequently observed in the metal-binding sites of other mono- and dinuclear Fe-, Mn-, and Zn-binding proteins. For example, the Lig-Xxx-Xxx-Xxx-Lig (in which Lig is His, Asp, or Glu) motif figures largely in the structures of hemerythrin and myohemerythrin, a class of diiron proteins involved in reversible oxygen binding and transport (63). Indeed, the backbone of myohemerythrin (64) may be described by a D_2-symmetric model to 1.2-Å rms deviation, by using only six adjustable parameters (translations in x, y, and z, rotation of the helix about its own axis, and its tilt and inclination relative to the axis of the bundle; Table 1.)

Deviations from this idealized motif are clearly important for function; for example, in the first helix of myohemerythrin, a liganding group is replaced by a hydrophobic side chain. The omission of a side-chain ligand provides a single free ligation site that binds oxygen in an end-on manner.

Retrostructural Analysis of EXXH Diiron Proteins. A second, functionally diverse class of carboxylate-bridged diiron proteins (2–4, 65) features a helical Glu-Xxx-Xxx-His (EXXH) motif, with two rather than three residues between the liganding side chains. We have conducted a retrostructural analysis (27) of six members of this class, including three ferroxidases [ferritin (66), bacterioferritin (67), and rubrerythrin (68, 69)], ribonucleotide reductase R2 subunit (R2) (70), Δ^9 ACP desaturase (71), and the catalytic subunit of methane monooxygenase (72). Although there is less than 5% sequence identity common to all members of this class, their active sites are housed within a very simple pseudo-222-symmetric four-helix bundle. As in the hemerythrins, the geometry of the active sites of these proteins could be described with a remarkably high degree of accuracy (approximately 1 Å rms deviation) with a D_2-symmetric model and six adjustable parameters. Further, the parameters for the individual proteins are tightly clustered (Table 1), allowing one to define an average primordial 222 symmetric bundle that is representative of the entire class.

Many members of this class of proteins display a highly symmetric arrangement of liganding side chains, particularly in the di-Fe(II) or di-Mn(II) states. The sites generally have four Glu residues (Fig. 2) that project toward the center of the bundle: two bridge both metal ions, whereas the other two carboxylates interact with a single metal ion in a monodentate or bidentate, chelating interaction. The two His residues at position i + 3 relative to the two bridging Glu side chains form additional monodentate ligands. The absence of a His ligand in the other two helices leaves two adjacent coordination sites available for interaction with O_2.

To facilitate the discussion of the structure of diiron proteins, we adopt the heptad nomenclature typically applied to coiled coils, which appears to be valid within approximately ± 7 Å of the diiron site. The liganding Glu and His residues respectively

Table 1. Parameters derived from retrostructural analysis of natural proteins

Protein name	rms deviation (Å)	Atoms superimposed	Symmetry	α, degrees	β, degrees	γ, degrees	Xtrans, Å	Ytrans, Å	Ztrans, Å
Zif268[a]	0.66	22	N/A	−35.8	−18.7	−3.1	−1.8	8.8	5.2
Rubredoxin[b]	0.78	20	C2	21.5	−13.8	12.2	0.8	−5.4	N/A
Bacterioferritin[c]	1.03	48	D2	8.5	13.6	−4.1	4.5	5.6	7.5
Rubrerythrin[d]	0.89	48	D2	10.3	11.2	−3.7	4.2	6.0	7.3
H Ferritin[e]	1.37	48	D2	6.7	8.5	−4.1	4.3	6.3	8.0
Ribonucleotide reductase (R2)[f]	1.11	48	D2	3.9	18.3	−1.4	4.6	5.7	6.9
Methane monooxygenase[g]	0.96	36	D2	0.1	19.9	2.2	4.1	5.4	6.5
Δ9 ACP desaturase[h]	1.29	36	D2	−11.9	14.8	−2.2	4.5	5.7	9.1
Myohemerythrin[i]	1.23	74	D2	4.8	10.0	4.5	5.3	5.2	7.3
Hemerythrin[j]	1.45	74	D2	0.2	9.6	3.7	5.2	5.2	7.2
Hemocyanin[k]	2.09	42	D2	5.5	14.4	−17.1	6.9	6.1	11.2

