



Account / Revue

## Diiron-containing metalloproteins: Developing functional models

Ornella Maglio<sup>a</sup>, Flavia Nastri<sup>a</sup>, Rafael Torres Martin de Rosales<sup>b</sup>,  
Marina Faiella<sup>a</sup>, Vincenzo Pavone<sup>a</sup>, William F. DeGrado<sup>c</sup>,  
Angela Lombardi<sup>a,\*</sup>

<sup>a</sup> Department of Chemistry, University Federico II of Naples, Complesso Universitario Monte S. Angelo, Via Cintia, 80126 Napoli, Italy

<sup>b</sup> Department of Chemistry, Imperial College London, Exhibition Road, South Kensington, London SW7 2AY, UK

<sup>c</sup> Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6059, USA

Received 23 December 2006; accepted after revision 20 March 2007

Available online 22 May 2007

### Abstract

A major objective in protein science is the design of enzymes with novel catalytic activities that are tailored to specific applications. Such enzymes may have great potential in biocatalysis and biosensor technology, such as in degradation of pollutants and biomass, and in drug and food processing. To reach this objective, investigations into the basic biochemical functioning of metalloproteins are still required. In this perspective, metalloprotein design provides a powerful approach first to contribute to a more comprehensive understanding of the way metalloproteins function in biology, with the ultimate goal of developing novel biocatalysts and sensing devices. Metalloprotein mimetics have been developed through the introduction of novel metal-binding sites into naturally occurring proteins as well as through de novo protein design. We have approached the challenge of reproducing metalloprotein active sites by using a miniaturization process. We centered our attention on iron-containing proteins, and we developed models for heme proteins and diiron–oxo proteins. In this paper we summarize the results we obtained on the design, structural, and functional properties of DFs, a family of artificial diiron proteins. **To cite this article:** O. Maglio *et al.*, *C. R. Chimie* 10 (2007).

© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

### Résumé

Un objectif majeur de la science des protéines est la conception d'enzymes possédant de nouvelles activités catalytiques et conçues pour des applications spécifiques. De telles enzymes peuvent avoir un grand potentiel dans des technologies de la biocatalyse et des biocapteurs telles que la dégradation des polluants et de la biomasse, ainsi que la transformation d'aliments et de produits pharmaceutiques. Pour atteindre cet objectif, il est nécessaire de réaliser une étude du fonctionnement biochimique de base des métalloprotéines. Dans cette optique, la conception des métalloprotéines fournit une approche puissante, initialement pour contribuer à une meilleure compréhension de la façon dont les métalloprotéines fonctionnent dans les processus biologiques, avec le but final de développer de nouveaux biocatalyseurs et capteurs. Les mimétiques des métalloprotéines ont été développés avec

\* Corresponding author.

*E-mail addresses:* [ornella.maglio@unina.it](mailto:ornella.maglio@unina.it) (O. Maglio), [flavia.nastri@unina.it](mailto:flavia.nastri@unina.it) (F. Nastri), [r.torresm@imperial.ac.uk](mailto:r.torresm@imperial.ac.uk) (R.T. Martin de Rosales), [marinafaiella@yahoo.it](mailto:marinafaiella@yahoo.it) (M. Faiella), [vincenzo.pavone@unina.it](mailto:vincenzo.pavone@unina.it) (V. Pavone), [wdegrado@mail.med.upenn.edu](mailto:wdegrado@mail.med.upenn.edu) (W.F. DeGrado), [angelina.lombardi@unina.it](mailto:angelina.lombardi@unina.it) (A. Lombardi).

l'introduction de nouveaux sites de liaison des métaux au sein de protéines naturelles et ainsi qu'à travers la conception de novo de protéines. Nous avons relevé le défi de reproduire les sites actifs des métalloprotéines en utilisant un processus de miniaturisation. En particulier, nous avons focalisé notre attention sur les protéines contenant du fer et développé des modèles pour les protéines hémiques et les protéines contenant un noyau « difer » ponté par l'oxygène. Dans cet article, nous présentons les résultats que nous avons obtenus dans la conception et l'étude des propriétés structurales et fonctionnelles des DFs, une famille de protéines artificielles à site actif binucléaire de fer. **Pour citer cet article** : O. Maglio et al., C. R. Chimie 10 (2007).

© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

**Keywords**: De novo protein design; Computational protein design; Metalloprotein; Diiron protein; Structure and function

**Mots-clés** : Conception de nouvelles protéines ; Conception assistée par ordinateur de protéines ; Métalloprotéines ; Protéines à site binucléaire de fer ; Structure et fonction

## 1. Introduction

The application of catalytic reactions in industrial processes, in drug detoxification, in environment bioanalysis and in biodegradation of pollutants makes biocatalysts very attractive compounds to be studied. Among biocatalysts, metal containing enzymes play important roles in numerous chemical and biological transformations. Therefore, increasing attention is being devoted to the possibility of tailoring synthetic catalysts that mimic natural metalloenzymes [1–9].

In metalloenzymes, catalytic activity and stability of the metal cofactors are regulated by specific interactions with amino acid residues in the protein cavities [10,11]. First, the protein matrix shapes the primary coordination shell of the metal ion dictating the composition, number, and geometry of the ligands. In addition, residues in the second shell, the immediate surroundings of the primary coordination sphere, influence a variety of structural and chemical facets, including hydrogen-bonding interactions to the ligands,  $pK_a$  values of ligand groups, and the control of the metal center oxidation state and redox potential. Finally, neighbouring side chains also exert steric and chemical controls over the ability of the metal ion(s) to bind or discriminate the substrate and to accommodate conformational changes [10,11].

Thus, it is clearly evident that the structures of the metal-binding sites in proteins reflect a delicate interplay between opposing requirements for tight binding of the cofactor, which drives toward static, coordinately saturated geometries, versus function, which requires the cofactor to be coordinately unsaturated, sometimes with unusual geometries, and properly positioned for binding substrates and undergoing dynamic changes in configuration during catalysis.

Therefore, numerous efforts are continuously aimed to a better understanding of the factors that modulate

metalloprotein activity, with the ultimate goal of developing metalloenzymes with novel structures and functions [1–9]. The study of natural metalloenzymes and related mutants, generated by site-directed mutagenesis, as well as the study of smaller complexes containing designed organic ligands, has clarified several fundamental properties of natural proteins [12–17]. For example, mutation studies have shown that modest changes brought to an enzyme's structure can have large effects on its catalytic properties. Chemical modification of a single amino acid residue can change the fundamental activity of the enzyme, and a single site mutation can alter its enantioselectivity [12,13].

Even though impressive progress has been made with both strategies, the study of large proteins may be hampered by their extreme complexity, and it is often difficult to design small molecule complexes that can bind multiple substrates and metal ions in water. Therefore, due to the inherent complexity of enzyme structure, tailoring synthetic biocatalysts still is a daunting challenge, since it requires the development of sophisticated molecular architectures that distil the quintessential elements believed to be responsible for the activities.

This problem has been recently approached with the stratagem of synthesizing scaled-down polypeptides that self-assemble with metal ions and cofactors to form metalloprotein mimetics.

These structures are simple enough to avoid much of the ambiguity of interpretation associated with large proteins, and yet should simultaneously be of sufficient size and chemical diversity to allow the construction of functional sites. Thus, these mimetics — which stand at the crossroads of traditional small molecule models and large proteins — should allow one first to understand the properties of natural proteins, and second to construct economical catalysts and sensing devices.

The development of peptide-based models takes advantage of recent progress in both design and synthesis of peptides and proteins. Relatively simple rules have been established for the development of stable and well-defined protein and metalloprotein models, using either minimalist design strategies or de novo design approach [5–7,18,19]. We have approached the challenge of reproducing metalloprotein active sites using both strategies.

We centered our attention on iron-containing proteins, and we developed models for heme proteins (mimochromes) [20–26], iron–sulfur proteins [27,28], and diiron–oxo proteins (DFs) [29].

All these models have a common characteristic: they are fully or partially  $C_2$ -symmetric systems, intended to reproduce the quasi-symmetrical structure of a metalloprotein. The use of  $C_2$  symmetry is particularly advantageous: it simplifies the design, synthesis and structural characterization of the models, even though it leaves open the problem of possible diastereomeric forms.

A miniaturization process was applied in the development of the first two classes of artificial metalloproteins [20–28]. By a detailed analysis of the crystal structures of natural heme proteins and iron–sulfur proteins, we successfully selected the minimum set of constituents necessary for an accurate reconstruction of the active sites' structures. In particular, we selected the type and number of constituents to be assembled, to include all the elements of the primary and secondary coordination shell to the metal. The results obtained on these two class of metalloprotein models, indicated that these simple, structurally defined systems provide a good opportunity for investigating how changes in the polarity, solvent accessibility, and electrostatic potential surrounding the active site affect its electrochemical midpoint potential and chemical reactivity. These molecules are very promising systems in the field of metalloenzymes model compounds: they have a high degree of functional similarity with respect to their natural counterparts, and they are excellent scaffolds to engineer new peptide–metal complexes, with different metal cofactors.

In the field of diiron–oxo proteins, we used a de novo design approach for the development of artificial models. In particular, we developed the DF (Due Ferri) family of artificial proteins, as models of diiron and dimanganese metalloproteins, with the aim of reproducing the functional specificity of the natural systems and with the ultimate goal of developing new metalloenzymes with functions unprecedented in nature [29–42]. The DF proteins were developed through an iterative process of design and rigorous characterization, which

allowed us to diagnose the problems associated with the initial models, and to improve the models in subsequent designs. Thus, the problem of designing the DF family was approached through several steps. The first involved the design of stable, uniquely folded proteins that contain the metal binding site; we engineered DF1, which was an antiparallel dimer of *helix–loop–helix* motif, able to bind the dimetal cofactor near the center of the structure [30–32]. To improve the functional properties of the initial model, DF1 was subjected to several changes in the sequence, as well in the loop structure (DF1 and DF2 subsets) [33–40]. The design further evolved with the construction of versions in which the four helices are distinct chains that come together by non-covalent self assembly (DFtet) [41–43]. Each class has different advantages. For example, the symmetric nature of the dimeric derivatives simplifies interpretation of data, whereas analogues of the four-chain DFtet constructs can be mixed and matched to allow easy generation of combinatorial diversity.

In this paper, we will summarize all the results we obtained to date from our work on the DF family of proteins, by using both experimental and theoretical studies.

## 2. DF1: a minimal model for diiron proteins

Diiron sites are found in a functionally diverse class of proteins involved in oxygen binding and activation [44–48]. Diiron proteins catalyze hydroxylation, desaturation, and epoxidation reactions on a variety of alkyl and aryl substrates [44–51], and a diiron site in ribonucleotide reductase, which participates in DNA biosynthesis, is responsible for the formation of organic radicals [52]. The hemerythrin is responsible for the transport of  $O_2$ , and ferritins are important for iron oxidation (ferroxidase activity), storage and transport [53–57]. There are many structural and mechanistic commonalities amongst this class of diiron proteins.

Although the folds of these proteins are frequently complex, a simple four-helix bundle (Fig. 1) houses their active sites [58–62]. The helices have a slight left-handed tilt as in classical four-helix bundles [63] and antiparallel four-stranded coiled coils [64]. Thus, their geometries may be discussed in terms of the heptad nomenclature of coiled coils [64–66].

