Biopolymers

Spectroscopic and Metal Binding Properties of a De Novo Metalloprotein Binding a Tetrazinc Cluster

Journal:	*Biopolymers*
Manuscript ID	BIP-2018-0033.R1
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Chino, Marco; Universita degli Studi di Napoli Federico II, Department of Chemical Sciences Zhang, Shao-Qing; UCSF Cardiovascular Research Institute, Pharmaceutical Chemistry Pirro, Fabio; Universita degli Studi di Napoli Federico II, Department of Chemical Sciences Leone, Linda; Universita degli Studi di Napoli Federico II, Department of Chemical Sciences Maglio, Ornella; Universita degli Studi di Napoli Federico II, Department of Chemical Sciences; National Council of Researches, Institute of Biostructure and Bioimaging Lombardi, Angela; Universita degli Studi di Napoli Federico II, Department of Chemical Sciences DeGrado, William; UCSF Cardiovascular Research Institute, Pharmaceutical Chemistry
Keywords:	de novo protein design, multinuclear transition metal ion clusters, coiled coils, four-helix bundles, spectroscopic characterization

Napoli, May 18th, 2018

Detailed reply to reviewers:

Reviewer 1

This is a very nice, little paper from a highly-regarded and successful protein design group. It's technically sound and all the conclusions are justified. I would only suggest a couple of minor clarifications.

Q1. On page 5 the authors describe experiments at low pH, however, it is not clear if only one pH was tested.
A1: We thank the reviewer for allowing us to better clarify the experimental conditions. The experiments in the *apo* state were performed at pH 3.5. We have modified the text accordingly, both in the Results and Methods sections (pag 7 and 5/6, respectively).

Q2. Did the authors consider any complementary methods to verify the tetrameric nature of any of their peptides, native MS, native gels or simple size exclusion chromatography come into mind.
A2: We thank the reviewer for this suggestion. Accordingly, we performed size exclusion chromatography experiments. The SEC analysis has been added in the main text (Methods: pag. 6; Results: pag. 10; Discussion: pag. 11; Figure 7), and as supporting information (Figure S1 and Table S1).

Q3. Figure 4 - the caption could explain the error bars (ie. number of repeats)
A3: We changed the caption of Figure 4, and we added the following sentence: "The error bars indicate the standard deviation of three independent experiments".

Q4. Figure 5 looks somewhat sketchy which could be a simple reproduction problem
A4: We apologize that the figure could look not well resolved. We changed the layout of Figure 5. We lowered the number of levels in the DOSY contour plot and used dotted lines to highlight diffusing species of interest. Further, we have provided all the figures, including figure 5, in .tiff format at 600 dpi resolution.

Q5. Figure 6 - the caption could indicate the range of [Zn]/[protein] in (a) and clarify of [peptide] refers to the monomer or tetramer
A5: The caption of Figure 6 has been changed according to the reviewer suggestions: "Zn^{2+}-binding properties of 4DH3. a) CD spectral changes of a 40 μM 4DH3 solution (as monomer concentration) upon addition of $ZnSO_4$ (0.2 eq each) in 10 mM MES buffer pH 6.5. The black arrow indicates the observed changes by increasing the metal:peptide ratio up to 2. ……."

Q6. Figure 7 - indicate in the figure caption that these are simple molecular models
A6: The required clarification has been inserted in the caption of Figure 8 (previously named Figure 7): "Structural differences among the designed models of 4DH3, 4DH4 and 4DH5…...".

Reviewer 2

This paper describes a de novo design of metalloprotein 4DH3. The authors prepared three peptides and elucidated their solution properties by NMR and CD spectra. I recommend publishing on Biopolymers after consideration of following points.

Q: In the manuscript body, CD and NMR were measured under acidic condition pH 3.5 but in the Figure captions of Figures 2 and 3, 'in H2O' or 'in D2O' was shown. What are these differences?

A: We thank the reviewer for this comment that allowed us to better clarify the experimental conditions. Figures 2 and 3 refer to CD and NMR experiments of the *apo* form, performed at pH 3.5. We have modified the text accordingly, both in the Results and Methods sections (pag 7 and 5, respectively). For sake of clarity, this information has been also added in the figure captions.

Spectroscopic and Metal Binding Properties of a *De Novo* Metalloprotein Binding a Tetrazinc Cluster

Marco Chino[1], Shao-Qing Zhang[2,3], Fabio Pirro[1], Linda Leone[1], Ornella Maglio[1,4], Angela Lombardi[1,*] and William F. DeGrado[2,*]

[1] Department of Chemical Sciences, University of Napoli "Federico II", Via Cintia, 46, 80126 Napoli, Italy.

[2] Department of Pharmaceutical Chemistry and the Cardiovascular Research Institute, University of California at San Francisco, San Francisco, CA 94158-9001, United States.

[3] Department of Chemistry, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104-6396, United States.

[4] Institute of Biostructure and Bioimaging, National Research Council, via Mezzocannone, 16, 80134, Napoli, Italy.

[*] Correspondence should be address to A. L. (alombard@unina.it) and W.F.D. (william.degrado@ucsf.edu).

Abstract

De novo design provides an attractive approach, which allows one to test and refine the principles guiding metalloproteins in defining the geometry and reactivity of their metal ion cofactors. Although impressive progress has been made in designing proteins that bind transition metal ions including iron-sulphur clusters, the design of tetranuclear clusters with oxygen-rich environments remains in its infancy. In previous work, we described the design of homotetrameric four-helix bundles that bind tetra-Zn^{2+} clusters. The crystal structures of the helical proteins were in good agreement with the overall design, and the metal-binding and conformational properties of the helical bundles in solution were consistent with the crystal structures. However, the corresponding *apo*-proteins were not fully folded in solution. In this work, we design three peptides, based on the crystal structure of the original bundles. One of the peptides forms tetramers in aqueous solution in the absence of metal ions as assessed by CD and NMR. It also binds Zn^{2+} in the intended stoichiometry. These studies strongly suggest that the desired structure has been achieved in the *apo* state, providing evidence that the peptide is able to actively impart the designed geometry to the metal cluster.

Keywords: *de novo* protein design, multinuclear transition metal ion clusters, coiled coils; four-helix bundles, spectroscopic characterization.

Introduction

De novo design of metalloproteins is an important tool for understanding how metal binding affects protein folding and for building novel structural and enzymatic functions in unnatural protein scaffolds.[1-7] In the last years, there have been considerable successful *de novo* designs of non-heme metalloproteins utilizing helical bundles.[8-10] Indeed, design of helical bundles is advantageous as sequences of the constituting helices are usually based on heptad or other repeats, and the residues in the core of the bundles are generally converted to metal-binding sites.[11-13] Mononuclear metal-binding sites have been successfully incorporated into *de novo* designed three-helix bundles resulting in different functions, such as zinc-mediated hydrolysis and copper redox catalysis.[10,14] Di-metal sites have been built in *de novo* designed four-helix bundles with diverse catalytic functions.[9,15-17] Recently, we have redesigned this scaffold to stabilize the radical semiquinone form of cathecols within the dizinc analogue,[18] or to direct N-hydroxylation of arylamines or quinone trimerization by neat four-electron reduction of dioxygen at the diferric site.[19-22] Meanwhile, lanthanide ion binding sites have been built in the cores of three-helix bundles and have potential as MRI contrast regents.[23] It is noteworthy that uranyl ions can be extracted from seawater by binding to *de novo* designed binding site in a three-helix bundle.[24] Multinuclear iron-sulfur clusters as well as Cys-rich tetra-cadmium sites have been installed in *de novo* designed three-helix and four-helix bundles with programmable redox properties.[25-28] Recently, we have *de novo* designed the first oxygen-liganded tetranuclear zinc clusters in four-helix bundles.[29]

De novo design of multinuclear metal clusters in helical bundles has proven to be a difficult task, because the large size of multinuclear clusters affects the folding behavior of the helical bundles. Breathing and dissociative mechanisms have been shown to be crucial mechanisms of metal binding in helical bundles.[30] However, multinuclear metal cofactor binding inside helical bundles is complicated by many factors. Firstly, the helix association state and the nuclearity/geometry of the metal complex are structurally and thermodynamically coupled; secondly, structural rigidity of the helical bundles renders it difficult to predict the effects of the ligand geometry on the global fold; thirdly, how the different shells of interactions around the multinuclear metal coordination site influences folding of the helical bundles.[28]