Retrostructural parameterization of metal sites. Models of proteins based on idealized secondary structure motifs were generated and fit to experimental structures of metalloproteins. The best-fit Cα rms deviation and the parameters used to generate the tertiary structure fold are shown above. To create the model helical bundles, an ideal helix ($\phi = -65.0$, $\psi = -40.0$) was first placed in a Cartesian coordinate system with its axis directed along Z. Structures containing β-hairpins (Zif268 and rubredoxin) similarly used an idealized hairpin oriented with its β sheet along Z. The sheet was additionally aligned onto X-Z by minimizing the deviations between the Cα coordinates and the X-Z plane. The idealized hairpin was generated by averaging the ϕ and ψ angles of the hairpins in Ziff268 and rubredoxin. To create the bundles, the helix (or sheet) was then translated along X, Y, and Z (Xtrans, Ytrans, Ztrans) and rotated about its own axis by the angle α. The value of α is given relative to an ''ideal'' value expected if a vector originating at the center of the helix and bisecting the Cα atoms of the central ''a'' and ''d'' positions were directed precisely towards the bundle axis (in a projection onto the X-Y plane). Additionally, the secondary structure was then tilted (through the angle β) by rotating about an axis defined by the bundle origin and the intersection of the helix axis with the X-Y plane. Further, the inclination of the helix (γ) was produced by rotation about an axis in the X-Y plane orthogonal to that used to produce the β-tilt angle. Finally, the bundles were created by using the symmetry operation shown. Optimal values of the fitting parameters were obtained by using a genetic algorithm to optimize the superposition of the D_2 symmetric model onto each crystal structure. The optimal superposition and associated rms deviation were calculated by using the Cα coordinates. Calculations were run on an SGI Indigo2 workstation using code written in Fortran 77. The PDB codes are as follows: [a]Zif268, 1ZAA (61), residues 35–44, 47–59; [b]1BRF (84), residues 3–12, 36–45; [c]1BCF (67), residues 47–58, 14–25, 90–101, 123–134; [d]1RYT (68), residues 16–27, 49–60, 90–98, 100–102, 124–135; [e]2FHA (85), residues 23–34, 58–69, 103–114, 137–148; [f]1RIB (86), residues 80–91, 111–122, 234–245, 200–204, 206–212; [g]1MTY (72), residues 110–121, 140–151, 239–250; [h]1AFR (71), residues 139–150, 192–203, 222–233; [i]2MHR (64) residues 21–37, 40–65, 70–87, 94–106; [j]1HMO (87) residues 21–37, 40–65, 70–87, 89–101; [k]1OXY (88), residues 170–180, 198–207, 321–331, 358–367.

occupy ''a'' and ''d'' positions, which project toward the center of the bundle (Fig. 3). The diiron site is bounded by ''e'' and ''b'' residues that line two sides of the active site (the e/b interface); the other two sides are formed by ''c'' and ''g'' (the c/g interface). The residues along the b/e interface in diiron proteins appear to serve a primarily structural role; they are tightly packed, often in an ''Alacoil'' (73) interaction. In contrast, the residues at one c/g interface hydrogen bond to the liganding His residue, whereas the opposite c/g interface defines the entry to the active site. Finally, the residues at the ''a'' and ''d'' positions above and below the liganding residues form the top and bottom of the active site. These side chains are often hydrophobic, and their association provides part of the driving force for the association of the bundle. Residues in these positions are additionally important for function; for example, in proteins with ferroxidase activity, a Tyr residue at an ''a'' position often hydrogen bonds to the carboxylate oxygen of one of the liganding Glu residues (27, 68). Also, a Tyr at an analogous position forms a free radical in R2.

Design of a Mimic of Diiron Proteins. To test the above analysis, we have designed a minimal idealized version of a diiron protein. The modeling began with a D_2-symmetric four-helix bundle of identical unconnected 21-residue helices. To maintain a constant

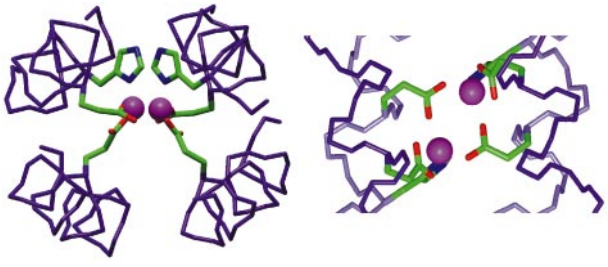

Fig. 2. Structure of dimetal ion site in an idealized diiron protein. Two Glu side chains form a bridging interaction between the metal ions, whereas the remaining two carboxylates form a one- or two-coordinate interaction with a single metal ion. Two His side chains are visible behind the ions. Two vacant sites face the viewer and are trans to the His ligands (*Right*). The figure shows the crystal structure of DF1; carbon atoms are green, nitrogens are blue, oxygens are red, and metal ions are magenta. The backbone trace is shown in purple.

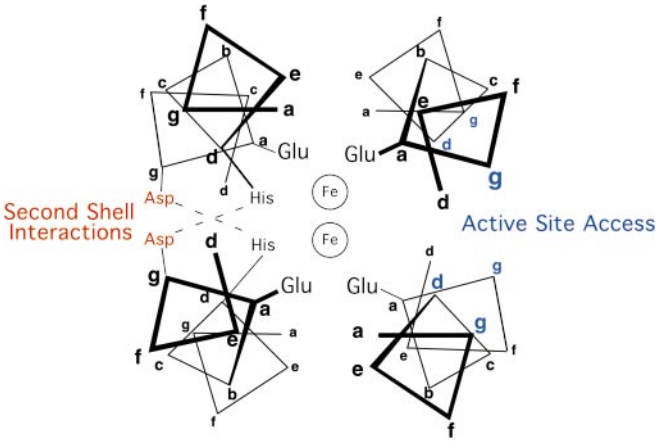

Fig. 3. Diagram of the active site of a diiron protein. Residue positions are labeled according to the heptad repeat generally applied to coiled coils. Four Glu side chains from ''a'' positions project toward the center of the bundle, as do two His side chains at ''d'' positions. The interfaces between helices are defined by ''b'' and ''e'' residues (*Upper* and *Lower*) or ''g'' and ''c'' residues (*Left* and *Right*). Second-shell hydrogen bonding residues are highlighted in red. Those residues that allow access to the open coordination sites of the metal ions are shown in blue.

