

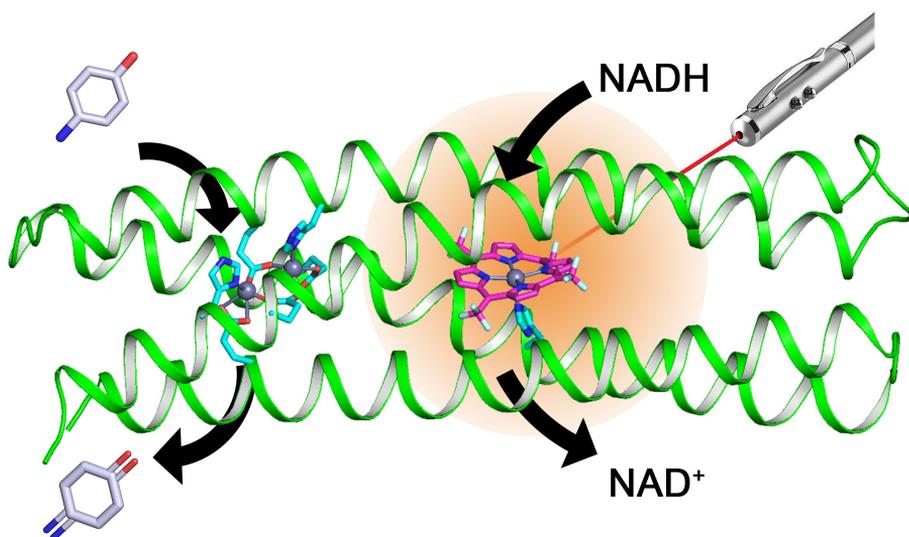
University of Naples Federico II Polytechnic and Basic Sciences School

Department of Chemical Sciences



Ph.D. in Chemical Sciences

De novo design of multi-domain metalloenzymes



Fabio Pirro

Advisor:
prof. Angela Lombardi

Examiner:
prof. Antonello Merlino

XXXIII Cycle 2018 – 2021
Coordinator: prof. Angela Lombardi

THIS PAGE IS INTENTIONALLY LEFT BLANK

*«'O frati,' dissì, 'che per cento milia
perigli siete giunti a l'occidente,
a questa tanto picciola vigilia*

*d'i nostri sensi ch'è del rimanente
non vogliate negar l'esperienza,
di retro al sol, del mondo senza gente.*

*Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza'.*»

Dante Alighieri, Divina Commedia, Inferno canto XXVI, 112-120

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Table of contents

Abstract	1
References	5
Chapter 1: Introduction	7
1.1 The protein language	7
1.1.1. The domain architecture: the word order	9
1.1.2. Evolutionary pathway for ‘words’ combination	10
1.2 Cooperativity in multi-domain protein	13
1.3 Protein Engineering of multi-domain proteins and allostery	14
1.4 De novo protein design	20
1.4.1. De novo design of metalloproteins	22
1.4.1.a. Due Ferri proteins: de novo designed metalloenzymes	27
1.4.2. Design of small-molecule binding proteins.....	35
1.5 Aim of the thesis	40
References	41
Chapter 2: Results and Discussion	54
2.1. Development of a new structural-based strategy for designing multi-domain proteins	54
2.1.1. Design of the first model DFP1.....	57
2.1.2. Structural characterization of DFP1.....	61
2.2. Biophysical characterization of the functional DFP3	63
2.2.1. Binding properties of DFP3 for its metal cofactors.....	66
2.2.2. Functional characterization of DFP3 dimetal site.....	71
2.2.2.a. Stabilization of radical semiquinone in the dizinc form.....	71
2.2.2.b. Allosteric modulation of ferroxidase and oxygenase activities by porphyrin binding	72
2.2.3. Probing allosteric bidirectionality in DFP3	78
2.2.3.a. ZnDP binding alters the thermodynamic stability of DFP3	80
2.2.3.b. Allosteric control on porphyrin binding from dimetal site.....	86
2.2.3.c. ZnDP-DFP structural model.....	87
2.2.4. Photosensitizing properties of zinc porphyrins in DFP3.....	90

Table of contents

References	97
Chapter 3: Conclusions	101
Chapter 4: Materials and Methods	103
4.1. Computational design of DFP proteins	103
4.1.1. Bundle axis determination, co-linear alignment of the two bundles, and generation of all structural poses	103
4.1.2. MASTER searches	104
4.1.2.a. Database of structures	104
4.1.2.b. Structural searches.....	104
4.1.3. Rosetta design.....	104
4.2. Experimental characterization of DFP proteins.....	105
4.2.1. Protein expression and purification	105
4.2.2. Pump-Probe Transient Absorption Spectroscopy	109
4.2.3. DFP1 crystallization and structure determination	109
4.2.4. Size Exclusion chromatography.....	112
4.2.5. Circular Dichroism	113
4.2.6. UV-Vis spectroscopy.....	114
4.2.7. Synthesis and characterization of Fmoc-His-OH	116
4.2.8. MD simulations.....	118
4.2.9. Docking and Minimization	119
4.2.10. Photochemical reaction	120
References	122
List of abbreviations.....	125

Abstract

The course of evolution required the recombination of protein domains to perform ever-growing complex functions. The presence of an additional domain in a multi-domain protein expands, alters, or modulates the functionality with respect to the isolated one-domain protein.¹ In particular, small molecule binding domains have shown a strong propensity to form multi-domain proteins and regulate enzymatic, transport, and signal-transducing domains.² This modulation is referred to as allostery (from Greek, other solid body), as the properties of a functional site are affected by a small molecule bound to a distinctive protein site.³ Taking inspiration from Nature, artificial proteins have been engineered combining different domains to develop bioinspired molecular machines, able to respond to external stimuli.⁴

This Ph.D. project, born from the collaboration of the Artificial Metallo-Enzyme Group and the DeGradoLab, was devoted to the development of a multi-domain protein (Figure 1). This represents the first example of an artificial multi-domain protein, in which allostery was designed completely from scratch.^{5,6}

DF (Due Ferri), a diiron phenol oxidase domain, and PS (Porphyrin-binding Sequence), a zinc porphyrin binding domain, were selected as starting proteins to be combined and give DFP (Due Ferri Porphyrin).⁷ The multiple junctions were exploited to link the two domains, and obtain a more extensive structural coupling between them. While the two metalloproteins present the same kind of domain, the two four-helix bundles are characterized by different geometrical parameters. Therefore, a structural-based methodology was firstly developed in order to identify the best colocalization and helical junctions to accommodate the changes in interhelical separation and registry between the bundles.

Abstract

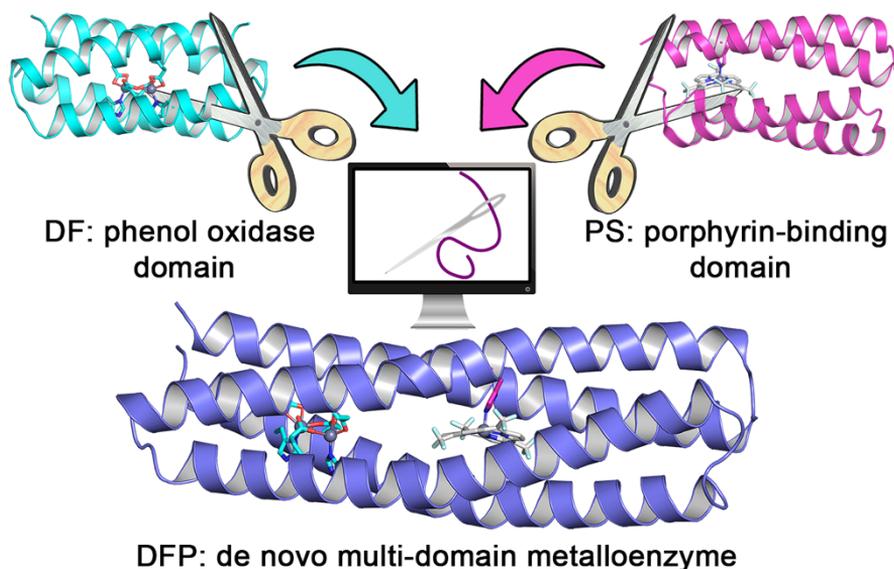


Figure 1. Graphical representation of this work: de novo design of multi-domain metalloenzyme

The x-ray structure of the first analogue, DFP1, was determined, bound to its metal cofactors. The superposition of the 120 residues comprising binding sites gave an excellent fit to the design model, with an overall backbone RMSD of less than 1.4 Å. However, DFP1 was designed to maximize structural stability with a tight and uniform packing, which hindered the access to organic substrates at the DF domain and, thus, its functional characterization. The channel-lining residues of the dimetal-binding site in DF domain were mutated in Gly residues to create a pocket for a substrate. The introduction of helix-breaking residues, that gave oligomerization promiscuity, required also the mutation of DF loop, leading to the final candidate DFP3.

An extensive spectroscopic characterization was performed to investigate the functional properties of the multi-domain proteins. DFP3 was demonstrated to bind the designed zinc porphyrin ZnP (Zn-meso-(trifluoromethyl)porphin) at the PS domain with nanomolar affinity. The strong negative Cotton Effect in the ZnP

Soret region confirmed the tight and single-mode binding in the rigid asymmetric protein core. On the other side of the multi-domain metalloprotein, cobalt binding experiments confirmed the preservation of the DF penta-coordinating environment. Indeed, the dizinc form was able to stabilize the semiquinone form of 3,5-ditertbutylcatechol/quinone couple, and DFP3 showed ferroxidase and phenoloxidase activities. Although these reactivities were still present upon ZnP binding, a modulation effect was observed. The catalytic characterization of 4-aminophenol oxidation demonstrated a Michaelis-Menten mechanism in the phenoloxidase activity, and high-lightened a 4-fold tighter K_m and a 7-fold decrease in k_{cat} upon binding of ZnP. Molecular Dynamics simulations suggested that the presence of ZnP restrains the conformational freedom of a second-shell Tyr, that have been previously shown to largely affect the reactivity of the diiron center.

Subsequently, the binding fitness of the zinc porphyrin was changed to investigate the bidirectionality of the allosteric regulation. In the presence of the different zinc porphyrin ZnDP (ZnDP, Zn-Deuteroporphyrin IX), DFP3 resulted to be more flexible, as demonstrated by thermal and chemical denaturations. Nevertheless, the dizinc center continued to stabilize the semiquinone, and the ferroxidase and phenol oxidase activities were still modulated by the presence of ZnDP. DFP3 showed an excellent affinity for ZnDP, only one order lower in magnitude compared to the designed ZnP. More importantly, the ZnDP affinity was modulated by the presence of zinc bound to DFP3, showing a 3-fold decrease in K_D , and demonstrating the presence of a back-regulation.

In final instance, the photosensitizing properties of zinc porphyrin-DFP3 complexes were tested in the oxidation of the biological redox cofactor NADH. The photocatalytic characterization highlighted the paramount role of the protein scaffold not only in increasing the reaction rate, but also in protecting

Abstract

the zinc porphyrins from highly reactive species. The lower binding fitness DFP3 towards ZnDP hindered this protection, enabling a major permeability of these species and leading to the zinc porphyrin photobleaching.

Although only a preliminary characterization of photocatalysis has been performed, the high reactivity and versatility of such systems are a promising starting point for the de novo design of artificial photosystems for the storage of light energy in chemical fuels.

References

1. Bashton, M. & Chothia, C. The Generation of New Protein Functions by the Combination of Domains. *Structure* **15**, 85–99 (2007).
2. Anantharaman, V., Koonin, E. V. & Aravind, L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains¹¹Edited by F. Cohen. *J. Mol. Biol.* **307**, 1271–1292 (2001).
3. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
4. Ostermeier, M. Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel.* **18**, 359–364 (2005).
5. Researchers design allosteric protein from scratch. *Chemical & Engineering News* <https://cen.acs.org/biological-chemistry/Researchers-design-allosteric-protein-scratch/98/i48>.
6. Pirro, F. *et al.* Allosteric cooperation in a de novo-designed two-domain protein. *Proc. Natl. Acad. Sci.* **117**, 33246–33253 (2020).
7. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).

Chapter 1: Introduction

1.1 The protein language

The entire human genome was identified in 2000 and it was immediately considered as 'The book of life'.¹ For the first time, a linear string of four letters 'ACGT', corresponding to the four nucleotides adenosine, cytosine, guanine, and thymine, respectively, contained all the information for the development, growth, reproduction, and regulation of the human cell. Actually, DNA sequences have been treated as a text since its structure was resolved in 1953.² However, this linguistic analogy was applied not only for DNA, but also for the other biological macromolecules.³ Although a 'translation' is possible, each biological macromolecule has its own language, with its vocabulary and grammar.

In particular, if protein language is considered (Figure 1.1):

- The amino acid residues are the letters, and, therefore, the primary sequence corresponds to the lexicon;
- The tertiary structure is the syntax, ruled by the grammar of the protein folding. Distant 'strings' in the primary sequence interact creating dependencies with each other;
- The semantics of this language is the protein function, that has changed in a diachronic way during evolution;
- The protein role, its pragmatism, depends on the larger context of the cell compartment in which is localized.⁴

Chapter 1

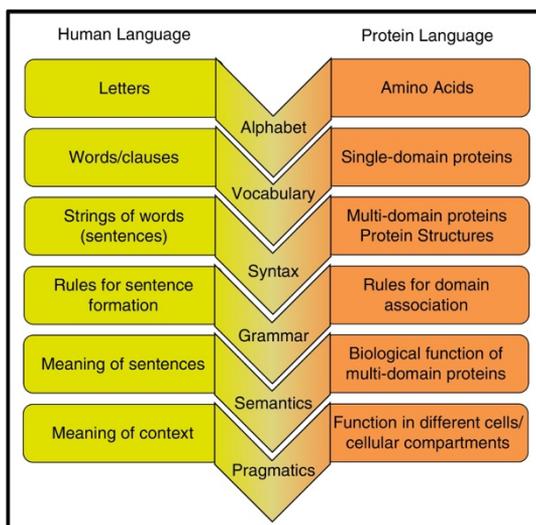


Figure 1.1. Analogy between human and protein languages. Adapted with permission from reference ⁴. *Copyright © 2015 Elsevier Ltd.*

In the protein ‘sentence’, the amino acids represent only the letters. Therefore, they are not able to carry the semantic information. In linguistic, the smallest meaningful elements of speech are words,⁵ and this role is taken by the domains in the protein language. The polypeptide ‘sentence’ can be made up of a single domain or a combination of multiple ones. A domain is a distinct, compact and stable protein structural unit, able to fold independently and highly conserved during evolution.⁶ The structural similarity among domains characterized by different functions underlined a ‘semantic plasticity’, more similar to ‘clauses’ than to ‘words’.⁷ However, the frequency distribution of domains encoded in a genome follows the Zip’s power-law, with a slope very similar to that observed for the frequency distribution of words in natural languages.⁸ This confirmed the close analogy between words and domains and their paramount role in their own languages.

1.1.1. The domain architecture: the word order

A protein domain architecture is the order of the domains along the polypeptide sequence from the N- to the C- terminus.⁹ Therefore, it can be considered as the ‘word order’ of the protein language. Single and multi-domain proteins can be found in Nature. In the same way, single and multi-domain architectures (SDA and MDA, respectively) can be identified. In turn, two topological architectures are more common for MDAs: *i)* the C- terminus of the previous domain is linked to the N-terminus of the following one, in an end-to-end fashion (Figure 1.2, left); *ii)* a domain is inserted in the other one, creating a continuous and a discontinuous domain (Figure 1.2, right).¹⁰

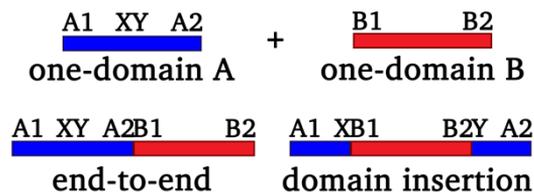


Figure 1.2. Representation of the multi-domain architectures found in Nature: end-to-end linkage (on the left) and domain insertion (on the right).

In particular, for the second topology was observed that:

- The inserted domain is usually the smallest among the ones in the MDA;
- The location of the insertion occurred in the last third of the protein sequence, and specifically in loop regions in the 3D structure of the parent domain;
- The ends of the inserted domain are within 8 Å of each other in the crystal structure.¹¹

With the determination of an increasing number of genomes, clear mapping of the differences among the superkingdoms has been accessible. A large fraction

Chapter 1

of SDAs (more than the 50%) is shared between Prokaryotes and Eukaryotes, in contrast to the small percentage of MDAs (under the 2%).¹² Therefore, it can be considered as the two superkingdoms have in common the same ‘words’, but are combined in different ways in the ‘sentence’.

Furthermore, the propensity of protein domains to form different combinations of MDAs, the so-called domain promiscuity or versatility,¹³ increases with the organism complexity: multi-domain proteins represent only the two-thirds of the proteome in Prokaryotes and more than the four-fifth in Eukaryotes.¹⁴ Moreover, MDAs have been demonstrated to be easily distinguished between the major divisions of Life, corresponding to clade-specific functions for the survival and differentiation of the cell.¹⁵ These findings highlighted the importance of the domain combinations during evolution.

1.1.2. Evolutionary pathway for ‘words’ combination

The population of the protein universe is not evenly distributed. The number of members in the different families has been approximated by the power-law ‘Pareto distribution’.¹⁶ This power-law was applied for the first time in economics to describe the wealth distribution, in which most wealth is held by a small fraction of the population.¹⁷

In the same way, few protein families hold the greatest number of members.¹⁸ Birth, Death and Innovation Model (BDIM) accurately fitted the distributions of domain family size (Figure 1.3): *i*) family growth via domain birth; *ii*) domain death as a result of inactivation and loss; *iii*) innovation or emergence of a new family.¹⁹ Moreover, BDIM did not exclude changes in the birth, death and innovation parameters due to burst of evolutionary activity, rapidly relaxing the domain distribution to a new equilibrium.²⁰

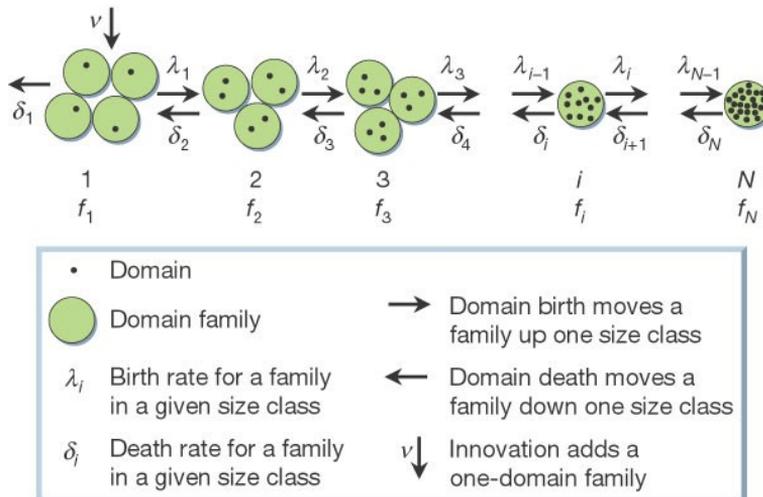


Figure 1.3. A general representation of the BDIM to explain the distribution of domain family size. Adapted with permission from reference ²⁰. Copyright © 2002 Nature Publishing Group.

Not only the frequency distribution of domain family sizes, but also the distribution of domain promiscuity was best approximated by a power-law function.²¹ Even in this case, it was observed that few protein families combined with many others and most families had one or few partners.

The combination of different domains required genetic recombination events during evolution. Although the most common are point mutations (substitution, insertion or deletions of single nucleotides), different mechanisms were recognized to allow great modifications in the genome:

- Homologous and Nonhomologous recombination: DNA lesions, like double-stranded DNA breaks and broken replication forks, can be the cause of the massive loss of genetic information up to cell death. These lesions can arise from ionizing radiation, reactive oxygen species (ROS), DNA replication errors and inadvertent cleavage by nuclear enzymes.²² Homologous recombination uses undamaged DNA templates to repair

Chapter 1

accurately the breaks.²³ In contrast, nonhomologous recombination just modifies and joins the ends.²⁴ Therefore, in the latter mechanism the generation of deletions and/or insertion is unavoidable;

- Transposition and Retrotransposition: transposable elements are DNA sequences able to move from one location to another in the genome. They represent a high portion of a genome and were identified both in Prokaryotes and Eukaryotes.²⁵ The transposition involves fragments of DNA, mobilized via a DNA intermediate with either a 'cut-and-copy' or 'peel-and-paste' mechanism.²⁶ In retrotransposition, an RNA intermediate is reverse-transcribed into a cDNA copy that is integrated elsewhere in the genome.²⁷ In particular, retrotransposition of maturated mRNAs is considered one of the major mechanisms in the gene duplication, that provides new genetic material for mutation, drift, and selection to act upon;²⁸
- Exon shuffling: Eukaryotes DNA is made up of protein-coding gene regions interspersed with non-coding ones, lost during the splicing in the mRNA maturation, called exon and introns, respectively.²⁹ In introns, transposable elements are highly present. Therefore, they promote the formation of rearranged genes by recombination of different exons.³⁰ This process is referred to as Exon shuffling.³¹ Moreover, point mutations can change the splicing pattern, resulting in the addition or deletion of a whole sequence of coding DNA.

The power-law relationship between abundance and promiscuity was consistent with a random process of domain combination in multidomain architectures. However, a simple stochastic model was not able to describe the observed domain versatility in genomes. In particular, it could not take into account the

high propensity of domain combinations, called supra-domains, in the duplication process.³² Moreover, the number of observed combinations was only a small fraction of all the possible combinations out of all families.³³ Thus, selection occurred in order to retain particular domains and supra-domains.

1.2 Cooperativity in multi-domain protein

The promiscuity of a domain depends on the frequency of genetic events that can lead to its genome recombination. However, the proliferation of such domains was dictated by their features in:

- Structural stability: the independence in the domain folding ensured the minimal influence of the neighboring domains in MDAs;³⁴
- Functional utility: these domains maximize the number of interacting partners and, thus, mediate protein-protein interactions, which played a key role in the creation of upgraded organizational units, like in transcription and signaling networks.³⁵

The functionality of the module can be preserved compared to the homologous one-domain protein. However, severe changes in function have emerged joining different domains, particularly at the newly high-susceptible interfaces.³⁶ The interactions between domains at the interfaces are communication vehicles within the protein and are a requisite for protein function and regulation. The physical-chemical analysis of domain-domain interfaces showed intermediate features between permanent and non-obligate ones. Moreover, the nature of amino acid residues at the interfaces seemed more similar to those at the domain surfaces compared to the cores. These suggested an independent

Chapter 1

folding pathway for the separate domain and a subsequent collapse, mediated principally by polar residues.³⁷

Moreover, domain interfaces have been identified as the primary location of regulatory sites in catalytic proteins and binding sites for signaling molecules and, thus, play a key role in allostery.³⁸ In allosteric regulation, the binding of a small molecule in one site affects the properties of a distinctive functional site.³⁹ In this perspective, small molecule binding domains have been identified and their analysis revealed a strong tendency to form multidomain proteins by fusing with enzymatic, transport or signal-transducing domains.⁴⁰

More in general, a modular domain can allosterically control the activity of an 'output' module with a steric or conformational mechanism.⁴¹ In steric allostery, the modular domain physically interacts with the output module, inhibiting it. The conformational allostery resembles more to the traditional model, in which the modular domain has an impact on the conformational equilibrium of the output domain between its active and inactive states.⁴²

1.3 Protein Engineering of multi-domain proteins and allostery

The presence of additional domains can lead to an alteration, an expansion or a modulation in the activity of a catalytic module compared to the one-domain protein.⁴³ This modulation effect can be exploited to design artificial multi-domain proteins with desired characteristics. Protein engineers have applied domain-fusion in order to obtain improved physio-chemical properties (thermal stability, solubility, expression and purification yield) and to create nanostructured and sensor devices.^{44–46} However, the design of new functional

proteins, efficiently controlled by an external stimulus, is still considered a challenging task.⁴⁷

To the endeavor of designing multi-domain proteins, different methodologies have been developed:

- Natural or designed linkers: the domains in natural multi-domain protein are joined by peptide linkers. Naturally-occurring linkers were grouped in small, medium and large, ranging from 3 to 30 amino acid residues, to enable different degrees of spacing.⁴⁸ The direct fusion between domains may cause the misfolding of the chimeric protein, low yield in expression or functional inactivity.⁴⁹ Moreover, domain linkers have been designed in function of the degree of freedom to achieve between the domain.⁵⁰ Flexible linkers, rich in small polar or polar residue, have been used to obtain a certain degree of movement or interaction (Figure 1.4.A). On the other hand, structurally rigid linker, either helical or proline-rich, allowed the effective separation between the domain to avoid mutual interference (Figure 1.4.B). Although the linker has to be enough long to prevent mutual exclusion between the two domains, it should not be too long to prevent coupling and let the domains be two discrete entities. Parker and coworkers identified a flexible 19-residue linker, enabling the coupling between an ACT domain and a 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAH7PS) protein without disrupting their folding.⁵¹ The former is a ligand-binding regulatory module in several proteins of diverse function, the second is an enzyme involved in the biosynthesis of the aromatic amino acids. The catalytic performance of the chimeric enzyme is allosterically modulated in presence of tyrosine or phenylalanine. The ACT domain undergoes a

Chapter 1

conformational reorganization upon binding of these effectors, preventing access to the DAH7PS domain;

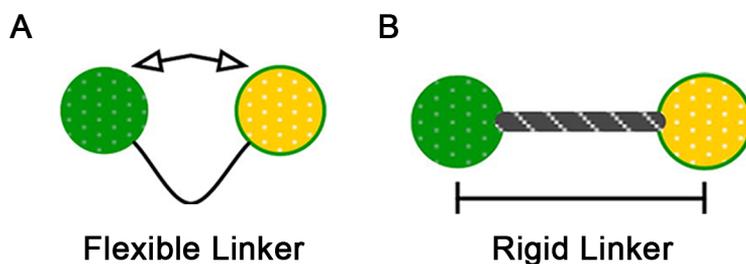


Figure 1.4. Representation of the two classes of protein domain linkers: **A.** flexible, to favor the interaction of the two domains; **B.** rigid, to create a stable distancing between them. Adapted with permission from reference ⁴⁹. *Copyright © 2012 Elsevier B.V.*

- Shared helix method: the helical terminus of the two domains are end-to-end fused, superimposing 1-2 turns to ensure helical continuity, without creating any steric clashes between other regions in the two domains (Figure 1.5).⁵² This type of superposition ensures that the three-dimensional spacing between all structural elements adjacent in the primary sequence.⁵³ This method has been demonstrated to be efficient in developing large symmetric protein complexes, well-behaved folded multi-domain and artificial repeat proteins.^{54,55} In particular, Sosnick and coworkers developed a multi-domain DNA-binding protein, allosteric modulated upon photo-excitation.⁵⁶ The shared helix, well packed in both domains in the resting state, was used to propagate the conformation changes of a photoactive LOV (Light, Oxygen, or Voltage) domain to the bacterial transcription factor trp repressor. The difference in the helix packing between the two domains, in the resting and photo-excited state, were successfully used to modulate the affinity of the 'output' domain for the substrate;

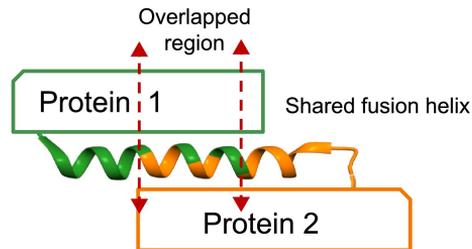


Figure 1.5. Representation of the “shared fusion helix” methodology: the helical ends of the two proteins are superimposed for one or two turns to create the single construct. Adapted with permission from reference ⁵². Copyright © 2019 Elsevier Ltd.

- Domain insertion: The insertion of one gene in another is not only used in Nature, but also in protein engineering to develop allosteric switches and enzymes. Many proteins have their termini in close contact.⁵⁷ Therefore, a protein can be inserted in a loop of another one by computational design or random gene fusion.⁵⁸ With this methodology, fluorescent sensors, molecular switches, allosteric enzymes have been developed.^{59–62} Ostermeier and colleagues gave a great contribution in this field. In particular, they have inserted different β -lactamases in the maltose-binding protein.^{63,64} Not only they successfully modulated the activity of the inserted domain by the binding of maltose in the discontinuous one, but they also observed that zinc ions were negative effectors (Figure 1.6). The chimera has a high-affinity binding site, while the isolated parental proteins were found to not bind zinc significantly. Therefore, it was shown here that domain fusion can result in emergent properties, thanks to the high coupling between domains.⁶⁵

Chapter 1

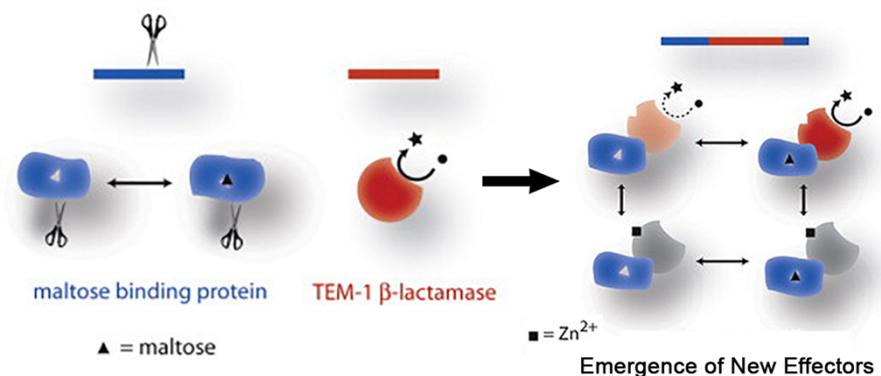


Figure 1.6. Schematic illustration of the TEM1 β -lactamase domain insertion into maltose-binding protein. The resulting multi-domain protein is able to bind zinc ions, a function absent in the parental one-domain proteins. Adapted with permission from reference ⁶². Copyright © 2009 Elsevier Ltd.

- **Circular Permutation:** A peptide linker can be used to join the starting N- with the C-terminus, and a cleavage made in a different location of the protein sequence to generate the new termini. This process is referred as to Circular Permutation and does not result in any amino acid substitutions, but only in a reorganization of the primary sequence.⁶⁶ Although some permutation sites prevent the protein from folding correctly, many permutants have been created with nearly identical structure and function to the original protein.⁶⁷ This methodology has been successfully used for domain insertion. The new termini of the continuous protein can be used as tether points for the insertion in a loop of the discontinuous domain. More importantly, multiple choices for the new termini position can be performed to optimize the relative orientations of the two domains. Such rearrangements can finely colocalize the two domains to achieve the desired coupling between them.⁶⁸ The most impressive fluorescent sensor created with this method is a circular permuted Green Fluorescent Protein (GFP)

inserted into a *Xenopus* calmodulin (a calcium-binding protein).⁶⁷ The calmodulin domain undergoes a large conformational change upon binding of calcium ions that, in the fusion, results in a change in the protonation state of the fluorophore and the consequent increase in the GFP fluorescence (Figure 1.7).

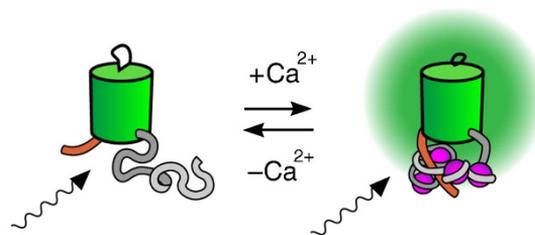


Figure 1.7. Graphical representation of the conformational and photophysical properties of a GFP-calmodulin chimera upon binding of calcium ions. Adapted with permission from reference ⁶⁹. Copyright © 2013 Elsevier B.V.