In our recent *de novo* design of tetranuclear zinc clusters in four-helix bundles, we showed by crystal structure analysis that two tetrameric bundles, 4DH1 and 4DH2, accommodate a tetranuclear zinc cluster in their core.[29] However, they exhibit different folding features in solution: the signatures of helical structure inferred from far UV circular dichroism spectroscopy (CD) of 4DH2 increases after addition of zinc ions, while 4DH1 appears to be largely pre-organized and hence insensitive to zinc ions. Moreover, 4DH1 displays different oligomerization state by analytical ultracentrifugation from the other designed helical bundles.[29]

Based on these findings, in this work we design three new 4DH analogues with the aim of examining the features of this class of metalloproteins in solution. Unexpectedly, only one out of three, namely 4DH3, showed the expected folding and metal-binding ability as assessed by CD and Nuclear Magnetic Resonance (NMR) spectroscopies. We discuss the two unsuccessful designs in terms of helix-helix interfaces in coiled coils, suggesting the reasons for misfolding.

The solution study of *apo*-4DH3 showed a monomer/tetramer equilibrium, which demonstrates the hydrogen bond network, observed by X-ray crystallography in our previous work, is established even in the absence of the metal cofactor. Moreover, the tetramer is stable in solution in the absence of metal ions, and hence ready to host the tetranuclear zinc cluster. Thus, the peptide is able to actively impart the designed geometry to the metal cluster.

Methods

Protein design The starting backbone was generated by Coiled Coil Crick Parametrization program (CCCP, arteni.cs.dartmouth.edu/cccp/). 4DH1 crystallographic structure (PDB: 5WLL)[29] was used as template in the fitting of backbone coordinates. Metal binding residues, second/third shell residues and hydrophobic core (*a* and *d* positions) residues were kept accordingly. Peripheral residues were first screened for aromatic (Trp and Tyr) residues insertion by manual screening through Rosetta fixed-backbone (fixbb) symmetry design module, and the best Trp/Tyr pairs were chosen according to their attractive and repulsive van der Waals terms.[31] Subsequently, a full design

run was accomplished excluding Cys and Met residues for *b*, *c, e, f* and *g* positions between residues 1-10 and 19-26 and *b*, *e, f* and *g* positions between residues 12-18.

Peptide Synthesis The 26 amino acid peptides were synthesized by automatic solid-phase synthesis using an ABI 433A peptide synthesizer (Applied Biosystem, Foster City, CA, USA) with Fastmoc-protocols on a 0.2 mmol scale. The resin was a H-Rink amide ChemMatrix (Sigma), with a substitution level of 0.4 mmol/g. Fmoc group was removed with two 20% piperidine solution washings (3 and 7 minutes), and conditional washings were added upon conductivity measurements. Coupling was performed with a 5:5:10 equivalents of Fmoc-amino acid:HCTU:DIPEA solution in NMP for 40 minutes. Capping was performed after each coupling step with Ac_2O/HOBt/DIPEA solution in NMP. Peptide cleavage was carried out with a 95:2.5:2.5 (v/v/v 20 mL) solution of TFA:TIS:H_2O for two hours. Resin was finally washed three times with pure TFA and the eluates concentrated by rotary evaporation. The peptides were precipitated by adding 5 volumes of cold diethyl ether and separated by centrifugation. Purification was accomplished by preparative reverse-phase HPLC on a Vydac C4 column (22x250 mm) with water and acetonitrile (0.1 % TFA) as eluents. Identity and purity of the peptides was assessed by analytical HPLC coupled to a Shimadzu ESI IT-TOF mass spectrometer.

Circular Dichroism CD measurements were performed using a J-815 spectropolarimeter equipped with a thermostated cell holder (JASCO, Easton, MD, USA). CD spectra were collected at 25°C, from 260 to 200 nm at 0.2 nm intervals with a 20 nm min^{-1} scan speed, at 2.5 nm band width and at 16 s response. Cells of 0.5, 0.1 and 0.01 cm path length were used in the measurements. Mean residue ellipticities θ were calculated using the equation $\theta = \theta_{obs}/(10 \cdot l \cdot C \cdot n)$, in which θ_{obs} is the ellipticity measured in millidegrees, l is the path length of the cell in centimeters, C is the concentration in moles per liter, and n is the number of residues in the protein. Protein (monomer) concentration was determined by UV-Vis spectroscopy, using ε_{280} 6990 $M^{-1}cm^{-1}$.[32] Protein stock solutions were freshly prepared in H_2O at 1 mM concentration (pH 3.5). Upon dilution, the pH was

adjusted to 3.5 with trifluoroacetic acid. To conduct oligomerization studies, the peptide concentration ranged from 1 μM to 177 μM (as monomer) in 20 mM acetate buffer at pH 4.6, and prepared by dilution from a 0.5 mM stock solution. Mean residue ellipticity was fit as a function of peptide concentration using the program OriginPro. The equilibrium constant K_{diss} is defined as $[M]^n/[T]$, in which [T] is the concentration of the oligomeric species, [M] is the concentration of monomer, and n is the association stoichiometry. Thus, $K_{diss} = n(\alpha^n)(C^{n-1})/(1 - \alpha)$, where α is the fraction of monomer $[(\theta_{obsd} - \theta_{assoc})/(\theta_{mon} - \theta_{assoc})]$ and C is the total peptide concentration. The values of θ_{assoc} and θ_{mon} were varied during fitting.

NMR All solution for 1D-NMR analysis were prepared in H_2O/ D_2O (90:10 *v/v*). The *apo* form of the three analogues were analyzed at 1 mM concentration (pH 3.5). 1D-NMR spectra were acquired at 25 °C, on a Bruker Avance 600 spectrometer equipped with a triple resonance cryo-probe. Suppression of the solvent signal was accomplished by excitation sculpting sequence. 2D-DOSY spectrum of the 4DH3 analogue (0.5 mM) was obtained in D_2O. Exponential fitting of the signal decay was made with the Bruker proprietary software TopSpin. 128 scans were recorded for each FID, with total acquisition time 4h for the experiment. The theoretical Stoke's radius for *holo*-protein was calculated with the program HYDROPRO,[33] using default parameters except for: radius of the atomic probe 3.2 Å, temperature 298 K, solvent viscosity 0.012 poise.[34]

Size-exclusion Chromatography Analytical size-exclusion chromatography (SEC) was carried out for the 4DH3 peptide on an AKTA FPLC (GE) fitted with a Superdex 75 column (GE; cutoff 3000-70000) and eluted at 10 °C with MES (20 mM, pH 6.5) / NaCl (200 mM) buffer solution, at a 0.5 mL/min flow rate. To analyze the association state in the *holo* form, the buffer solution contained also 50 μM $ZnCl_2$. Peptide samples were prepared at 200 μM and 200 μL were injected. Four size standards were employed for calibration: blue dextran (2,000 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa), Aprotinin (6.5 kDa). Standard curve (Figure S1) and elution parameters (Table S1) are reported in the supporting information.