interhelical distance throughout the bundle, the helices were slightly curved by providing a left-handed superhelical twist (pitch = 190 Å). To provide a Glu_4His_2 liganding environment, the D_2 symmetry was relaxed to C_2 leading to a homodimer of helix-loop-helix motifs. One helix in each monomer included a Glu-Xxx-Xxx-His motif, whereas the other contained a single Glu ligand. Liganding side chains were placed in the appropriate rotamers to allow interaction with a diiron center. The final positions of the helices were dictated by three requirements: (*i*) The geometry of the liganding site was restrained to bind diiron with two bridging Glu carboxylates, two nonbridging Glu side chains, and the δ-N of two His side chains; (*ii*) the helical packing angles and distances were constrained to match those typically observed in the active sites of diiron proteins; and (*iii*) precise 2-fold symmetry between the two pairs of helices was enforced.

Next, residues that might serve a functional role were added. An Asp residue at position 35 was included to hydrogen bond to the liganding His residue of a neighboring helix. Also, Tyr-17 at a "d" position was included to form a hydrogen bond with a nonbridging Glu ligand from a neighboring helix and possibly also to form a free radical under appropriate conditions. The packing of these residues provided geometric constraints, such that the side chains at the remaining solvent-inaccessible "a" and "d" core positions could be chosen by visual inspection. Almost precisely the same residues were automatically chosen by using a side-chain repacking algorithm (74). A collection of hydrophobic and polar residues was included at the remaining interfacial positions by using criteria discussed previously (49). Finally, helix-stabilizing polar residues were placed at the most exposed positions to provide water solubility. An idealized γ-α_L-β interhelical loop (8, 75) was included between the two pairs of helices. The resulting protein is designated "Due Ferro 1" (DF1), which means "two iron" in Italian. The amino acid sequence of DF1 from its N to C terminus is DYLRELLKLELQLIKQYREALEYVKLPVLAKILEDEEKHIEWLETILG.

Solution Characteristics of DF1. The 48-residue DF1 protein was prepared in good yield by standard solid phase methods used previously for the synthesis of similarly sized designed proteins (37). DF1 adopts a folded helical conformation in aqueous solution, irrespective of its ligation state. The CD spectrum for both the apo- and Co(II)-reconstituted protein ($\theta_{222} = -22{,}300$ deg·cm^2·dmol^{-1}·res^{-1}; $\theta_{209} = -21{,}100$ deg·cm^2·dmol^{-1}·res^{-1}) is consistent with its proposed helical structure. Analytical ultracentrifugation [10 mM Mops buffer, pH 6.5/60 μM protein/120 μM Co(II)] indicated that the protein sedimented as a single homogeneous species with an apparent molecular weight of 11,600 Da, consistent with the value expected for a dimer (11,760 Da). Reversible urea-induced unfolding curves (monitored by CD spectroscopy) showed that DF1 exists in an equilibrium between a folded helical dimer and an unfolded monomer. The secondary structure is lost in a single transition, with a midpoint that depends on the total peptide concentration, as expected for a reversible monomer–dimer equilibrium. The apparent free energy of dimerization extrapolated to zero denaturant concentration was -12.8 ± 0.6 kcal/mol (1 M standard state), corresponding to a dissociation constant of 0.41 nM (10 mM phosphate buffer, pH 6.5).

DF1 has been demonstrated to bind Zn(II), Co(II), and Fe(II). The properties of the iron complex will be presented in a subsequent paper. The spectrum of the di-Co(II) complex shows absorption maxima at 524, 574, and 594 nm. The position and the intensities (see *Materials and Methods*) of these bands are in reasonable agreement with the values reported in the literature for the Co(II) derivative of bacterioferritin (76). This finding supports the proposal that the coordination of the cobalt at the dinuclear site is in approximately the designed geometry.

Table 2. Summary of the data collection and refinement statistics for DF1

Data collection	
Resolution range, Å	44.7–2.5
No. of observations	23,396
No. of unique reflections	4,717
Completeness, %	
All data	95.5
Last shell (2.64–2.50 Å)	96.7
Multiplicity	5.2
R_{merge}*	0.107
Space group	$C222_1$
Cell dimension, Å	
a	36.07
b	89.16
c	79.89
V, Å^3	256,900
Final refinement parameters	
No. of reflections used†	4,270
No. of protein atoms	1,248
No. of metal atoms	3
No. of solvent atoms	32
R_{conv}/R_{free}‡	0.237/0.304
rms deviation from ideal geometry, Å	
Bond lengths	0.009
Bond angle distances	0.036
Planar groups	0.020
Average B factors, Å^2	
Main chain	40.47
Side chain	41.41
Metal	32.91
Solvent	48.66

*$R_{merge} = \Sigma_h\Sigma_i|I_{h,i} - I_h|/\Sigma_h\Sigma_i I_{h,i}$, where $I_{h,i}$ and I_h are the *i*th measurement and mean measurement of reflection *h*, respectively, and the sum is over all reflections for which more than one measurement is recorded.