- **Fragment Assembly Approach:** in this methodology, a large number of backbone junctions were generated with a Rosetta Monte Carlo protocol to link the two domains (Figure 1.8). These fragments consisted of a combination of small secondary structure elements, ranging between 2 and 64 amino acid residues. Subsequently, Rosetta design was used to design the amino acid sequence in the junctions and their neighboring residues in the starting domains. Finally, models containing buried unsatisfied hydrogen bonds or low interacting interfaces were filtered out. This protocol enabled the design of multi-domain proteins with a wide range of shapes through a modular combination of blocks.⁷⁰ However, the development of allosteric modulated multi-domain protein with this methodology has not been reported yet.

Chapter 1

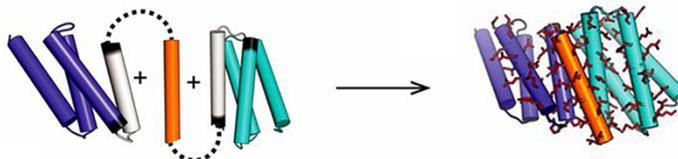


Figure 1.8. Representation of the “Fragment Assembly Approach” methodology: secondary structure fragments are randomly built to join the two structures; subsequently sequence design is performed to nicely pack the linking fragment with the starting proteins. Adapted with permission from reference ⁷⁰. *Copyright © 2020 National Academy of Sciences.*

In each case, the domain architectures of the artificial multi-domain proteins fell into either end-to-end or domain insertion topology.

1.4 De novo protein design

Protein folding is a purely physical process that depends only on the specific amino acid sequence of the protein and the surrounding solvent. The available protein sequences and structures are often referred to as the “Protein Universe”.⁷¹ Considering Anfinsen’s hypothesis, it should be possible to predict the structure of a protein if a model of the free energy is available and if the global minimum of this function can be identified.⁷² Therefore, Protein Design is a valuable tool for understanding the fundamental laws that correlate protein sequence to structure and function.⁷³

De novo design involves the construction of a polypeptide that is not directly related to any natural protein, yet folds into a predicted well-defined three-dimensional structure.⁷⁴ Evolution has explored only tiny regions of the Protein Universe, which are not uniformly widespread as evolution proceeds by

incremental mutation. The huge space that is unlikely to be sampled during evolution is thus accessible only by de novo protein design.⁷⁵ Following the de novo design approach, Top7, the first designed protein, the fold of which has not yet been found in nature, has been developed.⁷⁶

De novo design has been successfully applied to the development of two classes of proteins, crucial for life: metalloproteins and small molecule-binding proteins. Metalloproteins account for nearly half of all proteins in nature and are responsible for catalyzing important biological processes, such as photosynthesis, respiration, water oxidation, molecular oxygen reduction and nitrogen fixation. In order to understand the laws that regulate these reactions and to develop artificial metalloenzymes that exceed the performance of natural counterparts, metalloproteins have become the main target of protein de novo design.^{77,78} Due to the redox properties of their cofactor, the most studied metalloenzymes contain iron, primarily, and copper, secondly.⁷⁹

The critical issues that have to be addressed in designing metalloenzymes are: i) the correct fold of the protein scaffold; ii) the coordination requirements of the metal ion; iii) the stability/activity trade-off, which is essential for the structure and the function.⁸⁰

The small molecule-binding proteins are also of paramount importance. Indeed, cell survival requires the regulation of many functions of proteins by their interactions with small-molecule ligands such as metabolites or drugs.⁸¹ The de novo design of small molecule binding protein can be considered as the inverse of drug design.⁸² While in the drug design, a small molecule is developed to bind with high affinity a target protein, in the inverse drug design a protein scaffold is built with a pocket able to bind a specific target. Therefore, if small molecule-binding proteins can be designed completely from scratch, a deeper

Chapter 1

understanding of the relation between the protein matrix and the target molecule can be led to great advancement of basic science and drug development.⁸³

In the design of small molecule-binding proteins, the binding site should have: *i)* specific hydrogen-bonding and van der Waals interactions with the ligand; *ii)* overall complementarity to the ligand; *iii)* structural pre-organization in the unbound protein state, which minimizes entropy loss after ligand binding.⁸⁴

1.4.1. De novo design of metalloproteins

The following paragraph describes some outstanding examples of de novo designed metalloproteins, containing different metal centers, in order to highlight the progress made in the field.

A metal-binding site can be carved inside the protein scaffold designing: *i)* a perfect preformed coordination environment for the metal, which must therefore generate the highest affinity; *ii)* a precursor structure, in which the cofactor site is created by induced fit at the expense of binding energy resulting in an energized environment.⁸⁵

Copper proteins have been among the subjects of choice for the protein design to test how deep is our understanding of metalloproteins structure and function. Indeed, the differences in the coordination preferences between the two oxidation states require a subtle control on the protein scaffold.

An early example of de novo copper proteins was reported by Tanaka and coworkers.⁸⁶ A single chain four-helix bundle, AM2, was designed based on the parallel GCN4-pLI tetramer, and demonstrated to bind a copper ion in a two-His-one-carboxylate environment. The AM2 scaffold was demonstrated to be highly tolerant to mutations. The single mutation of the carboxylate amino acid to

cysteine led to AM2C (Figure 1.9.A),⁸⁷ a mimetic of the Type I copper proteins, referred as to cupredoxins, as they have the predominant role to transfer electrons.⁸⁸ In a subsequent redesign,⁸⁹ the single copper ion binding environment was duplicated in bi-AM2C to reproduce the Cu_A center of cytochrome c oxidase (Figure 1.9.B), involved in the electron pumping for the 4-electron reduction of oxygen to water.⁹⁰

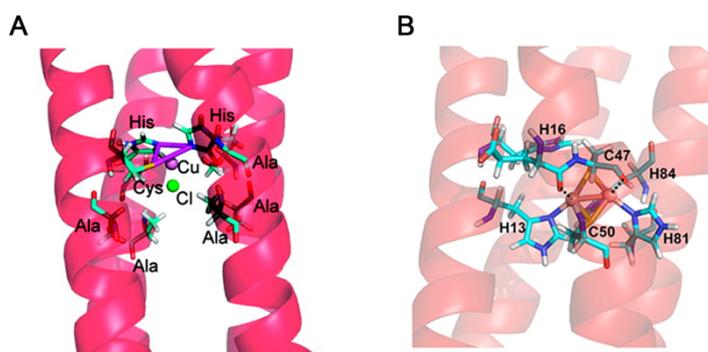


Figure 1.9. Computational models of: **A.** Type I copper site inside AM2C; **B.** Cu_A site in bi-AM2C. Backbone is represented as cartoon, the residues in the metal site as sticks, the ions as spheres. Adapted with permission from references ⁸⁷ and ⁸⁹, respectively. Copyright © 2010 and 2012 American Chemical Society

A great contribution to the design of copper metalloproteins has been made by Pecoraro and coworkers.⁹¹ In several three-helix bundles, copper type I and II were engineered.^{92,93} The three-helix bundles were demonstrated a simple but stable scaffold to disentangle the effect of mutations of the protein matrix around the metal center, to finely modulate copper-binding affinity, reduction potentials, protonation equilibria and catalytic performances of the metal site (Figure 1.10).^{94–97}

Chapter 1

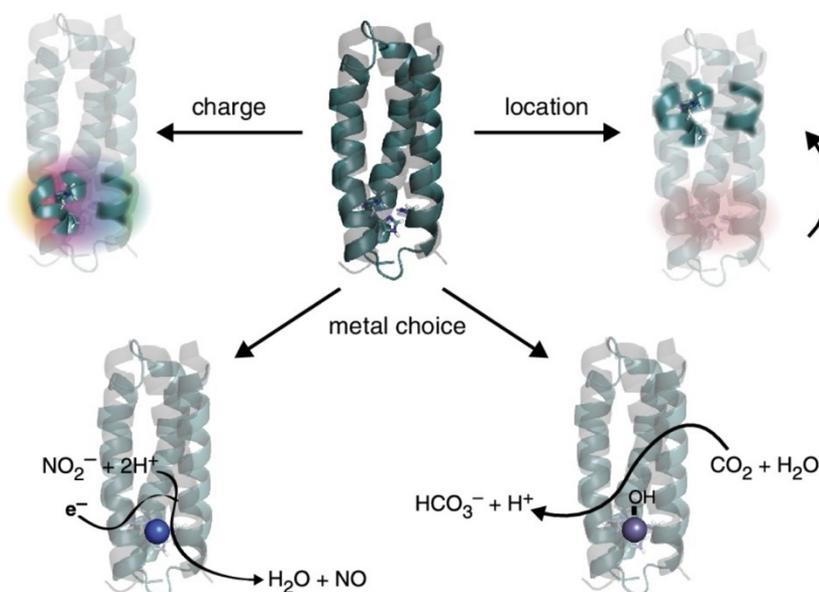


Figure 1.10. The structural robustness of three-helix bundles enables the studies of the effects of charge, location and metal nature on the catalysis. Adapted with permission from reference ⁹⁸. Copyright © 2015 Elsevier Ltd.

Peacock and coworkers have engineered a lanthanide(III)-binding site in a similar three-helix scaffold.⁹⁹ The presence of three negatively charged aspartic residues in the first coordination sphere enabled MB1 peptide to selectively bind several lanthanide (III) ions with micromolar affinity. In particular, Gd(III)-MB1 was demonstrated to be a promising contrast agent in magnetic resonance imaging (MRI), enhancing the transverse magnetic resonance relaxation times of water protons. Moreover, the location of the metal-binding site and the mutation of second coordination sphere residues were used to modulate the hydration state of the lanthanide ions and, thus, its MRI efficiency.^{100,101}

Moreover, Pecoraro and colleagues engineered in the three-helix bundle a catalytic active zinc site.^{102,103} In Nature, the Lewis acid properties of zinc ion are exploited in several metal-dependent Hydrolases, such as Carbonic Anhydrase, Esterase, Phosphatases, Peptidases, and Metalloproteases.¹⁰⁴ The zinc ion

promotes the deprotonation of a bound water molecule to generate hydroxide ion, a potent nucleophile for many hydrolytic reactions. The three-helix bundle developed was made up of a $\text{Zn}^{2+}\text{His}_3\text{OH}$ tetrahedral coordination site for the catalysis and a distinct $\text{Hg}^{2+}\text{Cys}_3$ site for structural stability. The artificial zinc metalloenzyme showed unprecedented catalytic efficiency in CO_2 hydration, with an efficiency within 500-fold that of the fastest isozyme Carbonic Anhydrase II.¹⁰⁵

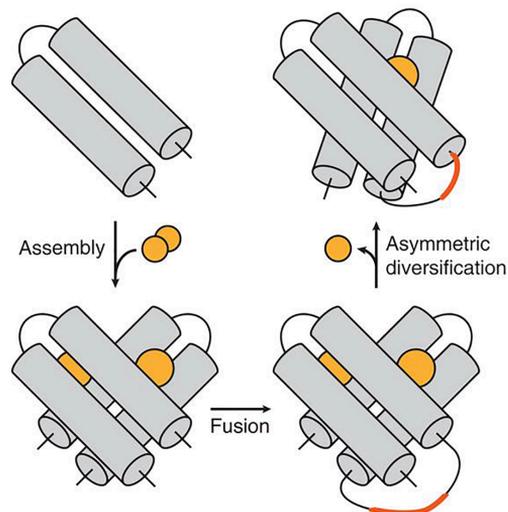


Figure 1.11. Representation of MID1sc design: starting from the zinc-mediated assembly of helix-turn-helix chains, the single-chain construct was obtained by fusion and asymmetric diversification, resulting in a single zinc-binding site. Adapted with permission from reference ¹⁰⁶. Reprinted with permission from AAAS.

With the combination of de novo design and directed evolution, Hilvert and coworkers developed a highly reactive and efficient esterase.¹⁰⁶ They started from a zinc-binding helix-turn-helix homodimer and joined the C-terminus of one chain with the N-terminus of the other one. Diversifying one of the two chains, the single-chain protein, MID1sc housed a single Zn(II)His_3 active site (Figure

Chapter 1

1.11). Several rounds of directed evolution led to the leading protein MID1sc10, which was demonstrated to catalyze the hydrolysis of p-nitrophenyl acetate with enhanced performances with respect to the starting protein and several artificial Esterases and Carbonic Anhydrases for the same substrate.

The de novo design metalloproteins containing iron ions, not bound in any cofactor such as porphyrins or iron-sulfur clusters,^{107,108} has been mainly focused on mimicking rubredoxins. Rubredoxins are small electron transfer proteins coordinating one single iron by four cysteines.¹⁰⁹ The tetrahedral environment around iron is the canonical building unit of iron-sulfur clusters and, therefore, rubredoxins represent the model proteins of this class.

Several groups have managed to reproduce the metal-binding properties of this class with different approaches and scaffolds. Pavone and coworkers developed an undecapeptide homodimer with a miniaturization approach, that folded around the metal with two alpha turns.¹¹⁰ DeGrado and coworkers, instead, used a de novo designed scaffold made up of two beta hairpins linked by the tryptophan zipper hairpin motif.¹¹¹ Jacques et al. reported a cyclic peptide of eighteen residues with a linear tail in a zinc ribbon motif.¹¹² Pecoraro and coworkers engineered the rubredoxin coordination environment in a single chain three-helix bundle, utilizing a secondary structure far from that observed in the natural proteins.¹¹³

Lombardi, DeGrado and coworkers significantly contributed to the de novo design of iron-containing metalloproteins, by developing Due Ferri (two-iron in Italian, DF) proteins. They represent a class of oxygen-activating proteins, whose progress allowed crossing special milestones in de novo protein design in their seminal work.¹¹⁴

1.4.1.a. Due Ferri proteins: de novo designed metalloenzymes

The DFs are inspired by the diiron-oxo proteins. This class of proteins shares, in a four-helix bundle, the ExxH motif, where a glutamate residue and a histidine residue are involved in the binding of two iron ions (Figures 1.12.A and 1.12.B).¹¹⁵ This motif performs several functions, ranging from dioxygen transport and activation, to phosphoryl transfer, to iron storage.¹¹⁶

In order to design a minimal four-helix diiron protein, the authors started from first principle with a parametric approach. After examination of the backbone parameters that define the geometry of the active site bundles of natural diiron proteins, a D_2 -symmetric four-helix bundle of identical, unconnected 21-residue helices was built using just six adjustable parameters (Figures 1.12.C and 1.12.D).¹¹⁷ To provide a Glu_4His_2 coordinating environment, the ExxH was engineered in two helices and a single glutamic amino acid in the remaining two. This resulted in a relaxation of the D_2 symmetry to C_2 , leading to homodimeric helix–loop–helix motifs as the most straightforward choice.

Subsequently, keystone residues in the second coordination shell were identified to stabilize the polar metal-binding site in the middle of the protein as in the natural counterparts. Thus, a Lys/Asp/His H-bond network was designed similar to the Arg/Asp/His pattern found in Bacterial Multicomponent Monooxygenases superfamily, while a Tyr residue donates a proton to the non-bridging carboxylate ligand (Figures 1.12.E and 1.12.F). This extensive network of hydrogen bonds stabilizes the primary ligands and modulates the reactivity of the cluster.

Chapter 1

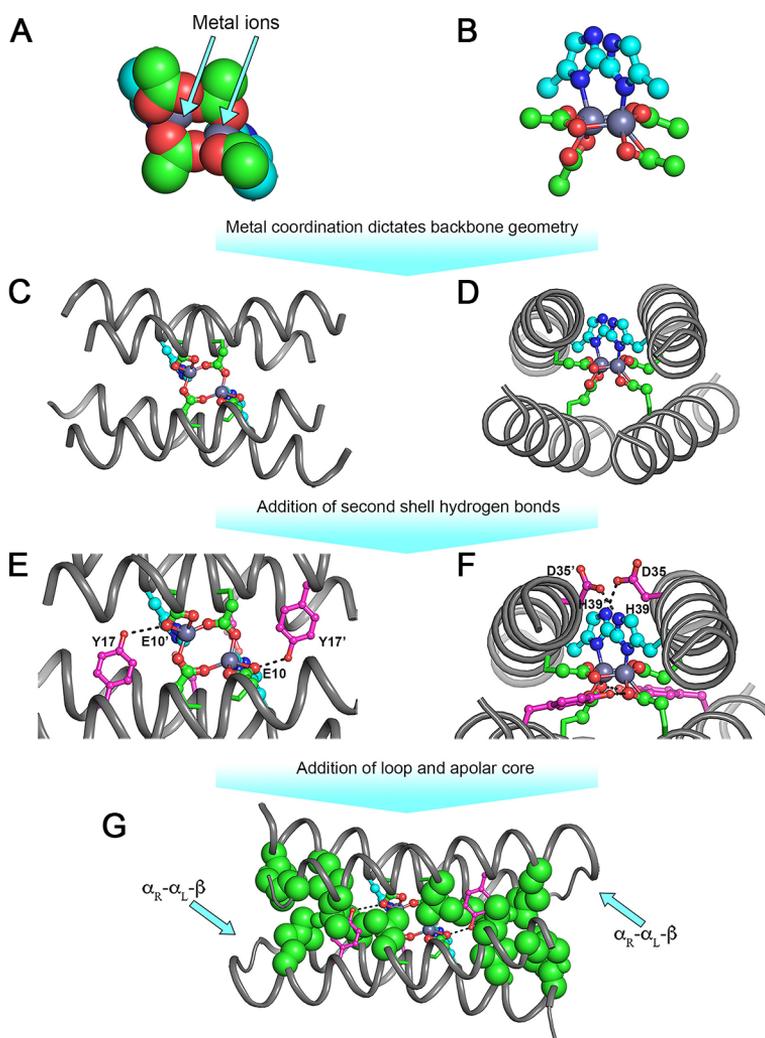


Figure 1.12. Crucial steps in DF1 de novo design: **A.** and **B.** first coordination sphere of the diiron-binding site; **C.** and **D.** parametrization of the backbone able to host the metal site; **E.** and **F.** second shell H-bond interactions; **G.** inclusion of the hydrophobic core and interhelical loop. Adapted with permission from reference ¹¹⁴. Copyright © 2019 American Chemical Society.

Subsequently, well-packed hydrophobic side chains were chosen at the remaining core positions to stabilize the folded state of the protein. In the final steps, an interhelical loop and the remaining surface-facing side chains were introduced, leading to DF1, the first member of DFs (Figure 1.12.G). For the first

time, a protein, completely designed from scratch, showed an experimental structure in perfect agreement with the intended design. Di-Zn²⁺-DF1 crystal structure deviates only 1.6 Å from the designed model, and only 1.0 Å if we consider the closest helical turns to the metal site. Moreover, the resolution of the NMR structure of the apo form showed only slight differences from the holo-form, with the main stabilizing interactions for the binding residues being preserved even in the absence of the metal ion.¹¹⁸ This confirmed that designing a sufficient set of hydrogen bonds and salt bridges was possible to achieve a pre-organized binding site for the metal cofactor in the apolar cavity of the protein. Though DF1 and its more soluble analogues DF2 and DF2t were able to accommodate different metal ions,¹¹⁹ no one of them was able to bind a reducing substrate.

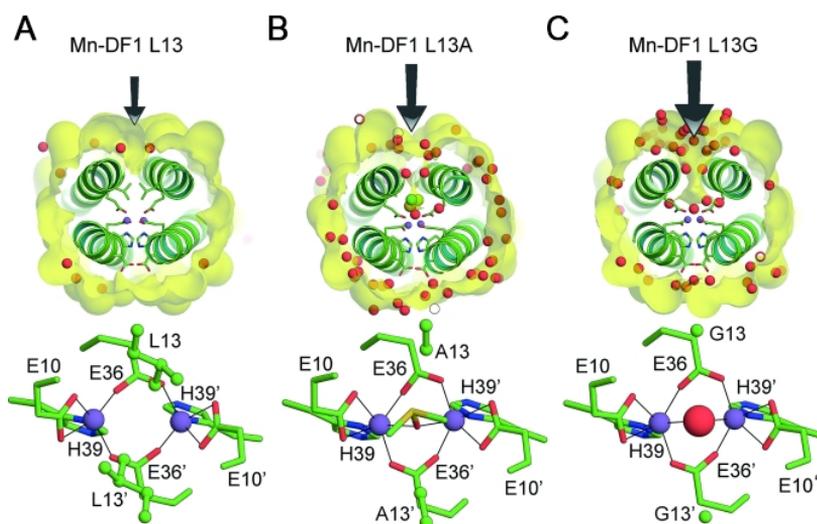


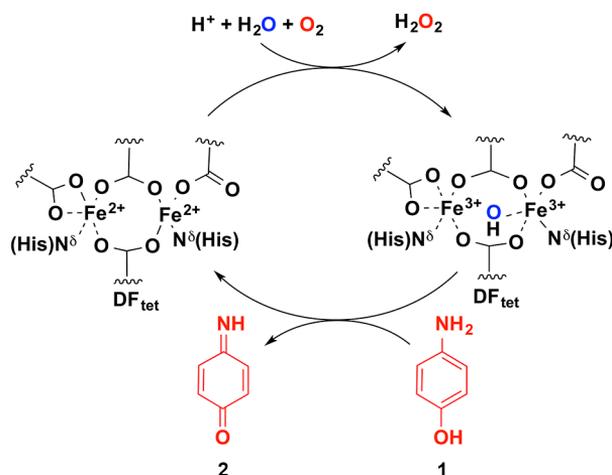
Figure 1.13. Effect of DF1 residue 13 on the metal site accessibility in **A.** di-Mn²⁺-DF1, **B.** di-Mn²⁺-DF1-L13A and **C.** di-Mn²⁺-DF1-L13G. The access channel, indicated by black arrows, gradually expands from leucine to glycine. The red spheres indicate the water molecules and the yellow spheres (in **B.**) indicate the DMSO molecule. The different residues in position 13 are depicted as balls and sticks. Adapted with permission from reference ¹¹⁵. Copyright © 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Chapter 1

Crystal structures of DF1 mutants at position lining at the dimetal site (positions 9 and 13) delineated a way to carve out a binding pocket: the substitution of Leu13 to Ala resulted in the binding of a dimethyl sulfoxide (DMSO) molecule at the metal center,¹²⁰ while further bulk-decreasing mutation to Gly led to a large increase in hydration and a consequent expansion of the interhelix distance (Figure 1.13).¹²¹

However, more room was needed to effectively bind more complex substrates. A combinatorial work on the simplified α -tetrameric scaffold, called DF_{tet}, was afforded to find the best combination of residues leading to iron-catalyzed oxidation of 4-aminophenol (4AP, **1** in Scheme 1.1) to 4-benzoquinone-monoimine (4BQM, **2** in Scheme 1.1), adopted as target reaction.

Scheme 1.1 Proposed Catalytic Cycle for the 4-Aminophenol Oxidation by DF_{tet}.



An unprecedented “negative design” protocol was adopted to prevent the formation of undesired parallel topologies. Firstly, slightly longer 33-long helices were adopted to extend the hydrophobic core, secondly, explicit evaluation of the electrostatic interactions in both parallel and anti-parallel topologies led to

the best combination of charged residues at the solvent-exposed interfaces.¹²²

A series of asymmetrical mutations were introduced at positions 15 and 19 (corresponding to the positions 9 and 13, respectively, of DF1) in the two helices responsible for the substrate accessibility. All the tetramers were found thermodynamically stable, and allowed for the catalytic screening. The fastest analogue (~1000-fold faster than the background oxidation) was the one bearing four glycine residues, corresponding to the wider access to the active site.¹²³

Lessons learned from the DFtet subset were applied into the well-characterized helix-loop-helix scaffold and into a single-chain protein.

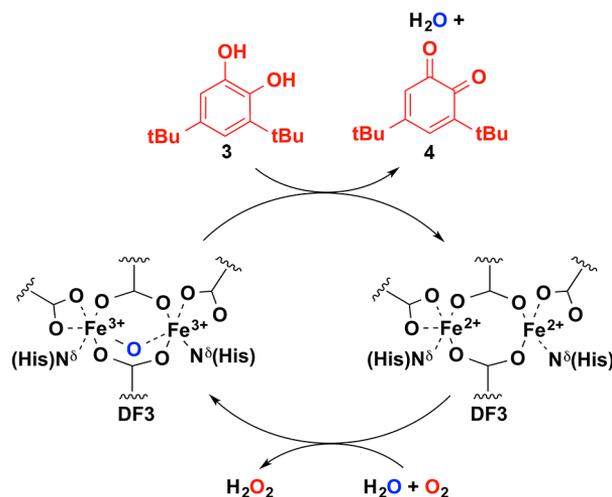
The four Gly residues were placed at the position 9 and 13 in DF1 to gain active site accessibility. Since only one Gly destabilizes DF1 by 10.8 kcal/mol,¹¹⁸ the interhelical loop was carefully redesigned in order to overcome the instability induced by the introduction of destabilizing residues. The original Val24-Lys25-Leu26 of DF1 was changed to Thr24-His25-Asn26, which adopts an α_R - α_L - β conformation.¹²⁴

The newly designed DF3 exhibited enhanced solubility (up to 3 mM in water) and active site accessibility, while retaining the intended global fold in the holo-form as confirmed by the NMR solution structural characterization.^{125,126}

Moreover, DF3 was the first de novo designed metalloenzyme, for which a high-resolution structure was determined.¹²⁵ Notably, its catalytic efficiency in the 3,5-di-tert-butyl-cathechol (DTBC, **3** in Scheme 1.2) oxidation to the corresponding o-quinone (DTBQ, **4** in Scheme 1.1) approaches that observed for natural enzymes ($k_{cat}/K_m = 6315 \text{ M}^{-1}\text{min}^{-1}$). DF3 is able to oxidize other aromatic substrates like 4AP and p-phenylene diamine, showing that its active site is able to host different moieties, though surprisingly o-phenylene diamine could not be converted.

Chapter 1

Scheme 1.2 Proposed Catalytic Cycle for Catechol Oxidation by DF3.



The DF_{tet} helices were used as template to design a 114-residue four-helix bundle, DF_{sc}.¹²⁷ The first analogue presented four alanine residues at the entrance of the metal-binding site, but it was not able to convert any substrate. On the other hand, the four-glycine analogue of DF_{sc}, 4G-DF_{sc}, was able to oxidize 4AP to 4BQM. Detailed spectroscopic studies helped in clarifying the role of the wider access channel for this class of enzymes.^{128,129} In particular, the binding of the phenol substrate to one of the two iron ions triggers the observed oxidase activity, which is afforded by reducing dioxygen to hydrogen peroxide on the other iron by the formation of an end-on hydroperoxo ferric intermediate. This finding supports the idea that the wider access, granted by the four glycine residues, let both ions be six-coordinated, and water protonates the hydroperoxo intermediate.

The following advancement in DF development was to emulate natural metalloenzymes in terms of their selectivity towards a certain reaction pathway

and product. Therefore, a highly specific and asymmetric environment was needed to perfectly orientate the substrate, and stabilize reaction intermediates.

A very general approach of covalent heterodimer design through copper-catalyzed alkyne-azide click chemistry was used to build a stable heterodimer.¹³⁰ DF-Click1 (DF-C1) was designed to tightly bury the reactive 4BQM, and induce its dimerization in a highly specific manner.¹³¹ The design consisted of the mutation of the active site access residues (4 Gly, 2 Gly 1 Phe, 1 Thr) in order to build an asymmetric binding site for the substrate. DF-C1 was demonstrated to perform 4-electron reduction of dioxygen in two successive two-electron steps by sequestering any reactive intermediate inside the protein. Whilst DF3 produces a plethora of products during 4AP oxidation, due to unspecific oxidative side-reactions and hydrogen peroxide release, DF-C1 leads to the formation of the Brandowsky's dye as the sole product, without any detectable release in solution of the 4BQM intermediate or H₂O₂. This result filled a gap with natural proteins approaching their unrivaled specificity and control.

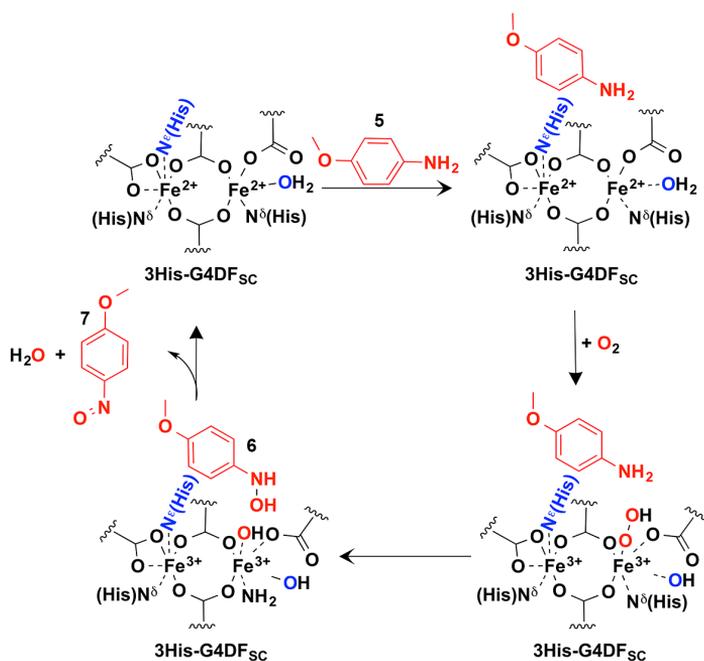
Finally, the direct mutation of the first coordination sphere of G4DFsc to His/4-carboxylate ligand set allowed the arylamines hydroxylation, as observed in the natural p-Aminobenzoate N-oxygenase (AurF).¹³² Site-specific mutations were required to insert the additional histidine residue within the hydrophobic core. Isoleucine 100 was mutated in the coordination histidine, while multiple substitutions (Y18L, I37N, and L81H) prevented steric clashes and favored stabilizing H-bond interactions. The resulting protein, 3His-G4DFsc, has very weak ferroxidase and oxidase activity, but catalyzes multiple rounds of conversion of p-anisidine (**5** in Scheme 1.3) to p-hydroxylamino-anisole (**7** in Scheme 1.1). The diferrous site in 3His-G4DFsc is very weakly antiferromagnetic coupled and this explains its slow ferroxidase activity. However, the binding of

Chapter 1

p-anisidine and interactions with the protein scaffold increases the antiferromagnetic coupling and allows the formation of the oxidative intermediate. Finally, the electrophilic attack on the amine (**6** in Scheme 1.3) and the further oxidation leads to hydroxylamino-anisole and the diiron resting state.^{128,129}

Moreover, DF scaffold is able to stabilize high reactive species in aqueous solution. In particular, the zinc-derivative of 3His-A2DFsc was demonstrated to bind the semiquinonic form of DTBC. Spectroelectrochemical redox titrations highlighted the stabilizing role of the protein matrix, reducing the one-electron oxidation potential of the catechol by approximately 400 mV (9 kcal/mol).¹³³

Scheme 1.3 Proposed Catalytic Cycle for the p-Anisidine Oxygenation by 3His-G4DFsc.



1.4.2. Design of small-molecule binding proteins

The de novo design of small-molecule binding proteins is very closely related to metalloproteins. Several examples of the former class involve the binding of metalloporphyrins.¹³⁴ Porphyrin-binding proteins are widely spread in Nature, thanks to the large range of their functions including oxygen transport, electron transfer/transport and catalysis.¹³⁵

The first example of de novo designed heme-binding protein was reported by Choma et al.¹³⁶ This was the first example of a designed protein able to tightly bind a cofactor. The protein scaffold was closely related to a backbone parametrized tetrameric four-helix bundle. Keystone residues were positioned for the first- and second-shell ligation as well as steric packing. This approach was subsequently expanded to enable design a variety of proteins able to bind porphyrin containing different metals.^{137–139}

DeGrado and Dutton applied a sequence-based approach to develop four-helix bundles able to perform mono- and multiheme binding, denominated 'maquettes'. A 31-residue peptide was designed to mimic the structural features of cytochrome bc1 and self-assembled in a four-helix bundle in presence of four hemins.¹⁴⁰ The structural stability of the maquette scaffold enabled the selection of the cofactors and ligands, their positioning inside the protein, and redox potential modulation, supporting diverse protein functionalities ranging from light capture to electron transfer, from reversible dioxygen binding to catalysis.^{141–144} Starting from a dioxygen-binding maquette, Anderson and coworkers engineered peroxidase activity. Only three amino acid substitutions were required to decrease the conformational flexibility and leaving only one heme binding site.¹⁴⁵ The resulting C45 promoted the hydrogen peroxide-dependent oxidation of diverse substrates with catalytic efficiencies

Chapter 1

approaching those of horseradish peroxidase (HRP). Moreover, this artificial heme protein was also reported to be active as carbene transferase, showing an exceptional diverse functionality.¹⁴⁶

Although the plethora of de novo designed porphyrin-binding proteins, structural data were not available preventing validation of these designs. The successful design of a porphyrin-binding protein with sub-Ångstrom accuracy has been reported only recently.

