

Syracuse University

SURFACE at Syracuse University

Dissertations - ALL

SURFACE at Syracuse University

8-26-2022

Design of novel metalloenzymes and investigation of protein-metal interactions

Alona Kulesha

Syracuse University, alyona.kulesha@gmail.com

Follow this and additional works at: <https://surface.syr.edu/etd>



Part of the [Biochemistry Commons](#), [Biomedical Engineering and Bioengineering Commons](#), and the [Chemistry Commons](#)

Recommended Citation

Kulesha, Alona, "Design of novel metalloenzymes and investigation of protein-metal interactions" (2022). *Dissertations - ALL*. 1624.
<https://surface.syr.edu/etd/1624>

This Dissertation is brought to you for free and open access by the SURFACE at Syracuse University at SURFACE at Syracuse University. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.

Abstract

Metalloproteins are the proteins that utilize metals or metal complexes as cofactors and exhibit a wide range of functions from oxygen transport and regulation of transcription to hydroxylation of alkanes and water splitting, even though they employ a limited set of metals and metal-binding ligands. The investigation of metalloprotein amino acid sequence, structure and function relationships can uncover the principles behind natural metalloprotein design and opens the possibility of their implication as scaffolds for the design of novel catalysts and biomaterials. Protein engineering is a set of techniques that allows to install new functionalities into existing proteins or to design new proteins from scratch (*de novo* design) for various applications.

Chapter 1 of this dissertation will summarize the principles behind protein-metal interactions and metalloprotein design by discussing current methods in protein engineering and the successful examples of engineered metalloproteins.

The work presented in Chapter 2 provides insight into the extraordinary metal affinity to UFsc, a single-metal-binding *de novo* designed protein. The introduction of a single mutation in a two-metal-binding site of DFsc, a *de novo* designed protein, resulted in the elimination of one coordination site. The rearrangement of the active site altered protein's metal affinity to selected transition metals and resulted in the high affinity to zinc (II) ion with the dissociation constant in the picomolar range. We have shown that the fine tuning of the protein scaffold is important for effective metal binding in designed proteins.

Chapter 3 will discuss the development of new protein engineering method, NMR-guided directed evolution. Directed evolution is a protein engineering technique

used to identify proteins with improved functions by generating large libraries of mutants with further high-throughput screening in a functional assay. Our work was focused on establishing a method that will help to identify the positions for potential productive mutations to improve the output of directed evolution. We have found a strong correlation between the HSQC NMR chemical shift perturbations of backbone amide resonances of amino acid residues interacting with the substrate transition state analog and the probability of finding a beneficial mutation in the vicinity of those residues. The combination of three mutations identified by NMR-guided directed evolution resulted in the design of FerrEiCat, the most effective designed Kemp eliminase reported to date.

The design of a metalloprotein-based PET tracer, presented in Chapter 4, will provide an example of the application of metalloproteins for theranostic purposes. The introduction of one mutation in the calcium-binding site of AlleyCat7, a calmodulin-derived Kemp eliminase, completely eliminated protein's affinity to calcium (II) but preserved affinity to yttrium (III). This protein, fused to a liver-cancer-specific antibody, can be used to deliver metals to the cancer cells for PET imaging and radiotherapy.

Finally, the application of a minimalist approach to the design of amyloid-binding peptides that can reduce the infectivity of HIV virions will be discussed in Chapter 5. The amyloid fibrils, formed by the peptides derived from PAP and SEM proteins, have been found in human semen and shown to enhance HIV infection. We have designed a set of peptides to bind to these fibrils specifically and identified the promising candidates that differentiate between fibrillated and monomeric forms of amyloidogenic peptides. The identified peptides also showed sequence specificity. The developed peptides can potentially be used as microbicides for the prevention of HIV infection.

Design of novel metalloenzymes and
investigation of protein-metal interactions

by

Alona Kulesha

B.Sc., Kyiv National Taras Shevchenko University, 2015

M.Sc., Kyiv National Taras Shevchenko University, 2017

M. Phil., Syracuse University, 2019

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry

Syracuse University

August 2022

Copyright © Alona Kulesha, 2022

All Rights Reserved

Acknowledgements

First, I would like to thank my advisor, Dr. Ivan Korendovych, for giving me the opportunity to perform research in his lab. Working on many different projects under the guidance of Dr. Korendovych, I was forced to get out of my comfort zone and learn to think fast while finding solutions to various problems. This experience allowed me to gain many valuable transferable skills and successfully apply them in problem-solving, also outside my research. Dr. Korendovych always encouraged me to explore and move forward, and I am thankful for that. I also would like to thank Dr. Olga Makhlynets for her support and the opportunity to contribute to the projects where I could share my knowledge and gain new experience.

I would like to thank Dr. Liviu Movileanu, Dr. Timothy Korter and Dr. Mathew Maye for expressing interest and finding time to join my examination committee. Special thanks to Dr. Robert Doyle and Dr. Carlos Castaneda, my departmental committee, for their support and encouragement throughout my PhD journey.

During the past five years, I also had an opportunity to work as a teaching assistant in the Chemistry Department. I am grateful for the support and guidance provided by all instructors I worked with but especially I would like to thank Prof. Jonathan French for his trust and confidence in me. I greatly appreciate all discussions, feedback sessions and trainings organized by Prof. French as they helped me to learn how to teach. I am grateful for your support over the years and the Syracuse University Outstanding Teaching Assistant Award I received is definitely your accomplishment too.

I am extremely grateful to all my lab mates for their support, mentoring and encouragement throughout these years. I would like to thank the former members, Dr.

Michelle Takacs, Dr. Zsofia Lengyel-Zhand, Dr. Martin Dolan, Dr. Oleksii Zozulia, Dr. Areetha D'Souza, Dr. Megha Jayachandran, Dr. Inhye Kim, the current members, Dr. Liam Marshall, Sagar Bhattacharya, Dona Imanga Edirisinghe and Cijun Zhang, and all undergraduate and summer students of Korendovych and Makhlynets labs for sharing their knowledge with me, supporting my personal and professional growth, and allowing me to be a part of their professional lives. Working with all of you provided me with a valuable experience I will carry through life. I appreciate all moments, including boba tea trips, long group meetings and even disagreements we had in the lab, as those moments shaped me into a person I am today. Thank you for your patience while I was learning, for your support in times I failed, and for your honest feedback when I was wrong.