†The number of reflections used in the calculation of R_{conv}, 202 reflections, 5% of the data, were selected at random at the beginning of the structure refinement for the calculation of R_{free}.

‡R_{conv}, $R_{free} = \Sigma_h||F_{obs_h}| - |F_{calc_h}||/\Sigma_h|F_{obs_h}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

Description of the Structure of DF1. The di-Zn(II) form of DF1 was crystallized and the structure determined by molecular replacement (Table 2). There are three monomers in the asymmetric unit cell of DF1 crystal. Two pack into a dimer (the quasisymmetric dimer) whose monomeric units are related by a pseudo-2-fold axis. The third monomer forms a symmetric dimer along one of the crystallographic 2-fold axes. The symmetric and quasisymmetric dimers have closely related structures, and both bind Zn(II) near the central 2-fold axis. The rms deviation between the two dimers is 1.36 Å for all atoms or 0.55 Å if only the backbones are compared. If the comparison is restricted to the core side chains and the backbone atoms within 10 Å of the Zn(II) site (the "structural core"), the rms deviation reduces to 0.41 Å.

The overall structure is very similar to the intended designed model, consisting of an antiparallel pair of helical hairpins with the desired topology (Figs. 2 and 4). The rms deviations between the backbones of the model and either the symmetric or quasisymmetric dimer are 1.64 Å and 1.65 Å, respectively. Again, the deviations are greatest near the ends of helices (far from the binding site). If the structural cores are compared with the model, the rms deviations are 1.0 Å for both the quasisymmetric and the symmetric dimer. Thus, the deviation between the observed structures and the model is only 0.6 Å greater than that observed between the two crystallographically determined dimers. The major source of difference between the observed

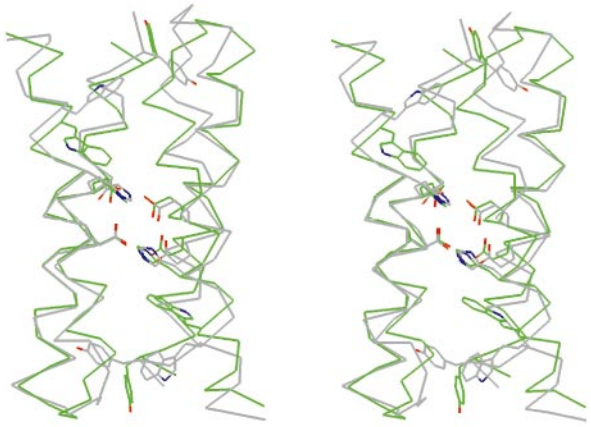

Fig. 4. Stereo comparison of 2.5 Å di-Zn-DF1 structure with designed model. The superposition of the crystal structure symmetric dimer (green) and the designed model (gray) shows the liganding Glu and His residues. Note that the dimetal-binding site is nearly identical between the model and the crystal structure. However, conformation of the Tyr-2 and Trp-42 side chains in the crystal structure differs markedly from that in the design.

structures and the model is associated primarily with differences in the conformations of two aromatic side chains, Tyr-2 and Trp-42. In the model, Tyr-2 projects away from the protein core, whereas in the crystallographic structures it drapes over the top of the bundle, helping to sequester hydrophobic residues from contact with solvent. Also, Trp-42, which was designed to tuck between two helices, instead occupies a more exposed interfacial orientation.

The Metal Ion-Binding Site. A binuclear metal-binding site appears as a pair of peaks of electron density (>10 σ) in the $2F_o - F_c$ map. No bridging water or hydroxide ions were observed at this resolution. The electron density of the liganding side chains is very well defined in this region of the map (Fig. 5) and allowed the unambiguous positioning of the side chains. The orientations of the liganding side chains are remarkably similar in the model and the experimental structure. Each metal ion is 5-coordinate; Glu-10 and 10′ (prime numbers refer to symmetry-related residues in the symmetric and quasisymmetric dimers) interact with both zinc ions in a 1,3 *syn-syn* bidentate bridging interaction, whereas Glu-36 and 36′ interact with individual ions in a chelating manner. His residues 39 and 39′ complete the liganding environment about the dimetal site by Nδ atom coordination to individual ions. The Zn(II)-Zn(II) distance is 3.9 Å, close to the distance observed between metal ions in the di-Mn(II) forms of bacterioferritin (4.0 Å) and R2 (3.6–3.7 Å), the diferrous forms of Δ^9 ACP desaturase (4.2 Å) and ribonucleotide reductase R2 (3.8 Å), and the Fe-Zn(II) form of rubrerythrin (3.7 Å) (68, 69). The pentacoordinate ligand arrangement with two 1,3 bridging carboxylates and two chelating carboxylates is also nearly identical to that observed in di-Fe(II) Δ^9 ACP desaturase (71) and di-Mn(II) bacterioferritin (67). In these proteins, the ligands surround the metal ions except for a vacant pair of sites along one face of the dimetal center. The vacant sites are oriented trans to the His ligands and are well oriented for interaction with bridging ligands such as a 1,2 peroxide. If the diiron form of DF1 adopts a similar geometry as the di-Zn(II) structure, its dimetal site would also be well suited for reaction with O_2.