The step forward was made considering the hydrophobic core and the ligand-binding as inseparable units, in contrast to as separate sectors. DeGrado and coworkers performed a flexible backbone design on a parametrically defined protein template, allowing simultaneously the packing of the hydrophobic core in the region both proximal to and remote from the porphyrin-binding site.¹⁴⁷ In particular, the resulting protein PS1 (Porphyrin-binding Sequence) was designed to bind the abiotic Zn-meso-(trifluoromethyl)porphin (ZnP) and the biophysical characterization proved the high affinity of the protein scaffold to the small molecule. Moreover, the solution NMR structure of PS1 was resolved both in the apo- and in the holo-form (Figure 1.14A). The packing of the hydrophobic core in the distal region was nearly identical in the two forms. On the other side, the apo-PS1 was found to be more flexible and ready to bind ZnP in the desired geometry.

The same group used these principles to design a non-natural manganese diphenylporphyrin-binding protein (MPP, Figure 1.14B).¹⁴⁸ Not only the porphyrin, but also a hydroperoxide molecule bound to the metal was considered during the design. The designed metalloprotein was able to tightly bind the manganese porphyrin and to access a Mn(V)-oxo species and perform O atom transfer to thioanisole. Moreover, the x-ray crystal structure showed the

presence of two interactive water molecules in the exact region for the designed hydroperoxo unit. This demonstrated the ability not only to design small-molecule binding protein with high accuracy, but also to include at the same time accessible channel for the substrate to develop functional enzyme.

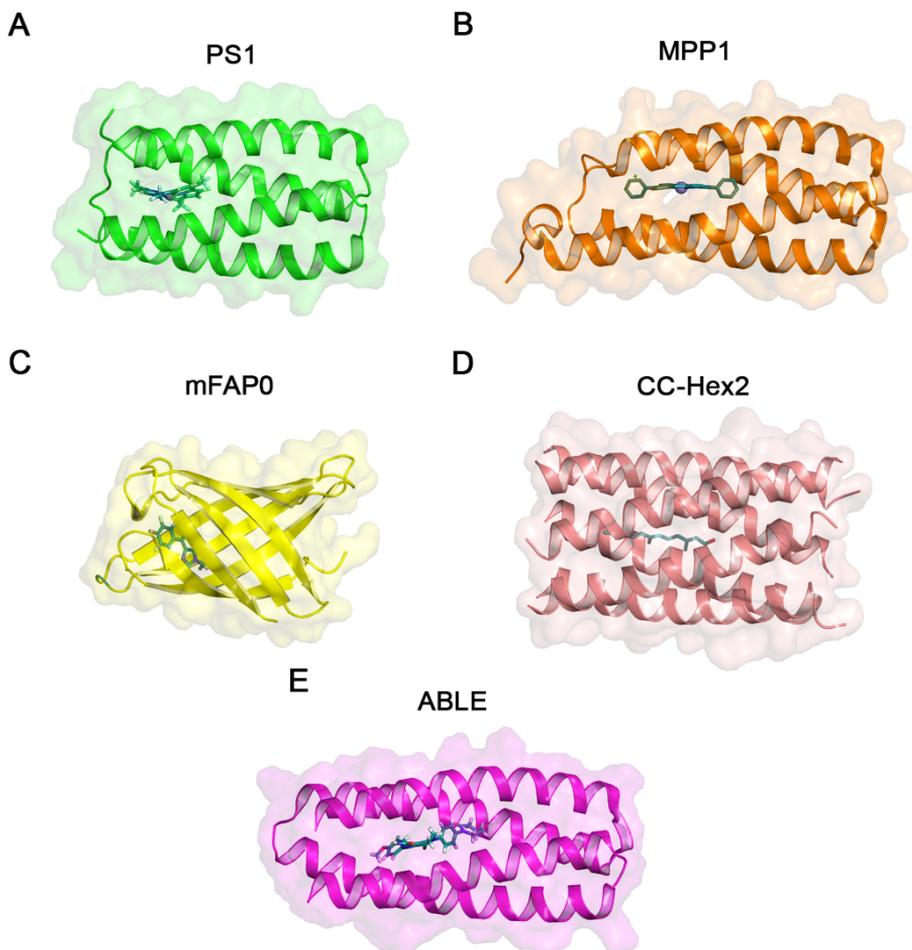


Figure 1.14. De novo designed small molecule binding protein. Resolved structures of: **A.** PS1 (PDB ID: 5TGY); **B.** MMP1 (PDB ID: 7JRQ); **C.** mFAP0 (PDB ID: 6CZH); **D.** CC-Hex2 (PDB ID: 6EIZ); **E.** ABLE (PDB ID: 6W70). Protein backbones are shown as cartoon, surfaces in transparency, small molecules as sticks and metal ions as spheres.

Chapter 1

Parametrized backbone design was used also for β barrels. Baker and coworkers disentangled the backbone requirements to develop the first de novo water-soluble β -barrel protein.¹⁴⁹ Moreover, once determined the principles to generate a family of backbones, they engineered the binding site of the fluorogenic compound DFHBI ((5Z)-5-[(3,5-difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one). They used an innovative 'rotamer interaction field' docking method to sample at the same time the rigid-body degrees of freedom of the ligand, and the sequence identities of the surrounding amino acids. The micromolar affinity for the substrate of the protein (mFAPO, Figure 1.14C), designed with this protocol, was further increase with a combinatorial yeast cell display and led to *in vivo* fluorescence activation.

Woolfson and coworkers, instead, have expanded the parametrization of coiled-coils to design helical barrels with a high degree of oligomerization (from five to seven helices). They designed de novo the first pentameric α -helical coiled and the first parallel, blunt-ended heptamer.¹⁵⁰ The x-ray structure highlighted the presence of accessible and chemically defined channels. As a matter of fact, the de novo helical barrels were demonstrated to selectively bind lipophilic and charged small molecules, as reported for farnesol in CC-Hex2 (Figure 1.14D).¹⁵¹ The different affinities and binding modes were rationalized considering the internal shapes of the channel. The channels are characterized by the alternation of widened voids with narrower sections. The length and presence of bulky groups can prevent a favorable and stable binding.

The example reported until now were involving rigid, flat, hydrophobic dyes or lipidic small molecules. Recently, Polizzi and DeGrado have managed in the challenging design of the de novo proteins binding polar flexible molecules.¹⁵²

They designed proteins, called ABLE, for the binding with submicromolar affinity the factor Xa inhibitor apixaban, characterized by five rotatable bonds and eight heteroatoms. To this endeavor, the authors have developed a new protein structure unit analogous to rotamers, the van der Mer (vdM). The vdMs sample locations of a defined chemical group that have been experimentally demonstrated to favorably interact with the sidechain of the considered backbone. Therefore, the vdMs are backbone dependent as the rotamers are. However, the degrees of freedom to sample between the protein and target are greatly reduced.

The design strategy involved: *i)* the definition of the target chemical groups in the small molecule; *ii)* the creation of an ensemble of parametrized scaffolds; *iii)* the identification of the vdM positions that collectively engage each of the targeted chemical groups of the small molecule; *iv)* the simultaneous packing of the residues within the binding site and in the distant protein core.

The x-ray structure of the apo-form showed an open, preorganized binding pocket. In particular, the unoccupied binding site was solvated by nine ordered water molecules and a buffer acetate molecule. The binding of the apixaban is favored by the displacement of the ordered solvent molecule and releasing the local frustration (Figure 1.14E).

These results suggest the intrinsic relationship between the three-dimensional protein structure and the target small molecule.

Chapter 1

1.5 Aim of the thesis

The development of DFs can be considered as Rational Evolution, involving three main steps: *i)* the first member DF1 was a single gene homodimer; *ii)* the duplication of this gene and the fusion led to a single construct; *iii)* the site-specific mutations optimized the sequence for the function, up to the substrate-gated four-electron reduction of O₂ to water.

This Ph.D. project was aimed at increasing the evolutionary complexity of DFs, introducing an allosteric regulation on the oxidase activity of the diiron center. To this endeavor, a novel MDA strategy was developed to couple the DF domain with a small molecule binding domain. In particular, PS1 protein was chosen as the second domain, considering the differences in flexibility of the small molecule binding site, before and after the binding of the abiotic zinc-porphyrin. Such difference was hypothesized to be inducible to the reactivity of the catalytic domain and enable the design of allosterically modulated multi-domain protein completely from scratch.¹⁵³ A further objective was to explore the binding specificity towards the abiotic zinc porphyrin and the effect on the functional properties of the system upon substitution with a different metalloporphyrin. Finally, as a first step for the development of de novo designed light-harvesting systems, the photosensitizing properties of the zinc porphyrins were investigated, both bound to PS1 domain and incorporated in micelles, to understand the role of the protein matrix.

References

1. Searls, D. B. Reading the book of life. *Bioinformatics* **17**, 579–580 (2001).
2. Searls, D. B. The Linguistics of DNA. *Am. Sci.* **80**, 579–591 (1992).
3. Searls, D. B. Linguistic approaches to biological sequences. *Bioinformatics* **13**, 333–344 (1997).
4. Scaiewicz, A. & Levitt, M. The language of the protein universe. *Curr. Opin. Genet. Dev.* **35**, 50–56 (2015).
5. Goldrick, M., Ferreira, V. S., Miozzo, M. & Blevins, J. P. The Morphology of Words. in *The Oxford Handbook of Language Production* (eds. Goldrick, M., Ferreira, V. S. & Miozzo, M.) (Oxford University Press, 2014). doi:10.1093/oxfordhb/9780199735471.013.018.
6. Yegambaram, K., Bulloch, E. M. M. & Kingston, R. L. Protein domain definition should allow for conditional disorder: Protein Conformational Stability and Domain Definition. *Protein Sci.* **22**, 1502–1518 (2013).
7. Gimona, M. Protein linguistics — a grammar for modular protein assembly? *Nat. Rev. Mol. Cell Biol.* **7**, 68–73 (2006).
8. Searls, D. B. The language of genes. *Nature* **420**, 211–217 (2002).
9. Hsu, C.-H., Chen, C.-K. & Hwang, M.-J. The architectural design of networks of protein domain architectures. *Biol. Lett.* **9**, 20130268 (2013).
10. Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. Domain Insertions in Protein Structures. *J. Mol. Biol.* **338**, 633–641 (2004).
11. Berrondo, M., Ostermeier, M. & Gray, J. J. Structure Prediction of Domain Insertion Proteins from Structures of Individual Domains. *Structure* **16**, 513–527 (2008).
12. Scaiewicz, A. & Levitt, M. Unique function words characterize genomic proteins. *Proc. Natl. Acad. Sci.* **115**, 6703–6708 (2018).
13. Basu, M. K., Carmel, L., Rogozin, I. B. & Koonin, E. V. Evolution of protein domain promiscuity in eukaryotes. *Genome Res.* **18**, 449–461 (2008).

Chapter 1

14. Han, J.-H., Batey, S., Nickson, A. A., Teichmann, S. A. & Clarke, J. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* **8**, 319–330 (2007).
15. Yu, L. *et al.* Grammar of protein domain architectures. *Proc. Natl. Acad. Sci.* **116**, 3636–3645 (2019).
16. Qian, J., Luscombe, N. M. & Gerstein, M. Protein family and fold occurrence in genomes: power-law behaviour and evolutionary model. *J. Mol. Biol.* **313**, 673–681 (2001).
17. Jones, C. I. Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality. *J. Econ. Perspect.* **29**, 29–46 (2015).
18. Chothia, C. Evolution of the Protein Repertoire. *Science* **300**, 1701–1703 (2003).
19. Karev, G. P., Wolf, Y. I., Rzhetsky, A. Y., Berezovskaya, F. S. & Koonin, E. V. Birth and death of protein domains: A simple model of evolution explains power law behavior. *BMC Evol. Biol.* **2**, 18 (2002).
20. Koonin, E. V., Wolf, Y. I. & Karev, G. P. The structure of the protein universe and genome evolution. *Nature* **420**, 218–223 (2002).
21. Apic, G., Gough, J. & Teichmann, S. A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310**, 311–325 (2001).
22. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* **7**, 1765–1771 (2008).
23. Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat. Rev. Mol. Cell Biol.* **7**, 739–750 (2006).
24. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).
25. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, (2018).
26. Feschotte, C. & Pritham, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007).

27. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. Ty elements transpose through an RNA intermediate. *Cell* **40**, 491–500 (1985).
28. Crow, K. D. What Is the Role of Genome Duplication in the Evolution of Complexity and Diversity? *Mol. Biol. Evol.* **23**, 887–892 (2006).
29. Gilbert, W. The Exon Theory of Genes. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 901–905 (1987).
30. Kolkman, J. A. & Stemmer, W. P. C. Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* **19**, 423–428 (2001).
31. Long, M., de Souza, S. J., Rosenberg, C. & Gilbert, W. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc. Natl. Acad. Sci.* **93**, 7727–7731 (1996).
32. Vogel, C., Teichmann, S. A. & Pereira-Leal, J. The Relationship Between Domain Duplication and Recombination. *J. Mol. Biol.* **346**, 355–365 (2005).
33. Apic, G., Huber, W. & Teichmann, S. A. Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination. *J. Struct. Funct. Genomics* **4**, 67–78 (2003).
34. Doolittle, R. F. & Bork, P. Evolutionarily Mobile Modules in Proteins. *Sci. Am.* **269**, 50–56 (1993).
35. Bornberg-Bauer, E., Beaussart, F., Kummerfeld, S. K., Teichmann, S. A. & Weiner, J. The evolution of domain arrangements in proteins and interaction networks. *Cell. Mol. Life Sci. CMLS* **62**, 435–445 (2005).
36. Vishwanath, S., de Brevern, A. G. & Srinivasan, N. Same but not alike: Structure, flexibility and energetics of domains in multi-domain proteins are influenced by the presence of other domains. *PLOS Comput. Biol.* **14**, e1006008 (2018).
37. Jones, S., Marin, A. & Thornton, J. Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng. Des. Sel.* **13**, 77–82 (2000).
38. Changeux, J.-P. Allosteric Mechanisms of Signal Transduction. *Science* **308**, 1424–1428 (2005).

Chapter 1

39. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
40. Anantharaman, V., Koonin, E. V. & Aravind, L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains¹¹Edited by F. Cohen. *J. Mol. Biol.* **307**, 1271–1292 (2001).
41. Dueber, J. E., Yeh, B. J., Bhattacharyya, R. P. & Lim, W. A. Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Curr. Opin. Struct. Biol.* **14**, 690–699 (2004).
42. Koide, S. Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules. *Curr. Opin. Biotechnol.* **20**, 398–404 (2009).
43. Bashton, M. & Chothia, C. The Generation of New Protein Functions by the Combination of Domains. *Structure* **15**, 85–99 (2007).
44. Yang, H., Liu, L. & Xu, F. The promises and challenges of fusion constructs in protein biochemistry and enzymology. *Appl. Microbiol. Biotechnol.* **100**, 8273–8281 (2016).
45. Matthaei, J. F. *et al.* Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations. *Nano Lett.* **15**, 5235–5239 (2015).
46. Stein, V. & Alexandrov, K. Synthetic protein switches: design principles and applications. *Trends Biotechnol.* **33**, 101–110 (2015).
47. Makhlynets, O. V., Raymond, E. A. & Korendovych, I. V. Design of Allosterically Regulated Protein Catalysts. *Biochemistry* **54**, 1444–1456 (2015).
48. George, R. A. & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng. Des. Sel.* **15**, 871–879 (2002).
49. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
50. Klein, J. S., Jiang, S., Galimidi, R. P., Keeffe, J. R. & Bjorkman, P. J. Design and characterization of structured protein linkers with differing flexibilities. *Protein Eng. Des. Sel.* **27**, 325–330 (2014).

51. Cross, P. J., Allison, T. M., Dobson, R. C. J., Jameson, G. B. & Parker, E. J. Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. *Proc. Natl. Acad. Sci.* **110**, 2111–2116 (2013).
52. Kwon, N.-Y., Kim, Y. & Lee, J.-O. The application of helix fusion methods in structural biology. *Curr. Opin. Struct. Biol.* **60**, 110–116 (2020).
53. Jacobs, T. M. *et al.* Design of structurally distinct proteins using strategies inspired by evolution. *Science* **352**, 687–690 (2016).
54. Youn, S.-J. *et al.* Construction of novel repeat proteins with rigid and predictable structures using a shared helix method. *Sci. Rep.* **7**, (2017).
55. Zhou, Z., Feng, H., Zhou, H., Zhou, Y. & Bai, Y. Design and Folding of a Multidomain Protein. *Biochemistry* **44**, 12107–12112 (2005).
56. Strickland, D., Moffat, K. & Sosnick, T. R. Light-activated DNA binding in a designed allosteric protein. *Proc. Natl. Acad. Sci.* **105**, 10709–10714 (2008).
57. Thornton, J. M. & Sibanda, B. L. Amino and carboxy-terminal regions in globular proteins. *J. Mol. Biol.* **167**, 443–460 (1983).
58. Kanwar, M., Wright, R. C., Date, A., Tullman, J. & Ostermeier, M. Protein Switch Engineering by Domain Insertion. in *Methods in Enzymology* vol. 523 369–388 (Elsevier, 2013).
59. Guntas, G. & Ostermeier, M. Creation of an Allosteric Enzyme by Domain Insertion. *J. Mol. Biol.* **336**, 263–273 (2004).
60. Ostermeier, M. Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel.* **18**, 359–364 (2005).
61. Gorman, S. D., D’Amico, R. N., Winston, D. S. & Boehr, D. D. Engineering Allostery into Proteins. in *Protein Allostery in Drug Discovery* (eds. Zhang, J. & Nussinov, R.) vol. 1163 359–384 (Springer Singapore, 2019).
62. Ostermeier, M. Designing switchable enzymes. *Curr. Opin. Struct. Biol.* **19**, 442–448 (2009).
63. Guntas, G., Mitchell, S. F. & Ostermeier, M. A Molecular Switch Created by In Vitro Recombination of Nonhomologous Genes. *Chem. Biol.* **11**, 1483–1487 (2004).

Chapter 1

64. Guntas, G., Mansell, T. J., Kim, J. R. & Ostermeier, M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci.* **102**, 11224–11229 (2005).
65. Liang, J., Kim, J. R., Boock, J. T., Mansell, T. J. & Ostermeier, M. Ligand binding and allostery can emerge simultaneously. *Protein Sci.* **16**, 929–937 (2007).
66. Bliven, S. & Prlić, A. Circular Permutation in Proteins. *PLoS Comput. Biol.* **8**, e1002445 (2012).
67. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci.* **96**, 11241–11246 (1999).
68. Yu, Y. & Lutz, S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.* **29**, 18–25 (2011).
69. Pérez Koldenkova, V. & Nagai, T. Genetically encoded Ca²⁺ indicators: Properties and evaluation. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1833**, 1787–1797 (2013).
70. Brunette, T. *et al.* Modular repeat protein sculpting using rigid helical junctions. *Proc. Natl. Acad. Sci.* **117**, 8870–8875 (2020).
71. Kolodny, R., Pereyaslavets, L., Samson, A. O. & Levitt, M. On the Universe of Protein Folds. *Annu. Rev. Biophys.* **42**, 559–582 (2013).
72. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230 (1973).
73. Korendovych, I. V. & DeGrado, W. F. *De novo* protein design, a retrospective. *Q. Rev. Biophys.* **53**, (2020).
74. Regan, L. & DeGrado, W. Characterization of a helical protein designed from first principles. *Science* **241**, 976–978 (1988).
75. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of *de novo* protein design. *Nature* **537**, 320–327 (2016).
76. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).

77. Lu, Y., Yeung, N., Sieracki, N. & Marshall, N. M. Design of functional metalloproteins. *Nature* **460**, 855–862 (2009).
78. Nastri, F. *et al.* Engineering Metalloprotein Functions in Designed and Native Scaffolds. *Trends Biochem. Sci.* (2019) doi:10.1016/j.tibs.2019.06.006.
79. Liu, J. *et al.* Metalloproteins Containing Cytochrome, Iron–Sulfur, or Copper Redox Centers. *Chem. Rev.* **114**, 4366–4469 (2014).
80. Nastri, F. *et al.* Design and engineering of artificial oxygen-activating metalloenzymes. *Chem. Soc. Rev.* **45**, 5020–5054 (2016).
81. Gao, M. & Skolnick, J. A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. *PLoS Comput. Biol.* **9**, e1003302 (2013).
82. Mortenson, D. E. *et al.* “Inverse Drug Discovery” Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates. *J. Am. Chem. Soc.* **140**, 200–210 (2018).
83. McFedries, A., Schwaid, A. & Saghatelian, A. Methods for the Elucidation of Protein-Small Molecule Interactions. *Chem. Biol.* **20**, 667–673 (2013).
84. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
85. Barker, P. D. Designing redox metalloproteins from bottom-up and top-down perspectives. *Curr. Opin. Struct. Biol.* **13**, 490–499 (2003).
86. Shiga, D. *et al.* The effect of the side chain length of Asp and Glu on coordination structure of Cu(2+) in a de novo designed protein. *Biopolymers* **91**, 907–916 (2009).
87. Shiga, D. *et al.* Creation of a Type 1 Blue Copper Site within a de Novo Coiled-Coil Protein Scaffold. *J. Am. Chem. Soc.* **132**, 18191–18198 (2010).
88. Warren, J. J., Lancaster, K. M., Richards, J. H. & Gray, H. B. Inner- and Outer-Sphere Metal Coordination in Blue Copper Proteins. *J. Inorg. Biochem.* **115**, 119–126 (2012).
89. Shiga, D. *et al.* Creation of a binuclear purple copper site within a de novo coiled-coil protein. *Biochemistry* **51**, 7901–7907 (2012).

Chapter 1

90. Tsukihara, T. *et al.* Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 Å. *Science* **269**, 1069–1074 (1995).
91. Pinter, T. B. J., Koebke, K. J. & Pecoraro, V. L. Catalysis and Electron Transfer in De Novo Designed Helical Scaffolds. *Angew. Chem. Int. Ed.* **59**, 7678–7699 (2020).
92. Koebke, K. J. & Pecoraro, V. L. Development of de Novo Copper Nitrite Reductases: Where We Are and Where We Need To Go. *ACS Catal.* **8**, 8046–8057 (2018).
93. Plegaria, J. S. *et al.* De Novo Design and Characterization of Copper Metallopeptides Inspired by Native Cupredoxins. *Inorg. Chem.* **54**, 9470–9482 (2015).
94. Koebke, K. J. *et al.* Clarifying the Copper Coordination Environment in a de Novo Designed Red Copper Protein. *Inorg. Chem.* **57**, 12291–12302 (2018).
95. Yu, F., Penner-Hahn, J. E. & Pecoraro, V. L. De Novo-Designed Metallopeptides with Type 2 Copper Centers: Modulation of Reduction Potentials and Nitrite Reductase Activities. *J. Am. Chem. Soc.* **135**, 18096–18107 (2013).
96. Koebke, K. J. *et al.* Modifying the Steric Properties in the Second Coordination Sphere of Designed Peptides Leads to Enhancement of Nitrite Reductase Activity. *Angew. Chem. Int. Ed.* **57**, 3954–3957 (2018).
97. Boyle, A. L. *et al.* Selective coordination of three transition metal ions within a coiled-coil peptide scaffold. *Chem. Sci.* **10**, 7456–7465 (2019).
98. Tebo, A. G. & Pecoraro, V. L. Artificial metalloenzymes derived from three-helix bundles. *Curr. Opin. Chem. Biol.* **25**, 65–70 (2015).
99. Berwick, M. R. *et al.* De Novo Design of Ln(III) Coiled Coils for Imaging Applications. *J. Am. Chem. Soc.* **136**, 1166–1169 (2014).
100. Berwick, M. R. *et al.* Location dependent coordination chemistry and MRI relaxivity, in de novo designed lanthanide coiled coils. *Chem Sci* **7**, 2207–2216 (2016).
101. N. Slope, L. *et al.* Tuning coordination chemistry through the second sphere in designed metallocoiled coils. *Chem. Commun.* **56**, 3729–3732 (2020).

102. Zastrow, M. L. & Pecoraro, V. L. Designing Hydrolytic Zinc Metalloenzymes. *Biochemistry* **53**, 957–978 (2014).
103. Cangelosi, V. M., Deb, A., Penner-Hahn, J. E. & Pecoraro, V. L. A De Novo Designed Metalloenzyme for the Hydration of CO₂. *Angew. Chem. Int. Ed.* **53**, 7900–7903 (2014).
104. Maret, W. New perspectives of zinc coordination environments in proteins. *J. Inorg. Biochem.* **111**, 110–116 (2012).
105. Zastrow, M. L., Peacock, A. F. A., Stuckey, J. A. & Pecoraro, V. L. Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **4**, 118–123 (2012).
106. Studer, S. *et al.* Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288 (2018).
107. Chino, M. *et al.* Oxidation catalysis by iron and manganese porphyrins within enzyme-like cages. *Biopolymers* **109**, e23107 (2018).
108. Dizicheh, Z. B., Halloran, N., Asma, W. & Ghirlanda, G. De Novo Design of Iron–Sulfur Proteins. in *Methods in Enzymology* vol. 595 33–53 (Elsevier, 2017).
109. Meyer, J. & Moulis, J.-M. Rubredoxin. in *Handbook of Metalloproteins* (eds. Messerschmidt, A. *et al.*) (John Wiley & Sons, Ltd, 2006). doi:10.1002/0470028637.met135.
110. Lombardi, A. *et al.* Miniaturized metalloproteins: Application to iron–sulfur proteins. *Proc. Natl. Acad. Sci.* **97**, 11922–11927 (2000).
111. Nanda, V. *et al.* De Novo Design of a Redox-Active Minimal Rubredoxin Mimic. *J. Am. Chem. Soc.* **127**, 5804–5805 (2005).
112. Jacques, A. *et al.* A cyclic peptide-based redox-active model of rubredoxin. *Chem. Commun.* **49**, 2915 (2013).
113. Tebo, A. G. *et al.* Development of a Rubredoxin-Type Center Embedded in a *de Novo* -Designed Three-Helix Bundle. *Biochemistry* **57**, 2308–2316 (2018).

Chapter 1

114. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).
115. Chino, M. *et al.* Artificial Diiron Enzymes with a De Novo Designed Four-Helix Bundle Structure. *Eur. J. Inorg. Chem.* **2015**, 3352–3352 (2015).
116. Kurtz, D. M. Structural similarity and functional diversity in diiron-oxo proteins. *JBIC J. Biol. Inorg. Chem.* **2**, 159–167 (1997).
117. Lombardi, A. *et al.* Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci.* **97**, 6298–6305 (2000).
118. Maglio, O., Nastri, F., Pavone, V., Lombardi, A. & DeGrado, W. F. Preorganization of molecular binding sites in designed diiron proteins. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3772–3777 (2003).
119. Maglio, O. *et al.* Diiron-containing metalloproteins: Developing functional models. *Comptes Rendus Chim.* **10**, 703–720 (2007).
120. Di Costanzo, L. *et al.* Toward the de Novo Design of a Catalytically Active Helix Bundle: A Substrate-Accessible Carboxylate-Bridged Dinuclear Metal Center. *J. Am. Chem. Soc.* **123**, 12749–12757 (2001).
121. DeGrado, W. F. *et al.* Sliding Helix and Change of Coordination Geometry in a Model Di-MnII Protein. *Angew. Chem. Int. Ed.* **42**, 417–420 (2003).
122. Summa, C. M., Rosenblatt, M. M., Hong, J.-K., Lear, J. D. & DeGrado, W. F. Computational de novo Design, and Characterization of an A2B2 Diiron Protein. *J. Mol. Biol.* **321**, 923–938 (2002).
123. Kaplan, J. & DeGrado, W. F. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11566–11570 (2004).
124. Lahr, S. J. *et al.* Analysis and Design of Turns in α -Helical Hairpins. *J. Mol. Biol.* **346**, 1441–1454 (2005).
125. Faiella, M. *et al.* An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **5**, 882–884 (2009).

126. Torres Martin de Rosales, R. *et al.* Spectroscopic and metal-binding properties of DF3: an artificial protein able to accommodate different metal ions. *JBIC J. Biol. Inorg. Chem.* **15**, 717–728 (2010).
127. Calhoun, J. R. *et al.* Computational Design and Characterization of a Monomeric Helical Dinuclear Metalloprotein. *J. Mol. Biol.* **334**, 1101–1115 (2003).
128. Snyder, R. A., Butch, S. E., Reig, A. J., DeGrado, W. F. & Solomon, E. I. Molecular-Level Insight into the Differential Oxidase and Oxygenase Reactivities of de Novo Due Ferri Proteins. *J. Am. Chem. Soc.* **137**, 9302–9314 (2015).
129. Snyder, R. A. *et al.* Systematic Perturbations of Binuclear Non-heme Iron Sites: Structure and Dioxygen Reactivity of de Novo Due Ferri Proteins. *Biochemistry* **54**, 4637–4651 (2015).
130. Chino, M., Leone, L., Maglio, O. & Lombardi, A. Designing Covalently Linked Heterodimeric Four-Helix Bundles. in *Methods in Enzymology* vol. 580 471–499 (Elsevier, 2016).
131. Chino, M. *et al.* A De Novo Heterodimeric Due Ferri Protein Minimizes the Release of Reactive Intermediates in Dioxygen-Dependent Oxidation. *Angew. Chem. Int. Ed.* **56**, 15580–15583 (2017).
132. Reig, A. J. *et al.* Altering the O₂-Dependent Reactivity of de novo Due Ferri Proteins. *Nat. Chem.* **4**, 900–906 (2012).
133. Ulas, G., Lemmin, T., Wu, Y., Gassner, G. T. & DeGrado, W. F. Designed metalloprotein stabilizes a semiquinone radical. *Nat. Chem.* **8**, 354–359 (2016).
134. Peacock, A. F. Incorporating metals into de novo proteins. *Curr. Opin. Chem. Biol.* **17**, 934–939 (2013).
135. Li, T., Bonkovsky, H. L. & Guo, J. Structural analysis of heme proteins: implications for design and prediction. *BMC Struct. Biol.* **11**, 13 (2011).
136. Choma, C. T. *et al.* Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **116**, 856–865 (1994).
137. Fry, H. C., Lehmann, A., Saven, J. G., DeGrado, W. F. & Therien, M. J. Computational design and elaboration of a de novo heterotetrameric alpha-

Chapter 1

- helical protein that selectively binds an emissive abiological (porphinato)zinc chromophore. *J. Am. Chem. Soc.* **132**, 3997–4005 (2010).
138. Fry, H. C. *et al.* Computational de Novo Design and Characterization of a Protein That Selectively Binds a Highly Hyperpolarizable Abiological Chromophore. *J. Am. Chem. Soc.* **135**, 13914–13926 (2013).
139. Korendovych, I. V. *et al.* De Novo Design and Molecular Assembly of a Transmembrane Diporphyrin-Binding Protein Complex. *J. Am. Chem. Soc.* **132**, 15516–15518 (2010).
140. Robertson, D. E. *et al.* Design and synthesis of multi-haem proteins. *Nature* **368**, 425–432 (1994).
141. Farid, T. A. *et al.* Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **9**, 826–833 (2013).
142. Koder, R. L. *et al.* Design and engineering of an O₂ transport protein. *Nature* **458**, 305–309 (2009).
143. Lichtenstein, B. R. *et al.* Engineering oxidoreductases: maquette proteins designed from scratch. *Biochem. Soc. Trans.* **40**, 561–566 (2012).
144. Grayson, K. J. & Anderson, J. R. The ascent of man(made oxidoreductases). *Curr. Opin. Struct. Biol.* **51**, 149–155 (2018).
145. Watkins, D. W. *et al.* Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. *Nat. Commun.* **8**, 358 (2017).
146. Stenner, R., Steventon, J. W., Seddon, A. & Anderson, J. L. R. A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase. *Proc. Natl. Acad. Sci.* **117**, 1419–1428 (2020).
147. Polizzi, N. F. *et al.* De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
148. Mann, S. I., Nayak, A., Gassner, G. T., Therien, M. J. & DeGrado, W. F. De Novo Design, Solution Characterization, and Crystallographic Structure of an Abiological Mn–Porphyrin-Binding Protein Capable of Stabilizing a Mn(V) Species. *J. Am. Chem. Soc.* (2020) doi:10.1021/jacs.0c10136.