There are no words to convey my deep gratitude to my parents, Viktor and Tamara, for their love, support and all sacrifices they made to give me a chance to be where I am today. My sincere thanks to all other family members, especially my siblings, Dariya and Ivan, my aunt, Tetyana, and cousin, Yana, for believing in me and always cheering for my accomplishments. Having a first-generation graduate student in the family was confusing for all of us but I am grateful for your attempts to figure out the best ways to support me during this journey and learn about my work.

I would like to express my appreciation to my friend, my best roommate and my loving husband, Tony. Thank you for being supportive of my career goals and encouraging me to move forward. Thank you for taking care of me and bearing more responsibilities in this relationship than you should have. Thank you for challenging me to be the best version of myself. You keep me young and that's how I wanna be.

Special thanks go to all my friends outside the lab. I want to express my gratitude to my former roommates, Dariya and Kaira, for always being there for me, and to my chemistry friends, Mariah, Kylie and Sadie, for always checking in and making sure I stay social. I cherish every moment of our friendship and will make sure that it continues after we all become PhDs. I also would like to thank the Women in Science and Engineering (WISE) Program which gave me a chance to meet many great people and build life-long friendships. I greatly appreciate the understanding and support I receive from all my friends in Ukraine, especially Olha and Tetyana. Thank you for being there for me through my academic journey.

Five years ago, my decision to pursue a doctoral degree was majorly supported by my manager and friend Dr. Yurii Moroz. I greatly appreciate all our discussions that fueled my interest in scientific research, as well as the guidance, advice and help I received from you during my time in Bienta.

Last but not least, I want to thank all our collaborators from the Chemistry Department, the Syracuse Biomaterials Institute and outside Syracuse University for their contributions to our projects, assistance in experiments and allowing me to use their facilities. Also, thank you to the administration of the Chemistry Department for the support and assistance during these five years.

Table of Contents

List of Figures	xii
List of Tables	xxiii
List of Abbreviations	xxv
Chapter 1. Metalloproteins and Protein Engineering	1
1.1. Introduction	2
1.2. Natural metalloenzymes.....	3
1.3. Protein-metal interactions	6
1.3.1. <i>Metals in biological systems</i>	7
1.3.2. <i>Metal coordination in proteins: ligand aspects</i>	10
1.3.3. <i>Metal coordination in proteins: metal aspects</i>	13
1.4. Methods in protein engineering and design of metalloproteins	16
1.4.1. <i>Rational design</i>	17
1.4.2. <i>Directed evolution</i>	20
1.4.3. <i>Semi-rational design</i>	23
1.5. References.....	26
Chapter 2. The origins of tight metal binding to UFsc, a <i>de novo</i> designed protein that has extraordinary affinity for metal ions	41
2.1. Introduction	42
2.2. Results and Discussion.....	46
2.2.1 <i>Design and characterization of metal affinity of UFsc, a single-metal binding protein</i>	46
2.2.2 <i>The importance of His77 for metal binding by 4G-UFsc</i>	52

2.2.3. Alteration of the bridging Glu74 leads to the decrease in metal binding affinity	57
2.2.4. Mutations of the non-bridging glutamate ligands have a negative effect on metal binding	62
2.2.5. Copper coordination by UFsc	70
2.3. Conclusions	73
2.4. Experimental	74
2.5. References.....	81
Chapter 3. Design of the efficient Kemp eliminase using NMR-guided directed evolution	88
3.1. Introduction	89
3.2. Results and Discussion.....	92
3.2.1. Method development for NMR-guided directed evolution	92
3.2.2. The catalytic activity of enzymes designed using NMR-guided directed evolution	96
3.2.3. FerrEICat, the most active designed Kemp eliminase.....	100
3.2.4. NMR-guided directed evolution allowed to improve affinity of AlleyCat to the inhibitor	106
3.3. Conclusions	111
3.4. Experimental	112
3.5. References.....	124
Chapter 4. Developing a Probe for Noninvasive Detection of Hepatocellular Carcinoma by Positron Emission Tomography	130

4.1. Introduction	131
4.1.1. <i>Detection of liver cancer and positron emission tomography</i>	131
4.1.2. <i>Metal-based PET tracers</i>	133
4.1.3. <i>Antibody-based delivery system for radiometals</i>	134
4.2. Results and Discussion.....	138
4.2.1. <i>Design of yttrium-binding protein</i>	138
4.2.2. <i>Determination of the binding affinity of HolIEE to rare earth metals</i>	142
4.2.3. <i>Development of 2E10-HolIEE radiotracer</i>	149
4.3. Conclusions and Future Work.....	158
4.4. Experimental	159
4.5. References.....	170
Chapter 5. Design of peptides to block association of HIV virus with human	
cells	177
5.1. Introduction	178
5.1.1 <i>HIV and AIDS</i>	178
5.1.2. <i>Amyloid fibrils promote HIV infection</i>	180
5.2. Results and Discussion.....	185
5.2.1. <i>Design of amyloid binding peptides</i>	185
5.2.2. <i>Designed peptides selectively bind cationic amyloid fibrils</i>	187
5.2.3. <i>FEV as a promising scaffold for the HIV microbicide development</i>	193
5.3. Conclusions and Future Work.....	199
5.4. Experimental	200
5.5. References.....	206

Appendix I.....	212
Appendix II.....	225
Appendix III.....	232
Curriculum Vitae.....	233