A detailed analysis of the active site of several natural diiron proteins (see Fig. 1) reveals a highly symmetric arrangement of the liganding side chains [58–62]. Almost invariably, each helix contains a single coordinating Glu residue in an “a” position of the heptad motif,

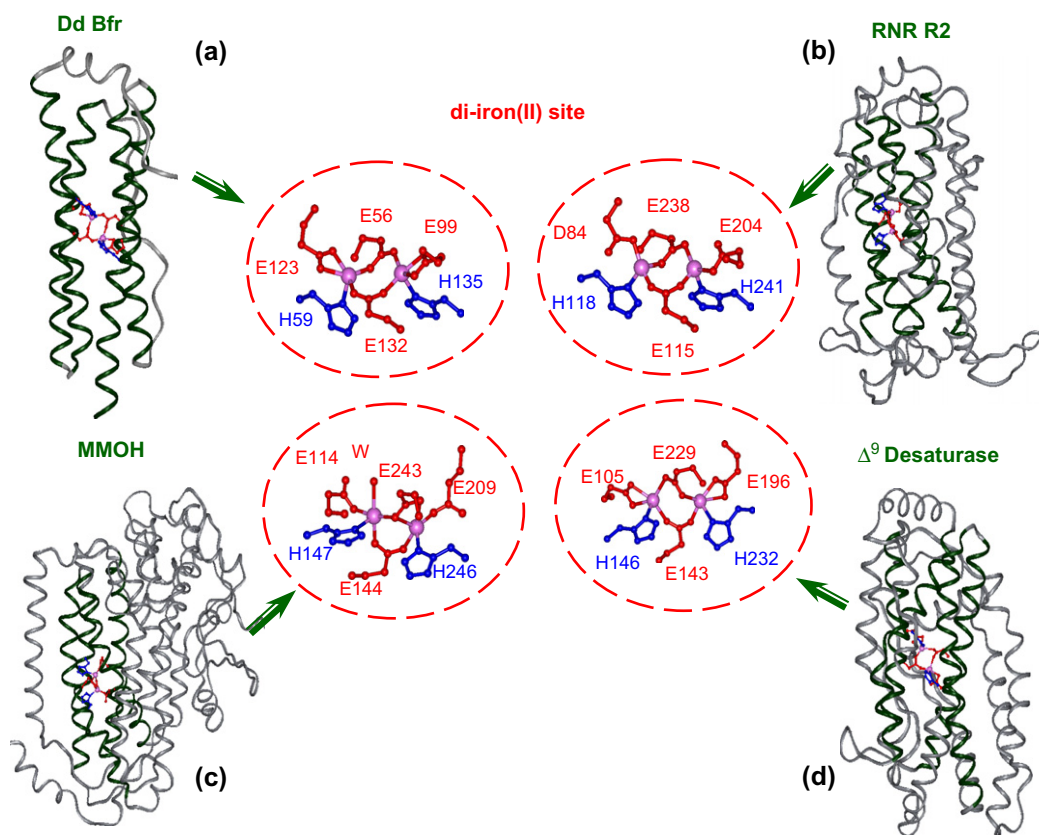


Fig. 1. Four-helix bundle motif housing the diiron site and details of the metal binding site in: (a) bacterioferritin (PDB code 1NF4), (b) ribonucleotide reductase (PDB code: 1XIK), (c) hydroxylase component of methane monooxygenase (PDB code: 1FYZ), (d)  $\Delta^9$ -ACP desaturase (PDB code: 1AFR), as derived from X-ray crystallography.

which projects towards the center of the bundle. Two carboxylates bridge both metal ions, while the other two interact with a single metal ion in a monodentate or bidentate chelating interaction. Two His residues at “d” position, three residues from the two bridging Glu side chains coordinate the iron as monodentate ligands, thus forming the Glu–Xxx–Xxx–His binding motif common to this class of proteins. Small side chains generally occur at the analogous “d” positions of the other two helices. Importantly, these side chains control the access of oxygen and substrates to the diiron center.

Other residues also help to shape the active sites of diiron proteins. The top and bottom of the active site are defined by the side chains projecting from “d” and “a” of neighboring heptads. The sides of the active site are defined by residues at positions “g” and “c” on two faces of the four-helix bundle, and “b” and “e” on the other two faces (Fig. 2). The residues at the b/e faces of the helical bundle tend to be tightly packed and appear to serve a structural role. In contrast, the c/g faces appear to be important for function; one g/c face presents side chains that help to define the entry and binding

of substrates. The opposite c/g face may help to tune the properties of the coordination site by hydrogen bonding to the coordinating His residues, thus forming second-shell interactions.

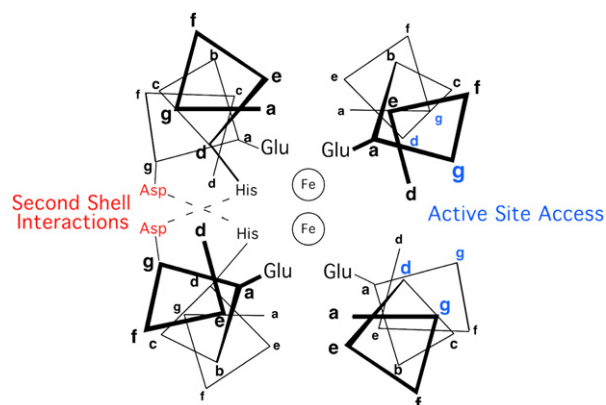


Fig. 2. Helix-wheel representation of the antiparallel four-stranded coiled coil in proximity of the metal-binding site, in diiron protein. Residue positions are labeled according to the heptad repeat generally applied to coiled coils.

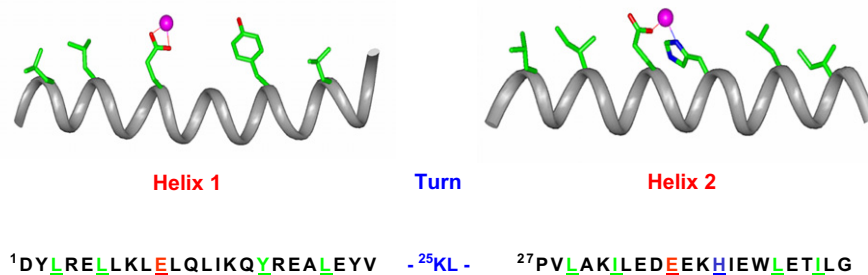


Fig. 3. The sequence of DF1. The DF1 sequence is indicated, together with the intended secondary structure. Residues that are expected to be essential for folding and/or function are shown in stick representation on the helices and are underlined in the sequence.

On the basis of this analysis, we designed the first model, DF1, which is made up of two 48-residue *helix–loop–helix* ( $\alpha_2$ ) motifs, able to specifically self-assemble into an antiparallel four-helix bundle [30–32]. Fig. 3 reports the amino acid sequence of the  $\alpha_2$  motif together with the intended helical secondary structure. To provide a  $\text{Glu}_4\text{His}_2$  liganding environment for the diiron center, each  $\alpha_2$  subunit contains a  $\text{Glu}_a\text{—Xxx}_b\text{—Xxx}_c\text{—His}_d$  sequence in helix 2, which donates a His side chain (His39/His39') as well as a bridging Glu carboxylate (Glu36/Glu36') to the site. A second Glu carboxylate ligand on helix 1 (Glu10/Glu10') provides a fourth protein ligand per metal ion. 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 different requirements: (1) the geometry of the liganding site was restrained to bind diiron with two bridging Glu carboxylates, two non-bridging Glu side chains, and the N $\delta$  of two His side chains; (2) the helical packing angles and distances were constrained to match those typically observed in the active sites of diiron proteins; (3) precise 2-fold symmetry between the two pairs of helices was enforced.

Each metal ion is five-coordinate, and a sixth vacant site lies on adjoining faces of the two metal ions, providing a potential site for binding of exogenous ligands. Satisfaction of side chain packing requirements was ensured by positioning hydrophobic side chains to fill the core as dictated by the steric environment of the backbone structure. In addition, Glu and Lys residues were chosen for helix-favoring, solvent-accessible sites, to provide water solubility and to drive the assembly into the desired antiparallel topology. Finally, an idealized  $\gamma\text{—}\alpha_L\text{—}\beta$  inter-helical loop was included between the two pairs of helices.

The DF1 sequence was also carefully engineered to include second-shell interactions, which are crucial for defining structural and functional properties of metal-binding sites. Thus, in DF1 a Tyr residue at

position 17 donates a second-shell hydrogen bond to the non-bridging E10' of the other monomer in each of the  $\alpha_2$  subunits (the same interaction exists between the symmetrically related pair, Y17' and E10). Similarly, an Asp residue at position 35' forms a hydrogen bond with the imidazole N $\epsilon$  of the His ligand in the neighboring helix of the dimer. This Asp is further involved in a salt bridge interaction with a Lys at position 38. This hydrogen-bond network consisting of Lys/Asp/His in DF1 is similar to that observed in the active sites of natural proteins, such as methane monooxygenase, where the lysine residue is replaced by an arginine, to form an Arg/Asp/His cluster (see Fig. 4).

DF1 was prepared in good yield by standard solid-phase methodology. The design effectively resulted in the construction of a protein that adopts a folded, native-like helical conformation in aqueous solution, able to bind divalent metal ions such as Zn(II), Co(II), and Fe(II) [31]. The successful assembly of the protein in the desired fold was proven by several techniques. The CD spectrum for both the *apo*- and metal-reconstituted protein is consistent with its proposed helical structure. Protein dimerization was ascertained by analytical ultracentrifugation analysis, which 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).

Fig. 5 shows the visible spectrum of the di-Co(II)—DF1 complex, and Table 1 reports spectroscopic data of several Co(II)-substituted DF derivatives. Data for the di-Co(II) derivative of bacterioferritin are also reported, for comparison. Inspection of Table 1 shows that all Co(II)-substituted DF derivatives exhibit absorption bands around 520 nm, 550 nm, and 600 nm, with associated molar extinction coefficients per Co(II) ion of approximately  $150\text{ M}^{-1}\text{ cm}^{-1}$  for the most intense 550-nm band. These values are in good agreement with those reported in the literature for the Co(II) derivative of bacterioferritin that contains an identical 4-Glu,

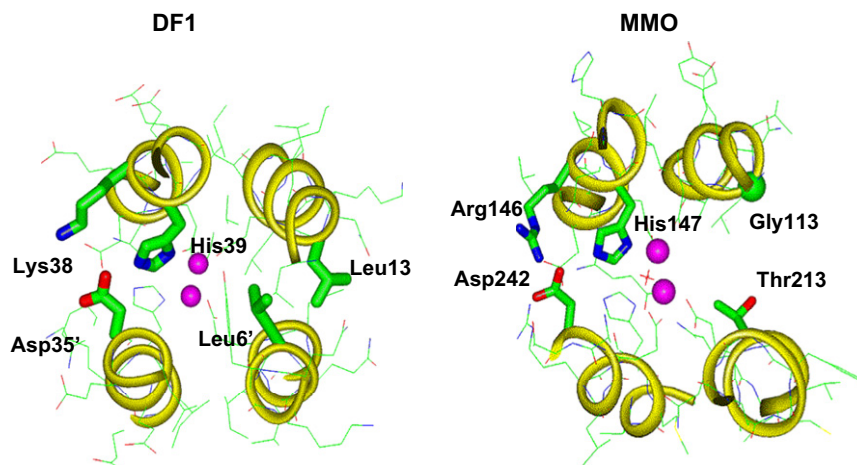


Fig. 4. Hydrogen-bonded network consisting of Lys/Asp/His in DF1 (left; PDB code 1EC5) versus Arg/Asp/His, as observed in MMO (right; PDB code 1MTY).

