An artificial di-iron oxo-protein with phenol oxidase activity

Marina Faiella, Concetta Andreozzi, Rafael Torres Martin de Rosales, Vincenzo Pavone, Ornella Maglio, Flavia Nastri, William F. DeGrado & Angela Lombardi

Supplementary Information

SUPPLEMENTARY METHODS

Materials

DF3 was chemically synthesized by using standard Fmoc protocol and purified to homogeneity by RP-HPLC; MALDI-TOF spectrometry gave the expected molecular weight (5779 a.m.u. for the monomeric unit). All the reagents were commercially available and were used without further purification. Ultrapure water was used in all experiments. Amino acids and coupling reagents for peptide synthesis were purchased from NovaBiochem. The substrates used in the catalytic assays (4-AP (**1**), 3,5-DTBC (**3**), PPD (**5**) and OPD (6)) were purchased from Sigma-Aldrich. 3,5-DTBQ (**4**) and all the standard chemicals were obtained from Fluka.

Design of DF3

The DF3 design was based on DF1 scaffold. Its sequence was first modified by introducing four Gly residues in both position 9 and 13 of each α_2 monomer, in order to increase the active site cavity and create a binding site for organic cofactors, according to earlier studies on DF tetrameric models[1-3]. Further, we re-design the loop sequence, in order to optimize its conformation.

In DF1, the loop sequence Val-Lys-Leu was designed to idealize a γ-α_L-β loop, a common inter-helical linker in proteins[4,5]. A careful examination of the loop region in the crystal and solution structures of DF1 and several variants revealed that the overall structural context was not optimal. A Rose-like α_L-β conformation was observed in only about one third of the structures; in the remaining structures, the loop residues adopted alternative conformations[6]. Thus, the designed conformation of the turn did not appear to be as stable as expected, and could be differently shaped, depending on the crystal environment[7,8].

In the subsequent analogue DF2[9], the former loop sequence Val-Lys-Leu was replaced by Val-Gly-Asp, according to results emerged from earlier studies on α-α hairpin sequence pattern[4,5]. The

γ position is most often occupied by hydrophobic residues, the α_L position is predominantly stabilized by Gly or hydrophilic residue with flexible side chain and the β position is occupied by hydrophilic or small residues. Although these mutations improved some DF2 properties, respect to DF1, the loop was still flexible, as revealed by the structural characterization both in the solid state[6] and in solution[10]. The di-Cd(II) complex showed in the crystal structure a ''Rose-like'' α_L-β inter-helical turn[11]; however, the relatively high B-factor values of the loop residues indicated some flexibility[6]. This finding was confirmed by the NMR structure of di-Zn(II) complex**,** which unexpectedly showed in solution two conformational families of the loop region[10].

In order to further stabilize the desired "Rose-like'' α_L-β inter-helical turn, we introduced in DF3 the sequence Thr-His-Asn, as suggested by the statistical analysis of the structures and sequence preferences of inter-helical hairpin[6].

Histidine was included to stabilize an α_L conformation, which was necessary for a Rose motif. This residue occupies in the loop a Ccap position. Although glycine is a good helix-terminating amino acid for its intrinsic flexibility, histidine is also highly favoured at this position: its positive dihedral angles allow a sharp turn in the peptide backbone and prevent helix propagation[12]. Asparagine, located near the N-terminus of the second helix, was selected as N-capping residue[13,14].

UV-Visible Studies

UV-Visible measurements were performed at 25 °C in a Varian Cary 5000 UV-Vis-NIR spectrophotometer fitted with a thermostated cell compartment, by using quartz cells of 1 cm path length. Wavelength scans were performed at room temperature, from 200 to 700 nm, with a 60 nm/min scan speed. Anaerobic spectra were recorded in rubber-sealed cuvettes under argon. Kinetic experiments were recorded using a Varian Cary 50 spectrophotometer, by measuring the appearance of the products in the reaction medium.

Chemical Denaturation

Gdn·HCl denaturation curves were obtained by monitoring the circular dichroism (CD) at 222 nm by using a Jasco (Tokyo) J-715 dichrograph with a 60 seconds averaging time, and the data were analyzed as described in ref. 15.

Iron complex preparation

Both protein and iron stock solutions were freshly prepared. All the DF3 concentrations are given for the dimeric form. The initial DF3 concentration was determined spectrophotometrically by using ε_{280} 19,060 M^{-1} cm^{-1} (per dimer). Iron content in the ferrous ammonium sulfate (Mohr's salt) stock solutions and in the pure DF3 complex were measured by a Varian Spectra AA 220 atomic absorption spectrometer, equipped with a MK7 burner. The di-Fe(III)-DF3 complex was prepared as follows. 1.05 equivalents of Mohr's salt (in 0.1% H_2SO_4 water solution) were added dropwise to 1.0 ml of aqueous protein solution (≈ 0.25 mM). The solution was then aerobically diluted by slow addition of ≈ 1.0 ml of a buffer solution (200 mM HEPES, 200 mM NaCl, pH 7.0), and incubated for 1 hour at room temperature under stirring, in atmospheric oxygen. During the incubation, the protein solution gradually became pink. Precipitated iron oxides and excess reagents were removed by centrifugation, followed by filtration and concentration on Amicon ULTRA filters (5,000 MW cut off). Pure complexes were obtained after 3 cycles of washing with the appropriate buffer solution. Complete conversion of the di-Fe(III)-DF3 complex to the reduced form was accomplished by the addition of sodium dithionite. A solution of di-Fe(III)-DF3 complex in buffer was made anaerobic by several cycles of flushing with argon. 1.5-fold excess of a dithionite solution was added, and the resulting solution was incubated under argon for 30 min. Re-oxidation of the diferrous complex was easily obtained upon exposure to atmospheric oxygen under stirring.

The UV-Vis spectra of the DF3 di-iron complexes in the ferric and ferrous forms are illustrated in the **Supplementary Fig. 1**.

Catalytic assays

All the catalytic assays were performed in 100 mM HEPES, 100 mM NaCl, pH 7.0 at 25 °C. The experiments were followed using a Varian Cary 50 spectrophotometer, by measuring the appearance of the products in the reaction medium. Reference cuvettes contained all of the components, except the substrate, in a final volume of 0.5 ml.