List of Figures

- Figure 1.1. Metal-binding proteins are found among all classes of enzymes.** The metals utilized by selected enzymes are presented as colored spheres. The structures of the enzymes were visualized in Pymol..... 4
- Figure 1.2. Elements used as cofactors by natural metalloenzymes.** The figure is reproduced from ref. 4 with the permission from Springer Nature, copyright 2009. 8
- Figure 1.3. The side chains of common amino acids used for metal coordination in proteins.** The amino acids are labeled with the three letter code, where Glu is glutamic acid, Asp is aspartic acid, His is histidine, Met is methionine, Cys is cysteine.12
- Figure 1.4. Metal coordination modes of carboxylate ligands.** The metal ion is depicted as a grey sphere. 13
- Figure 2.1. DFsc is a four-helix bundle with a divalent metal binding site (left).** The metal binding site consists of two histidine and four glutamic acid residues (right). Images generated based on the protein NMR structure (PDB ID 2HZ8)..... 43
- Figure 2.2. UFsc and 4G-UFsc were created by introducing E104H mutation in the metal binding site (right).** 4G-UFsc has wider channel (left) due to additional alanine-to-glycine mutations (DFsc PDB ID 2HZ8 was used as scaffold for mutagenesis using PyMol) 45
- Figure 2.3. Metal binding sites of DFsc (A) and UFsc (B).** Cobalt titration experiment showing the metal-to-protein ratio in DFsc (C) and UFsc (D). The 130 μM DFsc (C) and 150 μM UFsc (D) were titrated with aliquots of 1 mM cobalt (II) chloride solution. The solutions were prepared in 25 mM HEPES (pH 7.6), 100 mM NaCl and incubated at 25 $^{\circ}\text{C}$ 47

Figure 2.4. The introduction of four Ala-to-Gly mutations widens the channel leading to the active site (PDB 2HZ8 was used as a scaffold for mutagenesis).... 48

Figure 2.5. A. The structure of triethylenetetramine (TETA). The potential ligands for metal coordination are highlighted in blue. **B.** Proposed coordination configuration in TETA-zinc (II) complex. **C and D.** Baseline- and offset-corrected ITC titration curve and binding isotherm, respectively, for the titration of 200 μM zinc (II) with 2 mM TETA at pH 7.6. The dissociation constant for the TETA-zinc (II) complex was used for further determination of protein-zinc (II) dissociation constants in competition titration experiments..... 50

Figure 2.6. Competition titration of 4G-UFsc and Mag-Fura 2 with ZnCl_2 . (A) Absorbance spectra of Mag-Fura 2 (18.7 μM) and 4G-UFsc (10.1 μM) titrated with ZnCl_2 . The coordination of metal reduces absorbance intensity of the dye at 366 nm. **(B)** Absorbance at 366 nm as a function of metal concentration was fit using DYNAFIT to determine dissociation constant of zinc (II)-4G-UFsc complex..... 52

Figure 2.7. The structure of UFsc metal binding site. D73 and D103 (shown in green) provide stabilization and ensure proper orientation of H77 and H107, respectively. The grey spheres represent the locations of possible metal-binding sites in UFsc. The figure was prepared using PDB 2HZ8 as a scaffold. 53

Figure 2.8. CD spectroscopy analysis of 4G-UFsc H107Q (A) and H77Q (B) showed that the mutation did not influence the degree of protein helicity. Proteins were analyzed at 20 μM in 5 mM HEPES, 20 mM NaCl (pH 7.6). Protein-metal complex was prepared by mixing 20 μM protein with 200 μM zinc (II) chloride. 54

Figure 2.9. ITC traces for the titration of H107Q (A) and H77Q (C) with cobalt (II) and the binding isotherms (B and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 25 mM HEPES, 100 mM NaCl, pH 7.6 at 25 °C..... 56

Figure 2.10. The mutations introduced at position 74 allowed to test the importance of the size, charge and identity of this ligand. 58

Figure 2.11. CD spectroscopy analysis showed that mutations at the position 74 (A – E74D, B – E74Q, C – E74H) have different effect on protein structure. All apoprotein spectra were recorded at 20 µM and zinc-bound spectra were recorded at 20 µM protein with 200 µM zinc (II) chloride concentration in 5 mM HEPES, 20 mM NaCl (pH 7.6) at room temperature..... 58

Figure 2.12. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the competition titration of zinc (II) – triethylenetetramine (TETA) complex with Glu74 mutants. The complex of 90 µM zinc (II) with 200 µM TETA was titrated with: A – 1 mM 4G-UFsc E74D (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 210 s); C – 1.04 mM 4G-UFsc E74Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s); E – 1.34 mM 4G-UFsc (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 210 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively. 60

Figure 2.13. The structure of UFsc metal binding site. Y51 and Y18 (shown in green) provide stabilization and ensure proper orientation of E11 and E44,

respectively. The grey spheres represent the locations of possible metal binding sites in UFsc. The figure was prepared using PDB 2HZ8 as a scaffold..... 62

Figure 2.14. CD spectroscopy analysis of 4G-UFsc mutants showed an increase in the degree of helicity upon metal binding. The solution of 20 μM protein in 5 mM HEPES, 20 mM NaCl (pH 7.6) was used as an apoprotein sample. The metal-bound form was prepared by addition of 200 μM zinc (II) chloride..... 63

Figure 2.15. ITC traces for the titration of 4G-UFsc (A) and 4G-UFsc E11Q (C) with manganese (II) and the binding isotherms (B and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 25 mM HEPES, 100 mM NaCl, pH 7.6 at 25 $^{\circ}\text{C}$ 66

Figure 2.16. CD spectroscopy analysis of UFsc mutants (A – E44Q, B – E44L/Y18F) showed an increase in the degree of helicity upon metal binding. The solution of 20 μM protein in 5 mM HEPES, 20 mM NaCl (pH 7.6) was used as an apoprotein sample. The metal-bound form was prepared by addition of 200 μM zinc (II) chloride..... 66

Figure 2.17. Molar extinction coefficients of Co(II)-bound proteins UFsc (A) and 4G-UFsc (B) determined from the absorbance spectra. Samples contained 150 μM protein in buffer (25 mM HEPES, 100 mM NaCl, pH 7.6) with 300 μM CoCl_2 . The spectra were corrected for the absorbance at 700 nm and shown with the molar extinction coefficient on Y axis. 68

Figure 2.18. Competition titration of UFsc and Mag-Fura 2 with CuCl_2 . (A) Absorbance spectra of Mag-Fura 2 (17.6 μM) and UFsc (9.7 μM) titrated with CuCl_2 . The coordination of metal reduces absorbance intensity of the dye at 366 nm. (B)

Absorbance at 366 nm as a function of metal concentration was fit using DYNAFIT to determine dissociation constant of Cu(II)-UFsc complex. 70