The design of DF1 included four second-shell interactions involving hydrogen bonds to the liganding side chains, which were indeed observed in the crystallographic structure. Tyr-17 and Tyr-17′ lie within hydrogen bonding distance (2.6, 3.1, and 3.4 Å in the three asymmetric monomers) of the nonbridging Glu-36 (and Glu-36′, respectively). Also, one of the carboxylate oxygen atoms of Asp-35 and -35′ come to within 2.5–2.7 Å of the liganding His side chain. In summary, the crystal structure of DF1 is in excellent agreement with the designed model and demonstrates all specific H-bonded and metal ligand interactions included in the model.

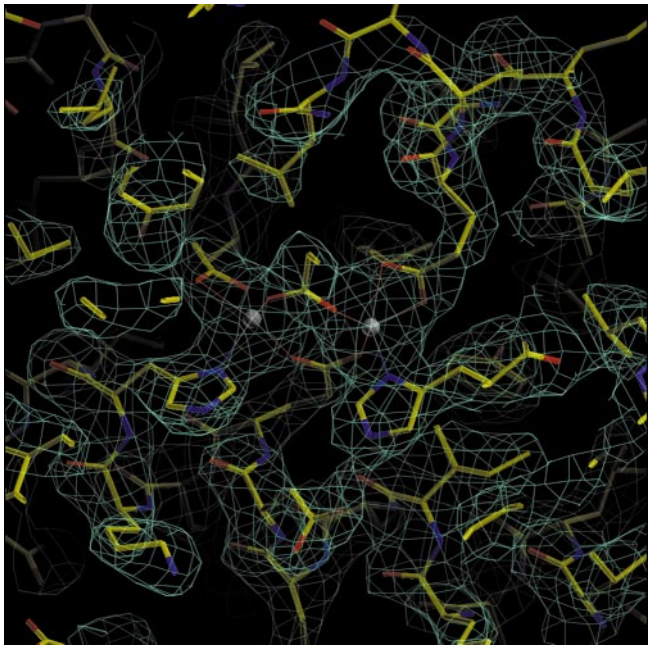

Fig. 5. Electron density map near the dinuclear metal-binding site. Contour levels are 1σ corresponding to 0.39 e/Å^3 for the $2F_o - F_c$ map.

Discussion

The *de novo* design of metalloproteins is an important step toward the engineering of novel materials, catalysts, and sensors. To facilitate the design of structurally defined metalloproteins, we have developed a framework for the geometric parameterization of both symmetric and asymmetric transition metal ion-binding sites. This approach was then tested through the design of a dimetal-binding protein. This design was particularly challenging, because it required the burial of six polar ionizable groups in the hydrophobic core, which were stabilized via second-shell hydrogen bonded interactions. Such second-shell interactions may prove to be essential for high-affinity binding and fine tuning of function (77, 78) and have not been considered in previous attempts to design metalloproteins. The success described here in the design and structure elucidation of a complex metalloprotein has significant implications for the evolution of natural proteins, the understanding of dimetal centers, and the automated design of metalloproteins.

Evolution of Natural and Nonnatural Metalloproteins. Our retrostructural analysis suggests that many complex metalloproteins may have evolved from primordial precursors that were formed by the noncovalent self assembly of simple secondary-structure motifs (59, 79, 80). Covalent connection of the individual secondary structures may have occurred at a later date, via either random shuffling of nonidentical units to give asymmetric proteins or gene duplication to give a symmetric protein. In both cases, the covalent connection would allow independent asymmetric mutation of each element of secondary structure. For example, individual members of the EXXH family of diiron

proteins may have evolved from a primordial precursor, consisting of an α-helix containing a Glu-Xxx-Xxx-His sequence. Subsequent covalent connection may have produced a helix-loop-helix motif that associated to form a dimetal-binding four-helix bundle. Loss of a single His residue from one of the helices in the helix-turn-helix motif would have provided a coordinately unsaturated position for binding oxygen. Interestingly, in each member of the EXXH family of proteins, the missing His residue occurs exclusively in the first helix of their two helix-turn-helix motifs, suggesting that the entire class of functionally divergent proteins may have evolved from a common helix-loop-helix precursor, similar to DF1.

Although the above evolutionary mechanism is speculative vis-à-vis natural proteins, it nevertheless provides a paradigm for the evolution of nonbiological metalloproteins. It should be possible to design and synthesize a small focused library of ligand-displaying secondary structure units. Individual members of this library could be mixed in discreet combinations and stoichiometries to provide a large combinatorial library of potential metalloproteins. Subsequent introduction of loops between combinations that show desirable metal binding or catalytic properties would provide single-chain proteins as candidates for further optimization.