The substrates used in the catalytic assays were 3,5-DTBC, 4-AP, PPD and OPD. Stock solutions of each substrate were prepared in different solvents, depending on their solubility. The appropriate amount of the substrate solution was added to the di-Fe(III)-DF3 solution, and the starting time was set at this point. The experiments were followed at the wavelength corresponding to the absorption maximum of the product. In details, the experimental conditions for the catalytic assays with the different substrates are as follows.

4-aminophenol (4-AP): di-Fe(III)-DF3 2.0 μM; stock solution of 4-AP was prepared in dimethylformamide in the range of 0.1 - 3 mM; the experiments were done in the presence of 10 mM *m*-phenylendiamine[16] (**7**). 4-amino-phenol was oxidized to the corresponding benzoquinone mono-imine (**2**). Benzoquinone mono-imine was detected as aminoindoaniline (**8**) by reaction with *m*-phenylendiamine.

The formation of the aminoindoaniline dye was measured at $\lambda_{max} = 528$ nm, using the ε value of 10,700 M^{-1} cm^{-1} , taken from the literature[17]. For this substrate, di-Fe(III)-DF3 protein showed multiple turnover kinetic. Measurement of the reaction over the course of an hour indicated that the protein was capable of at least 50 turnovers.

3,5-di-tert-butylcathecol (3,5-DTBC): di-Fe(III)-DF3 13 μM; stock solution of 3,5-DTBC was prepared in methanol in the range of 0.1 mM - 2 mM; the formation of 3,5-ditert-butylquinone (3,5-

DTBQ) was measured at λ_{max} = 400 nm, using the ε value of 1,040 M^{-1} cm^{-1}, experimentally determined in the kinetics conditions, using commercially available 3,5-DTBQ.

***p*-phenylenediamine (PPD):** di-Fe(III)-DF3 2.0 μM; stock solution of PPD was prepared in dimethylformamide in the range of 0.2 - 5 mM; the experiments were done in the presence of 10 mM aniline[17] **(9)**. The formation of phenosafranine (**10**) was measured at λ_{max} = 519 nm, using the ε value of 29,313 M^{-1} cm^{-1}, experimentally determined in the kinetics conditions.

Initial velocities were extracted from the steady-state portion of the progress curves (first 30 s after addition of the substrate). Kinetic parameters (V_{max}, K_m and k_{cat}) were obtained assuming a Michaelis-Menten model and determined from a Lineweaver-Burk plot (see **Supplementary Fig. 2**). The oxidation products (3,5-DTBQ and phenosafranine) were also identified by ESI-MS spectrometry.

ESI-MS analysis of the oxidation products

The oxidation products derived from the kinetic measurements of 3,5-DTBC and PPD (3,5-DTBQ and phenosafranine) were identified by ESI-MS spectrometry. ESI-MS spectra were recorded on a Shimadzu LCMS-2010EV system with ESI interface, and Shimadzu LCMS solution Workstation software for the data processing. A Q-array-octapole-quadrupole mass analyzer was used as the detector. The mobile phase, at a flow rate of 0.2ml/min, was made up of acetonitrile containing 0.1% formic acid. The optimized MS parameters were selected as followed: CDL (curved desolvation line) temperature 250 °C; the block temperature 250 °C; the probe temperature 250 °C; detector gain 1.6kV; probe voltage +4.5kV; CDL voltage -15V. Nitrogen served as nebulizer gas (flow rate: 1.5L/min). In order to avoid perturbation of the ESI-MS analysis by HEPES buffer, the reaction products were extracted from the reaction mixture with CH_2Cl_2. The organic solution was dried, and the product dissolved in methanol. ESI-MS spectrum for the 3,5-DTBQ shows a peak m/z, 220.85, which corresponds to the expected mass for the quinone. The ESI-MS spectrum for the oxidation product of PPD shows the most abundant peak at m/z 286.95, which corresponds to the expected mass for phenosafranine.

NMR Data Collection and Analysis.

Samples for NMR spectroscopy were prepared using the following conditions: to a 0.25 mM DF3 solution in phosphate buffer (50 mM, in 90% H_2O/10% D_2O, pH 6.0) were added dropwise 24 μl of a $ZnCl_2$ aqueous solution (0.14 mM, 10 % fold excess); 3-(trimethylsilyl) propionic-2,2,3,3-D4 acid sodium salt (TSP) was added as internal reference. All NMR spectra were acquired at 25 °C on a Bruker Avance 600 spectrometer equipped with a triple resonance cryo-probe. Suppression of the water signal was accomplished by excitation sculpting sequence[18]. Two dimensional NOESY[19,20], TOCSY[21] and DQF-COSY[22] spectra were acquired at very high digital resolution, using standard pulse sequences. NOESY spectra were acquired using mixing times ranging from 80-150 ms. TOCSY experiment was performed with spin lock applied for 60 ms. Spectra were processed using Bruker TOPSPIN software and analyzed on SGI Octane workstation with the CARA program[23]. The ^1H chemical shifts (in ppm) of di-Zn(II)-DF3 complex are reported in **Supplementary Table 1**.

NMR Structure Calculations and Refinement.

Intensities of dipolar connectivities in the two-dimensional NOESY spectrum, obtained using a 150 ms mixing time, were measured using the integration subroutine of the CARA program. Peak volumes were converted into upper limits of interatomic distances, to be used as input for structure calculations, by following the methodology of the program CALIBA[24]. The di-Zn(II)-DF3 structure was calculated with the program CYANA[25] (version 2.1), by using a total of 1,402 experimental restraints (15 per residue, including 2 x 230 intraresidue, 2 x 185 sequential, 2 x 152 medium-range, 174 intermonomer, and 2 x 47 intramonomer NOEs) (**Supplementary Table 2**). The di-Zn(II) site was treated the same as previous dimeric DF structures[6,10]: the CYANA standard library was modified by defining an artificial His residue containing a zinc atom covalently bonded to its N^{δ}. The geometry of the other coordinating residues (Glu10 and Glu36) was restrained by upper and lower distance limits. Typical CYANA runs were performed on 200 randomly generated starting structures with 10,000 torsion angle dynamics steps. The 40 CYANA structures with the lowest target function were then subject to further refinement by using the Sander module of AMBER 7.0 with the 1999 version of all-atom force field[26]. During the energy minimizations, the solvation effects were implicitly included using the pairwise GB-model as implemented in Amber 7[27]. Distance restraints were applied as a flat well with parabolic penalty within 0.5 Å outside the upper bond, and a linear function beyond 0.5 Å, using a force constant of 32 kcal mol^{-1} $Å^{-2}$ for the distance constraints. The best 30 structures (Fig. 1f, main text) among those with a total Amber energy lower than -4,801.2 kcal/mol and with a residual distance-constraint energy lower than 56.4 kcal/mol were selected to represent the solution structure for di-Zn(II)-DF3. The average Amber energies of the NMR bundle were: total energy: -4,904.21 ± 50.21 kcal mol^{-1}; distance constraints: 42.56 ± 4.18 kcal mol^{-1}; Van der Waals -668.49 ± 13.02 kcal mol^{-1}.