Results

Protein design

In this study, three new sequences were designed based on the 4DH series. Basically, four helices were arranged with D_2 symmetry, using Crick parameters from our previous designs.[29] The 26-mer helix bears 4 heptad repeats if we consider acetyl and amide capping at N- and C- termini, respectively. The central heptad repeat is occupied by the DXXH metal binding motif, where Asp and His occupy *a* and *d* positions, respectively (Figure 1a-b). Previous structural analysis showed that this binding motif forms a cuboidal zinc metal cluster of the form $[Zn^{2+}_4(N\delta\bullet His)_4(\mu^3\text{-}\eta^1\text{:}\eta^3\text{-}O_2C\bullet Asp)_4]$ (Figure 1c). The tetranuclear metal ion site is surrounded by the hydrophobic core composed by Leu and Ile residues in *a* and *d* positions.[35,36] In our previous paper,[29] we were interested in the design of residues at positions *c* and *g*, which form extensive hydrogen-bonding networks, acting as second and third-shell ligands to the primary ligands in the active site (Figure 1d). It is also known that the residues at the *b* and *e* positions of coiled coils play significant roles in stabilizing coiled coils.[37] While these residues were primarily apolar in the region adjacent to the active site in our previously designed peptides 4DH1 and 4DH2 (Figure 1e), in this work we explored the introduction of polar and electrostatic interactions. Therefore, polar residues were screened at several *b* and *e* positions, as well as *c* and *g* positions distant to the active site, using the Rosetta symmetric fixbb module.[31] Finally, aromatic residues, included to facilitate peptide concentration measurements, were allowed to vary at *c, f* and *g* positions. Three sequences were chosen based on the rank order of their Rosetta-computed energies (Figure 1e). Here, we present the solution properties of these variants.

Initial screening of the designed peptides

Circular dichroism spectroscopy reveals that only 4DH3 and 4DH4 fold into helical structures. CD spectroscopy was adopted to verify the helical global folding of the three analogues. The peptides were screened in the *apo*-state at pH 3.5, which allows protonation of the Asp and His residues and improves

solubility. Only 4DH3 has a significant α-helical content at this pH as determined by mean residue ellipticities (θ) at 192, 208 and 222 nm (Table 1 and Figure 2a, black). 4DH4 shows a lower helical content with respect to 4DH3 at 10 μM concentration (Figure 2a, red), whereas 4DH5 has random coil structure in these experimental conditions (Figure 2a, blue). Preliminary analysis of the three peptides at higher concentrations (0.1, 1 mM), reveals that only 4DH3 preserves its helical character, with concentration-dependent increase in helical content, as expected for coiled coil formation (Figure 2b-c, black). 4DH4 shows deepening of the minima centered at 205 and 225 nm as the concentration increases, characteristic of helical aggregates (Figure 2b-c, red).[38] 4DH5 exists as a random coil in these experimental conditions (Figure 2b-c, blue).

^{1}H-NMR screening suggests the equilibrium between different 4DH3 oligomeric states in water. NMR spectra of the three analogues were acquired in water (pH 3.5) at 1 mM concentration. Spectra analysis substantially confirms the results obtained by the preliminary CD analysis.

4DH5 shows sharp signals in the NH region, but they are not dispersed in the NH and Cα region, as would be expected for well-folded globular proteins (Figure 3, blue). These data are consistent with the hypothesis of a random coil structure for this analogue. 4DH4 shows a poorly defined and dispersed spectrum that suggests the presence of an oligomeric aggregate (Figure 3, red).

High field shifts for methyl protons hint the formation of a well-defined hydrophobic core in 4DH3 (Figure 3b, black). Amide signals appear well-defined and well-dispersed from 6 to 9 ppm (Figure 3a, black). These findings indicate the formation of a globular fold that consists of a defined number of associated helices, sequestering hydrophobic residues from the solvent.[39]

Solution analysis of 4DH3

Apo-4DH3 is in equilibrium between the monomer and the tetramer. Based on the above reported CD and NMR results, further analyses were carried out only on 4DH3. CD spectroscopy was adopted to evaluate the number of helices involved in the association, and to determine the thermodynamic driving force. At 1 mM concentration, the CD spectrum is consistent with nearly complete helix formation,

whereas at lower concentrations, the spectra are consistent with a combination of α-helix and random coil conformations (Figure 2, black).

Fitting of the concentration dependence of the θ at 222 nm, with a simple tetramerization model, gave a K_{diss} of 3.8 x 10^{-16} M^3 in 50 mM buffer acetate at pH 4.6, consistent with previously characterized tetrameric designed coiled coils (Figure 4)[40,41] Fitting of the experimental data for oligomeric states other than four gave worse R^2 values.

As expected, when the same experiment was performed in 10 mM MES buffer at pH 6.5, only slight changes in the θ_{222} values were observed upon concentration increasing (data not shown). This finding reflects the behavior of the DF class of proteins, in which the thermodynamic cost of hindering four negatively charged residues is too high in the absence of the metal cofactor.[42]

To confirm the hypothesis that 4DH3 is in equilibrium between the monomer and the tetramer states, DOSY-NMR analysis was performed to evaluate the diffusional properties of the associated oligomer. As shown in Figure 5, a single species can be observed in a 0.5 mM 4DH3 solution (in D_2O, pD = 3.91), corresponding to a diffusion coefficient of 1.2±0.1 10^{-10} m^2/s. Based on the hydrodynamic radius of the glycerol, used as an internal standard, the observed radius for 4DH3 is 16±2 Å.[21] Experimental value has been compared to the hydrodynamic radius as calculated directly from the designed model of the tetramer. The theoretical radius of the *holo*-form has been calculated to be 19 Å from the designed model.[33] Discrepancy between *apo* and *holo* forms is expected for this class of compounds, which reach the optimal fold upon metal binding at neutral pH.

4DH3 tetramer binds up to four zinc ions. The metal binding properties of 4DH3 were analyzed by following the increase of helical content upon addition of zinc.[42] The minimum at 222 nm of a 40 uM solution of 4DH3, in a 10 mM MES buffer at pH 6.5, decreased smoothly upon addition of $ZnSO_4$, resembling a saturation mechanism of binding (Figure 6a). Thus, the CD titration data were fit by linear regression of the initial and final linear regions of the curves before and after binding saturation (Figure 6b). The intercept of the two lines gives the peptide:metal expected ratio of 1:1 (corresponding to four

zinc ions for each tetramer). Protein aggregation and precipitation at higher concentrations, however, prevented investigation of the oligomeric state by NMR diffusion-oriented experiments. The association state of the designed peptide 4DH3 both in the *apo* and *holo* forms was then investigated by SEC. Figure 7 summarizes the SEC results with and without $ZnCl_2$. in the elution buffer. 4DH3 produced two well-defined chromatographic peaks when eluted at pH 6.5 in the absence of zinc ion. The first, less intense, eluted peak (elution volume 15.6 mL), corresponds to an apparent molecular weight approximately 2-fold higher than the monomer molecular weight (apparent M_w was 6000 Da). The second peak, 5-fold more intense than the previous, corresponds to the unfolded monomer eluting with the column volume, as its weight is close to the lower cutoff (3000-70000 Da), confirming the CD data under the same pH conditions. In the presence of zinc, 4DH3 produces a very broad, tailing peak, attributable to dissociation equilibria taking place into the column. Multi-peak fitting analysis was performed to elucidate the intermediate species. Three Gaussian curves were needed to fit the observed peak, corresponding to oligomerization state n = 3.4, 2.0, 0.9, respectively.

Discussion

In this work, we designed three new analogues of the 4DH series. We have previously shown that this class of peptides is able to fold into a homotetrameric four-helix bundle, and to host, in their core, an abiological metal cofactor composed by four Zn^{2+} ions bridged by four aspartates residues. In our two previous designs, solution characterization was limited by the lack of defined structure in the *apo* form. Here, we designed three new sequences by increasing the designable space of the two non-equivalent interfaces of the antiparallel four-helix bundle. This necessarily increases the number of sequences that could erroneously satisfy the desired fold, thus raising the risk of failure in such a difficult design task. In fact, only two out of three were found to be helical in the *apo* form, and only one was stable as a tetramer without aggregation by CD and NMR spectroscopies.