149. Dou, J. *et al.* De novo design of a fluorescence-activating β -barrel. *Nature* **561**, 485 (2018).
150. Thomson, A. R. *et al.* Computational design of water-soluble α -helical barrels. *Science* **346**, 485–488 (2014).
151. Thomas, F. *et al.* De Novo -Designed α -Helical Barrels as Receptors for Small Molecules. *ACS Synth. Biol.* **7**, 1808–1816 (2018).
152. Polizzi, N. F. & DeGrado, W. F. A defined structural unit enables de novo design of small-molecule-binding proteins. *Science* **369**, 1227–1233 (2020).
153. Pirro, F. *et al.* Allosteric cooperation in a de novo-designed two-domain protein. *Proc. Natl. Acad. Sci.* **117**, 33246–33253 (2020).

Chapter 2: Results and Discussion

2.1. Development of a new structural-based strategy for designing multi-domain proteins

As outlined in the previous chapter, the DF and PS scaffolds have in common the same domain, a four-helix bundle.¹ However, they differ in the interhelical separation and registry of their domains (Figure 2.1). In particular, the main differences can be identified in the superhelical radius (R_0), the superhelical phase offset ($\Delta\phi_0$) and the axial offset ($\Delta Z_{aa'}$) from the Coiled Coil Crick Parametrization (CCCP) analysis (Table 2.1).² In order to design a multi-domain protein, combining DF and PS bundles, it was necessary to determine the best linkers to connect and accommodate the transition between them in an energetically reasonable manner. To accomplish this, the helical bundles were connected using a structure-based design strategy involving two inter-related steps: first, the optimal relative translation and orientation between the bundles, namely the bundle placement, were determined; second, a structural solution interconnecting each of the disjointed helices between the domains was identified.

Table 2.1. Four-helix bundles geometrical parameters of the starting proteins.^[a]

	RMDS [Å]	Coil radius R_0 [Å]	Coil frequency ω_0 [°/res]	Pitch angle α [°]	Coil phase offset Δf_0 [°]	Rise per residue [Å]	Axial offset, $\Delta Z_{aa'}$ [Å]
DF1 ^[b]	0.88	7.44	-3.07	-15.27	-98.42	1.51	2.61
PS1 ^[c]	1.87	7.98	-2.90	-15.87	-77.68	1.48	1.37

^[a] Coiled Crick parametrization, enforcing D_2 for a 15-residue bundle section around coordinating residues. ^[b]1JMB. ^[c]5TGY.

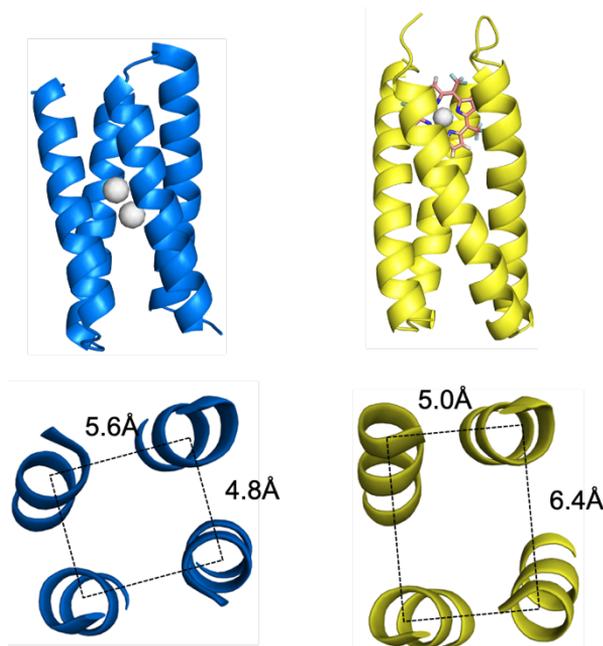


Figure 2.1. Differences between helical bundle geometries. DF1 (left) and PS1 (right) are both 4-helix bundle proteins. The bundle geometry at the planes of connection between DF1 and PS1 (bottom left and right respectively) are substantially different from one another.

Considering the two bundles as rigid bodies, their colocalization can be described by six coordinates, corresponding to the three translational and three rotational degrees of freedom. In many helical bundles, including those considered here,^{3,4} the helices supercoil about a central or bundle axis. If the two bundles share the same bundle axis (Figure 2.2), the bundle placement problem is simplified, and becomes a search in a two-dimensional space: the translational displacement along the bundle axis, ΔZ , and the rotational displacement about the bundle axis, $\Delta\Phi$. Each bundle axis is exclusively identified by its helical chains. Thus, rotation and translation matrices can be applied to the atomic coordinates of each structure to align the two axes and explore the $(\Delta Z, \Delta\Phi)$ space.

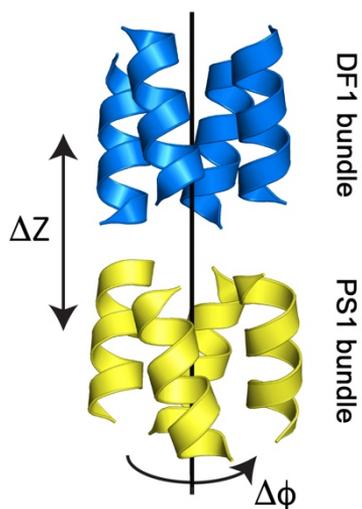


Figure 2.2. Definition of the two degrees of freedom. The helical bundles are first aligned along a common axis. In both bundles, the helical segments surrounding the dimetal and porphyrin ligands are identified (colored in blue and yellow, respectively) and used to determine their bundle axis.

With this parameterization, the optimal bundle placement became a set of $(\Delta Z, \Delta\Phi)$ coordinates, where four helical junctions could be favorably and simultaneously engineered to fill the gaps between the bundle. The linkers were selected from the database individually. The i^{th} bundle pose, defined by coordinates $(\Delta Z_i, \Delta\Phi_i)$, was decomposed into four poses corresponding to each pair of disjointed helices from the two bundles (Figure 2.3). Poses were scored using a structural designability criterion implemented through the structural similarity search program MASTER.⁵ In MASTER, an input structural element ‘query’ is compared to a collection of three-dimensional protein structures ‘library’ using a similarity metric of the backbone root-mean-square deviation ($\text{RMSD} < 1.0 \text{ \AA}$). Library structural fragments that match the query are enumerated and the designability of a query is quantified by the total number of structural matches found in the library. Operationally, for coordinates $(\Delta Z_i, \Delta\Phi_i)$

the four poses from each pair of disjointed helices became four distinct queries that were scored by their designabilities, $H_n(\Delta Z_i, \Delta\Phi_i)$, with $n=1, 2, 3, 4$. The search through $(\Delta Z, \Delta\Phi)$ space was readily visualized by a 2D contour plot (Figure 2.3), with designable regions indicated as peaks surrounded by less designable valleys.

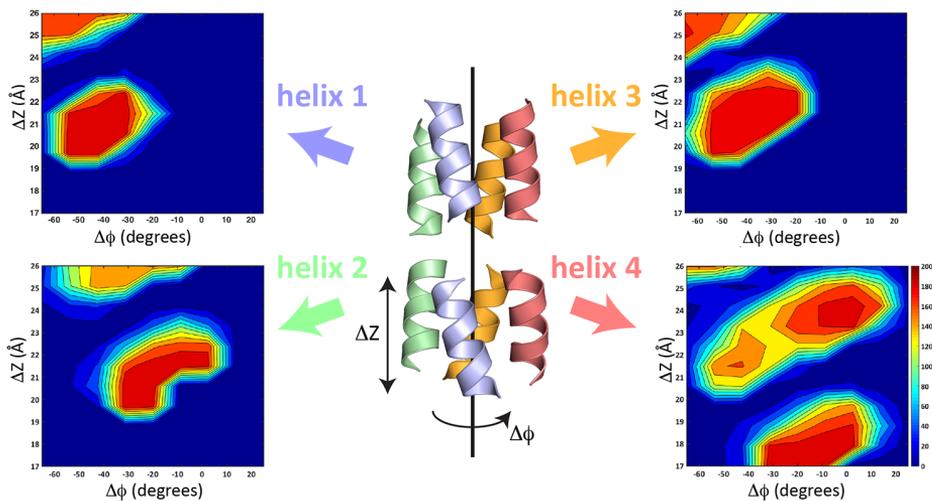


Figure 2.3. Structural designability of the four pairs of helices between bundles. The bundle (middle) is decomposed into four searches for each disjointed helical pair, helices 1, 2, 3 and 4 (same color), which are considered separately (the four panel corners). In each search one helix is fixed (top) and the other is translated, ΔZ , and rotated, $\Delta\Phi$, about the bundle axis (depicted on bottom). The four helix designability landscapes show designable ‘hotspots’ with ~ 200 structural matches surrounded by vast non-designable regions with 0 structural matches (scale included in helix 4 landscape).

2.1.1. Design of the first model DFP1

The bundle designability landscape $B(\Delta Z, \Delta\Phi)$ was considered as the product of the four helices designability landscape, determined by their distinct searches: $B(\Delta Z, \Delta\Phi) = H_1(\Delta Z, \Delta\Phi) \times H_2(\Delta Z, \Delta\Phi) \times H_3(\Delta Z, \Delta\Phi) \times H_4(\Delta Z, \Delta\Phi)$ for all $\Delta Z, \Delta\Phi$. This bundle designability landscape showed a single maximum ‘hotspot’ at $(\Delta Z,$

Chapter 2

$\Delta\Phi = (21.5 \text{ \AA}, -33.8^\circ)$ (Figure 2.4A), in which each pair of helices were offset in both translation and rotation to accommodate the gentle left-handed supertwist of the bundles.

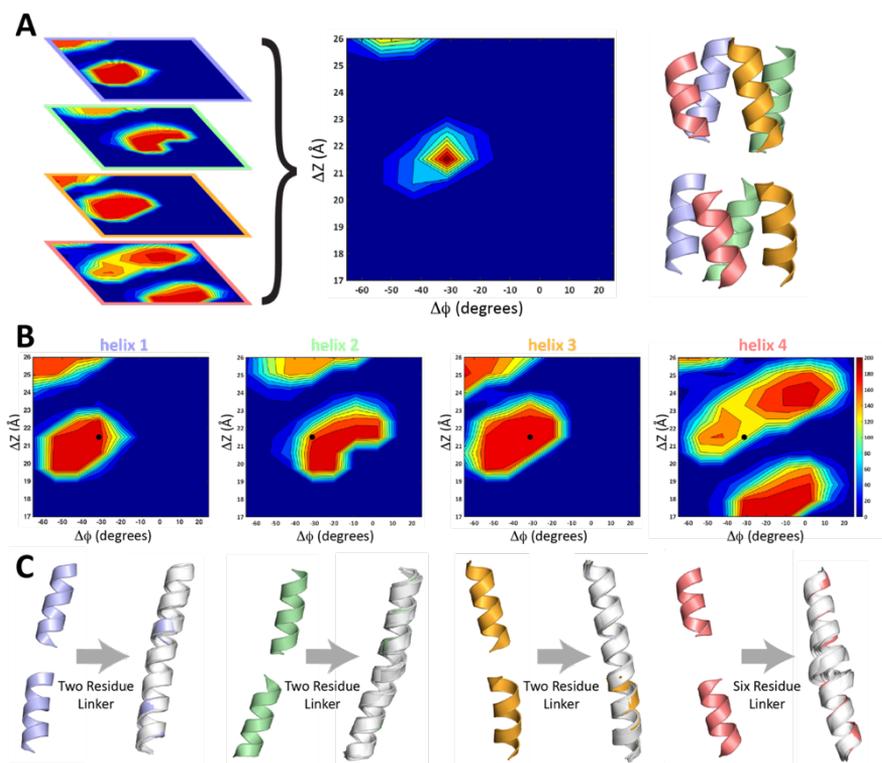


Figure 2.4. Structural solution from designability landscapes. **A.** The bundle designability landscape shows a single maximum corresponding to a specific placement of bundles (shown on the right). **B.** The four helix designability landscapes, with black dots marking the location of the bundle designability maximum. **C.** Structures of the four disjoint helix structures (colored) and their top 10 structural matches (in white). The residue length of the linker necessary to connect the two helices is specified.

Inspection of the designability landscapes of the four backbone connections revealed the positions of their hotspots had some degree of similarity (Figure 2.4B), yet they also differed due to underlying differences between bundle geometries (Figure 2.4C). As expected, structural matches occurred when the

Results and Discussion

translational and rotational offset between helical fragments were compatible with polypeptide linkers that adopted designable helix geometries. Helices 1-3 had a large designability hotspot for 2-residue linker lengths at roughly the same $(\Delta Z, \Delta\Phi)$ coordinates, whereas the 2-residue linker hotspot for helix 4 was in an entirely different location in its landscape. However, the helix 4 designability landscapes for higher residue-length (in particular 5-residue and 6-residue linkers) had hotspots that well overlapped with the hotspots of helices 1-3 with 2-residue linker lengths. Therefore, 2-residue helical junctions were chosen for helices 1-3, whereas 6-residue for helix 4 (Figure 2.5).

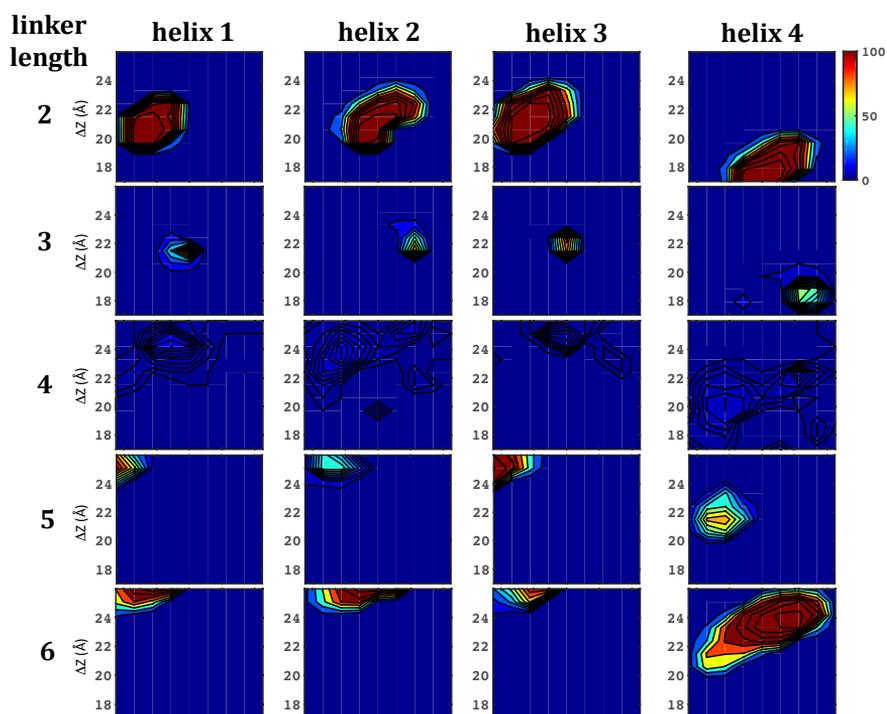


Figure 2.5. Designability landscapes for the four disjointed helices separated by connecting linker length. A bundle placement at $(\Delta Z, \Delta\Phi) = (21.5\text{\AA}, -33.8^\circ)$, with 2-residue linkers for helices 1-3 and 6-residue linker for helix 4, is the best option over the range of $(\Delta Z, \Delta\Phi)$ considered here. Scale in upper right relates the number of matches to contour plot color.

Chapter 2

To construct the final backbone structure, the four helical backbone fragments were connected into a single chain by incorporating the N- and C- termini plus the loop from the DF structure (Figure 2.6, colored blue), and the loops and helical regions comprising the folded core from the porphyrin-binding structure (Figure 2.6, colored yellow). Sequence design was restricted to the helical segments where the distinct bundles were connected, at residue positions that do not have side chains within the first- and second-shell of the dimetal-binding and porphyrin-binding sites. Backrub within Rosetta was used to sample small structural changes around the connections in conjunction with alternating loops of fixed-backbone sequence design and backbone/sidechain minimization. The results from Rosetta were compared with sequence information from the MASTER searches and these data sources along with visual inspection of the model were used to finalize the primary structure.

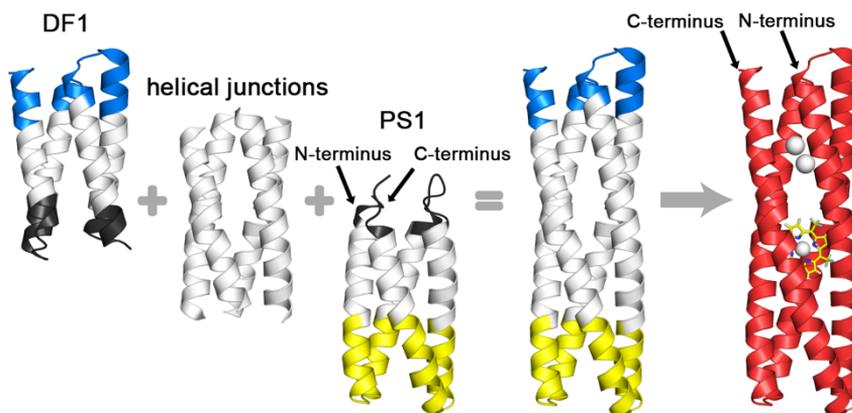


Figure 2.6. Final construction of the multi-domain protein. The bottom/top parts of the metal/porphyrin-binding proteins (black) are removed, while the top/bottom portions (blue/yellow) are reconnected to the designed helical junctions from the two bundles. The resulting DFP1 design, with metal and porphyrin ligands, is shown at the far right.

2.1.2. Structural characterization of DFP1

DFP1 was cloned and expressed as described in Chapter 4. For spectroscopic and structural characterization, the protein was reconstituted with ZnP in the porphyrin-binding site. Zn^{2+} was introduced into the dimetal site as a redox-inactive mimic of ferrous ions, as in previous studies of DF proteins. Spectral titration with ZnP demonstrated that the protein bound tightly to the cofactor. In collaboration with the Therien Lab at the Duke University (North Carolina, USA), time-resolved transient absorption spectroscopic data were acquired for ZnP-DFP1, and evinced spectral features and excited-state dynamics characteristic of the benchmark ZnP-PS1 holo-protein (Figure 2.7A).⁴

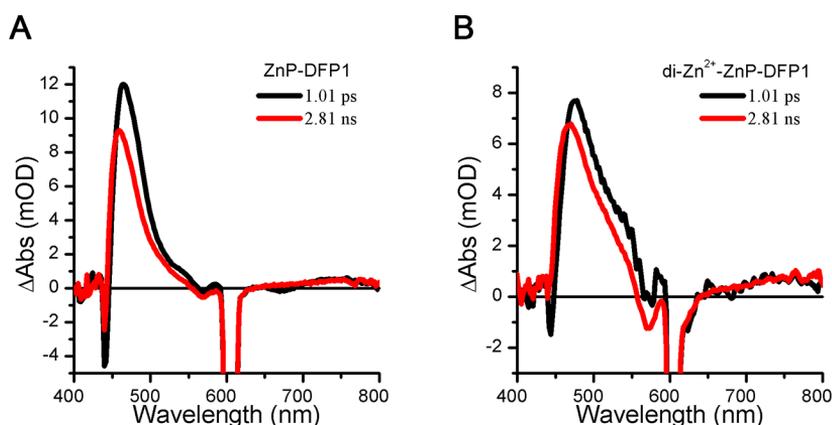


Figure 2.7. Ultrafast pump-probe spectroscopy of DFP1. **A.** Pump-probe transient absorption spectra acquired for ZnP bound in the interior of DFP1 (ZnP-DFP1) following excitation at 600 nm. The ps and ns time scale transient absorption spectra are characteristic of the respective $S_1 \rightarrow S_N$ and $T_1 \rightarrow T_N$ absorptions of the ZnP chromophore. **B.** Corresponding pump-probe transient absorption spectral data acquired for di- Zn^{2+} -ZnP-DFP1. **Experimental conditions:** solvent = 50 mM NaPi, 100 mM NaCl, pH 7.5; excitation wavelength = 600 ± 5 nm; magic-angle polarization between pump and probe pulses; pump-probe cross-correlation of ~ 250 fs.

Chapter 2

Identical experiments carried out with di-Zn²⁺-ZnP-DFP1 indicated that Zn²⁺ occupancy of protein dimetal binding site does not perturb the excited-state relaxation dynamics of the ZnP chromophore or its characteristic electronically excited singlet (S₁)- and triplet (T₁)-state absorptions (Figure 2.7B). Because S₁→T₁ intersystem crossing rate constants of electronically excited porphyrins are known to be sensitive to both macrocycle structure and the local environment,⁶ these experiments demonstrated that the ZnP binding site in PS1 is faithfully reproduced in DFP1.

The structure of the holoprotein was solved by molecular replacement to 3.5 Å resolution (Table 4.1). At this resolution, the metal ions are very well resolved, and OMIT maps allowed placement of the porphyrin macrocycle and the zinc ions (Figure 2.8).

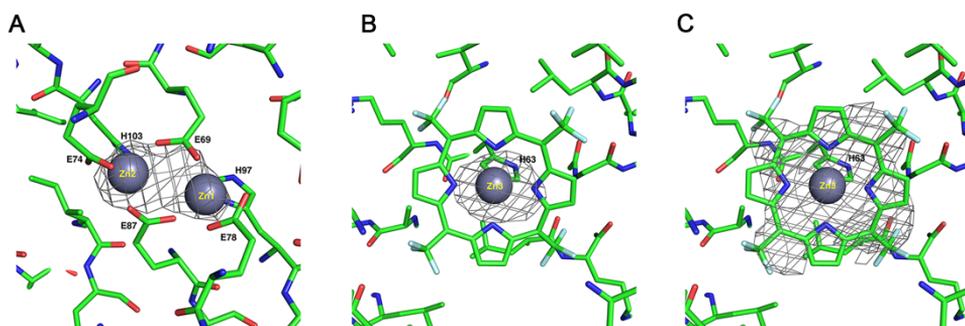


Figure 2.8. The 2Fo-Fc omit maps of the di-Zinc cluster and the ZnP bound to DFP1. The omit maps for **A.** Zn1 and Zn2, **B.** Zn3 of ZnP, and **C.** ZnP are contoured at 3.0 sigma, 3.0 sigma and 1.0 sigma, respectively. To calculate the 2Fo-Fc omit maps, the zinc ions as well as ZnP were removed from the model, and the model was refined with simulated annealing to eliminate the bias before map calculation with PHENIX.

DFP1 preserved the structural features of the starting single-domain proteins. Indeed, an excellent agreement was determined considering individually the two

domains: 0.74 Å RMSD for superposition of the ZnP domain versus PS1, and 0.52 Å for the dimetal binding domain versus DF1 (Figure 2.9).

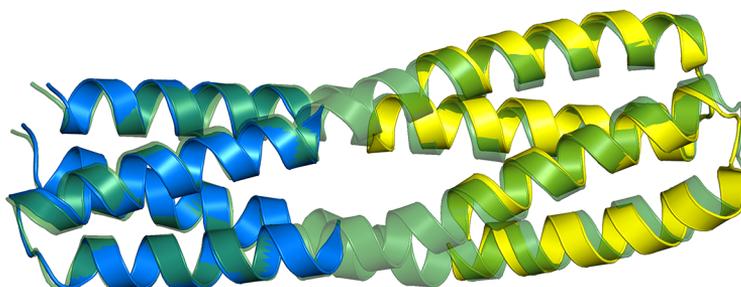


Figure 2.9. Structures of di-Mn²⁺-DF1 (PDB ID: 1JMB) and ZnP-PS1 (PDB ID: 5TGY), shown as solid blue and yellow cartoon, respectively, superimposed onto di-Zn²⁺-ZnP-DFP1 (PDB ID: 7JH6, in green), shown as green cartoon in transparency. For di-Mn²⁺-DF1 and ZnP-PS1, only the residues considered for the structural alignment are displayed.

2.2. Biophysical characterization of the functional *DFP3*

DFP1 was designed for maximal thermodynamic stability, and its interior is well-packed with apolar sidechains throughout the bundle. The resulting tight and uniform packing provided high stability, but did not leave room for organic substrates to access the dimetal-binding site.

Therefore, the four interacting Leu and Ala residues located just above the dimetal site were substituted to Gly residues (Figures 2.10A and 2.10B, respectively), resulting in a deeply invaginated substrate access cavity. The resulting 4-site mutant with Gly substitutions at positions 10, 14, 71 and 74

Chapter 2

(designated DFP2, sequence in Figure 2.11) was expressed and remained folded despite the presence of four helix-destabilizing mutations.

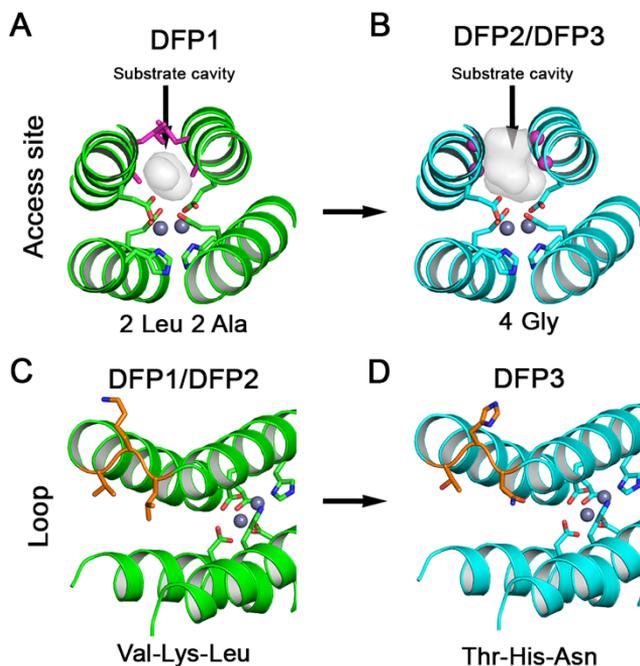


Figure 2.10. Engineering functionality in DFP family. The substrate cavity at the dimetal site was broadened mutating the 2 Leu and 2 Ala residues in DFP1 (shown as magenta sticks in **A**) in 4 Gly residues in DFP2 and DFP3 (shown as magenta spheres in **B**). To improve the oligomeric behavior, we mutated the Val-Lys-Leu loop in DFP1 and DFP2 (shown as orange sticks in **C**) in Thr-His-Asn in DFP3 (shown as orange sticks in **D**). The backbone is shown as green and cyan cartoon before and after the mutations, respectively. The coordinating residues of the dimetal center are shown as sticks, and the metal ions in grey spheres.

Nevertheless, it was found to exist in a monomer-dimer equilibrium, similar to DFP1. Size exclusion chromatography (SEC) showed the presence of two peaks of approximately equal integrated intensity at roughly the positions expected for a monomer and dimer. Upon addition of a single equivalent of the porphyrin, both peaks sharpened and shifted to a longer retention time, indicative of a

Results and Discussion

more compact conformation (Figure 2.12A). Additionally, the intensity of the monomeric peak increased, although dimers still accounted for about 20% of the total.

Sequence		
DF1	-----	0
DFP1	GDYLRELLK LELQAIKQY EKLRQTGDELVQAFQRLREIFDKGDDDSLEQV L	51
DFP2	GDYLRELLK GELQGIKQY EKLRQTGDELVQAFQRLREIFDKGDDDSLEQV L	51
DFP3	GDYLRELLK GELQGIKQY EKLRQTGDELVQAFQRLREIFDKGDDDSLEQV L	51
PS1	-----SEFEKLRQTGDELVQAFQRLREIFDKGDDDSLEQV L	36
DF1	-----DYLR ELPKLELQLIKQY REALEY VKLPV LAKILEDEEKH	39
DFP1	EEIEELIQK HRQLASEL PK LELQAIKQY REALEY VKLPV LAKILEDEEKH	101
DFP2	EEIEELIQK HRQLASEL PK GELQGIKQY REALEY VKLPV LAKILEDEEKH	101
DFP3	EEIEELIQK HRQLASEL PK GELQGIKQY REALEY THNPV LAKILEDEEKH	101
PS1	EEIEELIQK HRQLFDNRQE -----	55
DF1	IEWLE TILG -----	48
DFP1	IEWL KEAAKQGDQWVQLFQR FREAI DKGDKS LEQLLEELEQALQKIREL	151
DFP2	IEWL KEAAKQGDQWVQLFQR FREAI DKGDKS LEQLLEELEQALQKIREL	151
DFP3	IEWL KEAAKQGDQWVQLFQR FREAI DKGDKS LEQLLEELEQALQKIREL	151
PS1	-AA TEAAKQGDQWVQLFQR FREAI DKGDKS LEQLLEELEQALQKIREL	104
DF1	-----	48
DFP1	TEKTGRKILEDEEKHIEWLE TILG	175
DFP2	TEKTGRKILEDEEKHIEWLE TILG	175
DFP3	TEKTGRKILEDEEKHIEWLE TILG	175
PS1	AEK KN -----	109

Figure 2.11 Protein sequences of DFP analogues. DF1 and PS1 fragments are shown with the background in blue and yellow, respectively. The other color backgrounds represent the designed helical junctions. The residues lining the active site channel and at the loop of DF domain are represented in bold magenta and red, respectively.

The dimeric peak might represent an elongated domain-swapped-dimer of the desired protein, as previously observed in several cytochromes and myoglobins.⁷⁻⁹ The dimeric state might have similar binding properties, but it would complicate solution analysis. To address this problem, a helix-breaking sequence was introduced to stabilize the intended α_R - α_L - β inter-helical loop used in DFP1.

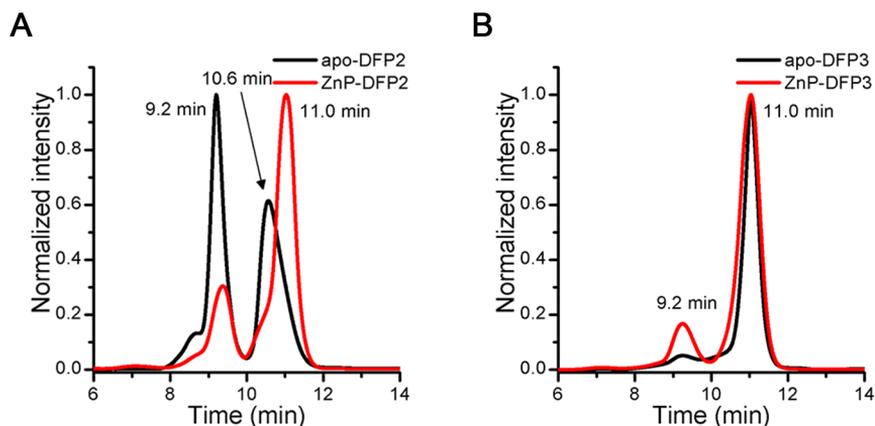


Figure 2.12. Effect of mutations on DFP3 oligomerization state. **A.** SEC of apo-DFP2 (in black) and ZnP-DFP2 (in red) at pH 7 (HEPES 50 mM, NaCl 100 mM), followed at 280 nm. **B.** SEC of apo-DFP3 (in black) and ZnP-DFP3 (in red) at pH 7 (HEPES 50 mM, NaCl 100 mM), was followed at 280 nm. Retention time peaks at 10.6 min and 11.0 min corresponded to hydrodynamic radii of 27 Å and 21 Å, respectively.