The quality of the obtained structures was evaluated in terms of deviations from ideal lengths and bond angles through a Ramachandran plot obtained using the program PROCHECK-NMR[28]. Analysis of the ensemble of the structures revealed that 90.0 % of backbone angles fall within the most favored regions of Ramachandran plot, 9.5 % in additional allowed regions and 0.50 % in the generously allowed regions. The structural statistics for the 30 final NMR structures are reported in **Supplementary Table 2**.

Supplementary Table 1 ^{1}H chemical shifts (in ppm) of di-Zn(II)-DF3 at 298 K and pH 6.1[a].

			HN	**Hα**	**Hβ**	**Hγ**	**Hδ**	**others**
	Ace			1,84				
1	**ASP**	D	8.38	4,37	2,64 2,47			
2	**TYR**	Y	7.70	4.17	3.00 2.91			$H\delta_{1,2}$ 6.94 $H\varepsilon_{1,2}$ 6.61
3	**LEU**	L	7.31	3.86	1.79	1.18		δ,δ' CH_3 0.79
4	**ARG**	R	7.69	3.83	1.88 1.76	1.50	3.13	$H_N\varepsilon$ 7.35
5	**GLU**	E	7.84	3.99	1.96 1.82	2.22		
6	**LEU**	L	7.72	1.92	1.68 1.16	1.68		δ,δ' CH_3 0.60, -0.12
7	**LEU**	L	7.91	3.84	1.87	1.45		δ,δ' CH_3 0.92, 0.86
8	**LYS**	K	7.95	3.85	1.93	1.32	1.65	εε' CH_2 2.89
9	**GLY**	G	7.96	4.11 3.83				
10	**GLU**	E	8.31	4.43	2.14	2.58 2.32		
11	**LEU**	L	8.57	4.00	1.91 1.71	1.39		δ,δ' CH_3 0.73
12	**GLN**	Q	7.88	4.06	2.14	2.38		$H_N\varepsilon_{1,2}$ 7.43, 6.82
13	**GLY**	G	8.19	3.83				
14	**ILE**	I	8.43	3.79	1.89	1.64 1.05		γCH_3 0.88 δ,δ' CH_3 0.72
15	**LYS**	K	7.26	3.89	2.08	1.49	1.79	
16	**GLN**	Q	7.52	4.06	2.09 1.93	2.43 2.29		$H_N\varepsilon_{1,2}$ 7.28, 6.57
17	**TYR**	Y	9.21	4.00	2.96 2.74			$H\delta_{1,2}$ 6.85 $H\varepsilon_{1,2}$ 6.58 Hη 9.05
18	**ARG**	R	8.67	3.79	1.85 1.76	1.55	3.05	$H_N\varepsilon$ 7.05
19	**GLU**	E	7.62	3.86	2.02 1.88	2.32		
20	**ALA**	A	7.95	3.71	1.39			
21	**LEU**	L	8.59	3.98	1.76 1.49	1.90		δ,δ' CH_3 0.78, 0.73
22	**GLU**	E	7.51	3.78	1.86 1.46	2.05		
23	**TYR**	Y	7.20	4.38	2.84			$H\delta_{1,2}$ 6.81 $H\varepsilon_{1,2}$ 6.63
24	**THR**	T	7.74	4.28	4.13	1.22		
25	**HIS**	H	8.70	4.22	3.24			Hδ2 7.04 Hε1 8.26
26	**ASN**	N	7.28	4.81	2.65			$H_N\delta_{1,2}$ 8.13, 6.71
27	**PRO**	P		4.20	2.26	1.95 1.86	4.00 3.74	
28	**VAL**	V	8.05	3.36	1.81			γ,γ' CH_3 0.74, -0.67
29	**LEU**	L	7.33	3.87	2.02 1.63	1.36	0.88	
30	**ALA**	A	7.16	3.87	1.35			
31	**LYS**	K	7.80	3.83	1.74	1.84 1.28	1.58	ε,ε' CH_2 2.82
32	**ILE**	I	7.84	2.86	1.18	0.94		γCH_3 -0.49 δCH_3 0.47
33	**LEU**	L	7.82	3.34	1.65	1.36		δ,δ' CH_3 0.80
34	**GLU**	E	7.24	3.98	2.03 1.77	2.34		
35	**ASP**	D	7.41	4.00	3.09 1.97			
36	**GLU**	E	8.20	4.14	1.93 1.60	2.55 2.22		
37	**GLU**	E	8.05	3.70	1.95 1.78	2.39		
38	**LYS**	K	6.99	3.64	1.09	0.50	1.46	εε' CH_2 2.70, 2.45 $H_N\zeta$ 7.46
39	**HIS**	H	7.92	3.90	3.49 2.85			Hδ2 6.09 Hε1 7.34 $H_N\varepsilon 2$ 15.16
40	**ILE**	I	7.79	3.58	1.90	1.54 1.19		γCH_3 0.81 δCH_3 0.73
41	**GLU**	E	7.15	4.04	2.09	2.37		
42	**TRP**	W	8.81	4.65	3.46 3.05			Hδ1 7.14 Hε1 10.17 Hε3 7.25 Hζ3 6.71 Hζ2 7.46 Hη 7.21
43	**LEU**	L	8.77	4.03	2.03	1.16		δ,δ' CH_3 0.93, 0.86
44	**GLU**	E	8.80	3.93	2.18 1.92	2.53 2.11		
45	**THR**	T	8.17	4.00	4.46	1.22		
46	**ILE**	I	7.72	3.79	1.71	1.15		γCH_3 0.86 δCH_3 0.79
47	**LEU**	L	8.15	3.98	1.73	1.53		δ,δ' CH_3 0.86, 0.75
48	**GLY**	G	7.80	3.86				
49	**NH_2**							