Figure 2.19. NO production by UFsc-Cu(II) complex at pH 7.6. The traces show two experiments where 400 μ M copper (II) (dashed lines) or 200 μ M UFsc with 400 μ M copper (II) (solid lines) were incubated with 20 mM sodium nitrite for 4H at 25 $^{\circ}$ C in 25 mM HEPES (pH 7.6) with 100 mM NaCl. The presence of produced nitric oxide is correlated with the amount of $[\text{Fe}(\text{NO})\text{EDTA}]^{2-}$ complex, which can be detected at 432 nm. 72

Figure 3.1. The reaction schemes for Kemp elimination promoted by (A) acid-base or (B) redox-mediated mechanisms. 93

Figure 3.2. The structures of Kemp substrate, 5-nitrobenzoxazole, and substrate analog, 6-nitrobenzotriazole, used for NMR experiments. 94

Figure 3.3. A. The positions for productive mutations were identified using a Z score. Out of 18 tested positions with the Z score >1 , all produced productive mutations. **B.** The identified hot spots mapped on the structure of myoglobin with the active site highlighted with the yellow sphere (PDB code 6G5T). 96

Figure 3.4. A. The kinetic traces for Kemp elimination reaction catalyzed by Mb-H64G/V68A with (black) and without (red) SAC buffer. **B.** The spectra collected for the kinetic analysis of Mb-H64G/V68A with (black) and without (red) SAC buffer..... 98

Figure 3.5. Background Kemp elimination rates in 20 mM Tris, pH 8.0 (black) and 20 mM Tris, pH 8.0 with 1 mM ascorbate, 0.1 μ M SOD, and 20 nM catalase (red). Presented traces are the average of three runs. 98

Figure 3.6. A. The kinetic traces for Kemp elimination reaction catalyzed by Mb-H64G/V68A with oxygenated (black) or deoxygenated (red) substrate solution. **B.** The spectra collected for the kinetic analysis of Mb-H64G/V68A with oxygenated (black) or deoxygenated (red) substrate solution. 99

Figure 3.7. The structure of the active site in myoglobin upon mutagenesis at position 64. H64 (left, PDB 1MBN) facilitates oxygen binding, V64 (center, PDB 6G5T) reduced oxygen-binding affinity, G64 (right, the structure was determined as part of this project and deposited to wwPDB) opens the active site for better substrate access. Iron ion in heme is presented as an orange sphere. 102

Figure 3.8. A. Michaelis-Menten traces for Mb-H64V and derived mutants showing the improvement in activity after the introduction of three mutations. **B.** The overlay of Mb-H64V (PDB 6CF0) and FerrEiCat (the structure was determined by Dr. Jeremy Tame's group and deposited to the wwPDB) showing the minor structural changes after mutagenesis..... 103

Figure 3.9. Spectroelectrochemical determination of reduction potentials of selected myoglobin mutants. The proteins were analyzed in 20 mM TRIS-HCl (pH 8.0) at 20°C in presence of the mediator (100 μM phenazine sulfate). The determined reduction potentials (vs Ag/AgCl) for one-electron transfer were as follows: -117.06 ± 2.40 mV (Mb-H64V), -163.81 ± 0.43 mV (Mb-H64G/V68A), -117.04 ± 0.94 mV (FerrEiCat). 104

Figure 3.10. A. The absorbance spectrum for zinc-porphyrin-substituted FerrEiCat. **B.** Michaelis-Menten kinetic trace for the Kemp elimination reaction catalyzed by FerrEiCat-Zn-P at pH 8.0. 105

Figure 3.11. pH dependence of FerrEiCat's activity in Kemp Elimination. Protein was analyzed at 5 nM with 35 μ M substrate in 1 mM L-ascorbic acid, 0.1 μ M SOD, 20 nM catalase, 1.5% acetonitrile, 20 mM Tris (pH 8.0).....	106
Figure 3.12. The mutations identified in 10 rounds of directed evolution of AlleyCat scaffold. The F92E mutation is colored in yellow, the mutations identified with epPCR and gene shuffling are in green, the positions identified using NMR-guided directed evolution are highlighted in red. The calcium ions are presented as grey spheres. The mutations were mapped on AlleyCat scaffold (PDB 2KZ2).	108
Figure 4.1. Distribution of cases and deaths for the top 10 most common cancers in 2020. Reproduced from ref 1 with the permission from American Cancer Society, copyright 2021.....	131
Figure 4.2. A. The structure of 18F-fluorodeoxyglucose (18F-FDG). B. The uptake of FDG by cells (reproduced from ref. 9 with permission from Elsevier Ireland Ltd., copyright 2003).	133
Figure 4.3. The schematic representation of different types of engineered antibodies. scFv stands for single-chain variable domain. The antibody structure was visualized by Pymol (PDB 1IGY).....	136
Figure 4.4. The structure of the EF-hand, a metal binding site present in AlleyCat7. The EF-hand consists of 12 canonical amino acids. The protein structure was prepared using Pymol and PDB 2KZ2.	140
Figure 4.5. Kemp elimination activity of HolIEE. The protein's activity was tested at 1 μ M in the buffer containing 20 mM MOPS (pH 7.0), 100 mM NaCl and 80 μ M EDTA, Ca	

(II) or Y (III)). The catalytic efficiency of HolIEE in the presence of yttrium was measured as $161 \pm 14 \text{ M}^{-1}\text{s}^{-1}$ 141

Figure 4.6. Circular dichroism spectra for the analysis of HolIEE upon metal binding. All samples were analyzed using 20 μM protein and 80 μM metal (Ca(II) or Y(III)) in 30 mM MOPS (pH 6.5), 100 mM KCl at 25°C..... 142

Figure 4.7. ITC traces for the titration of 198 μM AlleCat7 (A) and 195 μM HolIEE (B) with 3 mM calcium (II) and the binding isotherms (C and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 100 mM MES, 100 mM KCl, pH 6.5 at 25 °C. 144

Figure 4.8. ITC traces for the titration of 198 μM AlleCat7 (A) and 195 μM HolIEE (B) with 4 mM yttrium (III) chloride and the binding isotherms (C and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 100 mM MES, 100 mM KCl, pH 6.5 at 25 °C. The photos attached to panels A and B show the precipitation of HolIEE upon addition of yttrium (B)..... 145