Based on a previous survey of the protein databank, the Val-Lys-Leu at positions 88-90 of DFP2 was substituted to Thr-His-Asn in DFP3 (Figures 2.10C and 2.10D, respectively).¹⁰ The resulting protein (sequence in Figure 2.11) formed only a small amount of dimer in the apo-state and when bound in the ZnP-bound state (Figure 2.12B). Moreover, the retention times were identical and as expected from the crystal structure of DFP1 (Table 4.2). This finding strongly suggested that the DFP3 was largely folded and preorganized in absence of any cofactor.

2.2.1. Binding properties of DFP3 for its metal cofactors

To evaluate the binding of ZnP to DFP3, the spectral shifts in the ZnP Soret band were used, as previously seen in PS1.⁴ The spectrum of ZnP solubilized in 1% w/v octyl- β -D-glucopyranoside micelles sharpened and shifted from 415 to 423 nm when the cofactor was titrated with increasing concentrations of DFP3 (Figure

2.13A). Nonlinear least squares fitting of the resulting spectral titration revealed a 1:1 binding stoichiometry with an apparent dissociation constant of $K_D=10\pm 2$ nM. This value represents an upper limit of the true binding constant, as it reflects the process of transfer of the porphyrin derivative from a micelle to the protein interior rather than from water to protein. A similar $K_D = 13\pm 3$ nM was observed in the presence of excess Zn^{2+} to saturate the dimetal-binding site (Figure 2.13B). Thus, DFP3 binds ZnP with exceptionally high affinity both in the apo and in the dimetal occupied states.

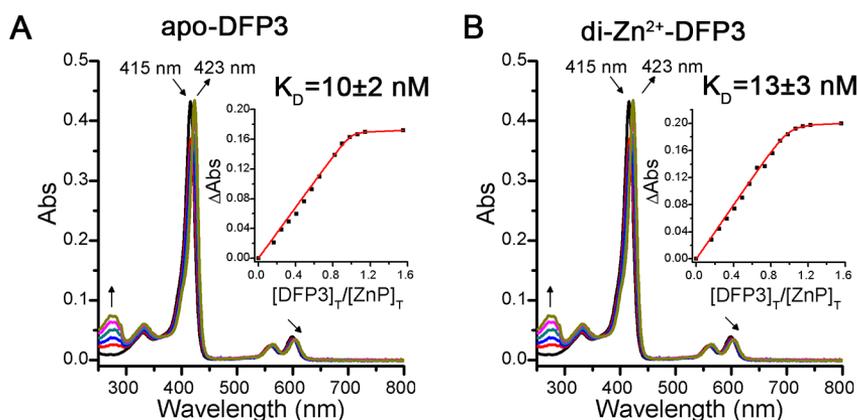


Figure 2.13. DFP3 binds ZnP with high affinity. UV-Vis spectral changes of ZnDP solutions upon addition of: **A.** apo-DFP3 and **B.** di- Zn^{2+} -DFP3 at pH 7 (HEPES 50 mM, NaCl 100 mM octyl- β -D-glucopyranoside 1% w/v). Inset: K_D determination of ZnP-DFP3 complex (Δ Abs followed at 423 nm). The smooth curve represents the best fit for a binding isotherm with a 1:1 (DFP3:ZnP) stoichiometry.

Although ZnP is achiral and hence not optically active, its ruffled conformation is rendered chiral when bound with an axial His ligand in the asymmetric binding site environment, as observed in DFP1 crystal structure. To confirm this feature in the related DFP3 protein, we examined the circular dichroism (CD) of its Soret electronic transition. When ZnP was dissolved in dichloromethane (DCM), its UV-

Chapter 2

Visible (UV-Vis) spectrum was similar to that observed in micelles with the Soret band at 415 nm (Figure 2.14A, in red). As expected, no spectral features were observed in CD spectrum in the visible region (Figure 2.14B, in red). Upon addition of an L-histidine derivative (with the N-terminus protected with the fluorenylmethoxycarbonyl group, Fmoc-HIS-OH), a shift of the Soret band at 423 nm was detected due to the zinc coordination (Figure 2.14A, in blue) by histidine. Absorption at the same wavelength was observed for the ZnP when bound by DFP3 (Figure 2.14A, in black). However, the presence of the chiral His ligand did not affect the CD spectrum of ZnP (Figure 2.14B, in blue). In contrast, ZnP-DFP3 complex displayed a very intense negative Cotton effect at ~ 415 nm (Figure 2.14B, in black), confirming that ZnP is bound in a relatively rigid asymmetric environment both in DFP1 and DFP3 proteins.

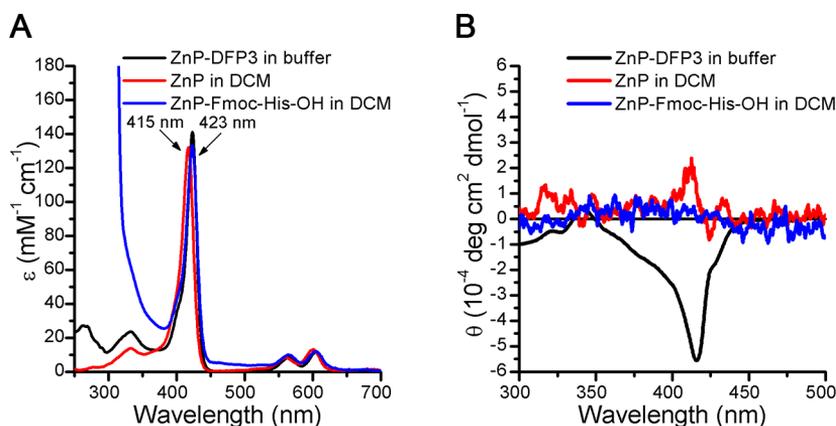


Figure 2.14. ZnP Cotton effect arises from binding with DFP3 **A.** UV-Vis spectrum of ZnP-DFP3 (in black) at pH 7 (HEPES 50 mM, NaCl 100 mM), ZnP (in red) in DCM and ZnP bound to Fmoc-His-OH in DCM. **B.** CD spectrum in the Soret region ZnP-DFP3 (in black) at pH 7 (HEPES 50 mM, NaCl 100 mM), ZnP (in red) in DCM and ZnP bound to Fmoc-His-OH in DCM.

While the crystal structure of holo-DFP1 demonstrated the binding of two Zn^{2+} in the expected geometry, the ability of apo-DFP3 to bind the first-row transition metal Co^{2+} was investigated in aqueous solution. The position and intensity of the d-d bands of Co^{2+} ions are very sensitive to the stoichiometry and geometry of the ligand environment, providing a convenient probe of ligation in solution.¹¹ The observed molar extinction coefficients ($\epsilon_{524\text{nm}}=146 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{547\text{nm}}=151 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{600\text{nm}}=82 \text{ M}^{-1} \text{ cm}^{-1}$) of the di- Co^{2+} -DFP3 complex are indicative of a pentacoordinate geometry, as seen in the earlier observed di- Co^{2+} -DF complexes (Figure 2.15A).^{3,12-14} Unfortunately, the presence of the ZnP Q-bands in the region between 550 nm and 620 nm did not allow the detection of the cobalt binding in presence of the ZnP (absorption coefficient of Q bands and cobalt d-d transition, $\sim 12000 \text{ M}^{-1}\text{cm}^{-1}$ and $150 \text{ M}^{-1}\text{cm}^{-1}$, respectively).

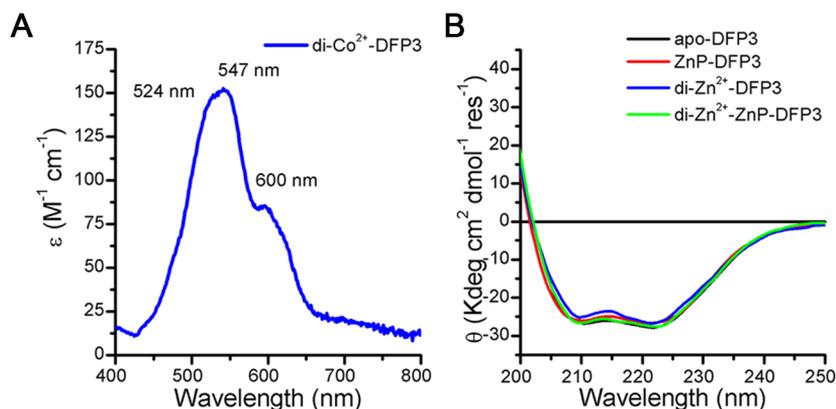


Figure 2.15. Effect of divalent ion binding. **A.** Electronic spectrum in the visible region of di- Co^{2+} -DFP3 complex at pH 7 (HEPES 50 mM, NaCl 100 mM). **B** Far-UV CD spectrum of apo-DFP3 (in black), ZnP-DFP3 (in red), di- Zn^{2+} -DFP3 (in blue) and di- Zn^{2+} -ZnP-DFP3 (in green) at pH 7 (HEPES 50 mM, NaCl 100 mM).

Finally, we examined the influence of cofactor-binding on the thermodynamic stability of the protein to heat and guanidine hydrochloride (Gdn-HCl) induced unfolding. The apo-form of DFP3 shows a far UV CD spectrum typical of an α -

Chapter 2

helical protein (Figure 2.15B), and its spectrum is exceptionally stable to thermal denaturation showing only the beginning of an unfolding transition near 90 °C (Figure 2.16A). The di-Zn²⁺, ZnP and doubly loaded di-Zn²⁺-ZnP proteins showed nearly identical far UV spectra with high stability to thermal unfolding up to 95 °C (Figures 2.16A and 2.16B).

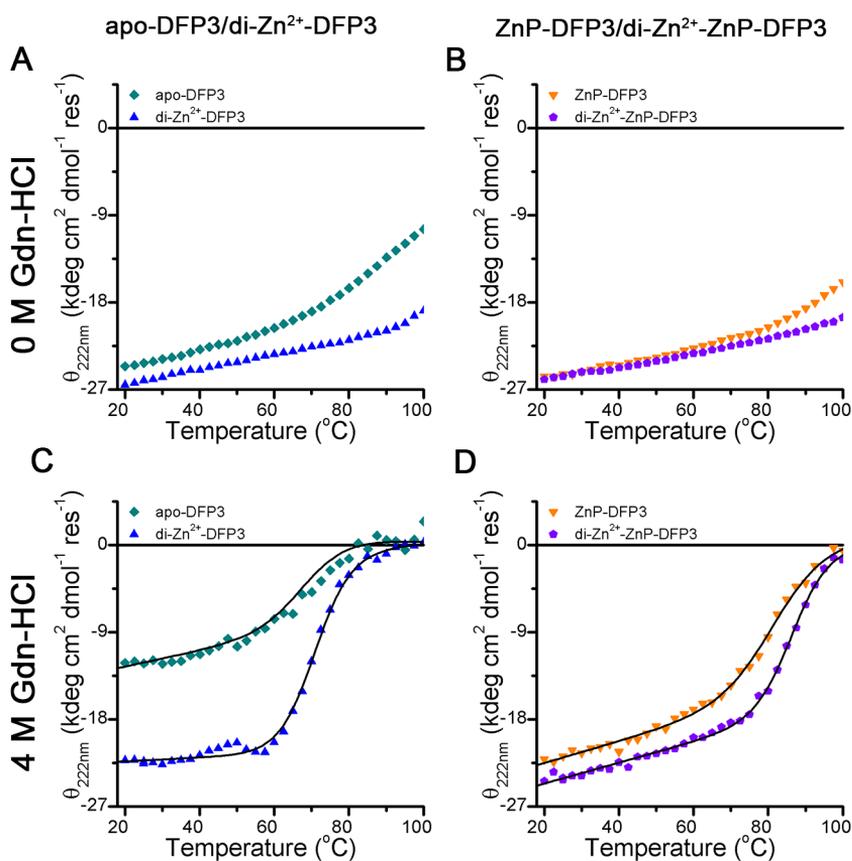


Figure 2.16. DFP3 shows enhanced thermal stability. Thermal denaturation curves of: **A.** apo-DFP3 (cyan diamonds) and di-Zn²⁺-DFP3 (blue triangles); **B.** ZnP-DFP3 (orange triangles) and di-Zn²⁺-ZnP-DFP3 (violet pentangles) at 10 μM protein concentration and pH 7 (HEPES 5 mM NaCl 10 mM). The thermal denaturations were performed also in presence of 4M Gdn-HCl (**C.** apo-DFP3 and di-Zn²⁺-DFP3, **D.** ZnP-DFP3 and di-Zn²⁺-ZnP-

DFP3). Where possible, the van't Hoff analysis (black line) was performed considering a two-state transition.

To resolve the differences in stabilities of the various forms of the protein, thermal unfolding was repeated in the presence of denaturant, 4M Gdn-HCl (Figures 2.16C and 2.16D). The observed midpoints of the unfolding curves were: di-Zn²⁺-DFP3 = 71.0 ± 0.2 °C, ZnP-DFP3 = 83.2 ± 0.7 °C and di-Zn²⁺-ZnP-DFP3 = 87.3 ± 0.4 °C. The exceptional stability is particularly important for future studies that might require destabilizing substitutions to support function.

2.2.2. Functional characterization of DFP3 dimetal site

2.2.2.a. Stabilization of radical semiquinone in the dizinc form

We next evaluated the ability of DFP3 to bind to and stabilize a reactive substrate at its dimetal active site. We previously demonstrated the redox-inert di-Zn²⁺ form of a single-domain DF protein is capable of binding to and stabilizing the otherwise highly reactive radical semiquinone form of 3,5-di-tert-butylcatechol (DTBC).^{15,16} In aqueous solution, the semiquinone form of DTBC is much less stable than either the corresponding catechol or quinone (DTBQ). Thus, the semiquinone form (DTBSQ) is present only in trace amounts when an equimolar mixture of DTBC and DTBQ is mixed in aqueous solution. However, upon addition of stoichiometric di-Zn²⁺-DFP3 or doubly loaded di-Zn²⁺-ZnP-DFP3 the quinone and catechol species slowly con-proportionate to generate the semiquinone, which is then tightly bound to the protein. The accumulation of DTBSQ is clearly visible based on the appearance of strong bands in the near IR at 740 nm and 800 nm ($\epsilon \sim 5000 \text{ M}^{-1}\text{cm}^{-1}$, Figure 2.17). The resulting radical semiquinone complex showed no signs of decomposition over the course of a day. Thus, DFP3

Chapter 2

uses binding energy to stabilize a substrate in a radical semiquinone form that is otherwise reactive and energetically inaccessible at ambient temperature.

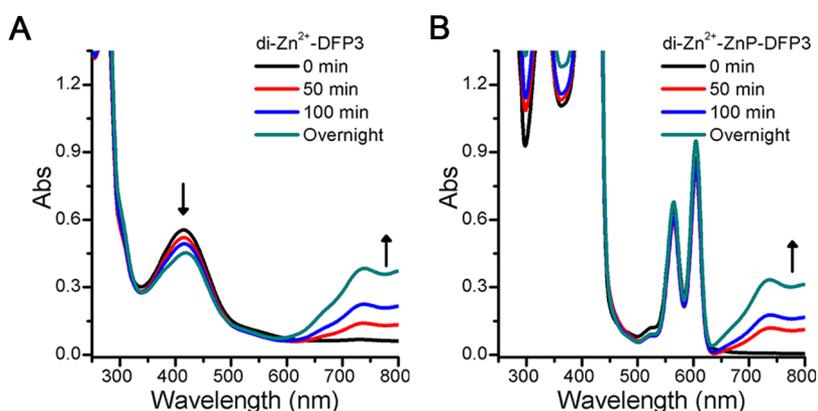


Figure 2.17. DFP3 stabilizes highly reactive semiquinone. UV/vis spectral changes upon addition of DTBC and DTBQ mixture, to generate DTBSQ *in situ*, to a solution of: **A.** di-Zn²⁺-DFP3; **B.** di-Zn²⁺-ZnP-DFP3.

2.2.2.b. Allosteric modulation of ferroxidase and oxygenase activities by porphyrin binding

Many diiron proteins catalyze the ferroxidase reaction, which involves 2-electron oxidation of ferrous (Fe²⁺) to ferric (Fe³⁺) ions with concomitant reduction of molecular oxygen to hydrogen peroxide.^{17–19} The kinetics of the reaction can be easily evaluated by monitoring the appearance of the resulting oxo-to ferric charge-transfer (LMCT) transition near 360 nm. In the presence of molecular oxygen, both apo-DFP3 and ZnP-DFP3 catalyzed the ferroxidase reaction, as evidenced by the formation of LMCT bands, with extinction coefficients 5000 and ~3750 M⁻¹ cm⁻¹ per diiron site for the resulting di-Fe³⁺-DFP3 and di-Fe³⁺-ZnP-DFP3 proteins, respectively (Figure 2.18). However, the kinetics differed

markedly between the two proteins. When apo-DFP3 was incubated with stoichiometric Fe^{2+} ions the diferric product was fully formed within 1 min (Figure 2.18A). The slow increase in absorbance at longer times in Fig. 6 is likely due to light scattering from aggregation or ferric ion precipitation, which leads to non-specific increases in absorption at wavelengths well outside the LMCT region. By contrast, a much slower process was observed when ZnP-bound DFP3 was incubated with ferrous ions under identical conditions, and conversion to the diferric form was not fully complete after 30 min (Figure 2.18B). Thus, ZnP acts as an allosteric regulator of the dimetal catalytic site of DFP3, having a profound influence on the reactivity of the diferrous state towards molecular oxygen as well as the long-term stability of the bound diferric state of the cofactor.

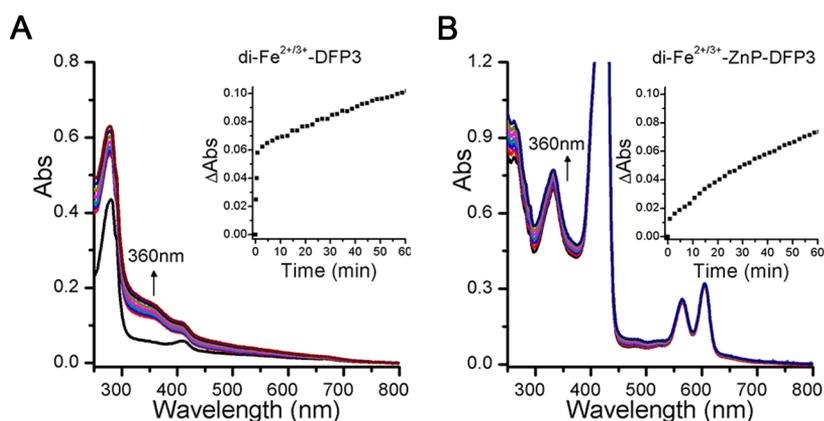


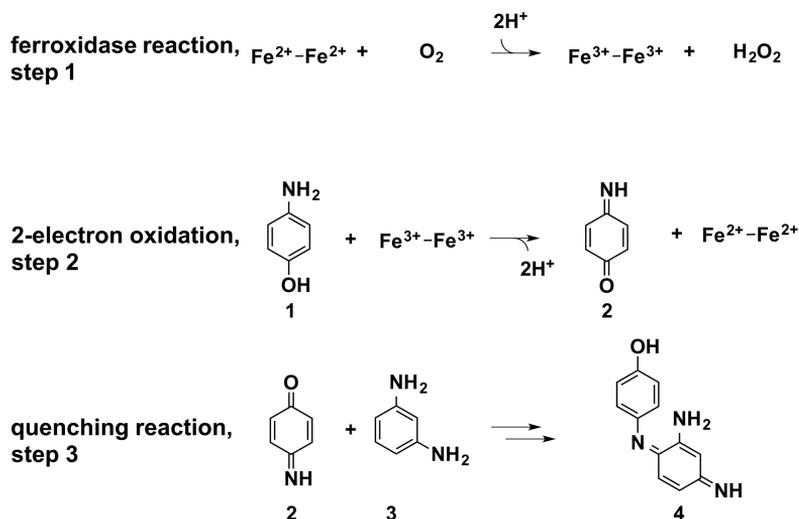
Figure 2.18. DFP3 catalyzes diiron 2-electron oxidation. A. Ferroxidase activity of apo-DFP3 at pH 7 (HEPES 50 mM, NaCl 100 mM). **B.** Ferroxidase activity of ZnP-DFP3 at pH 7 (HEPES 50 mM, NaCl 100 mM). In **A.** and **B.** the spectra display an increase in the absorbance of the typical oxo- Fe^{3+} LMCT band near 360 nm. Inset: time course of the ferroxidase reaction followed at 360 nm.

Subsequently, the ability of DFP3 to catalyze the two-electron oxidation of 4-aminophenol (4AP) to the corresponding 4-benzoquinone monoimine (4BQM)

Chapter 2

was evaluated, as studied previously for DF proteins.^{13,14,20,21} The catalytic cycle (Scheme 2.1) involves the oxidation of the diferrous protein to a diferric species by O₂ (the ferroxidase reaction, Scheme 2.1a). The diferric protein then binds and performs a two-electron oxidation of the substrate 4AP to produce 4BQM (**1** and **2**, in Scheme 2.1b, respectively). The produced diferrous form is then oxidized by O₂, thereby initiating another catalytic cycle. To allow easy detection of the quinone imine, it is rapidly quenched with 1,3-diaminobenzene (MPD) to form an aminoindoaniline dye (**3** and **4**, in Scheme 2.1c), with λ_{max} at 486 nm at pH 7 in a reaction first studied in 1879 by Witt.^{22,23}

Scheme 2.1. Oxidative coupling reactions involving 4-aminophenol



The diiron forms of both di-Fe-DFP3 and di-Fe-ZnP-DFP3 showed significant phenol oxidase activity, which was strongly modulated by the binding of ZnP (Figures 2.19A, 2.19B and 2.19C). Di-Fe-DFP3 displayed saturation kinetics (Figures 2.19D and 4.2), and a non-linear least squares fit to the data provided Michaelis-Menten parameters: $K_m = 2.9 \pm 0.3$ mM; $k_{\text{cat}} = 0.70 \pm 0.04$ min⁻¹; k_{cat}/K_m

Results and Discussion

= $2.4 \cdot 10^2 \text{ M}^{-1} \text{ min}^{-1}$. Compared to DF3 ($K_m=1.97 \pm 0.27 \text{ mM}$ and $k_{cat}=2.72 \pm 0.19 \text{ min}^{-1}$), a slight increase in K_m and a decrease in k_{cat} were observed.²⁰ The reaction progressed over multiple turnovers, and a total of 10 turnovers were observed at $[4AP] = 1 \text{ mM}$ and $[\text{di-Fe-DFP3}] = 20 \mu\text{M}$.

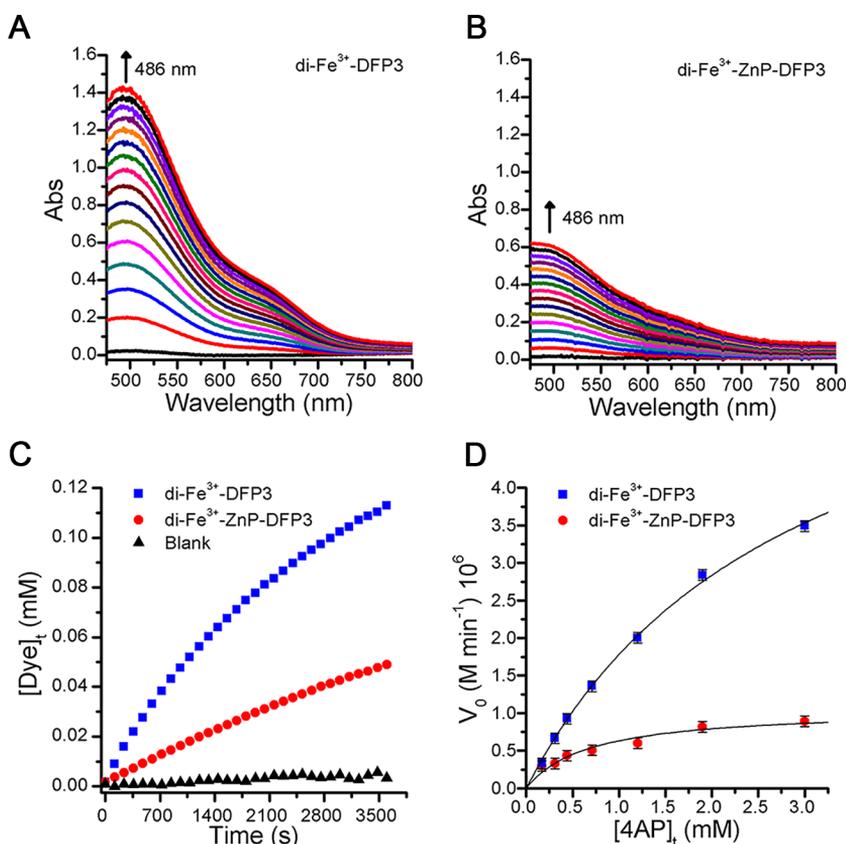


Figure 2.19. DFP3 shows allosterically regulated phenol oxidase activity. 2-electron oxidation of 4AP in presence of MPD by: **A.** 20 μM di-Fe³⁺-DFP3; **B.** 20 μM di-Fe³⁺-ZnP-DFP3 at pH 7 (HEPES 50 mM, NaCl 100 mM, 10% DMF). **C.** Time courses of 4AP oxidation, by di-Fe³⁺-DFP3 (in blue), di-Fe³⁺-ZnP-DFP3 (in red) and blank (in black), followed at 486 nm. The absorbance values were converted to aminoindoaniline dye concentration. **D.** Initial rate of the oxidation in function of 4AP concentration catalyzed by di-Fe³⁺-DFP3 (in blue) and di-Fe³⁺-ZnP-DFP3 (in red). Kinetics parameters (k_{cat} and K_m) were determined by non-linear least squares fit of the Michaelis-Menten equation.

Chapter 2

The Michaelis-Menten parameters for the corresponding ZnP-bound diiron DFP3 protein were strongly affected by the binding of the allosteric modulator ZnP, resulting in a 4-fold tighter K_m (0.68 ± 0.15 mM) and a 7-fold decrease in k_{cat} (0.106 ± 0.009 min⁻¹), corresponding to a k_{cat}/K_m of $1.6 \cdot 10^2$ M⁻¹ min⁻¹.

To provide insight into the allosteric modulation, we conducted 1 μ s all-atom molecular dynamic (MD) simulations of di-Fe²⁺-DFP3 and di-Fe²⁺-ZnP-DFP3 in triplicate. The most striking difference in the structural ensembles was a shift in the second-shell ligand, Tyr18 (Figure 2.20). In the starting model of di-Fe²⁺-ZnP-DFP3, this buried residue donates a hydrogen-bond to the metal ion ligand Glu72.

Across the three simulations, Tyr18 largely maintains this major conformation, which allows direct and water-mediated hydrogen bonds to Glu72 (Figures 2.20A and 2.20B). In a second minor conformer, the Tyr sidechain transiently points towards the surface (Figures 2.20C and 2.20D). In absence of ZnP, Tyr18 shifts and prefers this alternate sidechain conformation, in which its hydroxyl group forms hydrogen bonds with the backbone carbonyl and surface sidechains of a neighboring helix, suggesting that steric blocking of the rotation by ZnP assists the maintenance of the Tyr18-Glu72 interaction in di-Fe²⁺-ZnP-DFP3 (Figures 2.20E and 2.20F).

Throughout the simulations, there was little evidence of additional differences in the dynamics or structure of the active site or pore between the *apo*- and *holo*-states. The use of a fixed-charge atomic force field, while able to capture the shift in Tyr18 orientation (Figure 2.21), is unlikely to capture a potential change in the energetics of the active site resulting from persistent hydrogen bonding of one of the ligating carboxylate oxygens of Glu72, which may produce the shift in catalytic activity.

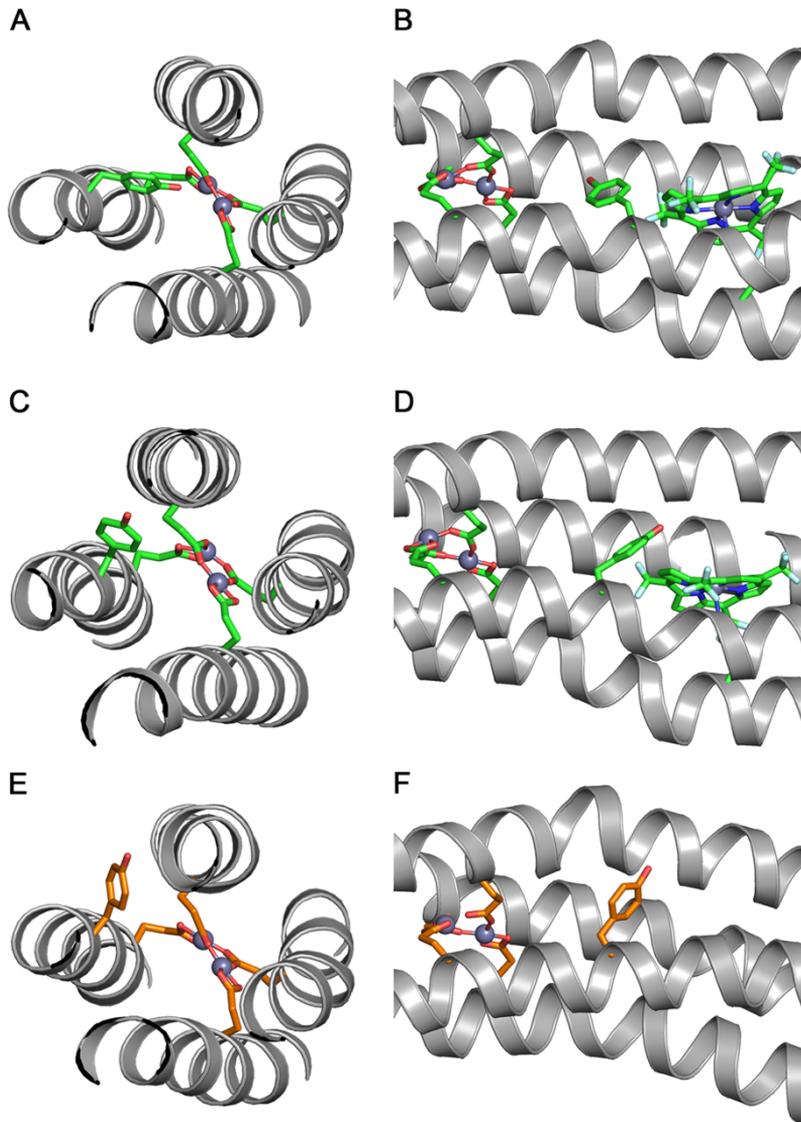


Figure 2.20. Tyr 18 shows conformation variability. In di-Zn²⁺-ZnP-DFP3, Tyr18 sidechain assumes two different orientations: in the main conformation, Tyr18 maintains a water-bridged or direct interaction with Glu72 (**A**. front and **B**. side views); in the second one, Tyr18 transiently exposes its sidechain to the surface (**C**. front and **D**. side views). In di-Zn²⁺-DFP3, Tyr18 adopts mainly the former: **E**. front and **F**. side views. Zn²⁺ ions are represented as gray balls; active side residues are represented as green or orange sticks in di-Zn²⁺-ZnP-DFP3 (**A-D**) and di-Zn²⁺-DFP3 (**E - F**), respectively.

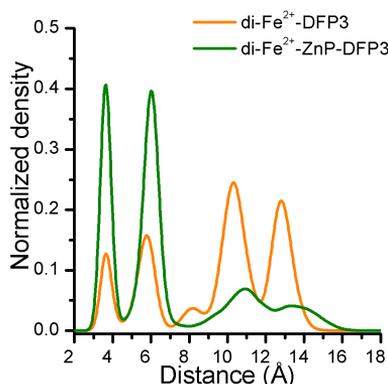


Figure 2.21. Summed and renormalized population densities of the Tyr18-Glu72 hydroxyl-carboxylate distance across the three independent simulations of di-Fe²⁺-DFP3 and di-Zn²⁺-ZnP-DFP3, in orange and green respectively. A gaussian kernel estimator was used to smooth data from the trajectories.