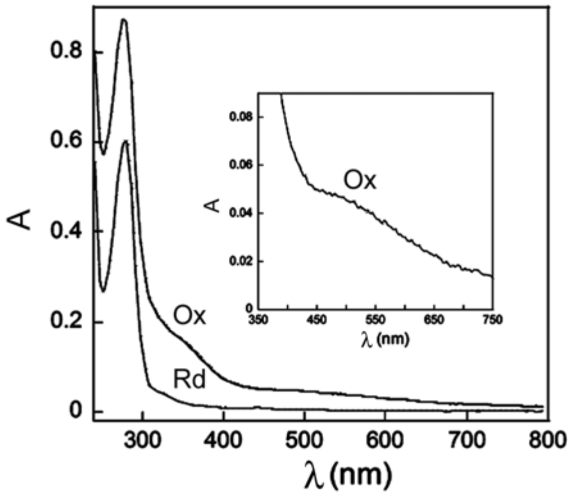
[a] *Chemical shifts were referenced to the internal TSP.*

Supplementary Table 2 NMR and refinement statistics the 30 final NMR structures

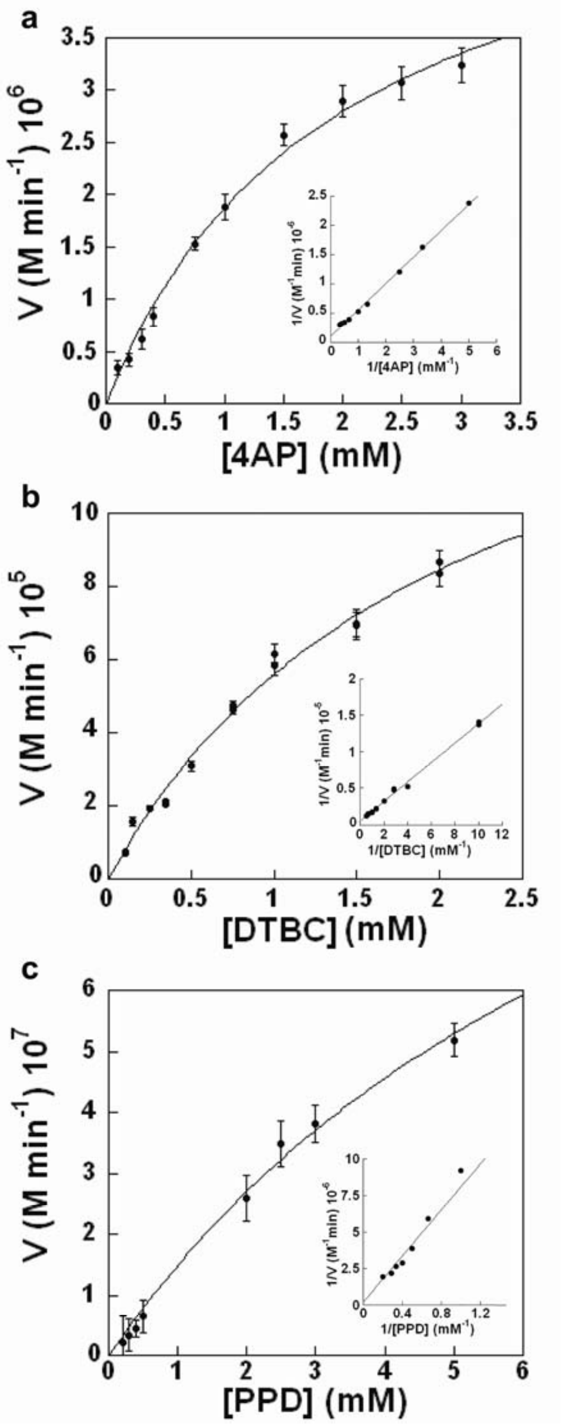
	di-Zn(II)-DF3
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1,402
Intra-residue	460
Inter-residue	942
Sequential ($\|i - j\| = 1$)	370
Medium-range ($\|i - j\| < 4$)	304
Long-range ($\|i - j\| > 5$)	94
Intermonomer	174
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	
$0.1 < d \leq 0.2$	36.07 ± 4.35
$0.2 < d \leq 0.3$	2.43 ± 1.36
$0.3 < d \leq 0.4$	0.07 ± 0.25
$0.4 < d \leq 0.5$	0.03 ± 0.18
Max. distance constraint violation (Å)	0.25 ± 0.05
Deviations from idealized geometry*	
Bond lengths (Å)	0.013
Bond angles (º)	2.1
Close contact	0
Average pairwise r.m.s. deviation** (Å)	
Heavy	1.39 ± 0.16
Backbone (2-46)	0.68 ± 0.12

* From the PDB validation software

** Pairwise r.m.s. deviation was calculated among the best 30 energy minimized conformers after superposition for the best fit of the atoms of the residues indicated in parentheses.



Supplementary Figure 1. Spectral features of the DF3 di-iron complex in the two oxidation states. The UV-visible spectrum of the diferric complex (indicated as Ox), displays the typical oxo-Fe(III) ligand-to-metal charge transfer bands from 300 to 350 nm (ε_{350nm} 5,270 M^{-1} cm^{-1}). The inset shows the magnified spectrum in the visible region. The observed band (ε_{500nm} 1,200 M^{-1} cm^{-1}) can be attributed to a phenolate-to-Fe(III) charge transfer transition, similarly to DFsc[29,30]. The UV-visible spectrum of the diferrous complex (indicated as Rd) was obtained upon reduction of di-Fe(III)-DF3 complex with sodium dithionite. Upon exposure to air, complete reformation of the di-Fe(III) species was observed. Repetition of this process over several cycles demonstrated the ability of the protein to oscillate between Fe(II)/Fe(III).



Supplementary Figure 2. Catalytic activity of di-Fe(III)-DF3. Initial rate of the oxidation versus substrate concentration for a) 4-AP, b) 3,5-DTBC, c) PPD. Di-Fe(III)-DF3 concentration was 2 μM in a) and c) and 13 μM in b). Kinetic parameters (k_{cat} and K_m) were determined from the Lineweaver Burk analysis (see insets). The analyzed substrates undergo a similar oxidation reaction, since they bear in their molecular structures two ortho- or para- related electron supplying groups (-OH or -NH_2), giving rise to the corresponding benzoquinone monoimines, quinones and di-imines, respectively.

SUPPLEMENTARY RESULTS

NMR structure of di-Zn(II)-DF3

The NMR solution structure of di-Zn(II)-DF3 was determined using a strategy similar to that reported for previous dimeric DF proteins[6,10,15].