Implication for Understanding Function of Dimetal Centers. Although natural dimetal ion proteins display a wide range of functions, their active sites are often remarkably similar. The mechanisms through which diiron proteins modulate the properties of the dimetal site may now be addressed by using variants of DF1. For example, although its structure was solved with zinc ions at the active site, it has been possible to use this structure to design a variant that forms a redox-active diiron complex (unpublished results). In DF1, Leu-13 and Leu-13′ occupy the space proximal to the vacant liganding sites of the dimetal center, preventing rapid access of O_2 and substrates. In subsequent designs, these residues have been mutated to smaller side chains, allowing more ready access of metal ions and O_2.

The role of second-shell interactions in defining the structures, reactivities, and function of diiron proteins is currently unknown. In the EXXH family of diiron proteins, a variety of side chains form H-bonds to the His ligands. This interaction may impart partial imidazolate character to the His ligand, thereby helping to stabilize high-valent Fe intermediates and additionally help fine tune the geometry of the imidazole rings. Tyrosine side chains often are found within H-bonding distance of one or more of the Glu ligands (in bacterioferritin, rubrerythrin, ferritin, myohemerythrin, and hemerythrin). A Tyr in R2 forms a free radical, which is essential for the activity of the catalytic subunit. The role of other Tyr residues that occupy similar positions in other proteins is unknown, although tyrosyl radicals have been detected in bacterioferritin (81) and H-ferritin (82). DF1 may be an ideal system for studying the roles of these second shell residues.

Automated Protein Design. This work also has implications for the automated *de novo* design of proteins with functional binding sites. The full automation of protein design requires computational variation of the backbone conformation as well as the identities and rotamers of the side chains. Previous workers used a mathematical parameterization of the backbone to design coiled coils (83), or four-helix bundles (49). Because the targets had precisely repeating geometries, it was necessary to vary the identities of only two or three core residues in these earlier studies. Here we demonstrate that a different parameterization can be used to design a much more complex protein, featuring a precisely defined active site.

The design of DF1 was accomplished in a hierarchic manner: the backbone conformation of the helical bundle was dictated by the desired chelation geometry, the symmetry of the overall structure, and low energy torsional angles of the liganding side chains. Next, residues important for stabilizing the chelating side chains were included, followed by the addition of the hydrophobic core. Polar residues and an interhelical loop were added in the last step of the design. Although each step was accomplished separately, we are currently automating the entire process. Although DF1 is a symmetric dimer, this approach should also be generalized easily to asymmetric structures.

The method described in this manuscript—in which the geometry of the active site helps dictate the overall fold of the protein—differs from the computational approaches of Hellinga (15, 17) and Clark (16, 18). Their methods instead begin with a fixed backbone structure (generally of a natural protein) and then search for convenient locations for introducing a metal ion-binding site. Over the last decade, these programs have been used successfully to graft a variety of structural and redox-active metal-binding sites into several different natural and designed proteins (6, 8, 17). However, because crystallographic or NMR structures are not available for any of these grafted sites, it is somewhat difficult to assess the merits of these methods relative to the *de novo* design approach described here. Clearly, both approaches have considerable potential, and further studies will be required to determine their full applicability.

The authors thank Stefania Galdiero and Giuseppina De Simone for protein crystallization and data collection, G. Patrick Brady for providing F77 code for the genetic algorithm, James Lear for analytical ultracentrifugation, and Stephen Lippard, Flavia Nastri, and Ornella Maglio for useful discussions. This work was supported by National Institutes of Health grant GM54616. We also acknowledge partial support from the Italian Ministry of University and Scientific Research (PRIN 9803184222) and the Materials Research Science and Engineering Center Program of the National Science Foundation (DMR96–32598).