2.2.3. Probing allosteric bidirectionality in DFP3

The possibility to back-regulate the activity of the porphyrin-binding domain from the dimetal site was investigated changing the nature of the zinc porphyrin itself, reducing its fitness for the PS binding site. Two candidates were selected to substitute ZnP: i) Zn-tetraphenylporphyrin (ZnTPP, Figure 2.22A), in which the trifluoromethyls are substituted with phenyl groups at meso positions; ii) Zn-Deuteroporphyrin IX (ZnDP, Figure 2.22B), a less hindered analogue of the common Protoporphyrin IX.

The porphyrin-binding ability of DFP3 was evaluated by SEC. Upon binding with ZnTPP, DFP3 oligomerization equilibrium was severely affected (Figure 2.22C). The main species was dimeric (Rt=9.2 min), and only a small fraction retained the monomeric state (Rt=11.0 min). Moreover, a soluble ZnTPP aggregate was detected at 6.5 min. The UV-Vis spectrum of the monomeric fraction was characterized by the characteristic spectral features of an imidazole-bound ZnTPP (Soret band at 428 nm, Q bands at 565 and 607 nm, Figure 2.22E).²⁴

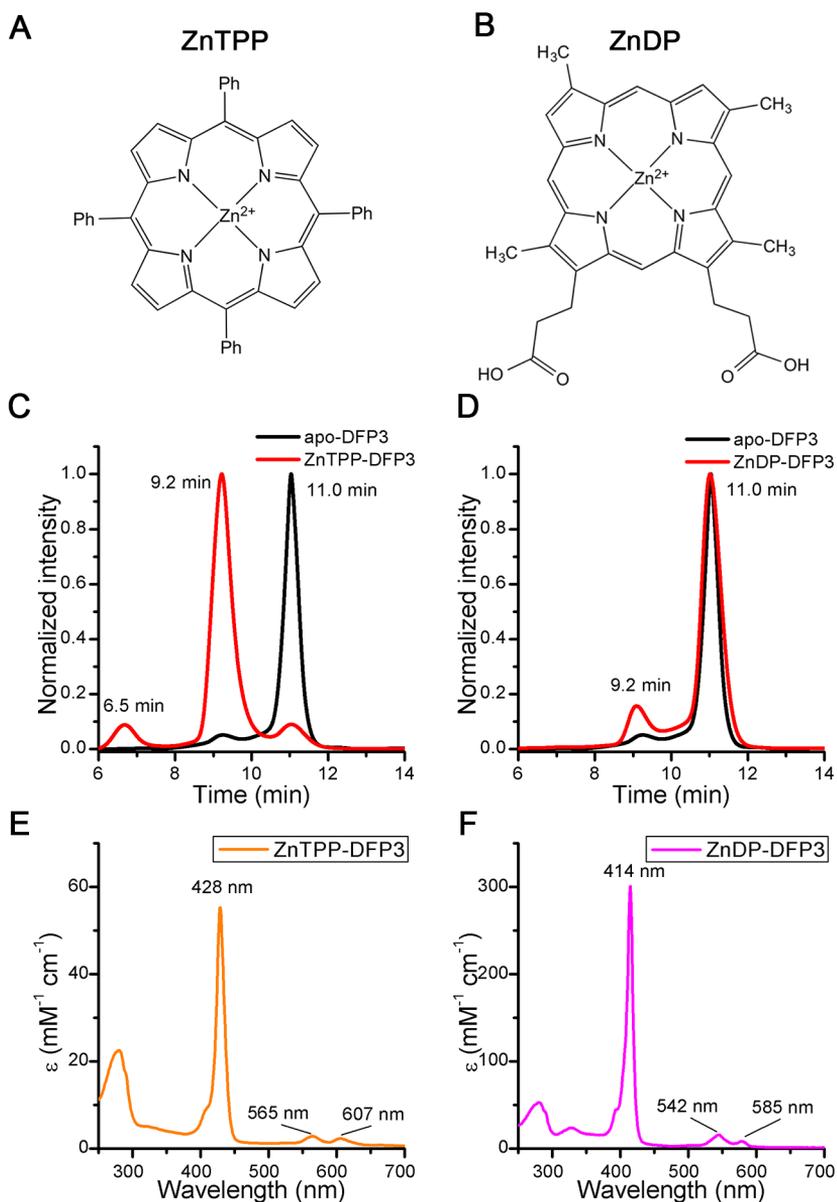


Figure 2.22. DFP3 is able to bind different zinc porphyrins. Chemical structures of the zinc porphyrins tested in the binding of DFP3: **A**. Zn tetraphenylporphyrin (ZnTPP); **B**. Zn-Deuteroporphyrin IX (ZnDP). SEC of apo-DFP3 (in black) and zinc porphyrin-reconstituted DFP3 (in red): **C**. ZnTPP; **D**. ZnDP. UV-Vis spectrum of the monomeric fraction of: **E**. ZnTPP-DFP3; **F**. ZnDP-DFP3.

Chapter 2

However, the porphyrin/protein ratio ($A_{428\text{nm}}/A_{280\text{nm}} < 3$) was non-compatible with a completely ZnTPP-DFP3 bound complex, considering the expecting absorption coefficient of the histidine-bound ZnTPP (experimental $\epsilon_{428\text{nm}} \sim 55 \text{ mM}^{-1} \text{ cm}^{-1}$; expected $\epsilon_{428\text{nm}} \sim 540 \text{ mM}^{-1} \text{ cm}^{-1}$). This result was not completely unexpected. It confirms that the hydrophobic core of the PS domain is incompatible with the presence of phenyl substituents in the porphyrin ring, as already demonstrated by the inability of the parental protein PS1 to bind a Mn diphenylporphyrin.²⁵ On the other hand, the binding of ZnDP to DFP3 left almost unaffected DFP3 oligomerization speciation, with the monomeric state as main species (Figure 2.22D). Moreover, the UV-Vis spectrum was similar to that reported for ZnDP-reconstituted Myoglobin, with the Soret band at 414 nm and Q bands at 542nm and 580 nm (Figure 2.22F).²⁶ The expected ratio $A_{414\text{nm}}/A_{280\text{nm}} > 7$ was observed, suggesting a complete formation of the ZnDP-DFP3 complex. Therefore, a more complete characterization was performed only on ZnDP-DFP3.

2.2.3.a. ZnDP binding alters the thermodynamic stability of DFP3

The binding of ZnDP did not significantly alter the α -helical content of DFP3 (Figure 2.23A). Moreover, the presence of Cotton effect in ZnDP Soret region suggested a tight binding in the asymmetric hydrophobic core (Figure 2.23B). Its sign was opposite in chirality compared to ZnP. This may be expected, considering the presence of substituents on different positions of the porphyrin ring (β and meso for ZnDP and ZnP, respectively).

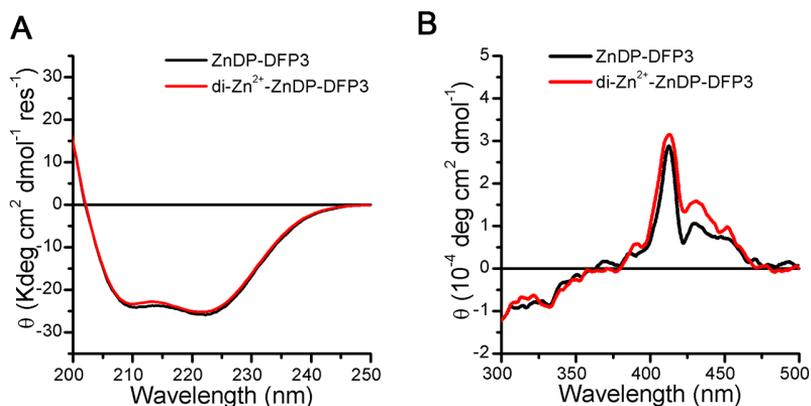


Figure 2.23. Reciprocal effect of ZnDP and DFP3 reorganization Far-UV (A.) and Visible (B.) CD spectrum of ZnDP-DFP3 (in black) and di-Zn²⁺-ZnDP-DFP3 (in red) at pH 7 (HEPES 50 mM, NaCl 100 mM).

However, the presence of ZnDP resulted in a severe change of DFP3 thermal stability (Figure 2.24A, in black). Indeed, a clear native-unfolded transition was observed at 67.1 ± 0.8 °C in ZnDP-DFP3 thermal denaturation, followed by CD spectroscopy. This represented a decrease in the melting temperature of over 30 °C, compared to the apo-state in absence of Gdn-HCl. However, the enhanced stability was restored when zinc ions were bound at the DF domain in di-Zn²⁺-ZnDP-DFP3 complex (Figure 2.24A, in red). This showed the stabilizing role of zinc ions bound at the DF domain to the overall folding, which was previously moved to the background by ZnP. Thermal unfolding was performed also in presence of 4 M Gdn-HCl, to detect the melting transition of di-Zn²⁺-ZnDP-DFP3 (Figure 2.24B). In this case, a 20 °C decrease was observed in melting temperature compared to di-Zn²⁺-ZnP-DFP3, confirming the destabilizing role of ZnDP on thermal stability.

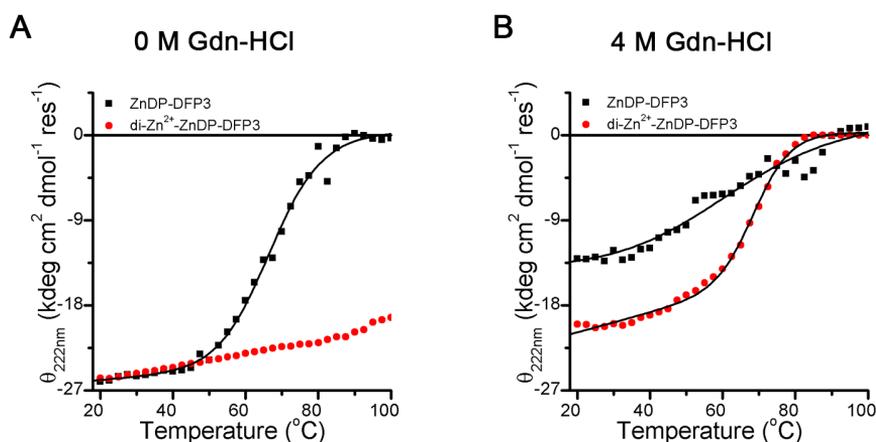


Figure 2.24. Effect of ZnDP on DFP3 thermal stability Thermal denaturation curves of ZnDP-DFP3 (black squares) and di-Zn²⁺-ZnDP-DFP3 (red circles) at 10 μ M protein concentration and pH 7 (HEPES 5 mM: **A.** in absence of Gdn-HCl and **B.** in presence of 4M Gdn-HCl. Van't Hoff analysis was performed considering a two-state transition.

Considering the different metal cofactor contributions to thermodynamic stability, chemical denaturations were also performed. The apparent free energy of unfolding (ΔG_0) extrapolated to zero denaturant concentration and $C_{1/2}$, the Gdn-HCl concentration at half denaturation, were determined for DFP3 bound to all its metal cofactors at 25 °C (Table 2.2).

Table 2.2. Thermodynamic parameters derived from the fit of the guanidine hydrochloride induced unfolding curves for DFP3, bound to its different metal cofactors.

	ΔG_0 (kcal mol ⁻¹) ^[a]	m (Kcal mol ⁻¹ M ⁻¹) ^[b]	$C_{1/2}$ (M) ^[c]
apo-DFP3	3.9 \pm 0.4	1.05 \pm 0.10	3.71
di-Zn²⁺-DFP3	7.4 \pm 1.1	1.9 \pm 0.3	3.89
ZnP-DFP3	5.1 \pm 0.5	1.16 \pm 0.13	4.40
di-Zn²⁺-ZnP-DFP3	8.3 \pm 0.9	1.8 \pm 0.2	4.61
ZnDP-DFP3	5.5 \pm 0.5	1.44 \pm 0.13	3.93
di-Zn²⁺-ZnDP-DFP3	8.0 \pm 1.0	1.9 \pm 0.3	4.21

^[a] Apparent free energy of unfolding extrapolated to zero denaturant concentration. ^[b] Linear coefficient determined linear extrapolation method. ^[c] Gdn-HCl concentration at half denaturation.

Results and Discussion

In this case, apo-DFP3 was the least stable protein with a ΔG_0 of 3.9 ± 0.4 kcal mol⁻¹ and $C_{1/2}$ of 3.7 M (Figure 2.25A, in cyan). ΔG_0 for ZnP-DFP3 and ZnDP-DFP3 were comparable, 5.1 ± 0.5 kcal mol⁻¹ and 5.5 ± 0.5 kcal mol⁻¹, respectively (Figures 2.25B in orange and 2.25C in black, respectively). However, a net difference in the $C_{1/2}$ values was observed: ZnP increases the stability to denaturant concentration to 4.4 M, in contrast ZnDP only to 3.9 M.

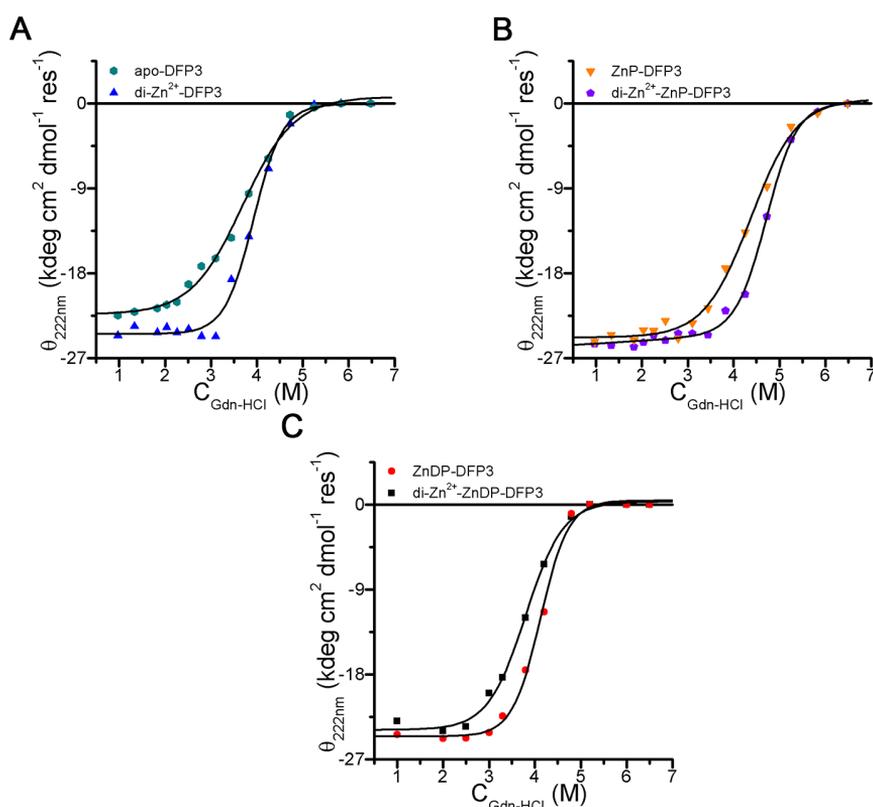


Figure 2.25. Chemical denaturation on DFP3 bound to its metal cofactors. Effect of metal cofactors on the thermodynamic stability of DFP3 was determined in Gdn-HCl denaturation: **A.** apo-DFP3 (cyan diamonds) and di-Zn²⁺-DFP3 (blue triangles); **B.** ZnP-DFP3 (orange triangles) and di-Zn²⁺-ZnP-DFP3 (violet pentangles); **C.** ZnDP-DFP3 (black squares) and di-Zn²⁺-ZnDP-DFP3 (red circles) at 10 μ M protein concentration and pH 7 (HEPES 5 mM NaCl 10 mM). The thermodynamic parameters were determined with the linear extrapolation method.

Chapter 2

Instead, the binding of zinc at the DF domain was determined to give a thermodynamic boost of $\sim 3 \text{ kcal mol}^{-1}$ in the ΔG_0 and 0.2 M in $C_{1/2}$ in all cases, compared to the DFP3-state in its absence (in di-Zn²⁺-DFP3 $\Delta G_0 = 7.4 \pm 1.1 \text{ kcal mol}^{-1}$ and $C_{1/2} = 3.9 \text{ M}$; in di-Zn²⁺-ZnP-DFP3 $\Delta G_0 = 8.3 \pm 0.9 \text{ kcal mol}^{-1}$ and $C_{1/2} = 4.6 \text{ M}$; in di-Zn²⁺-ZnDP-DFP3 $\Delta G_0 = 8.0 \pm 1.0 \text{ kcal mol}^{-1}$ and $C_{1/2} = 4.2 \text{ M}$). Moreover, the *m* values derived from the denaturation curve fitting of the di-Zn²⁺ reconstituted proteins were very similar. The *m* value is typically associated to the difference in solvent-accessible surface area (ΔSASA) between the unfolded and the folded state of a protein.²⁷ This suggests the paramount role of the zinc-binding in stabilizing the DF domain, which likely gives the main contribution to the ΔSASA .

The reactivity of the dimetal center was investigated to probe the effect of ZnDP binding. Upon air oxidation of DTBC in presence of di-Zn²⁺-ZnDP-DFP3, the characteristic bands of DTBSQ were observed between 700 and 900 nm (Figure 2.26A). The major flexibility of the protein scaffold did not hamper its ability to bind hydrophobic small molecules. The additional feature at 814 nm was likely caused by partial oxidation of ZnDP. Subsequently, ferroxidase and phenol oxidase reactivity were evaluated in order to confirm ZnDP to be an allosteric effector of the diiron center. Upon addition of Fe²⁺ to ZnDP-DFP3, the formation of the oxo-to-ferric CT at 360 nm was observed. However, the iron oxidation rate was severely slowed down by the ZnDP, with an extinction coefficient at 360 nm of only $\sim 1000 \text{ M}^{-1} \text{ cm}^{-1}$ per diiron site after 1h (compared to 5000 and $\sim 3750 \text{ M}^{-1} \text{ cm}^{-1}$ per diiron site for the resulting di-Fe³⁺-DFP3 and di-Fe³⁺-ZnP-DFP3 proteins, respectively, Figure 2.26B). A similar slowdown was observed for the oxidation of the 4AP (~ 6 - and 2-fold compared to di-Fe³⁺-DFP3 and di-Fe³⁺-ZnP-DFP3,

respectively), in the DF standard assay (1mM 4AP and 10 mM MPD) (Figure 2.26C).

This demonstrated that the binding of any zinc porphyrin represents a molecular signal to modulate the reactivity of the diiron center, through the protein matrix.

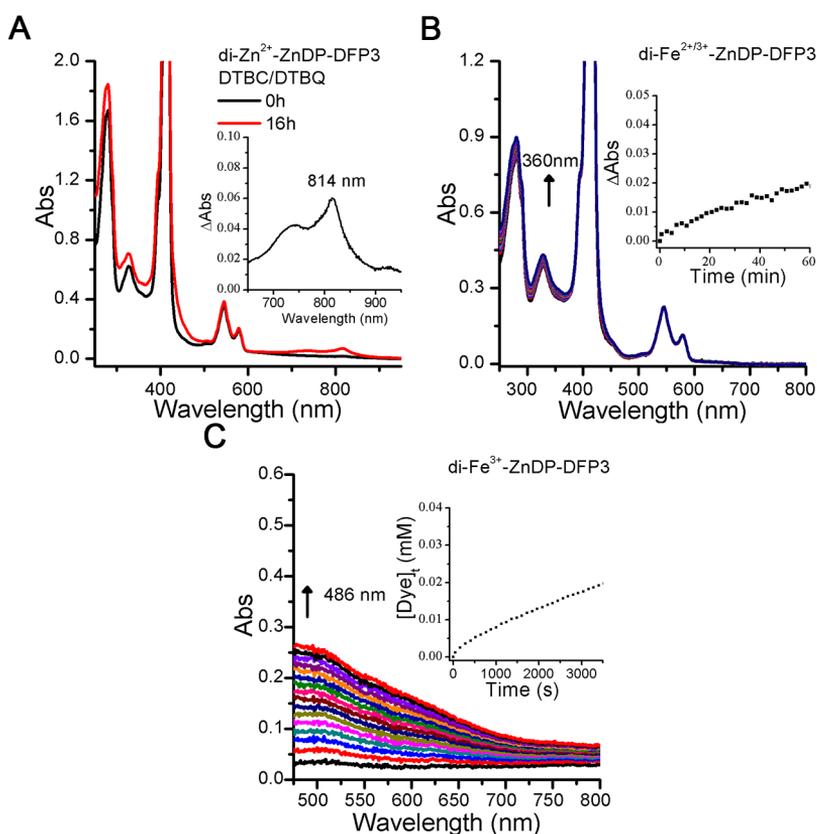


Figure 2.26. Functional characterization on the dimetal center **A.** UV-Vis spectral changes upon air oxidation of DTBC, to generate DTBSQ *in situ*, to a solution of di-Zn²⁺-ZnDP-DFP3. **B.** Ferroxidase activity at pH 7 (HEPES 50 mM, NaCl 100 mM) of: ZnDP-DFP3. The spectra display increase in the absorbance of the typical oxo-Fe³⁺ LMCT band near 360 nm. Inset: time course of the ferroxidase reaction followed at 360 nm. **C.** 2-electron oxidation of 4AP in presence of MPD by 20 μM di-Fe³⁺-ZnDP-DFP3 at pH 7 (HEPES 50 mM, NaCl 100 mM, 10% DMF). Inset: time course of 4AP oxidation, where the absorbance values were converted to aminoindole dye concentration.

Chapter 2

2.2.3.b. Allosteric control on porphyrin binding from dimetal site

Finally, the functional properties of PS domain were evaluated to have insights on allosteric regulation bidirectionality in DFP3.

Firstly, the binding properties were evaluated more extensively. Although ZnDP is charged at neutral pH, it is not soluble in buffer and aggregates. Binding affinity was determined in micellar condition (1% w/v octyl- β -D-glucopyranoside in buffer at pH 7) and, thus, was possible to compare it with ZnP. ZnDP Soret band smoothly underwent to a 406-414 nm transition upon addition of DFP3 (Figure 2.27A).

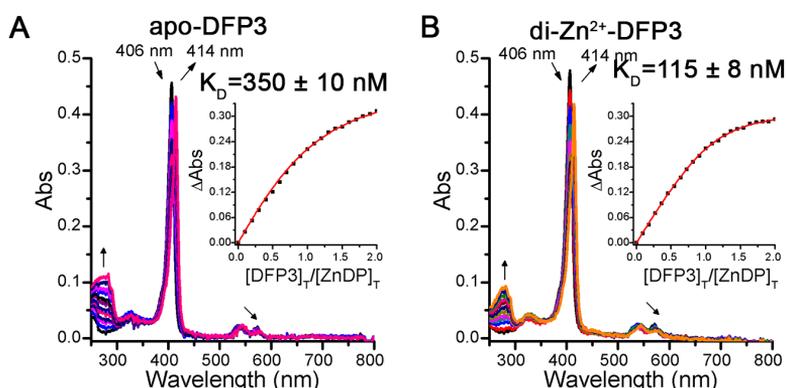


Figure 2.27. DFP3 binds ZnDP with high affinity. UV-Vis spectral changes of ZnDP solutions upon addition of: **A.** apo-DFP3 and **B.** di-Zn²⁺-DFP3 at pH 7 (HEPES 50 mM, NaCl 100 mM octyl- β -D-glucopyranoside 1% w/v). Inset: K_D determination of ZnDP-DFP3 complex (Δ Abs followed at 414 nm). The smooth curve represents the best fit for a binding isotherm with a 1:1 (DFP3:ZnDP) stoichiometry.

Even in this case, the spectral titration could be fitted with 1:1 binding stoichiometry, with an apparent dissociation constant K_D of 350 ± 10 nM. The affinity of DFP3 for ZnP is higher, as expected considering that PS domain was designed to bind this zinc porphyrin. However, the affinity for ZnDP is only one

order of magnitude lower. The presence of zinc ions bound at the DF domain had no influence at all on the binding of ZnP. Here, a three-fold improvement was observed for ZnDP with a K_D of 115 ± 8 nM (Figure 2.27B). The reorganization of the DF domain, thanks to the binding of zinc ions, shifts DFP3 equilibrium in a conformational state more eager to bind ZnDP. This demonstrated the mutual allosteric communication between the two functional sites in DFP3.

2.2.3.c. ZnDP-DFP structural model

Starting from DFP1 x-ray structure (Figures 2.28A and 2.28B), it was possible to build a model of ZnDP-DFP3 complex. ZnDP was docked into the empty porphyrin binding site of DFP1 x-ray crystal structures using AutoDock Vina.²⁸ From the seven binding modes, only the second, third and fourth in affinity (between -11.0 and -12.0 kcal mol⁻¹, as calculated by the AutoDock Vina scoring function) presented ZnDP zinc ion at binding distance from His61 (Table 2.3). The three binding poses were subsequently minimized in explicit solvent, to relax the overall protein structure that was rigid in the previous step, in particular the hydrophobic core residues near the porphyrin binding site. The energies of the minimized structures, excluding the contribution of water molecules and neutralizing ions, retained docking score trend, confirming the second binding pose as the highest in affinity and lowest in energy (Figures 2.28C and 2.28D). ZnDP porphyrin ring well superimposed on ZnP one, upon structural alignment considering only the protein backbone (Figures 2.28E).

Chapter 2

Table 2.3. Docking and minimization results of ZnDP-DFP3 structural models

Binding mode	<i>Affinity (kcal mol⁻¹)</i> ^[a]	<i>N^ε_{His61}-Zn_{ZnDP} (Å)</i>	<i>Minimized Energy (kcal mol⁻¹)</i> ^[b]
1	-12.4	5.3	
2	-12.0	2.3	-4882
3	-11.9	2.3	-4807
4	-11.6	2.3	-4760
5	-11.0	2.7	
6	-10.8	4.5	
7	-9.6	5.4	

^[a] Affinity energy as calculated by the AutoDock Vina scoring function. ^[b] Minimized energy as calculated by NAMD with charmm36m forcefield, considering only protein, ZnDP and zinc ions at DF domain.

The main differences are comprised in the region between helix 1 and helix 4, where porphyrin propionic groups interact with protein positively charged residues (Arg22 and Arg149). Moreover, one of the propionic groups is located near DF second-coordination residue Tyr18, blocking the possibility of conformation shift of the latter. This may explain the lowest ferroxidase and phenoloxidase reactivities, observed in di-Fe-ZnDP-DFP3.

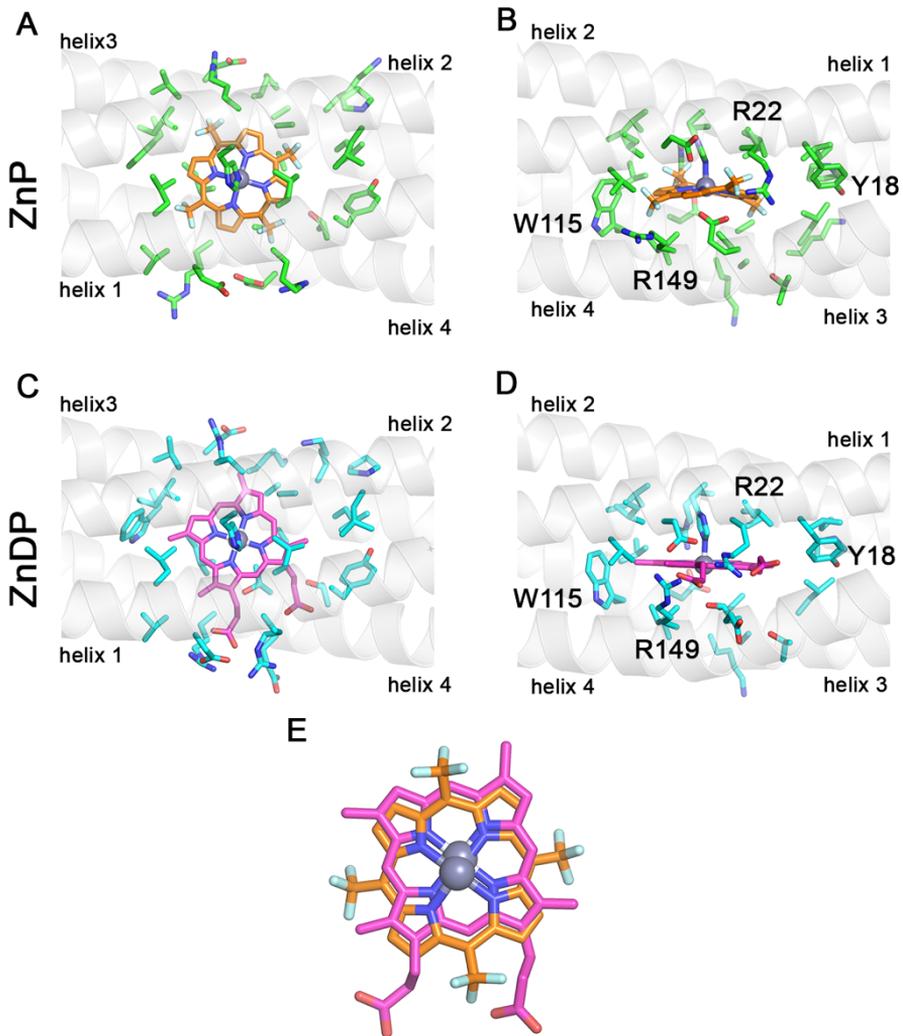


Figure 2.28. DFP3 is able to accommodate ZnDP in the porphyrin binding site. From the x-ray structure of di-Zn²⁺-ZnP-DFP1 (PDB ID: 7JH6, **A.** front and **B.** side views), a model of ZnDP-DFP3 complex was built (**C.** front and **D.** side views). **E.** Superimposition of ZnDP on ZnP. Zn²⁺ ions are represented as gray balls; protein residues are represented as green or cyan sticks in ZnP-DFP3 (**A.** and **B.**) and ZnDP-DFP3 (**C.** and **D.**), respectively. ZnP and ZnDP are represented as orange and magenta sticks, respectively.

Chapter 2

2.2.4. Photosensitizing properties of zinc porphyrins in DFP3

In final instance, the photocatalytic properties of zinc porphyrin-DFP3 complexes were investigated. Zinc porphyrins have been used as photosensitizers in dye-sensitized solar cells and photocatalysis thanks to their long-lived triplet state, generated by intersystem crossing upon excitation by visible light.^{29,30} Thus, the excited state is able to either photooxidize or photoreduce redox-active substrates, with the concomitant formation of zinc porphyrin radical anion or cation, respectively. Finally, the sacrificial oxidant/reducing agent restores the zinc porphyrin resting state.

Zinc porphyrin-reconstituted Myoglobins were demonstrated to photooxidize NADH to NAD⁺ (NAD⁺= nicotinamide adenine triphosphate),³¹ a biological redox couple that plays a key in metabolism: NAD⁺ is fundamental for the glucose degradation to produce energy. If NAD⁺ is not regenerated, glycolysis will halt.³² The photooxidation of NADH to NAD⁺ by molecular oxygen as sacrificial oxidant was tested in presence of ZnDP and ZnP.

When NADH was irradiated with Amber light (maximum of emission at 615 nm, 2.02 eV) for 24 h in blank condition, no significant decrease (~5 %) was detected at 340 nm due to the conversion to NAD⁺, compared to the control kept in dark (Figure 2.29A). In contrast, complete NADH oxidation (~80 % decrement at 340 nm) was observed in presence of ZnP-DFP3 under continued photoirradiation after 16h (Figure 2.30A). Such decrement at 340 nm was only negligible in dark conditions (Figure 2.30B).