Previous studies of dimeric DF1 and DF2 proteins indicate the di-Zn(II) substitution reproduces the structural properties of the di-Fe(II) and di-Mn(II) forms of the protein, having minimal effect on the tertiary structure or ligation geometry[31-33].

The well-dispersed resonances are indicative of a well-defined structure. In particular, the presence of several resonances shifted in the region -0.5 - 1.0 ppm is typically caused by the persistent interaction of aliphatic side chains with aromatic rings expected for a well-folded protein (**Supplementary Fig. 3**). The elements of secondary structure along di-Zn(II)-DF3 were first delineated using information provided by NOE data and $C_\alpha H$ conformational shifts (**Supplementary Fig. 4**). Two stretches of strong sequential $d_{NN(i,i+1)}$ NOEs, weak or absent $d_{\alpha N(i,i+1)}$, NOEs, medium-range $d_{\alpha N(i,i+3)}$, $d_{\alpha N(i,i+4)}$, and $d_{\alpha\beta(i,i+3)}$ NOEs, and large upfield conformational shifts indicate the presence of two α-helices spanning approximately residues 2-24 and 27-47. The secondary chemical shifts of the $C_\alpha H$ protons[34] show two stretches of negative values, characteristic of α-helical structures, in the regions encompassing the residues 1-24 and 27-47.

The good correlation between all the experimental data allows us to define two stable helical segments for di-Zn(II)-DF3 in solution: helix 1 (residues 2-24), and helix 2 (residues 27-47). The experimental data are also in agreement with the helix lengths determined from the ϕ and ψ angles of the final calculated structures.

Direct indication about the tertiary fold could be obtained from the unambiguous assignment of long-range (> *i, i+4*) NOE connectivities, using the well-resolved resonances of aromatic rings and methyl groups, as structural probes.

A family of structures was calculated using CYANA[25], with the di-Zn(II) site treated essentially the same as in the DF2 and DF2t structures[6,10]. The 40 CYANA structures, with the lowest target function, were then subjected to further refinement by restrained energy minimization (REM). A final ensemble of 30 refined structures was selected to represent the di-Zn(II)-DF3 solution structure (**Fig. 1f,** main text). This family of structures does not have a distance constraint violation larger than 0.5 Å (**Supplementary Table 2**). The overall stereochemical quality of the set of structures indicates that di-Zn(II)-DF3 structure is comparable to that of a 1.0 Å well-defined X-ray structure[28].

The solution structures of di-Zn(II)-DF3 clearly confirm all the aspects of the intended design. Despite the addition of four helix-destabilizing Gly residues, the protein retained the ability to fold correctly.

The di-Zn(II)-DF3 tertiary fold consists of an antiparallel four-helix bundle with an up-down-up-down topology. In the structure, helix 1 and helix 2 pack in an antiparallel manner against their symmetry related helix 1' and helix 2', respectively. As in native proteins, the apolar side chains are located in the core of the structure. The packing pattern of DF3 hydrophobic core spans the whole helices length. Most of the core residues in the protein adopt well-defined side chain conformation, with χ^1 angle values close to those intended in the design.

The helices within each monomer span residues 2-24 and 27-47, as defined by the pattern of hydrogen bonding as well as by the backbone torsion angles ϕ and ψ. As is often observed in the NMR structures of natural proteins, the ends of the helices show slightly greater distortion from ideal helical geometry than the remaining residues.

The loop region, comprising residues 24-26, is well defined and it is characterized, in all 30 selected structures, by an $\alpha_R\alpha_L\beta$ conformation. Thr24 adopts an α_R-conformation (ϕ= -112 ± 24°; ψ = -29 ± 6°); His25 adopts an α_L-conformation (ϕ= 43 ± 6°; ψ= 36 ± 12°); Asn26 is in a β-conformation (ϕ = -103 ± 19°; ψ = 149 ± 5°), consistent with its main-chain chemical shift index (**Supplementary Fig. 4**). The pattern of hydrogen bonding (**Fig. 1d,** main text) and dihedral angle values of the loop residues, are typical for a "Rose-like'' α_L-β inter-helical turn[11].

Interestingly, the NH group of His25 is involved in the C-capping motif[11,35,36]: the amide proton is hydrogen bonded to the Leu21 carbonyl group. The C-capping motif is further stabilized by two additional H-bonds: the first involves the Nε proton of His25 imidazole ring and the amide proton of Glu22, the second involves the Thr24 amide proton and the carbonyl group of Ala20. Asn26 adopts a β-conformation and is well situated to make N-cap interaction with the N-terminus of the helix 2[13,14]. In all selected structures, the Asn26 backbone carbonyl group forms an hydrogen bond to the amide proton of Leu29, three amino acids toward the C-terminus. The average $CO_i \leftarrow NH_{i+3}$ hydrogen bond distance is 2.83 ± 0.19 Å. The hydrophilic side chain of Asn26 is solvent exposed and, in about 50% of selected structures, it forms a well-directed hydrogen bond with the carbonyl oxygen of Thr24. Thus, the selected loop sequence encompasses all the requirements for stabilizing the desired Rose motif. As a consequence, aggregation phenomena are inhibited, and the solubility in water solution is highly improved (> 2 mM). The increased solubility also benefits from the polar nature of the loop: indeed, the side chain of the C-cap (His25) residue, as well as the carbonyl oxygens of C1 (Thr24) and C-cap are solvent exposed.