We therefore analyzed the designed models, by direct comparison of the bundle interfaces (Figure 8). Interfaces between residues at *b-e* positions are closely related in the three analogues; however, both

4DH4 and 4DH5 would expose a Val-Ala patch on both extremities of the bundle, which could account for the aggregation observed for 4DH4. 4DH5, instead, shows a great divergence in the *c*-*g* interface with respect to the other two analogues. The main difference lies for the pair of residues 4-21. Both 4DH3 and 4DH4 present two hydrophobic/aromatic residues (Tyr-Trp for 4DH3; Leu-Trp for 4DH4), whereas 4DH5 does not show any hydrophobic residue in *g* position. 4DH series is composed by only 4 heptad repeats, and one of them is dedicated to metal binding, thus limiting the thermodynamic driving force for folding given by the Leu/Ile hydrophobic core. The lack of hydrophobic residues for 4DH5 in the *c*-*g* interface is therefore not completely counterbalanced by the designed ion pairs (only -0.6 kcal/mol each).[41]

We thus calculated the free energy of tetramerization of *apo*-4DH3 at 25 °C, and we found it to be -21 kcal/mol, comparable to previously designed antiparallel tetrameric coiled coils.[36,40,41] This finding highlights that the *apo* form of this class of proteins is highly stable under acidic conditions and pre-organized for metal binding. We expect that a dense network of hydrogen bonds could stabilize the protonated aspartates also in the *apo* form, as already outlined in our previous structural study and in other brilliant examples.[43,44] At neutral pH, more than two oligomerization states are in equilibrium, as the metal-binding titration clearly shifted this equilibrium towards a well folded oligomeric state, with a saturation mechanism of binding. Interestingly, SEC analysis clearly suggests that the *apo* form is in equilibrium with a dimer intermediate at pH 6.5, which further moves to the tetrameric state upon zinc binding. Full conversion to the 4DH3 tetramer *holo* state was not observed under the explored conditions mostly because of dilution in the SEC column and successive dissociation equilibria. Protein precipitation prevented injection at higher concentration, and future designs will focus at increasing solubility of the *holo* state at neutral pH.

Conclusion

The primary goal of this work was to design new 4DH analogues, which would allow elucidating the behavior of this class of metalloproteins in solution. One of the three newly designed sequences, 4DH3,

adopts the expected folding in aqueous solution. In fact, 4DH3 is stable and tetrameric both in the *apo* and the *holo* forms, as assessed by CD and NMR spectroscopies. We also demonstrated that for this class of proteins zinc binds the tetramer in solution with the desired stoichiometry, thus recovering one of the main features of Photosystem II, which is the ability to release and recover the metal cofactor of the OEC, during turnovers. This finding is crucial for the simple and sustainable development of a bioinspired catalyst for solar energy uptake,[45] as well as the binding of abiological cofactors within *de novo* designed proteins.[46-48]

Acknowledgments

The authors thank Dr. Fabiana Paragliola for recording zinc titration data, and Fabrizia Sibillo for technical assistance. This work was primarily supported by NIH grant R35GM122603 to W.F.D., with additional support from the NSF (CHE1413295), the Research Department of Campania Region (STRAIN Project, POR FSE 2007/2013, grant number B25B0900000000 for a postdoctoral fellowship to M. C.), and University of Napoli "Federico II" for mobility grants to W.F.D. and M.C.

Dedication

We dedicate this work to Vincenzo Pavone, friend, mentor and inspiring scientist. His work in the areas of bioinorganic chemistry, peptide design and protein design has inspired multiple generations of chemists and will be valued for years to come. He pioneered the use of β-substituted amino acids in foldamer design long before it became fashionable, built a wide range of bioactive peptides, and contributed greatly to the first *de novo* design of catalytically active peptides and metalloproteins. We are deeply indebted to Vincenzo for his pioneering vision.

References

1. M. Chino, L. Leone, G. Zambrano, F. Pirro, D. D'Alonzo, V. Firpo, D. Aref, L. Lista, O. Maglio, F. Nastri, A. Lombardi, *Biopolymers* **2018**. DOI: 10.1002/bip.23107
2. M. H. Hecht, S. Zarzhitsky, C. Karas, S. Chari, *Curr. Opin. Struct. Biol.* **2018**, 48, 124.
3. F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Kohler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, 118, 142.
4. F. Nastri, M. Chino, O. Maglio, A. Bhagi-Damodaran, Y. Lu, A. Lombardi, *Chem. Soc. Rev.* **2016**, 45, 5020.
5. F. Yu, V. M. Cangelosi, M. L. Zastrow, M. Tegoni, J. S. Plegaria, A. G. Tebo, C. S. Mocny, L. Ruckthong, H. Qayyum, V. L. Pecoraro, *Chem. Rev.* **2014**, 114, 3495.
6. D. W. Watkins, C. T. Armstrong, J. L. Ross Anderson, *Curr. Opin. Chem. Biol.* **2014**, 19, 90.
7. F. Nastri, R. Bruni, O. Maglio, A. Lombardi, In *Coordination Chemistry in Protein Cages*, Wiley: Hoboken, NJ **2013**, pp 43.
8. A. F. Peacock, *Curr. Opin. Chem. Biol.* **2016**, 31, 160.
9. M. Chino, O. Maglio, F. Nastri, V. Pavone, W. F. DeGrado, A. Lombardi, *Eur. J. Inorg. Chem.* **2015**, 2015, 3371.
10. M. Tegoni, *Eur. J. Inorg. Chem.* **2014**, 2014, 2177.
11. O. Maglio, F. Nastri, A. Lombardi, In *Ionic Interactions in Natural and Synthetic Macromolecules*, Wiley: Hoboken, NJ **2012**, pp 361.
12. A. Lombardi, *Nat. Chem. Biol.* **2015**, 11, 760.

13. N. W. Schmidt, G. Grigoryan, W. F. DeGrado, *Protein Sci.* **2017**, 26, 414.

14. A. G. Tebo, V. L. Pecoraro, *Curr. Opin. Chem. Biol.* **2015**, 25, 65.

15. O. Maglio, F. Nastri, R. T. Martin de Rosales, M. Faiella, V. Pavone, W. F. DeGrado, A. Lombardi, *C. R. Chimie* **2007**, 10, 703.

16. D. Shiga, Y. Funahashi, H. Masuda, A. Kikuchi, M. Noda, S. Uchiyama, K. Fukui, K. Kanaori, K. Tajima, Y. Takano, H. Nakamura, M. Kamei, T. Tanaka, *Biochemistry* **2012**, 51, 7901.

17. T. L. Olson, E. Espiritu, S. Edwardraja, E. Canarie, M. Flores, J. C. Williams, G. Ghirlanda, J. P. Allen, *Biochim. Biophys. Acta, Bioenerg.* **2017**, 1858, 945.

18. G. Ulas, T. Lemmin, Y. Wu, G. T. Gassner, W. F. DeGrado, *Nat. Chem.* **2016**, 8, 354.

19. A. J. Reig, M. M. Pires, R. A. Snyder, Y. Wu, H. Jo, D. W. Kulp, S. E. Butch, J. R. Calhoun, T. Szyperski, E. I. Solomon, W. F. DeGrado, *Nat. Chem.* **2012**, 4, 900.

20. R. A. Snyder, S. E. Butch, A. J. Reig, W. F. DeGrado, E. I. Solomon, *J. Am. Chem. Soc.* **2015**, 137, 9302.

21. M. Chino, L. Leone, O. Maglio, D. D'Alonzo, F. Pirro, V. Pavone, F. Nastri, A. Lombardi, *Angew. Chem. Int. Ed.* **2017**, 56, 15580.

22. M. Chino, L. Leone, O. Maglio, A. Lombardi, *Methods Enzymol.* **2016**, 580, 471.

23. L. N. Slope, A. F. A. Peacock, *Chem. - Asian J.* **2016**, 11, 660.

24. L. Zhou, M. Bosscher, C. Zhang, S. Özçubukçu, L. Zhang, W. Zhang, C. J. Li, J. Liu, M. P. Jensen, L. Lai, C. He, *Nat. Chem.* **2014**, 6, 236.

25. C. E. Laplaza, R. H. Holm, *J. Am. Chem. Soc.* **2001**, 123, 10255.

26. D. J. Sommer, A. Roy, A. Astashkin, G. Ghirlanda, *Biopolymers* **2015**, 104, 412.

27. V. Nanda, S. Senn, D. H. Pike, A. Rodriguez-Granillo, W. A. Hansen, S. D. Khare, D. Noy, *Biochim. Biophys. Acta, Bioenerg.* **2016**, 1857, 531.

28. D. V. Zaytsev, V. A. Morozov, J. Fan, X. Zhu, M. Mukherjee, S. Ni, M. A. Kennedy, M. Y. Ogawa, *J. Inorg. Biochem.* **2013**, 119, 1.