2-His ligation environment [67]. It is well known that Co(II) complexes show distinctive UV–vis spectra, which are sensitive to the metal coordination geometry and to the nature of the ligands [68]. The molar extinction coefficients of the Co(II) d–d transitions increase as the coordination number decreases. Typical values range from approximately  $10\text{--}20\text{ M}^{-1}\text{ cm}^{-1}$  to  $100\text{--}150\text{ M}^{-1}\text{ cm}^{-1}$  to  $400\text{--}600\text{ M}^{-1}\text{ cm}^{-1}$  for hexa-, penta- and tetra-coordinate complexes, respectively [68]. Thus, for all the di-Co(II)–DF complexes, the positions and intensities of the d–d transitions in the visible region are consistent with a pentacoordinate geometry at the cobalt site.

Detailed information of the overall DF1 structure were obtained from the analysis of the crystal structure of the di-Zn(II) form of the protein (Fig. 6), as well as the NMR structure of the *apo*-form [31,32]. The

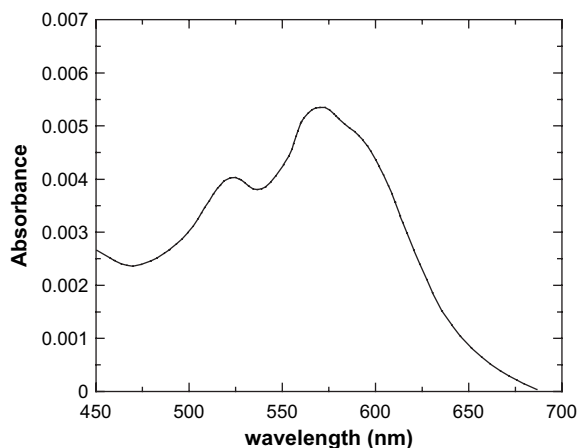


Fig. 5. Visible spectrum of DF1 cobalt complex ( $1.4 \times 10^{-5}\text{ M}$ , MOPS buffer pH = 7.0; 1.0-cm path-length cell).

overall structure is very similar to the intended designed model, consisting of an antiparallel pair of helical hairpins with the desired topology. The structure of the active site is very close to that found in the diferrous and dimanganous forms of naturally occurring proteins [60,69]. Each metal ion is 5-coordinate, with two Glu carboxylates (Glu36 and Glu36'), each interacting with both zinc ions in a 1,3 *syn–syn* bidentate bridging interaction, and two Glu side chains (Glu10 and Glu10'), each binding individual ions in a chelating manner (Fig. 6b). Two His residues complete the ligation environment about the dimetal site, by N $\delta$  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 dimanganous and diferrous forms of natural diiron proteins. The pentacoordinate ligand arrangement with two 1,3 bridging carboxylates and two chelating carboxylates is also nearly identical to that observed in diferrous  $\Delta^9$ -ACP desaturase [58] and di-Mn(II)–bacterioferritin [69].

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 exogenous ligands. Further, the intended second-shell interactions were all actually observed in the experimental structure (Fig. 6b).

DF1 was also characterized in its di-Mn(II) form, by X-ray analysis [35]. The three-dimensional structure is very similar to the di-Zn(II) derivative, both in the overall protein structure, and in the dimetal site coordination geometry.

All the results obtained on the prototype DF1 clearly showed that our approach for the development of

Table 1  
Spectroscopic data of Co(II)-substituted DF derivatives, compared to Co(II)-bacterioferritin

Protein	$\lambda$ (nm)	$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ )	$\lambda$ (nm)	$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ )	$\lambda$ (nm)	$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ )
DF1 <sup>a</sup>	524	143	574	190	594	165
DF2 <sup>b</sup>	520	140	550	155	600	90
DFtet <sup>c</sup>	520	125	550	140	597	80
DFsc <sup>d</sup>	529	124	550	136	600	86
Bacterioferritin <sup>e</sup>	520	126	555	155	600	107

<sup>a</sup> Data from Ref. [31].

<sup>b</sup> Data from Ref. [33].

<sup>c</sup> Data from Ref. [41].

<sup>d</sup> Data from Ref. [96].

<sup>e</sup> Data from Ref. [67].

artificial diiron proteins was successful. Further, the folding of DF1 was dictated neither by the binding of the metal cofactor, nor by the nature of the metal ion. In fact, solution structure determination by NMR analysis of the *apo*-form indicated that the four-helix bundle fold and the metal binding site are largely preorganized in the *apo*-protein [32]. The solution structure of *apo*-DF1 agrees well with the model structure, and it is nearly identical to the crystal structure of di-Zn(II) and di-Mn(II) forms [31,35].

Unfortunately, although DF1 behaves as a native-like protein irrespective of its ligation state, it had several limitations. To constitute the metalloprotein, DF1 has to be denatured and refolded in the presence of metal ions. Further, the diiron complex was not able to bind exogenous ligands. Therefore, the next step was redesigning DF1, with the aim of improving the

properties of the prototype and constructing catalytically active analogs.

### 3. Engineering an active site cavity into DF1

The design of DF1 was particularly challenging, because it required the burial of six polar ionizable groups in the hydrophobic core of a four-helix bundle. To drive the four-helix bundle formation, the interior of DF1 was efficiently packed with a large number of hydrophobic side chains, resulting in a protein with maximal stability. The requirements for a high degree of conformational stability are often contrary to those for binding and enzymatic functions, which in turn require a certain level of conformational freedom.

DF1 was not able to support any activity, probably because access to its dimetal center is blocked by a pair of symmetrically related hydrophobic Leu residues at position 13 of the sequence, L13 and L13'. In order to introduce an active site cavity into the protein, DF1 was re-designed: modeling suggested that replacement of Leu side chains at position 13 and 13' with smaller Ala or Gly residues might open up the active site, providing a cavity large enough to allow small molecules to access the metal center (see Fig. 7). Indeed, natural diiron and dimanganese proteins, with substrate-access channels, tend to have small side chains at these positions. Therefore two DF1 analogs, L13A-DF1 and L13G-DF1, were synthesized [32,34,35,37,38].

The analysis of the folding and binding properties of these new proteins allowed us to outline the stability/function tradeoffs associated with DF models. In order to determine the thermodynamic cost of carving an active site access channel within a protein, the thermodynamics of folding of the three mutants were analyzed [32]. The thermodynamic stabilities of DF1, L13A-DF1, and L13G-DF1 were determined by chemical denaturation with Gdn·HCl, by monitoring the loss of the

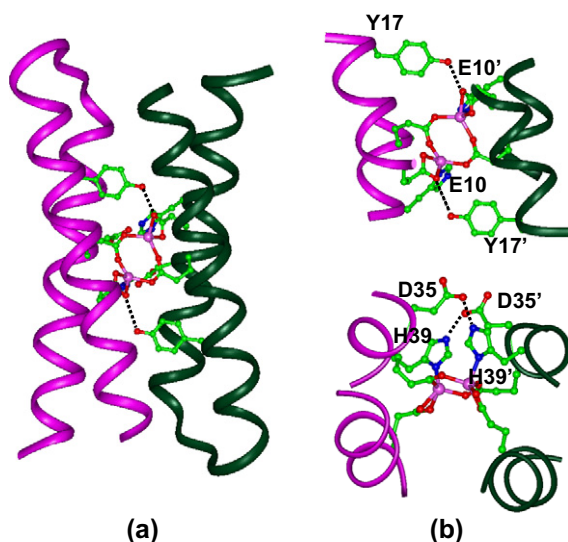


Fig. 6. (a) Crystal structure of di-Zn(II)-DF1. (b) Structure of the binding site, showing the second-shell hydrogen bonds between Tyr17 and Glu10' (top), and between Asp35 and His39' (bottom).

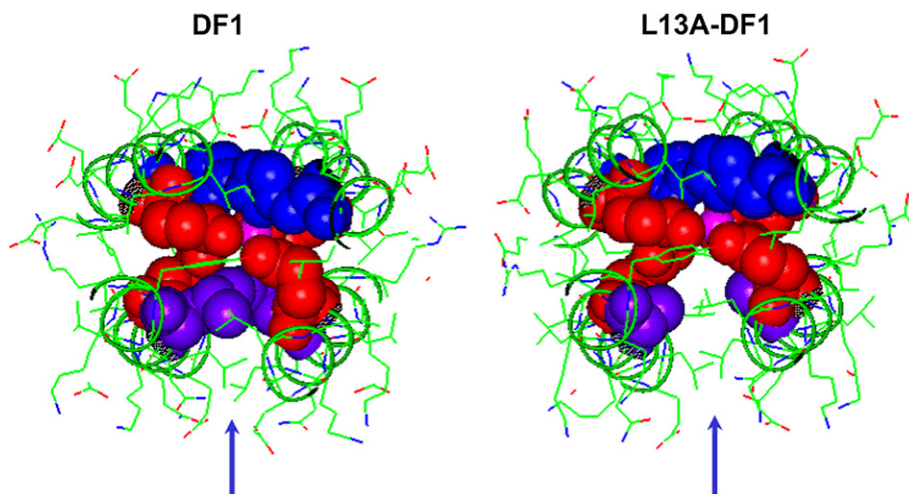


Fig. 7. Engineering an active site cavity into DF1. Top view of the active site cavity in DF1 (*left*) and L13A–DF1 (*right*). The liganding Glu (red spheres) and His (blue spheres) side chains are reported. The replacement of Leu13/13' (violet spheres) with Ala creates an access channel to the dimetal site. (For interpretation of the references to colour in figure legends, the reader is referred to the web version of this article.)

helical CD signal at 222 nm. The Gdn denaturation curves for DF1, L13A–DF1 and L13G–DF1 markedly depend on the total protein concentration, as expected for a monomer–dimer equilibrium. Globally fitting of the thermodynamic parameters to a two-state equilibrium, between folded dimers and unfolded monomers, gave the free energies of dimerization of the three proteins. DF1 has exceptional stability, with a dissociation constant of approximately 0.001 femtomolar (fM); the corresponding values for L13A–DF1 and L13G–DF1 were 100 fM and 0.6 nM. The difference in stability associated with mutation of Leu 13 to Ala [ $\Delta\Delta G(0)$ ] was found to be 5.6 kcal/mol dimer (2.8 kcal/mol monomer). This value is within the range expected for mutating a single buried Leu to an Ala in a native protein [71,72]. The mutation of position 13 from Ala to Gly resulted in a further destabilization of 5.2 kcal/mol (2.6 kcal/mol monomer), reflecting contributions from both a decrease in the helix propensity of Ala relative to Gly [ $\approx 1$  kcal/mol monomer [66,73] and a decreased hydrophobic driving force [ $\approx 1.3$  kcal/mol monomer [72]].