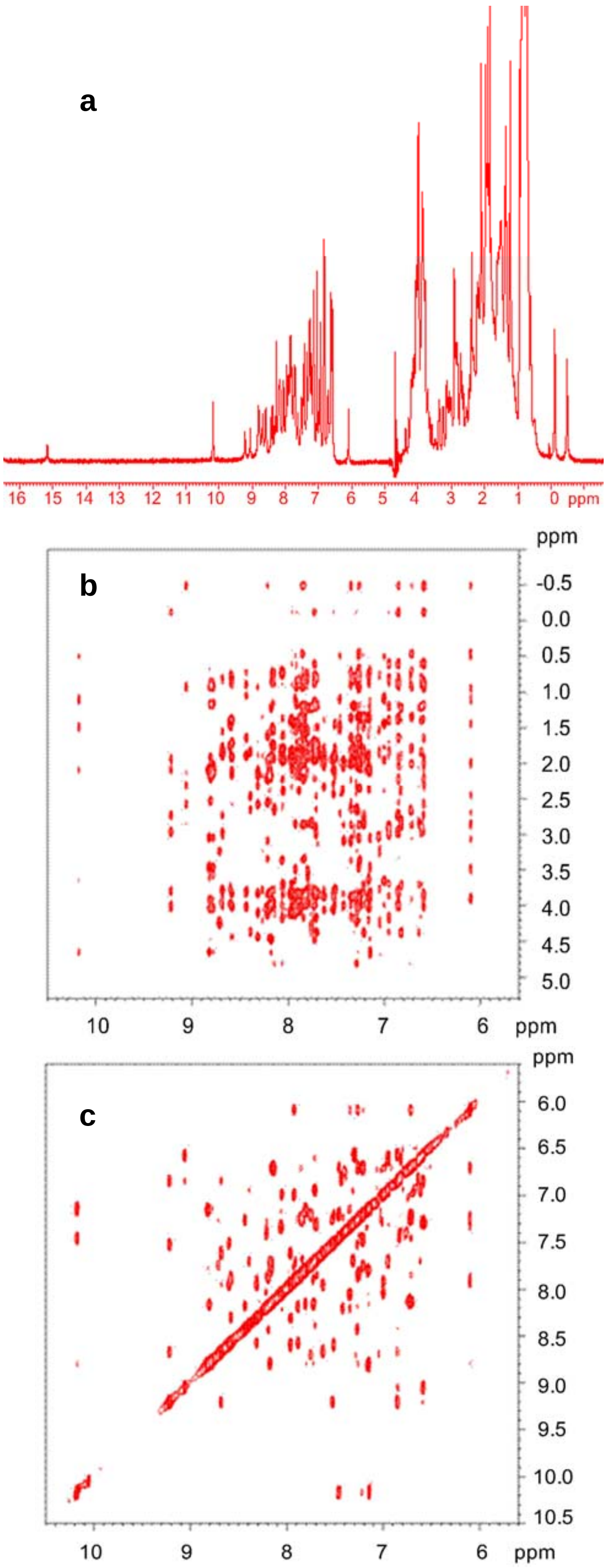
This hydrogen bond network, which stabilizes the DF3 inter-helical turn, was not observed in DF1, since the DF1 turn lacks a proper residue able to assume an α_L conformation, as well as an N-capping residue at the β-position[31].

The metal-binding site in di-Zn(II)-DF3 is stabilized by second-shell hydrogen bonds (**Supplementary Fig. 5and Fig. 6**), as already observed in previous structures of the dimeric DF proteins[32,33]. The primary ligand Glu10 accepts second-shell hydrogen bonds from Tyr17' of a neighbouring monomer, in all selected structures. The average distance between the tyrosine phenolic oxygen and the glutamate carboxylate oxygen (3.09 ± 0.28 Å) and the geometry are consistent with the formation of a hydrogen bond. Similarly, Asp35 accepts a second-shell hydrogen-bonded interaction from the primary ligand His39'. The average distance between the between the carboxylate oxygen of Asp35 and the Hε2 proton of His39' is 2.97 ± 0.12 Å. This distance is indicative of a reasonably strong hydrogen bond, possibly imparting partial imidazolate character to the histidine side chain[37-39], and it is in good agreement with the NMR data (His39 Hε2 chemical shift)[40-42].