1. Nordlund, P. & Eklund, H. (1995) *Curr. Opin. Struct. Biol.* **5,** 758–766.
2. Lange, S. J. & Que, L., Jr. (1998) *Curr. Opin. Chem. Biol.* **2,** 159–172.
3. Waller, B. J. & Lipscomb, J. D. (1996) *Chem. Rev.* **96,** 2625–2657.
4. Feig, A. L. & Lippard, S. J. (1994) *Chem. Rev.* **94,** 759–805.
5. Nastri, F., Lombardi, A., D'Andrea, L. D., Sanseverino, M., Maglio, O. & Pavone, V. (1998) *Biopolymers* **47,** 5–22.
6. Regan, L. (1995) *Trends Biochem. Sci.* **20,** 280–285.
7. Gibney, B. R., Rabanal, F. & Dutton, P. L. (1997) *Curr. Opin. Chem. Biol.* **1,** 537–542.
8. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. (1999) *Annu. Rev. Biochem.* **68,** 779–819.
9. Williamson, D. A. & Benson, D. R. (1998) *Chem. Commun.* **9,** 961–962.
10. Rojas, N. R., Kamtekar, S., Simons, C. T., McLean, J. E., Vogel, K. M., Spiro, T. G., Farid, R. S. & Hecht, M. H. (1997) *Protein Sci.* **6,** 2512–2524.
11. Suzuki, K., Hiroaki, H., Kohda, D., Nakamura, H. & Tanaka, T. (1998) *J. Am. Chem. Soc.* **120,** 13008–13015.
12. Sakamoto, S., Obataya, I., Ueno, A. & Mihara, H. (1999) *Chem. Commun.* **1999,** 1111–1112.
13. Huffman, D. L., Rosenblatt, M., M. & Suslick, K. (1998) *J. Am. Chem. Soc.* **120,** 6183–6184.
14. Willner, I., Heleg-Shabtai, V., Katz, E., Rau, H. K. & Haehnel, W. (1999) *J. Am. Chem. Soc.* **121,** 6455–6468.
15. Hellinga, H. W. & Richards, F. M. (1991) *J. Mol. Biol.* **222,** 763–785.
16. Clarke, N. D. & Yuan, S. M. (1995) *Proteins* **23,** 256–263.
17. Hellinga, H. W. (1998) *Folding Des.* **3,** R1–R8.
18. Desjarlais, J. R. & Clarke, N. D. (1998) *Curr. Opin. Struct. Biol.* **8,** 471–475.
19. Pinto, A., Hellinga, H. W. & Caradonna, J. P. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5562–5567.
20. Sigman, J. A., Kwok, B. C., Gegenbach, A. & Lu, Y. (1999) *J. Am. Chem. Soc.* **121,** 8949–8950.
21. Brittain, I. J., Huang, X. & Long, E. C. (1998) *Biochemistry* **37,** 12113–12120.
22. Hay, R. W., Hassan, M. M. & You-Quan, C. (1993) *J. Inorg. Biochem.* **52,** 17–25.
23. Stemmler, A. J. & Burrows, C. J. (1999) *J. Am. Chem. Soc.* **121,** 6956–6957.
24. Torrado, A., Walkup, G. K. & Imperiali, B. (1998) *J. Am. Chem. Soc.* **120,** 609–610.
25. Gibney, B. R., Mulholland, S. E., Rabanal, F. & Dutton, P. L. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 15041–15046.