In order to confirm the presence of the designed cavity, the crystal structures of the two mutants were solved. The L13A–DF1 mutant was crystallized as di-Mn(II) complex in the presence of excess of Mn(II) ions [38]. The asymmetric unit contains three dimers, whose structures are nearly identical to that of di-Zn(II)– and di-Mn(II)–DF1. Mn(II) ions are also involved in inter-subunit contacts. Dimers within a given unit cell are linked by five “external” crystallographically independent manganese ions. The electron density

of the metal binding site and the surrounding site is well-defined (Fig. 8). As intended in the design, the di-metal binding site lies at the bottom of a deep pocket water-filled. A trilobate peak of electron density was found in a depression formed between the coordinating glutamate residues. This density has been interpreted as a dimethyl sulfoxide molecule, derived from the crystallization buffer, bridging the two Mn ions. An exogenous bridging organic ligand was observed in the crystal structures of several natural diiron and dimanganese proteins. In addition to DMSO, a number of ordered water molecules fill the pocket leading into the di-Mn(II)

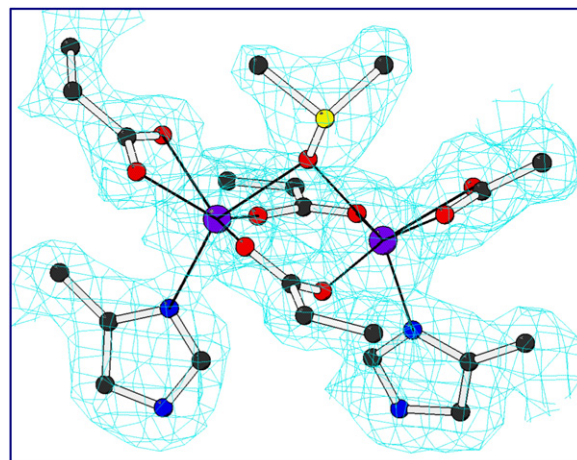


Fig. 8. Electron density map of the dinuclear metal binding site of di-Mn–L13A–DF1 ( $2F_o - F_c$  map (contour levels are  $1\sigma$ ) for the liganding residues including the exogenous DMSO ligand.



site. These water molecules fill the aqueous channel formed by the leucine replacements in DF1. Consistent with the design, all the second-shell interactions were retained in the crystal structure of the di-Mn(II)–L13A–DF1 complex.

The replacement of the two leucine residues in DF1 with the smallest amino acid, glycine, indeed further increases the channel dimensions, as observed by the analysis of the crystal structure of the di-Mn(II)–L13G–DF1 [37]. The structure displays four crystallographically independent dimers in the asymmetric unit (Fig. 9). Despite the structural similarity with di-Mn(II)–L13A–DF1, a significant difference was found in the electron-density map around the dimetal site. Each of the four crystallographically independent dimers has a larger number of water molecules without DMSO molecules in the protein core with respect to the Ala13 DF1 variant. Fig. 10 depicts the active site cavity in L13A–DF1 and L13G–DF1: the water-filled access channel has been expanded in the expected manner, upon substitution of Ala 13/13' with Gly.

The more striking feature found in the di-Mn(II)–L13G–DF1 structure is the presence of two different dimanganese coordination environments (Fig. 11). In three of the four crystallographically independent dimers (AB, CD and EF), a water molecule bridges the

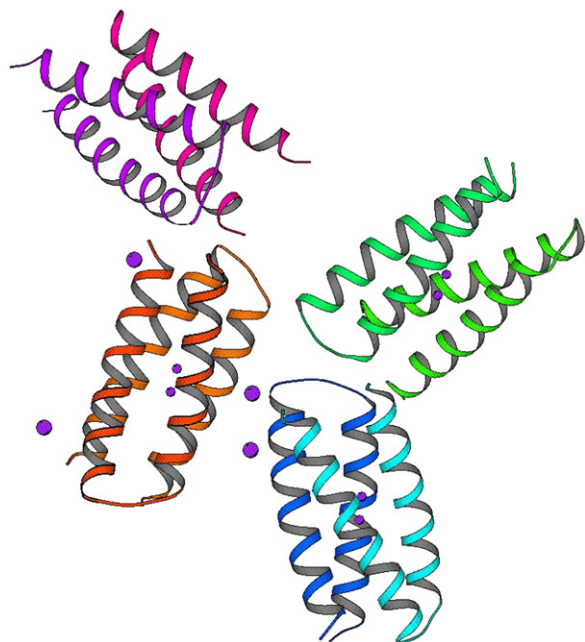


Fig. 9. di-Mn–L13G–DF1 crystal structure: the four crystallographically independent dimers of the asymmetric unit. Violet spheres indicated the Mn(II) ions.

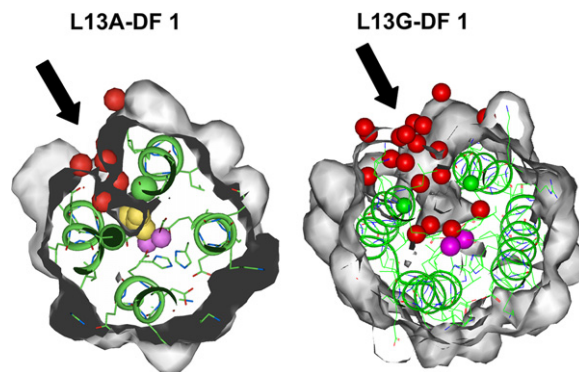


Fig. 10. Top view of the active site cavity in L13A–DF1 (*left*) and L13G–DF1 (*right*). The access channel, indicated by black arrows, expands upon Ala to Gly mutation. The yellow spheres (in L13A–DF1) indicate the dimethyl sulfoxide molecule, the red spheres indicate the water molecules, and the violet spheres indicate the metal ions.

two metal ions, while in the other dimer (GH) two terminal water molecules are coordinated to the two manganese ions *trans* to the histidine ligands. The inter-metal distance is shorter in three of the dimers (3.6 Å) and longer in the fourth (4.2 Å). These different modes of binding led to changes in the quaternary structure (Fig. 12). By superimposing the C-terminal helices (2 and 2') of the four crystallographic dimers, it can be seen that they are virtually invariant among the structures. In contrast, the N-terminal helices (1 and 1') occupy different positions in the GH dimer relative to the others. In particular, the two copies of helix 1 undergo a shift in opposite directions (approximately 0.7 Å) along the Z axis, away from the metal-binding site. Indeed, this shift increases the length of the metal–metal distance. This sliding-helix mechanism may be extended to natural metalloproteins, in order to accommodate changes in their coordination environment. However, such motions have not been observed in the parent natural metalloproteins, probably because it might be more difficult to be observed in a complex, large molecule.

In summary, the three-dimensional structures of the di-Zn(II)–DF1, di-Mn(II)–L13A–DF1 and di-Zn(II)–L13G–DF1 are very similar, indicating that the introduction of the access channel and changing the metal ion from Zn(II) to Mn(II) did not greatly affect the structure of the protein. Further, the analysis of the thermodynamic consequences of introducing a substrate-access channel into the structure of DF1 shows that the destabilization is extreme; the Leu-to-Gly mutation destabilizes the dimer by 10 kcal/mol. Thus, although this functional mutation extracts a large thermodynamic

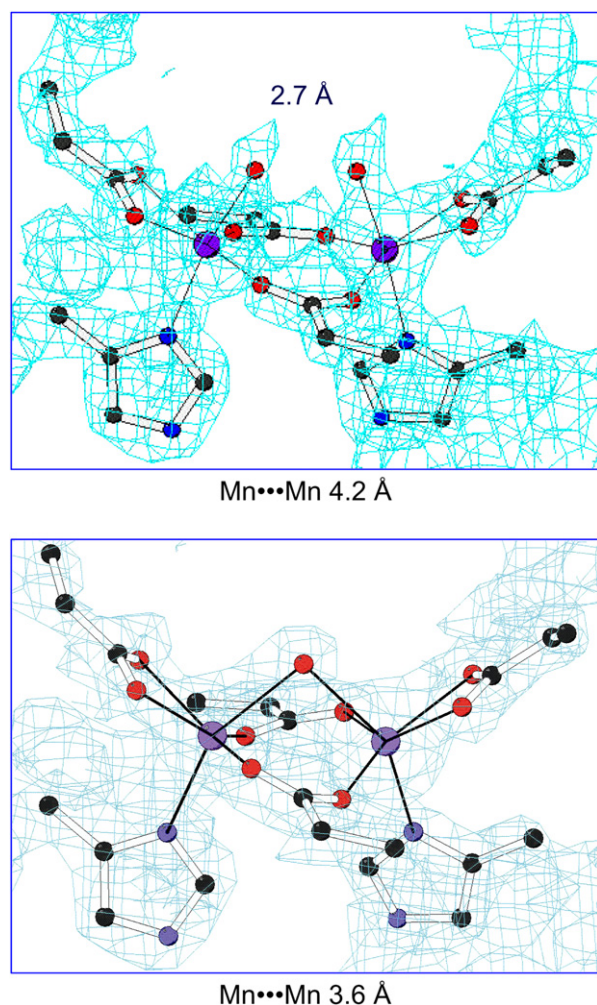


Fig. 11.  $2F_o-F_c$  electron density maps (contour levels are  $1.5\sigma$ ) of the dinuclear metal-binding site of di-Mn–L13G–DF1 in the AB dimer (top) and in the GH dimer (bottom).

price, the extreme stability of DF1 provides adequate stability to compensate for the mutation.

The limited solubility of DF1 analogs in aqueous solution has facilitated the formation of X-ray diffraction-quality crystals, thus allowing a deep investigation of their structures in the solid state; on the other hand, it has impeded further solution characterization in the absence of organic solvents. We therefore tried to improve the properties of DF1 molecules by designing two recombinant versions of L13A–DF1, DF2 and DF2t [33,34,36,39,40], that have several changes in their sequence, as well in the loop structures, with respect to DF1 (see Fig. 13a for a comparison of their amino acid sequences). In order to enhance the water solubility and to reduce the tendency to aggregate in solution, several hydrophobic residues on the surface were converted

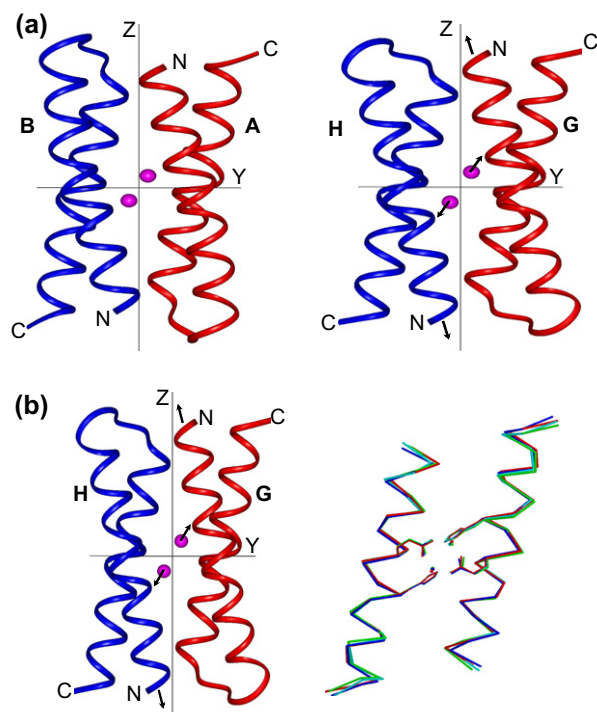


Fig. 12. (a) Crystallographically independent dimers of di-Mn–L13G–DF1, AB (left) and GH (right), viewed along the  $x$  axis of a common orthogonal reference system. The arrows in the GH dimer indicate the shifts of N-terminal helices and metal ions with respect to the other dimers. (b) Superimposition of helices 1 and 1' (left) and helices 2 and 2' (right) of the four crystallographically independent dimers of di-Mn<sup>II</sup>–L13G–DF1. Note the movement of helices 1 and 1' of the GH pair (red) relative to the remaining helices (AB, blue; CD, green; EF, cyan).

to polar residues in DF2. Further, a careful examination of the DF1 structure showed that the inter-helical turn adopts a strained conformation, which might account for these problems. Thus, the loop was re-designed in DF2.