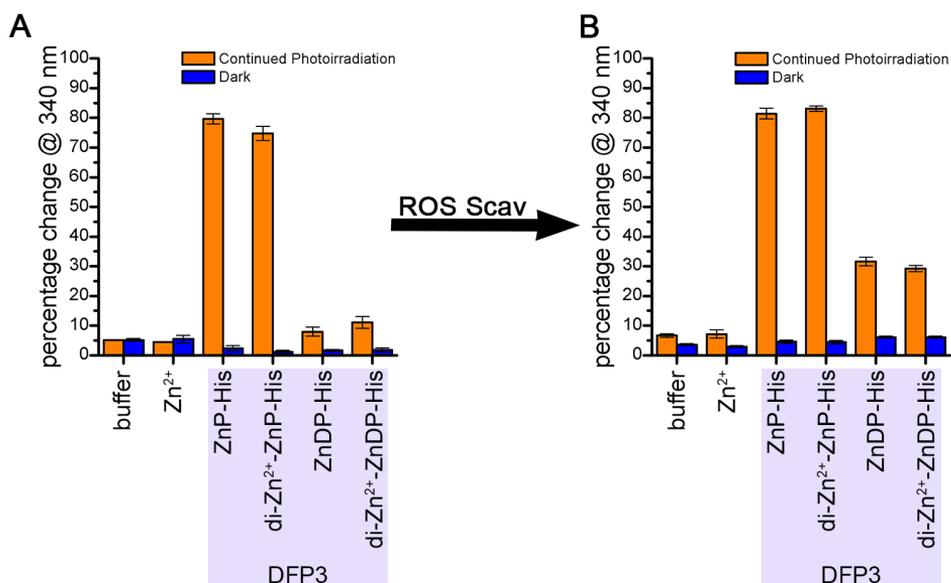


Figure 2.29. Effect of ROS scavenger on NADH photooxidation **A.** Decrement of the substrate absorbance (followed at 340 nm) observed after 24 h, under continued photoirradiation (in orange) and in dark conditions (in blue). **B.** Decrement of the substrate absorbance (followed at 340 nm) observed after 24 h, under continued photoirradiation (in orange) and in dark conditions (in blue), after addition of ROS scavenger.

As expected, the presence of zinc bound in di-Zn²⁺-ZnP-DFP3 did not significantly alter the photoreactivity (Figure 2.29A).

In presence of ZnDP, only a ~15% decrement was detected at 340 nm. Moreover, ZnDP showed significant photobleaching, considering the decrement in absorbance of the Soret band at 414 nm, and the concomitant formation of a band at 430 nm that suggested oxidative damage (Figure 2.30C). This bleaching was not observed in dark conditions (Figure 2.30D).

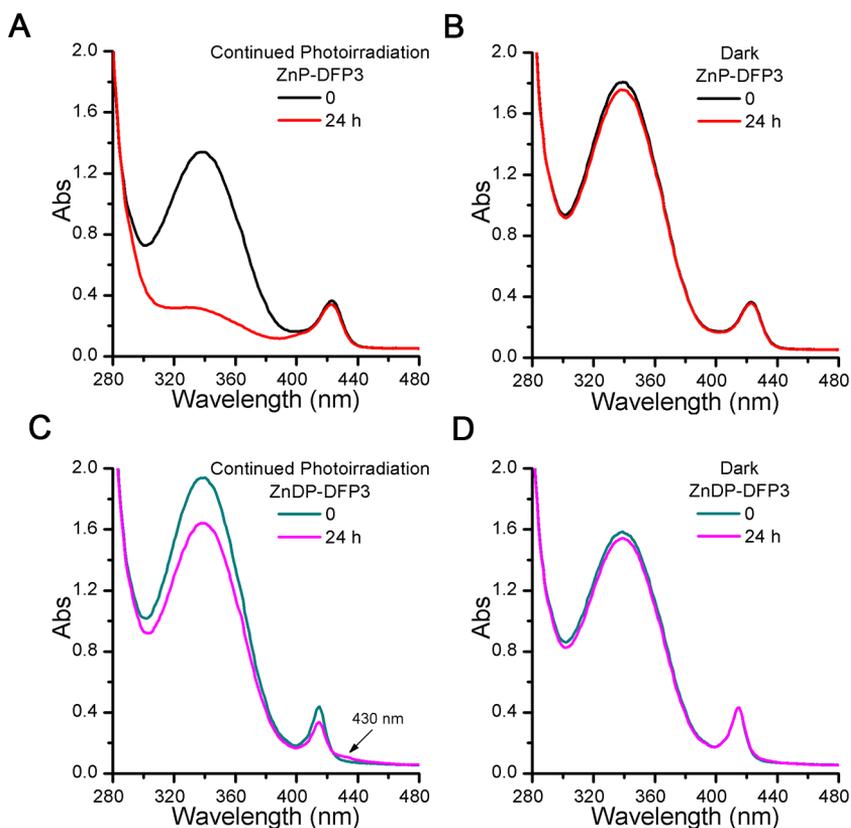


Figure 2.30. Porphyrin-bound DFP3 protoxidizes NADH. UV-Vis spectrum collected at the beginning (in black or cyan) and after 24 h (in red or pink) of the reaction in presence of di-Zn²⁺-ZnP-DFP3 **C.** under continued photoirradiation and **D.** in dark, and of di-Zn²⁺-ZnDP-DFP3: **E.** under continued photoirradiation and **F.** in dark. Under continued photoirradiation, photobleaching of ZnDP was observed at 430 nm.

Indeed, the two-electron reduction of molecular oxygen leads to hydrogen peroxide (H₂O₂). Moreover, the zinc porphyrins mediate with one-electron processes between the redox partners, suggesting the formation of superoxo (O₂^{-•}) radical anion. Both are referred to as Reactive Oxygen Species (ROS), able to oxidize organic molecules. Unexpectedly, a similar NADH photooxidation yield

Results and Discussion

and photobleaching degree was observed in di-Zn²⁺-ZnDP-DFP3, regardless of the tighter binding of ZnDP.

The same reaction was performed in presence of ROS scavengers (ROS Scav, Superoxo Dismutase and Catalase were used, Fig. 2.29B). No significant differences were detected in the blank conditions or in presence of ZnP. On the other hand, an increment to the ~30 % was observed for ZnDP, demonstrating the formation of ROS species during photocatalysis.

In order to have more insights on the effect of the protein matrix on zinc porphyrins photoreactivity, NADH photooxidation was performed in presence of zinc porphyrin bound to imidazole (Im) in micellar condition (1% w/v octyl- β -D-glucopyranoside in buffer at pH 7). Indeed, the position of the Soret band of ZnP and ZnDP resembled those of ZnP-DFP3 and ZnDP-DFP3 complexes (423 nm and 414 nm, respectively), confirming imidazole coordination (Figure 2.31).

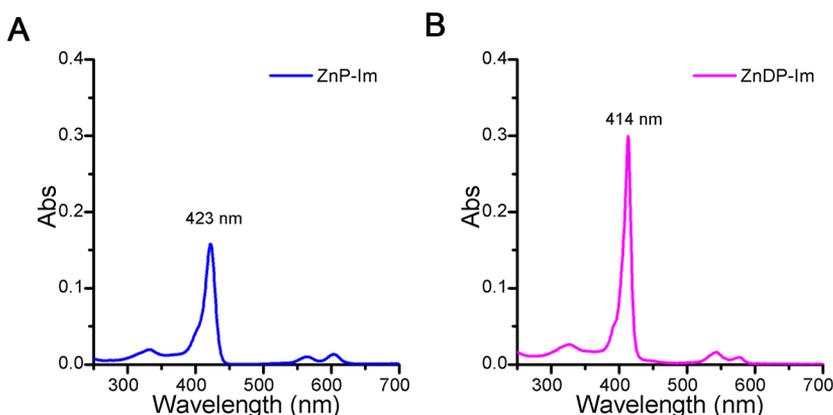


Figure 2.31. Micellar incorporation of zinc porphyrin-imidazole complexes. UV-Vis spectrum of **A.** ZnP-Im and **B.** ZnDP-Im, at 10 μ M concentration in micellar condition (HEPES 50 mM, NaCl 100 mM octyl- β -D-glucopyranoside 1% w/v).

ZnP is intrinsically more reactive than ZnDP in NADH photooxidation in the explored experimental condition. Indeed, comparing the percentage decrease at

Chapter 2

340 nm after 24 h, no differences were observed between ZnP-DFP3 and micellar-incorporated ZnP-Im, with a complete conversion of the substrate (Figure 2.32A). On the other hand, a lower NADH conversion was observed for ZnDP-Im compared to ZnDP-DFP3, suggesting an active role of the protein matrix (Figure 2.32B). The reactions performed in presence of ROS scavenger confirmed the trends observed previously.

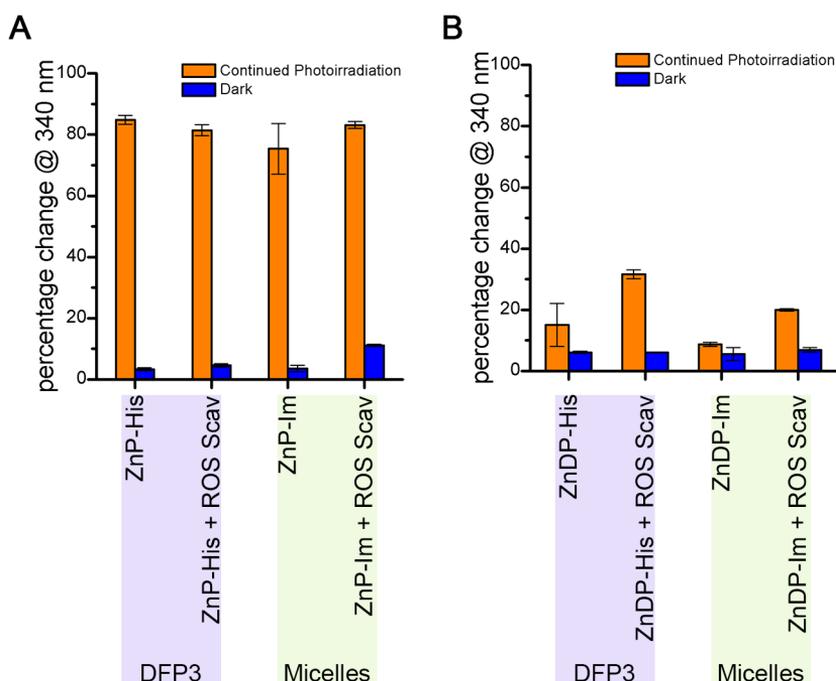


Figure 2.32. Comparison of NADH photooxidation between protein and micellar incorporation. Decrement of the substrate absorbance (followed at 340 nm) observed after 24 h, under continued photoirradiation (in orange) and in dark conditions (in blue) of **A. ZnP** and **B. ZnDP** complexes.

Subsequently, the photoreaction was followed at different times. The photoreaction rates were significantly slower when both the zinc porphyrins were incorporated in micelle (Figure 2.33, in blue). When the zinc porphyrins were bound to the protein scaffold, an increase in the conversion rates was

observed (Figure 2.33, in black). The presence of ROS scavengers gave a sensible boost to the reaction, confirming ROS formation during photocatalysis of both photosensitizers: for ZnP in particular in absence of the protein matrix (Figure 2.33A in, cyan); in contrast for ZnDP both in DFP3 presence and absence (Figure 2.33B, in red and cyan, respectively).

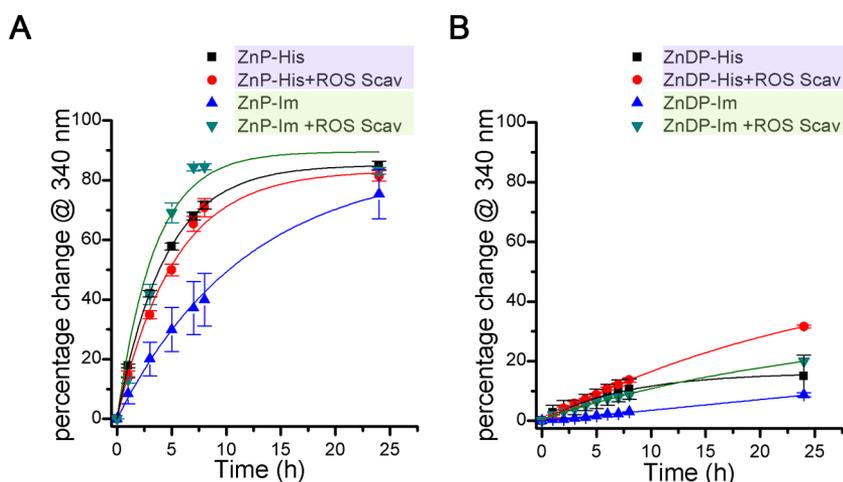


Figure 2.33. Kinetic effect of DFP3 and ROS Scav on NADH photooxidation. Time courses of NADH photooxidation followed at 340 nm by: **A.** ZnP and **B.** ZnDP. The reaction was performed with the zinc porphyrin bound to DFP3 (in black and red) and incorporated in micelles (in blue and cyan). The effect of ROS scavenger on kinetic was also investigated (in red and cyan). The experimental data were fitted with a mono-exponential function.

The protein matrix enhanced the reaction rate and, thus, the photosensitizer turnover frequency (TOF). ROS scavenger, regenerating molecular oxygen, the sacrificial reductant, and at the same time decreasing ROS concentration, led to an increase in photosensitizer TOF and turnover number (TON), respectively. Interestingly, when ZnDP was bound to imidazole in micellar buffer, no photobleaching and oxidation were observed after 24 h (Figure 2.34A), in contrast to ZnDP-DFP3 (Figure 2.34B). Therefore, the photoreactivity of ZnDP-Im

Chapter 2

was performed in absence of micelles, to investigate whether the absence of bleaching was due to the low permeability of H_2O_2 and $\text{O}_2^{\cdot-}$ in micelles and membranes.³³

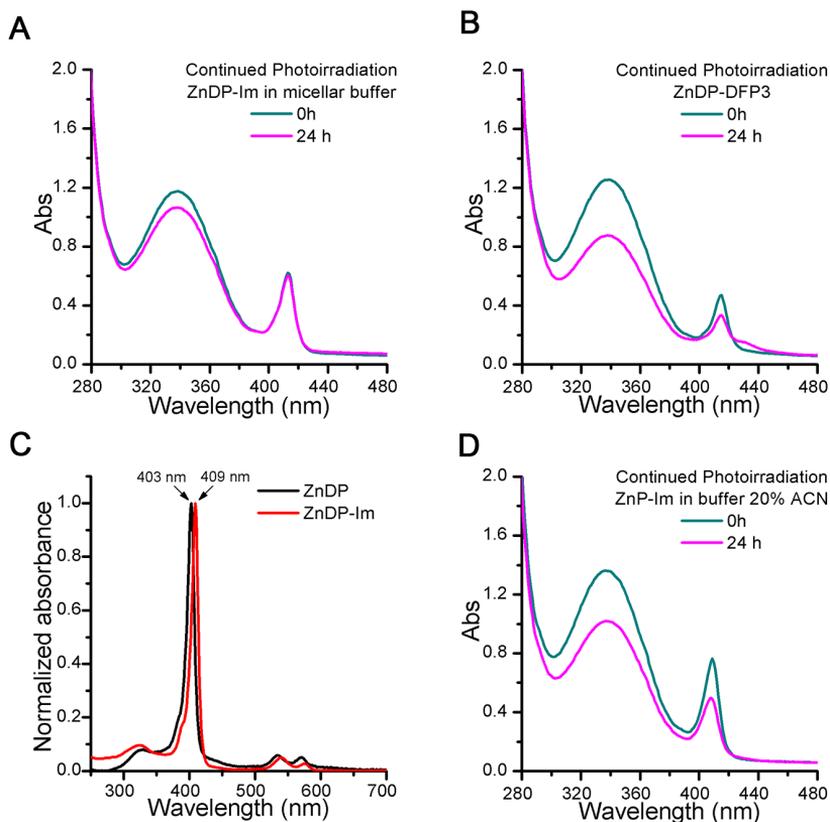


Figure 2.34. Comparison of the effect of photoirradiation on ZnDP-DFP3 and ZnDP-Im. UV-Vis spectrum collected at the beginning (in cyan) and after 24 h (in magenta) of the reaction under continued photoirradiation of: **A.** ZnDP-Im in micellar buffer and **B.** ZnDP-DFP3. **C.** UV-Vis spectra of ZnDP and ZnDP-Im (in black and red, respectively) in buffer % 20 ACN. **D.** UV-Vis spectrum collected at the beginning (in cyan) and after 24 h (in magenta) of the reaction under continued photoirradiation of ZnDP-Im in buffer % 20 ACN.

ZnDP was not soluble in buffer at pH 7 (HEPES 50 mM, NaCl 100 mM) even in presence of imidazole, but required the presence of acetonitrile (ACN) as co-

solvent, at 20% v/v concentration. The maximum of ZnDP Soret band was centered at 403 nm and shifted to 409 nm upon imidazole coordination (Figure 2.34C). After 24 h, a decrease in absorbance was detected both at NADH and ZnDP-Im band (Figure 2.34D). This confirmed that $\text{H}_2\text{O}_2/\text{O}_2^{\cdot-}$ formation was responsible for porphyrin bleaching. Moreover, this highlighted that the designed binding site of PS domain could protect only ZnP from the oxidative attack. The lower binding fitness and the higher flexibility of ZnDP-DFP3 complex allowed a greater permeability of ROS towards the porphyrin binding site, leading to ZnDP bleaching.

References

1. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).
2. Grigoryan, G. & DeGrado, W. F. Probing Designability via a Generalized Model of Helical Bundle Geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011).
3. Lombardi, A. *et al.* Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci.* **97**, 6298–6305 (2000).
4. Polizzi, N. F. *et al.* De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
5. Zhou, J. & Grigoryan, G. Rapid search for tertiary fragments reveals protein sequence–structure relationships. *Protein Sci.* **24**, 508–524 (2015).
6. Gentemann, S. *et al.* Variations and Temperature Dependence of the Excited State Properties of Conformationally and Electronically Perturbed Zinc and Free Base Porphyrins. *J. Phys. Chem. B* **101**, 1247–1254 (1997).
7. Hirota, S. Oligomerization of cytochrome c, myoglobin, and related heme proteins by 3D domain swapping. *J. Inorg. Biochem.* **194**, 170–179 (2019).

Chapter 2

8. Nagao, S. *et al.* Thermodynamic Control of Domain Swapping by Modulating the Helical Propensity in the Hinge Region of Myoglobin. *Chem. – Asian J.* **15**, 1743–1749 (2020).
9. Yoshida, N., Higashi, M., Motoki, H. & Hirota, S. Theoretical analysis of the domain-swapped dimerization of cytochrome c: An MD and 3D-RISM approach. *J. Chem. Phys.* **148**, 025102 (2018).
10. Lahr, S. J. *et al.* Analysis and Design of Turns in α -Helical Hairpins. *J. Mol. Biol.* **346**, 1441–1454 (2005).
11. Bertini, I. & Luchinat, C. High spin cobalt(II) as a probe for the investigation of metalloproteins. *Adv. Inorg. Biochem.* **6**, 71–111 (1984).
12. Torres Martin de Rosales, R. *et al.* Spectroscopic and metal-binding properties of DF3: an artificial protein able to accommodate different metal ions. *JBIC J. Biol. Inorg. Chem.* **15**, 717–728 (2010).
13. Chino, M. *et al.* A De Novo Heterodimeric Due Ferri Protein Minimizes the Release of Reactive Intermediates in Dioxygen-Dependent Oxidation. *Angew. Chem. Int. Ed.* **56**, 15580–15583 (2017).
14. Reig, A. J. *et al.* Alteration of the oxygen-dependent reactivity of de novo Due Ferri proteins. *Nat. Chem.* **4**, 900–906 (2012).
15. Ulas, G., Lemmin, T., Wu, Y., Gassner, G. T. & DeGrado, W. F. Designed metalloprotein stabilizes a semiquinone radical. *Nat. Chem.* **8**, 354–359 (2016).
16. Yoon, J. H. *et al.* Uno Ferro, a de novo Designed Protein, Binds Transition Metals with High Affinity and Stabilizes Semiquinone Radical Anion. *Chem. – Eur. J.* **25**, 15252–15256 (2019).
17. Wei, P., Skulan, A. J., Wade, H., DeGrado, W. F. & Solomon, E. I. Spectroscopic and Computational Studies of the de Novo Designed Protein DF2t: Correlation to the Biferrous Active Site of Ribonucleotide Reductase and Factors That Affect O₂ Reactivity. *J. Am. Chem. Soc.* **127**, 16098–16106 (2005).

18. Snyder, R. A., Butch, S. E., Reig, A. J., DeGrado, W. F. & Solomon, E. I. Molecular-Level Insight into the Differential Oxidase and Oxygenase Reactivities of de Novo Due Ferri Proteins. *J. Am. Chem. Soc.* **137**, 9302–9314 (2015).
19. Snyder, R. A. *et al.* Systematic Perturbations of Binuclear Non-heme Iron Sites: Structure and Dioxygen Reactivity of de Novo Due Ferri Proteins. *Biochemistry* **54**, 4637–4651 (2015).
20. Faiella, M. *et al.* An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **5**, 882–884 (2009).
21. Kaplan, J. & DeGrado, W. F. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11566–11570 (2004).
22. Brown, K. C. & Corbett, J. F. Benzoquinone imines. Part 16. Oxidation of p-aminophenol in aqueous solution. *J. Chem. Soc. Perkin Trans. 2* 308–311 (1979) doi:10.1039/P29790000308.
23. Witt, O. N. Ueber neue Farbstoffe. *Berichte Dtsch. Chem. Ges.* **12**, 931–939 (1879).
24. Nappa, M. & Valentine, J. S. The influence of axial ligands on metalloporphyrin visible absorption spectra. Complexes of tetraphenylporphinatozinc. *J. Am. Chem. Soc.* **100**, 5075–5080 (1978).
25. Mann, S. I., Nayak, A., Gassner, G. T., Therien, M. J. & DeGrado, W. F. De Novo Design, Solution Characterization, and Crystallographic Structure of an Abiological Mn–Porphyrin-Binding Protein Capable of Stabilizing a Mn(V) Species. *J. Am. Chem. Soc.* (2020) doi:10.1021/jacs.0c10136.
26. Nocek, J. M., Sishta, B. P., Cameron, J. C., Mauk, A. G. & Hoffman, B. M. Cyclic Electron Transfer within the [Zn-Myoglobin, Cytochrome *b₅*] Complex. *J. Am. Chem. Soc.* **119**, 2146–2155 (1997).
27. Myers, J. K., Nick Pace, C. & Martin Scholtz, J. Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).

Chapter 2

28. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
29. Chen, Y., Li, A., Huang, Z.-H., Wang, L.-N. & Kang, F. Porphyrin-Based Nanostructures for Photocatalytic Applications. *Nanomaterials* **6**, 51 (2016).
30. Birel, Ö., Nadeem, S. & Duman, H. Porphyrin-Based Dye-Sensitized Solar Cells (DSSCs): a Review. *J. Fluoresc.* **27**, 1075–1085 (2017).
31. Chang, C.-H. *et al.* Photoactivation Studies of Zinc Porphyrin-Myoglobin System and Its Application for Light-Chemical Energy Conversion. *Int. J. Biol. Sci.* **7**, 1203–1213 (2011).
32. Xie, N. *et al.* NAD⁺ metabolism: pathophysiologic mechanisms and therapeutic potential. *Signal Transduct. Target. Ther.* **5**, 227 (2020).
33. Cordeiro, R. M. Reactive oxygen species at phospholipid bilayers: Distribution, mobility and permeation. *Biochim. Biophys. Acta BBA - Biomembr.* **1838**, 438–444 (2014).

Chapter 3: Conclusions

The research activity carried out in the frame of this Ph.D. project was focused on expanding the functional complexity of de novo design metalloproteins. In particular, a new design method was developed to combine the functionalities of DF and PS protein in a single multi-domain scaffold. The strategy here reported differs from previous approaches, as multiple junctions were used to allow tight coupling of protein domains, in contrast to a one or two linker to fuse the two domains in end-to-end manner or through domain insertion, respectively.

The design strategy was successful not only in preserving the original structural and functional properties of DF and PS1 proteins, but also in allowing the allosteric communication between the domains to control the catalysis. The binding of ZnP into PS1 domain markedly decreased K_m and k_{cat} values in the phenol oxidation at DF domain. MD simulations suggested that it might be associated with a shift in a second-shell Tyr, which was previously found to stabilize the conformation of one Glu ligand in DF. Mutations of the second shell Tyr in DF proteins have been shown to largely affect the reactivity of the diiron center. Indeed, the ferroxidase/oxidase reactions require multiple shifts in carboxylate ligands, as the protein adjusts to changes in hydration and protonation that accompany binding of O_2 at the di- Fe^{2+} center.

Subsequently, the nature of the zinc porphyrin was demonstrated to have a great impact on the thermodynamic stability and degree of allostery achievable in DFP3. The substitution of ZnP, the designed target of PS domain, with ZnDP gave greater flexibility to the protein scaffold. However, such flexibility not only

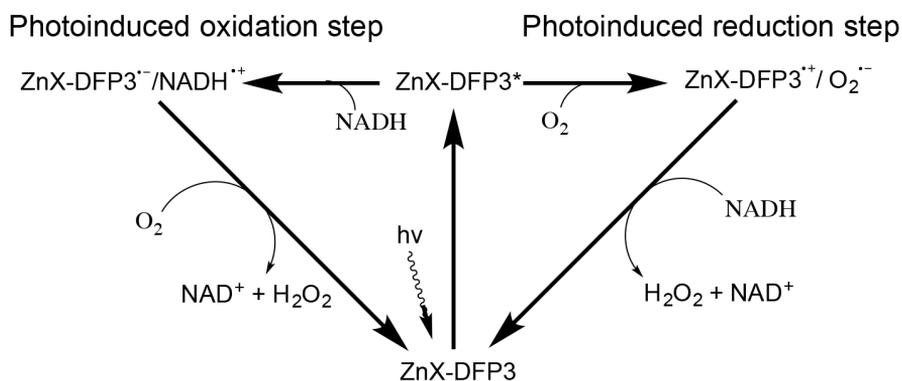
Chapter 3

did not hinder the allosteric effect on the dimetal site by the zinc porphyrin, but also allowed underlining the bidirectionality of this modulation: the affinity of DFP3 for ZnDP increased 3-fold when zinc ions were bound at DF domain.

Finally, ZnP- and ZnDP-DFP3 complexes were tested in the photooxidation of NADH by molecular oxygen. The preliminary characterization highlighted the paramount role of the protein matrix in increasing the conversion rate and controlling the permeability of harmful ROS at the porphyrin binding site. Moreover, ZnP was demonstrated to be intrinsically more reactive than ZnDP. Ultrafast spectroscopy studies will be required to investigate whether this difference in reactivity corresponds to a difference in NADH photooxidation mechanism: upon irradiation, ZnX-DFP3 (X=P or DP) complexes can oxidize NADH with the formation of porphyrin radical anion (Scheme 3.1, on the left), or reduce oxygen generating a radical cation on ZnX (Scheme 3.1, on the right).

Overall, the results reported in this Ph.D. thesis pave the way for the de novo design of allosterically regulated proteins and photosynthetic molecular machines.

Scheme 3.1. Possible mechanisms of NADH photooxidation with molecular oxygen as sacrificial reductant, mediated by ZnX-DFP3 complexes.



Chapter 4: Materials and Methods

4.1. Computational design of DFP proteins

4.1.1. Bundle axis determination, co-linear alignment of the two bundles, and generation of all structural poses

For all designs, the DF1 (PDB ID: 1JMB) and PS1 (PDB ID: 5TGY) crystal structures were used as starting template. We first identified 12 residue helical segments for each of the four chains in the bundle surrounding the dimetal or porphyrin binding sites. In DF1, helix 1 is chain B residues 5-16, helix 2 is chain C residues 5-16, helix 3 is chain C residues 31-42, and helix 4 is chain B residues 31-42. In PS1, helix 1 is chain A residues 5-16, helix 2 is chain A residues 38-49, helix 3 is chain A residues 60-71, helix 4 is chain A residues 93-104. In both bundles the $C\alpha$ positions of each helical segment were fit to a cylinder.¹ The fit provided the cylinder axis, which was used to approximate the helical axis of a segment. Finally, the four helical axes in the bundle were point-by-point averaged to calculate the bundle axis. Once the helical axes of both bundles were determined, they were aligned along the z-axis by applying the appropriate the translation and rotation matrices. Finally, pdb coordinates of all poses were generated by fixing DF1 and moving PS1 along their shared z-axis by $\Delta Z = 17$ to 26 \AA using 0.25 \AA steps, and rotating PS1 about the z-axis using angular offset $\Delta\Phi = -65^\circ$ to 25° using 2.5° steps. The coordinate positions of DF1 and PS1 were defined by the midpoint of the helical axis of helix 1 for both bundles, $(X_{DF1}, Y_{DF1}, Z_{DF1})$ and $(X_{PS1}, Y_{PS1}, Z_{PS1})$, respectively. The Z distance offset is $\Delta Z = ||Z_{DF1} - Z_{PS1}||$,

Chapter 4

and the angular offset is $\text{Cos}[\Delta\Phi] = [(\mathbf{X}_{\text{DF1}} - \mathbf{X}_{\text{PS1}})^2 + (\mathbf{Y}_{\text{DF1}} - \mathbf{Y}_{\text{PS1}})^2]^{0.5} / [(\mathbf{X}_{\text{DF1}}^2 + \mathbf{Y}_{\text{DF1}}^2)^{0.5} (\mathbf{X}_{\text{PS1}}^2 + \mathbf{Y}_{\text{PS1}}^2)^{0.5}]$.

4.1.2. MASTER searches

4.1.2.a. Database of structures

All structures were obtained from the Protein Database. Entries were queried for X-ray crystal structures with resolution less than 2.0 Å. In each retrieved pdb file, a single protein chain was extracted into a new coordinate file that was included in the database. Each of the coordinate files was converted into a binary structural file using MASTER (v1.3) with the createPDS executable. A total of 15,768 chains comprised the library.

4.1.2.b. Structural searches

Each pose consisting of a pair of disjointed helices was converted into a searchable file using MASTER (v1.3) with the createPDS executable. Next, the pose was used to query the database of structures. Matches were counted if the backbone RMSD of the query structure to a library structure fragment were less than 1.0 Angstrom and the structural match was a continuous chain. The designability of the query structure was quantified by the number of matches found in the library. For a given pair of disjointed helices the designability of 37steps ´ 37steps = 1369 distinct poses were scored.

4.1.3. Rosetta design

After constructing the DFP1 backbone, Rosetta was used to design the sequences of backbone fragments connecting the four helices comprising the

bundle.² All other residue positions were set to the amino acids at those positions in DF1 and PS1.

Backbone minimization was first performed with Ca positions fixed to remove any unphysical backbone geometries resulting from DFP1 construction. Next a round of design was conducted using the soft-repulsive force field ‘soft_rep_design’ with layer design where packing progressed from protein interior, to the interior-exterior boundary interface, and finally to the protein surface. This was followed with another round of design using the standard talaris2014 ‘hard’ force field with layer design. After this first design iteration, Rosetta backrub was performed on a 16-residue window around each of the four helices used to connect the helical bundles. Following backrub, the bundle was subjected to subsequent iterations of soft and hard force-field design using the GenericMonteCarlo mover. In total, 400 designs were generated, with the vast majority of sequences converging on the final protein sequence of DFP1.