Figure 4.9. The model of HolIEE created using the structure of AlleyCat as a scaffold (PDB 2KZ2). A. The view from the side of substrate binding site. B. The view from the side opposite to the binding site. The calculations were performed using APBS electrostatics plugin in Pymol for solvent excluded surface..... 146

Figure 4.10. A. CD spectra for the titration of 10 mM EDDS with the aliquots of 100 mM yttrium (III) chloride. **B.** The determination of metal binding stoichiometry and the concentrations necessary to prepare one-to-one EDDS:Y(III) complex. 148

Figure 4.11. A. CD spectra of the titration of 10 μ M HolIEE with yttrium (III) in yttrium-EDDS. **B.** The plot presenting MRE values at 220 nm plotted versus the concentration of free yttrium (III) did not show any dependence. 148

Figure 4.12. SDS-PAGE analysis of the fractions from test protein expression of 2E10-eGFP and 2E10-HolIEE in Origami 2 E. coli cells after IPTG induction in LB medium. 150

Figure 4.13. SDS-PAGE analysis of the fractions from test protein expression of 2E10-eGFP and 2E10-HolIEE in Rosetta pLysS E. coli cells after IPTG induction in LB medium...... 151

Figure 4.14. SDS-PAGE analysis of the fractions from test protein expression of 2E10 and 2E10-HolIEE in Rosetta pLysS E. coli cells in autoinduction medium. 152

Figure 4.15. SDS-PAGE analysis of the elution fractions which contain 2E10-HolIEE after purification from inclusion bodies. 153

Figure 4.16. SDS-PAGE analysis of the medium samples collected during the test secreted protein expression of 2E10 and 2E10-eGFP in Pichia pastoris. 154

Figure 4.17. SDS-PAGE analysis of the yeast cell extracts collected during the test secreted protein expression of 2E10 and 2E10-eGFP in Pichia pastoris...... 155

Figure 4.18. Photos of the insect cells infected with the viruses containing 2E10, 2E10-HolIEE and 2E10-eGFP coding regions. Sample containing 2E10-eGFP emitted green light indicating functional eGFP...... 156

Figure 4.19. SDS-PAGE analysis of flow-through and elution fractions collected during test purification of 2E10, 2E10-HolIEE and 2E10-eGFP from infected insect cells. 157

Figure 5.1. HIV virion life cycle. The figure is reproduced from ref. 3 with the permission from Elsevier Ltd., copyright, 2008..... 179

Figure 5.2. The SEVI amyloid fibrils are formed by the peptides derived from the prostatic acid phosphatase (PAP). The structure of PAP is modeled based on PDB 1CVI and the structure of PAP246-286 is from PDB 2L3H..... 181

Figure 5.3. The schematic representation of the activity of designed peptides. Upon binding to the cationic amyloid fibrils, the designed peptides will block the association of fibrils with HIV virions and reduce the rate of HIV infection. 185

Figure 5.4. The fluorescence spectra for the reaction containing dansylated GEV peptide and amyloid fibrils. The spectra were collected right after the mixing (A-C) and after the 24-hour incubation of the reactions at room temperature (D-F). 189

Figure 5.5. The fluorescence spectra for the reaction containing dansylated ILE peptide and amyloid fibrils. The spectra were collected right after the mixing (A-C) and after the 24-hour incubation of the reactions at room temperature (D-F). 190

Figure 5.6. The fluorescence spectra for the reaction containing dansylated ILE and amyloid fibrils before (A, zoomed Figure 5.5. B) and after (B) centrifugation. The binding of positively charged ILR (C) to the amyloid peptides showed no interaction. The fluorescence spectra were recorded at 25° in PBS buffer (pH 7.3). 191

Figure 5.7. The titration of fibrillated SEM1 (5 μM) with dansylated ILE peptide at pH 7.3. The fluorescence at 510 nm was recorded and lotted versus the peptide/fibril ratio. 191

Figure 5.8. The fluorescence spectra for the reaction containing dansylated FEV peptide and amyloid fibrils. The spectra were collected right after the mixing (A-C) and after the 24-hour incubation of the reactions at room temperature (D-F). 193

Figure 5.9. The test of FEV cytotoxicity towards HEK-293 mammalian cells. The viability of the cells treated with the peptide was normalized to the untreated cells. ... 194

Figure 5.10. A logo plot (a graphical representation of an amino acid multiple sequence alignment) for the peptides identified from the phage display library after the three rounds of panning. The plot was created using WebLogo.⁴¹ The height of single-letter amino acid notation denotes the relative frequency of each amino acid in that position..... 197

Figure 5.11. The fluorescence spectra for the reaction containing dansylated peptides, identified from the phage display library, and amyloid fibrils. 198

List of Tables

Table 1.1. Metals found in metalloproteins can be classified as hard or soft acids. The amino acids provide ligands for metal coordination that are classified as hard or soft bases.....	11
Table 1.2. Coordination geometries observed for selected transition metals found in metalloproteins. The number of ligands is specified in parentheses and the most common geometry is in bold.	15
Table 2.1. Dissociation constants (in nM) determined by ITC for metal complexes of DF family proteins.	49
Table 2.2. Thermodynamic parameters of divalent metal binding to 4G-UFsc H107Q and H77Q mutants determined by ITC.....	55
Table 2.3. Thermodynamic parameters of zinc binding to E74 mutants determined by ITC competition titration with triethylenetetramine.....	59
Table 2.4. Thermodynamic parameters of divalent metal binding to E74 mutants determined by ITC.....	61
Table 2.5. Thermodynamic parameters of divalent metal binding to 4G-UFsc E44Q, E44L/Y18F and E11Q mutants determined by ITC (zinc was tested in the competition experiment with TETA).....	64
Table 2.6. Thermodynamic parameters of divalent metal binding to UFsc E44Q and E44L/Y18F mutants determined by ITC (zinc was tested in the competition experiment with TETA).	67
Table 2.7. Dissociation constants for 4G-UFsc and its mutants determined using competition titration assay with Mag-Fura 2.	69