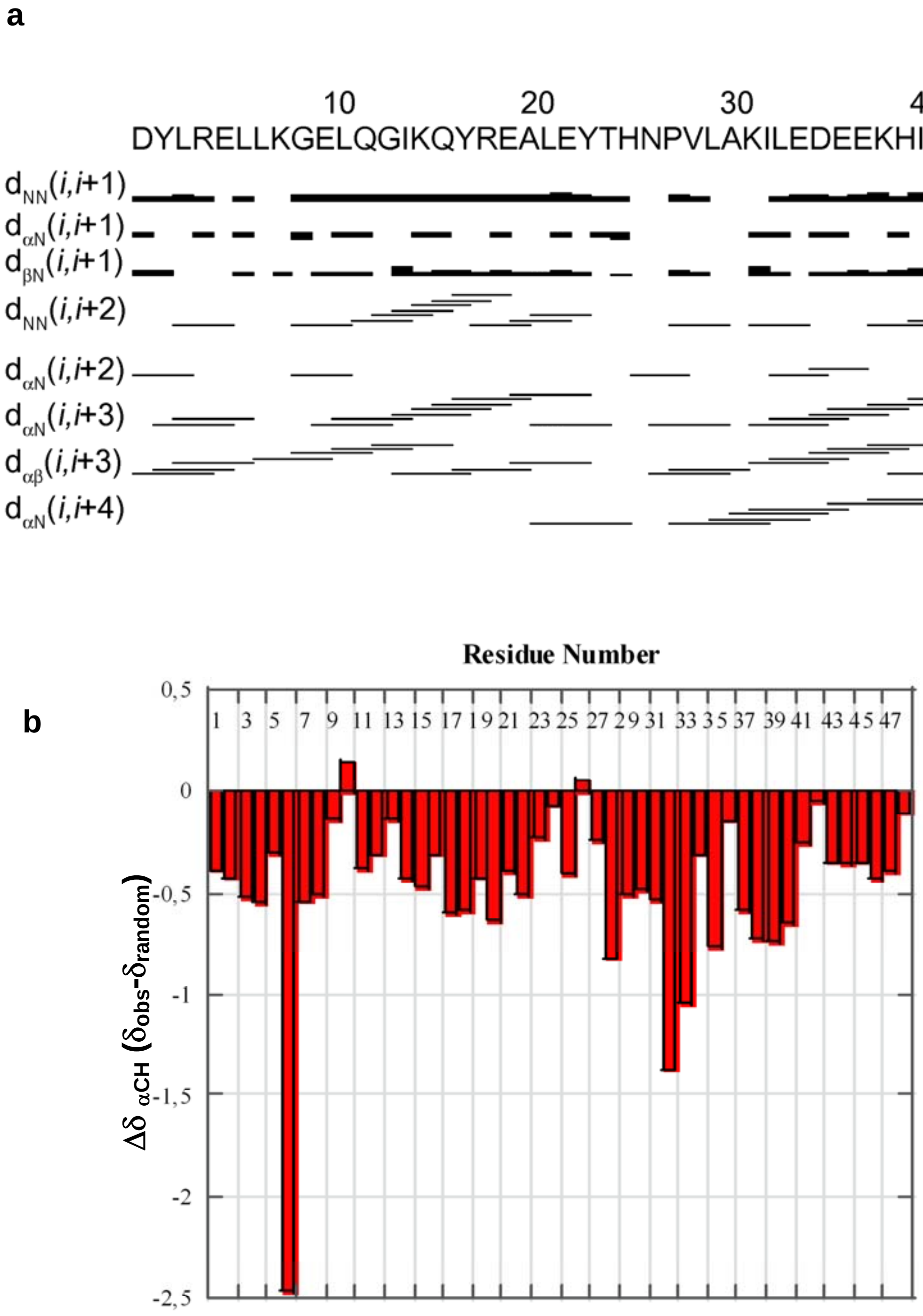
Thermodynamic analysis

The initial design of DF proteins focused on the construction of a thermodynamically stable framework, able to bind the di-metal cofactor, and to react with molecular oxygen[31]. In subsequent studies, the basic framework was re-engineered, with the aim of expanding the active site pocket, to allow binding of organic substrates[7-9,32,33]. Although the first two objectives were achieved, the expansion of the active site - with concomitant loss of packing and dehydration of apolar sidechains - proved to be unfeasible within the initially designed sequence. When the active site cavity was expanded, the desired fold became energetically unfavorable, leading to either unfolded protein or off-pathway formation of misfolded aggregated species[15]. In this work, we successfully increased the stability and conformational specificity of the protein through the redesign of a turn quite distant from the active site. In fact, the destabilization caused by introducing four Gly residues into the core structure of DF3 dimer was compensated by the increased stability of the inter-helical loop. Thus, the loop had a profound influence on the solubility (> 2 mM), stability, and ultimately on the possibility to introduce destabilizing mutations into the active site.

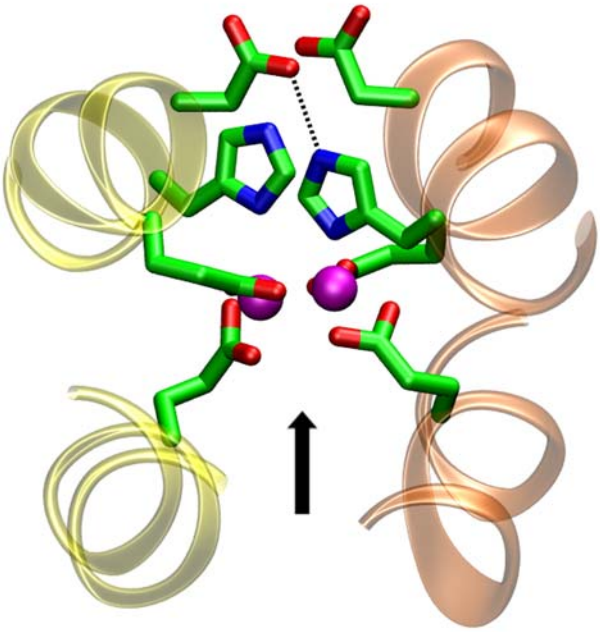
We compared the thermodynamics of folding of apo-DF3 with the apo forms of DF1 and L13G-DF1, as assessed by chemical denaturation in the absence of metal ions at pH 5.5[15]. Folding and dimerization are thermodynamically linked for this series of peptides, and the experimentally observed values for association/folding are $\Delta G_u^{H_2O}$ are 23.5 ± 0.3 kcal/mol, 12.7 ± 0.3 kcal/mol and 10.3 ± 0.6 kcal/mol for apo-DF1, apo-L13G-DF1 and DF3, respectively (corresponding to approximately 0.01 fM, 0.6 nM, and 0.03 μM). The Leu residues at positions 9 and 13 are almost completely inaccessible to solvent, and hence considerably influence the stability of the three proteins. In apo-DF1 replacement of Leu13 with Gly is quite deleterious, ($\Delta\Delta G$ = 11 kcal/mol), reflecting both the decrease in the helix propensity (approximately 2 kcal/mol dimer[17]) as well as the decreased contribution from burial of Leu13. It is therefore not surprising that the double replacement of Leu13 and Leu9 with Gly within the DF1 framework was problematic. The introduction of Gly at both 9 and 13 was, however, feasible in DF3; the value of $\Delta\Delta G$ for DF3 *vs.* L13G-DF1 is 2 kcal/mol. In this case $\Delta\Delta G$ reflects the destabilizing effects associated with the L9G substitution as well as the stabilizing effects of the improved loop; in balance the effect is far less drastic than that observed for L13G in DF1.



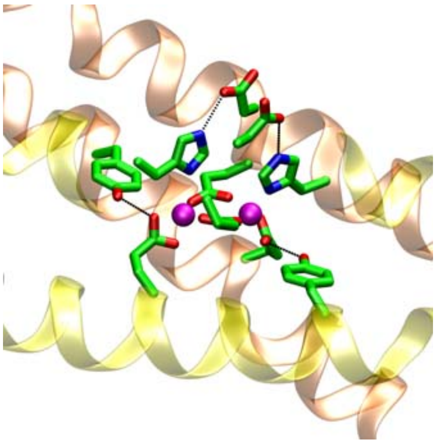
Supplementary Figure 3. NMR spectra of di-Zn(II)-DF3. **a**) The one-dimensional NMR spectrum. **b**) The two-dimensional NOESY spectrum in the amide/aromatic-aliphatic region. **c**) The two-dimensional NOESY spectra in the amide/aromatic-amide/aromatic region.



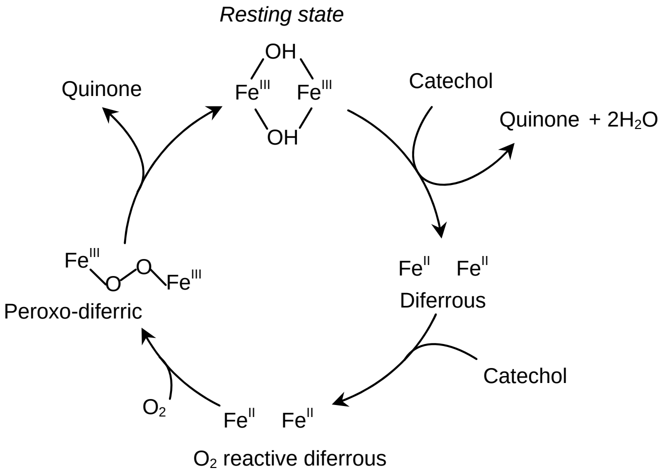
Supplementary Figure 4. Spectral data used for secondary structure determination for di-Zn(II)-DF3. **a**) NOE connectivities are indicated by a line drawn between two residues, and are classified as strong (■), medium (▬) and weak (—). Distinctive intensities of NOEs due to sequential and medium-range (i.e. residues less than 5 apart) contacts were used to define the elements of secondary structure in the protein. **b**) Chemical shift index (CSI) of α protons.