4.2. Experimental characterization of DFP proteins

4.2.1. Protein expression and purification

All chemicals were purchased from either Sigma or Fisher Scientific, with no further purification. The gene for DFP1 was introduced into vector pET28a (Novagen), with an N-terminal hexahistidine tag followed by a TEV protease site using NcoI and XhoI restriction site. The genes for DFP2 and DFP3 were introduced into vector pET11a (Novagen), with an N-terminal hexahistidine tag followed by a TEV protease cleavage site, via Gibson assembly. The cloned gene sequences were E. coli codon optimized by Genscript:

Chapter 4

DFP1

ATGCACCACCATCATCACCACGAGAACCTGTATTTCCAAGGCGATTACTTGCGCGAGC
TTTTAAACTGGAGCTGCAAGCAATTAACAATATGAAAACTTCGCCAACTGGAG
ATGAACTGGTCCAGGCTTTCCAGCGTCTGCGTGAAATCTTTGACAAGGGCGACGATG
ACTCCTTGGAACAAGTATTGGAAGAGATCGAGGAGTTGATTCAGAAGCACCGTCAAC
TTGCGTCTGAGTTACCAAAGCTGGAACCTCAAGCGATCAAACAGTACCGTGAGGCTTT
AGAGTACGTAAATTGCCCGTCTGGCGAAGATTCTGGAAGATGAAGAGAAACACAT
TGAGTGGCTTAAGGAAGCGGCCAAGCAAGGCGATCAGTGGGTACAACCTGTTTCAAC
GCTTCGCGAAGCCATCGACAAAGGTGATAAAGATAGTCTTGAGCAGCTGCTGGAGG
AACTGGAACAGGCTTTACAAAAGATTGCGGAATTGACCGAGAAAACCTGGCCGTAAAA
TCCTTGAAGACGAGGAAAAGCATATCGAGTGGTTGGAACAATCTTAGGGTAA

DFP2

ATGCACCACCATCATCACCACGAGAACCTGTATTTCCAAGGCGATTACCTGCGCGAAC
TGCTGAAGGGCGAACTGCAAGGGATCAAGCAGTACGAGAAGCTGCGTCAAACCGGT
GATGAACTGGTGCAGGCGTTCCAACGTCTGCGTGAGATCTTTGACAAGGGCGACGAT
GACAGCCTGGAACAGGTTCTGGAGGAAATCGAGGAACTGATTCAGAAACACCGTCA
ACTGGCGAGCGAGCTGCCGAAGGGGAACTGCAGGGTATTAACAATACCGTGAGG
CGCTGGAATATGTGAAGCTGCCGTTCTGGCGAAAATCCTGGAGGATGAAGAGAAG
CATATTGAGTGGCTGAAGGAAGCGGCGAAACAGGGTGATCAATGGGTGCAGCTGTT
CCAACGTTTTCTGTAAGCGATCGACAAGGGCGATAAAGACAGCCTGGAGCAGCTGCT
GGAGGAACTGGAACAGGCGCTGCAAAAGATTCTGTGAGCTGACCGAAAAACCGGTC
GCAAGATTCTGGAGGACGAGGAAAAACACATTGAGTGGCTGGAAACCATCTGGGT
TAA

DFP3

ATGCACCACCATCATCACCACGAGAACCTGTATTTCCAAGGCGATTACCTGCGCGAAC
TGCTGAAGGGCGAACTGCAAGGGATCAAGCAGTACGAGAAGCTGCGTCAAACCGGT
GATGAACTGGTGCAGGCGTTCCAACGTCTGCGTGAGATCTTTGACAAGGGCGACGAT
GACAGCCTGGAACAGGTTCTGGAGGAAATCGAGGAACTGATTCAGAAACACCGTCA
ACTGGCGAGCGAGCTGCCGAAGGGGAACTGCAGGGTATTAACAATACCGTGAGG
CGCTGGAATATACCCACAACCCGGTTCTGGCGAAAATCCTGGAGGATGAAGAGAAGC
ATATTGAGTGGCTGAAGGAAGCGGCGAAACAGGGTATCAATGGGTGCAGCTGTTC
CAACGTTTTCTGTAAGCGATCGACAAGGGCGATAAAGACAGCCTGGAGCAGCTGCT
GGAGGAACTGGAACAGGCGCTGCAAAAGATTCGTGAGCTGACCGAAAAAACCGGTC
GCAAGATTCTGGAGGACGAGGAAAAACACATTGAGTGGCTGGAAACCATTCTGGGT
TAA

The expressed protein sequences were finally:

DFP1

MHHHHHHENLYFQ/GDYLRELLKLELQAIKQYKLRQTGDELVQAFQRLREIFDKGDDDS
LEQVLEEIEELIQKHRQLASELPKLELQAIKQYREALEYVKLPVLAKILEDEEKHIEWLKEAAK
QGDQWVQLFQRFREAIKGDKDSLEQLLEELEQALQKIRELTEKTGRKILEDEEKHIEWLE
TILG

DFP2

MHHHHHHENLYFQ/GDYLRELLKGELQGIKQYKLRQTGDELVQAFQRLREIFDKGDDD
SLEQVLEEIEELIQKHRQLASELPKGELOGIKQYREALEYVKLPVLAKILEDEEKHIEWLKEAA
KQGDQWVQLFQRFREAIKGDKDSLEQLLEELEQALQKIRELTEKTGRKILEDEEKHIEWL
ETILG

Chapter 4

DFP3

MHHHHHHHENLYFQ/GDYLRRELLKGELQGIKQYEKLRQTGDELVQAFQRLREIFDKGDDD
SLEQVLEEIEELIQKHRQLASELPKGELQGIKQYREALYTHNPVLAKILEDEEKHIEWLKEA
AKQGDQWVQLFQRFREAIKDGKDSLEQLLEELEQALQKIRELTEKTGRKILEDEEKHIEW
LETILG

where the “/” defines the cleavage site of TEV protease.

Their sequences were confirmed (Genewiz San Francisco), and they were then transformed and expressed in One Shot BL21(DE3) chemically competent *E. coli* (Thermo Fisher Scientific). All the three proteins were expressed and purified as follows. BL21(DE3) cells were grown in LB broth (100 mg ml⁻¹ kanamycin) to optical density (OD) 0.6 - 0.8 at 37 °C, then induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for 4 hours (1L per construct). Cells were harvested by centrifugation and resuspended in 60 mL of buffer A (50 mM HEPES pH 7, 100 mM NaCl, 20 mM imidazole) lysed by sonication at 4 °C, and then centrifuged at 18,000 g at 4 °C for 25 min. The supernatant was applied to 2.5 mL of Ni-NTA beads pre-washed with buffer A and the His-tagged protein were eluted with buffer B (50 mM HEPES pH 7, 100 mM NaCl, 300 mM imidazole). The His-tag was subsequently cleaved with a His-tagged TEV protease in buffer C (50 mM TrisHCl pH 8, 0.5 mM EDTA, 1 mM DTT), incubated overnight at 4 °C. The unreacted His-tagged DFP protein and the His-tagged TEV protease were removed by application of 2.5 mL of Ni-NTA beads pre-washed with buffer C. Finally, the proteins were exchanged in the final buffer containing 50 mM HEPES 100 mM NaCl pH 7.

4.2.2. Pump-Probe Transient Absorption Spectroscopy

Ultrafast transient absorption experiments were performed following previously reported methods.³

4.2.3. DFP1 crystallization and structure determination

To prepare the doubly loaded di-Zn²⁺-ZnP-DFP1, Chelex 100 was used first to remove all bound metal ions to apo-DFP1. The di-Zn²⁺-ZnP-DFP1 was then reconstituted from the metal ion-free form as previously described and in excess of Zn²⁺. Finally, ZnCl₂ was added to an additional free Zn ion concentration of 100 μM. The protein was crystallized by vapor-diffusion hanging-drop method at 25 °C, with a 1:1 (v/v) mixture of protein solution (5mg/mL in 100 mM NaCl, 100 mM HEPES, pH 7.0 and 100 μM ZnCl₂) and reservoir solution (22% wt/vol PEG 4000, 100 mM MgCl₂, 100 mM Hepes, pH 7.0) equilibrated against the reservoir solution. The crystals were frozen in liquid nitrogen, and the diffraction data were collected at 100 K at the Beamline 8.3.1 of the Advanced Light Source (Berkeley, CA). The X-ray wavelength was 1.11583 Å. The data were processed with XDS.⁴ The structure was solved by molecular replacement with Phaser using the designed model as a search model. Four porphyrin-bound helical bundles exist in the crystallographic asymmetric unit. The structure refinement was done with REFMAC.⁵ In particular, rigid-body refinement was heavily performed with rigid-body domains down from single helical bundles to single helices. The restrained refinement was then done, with TLS refinement combined at late stage; for TLS refinement the TLS domains were set down to single helical bundles. During the restrained refinement, the non-crystallographic symmetry restraints were applied among the four bundles. The software COOT was used

Chapter 4

for structural model adjustments.⁶ The data processing and structural refinement statistics were shown in Table 4.1.

The superimposition of the single domains on DFP1 was performed considering only $C\alpha$ atoms. The total symmetric di-Mn²⁺-DF1 x-ray structure of (PDB ID: 1JMB), first NMR structure of ZnP-PS1 ensemble (PDB ID: 5TGY) and di-Zn²⁺-ZnP-DFP1 chain D (PDB ID: 7JH6) were considered in the following defined ranges:

- di-Mn²⁺-DF1: 1-17 and 31-48 chain A and 5 and 43 chain B
- ZnP-PS1: 6-46 and 69-101
- di-Zn²⁺-ZnP-DFP1: 2-18,67-105 and 158-175 for DF domain and 21-64 and 108-151 for PS1 domain

Table 4.1. Summary of the data collection and refinement statistics for DFP1

Wavelength	1.116
Resolution range	86.24 - 3.5 (3.625 - 3.5)
Space group	C2
Unit cell	172.478 27.825 188.494 90 117.162 90
Total reflections	66230 (6450)
Unique reflections	10708 (1008)
Multiplicity	6.2 (6.4)
Completeness (%)	99.63 (100.00)
Mean I/sigma(I)	6.49 (0.80)
Wilson B-factor	125.21
R-merge	0.1418 (2.184)
R-meas	0.1552 (2.379)
R-pim	0.06219 (0.931)
CC1/2	0.997 (0.322)
CC*	0.999 (0.698)
Reflections used in refinement	10692 (1008)
Reflections used for R-free	503 (52)
R-work	0.261
R-free	0.283
Number of non-hydrogen atoms	6081
macromolecules	5880
ligands	200
solvent	1
Protein residues	696
RMS(bonds)	0.005
RMS(angles)	0.97
Ramachandran favored (%)	96.37
Ramachandran allowed (%)	3.63
Average B-factor	154.5
macromolecules	155.0
ligands	138.5

Values in the parentheses are corresponding to the outmost resolution bin. Randomly selected 5% of the reflections were omitted from refinement but used for R-free factor calculation.

Chapter 4

4.2.4. Size Exclusion chromatography

The monomeric forms were purified by SEC on an AKTA FPLC (GE) fitted with a Superdex 75 Increase 10/300 or a Superdex 200 Increase 10/300 and eluted at 4 °C with HEPES (50 mM, pH 7)/NaCl (100 mM) buffer solution, at a 0.3 mL min⁻¹ flow rate. ZnP and ZnTPP were dissolved in DMSO to a final concentration of 12 mM. ZnDP was dissolved at the same concentration in basic water (10 mM NaOH). The ZnP-DFP, ZnTPP-DFP and ZnDP-DFP complexes were prepared adding a 2-fold excess of the porphyrin from a stock solution to a 50 mM HEPES, 100 mM NaCl, pH 7 buffer with apo-DFP3 protein (the final DMSO concentrations were kept to <5%). The solution was incubated for 15 min at 70 °C and subsequently filtered before injection. Four size standards were employed for calibration: blue dextran (2000 kDa), Conalbumin (75 kDa), Carbonic Anhydrase (29 kDa) and Ribonuclease A (13.7 kDa) (Figure 4.1 and Table 4.2) The theoretical Stokes radius values for DFP proteins was calculated by the shell model with hydration layer 3.2 Å of HYDRONMR,⁷ with parameters corrected according to experimental conditions, starting from the x-ray structure of DFP1.

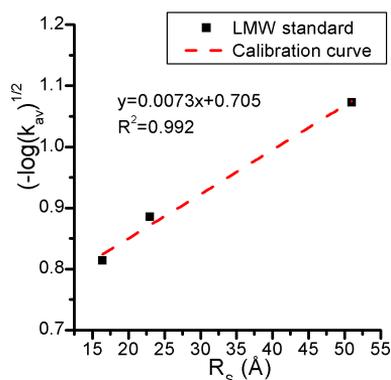


Figure 4.1. Calibration curve with standard proteins: ribonuclease A, carbonic anhydrase and conalbumin

Table 4.2. Stokes radius calibration curve parameters

	$v(-\log(K_{av}))$	RS (Å)
LMW standards		
Ribonuclease A	0.81	16.4
Carbonic Anhydrase	0.88	23
Conalbumin	1.07	51
DFP proteins		
apo -DFP2 (RT = 10.6min)	0.91	27 (21)
apo -DFP3 (RT = 11.0min)	0.86	21 (21)

Values in the parentheses are corresponding to the theoretical value calculated with HYDRONMR, starting from the x-ray structure of DFP1

4.2.5. 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, 500 and 300 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. Melting curves were performed at a total protein concentration of 10 μM in 5 mM HEPES 10mM NaCl pH 7, in a 0.1 cm cell.

Thermal denaturations were obtained by monitoring the CD signal at 222 nm as a function of temperature from 20 to 100 °C. The temperature was raised with a constant ramp of 0.2 °C min⁻¹. Points were collected every 2.5 °C, with a data averaging of 32 s and 5 nm bandwidth. The thermal unfolding in presence of 4M Gdn-HCl was performed under the same experimental conditions. The melting temperatures were determined considering a two-state transition of a monomer

Chapter 4

between folded and unfolded forms, with correcting the data for pre- and post-transition linear changes in ellipticity as a function of temperature.⁸

Chemical denaturation studies were performed at 10 μM protein concentration in 5 mM HEPES 10mM NaCl pH 7, at different Gdn-HCl concentration. Samples were left to equilibrate for 12 h before measurement. Gdn-HCl denaturation curves were obtained by monitoring the CD signal at 222 nm, with a data averaging of 32 s and 2 nm bandwidth. The thermodynamic parameters of the chemical denaturation were determined with the linear extrapolation method, considering a two-state transition of a monomer between folded and unfolded forms, with correcting the data for pre- and post-transition linear changes in ellipticity as a function of Gdn-HCl concentration.⁹

The CD signal in the UV region is expressed as mean residue ellipticities θ , 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 protein concentration in moles per liter, and n is the number of residues in the protein. In the Soret region, θ corresponds to the molar ellipticity, $\theta_{obs}/(10 \cdot l \cdot C \cdot n)$, in which θ_{obs} is the ellipticity measured in millidegrees, l is the path length of the cell in centimeters, C is the protein concentration in moles per liter

4.2.6. UV-Vis spectroscopy

UV-Vis measurements were performed using Cary 60 UV/vis spectrophotometer using quartz cuvettes of 1 cm. Wavelength scans were performed at 25 °C from 200 to 800 nm, with a 60 nm min⁻¹ scan speed.

2.2 μM of ZnP was solubilized in a 1.5 mL solution of 50 mM HEPES, 100 mM NaCl, pH 7 buffer by inclusion of 1% w/v octyl- β -D-glucopyranoside. 4 μL of a 75 μM stock of apo-DFP3 or di-Zn²⁺-DFP3 (0.09 μM aliquots) was titrated into the

Materials and Methods

1.5 mL solution containing the porphyrin, and an electronic absorption spectrum was measured until > 1.5 equivalents of protein were added. Absorbance changes at 423 nm, due to His-Zn coordination-induced spectral shifts of the porphyrin, were fit to a single-site, protein-ligand binding model.

The di-Co²⁺-DFP3 complex was prepared from a stock solution of apo-DFP3 in 50 mM HEPES buffer 100 mM NaCl pH 7.0 by adding 2 μ L of a solution of CoCl₂ (100 mM in H₂O).

The initial DFP3 concentration (500 μ M) was spectrophotometrically determined, by using $\epsilon(280\text{nm})= 22.460 \text{ cm}^{-1} \text{ M}^{-1}$.

Di-Zn²⁺-DFP3 and di-Zn²⁺-ZnP-DFP3 complexes were prepared in the same way as di-Co²⁺-DFP3 complex. However, the final concentration was 50 μ M of complex. DTBC and DTBQ, premixed in a separate vessel (in a 1:1 ratio, dissolved in dimethylformamide (DMF)), were added in ten molar equivalents to the protein solutions, and then quickly mixed by pipetting. For ZnDP experiments, DTBC was dissolved in DMF and added in ten molar equivalents to the di-zinc protein solution, and incubated overnight to air oxidize.

To study the ferroxidase reactivity, stock solutions of ferrous ammonium sulfate were prepared by dissolving the solid in 1% sulfuric acid and then diluted to a final concentration of 10mM (0.05% sulfuric acid). Reaction started upon addition of a stoichiometric amount (2 eq) of Fe²⁺ to a solution of the apo-protein (20 μ M) in aerobic conditions.

The oxidation of 4AP by preloaded protein diferric complexes were carried out in the presence of MPD, which led to the formation of the aminoindoaniline dye that allowed for spectrophotometric detection at the absorption maximum of 486 nm. The final solution contained 20 mM protein in 50 mM HEPES buffer 100 mM NaCl pH 7.0 10% DMF, 1 mM 4AP, 10 mM MPD.

Chapter 4

The effect of 4AP concentration on catalysis was tested in the range 0.1 mM – 3 mM. The experiments were performed always in presence of 10 mM MPD, and 10 mM diferric proteins in 50 mM HEPES buffer 100 mM NaCl pH 7.0 10% DMF. The formation of the aminoindole dye was followed at 528 nm at 25 °C and ϵ value of $10700 \text{ M}^{-1} \text{ cm}^{-1}$, taken from the literature (Figure 4.2).

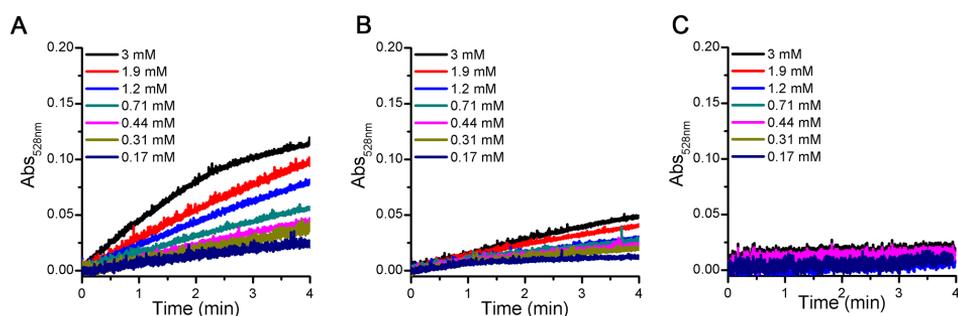


Figure 4.2. Dependency of the initial rate the phenol oxidation in function of substrate concentration. Kinetic curves of 4AP oxidation, at different substrate concentration, followed at 528 nm catalyzed by **A.** di-Fe³⁺-DFP3 and **B.** di-Fe³⁺-ZnP-DFP3, and **C.** in blank condition.

4.2.7. Synthesis and characterization of Fmoc-His-OH

Fmoc-His(trt)-OH (**R**, 0.620 g, 1 mmol) was dissolved in the side chain deprotection solution of 95% TFA, 2.5% TIS, 2.5% H₂O (v/v/v 10 mL). The reaction was carried out for one hour at 0 °C and one hour at room temperature, under a slight magnetic stirring. The TFA was removed by rotary evaporation to give an oily residue. The latter was dissolved in 10 mL of cold acetonitrile and the final product Fmoc-His-OH (**P**) was purified by crystallization.

The product was analyzed by analytical RP-HPLC, performed with a Shimadzu LC-10ADvp equipped with a SPDM10Avp diode-array detector. ESI-IT/TOF spectra were recorded on a Shimadzu LCMS-IT-TOF system with ESI interface and

Shimadzu LC-MS solution Workstation software for the data. The analysis was performed with a Vydac C18 column (2.1 mm x 100 mm; 5 μ m), eluted with an H₂O 0.05 % trifluoroacetic acid, TFA, (eluent A) and CH₃CN 0.05 % TFA (eluent B) in isocratic steps (30 % solvent B for 5 min, 50 % solvent B for 10 min) at 0.2 mL min⁻¹ flow rate. 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: 1.5 L min⁻¹).

The retention time of the starting **R** and the final **P** was 12.53 min and 2.65 min, respectively (Figure 4.3A). In the purified product chromatogram, we did observe only the peak of the latter (purity >92 %). Identity was ascertained by high resolution ESI-MS (Figure 4.3B): [P+H]⁺=378.15 Th (theoretical 378.145); [P+Na]⁺=400.129 Th (theoretical 400.127); [P-H⁺+2Na]⁺=422.108 Th (theoretical 422.108).

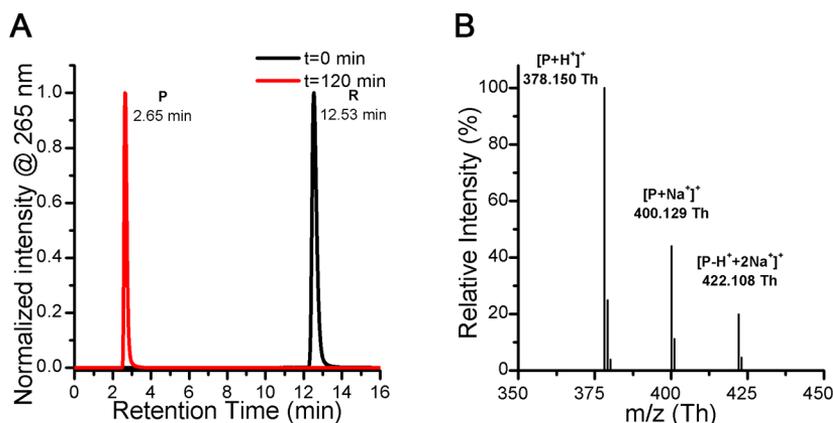


Figure 4.3. LC-MS characterization of Fmoc-His-OH A. RP-HPLC chromatogram (265 nm trace) of the Fmoc-His(trt)-OH (R) deprotection reaction in Fmoc-His-OH (P) at time 0 min (in black) and 120 min (in red). B. ESI-TOF spectrum relative to the purified **P**.

Chapter 4

ZnP was dissolved in DCM at a final concentration of 1mM. The maximum of the Soret band was at 415 nm, and we did not observe any peak in the CD spectrum, as expected. When Fmoc-His-OH and DIPEA were added at a final concentration of 2mM and 20mM, we observed a red shift of the Soret band from 415 nm to 423 nm, due to the coordination of the histidine to the zinc porphyrin. This shift was already observed upon coordination of DFP3 to ZnP solubilized in buffer in 1 % w/v octyl- β -D-glucopyranoside detergent. However, we did not observe any Cotton effect in the Fmoc-His-OH – ZnP complex.

4.2.8. MD simulations

The structure of ZnP was taken from the crystal structure, and beta hydrogens were added manually in PyMOL¹⁰. ZnP, without the coordinated zinc ion, was then parameterized using Antechamber and the general Amber force field (GAFF) with a net -2 charge to form a neutral complex with the zinc ion.¹¹ Structure preparation for the holo and apo states was then completed using the AmberTools18 program tleap. The protein was solvated in rectangular boxes with 1.5 nm padding of solvent on each side, and then solvent molecules were randomly replaced with ions to neutralize the system charge and reach 150 mM NaCl. The Amber ff14SB force field was used for the protein,¹² with the TIP3P water model,¹³ and Li/Merz divalent ion parameters for zinc and iron in the porphyrin and active site, respectively.¹⁴

Simulations were conducted using GPU-accelerated Amber18.¹⁵ Systems were minimized with 2000 kJ/mol nm² restraints on protein and ligand heavy atoms and zinc ions, then heated from 100 K to 310 K over 25 ps in the NVT ensemble with the same restraints. Temperature was controlled using a Langevin thermostat. Five rounds of NPT each 300 ps long were then run, lowering the

applied position restraints after each round: 2000, 1000, 400, 200, 20 kJ/mol nm². Final equilibration was run with no restraints for 1 ns. The Berendsen barostat was used to maintain pressure throughout NPT equilibration. Production was then run, using the Langevin thermostat with a Monte Carlo barostat, for 1 μ s using a 2 fs timestep with coordinates saved every 50 ps. A trial simulation of the porphyrin-bound state revealed that the zinc ion in the porphyrin distorted the protein structure by coordinating with the backbone oxygen of THR 152. A pair of distance restraints between the porphyrin zinc and the beta and carboxyl carbons of THR were applied to maintain normal structure of the protein backbone.

4.2.9. Docking and Minimization

X-ray structure of di-Zn²⁺-ZnP-DFP1 was removed of the ZnP from the PS domain and converted to the PDBQT format using AutoDockTools¹⁶. The docking grid parameters were chosen to include the porphyrin binding site within an 80 Å box. As a further restraint, the protein was considered as a rigid receptor, exception made for the charged residues facing between helix 1 – 4 (Arg22, Asp26, Arg149 and Glu143) and helix 2 – 3 (Arg62, Lys106 and Asp 113), which were considered as flexible. Torsional freedom was given only to the propionic groups for the ZnDP porphyrin ring. ZnDP was docked to DFP1 using AutoDock Vina with default parameters.¹⁷ Out of the seven binding poses identified, only three presented ZnDP zinc ion at coordinating distance from His61, and therefore considered for the subsequent minimization.

All minimizations were performed using NAMD2engine,¹⁸ with the CHARMM36m force field.¹⁹ For ZnDP, the standard Fe(II) protoporphyrin IX parameters were used as starting point. The vinyl groups were substituted with

Chapter 4

hydrogens (HGR51 atom type in cgenff),²⁰ with a partial charge of 0. Iron was substituted with Zn²⁺ ion already present in the force field, but keeping the same partial charge. The three complexes were hydrated within a pre-equilibrated octahedral box with an edge distance of 10 Å with the CHARMM-GUI interface.²¹ TIP3P water parameterization was used to describe the water molecules and the charge neutralized by sodium chloride at 150 mM concentration, placed with the Monte Carlo method. The periodic electrostatic interactions were computed using particle mesh Ewald summation with a grid spacing smaller than 1 Å. Subsequently, patches for the coordination of zinc by pyrrole nitrogens and histidine were inserted with VMD program, with the psfgen protocol.²² Water molecules and ions were firstly minimized in 50000 steps, keeping rigid all the complex. Subsequently, sidechains and ZnDP were minimized in 75000 steps with fixed backbone, and lastly 100000 steps minimized all the structure. The energies reported in Table 2.3 were calculated with VMD, excluding water and ion contributions.

4.2.10. Photochemical reaction

The reaction was performed in a 100 µL final volume reaction mixture containing 1 mM NADH and 10 µM photosensitizer. When used, ROS scavenger SOD and Cat were at 0.1 mg mL⁻¹ concentration. The solutions were placed in 96-well plate (Greiner PS Microplate, 96 Well, solid F-bottom (flat), chimney well with black sides) at 6 °C. The solutions were irradiated with Amber LEDs (producer TOP Led) with λ_{\max} at 615 nm. A single diode with an illuminance of $(8.5 \pm 0.3) 10^2$ lux, was placed above the solution in contact with the lid (Figure 4.4). UV-Vis spectra were collected in a TECAN Spark plate reader (Tecan Trading AG, Switzerland) between 280 and 480 nm, before and after 24 h irradiation.

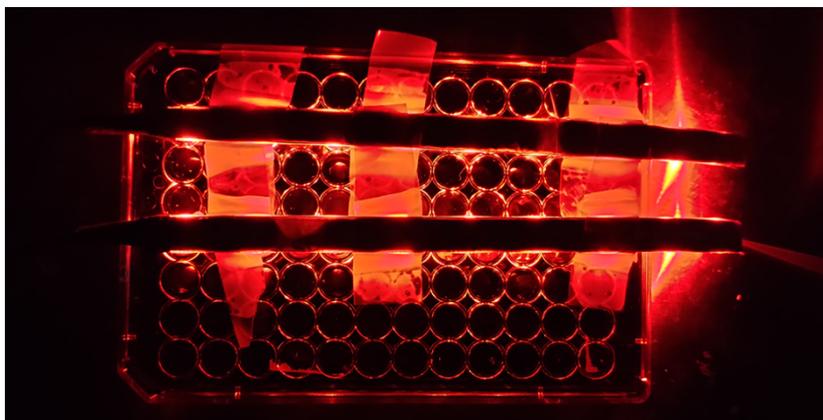


Figure 4.4. Experimental setup for NADH photooxidation. Solutions were placed only in the wells under LEDs diode: in the specific experiment line B and D; wells no. 1,3,5,7,10 e 12.

Chapter 4

References

1. Åqvist, J. A simple way to calculate the axis of an α -helix. *Comput. Chem.* **10**, 97–99 (1986).
2. Leaver-Fay, A. *et al.* Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. in *Methods in Enzymology* (eds. Johnson, M. L. & Brand, L.) vol. 487 545–574 (Academic Press, 2011).
3. Polizzi, N. F. *et al.* Photoinduced Electron Transfer Elicits a Change in the Static Dielectric Constant of a de Novo Designed Protein. *J. Am. Chem. Soc.* **138**, 2130–2133 (2016).
4. Kabsch, W. *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
5. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
6. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
7. García de la Torre, J., Huertas, M. L. & Carrasco, B. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J. Magn. Reson.* **147**, 138–146 (2000).
8. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–2535 (2006).
9. Greenfield, N. J. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat Protoc.* **1**, 2733–2741 (2007).
10. *The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.*
11. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

12. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
13. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
14. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **9**, 2733–2748 (2013).
15. Lee, T.-S. *et al.* GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **58**, 2043–2050 (2018).
16. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
17. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
18. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
19. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
20. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
21. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
22. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

List of abbreviations

4AP: 4-aminophenol

4BQM: 4-benzoquinone-monoimine

Abs: Absorbance

ACN: Acetonitrile

ACT: Aspartate kinase, Chorismate mutase and TyrA

Ala: Alanine

Arg: Arginine

Asn: Asparagine

AurF: p-Aminobenzoate N-oxygenase

BDIM: Birth, death and innovation model

Cat: Catalase

CCCP: Coiled Coil Crick Parametrization

CD: Circular Dichroism

cDNA: Complementary DNA

Cu_A: Copper A site

DAH7PS: 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase

DCM: Dichloromethane

DFHBI: (5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one

DF: Due Ferri

DF-C1: Due Ferri – Click1

DFP: Due Ferri Porphyrin

DMF: Dimethylformamide

List of abbreviations

DNA: Deoxyribonucleic acid

DTBC: 3,5-di-tert-butylcatechol

DTBQ: 3,5-di-tert-butylquinone

DTBSQ: 3,5-di-tert-butylsemiquinone

Gdn-HCl: Guanidine hydrochloride

GFP: Green Fluorescent Protein

Glu: Glutamic acid

Gly: Glycine

H₂O₂: hydrogen peroxide

His: Histidine

HRP: Horseradish Peroxidase

Im: Imidazole

IR: Infra Red

Leu: Leucine

LMCT: Ligand-to-metal charge transfer

Lys: Lysine

LOV: Light, Oxygen and Voltage

MDA: Multi domain architecture

MPP: Manganese diphenylporphyrin-binding protein

mRNA: messenger RNA

NAD⁺: Nicotinamide adenine triphosphate

NMR: Nuclear Magnetic Resonance

O₂^{-•}: superoxo radical anion

PS: Porphyrin-binding Sequence

RMSD: root-mean-square deviation

RNA: Ribonucleic acid

ROS: Reactive oxygen species

List of abbreviations

ROS Scav: Reactive oxygen species scavenger

Rt: Retention time

S₁: Singlet state

SASA: Solvent-accessible surface area

SEC: Size Exclusion Chromatography

SDA: Single domain architecture

SOD: Superoxo dismutase

T₁: Triplet state

Thr: Threonine

TOF: Turnover frequency

TON: Turnover number

Tyr: Tyrosine

USA: United States of America

UV-Vis: UV-Visible

Val: Valine

vdM: van del Mer

ZnDP: Zn-Deuteroporphyrin IX

ZnP: Zn-meso-(trifluoromethyl)porphin

ZnTPP: Zn-tetraphenylporphyrin