Table 3.1. Kinetic parameters for the Kemp elimination reaction catalyzed by myoglobin mutants. Experiments were conducted under anaerobic and reducing conditions at pH 8.0 (25°C).....	101
Table 3.2. Thermodynamic parameters of 6-nitrobenzotriazole binding to AlleyCat mutants determined by ITC. The titrations were performed in 20 mM HEPES, 100 mM NaCl, 10 mM CaCl ₂ (pH 7.0), 2% acetonitrile at 25°C. K _d stands for the dissociation constant in micromolar, ΔH is the change in enthalpy in kcal/mol.	109
Table 5.1. The sequences of peptides found in human semen and shown to form cationic amyloid fibrils that promote HIV infection. The residues that possess positive charge at physiological pH are colored in blue and those that have negative charge are shown in red.....	186
Table 5.2. Peptides used in this study and their theoretical pI values calculated using ExPASy ProtParam. The positively charged residues are highlighted in blue and the negatively charged residues are colored in red.....	187
Table 5.3. Amino acid sequences obtained after three rounds of panning using FEV-based phage library.....	196

List of Abbreviations

AIDS	Acquired immunodeficiency syndrome
APBS	Adaptive Poisson-Boltzmann Solver
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
CD	Circular dichroism
CSP	Chemical shift perturbation
CT	Computer tomography
CYP	Cytochrome
DF	Due Ferri
DIEA	N,N-diisopropylethylamine
DMEM	Dulbecco's modified eagle's medium
DMF	N,N-dimethylformamide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DOTA	2-[4,7,10-tris(carboxymethyl)-1,4,7,10-tetrazacyclododec-1-yl]acetic acid
DTT	Dithiothreitol
EDDS	Ethylenediamine-N,N'-disuccinic acid
EDTA	Ethylenediaminetetraacetic acid
FDA	Food and Drug Administration
FBS	Fetal bovine serum
FDG	Fluorodeoxyglucose

GFP	Green fluorescent protein
GPC3	Glypican-3
HCC	Hepatocellular carcinoma
HCTU	O-(6-chlorobenzotriazol-1-yl)-N,N,N',N'- tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HSAB	Hard and soft acid and base
HSQC	Heteronuclear Single-Quantum Correlation Spectroscopy
ICP-OES	Inductively coupled plasma – optical emission spectrometry
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Luria broth
MALDI -TOF	Matrix assisted laser desorption ionization – time of flight mass spectrometry
MES	2-ethanesulfonic acid
MOPS	3-morpholinopropane-1-sulfonic acid
MRE	Molar residue ellipticity
MRI	Magnetic resonance imaging
MWCO	Molecular weight cut off
NMR	Nuclear magnetic resonance
NTA	Nitriloacetic acid

PAP	Prostate acid phosphatase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Poly(ethylene glycol)
PES	Polyethersulfone
PET	Positron emission tomography
PIPES	1,4-Piperazinediethanesulfonic acid
PMSF	Phenylmethylsulfonyl fluoride
PP	Polypropylene
QM	Quantum mechanics
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEM	Semenogelin
SEVI	Semen-derived enhancer of viral infection
SOC	Super Optimal broth with Catabolite repression
SOD	Superoxide dismutase
SOE-PCR	Splicing by overlap extension polymerase chain reaction
SQ	Semiquinone
TETA	Triethylenetetramine
TEV	Tobacco etch virus

TFA	2,2,2-Trifluoroacetic acid
TIS	Triisopropylsilane
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
UF	Uno Ferro
UNAIDS	The Joint United Nations Programme on HIV/AIDS

Chapter 1. Metalloproteins and Protein Engineering

Abstract. Protein engineering is a process of utilizing a set of tools and strategies to develop novel proteins or improve the properties of existing proteins. Metalloproteins, proteins that use metals as cofactors, exhibit a wide range of functions from oxygen transport and regulation of transcription to hydroxylation of alkanes and water splitting. Even though metalloproteins have multiple activities, they employ a limited set of metals and metal-binding motifs for their functions. The investigation of protein-metal interactions provides essential information necessary for the successful metalloprotein engineering. This chapter will summarize the principles behind protein-metal interactions as well as the approaches used for successful design of metalloproteins.

1.1. Introduction

Proteins are a large group of biomolecules with a wide range of structures and properties. Proteins fulfill diverse functions starting from the structural support of cells to complex catalysis. The investigation of protein sequence, structure and function relationships opens a possibility to the implication of proteins as scaffolds for the design of novel biomaterials. Engineered proteins have been employed in many areas of modern life, such as green organic synthesis, food, pharmaceutical, biotechnology industries and more.¹⁻³ Proteins are linear biopolymers that consist of amino acids and folded into dynamic three-dimensional structures to accomplish a variety of functions. It has been estimated that from one-third to half of the identified natural proteins use metal ions as cofactors.⁴ Metalloproteins engage metal ions for proper structural folding, metal storage, transport, and catalysis. The investigation of the structure/function relationship in natural metalloenzymes provides an important information about enzymatic catalysis which allows to better understand the mechanisms of chemical reactions catalyzed by proteins, as well as provides a background for potential engineering of enzymes with desired activities.

To accommodate the variety of functions, natural metalloenzymes often have a complex multidomain structure, they also require accessory proteins for proper folding and certain secondary structure for metal binding and activity. Due to their complexity, natural metalloenzymes are challenging to study and simpler models that mimic the active units of natural metalloenzymes can provide an opportunity for more effective research. Protein engineering is a process of the installation of novel or improved characteristics into existing proteins or the design of completely new proteins for

specific functions. The tools of protein engineering can be used to design simple models of natural metalloproteins and study their properties for a better understanding of biocatalysis, as well as to use our knowledge of protein-metal interactions to create new catalysts and metal-binding proteins with predicted functions.

1.2. Natural metalloenzymes

Metalloproteins are proteins that use metals, metal organic complexes, like heme, or metal inorganic complexes, as iron-sulfur clusters, as cofactors to obtain specific characteristics. As the characterization of various proteins continues, it is possible that more metalloproteins will be identified in the future. Metalloproteins perform a variety of functions from oxygen transport, regulation of transcription and metal transport to hydroxylation of alkanes, water splitting and electron transfer. The wide range of functions makes metalloproteins targets for extensive structure-function relationship investigation and promising scaffolds for protein engineering.