29. S. Q. Zhang, M. Chino, L. Liu, Y. Tang, X. Hu, W. F. DeGrado, A. Lombardi, *J. Am. Chem. Soc.* **2018**, 140, 1294.

30. D. Ghosh, V. L. Pecoraro, *Inorg. Chem.* **2004**, 43, 7902.

31. F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, I. André, *PLoS One* **2011**, 6, e20450.

32. C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, *Protein Science* **1995**, 4, 2411.

33. A. Ortega, D. Amorós, J. García de la Torre, *Biophys. J.* **2011**, 101, 892.

34. R. C. Hardy, R. L. Cottington, *J. Res. Natl. Bur. Stand.* **1949**, 42, 573.

35. J. M. Fletcher, A. L. Boyle, M. Bruning, G. J. Bartlett, T. L. Vincent, N. R. Zaccai, C. T. Armstrong, E. H. Bromley, P. J. Booth, R. L. Brady, A. R. Thomson, D. N. Woolfson, *ACS Synth. Biol.* **2012**, 1, 240.

36. F. Thomas, A. L. Boyle, A. J. Burton, D. N. Woolfson, *J. Am. Chem. Soc.* **2013**, 135, 5161.

37. M. K. Yadav, L. J. Leman, D. J. Price, C. L. Brooks, 3rd, C. D. Stout, M. R. Ghadiri, *Biochemistry* **2006**, 45, 4463.

38. S. M. Kelly, T. J. Jess, N. C. Price, *Biochim. Biophys. Acta, Proteins Proteomics* **2005**, 1751, 119.

39. M. Faiella, C. Andreozzi, R. T. Martin de Rosales, V. Pavone, O. Maglio, F. Nastri, W. F. DeGrado, A. Lombardi, *Nat. Chem. Biol.* **2009**, 5, 882.

40. S. F. Betz, W. F. DeGrado, *Biochemistry* **1996**, 35, 6955.

41. R. Fairman, H.-G. Chao, T. B. Lavoie, J. J. Villafranca, G. R. Matsueda, J. Novotny, *Biochemistry* **1996**, 35, 2824.

42. R. T. Martin de Rosales, M. Faiella, E. Farquhar, L. Que, Jr., C. Andreozzi, V. Pavone, O. Maglio, F. Nastri, A. Lombardi, *J. Biol. Inorg. Chem.* **2010**, 15, 717.

43. S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. DiMaio, J. H. Pereira, B. Sankaran, G. Seelig, P. H. Zwart, D. Baker, *Science* **2016**, 352, 680.

44. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, *Nat. Chem.* **2016**, 8, 837.

45. J. D. Blakemore, R. H. Crabtree, G. W. Brudvig, *Chem. Rev.* **2015**, 115, 12974.

46. D. Roxbury, S. Q. Zhang, J. Mittal, W. F. DeGrado, A. Jagota, *J. Phys. Chem. C Nanomater. Interfaces* **2013**, 117, 26255.

47. K. H. Kim, D. K. Ko, Y. T. Kim, N. H. Kim, J. Paul, S. Q. Zhang, C. B. Murray, R. Acharya, W. F. DeGrado, Y. H. Kim, G. Grigoryan, *Nat. Commun.* **2016**, 7, 11429.

48. N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S. Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien, W. F. DeGrado, *Nat. Chem.* **2017**, 9, 1157.

Figure Legends

Figure 1. Design of tetranuclear clusters in helix bundles. a) D_2-symmetrical 4DH analogues, bearing the DXXH metal-binding motif, have four Asp and four His residues coordinating 4 ions at the core. Second-shell Asp residues at the *c-g* interface (shown in red) are included to interact with the first-shell His ligands. The positions of the three orthogonal two-fold axes are also indicated. b) Crystal structure of 4DH1 (PDB ID 5WLL): the main chain is presented as cartoon, zinc ions as spheres, and coordinating residues as sticks. c) Designed tetranuclear zinc cluster, showing the relative orientations among the carboxylates. d) Model structure of 4DH analogues, including the first-shell ligands Asp^{12} and His^{16}, the second-shell Asp^{11}, and the third-shell Arg^{14}. e) Sequence alignment of 4DH analogues. Coordinating residues are in bold, while *b* and *e* positions in proximity of the metal-binding site are underlined.

Figure 2. CD characterization of the three analogues as a function of peptide concentration. a) Comparison of far UV CD spectra of 4DH3 (black), 4DH4 (red) and 4DH5 (blue) at 10 μM concentration. b) Comparison of far UV CD spectra of 4DH3 (black), 4DH4 (red) and 4DH5 (blue) at 100 μM concentration. c) Comparison of far UV CD spectra of 4DH3 (black), 4DH4 (red) and 4DH5 (blue) at 1 mM concentration. All experiments were performed in H_2O at pH 3.5.

Figure 3. NMR characterization of the three analogues, 1 mM in H_2O/ D_2O (90:10 *v/v*), pH 3.5. a) The 1D ^1H NMR spectra in the NH region of 4DH3 (top in black), 4DH4 (middle in red) and 4DH5 (bottom in blue). b) The 1D ^1H NMR spectra in the αCH region of 4DH3 (top in black), 4DH4 (middle in red) and 4DH5 (bottom in blue).

Figure 4. Dependence of the mean residue molar ellipticity at 222 nm as a function of 4DH3 concentration in 50 mM buffer acetate at pH 4.6. The error bars indicate the standard deviation of three independent experiments. The smoothed curve represents the best fit for the oligomerization equation from monomer to tetramer, as reported in the methods section.

Figure 5. 2D ^1H DOSY spectrum of 500 μM 4DH3 in D_2O, in presence of glycerol as internal standard.

Figure 6. Zn^{2+}-binding properties of 4DH3. a) CD spectral changes of a 40 μM 4DH3 solution (as monomer concentration) upon addition of $ZnSO_4$ (0.2 eq each) in 10 mM MES buffer pH 6.5. The black arrow indicates the observed changes by increasing the metal:peptide ratio up to 2. b) Dependence of the mean residue molar ellipticity at 222 nm as a function of metal:peptide ratio. The dashed line intersects at the 1:1 ratio.

Figure 7. Size-exclusion chromatography of 4DH3 at pH 6.5 (MES 20 mM, NaCl 200 mM), both in the presence (black line) and in the absence (blue line) of 50 μM $ZnCl_2$. In the *apo* state, two single well-resolved peaks are observed, indicative of a momomer-dimer equilibrium. In the presence of zinc, a single poorly defined peak is produced. Multi-peak fitting (green lines and red dashed lines for the sum) led identification of the aggregation state of the three species under the

main peak, resulting in apparent molecular weights corresponding to oligomerization state n = 3.4; 2.0; 0.9, respectively.

Figure 8. Structural differences among the designed models of 4DH3, 4DH4 and 4DH5. The main chain is presented as cartoon, residues in the *b* and *e* (on the left) and *c* and *g* (on the right) positions as sticks. Hydrophobic, polar, negative- and positively-charged residues are coloured in green, magenta, red and blue, respectively. Helices *2* and *3* (on the left) and helices *3* and *4* (on the right) are transparent to allow a clearer view of the interfaces.

Table 1. Far-UV region CD parameters for 4DH3, 4DH4 and 4DH5 at 10 μM concentration in H_2O.

	θ_{192} [a]	θ_{208} [a]	θ_{222} [a]	θ_{ratio} [b]
4DH3	26.57	-14.72	-13.30	0.90
4DH4	14.49	-11.60	-9.24	0.80
4DH5	-2.36	-8.53	-3.96	0.46

[a] The values of θ are reported in kdeg cm^2 $dmol^{-1}$ res^{-1}.

[b] θ_{ratio} is the ratio $\theta_{222}/\theta_{208}$.