We first performed an analysis of the structures and sequence preferences of inter-helical hairpins, in a large database of crystallographically characterized proteins. A very limited number of inter-helical turns were found to occur with high frequency. Furthermore, these turns occur only within well-defined structural contexts. Three classes of turns were found to occur with high frequency: the Rose and Shellmann  $\alpha_L-\beta$ , which differs for the H-bonding pattern, and the longer  $\alpha_R-\beta-\alpha_R-\beta$ . In DF2, the sequence was changed to stabilize the  $\alpha_L-\beta$  inter-helical turn; in DF2t the longer  $\alpha_R-\beta-\alpha_R-\beta$  loop was introduced between the helices.

DF2 and DF2t were expressed in *Escherichia coli*. The solution properties of DF2 and DF2t were thoroughly studied by sedimentation equilibrium

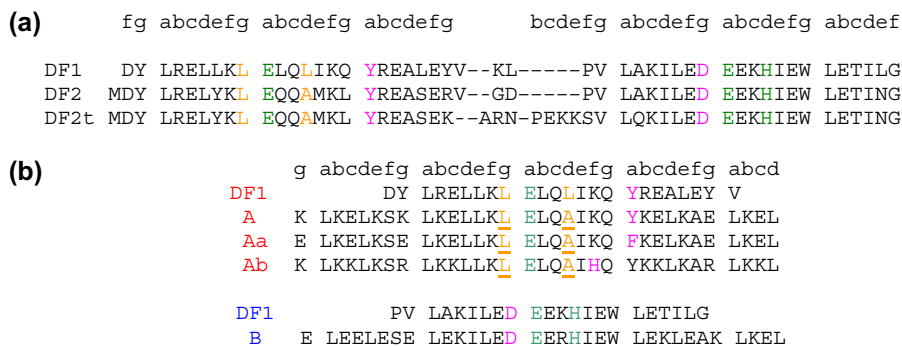


Fig. 13. Sequence alignment of DF family proteins respect to the prototype DF1. The liganding residues are highlighted in green, second-shell ligands in pink, and residues involved in the active site access in yellow. (a) DF2 subset and (b) DF<sub>let</sub> subset: the residues in the A-subunits that were varied are underlined.

ultracentrifugation, CD-monitored guanidine denaturation, and NMR solution characterization [34,40]. Thermodynamic measurements showed that DF2t was slightly more stable than DF2 in the *apo*-state. Both proteins are strongly stabilized by divalent cations, which they bind with a stoichiometry of two ions per four-helix bundle. Both had similar affinities for metal ions, although DF2t bound Mn(II) an order of magnitude greater than DF2 bound this ion.

The structural characterization in solution and in the crystal state of both proteins [36] confirmed that they formed the desired dimeric *helix–loop–helix* structure (Fig. 14). Small differences between the crystal and the solution structure are restricted to the loop regions, with the NMR structure showing two loop conformational families, in both the proteins. Analysis of the loop regions revealed that the 3-residue loop,  $\beta$ – $\alpha_L$ – $\beta$  loop, engineered in DF2t, was more defined than the 2-residue  $\alpha_L$ – $\beta$  loop, present in DF1 and DF2 (Fig. 15). This finding may be responsible for the higher stability and the better expression level of DF2t with respect to DF2.

With the aim of correlating structural with functional properties, we examined the ligand binding and spectroscopic properties of all the DF1 and DF2 analogs. In contrast to the prototype DF1, all the other members of DF1 and DF2 subsets, containing either Ala or Gly at the access site position, can be reconstituted with metal ions without unfolding the *apo*-protein, thus suggesting that the metal binding site is readily accessible to small molecules. All the DF1 analogues showed an absorption spectrum of the diferric form typical of oxo-bridged diferric systems, and all the proteins are able to bind azide (Fig. 16). Further, the UV–vis spectrum shows small changes upon addition of acetate, benzoate or 4-hydroxybenzoate anions, suggesting that the carboxylate groups of these ligands interact

with the metal center, similarly to natural diiron–oxo proteins [38].

This last finding was confirmed by the crystal structure resolution of the diferric complex of DF2t [39]. The diiron site is characterized by a cluster consisting of two  $\mu$ -1,3 bridging Glu carboxylates, two chelating Glu carboxylates and two N $\delta$  bound His ligands (Fig. 17). Further, the diiron(III) center is able to accommodate two additional non-protein ligands. In fact, two exogenous ligands were apparent in the electron density map: an oxo ligand was modeled at the  $\mu$ -bridging position, *cis* to both Fe–N $\delta$ His bonds, and an acetate ion, from the crystallization buffer, was modeled in the other

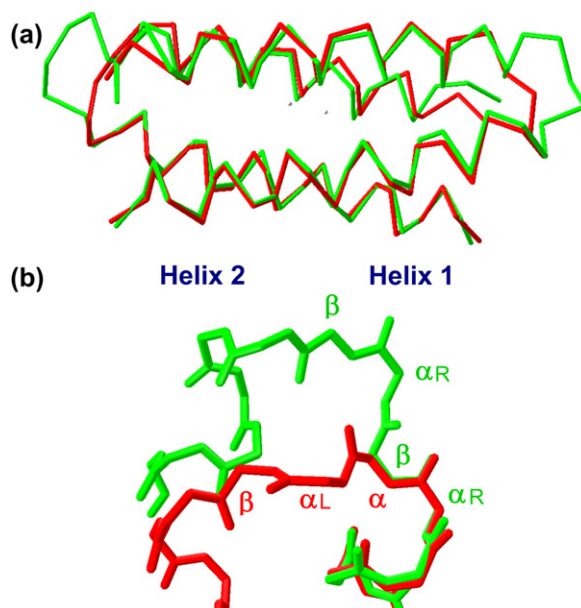


Fig. 14. (a) Comparison of DF2 (red) and DF2t (green) crystal structures. (b) Backbone structure of the turns in DF2 (red) and DF2t (green).

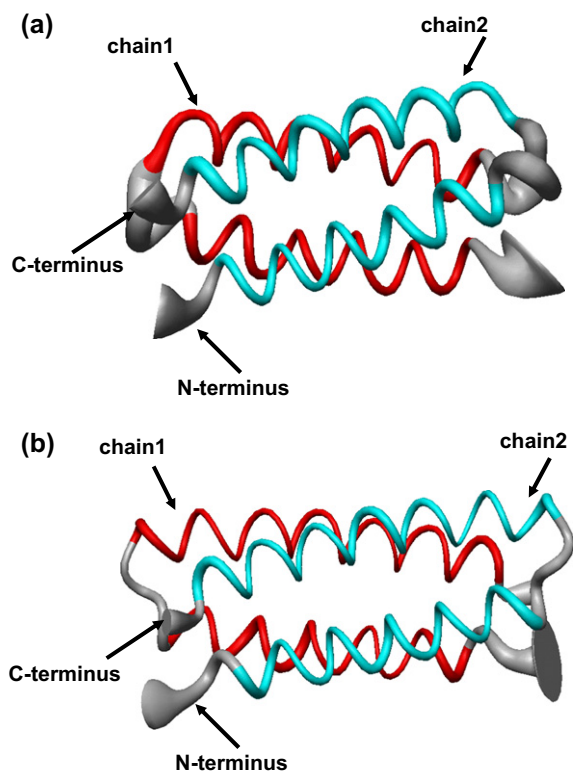


Fig. 15. Solution structure of (a) di-Zn(II)–DF2 and (b) di-Zn(II)–DF2t, represented as a tube of radius proportional to the rmsd of each residues. The thinner segment of the ribbon in the DF2t loop region evidences a comparatively lower positional variability of the backbone atoms with respect to that observed in DF2 (thicker region).

electron density peak. An extensive series of second-shell hydrogen bonds restrain the protein Glu ligands to nearly the same positions as those observed in divalent metal-bound DF proteins, despite the accommodation of exogenous  $\mu$ -oxo and acetate ligands. Thus, the diferric

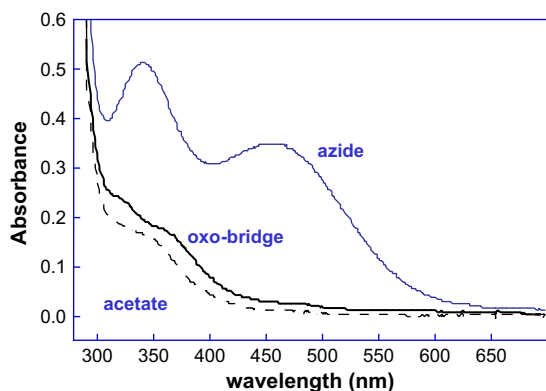


Fig. 16. Spectrum of di-Fe(III)–DF2 (solid black lines) and spectral changes that accompany the addition of the small molecules sodium azide (blue line) and sodium acetate (black dashed line).

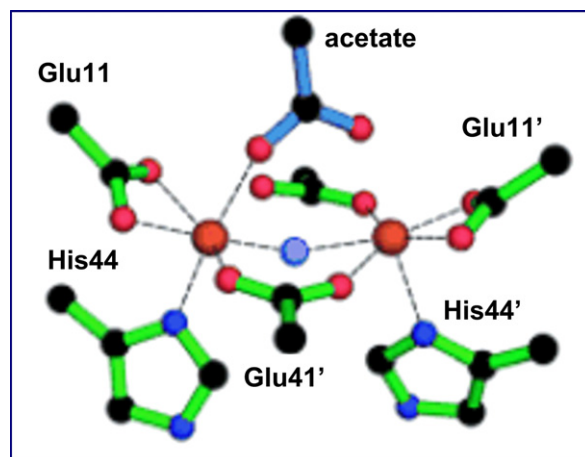


Fig. 17. The diiron(III) cluster ligands and  $\sigma_A$ -weighted Fo–Fc omit map peaks modeled by acetate and  $\mu$ -oxoligands ( $4.0\sigma$ ) as observed in the di-Fe(III)–DF2t crystal structure.

complex of DF2t does not show the same degree of flexibility of the carboxylate ligands, typically found in the natural diiron proteins and thought to be important for catalysis. In fact, changes in the coordination mode of active site ligands in natural diiron proteins are well predated, as they have been structurally characterized in methane monooxygenase and in the R2 subunit of ribonucleotide reductase [74,59]. In all cases, the ligating modes of the active site carboxylate ligands have been shown to be flexible (carboxylate shifts) [70,75]. This flexibility is not surprising, as carboxylate ligands are able to adopt several coordination modes.

In light of the catalytic relevance of the carboxylate shift phenomenon [76], a theoretical analysis was carried out on the di-Zn(II)–DF1 active site, in order to investigate the structural and dynamical behavior of the metal cofactor. First principles (Car-Parrinello) [77] and hybrid QM/MM molecular dynamic simulation [78] indicated that carboxylate shift from chelated to monodentate occurs during the simulations, at least on one of the two bridging glutamates. The calculations also showed that the second-shell interactions contribute significantly to the structural stability of the active site, and that the bulk solvent water molecules play a critical role in fine-tuning the dynamic of the system.

#### 4. Searching for functionality: DF<sub>tet</sub> an heterotetrameric system

The results we had obtained so far on the analysis of all DF analogs clearly indicated that three main aspects must be considered in order to obtain stable and functional models: (i) the correct design of inter-helical

turn; (ii) the residues responsible for the primary and second-shell ligands; (iii) the residues that define the active site access channel.