Metal-binding proteins exhibit a variety of different activities and examples have been found in all six classes of enzymes. Superoxide dismutases (SODs) are a group of metal-containing enzymes which catalyze the dismutation of superoxide anion (O_2^-) to hydrogen peroxide (H_2O_2) and oxygen (O_2).⁵ SODs are classified as oxidoreductases, meaning that they facilitate the transfer of electrons between electron donor and electron acceptor. SODs are found in many living organisms, have different intracellular and organ distribution, and differ in protein structure resulting in multiple isoforms but all require metal ions for their activity. SODs utilize copper, zinc, manganese and iron metal ions to perform catalysis and stabilize the structure of the active site (Figure 1.1).⁶

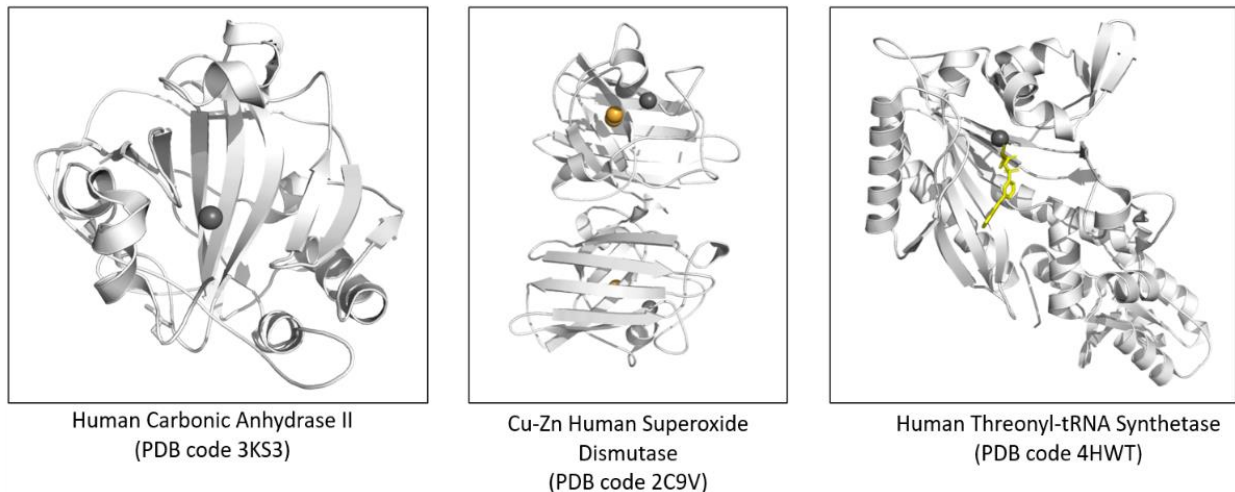


Figure 1.1. Metal-binding proteins are found among all classes of enzymes. The metals utilized by selected enzymes are presented as colored spheres. The structures of the enzymes were visualized in Pymol.

Polymerases, enzymes that are involved in DNA or RNA synthesis, are examples of metal-ion-dependent transferases. Metal ions aid the substrate stabilization in the active site of polymerases for further functional group transfer.^{7, 8}

Hydrolases are a class of enzymes that utilize water as a hydroxyl group donor during the substrate breakdown, and often require metal ions for their activity. The interaction of the metal ion with water results in the decrease of pK_a of water which leads to the generation of hydroxide ions at physiological pH. The produced hydroxide ion can later act as a nucleophile. Metal ions can also neutralize partially negative charges on the substrate or product.⁹ Hydrolases are represented by multiple cleaving enzymes as nucleases, lipases and proteases which utilize iron (II), zinc (II), nickel (II) and other divalent metals for their catalytic function.¹⁰

Carbonic anhydrase is the most studied representative of the lyase class, the group of enzymes that cleave C-C, C-O, C-N and other bonds using mechanisms other

than hydrolysis or oxidation to yield double bonds. These enzymes can also catalyze the addition of groups to double bonds.¹¹ The specific characteristic of these enzymes is that typically two substrates are involved in one reaction direction, but only one in the other direction. Carbonic anhydrase catalyzes the reversible hydration of gaseous CO₂ to carbonic acid (H₂CO₃), which then dissociates spontaneously to give bicarbonate (HCO₃⁻) and a proton.¹² The active site of carbonic anhydrase contains zinc ion which is essential for enzymatic activity (Figure 1.1).

Isomerases are the enzymes that catalyze the conversion of one isomer to another. Glucose isomerase, also known as xylose isomerase, catalyzes the reversible conversion of D-glucose and D-xylose to D-fructose and D-xylulose, respectively. Glucose isomerase requires divalent cations as cofactors for its isomerization activity and, depending on the source of the enzyme, can use magnesium, manganese or cobalt.¹³

Lastly, metalloenzymes are also found among ligases, the proteins that catalyze the joining of two molecules by using the energy of ATP. In the case of threonyl-tRNA synthetase activity, the enzyme uses a zinc ion to discriminate against the isosteric valine at the activation step. This metal is found in the active site of the enzyme's catalytic domain and is involved in the direct recognition of an amino acid by interacting with the amino and the hydroxyl groups (Figure 1.1).¹⁴

The widespread presence of the metalloproteins among all classes of enzymes suggests the impressive potential of these complexes and their significance for the proper functioning of living organisms.

1.3. Protein-metal interactions

The general analysis of metals used by proteins shows that natural metalloproteins utilize a limited set of metal ions. The placement of metal ions in different coordination environments controls the activities that metal can exhibit. For example, zinc is one of the most common metals used by proteins. Among others, zinc fingers present one of the canonical DNA-binding motifs found in proteins involved in the regulation of DNA replication and transcription. These small proteins require zinc ions for proper folding and coordinate metal through four cysteine and/or histidine residues.¹⁵ In DNA-binding zinc fingers, metal ion plays only structural role by providing stability to these small proteins. At the same time, while placed in similar environment in proteins like carbonic anhydrases, zinc becomes a central player in catalysis.¹² These observations suggest that the protein modulates the function of metal by placing it in the appropriate environment.