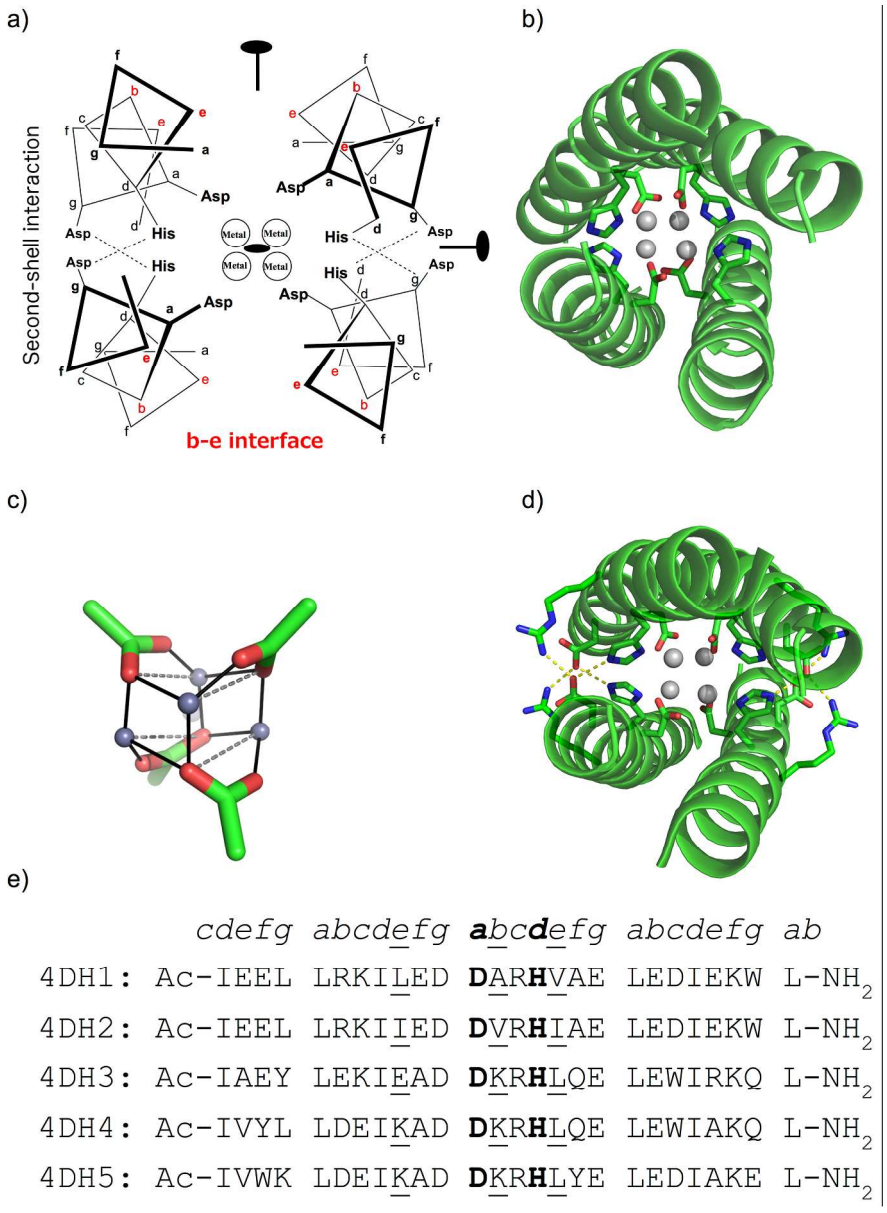


Figure 1. Design of tetranuclear clusters in helix bundles. a) D_2-symmetrical 4DH analogues, bearing the DXXH metal-binding motif, have four Asp and four His residues coordinating 4 ions at the core. Second-shell Asp residues at the *c-g* interface (shown in red) are included to interact with the first-shell His ligands. The positions of the three orthogonal two-fold axes are also indicated. b) Crystal structure of 4DH1 (PDB ID 5WLL): the main chain is presented as cartoon, zinc ions as spheres, and coordinating residues as sticks. c) Designed tetranuclear zinc cluster, showing the relative orientations among the carboxylates. d) Model structure of 4DH analogues, including the first-shell ligands Asp[12] and His[16], the second-shell Asp[11], and the third-shell Arg[14]. e) Sequence alignment of 4DH analogues. Coordinating residues are in bold, while *b* and *e* positions in proximity of the metal-binding site are underlined.

207x281mm (300 x 300 DPI)

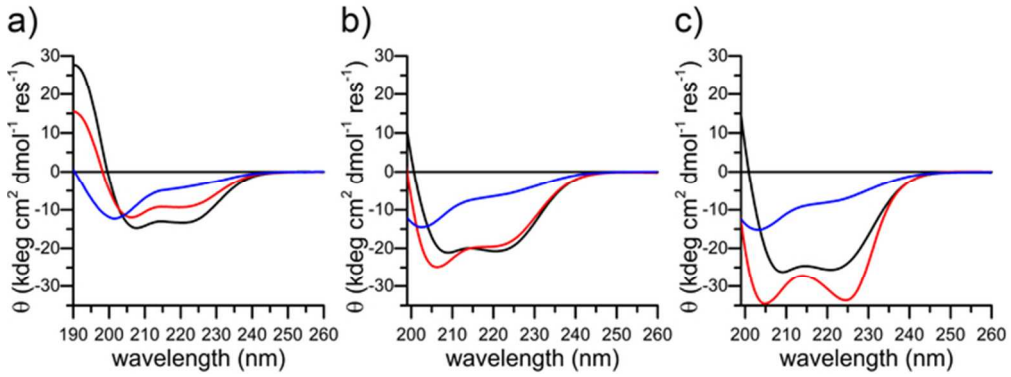


Figure 2. CD characterization of the three analogues as a function of peptide concentration. a) Comparison of far UV CD spectra of 4DH3 (black), 4DH4 (red) and 4DH5 (blue) at 10 μM concentration in H_2O. b) Comparison of far UV CD spectra of 4DH3 (black), 4DH4 (red) and 4DH5 (blue) at 100 μM concentration in H_2O. c) Comparison of far UV CD spectra of 4DH3 (black), 4DH4 (red) and 4DH5 (blue) at 1 mM concentration in H_2O. All experiments were performed in H_2O at pH 3.5.

61x23mm (300 x 300 DPI)

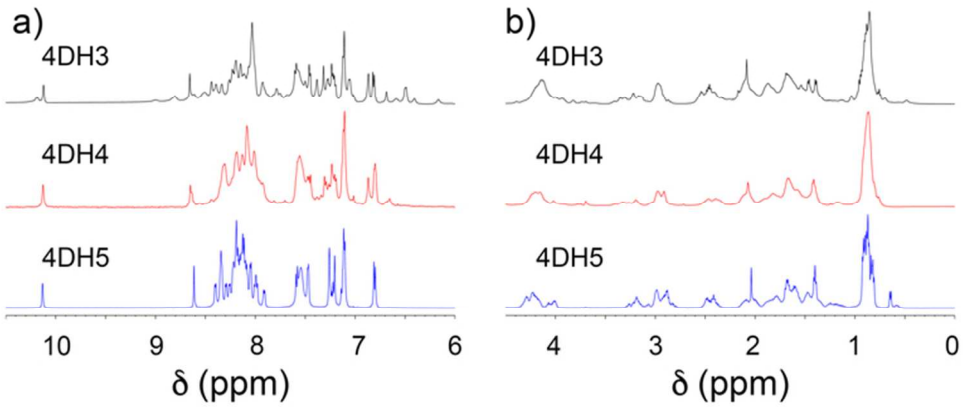


Figure 3. NMR characterization of the three analogues, 1 mM in H_2O/D_2O (90:10 *v/v*), pH 3.5. a) The 1D ^1H NMR spectra in the NH region of 4DH3 (top in black), 4DH4 (middle in red) and 4DH5 (bottom in blue). b) The 1D ^1H NMR spectra in the αCH region of 4DH3 (top in black), 4DH4 (middle in red) and 4DH5 (bottom in blue).