To search for a catalytically active system based on the DF structure, it is helpful to evaluate how systematic changes in the structure affect substrate-binding and catalytic properties.

DF1 is an extremely stable scaffold to tolerate significant changes in the amino acid sequence, without affecting its overall fold. The extreme thermodynamic stability of DF1 makes it a particularly apt framework for substitutions, which are important for function and for probing specific mechanistic questions, more difficult to analyze on a complex, large protein.

Although some DF molecules in the diferric forms are able to bind exogenous ligands, such as phenol and acetate, catalytic activity still requires changes in the second-shell ligands, in particular in defining the access to the diiron site. For example, modeling suggests that for the accessibility to the metal site, not only residues at position 13 play a critical role, as already described, but also residues at position 9 should be crucial. However, as the number of mutation increases, a very large number of combinations are generated. Thus an exhaustive analysis would require the preparation and purification of significant quantities of hundreds to thousands of variants, to be screened for binding of small molecules, and reactivity toward a variety of substrates.

To advance this goal, we tried to bridge the gap between “rational design” and combinatorial approaches through the development of four-helix bundle based on heterotetrameric systems [41]. A heterotetrameric system consisting of four disconnected helices, which could be separately synthesized, purified, and then combinatorially assembled, is uniquely suited for the production of an array of any desired helical bundles from a significantly smaller number of peptides. For an ABCD heterotetramer (in which each of the monomers was different),  $10^4$  combinations could be prepared by combining only 10 variants of each of the individual chains. Previously, strategies have been developed for the design of two-stranded coiled coils that specifically self-assemble to form AB heterodimers, with minimal formation of homodimeric species [79]; similarly, ABC heterotrimers [80–82] and  $A_2B_2$  heterotetrameric [83] coiled coils have been designed.

We have applied this strategy to the development of DF1 analogs, by designing either an  $A_2B_2$  heterotetrameric complex [41] and a three-component  $AA'B_2$  heterotetrameric complex [42]. We named these analogs

$DF_{tet}$ . The design of heterotetrameric systems presents significant challenges for protein design. Specifying a unique topology is a more complex problem for a heterotetrameric system, as in  $DF_{tet}$ , than it was in a homodimeric system, such as DF1 or DF2. In particular, it was important to destabilize both homooligomeric folds as well as undesired heterotetrameric topologies. To achieve this goal, a new computational design algorithm was developed. This design methodology included the concept of “negative design” [84–88,30] to prevent alternate topologies from occurring and as well as a “minimalist” approach [88,89] to minimize extraneous structural variables.

The algorithm was first applied for the design of  $A_2B_2$  heterotetramers (Fig. 13b), where A and B are 33-residue peptides. The first step in the design process involved the specification of the four-helix bundle backbone, which was generated using the parameters obtained from the analysis of DF1. The helices of  $DF_{tet}$  were extended relative to those of DF1 (33 residues in  $DF_{tet}$  versus 24 residues in DF1) to increase the stability of the system. Helix A contains only the Glu ligands, and helix B contains the EXXH metal-binding motif. By inspection of the sequence alignment reported in Fig. 13b, it can be seen that the residues within the region of the binding site are the same as the corresponding residues of DF1. Similarly to L13A–DF1, an Ala residue was inserted at position 19 (position 13 in DF1) to create an open cavity. Residue 18 in  $DF_{tet}$ –B (corresponding to residue 38 of DF1) was mutated from Lys to Arg, in order to strictly resemble the Asp–Glu–Xxx–Arg–His motif of natural diiron proteins such as MMO [90] and  $\Delta^9$ -ACP desaturase [60] (see Fig. 4).

The nature of the residues at the b/e and g/c interfacial positions (see Fig. 2 for the heptad nomenclature) was chosen to specifically stabilize only one of the possible topologies for an antiparallel  $A_2B_2$  heterotetramer. These positions were allowed to correspond to Lys and Glu amino acids, and the energy of the desired folds and of the alternatively folded structures was considered in the calculation. The final sequences for A and B peptides correspond to those which showed the minimum number of unfavorable contacts in the desired topology (Fig. 18).

The design of  $DF_{tet}$  was successful:  $DF_{tet}$ –A and  $DF_{tet}$ –B are unfolded at neutral pH; however, when mixed in a 1:1 ratio, they assemble into a helical bundle with the expected stoichiometry [41]. Furthermore, the peptide is tetrameric and stable over a wide temperature range, being thermally unfolded only at low concentrations and at an elevated pH.  $DF_{tet}$  also binds metal ions

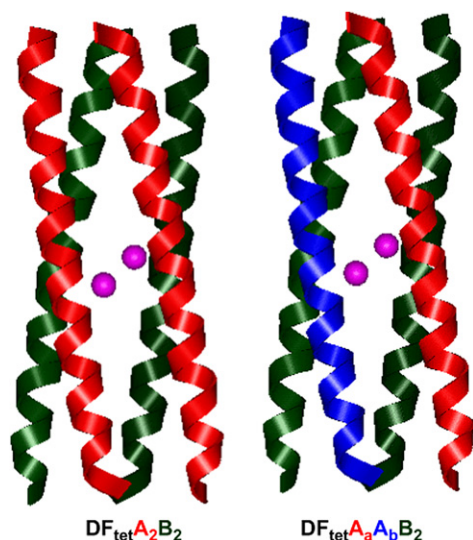


Fig. 18. Models of DF<sub>tet</sub>.

in the proper stoichiometry, and the spectroscopic properties of the Co(II) and Fe(III) derivatives strongly suggest that the ligand environment is precisely as intended. The match of the Co(II) spectrum between DF<sub>tet</sub> and DF1/DF2 analogs (see Table 1), and the identical kinetic scheme for Fe(II) oxidation provide good evidence that the metal-binding sites of these proteins are very similar. Thus, the intended topology appears to have been achieved.

The potentiality of DF<sub>tet</sub> for the preparation and screening of a library of DF variants has been demonstrated by the results we achieved with a less symmetrical version of the original heterotetramer DF<sub>tet</sub>. DF<sub>tet</sub> was re-designed for constructing a new system based on a three-component AA'B<sub>2</sub> complex [42]. In order to prepare this complex, two electrostatically complementary versions of the A peptide were prepared, which should specifically assemble with two equivalents of B to form the desired AA'B<sub>2</sub> heterotetramer. Along one face of the helix of the original A, negatively and positively charged groups had been placed near the N and C termini, respectively, to stabilize an antiparallel arrangement of the helices. The sequences of these two helices were varied such that one contains a strip of Glu along the entire length of the interface, and the other contains Lys and Arg at equivalent positions (Fig. 13b). The peptide with the acidic replacements was named helix A<sub>a</sub> and the other helix A<sub>b</sub>. Additional substitutions were also made in the second-shell interactions, to probe specific mechanistic questions. The non-bridging Glu residues in DF1 form second-shell interactions with two Tyr side chains (see Fig. 6), whereas in natural proteins only one of the structurally

equivalent Glu residues is generally constrained by a second-shell interaction.

The assembly of the three peptide chains, A<sub>a</sub>, A<sub>b</sub>, and B in the expected 1:1:2 stoichiometry to give the desired A<sub>a</sub>A<sub>b</sub>B<sub>2</sub> heterotetramer was successful, as proven both by CD and size-exclusion chromatography studies [42]. The substitutions made to the A<sub>a</sub>A<sub>b</sub>B<sub>2</sub> four-helix bundle, with respect to the prototype DF1, were interesting, as demonstrated by the results we obtained on the diiron complex of this new protein. On addition of ferrous ions and oxygen, the protein forms a product, entirely different from those formed in the original DF1 protein and DF<sub>tet</sub> complex.

Fig. 19 shows the UV–vis spectra of A<sub>a</sub>A<sub>b</sub>B<sub>2</sub> measured at various time after the addition of Fe(II) in the presence of oxygen. The spectra are characterized by the presence of a strong band centered near 625 nm with an extinction coefficient of approximately 1000 M<sup>-1</sup> cm<sup>-1</sup>. Although additional spectroscopic studies are necessary to confirm the identity of this species, its UV–vis spectrum is similar to that found in the *cis*-μ-1,2 peroxo-bridged diferric intermediates observed in a variety of diiron proteins, such as ferritin (λ<sub>max</sub> = 650 nm, ε = 1000 M<sup>-1</sup> cm<sup>-1</sup>) [91,92] but at somewhat shorter wavelength than the corresponding species in the enzymes Δ<sup>9</sup>-ACP desaturase (λ<sub>max</sub> = 700 nm, ε = 1200 M<sup>-1</sup> cm<sup>-1</sup>) [93], methane monooxygenase (λ<sub>max</sub> = 725 nm, ε = 1800 M<sup>-1</sup> cm<sup>-1</sup>) –for a review, see Ref. [75]–and a mutant of the R2

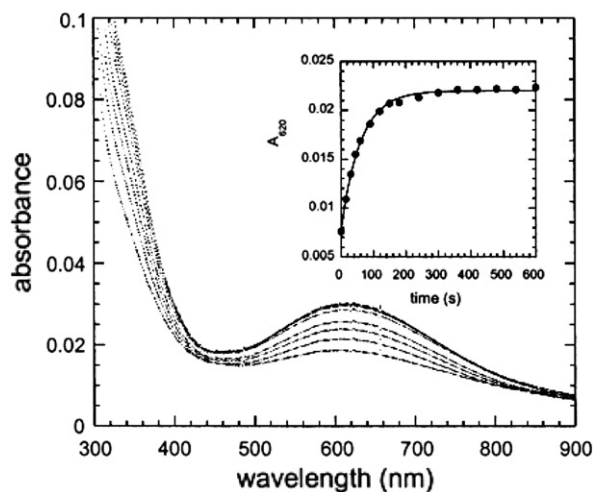


Fig. 19. Spectra of A<sub>a</sub>A<sub>b</sub>B<sub>2</sub> (25 μM in tetramer) measured at various times after the addition of ferrous ammonium sulfate, final concentration 50 μM, in air saturated buffer (150 mM Mes, pH 6.5/150 mM NaCl). The spectra were measured at 15, 30, 45, 60, 90, 120, 150, and 180 s after mixing. *Inset*: the kinetics of peroxide formation followed by the increase in absorbance at 620 nm.

subunit of ribonucleotide reductase ( $\lambda_{\max} = 700 \text{ nm}$ ,  $\epsilon = 1800 \text{ M}^{-1} \text{ cm}^{-1}$ ) [94].

The visible band could be also assigned to a tyrosinate–Fe(III) ligand-to-metal charge-transfer band [95]. Further studies are currently in progress.

Further design refinements to DF<sub>tet</sub> allowed us to develop an O<sub>2</sub>-dependent phenol oxidase [43]. The interior of the symmetrical DF<sub>tet</sub> A<sub>2</sub>B<sub>2</sub> was sculpted for creating a pocket capable of binding a substrate, the 4-aminophenol. Modeling suggested that when this molecule is inserted in the active site of the original DF<sub>tet</sub> A<sub>2</sub>B<sub>2</sub>, and modeled such that its phenolic oxygen bridges the metal ions, it makes unfavorable contacts with Leu 15 and Ala 19 (corresponding to positions 9 and 13, respectively, of DF1) (see Fig. 20a). The steric bulk of these residues was therefore reduced in variants in which Ala 19 is changed to Gly and Leu 15 is changed to either Ala or Gly in both of the A chains (see Fig. 20b).