Supplementary Figure 5. The structure of di-Zn(II)-DF3 active site. The second-shell hydrogen bond between Asp35 and His39' is indicated as dashed line. The active site access channel is highlighted with a black arrow. The structure was generated with VMD.



Supplementary Figure 6. Structure of the active site of di-Zn(II)-DF3. Second-shell hydrogen bonds between Tyr17 and Glu10’, and between Asp35 and His39’ are indicated as dashed lines. (prime numbers refer to residues belonging to symmetry-related chains). The figure was generated using the program VMD.



Scheme 1. Proposed catalytic cycle for the alternative oxidase. The resting diferric state is reduced by quinol to yield quinone. In the next step the diferrous state reduce O_2, to form a peroxo-diferric intermediate. Finally, the two oxidizing equivalents of this intermediate are used to convert the second quinol, thus affording the diferric resting state of the protein.

References

1. Summa, C.M., Rosenblatt, MM., Hong, J.K., Lear, J.D. & DeGrado, W.F. *J. Mol. Biol.* **321**, 923-938 (2002).
2. Marsh, E.N. & DeGrado, W.F. *Proc. Natl. Acad. Sci. USA* **99**, 5150-5154 (2002).
3. Kaplan, J. & DeGrado, W.F. *Proc. Natl. Acad. Sci. USA* **101**, 11566-11570 (2004).
4. Efimov, A.V. *Prog. Biophys. Mol. Biol.* **60**, 201-239 (1993).
5. Brazhnikov, E.V. & Efimov, A.V. *Mol. Biol.* **35**, 89-97 (2001).
6. Lahr, S.J., *et al. J. Mol. Biol.* **346**, 1441-1454 (2005).
7. Di Costanzo, L., *et al. J. Am. Chem. Soc.* **123**, 12749-12757 (2001).
8. Geremia, S., *et al. J. Am. Chem. Soc.* **127**, 17266-17276 (2005).
9. Pasternak, A., Kaplan, J., Lear, J. D. & DeGrado W. F. *Protein Sci.* **10**, 958-969 (2001).
10. Maglio, O., *et al. J. Biol. Inorg. Chem.* **10**, 539-549 (2005).
11. Aurora, R. & Rose, G.D. *Protein Sci.* **7**, 21-38 (1998).
12. Richardson, J.S. & Richardson, D.C. *Science* **240**, 1648-1652 (1988).
13. Doig A.J.& Baldwin R. L. *Protein Sci.* **4**, 1325-1335 (1995).
14. Wan, W-Y & Milner-White, E.J. *J. Mol. Biol.* **286**,1633-1649 (1999).
15. Maglio, O., Nastri, F., Pavone, V., Lombardi, A. & DeGrado, W.F. *Proc. Natl Acad Sci. USA* **100**, 3772-3777 (2003).
16. Corbett, J.F. *J. Chem. Soc. B*, 823-826 (1969).
17. Corbett, J.F. & Gamson, E.P. *J. Chem. Soc. Perkin 2*, 1531-1537 (1972).
18. Hwang, T.L. & Shaka, A.J. *J. Magn. Reson. Ser. A* **112**, 275-279 (1995).

19. Jeneer, J., Meier, B.H., Bachmann, P. & Ernst, R.R. *J. Chem. Phys.* **71**, 4546-4553 (1979).

20. Kumar, A.V., Ernst, R.R. & Wüthrich, K. *Biochem. Biophys. Res. Commun.* **95**, 1-6 (1980).

21. Bax, A. & Davis, D.G. *J. Magn. Reson.* **65**, 355-360 (1985).

22. Piantini, U., Sorensen, O.W. & Ernst, R.R. *J. Am. Chem. Soc.* **104**, 6800-6801 (1982).

23. Keller, R.L.J. The Computer-aided Resonance Assignment Tutorial (Cantina Verlag, 2004).

24. Güntert, P., Braun, W. & Wüthrich, K. *J. Mol. Biol.* **217**, 517-530 (1991).

25. Günter, P., Mumenthaler, C. & Wüthrich, K. *J. Mol. Biol.* **273**, 283–298 (1998).

26. Case, D.A., *et al.* AMBER 7, University of California, San Francisco (2002).

27. Tsui, V. & Case, D.A. *Biopolymers* **56**, 275-291 (2000).

28. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. & Thornthon, J.M. *J. Biomol. NMR* **8**, 477-486 (1996).

29. Bell, C.B. III, *et al. Biochemistry* **48**, 59-73 (2009).

30. Calhoun, J.R., *et al. J. Am. Chem. Soc.* **130**, 9188-9189 (2008).

31. Lombardi, A,. *et al. Proc. Natl. Acad. Sci.* **97**, 6298-6305 (2000).

32. Maglio, O., *et al. C. R. Chimie* **10**, 703-720 (2007).

33. Calhoun, J.R., *et al. Biopolymers* **80**, 264-278 (2005).

34. Wishart, D.S., Sykes, B.D. & Richards, F.M. *Biochemistry* **31**, 1647-1651 (1992).

35. Harper, E.T.& Rose, G.D. *Biochemistry* **32**, 7605-7609 (1993).

36. Zhou, H.X., Lyu, P., Wemmer, D.E. & Kallenbach, R. *Protein Struct. Funct. Genet.* **18**,1-7 (1994).

37. Christianson, D.W. & Alexander, R.S. *J. Am. Chem. Soc.* **111**, 6412-6419 (1989).

38. Goodin, D.B. & McRee, D.E. *Biochemistry* **32**, 3313-3324 (1993).
39. Valentine, J.S., Sheridan, R.P., Allen, L.C. & Kahn, P.C. *Proc. Natl. Acad. Sci. USA* **76**,1009-1013 (1979).
40. Viragh, C., *et al. Biochemistry* **39**,16200-16205 (2000).
41. Frey, P.A.. *Magn. Reson. Chem.* **39**, S190-S198 (2001).
42. Cassidy, C.S., Lin, J. & Frey, P.A. *Biochem. Biophys. Res. Commun.* **273**,789-792 (2000).