67x28mm (300 x 300 DPI)

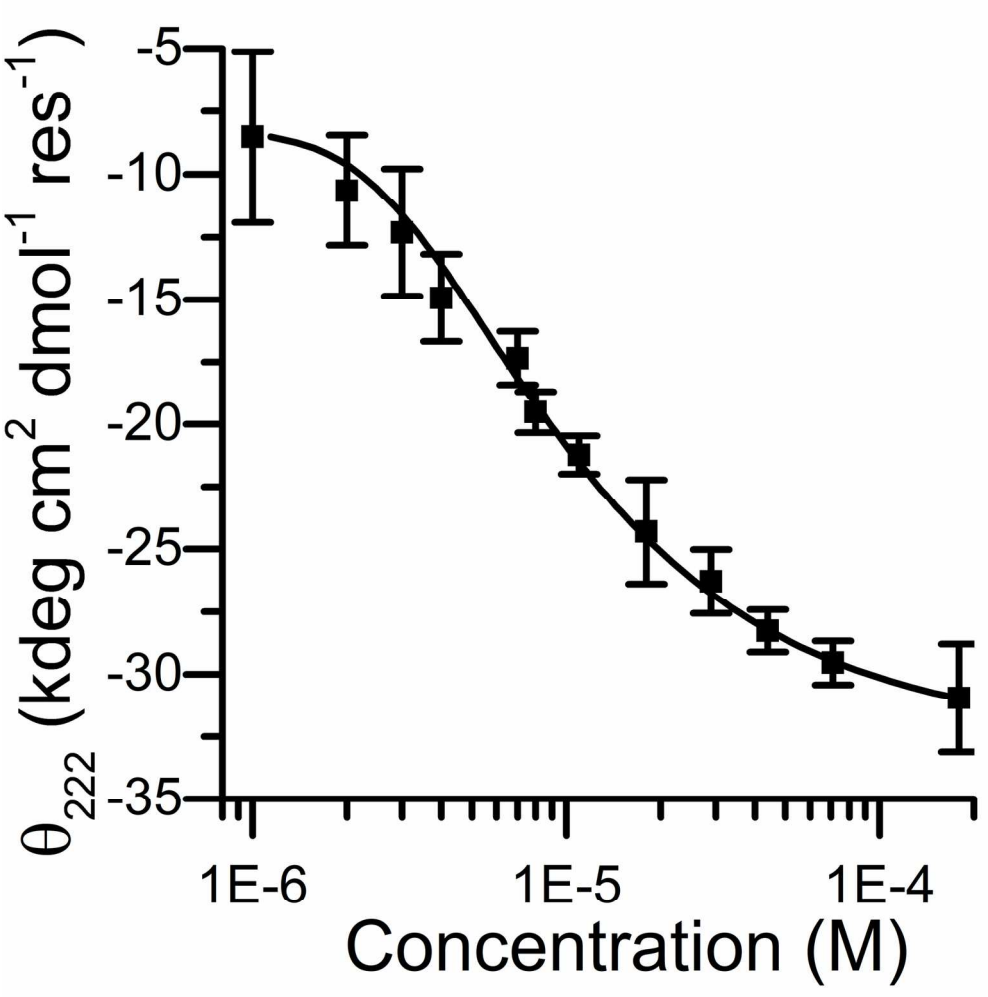


Figure 4. Dependence of the mean residue molar ellipticity at 222 nm as a function of 4DH3 concentration in 50 mM buffer acetate at pH 4.6. The error bars indicate the standard deviation of three independent experiments. The smoothed curve represents the best fit for the oligomerization equation from monomer to tetramer, reported in the methods section.

80x80mm (600 x 600 DPI)

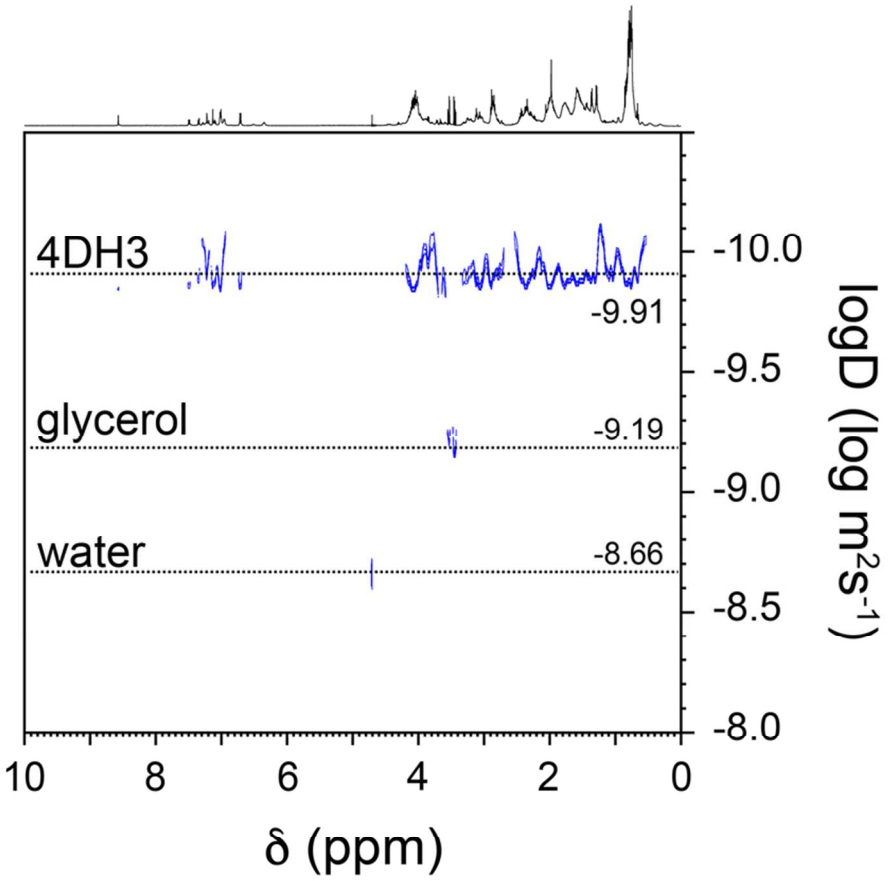


Figure 5. 2D 1H DOSY spectrum of 500 μM 4DH3 in D_2O, in presence of glycerol as internal standard.

80x80mm (300 x 300 DPI)

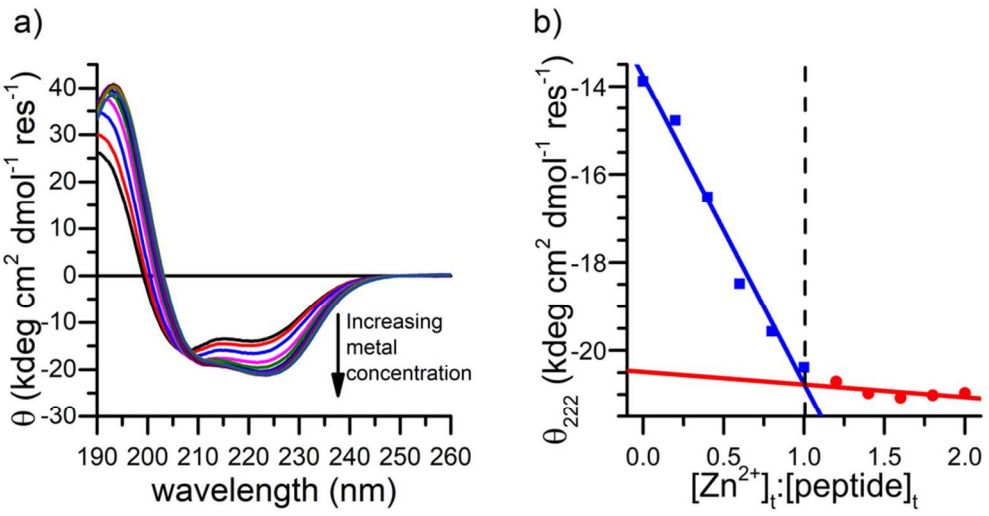


Figure 6. Zn^{2+}-binding properties of 4DH3. a) CD spectral changes of 40 μM 4DH3 solution (as monomer concentration) upon addition of $ZnSO_4$ (0.2 eq each) in 10 mM MES buffer pH 6.5. The black arrow indicates the observed changes by increasing the metal:peptide ratio up to 2. b) Dependence of the mean residue molar ellipticity at 222 nm as a function of metal:peptide ratio. The dashed line intersects at the 1:1 ratio.