26. Coldren, C., Hellinga, H. W. & Caradonna, J. P. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6635–6640.
27. Summa, C. M., Lombardi, A., Lewis, M. & DeGrado, W. F. (1999) *Curr. Opin. Struct. Biol.* **9,** 500–508.
28. Richardson, J., Richardson, D. C., Tweedy, N. B., Gernert, K. M., Quinn, T. P., Hecht, M. H., Erickson, B. W., Yan, Y., McClain, R. D. & Donlan, M. E. (1992) *Biophys. J.* **63,** 1186–1209.
29. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T. & DeGrado, W. F. (1995) *Science* **270,** 935–941.
30. Fezoui, Y., Connolly, P. J. & Osterhout, J. J. (1997) *Protein Sci.* **6,** 1869–1877.
31. Struthers, M. D., Cheng, R. P. & Imperiali, B. (1996) *Science* **271,** 342–345.
32. Dahiyat, B. I. & Mayo, S. L. (1997) *Science* **278,** 82–87.
33. Ilyina, E., Roongta, V. & Mayo, K. H. (1997) *Biochemistry* **36,** 5245–5250.
34. Hill, R. B. & DeGrado, W. F. (1998) *J. Am. Chem. Soc.* **120,** 1138–1145.
35. Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. & Stroud, R. M. (1997) *Nat. Struct. Biol.* **4,** 1039–1046.
36. Walsh, S. T. R., Cheng, H., Bryson, J. W., Roder, H. & DeGrado, W. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5486–5491.
37. Choma, C. T., Lear, J. D., Nelson, M. J., Dutton, P. L., Robertson, D. E. & DeGrado, W. F. (1994) *J. Am. Chem. Soc.* **116,** 856–865.
38. Gibney, B. R. & Dutton, P. L. (1999) *Protein Sci.* **8,** 1888–1898.
39. Handel, T. M., Williams, S. A. & DeGrado, W. F. (1993) *Science* **261,** 879–885.
40. Regan, L. & Clarke, N. D. (1990) *Biochemistry* **29,** 10878–10883.
41. D'Auria, G., Maglio, O., Nastri, F., Lombardi, A., Mazzeo, M., Morelli, G., Paolillo, L., Pedone, C. & Pavone, V. (1997) *Chem. Euro. J.* **3,** 350–362.
42. Skalicky, J., Gibney, B. R., Rabanal, F., Urbauer, R. J. B., Dutton, P. L. & Wand, A. J. (1999) *J. Am. Chem. Soc.* **121,** 4941–4951.
43. Dahiyat, B. I., Sarisky, C. A. & Mayo, S. L. (1997) *J. Mol. Biol.* **273,** 789–796.
44. Lazar, G. A. & Handel, T. M. (1998) *Curr. Opin. Chem. Biol.* **2,** 675–679.
45. Desjarlais, J. R. & Handel, T. M. (1999) *J. Mol. Biol.* **290,** 305–318.
46. Ghirlanda, G., Lear, J. D., Lombardi, A. & DeGrado, W. F. (1998) *J. Mol. Biol.* **281,** 379–391.
47. Su, A. & Mayo, S. L. (1997) *Protein Sci.* **6,** 1701–1707.
48. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. (1998) *Science* **282,** 1462–1467.
49. Berg, J. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 99–102.
50. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. (1996) *Structure (London)* **4,** 1171–1180.
51. Navaza, J. (1994) *Acta Crystallogr. A* **50,** 157–163.
52. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
53. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Crystallogr. D* **53,** 240–255.
54. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291.
55. Mann, C. J. & Matthews, C. R. (1993) *Biochemistry* **32,** 5282–5290.
56. Zhang, J. H., Kurtz, D. M., Jr., Xia, Y. M. & Debrunner, P. G. (1991) *Biochemistry* **30,** 583–589.
57. Zhang, J. H. & Kurtz, D. M., Jr. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7065–7069.
58. Keech, A. M., Le Brun, N. E., Wilson, M. T., Andrews, S. C., Moore, G. R. & Thomson, A. J. (1997) *J. Biol. Chem.* **272,** 422–429.
59. Betz, S. F. & DeGrado, W. F. (1996) *Biochemistry* **35,** 6955–6962.
60. Betz, S. F., Liebman, P. A. & DeGrado, W. F. (1997) *Biochemistry* **36,** 2450–2458.
61. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252,** 809–817.
62. Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19,** 405–421.
63. Stenkamp, R. E. (1994) *Chem. Rev.* **94,** 715–726.
64. Sheriff, S., Hendrickson, W. A. & Smith, J. L. (1987) *J. Mol. Biol.* **197,** 273–296.
65. Andersson, K., K. & Gräslund, A. (1995) *Adv. Inorg. Chem.* **43,** 359–403.
66. Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingston, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., *et al.* (1991) *Nature (London)* **349,** 541–544.
67. Frolow, F., Kalb (Gilboa), A. J. & Yariv, J. (1994) *Nat. Struct. Biol.* **1,** 453–460.
68. deMare, F., Kurtz, D. M., Jr. & Nordlund, P. (1996) *Nat. Struct. Biol* **3,** 539–546.
69. Sieker, L. C., Holmes, M., Le Trong, I., Turley, S., Santarsiero, B. D., Liu, M. Y., LeGall, J. & Stenkamp, R. E. (1999) *Nat. Struct. Biol.* **6,** 308–309.
70. Andersson, M. E., Högbom, M., Rinaldo-Matthis, A., Andersson, K. K., Sjöberg, B.-M. & Nordlund, P. (1999) *J. Am. Chem. Soc.* **121,** 2346–2352.
71. Lindqvist, Y., Huang, W., Schneider, G. & Shanklin, J. (1996) *EMBO J.* **15,** 4081–4092.
72. Rosenzweig, A. C., Brandstetter, H., Whittington, D. A., Nordlund, P., Lippard, S. J. & Frederick, C. A. (1997) *Proteins* **29,** 141–152.
73. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S. & Richardson, D. C. (1995) *Protein Sci.* **4,** 2252–2260.
74. Desjarlais, J. R. & Handel, T. M. (1995) *Protein Sci.* **4,** 2006–2018.
75. Efimov, A. V. (1991) *Protein Eng.* **4,** 245–250.
76. Le Brun, N. E., Keech, A. M., Mauk, M. R., Mauk, A. G., Andrews, S. C., Thomson, A. J. & Moore, G. R. (1996) *FEBS Lett.* **397,** 159–163.
77. Huang, C. C., Lesburg, C. A., Kiefer, L. L., Fierke, C. A. & Christianson, D. W. (1996) *Biochemistry* **35,** 3439–3446.
78. Karlin, S., Zhu, Z. Y. & Karlin, K. D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14225–14230.
79. McLachlan, A. D. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **LII**.
80. Kinoshita, K., Kidera, A. & Go, N. (1999) *Protein Sci.* **8,** 1210–1217.
81. Cheesman, M. R., Le Brun, N. E., Kadir, F. H., Thomson, A. J., Moore, G. R., Andrews, S. C., Guest, J. R., Harrison, P. M., Smith, J. M. & Yewdall, S. J. (1993) *Biochem. J.* **292,** 47–56.
82. Chen-Barrett, Y., Harrison, P. M., Treffry, A., Quail, M. A., Arosio, P., Santambrogio, P. & Chasteen, N. D. (1995) *Biochemistry* **34,** 7847–7853.
83. Harbury, P. B., Tidor, B. & Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8408–8412.
84. Bau, R., Rees, D. C., Kurtz, D. M., Scott, R. A., Huang, H., Adams, M. W. W. & Eidsness, M. K. (1998) *J. Biol. Inorg. Chem.* **3,** 484–493.
85. Hempstead, P. D., Yewdall, S. J., Fernie, A. R., Lawson, D. M., Artymiuk, P. J., Rice, D. W., Ford, G. C. & Harrison, P. M. (1997) *J. Mol. Biol.* **268,** 424–448.
86. Nordlund, P. & Eklund, H. (1993) *J. Mol. Biol.* **232,** 123–164.
87. Holmes, M. A., Le Trong, I., Turley, S., Sieker, L. C. & Stenkamp, R. E. (1991) *J. Mol. Biol.* **218,** 583–593.
88. Magnus, K. A., Hazes, B., Ton-That, H., Bonaventura, C., Bonaventura, J. & Hol, W. G. (1994) *Proteins* **19,** 302–309.