A number of symmetrical variants, G<sub>4</sub>–DF<sub>tet</sub> (in which Leu 15 and Ala 19 of both A chains were substituted with Gly), L<sub>2</sub>G<sub>2</sub>–DF<sub>tet</sub> (in which Leu 15 was retained and Ala 19 was changed to Gly), A<sub>2</sub>G<sub>2</sub>–DF<sub>tet</sub> and G<sub>2</sub>A<sub>2</sub>–DF<sub>tet</sub> (which have Gly or Ala at the indicated positions) were screened for activity. In particular, all the variants were screened for their ability to react with Fe(II) and O<sub>2</sub>. Some variants produced a relatively stable intermediate, resembling the same diferric peroxo species formed in the DF<sub>tet</sub> A<sub>a</sub>A<sub>b</sub>B<sub>2</sub>. The variant with the fewest steric restrictions, namely G<sub>4</sub>–DF<sub>tet</sub>, showed increasingly rapid formation of the oxo-species with no detectable intermediates. The same variant showed the greatest binding ability toward phenol, which binds to the diferric site with high affinity.

Finally, we measured the two-electron oxidation of 4-aminophenol to the corresponding quinone monimine, catalyzed by the proteins in atmospheric O<sub>2</sub>. Again, G<sub>4</sub>–DF<sub>tet</sub> showed a  $\approx 1000$ -fold rate enhancement, relative to the background reaction when the initial rate of the reaction in the presence and absence of the protein was compared. The G<sub>4</sub>–DF<sub>tet</sub> catalyzed this reaction for at least 100 turnovers. Changing either of the Gly residues at positions 19 or 15 to Ala gave a protein whose rate was decreased between 2.5- and 5-fold. All the results on DF<sub>tet</sub> variant clearly showed that the catalytic efficiency is sensitive to changes in the size of a methyl group in the protein, illustrating the specificity of the design.

All these data demonstrate the promise of the combinatorial approach for identifying potential candidates for defined applications. Having tested the potential of the system, the next step is to prepare a library of

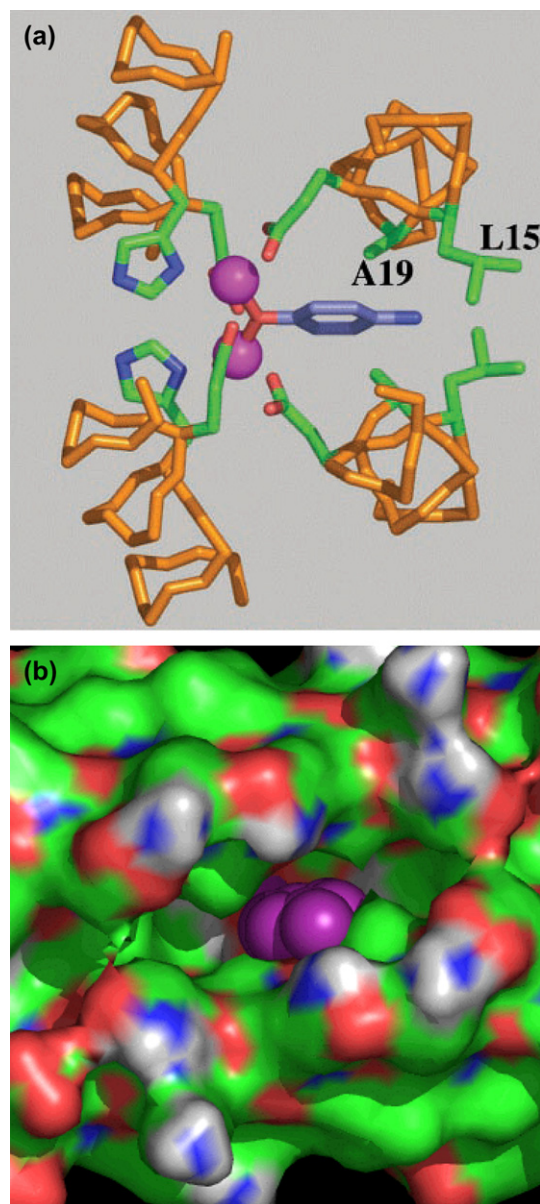


Fig. 20. (a) Computer model of 4-aminophenol bound to the active site of DF<sub>tet</sub>A<sub>2</sub>B<sub>2</sub>. (b) Solvent-accessible surface associated with G<sub>4</sub>–DF<sub>tet</sub>. The aminophenol ring into the pocket is represented in purple.

variants, with the active site cavity designed for a specific functionality.

## 5. Conclusions

Principles and methods for protein design have matured, and protein design has deeply contributed to our understanding of the principles of protein folding and stabilization. It has now become possible to design

structurally defined protein models. Our studies in the design of diiron proteins suggest that it should now be possible to approach the de novo design of highly functionalized metalloproteins with well-defined three-dimensional structures.

We applied different and complementary strategies for the construction of diiron proteins' model systems. To create the prototype scaffold we followed the symmetry concept, trying to reproduce in a symmetric model the quasi-symmetrical structure of a metalloprotein. The use of  $C_2$  symmetry has been shown to be particularly advantageous, because it simplifies the design, reduces the size of the molecules to be synthesized, and simplifies their structural characterization.

The  $C_2$  symmetric DF1 protein, based on the dimerization of a *helix–loop–helix* motif, has set the stage for the development of the other proteins. The studies on DF1 analogs and on the more soluble versions DF2 and DF2t, revealed the crucial features required for defining the properties of an artificial diiron protein. What we have learned by the lesson on DF molecules based on the *helix–loop–helix* motif can be summarized as follows.

The designed four-helix bundle motif allows the proper burial of the first-coordination-sphere ligands. All of our designed DF proteins bind divalent metal ions in the proper stoichiometry of two per protein. Further, all the spectroscopic and structural data suggest that all these proteins share the same pentacoordinate ligand environment, with each metal ion having three carboxylate ligands and one His ligand.

Of crucial importance in modulating the functional properties of the proteins are the residues at positions “g” and “d” that define the shape and accessibility of the active site cavity (see Fig. 2). Residues with different steric hindrance at the above-mentioned positions influence both stability and activity. The analysis of the thermodynamics of folding showed that stability decreases as the bulk of the residues at positions “d” decreases from Leu to Ala to Gly. On the opposite, as the active site cavity increases the chemical properties of the proteins increasingly resembles those of the natural enzymes. In fact, L13G-DF1 and L13A-DF1 variants, as well as the DF2 subset, bind exogenous ligands and display ferroxidase activity, in contrast to DF1, which was extremely stable and not able to support any function. Further evidences of the correlation between the catalytic activity and the size of the active site cavity came from the results obtained on the DF<sub>tet</sub> subset, in which the influence of the residues at the “g” positions was also evaluated. The results obtained on the variant with the fewest steric restrictions, namely G<sub>4</sub>-DF<sub>tet</sub>,

reveal that changes, as small as the size of a single methyl group in the active site residues, have a significant effect on the protein functional properties. In fact, G<sub>4</sub>-DF<sub>tet</sub> shows exceptional substrate specificity: the substrate 4-methoxyaniline, in which the hydroxyl group of 4AP is converted to a methoxy, was not a substrate for the protein. Also, 4-aminoaniline, in which the phenolic hydroxyl is replaced by an amino group, was oxidized at a rate only 2-fold greater than the background reaction.

All the results we obtained so far were included in the design of the last protein, DF3, which belongs to the DF1 subset. This new symmetric *helix–loop–helix* dimer encompasses all the main features thought to be crucial for function: (i) Gly residue at both positions 9 and 13; (ii) new inter-helical loop with respect to DF1. The new protein shows improved solubility and active site accessibility, while retaining the unique native-like structure, as assessed by NMR structural characterization (manuscript in preparation).

In conclusion, throughout this paper we tried to illustrate the importance of using a very stable and highly optimized protein model for functional design, which can withstand multiple changes into the sequence without affecting the overall fold. DFs encompass these properties: their thermodynamic stability makes them apt framework for substitutions, which are important for function and for probing specific mechanistic questions, more difficult to analyze on a complex, large protein.

## Acknowledgments

The authors wish to thank the European Commission for a Marie Curie postdoctoral fellowship to RTMT.

## References

- [1] K.D. Karlin, *Science* 261 (1993) 701.
- [2] Y. Lu, *Inorg. Chem.* 45 (2006) 9930.
- [3] Y. Lu, in: R.B. King (Ed.), 2nd ed., *Encyclopedia of Inorganic Chemistry*, Vol. V, Wiley, Chichester, 2006, p. 3159.
- [4] H.W. Hellinga, *Curr. Opin. Biotechnol.* 7 (1996) 437.
- [5] H.W. Hellinga, *Fold. Design* 3 (1998) R1.
- [6] W.F. DeGrado, C.M. Summa, V. Pavone, F. Nastri, A. Lombardi, *Annu. Rev. Biochem.* 68 (1999) 779.
- [7] L. Baltzer, *Topics Curr. Chem.* 202 (1999) 39.
- [8] L. Regan, *Trends Biochem. Sci.* 20 (1995) 280.
- [9] Y. Lu, J.S. Valentine, *Curr. Opin. Struct. Biol.* 7 (1997) 495.
- [10] S.J. Lippard, J.M. Berg, *Principles of Bioinorganic Chemistry*, Univ. Sci. Books, Mill Valley, CA, 1994.
- [11] R.H. Holm, P. Kennepohl, E.I. Solomon, *Chem. Rev.* 96 (1996) 2239.
- [12] Y. Lu, S.M. Berry, T.D. Pfisterand, *Chem. Rev.* 101 (2001) 3047.