87x45mm (300 x 300 DPI)

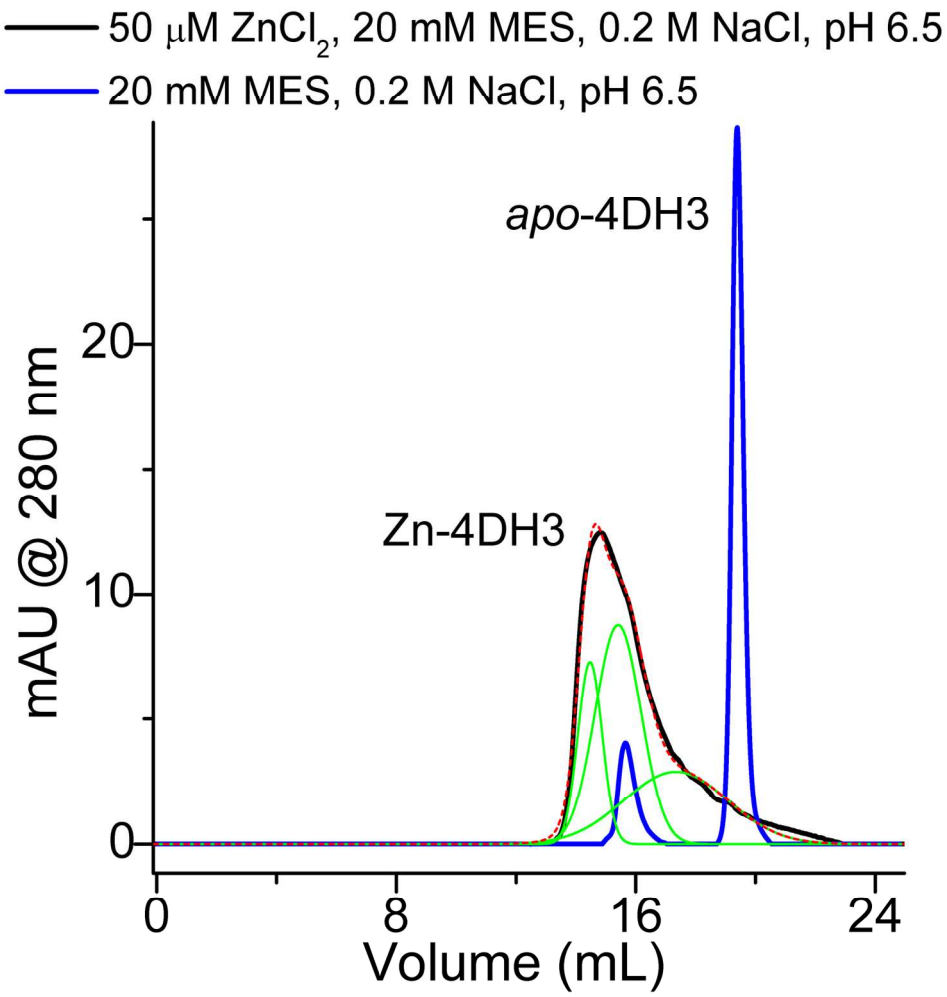


Figure 7. Size-exclusion chromatography of 4DH3 at pH 6.5 (MES 20 mM, NaCl 200 mM), both in the presence (black line) and in the absence (blue line) of 50 μM $ZnCl_2$. In the *apo* state, two single well-resolved peaks are observed, indicative of a momomer-dimer equilibrium. In the presence of zinc, a single poorly defined peak is produced. Multi-peak fitting (green lines and red dashed lines for the sum) led identification of the aggregation state of the three species under the main peak, resulting in apparent molecular weights corresponding to oligomerization state n = 3.4; 2.0; 0.9, respectively.

159x159mm (300 x 300 DPI)

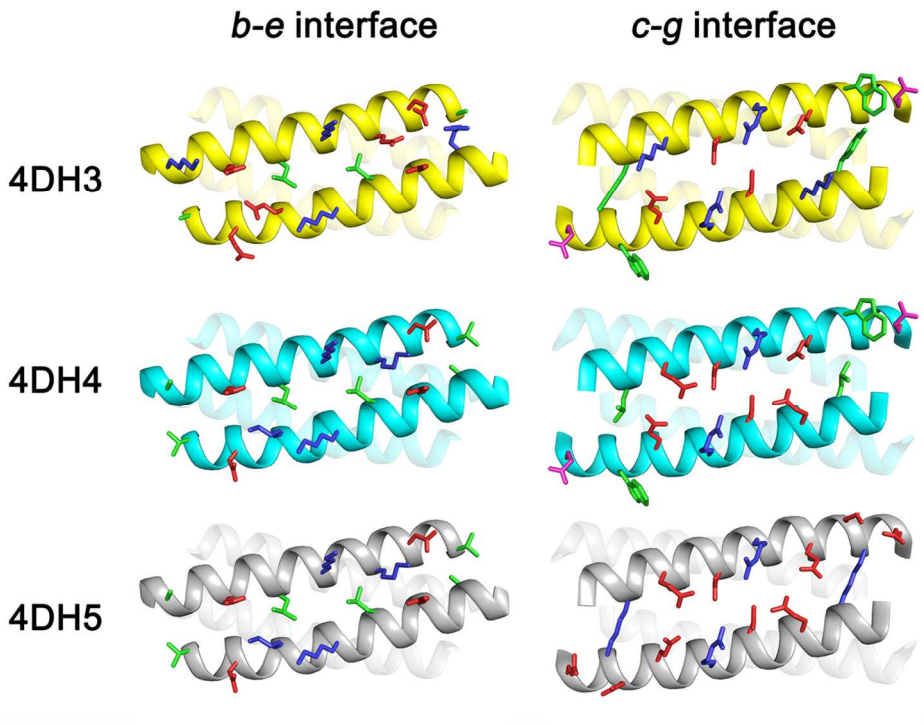
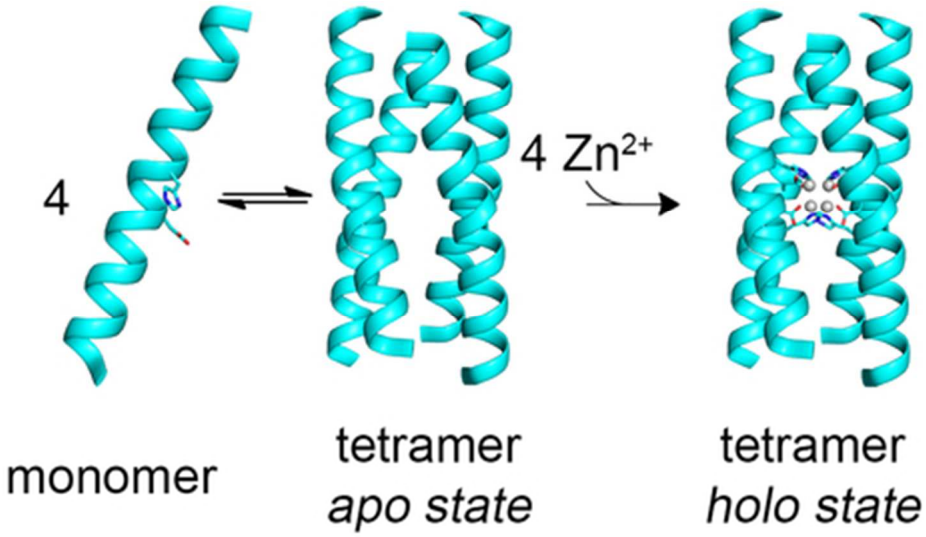


Figure 8. Structural differences among the designed models of 4DH3, 4DH4 and 4DH5. The main chain is presented as cartoon, residues in the *b* and *e* (on the left) and *c* and *g* (on the right) positions as sticks. Hydrophobic, polar, negative- and positively-charged residues are coloured in green, magenta, red and blue, respectively. Helices *2* and *3* (on the left) and helices *3* and *4* (on the right) are transparent to allow a clearer view of the interfaces.

119x89mm (300 x 300 DPI)



39x39mm (300 x 300 DPI)