- [13] T.M. Penning, J.M. Jez, *Chem. Rev.* 101 (2001) 3027.
- [14] W.B. Tolman, D.J.E. Spencer, *Curr. Opin. Chem. Biol.* 5 (2001) 188.
- [15] E.Y. Tshuva, S.J. Lippard, *Chem. Rev.* 104 (2004) 987.
- [16] P.V. Rao, R.H. Holm, *Chem. Rev.* 104 (2004) 527.
- [17] M. Costas, M.P. Mehn, M.P. Jensen, L. Que, *Chem. Rev.* 104 (2004) 939.
- [18] L. Baltzer, H. Nilsson, J. Nilsson, *Chem. Rev.* 101 (2001) 3153.
- [19] D.A. Moffet, M.H. Hecht, *Chem. Rev.* 101 (2001) 3191.
- [20] F. Nastro, A. Lombardi, G. Morelli, O. Maglio, G. D'Auria, C. Pedone, V. Pavone, *Chem. Eur. J.* 3 (1997) 340.
- [21] G. D'Auria, O. Maglio, F. Nastro, A. Lombardi, M. Mazzeo, G. Morelli, L. Paolillo, C. Pedone, V. Pavone, *Chem. Eur. J.* 3 (1997) 350.
- [22] A. Lombardi, F. Nastro, M. Sanseverino, O. Maglio, C. Pedone, V. Pavone, *Inorg. Chim. Acta* 301 (1998) 275.
- [23] F. Nastro, A. Lombardi, G. Morelli, C. Pedone, V. Pavone, G. Chottart, P. Battioni, D. Mansuy, *J. Biol. Inorg. Chem.* 3 (1998) 671.
- [24] L. Di Costanzo, S. Geremia, L. Randaccio, F. Nastro, O. Maglio, A. Lombardi, V. Pavone, *J. Biol. Inorg. Chem.* 9 (2004) 1017.
- [25] A. Lombardi, F. Nastro, D. Marasco, O. Maglio, G. De Sanctis, F. Sinibaldi, R. Santucci, M. Coletta, V. Pavone, *Chem. Eur. J.* 9 (2003) 5643.
- [26] A. Lombardi, F. Nastro, V. Pavone, *Chem. Rev.* 101 (2001) 3165.
- [27] A. Lombardi, D. Marasco, O. Maglio, L. Di Costanzo, F. Nastro, V. Pavone, *Proc. Natl Acad. Sci. USA* 97 (2000) 11922.
- [28] V. Nanda, M.M. Rosenblatt, A. Osyczka, H. Kono, Z. Getahun, P.L. Dutton, J.G. Saven, W.F. DeGrado, *J. Am. Chem. Soc.* 127 (2005) 5804.
- [29] J.R. Calhoun, F. Nastro, O. Maglio, V. Pavone, A. Lombardi, W.F. DeGrado, *Peptide Sci.* 80 (2005) 264.
- [30] B. Hill, D.P. Raleigh, A. Lombardi, W.F. DeGrado, *Acc. Chem. Res.* 33 (2000) 745.
- [31] A. Lombardi, C. Summa, S. Geremia, L. Randaccio, V. Pavone, W.F. DeGrado, *Proc. Natl Acad. Sci. USA* 97 (2000) 6298.
- [32] O. Maglio, F. Nastro, V. Pavone, A. Lombardi, W.F. DeGrado, *Proc. Natl Acad. Sci. USA* 100 (2003) 3772.
- [33] A. Pasternak, J. Kaplan, J.D. Lear, W.F. DeGrado, *Protein Sci.* 10 (2001) 958.
- [34] O. Maglio, F. Nastro, J.R. Calhoun, S. Lahr, V. Pavone, W.F. DeGrado, A. Lombardi, *J. Biol. Inorg. Chem.* 10 (2005) 539.
- [35] S. Geremia, L. Di Costanzo, L. Randaccio, D.E. Engel, A. Lombardi, F. Nastro, W.F. DeGrado, *J. Am. Chem. Soc.* 127 (2005) 17266.
- [36] S.J. Lahr, D.E. Engel, S.E. Stayrook, O. Maglio, B. North, S. Geremia, A. Lombardi, W.F. DeGrado, *J. Mol. Biol.* 346 (2005) 1441.
- [37] W.F. DeGrado, L. Di Costanzo, S. Geremia, A. Lombardi, V. Pavone, L. Randaccio, *Angew. Chem., Int. Ed.* 42 (2003) 417.
- [38] L. Di Costanzo, H. Wade, S. Geremia, L. Randaccio, W.F. DeGrado, V. Pavone, A. Lombardi, *J. Am. Chem. Soc.* 123 (2001) 12749.
- [39] H. Wade, S.E. Stayrook, W.F. DeGrado, *Angew. Chem., Int. Ed.* 45 (2006) 1.
- [40] P. Wei, A.J. Skulan, H. Wade, W.F. DeGrado, E.I. Solomon, *J. Am. Chem. Soc.* 127 (2005) 16098.
- [41] C.M. Summa, M.M. Rosenblatt, J.K. Hong, J.D. Lear, W.F. DeGrado, *J. Mol. Biol.* 321 (2002) 923.
- [42] E.N. Marsh, W.F. DeGrado, *Proc. Natl Acad. Sci. USA* 99 (2002) 5150.
- [43] J. Kaplan, W.F. DeGrado, *Proc. Natl Acad. Sci. USA* 101 (2004) 11566.
- [44] A.L. Feig, S.J. Lippard, *Chem. Rev.* 94 (1994) 759.
- [45] B.J. Waller, J.D. Lipscomb, *Chem. Rev.* 96 (1996) 2625.
- [46] P. Nordlund, B.-M. Sjöberg, H. Eklund, *Nature* 345 (1990) 593.
- [47] S.J. Lange, L. Que, *Curr. Opin. Chem. Biol.* 2 (1998) 159.
- [48] K.K. Andersson, A. Gräslund, *Adv. Inorg. Chem.* 43 (1995) 359.
- [49] S.C. Gallagher, A. George, H. Dalton, *Eur. J. Biochem.* 254 (1998) 480.
- [50] F. Bouvier, Y. Keller, A. D'Harlinghe, B. Cammara, *Biochim. Biophys. Acta* 1391 (1998) 320.
- [51] M. Lee, M. Lenman, A. Banas, M. Bafor, S. Singh, M. Schweizer, R. Nilsson, C. Liljeborg, A. Dahlqvist, P.O. Gummeson, S. Sjö Dahl, A. Green, S. Stymne, *Science* 280 (1998) 915.
- [52] J.M. Bollinger, E. Hidalgo, *Methods Enzymol.* 258 (1995) 278.
- [53] R.E. Stenkamp, *Chem. Rev.* 94 (1994) 715.
- [54] A.K. Powell, D.A. Fish, in: H. Sigel (Ed.), *Metal Ions in Biological Systems*, vol. 35, Marcel Dekker, Inc., 1998, p. 515.
- [55] N.D. Chasteen, in: H. Sigel (Ed.), *Metal Ions in Biological Systems*, vol. 35, Marcel Dekker, Inc., 1998, p. 479.
- [56] P.M. Harrison, et al., in: H. Sigel (Ed.), *Metal Ions in Biological Systems*, vol. 35, Marcel Dekker, Inc., 1998, p. 435.
- [57] J.F. Briat, S. Lobreaux, in: H. Sigel (Ed.), *Metal Ions in Biological Systems*, vol. 35, Marcel Dekker, Inc., 1998, p. 563.
- [58] C.M. Summa, A. Lombardi, M. Lewis, W.F. DeGrado, *Curr. Opin. Struct. Biol.* 9 (1999) 500.
- [59] D.T. Logan, X.D. Su, A. Aberg, K. Regnstrom, J. Hajdu, H. Eklund, P. Nordlund, *Structure* 4 (1996) 1053.
- [60] Y. Lindqvist, W. Huang, G. Schneider, J. Shanklin, *Eur. Mol. Biol. J.* 15 (1996) 4081.
- [61] D.A. Whittington, S.J. Lippard, *J. Am. Chem. Soc.* 123 (2001) 827.
- [62] S. Macedo, C.V. Romão, E. Mitchell, P.M. Matias, M.Y. Liu, A.V. Xavier, J. LeGall, M. Teixeira, P. Lindley, M.A. Carrondo, *Nat. Struct. Biol.* 10 (2003) 285.
- [63] D.W. Banner, M. Kokkinidis, D. Tsernoglou, *J. Mol. Biol.* 196 (1987) 657.
- [64] F.H.C. Crick, *Acta Crystallogr* 6 (1953) 685.
- [65] C. Cohen, D.A.D. Parry, *Proteins* 7 (1990) 1.
- [66] J.W. Bryson, S.F. Betz, H.S. Lu, D.J. Suich, H.X. Zhou, K.T. O'Neil, W.F. DeGrado, *Science* 270 (1995) 935.
- [67] A.M. Keech, N.E. Le Brun, M.T. Wilson, S.C. Andrews, G.R. Moore, A.J. Thomson, *J. Biol. Chem.* 272 (1997) 422.
- [68] I. Bertini, C. Luchinat, *Adv. Inorg. Biochem.* 6 (1984) 71.
- [69] F. Frolow, A.J. Kalb (Gilboa), J. Yariv, *Nat. Struct. Biol.* 1 (1994) 453.
- [70] M.E. Andersson, M. Högbom, A. Rinaldo-Matthis, K.K. Andersson, B.-M. Sjöberg, P. Nordlund, *J. Am. Chem. Soc.* 121 (1999) 2346.
- [71] W.J. Xu, A. Baase, E. Baldwin, B.W. Matthews, *Protein Sci.* 7 (1998) 158.
- [72] C.N. Pace, *J. Mol. Biol.* 226 (1992) 29.
- [73] J.K. Myers, C.N. Pace, J.M. Scholtz, *Biochemistry* 36 (1997) 10923.
- [74] A.C. Rosenzweig, P. Nordlund, P.M. Takahara, C.A. Frederick, S.J. Lippard, *Chem. Biol.* 2 (1995) 409.
- [75] M. Merckx, D.A. Kopp, M.H. Sazinsky, J.L. Blazyk, J. Muller, S.J. Lippard, *Angew. Chem., Int. Ed.* 40 (2001) 2782.
- [76] R.L. Rardin, W.B. Tolman, S.J. Lippard, *New J. Chem.* 15 (1991) 417.

- [77] G.A. Papoian, W.F. DeGrado, M.L. Klein, *J. Am. Chem. Soc.* 125 (2003) 560.
- [78] A. Magistrato, W.F. DeGrado, A. Laio, U. Rothlisberger, J. VandeVondele, M.L. Klein, *J. Phys. Chem. B* 107 (2003) 4182.
- [79] K.J. Lumb, P.S. Kim, *Biochemistry* 34 (1995) 8642.
- [80] S. Nautiyal, T. Alber, *Protein Sci.* 8 (1999) 84.
- [81] A. Lombardi, J.W. Bryson, W.F. DeGrado, *Biopolymers* 40 (1996) 495.
- [82] S. Nautiyal, D.N. Woolfson, D.S. King, T. Alber, *Biochemistry* 34 (1995) 11645.
- [83] R. Fairman, H.G. Chao, T.B. Lavoie, J.J. Villafranca, G.R. Matsueda, J. Novotny, *Biochemistry* 35 (1996) 2824.
- [84] J.S. Richardson, D.C. Richardson, *Trends Biochem. Sci.* 14 (1989) 304.
- [85] H.W. Hellinga, F.M. Richards, *Proc. Natl Acad. Sci. USA* 91 (1994) 5803.
- [86] S.F. Betz, W.F. DeGrado, *Biochemistry* 35 (1996) 6955.
- [87] M. Struthers, J.J. Ottesen, B. Imperiali, *Fold. Des.* 3 (1998) 95.
- [88] W.F. DeGrado, Z.R. Wasserman, J.D. Lear, *Science* 243 (1989) 622.
- [89] B. North, C.M. Summa, G. Ghirlanda, W.F. DeGrado, *J. Mol. Biol.* 311 (2001) 1081.
- [90] A.C. Rosenzweig, H. Brandstetter, D.A. Whittington, P. Nordlund, S.J. Lippard, C.A. Frederick, *Proteins: Struct. Funct. Genet.* 29 (1997) 141.
- [91] A.S. Pereira, W. Small, C. Krebs, P. Tavares, D.E. Edmondson, E.C. Theil, B.H. Huynh, *Biochemistry* 37 (1998) 9871.
- [92] J. Hwang, C. Krebs, B.H. Huynh, D.E. Edmondson, E.C. Theil, J.E. Penner-Hahn, *Science* 287 (2001) 122.
- [93] J.A. Broadwater, J. Ai, T.M. Loehr, J. Sanders-Loehr, B.G. Fox, *Biochemistry* 37 (1998) 14664.
- [94] P. Moenne-Loccoz, J. Baldwin, B.A. Ley, T.M. Loehr, J.M. Bollinger Jr., *Biochemistry* 37 (1998) 14659.
- [95] D.T. Logan, F. deMare, B.O. Persson, A. Slaby, B.M. Sjoberg, P. Nordlund, *Biochemistry* 37 (1998) 10798.
- [96] J.R. Calhoun, H. Kono, S. Lahr, W. Wang, W.F. DeGrado, J.G. Saven, *J. Mol. Biol.* 334 (2003) 1101.