Copper-binding sites are another great example of metal's activity modulation by protein scaffold. Copper is a redox-active metal and upon reduction Cu(II) becomes Cu(I). Copper has the most diverse array of binding sites that are found in proteins with multiple enzymatic and electron-transport activities. Copper enzymes act as oxidases, mono- and dioxygenases, superoxide decomposing enzymes and nitrogen oxide (NO_x) reductases.^{16, 17} Based on the needs of the catalyzed reaction, the enzymes adopt different mononuclear (type-1 and type-2), binuclear (type-3, Cu_A and heme-Cu_B), trinuclear (coupled type-2/type-3) and tetranuclear (Cu₂) copper centers.^{18, 19} The information learned from the natural proteins allows us to identify the factors that

contribute to such diverse functions of the same metal and utilize them in our protein design efforts.

Metal-binding proteins are not only limited by the metal selection but also by the choice of ligands. Proteins consist of 20 proteinogenic amino acids but only a handful of them can be used as potential ligands. These limitations create situations where the metal-binding site can accommodate multiple metals with similar affinity or the slight change in the coordination environment can completely eliminate binding to the specific metal. The investigation of protein-metal interactions helps to address multiple questions, such as protein mismetallation which attracts more attention as more physiologically relevant enzymes are shown to malfunction to cause diseases because of wrong or inefficient metal coordination.²⁰⁻²³ The characterization of metal-binding sites in natural proteins allows us to understand the biochemical processes and utilize that information for protein engineering. Also, the identification of the patterns in protein-metal binding will help us to predict the locations of possible metal binding sites in the newly discovered proteins or to deduce their activity based on the known examples.

1.3.1. Metals in biological systems

Metal ions are widely present in biological systems and perform a variety of functions. Alkali metal ions, Na^+ and K^+ , are responsible for the maintenance of electrolyte balance, acid-base homeostasis, and maintenance of cell membrane potential. Alkaline earth metal ions, Ca^{2+} and Mg^{2+} , are important for cell signaling and the structural arrangement of many proteins, in some cases they are also involved in catalysis. Transition metals such as iron, zinc and copper are mostly involved in

catalysis and electron transfer. Although metal ions are critical for sustaining the functions of biological systems, their concentration must be maintained within a proper range as metals are often toxic or too reactive to exist in their free forms and require well-regulated transport and storage systems.

The analysis of known metalloproteins showed that natural enzymes use a limited number of metals available in Nature. Magnesium is the most common metal cofactor in natural metalloenzymes (Figure 1.2) and is involved in the stabilization of enzyme structure or facilitates substrate binding by forming a magnesium-substrate complex.⁴

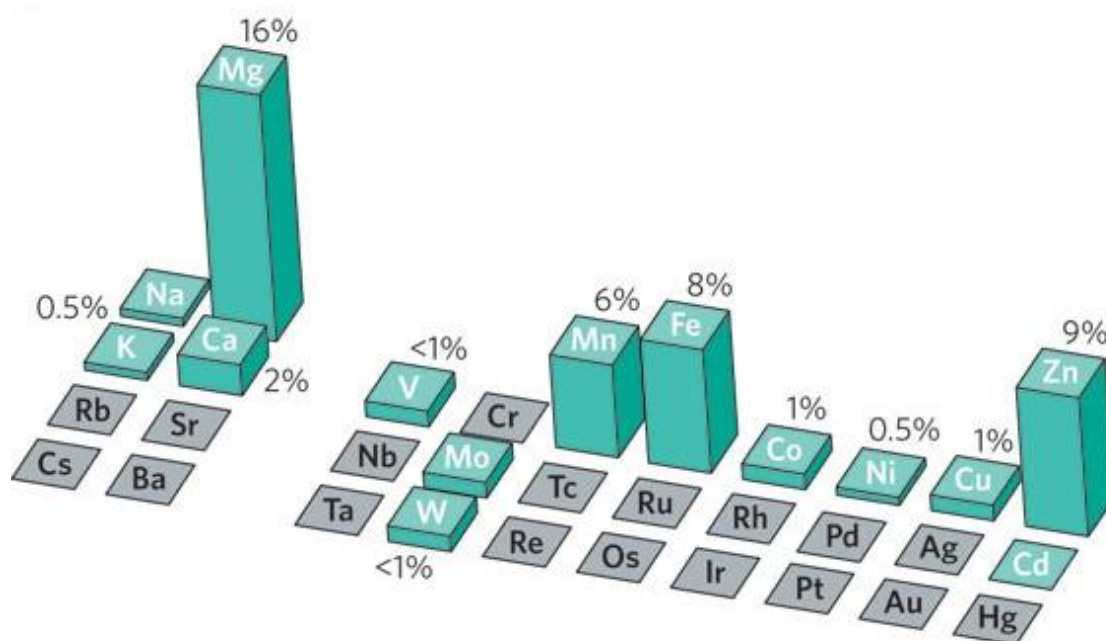


Figure 1.2. Elements used as cofactors by natural metalloenzymes. The figure is reproduced from ref. 4 with the permission from Springer Nature, copyright 2009.

Magnesium is the most abundant divalent metal and, possibly, because of its availability many enzymes evolved to utilize this metal.²⁴ Besides the availability, this

metal also has unique physicochemical properties, which makes it the best candidate for a specific function. Magnesium has a small ionic radius, which results in a high charge density, and slow solvent exchange rate, meaning that the ion prefers water to bulky ligands. These characteristics define magnesium binding to other molecules: the ion prefers octahedral geometry and makes four or fewer coordination bonds with an enzyme or a substrate, while other coordination positions are filled with water. Such coordination preferences are important for fast ligand exchange in magnesium-containing enzymes.²⁵

Proteins have adopted diverse metal-binding sites to accommodate metal ions in their free form or metal complexes. The most common metal organic complexes found in proteins are based on porphyrin, the heterocyclic macrocycle that provides four nitrogen atoms for metal coordination. Heme, chlorophyll and cobalamin (vitamin B₁₂) are well-known representatives of this group that coordinate to iron, magnesium and cobalt, respectively, and perform a variety of functions. Heme-containing enzymes are especially interesting as the redox-active nature of central iron atoms allows these proteins to participate in multiple enzymatic reactions (the decomposition of hydrogen peroxide by catalase and oxidation of metabolites by cytochrome p450)²⁶, non-enzymatic activities (oxygen transfer by myoglobin)²⁷ and electron-transfer chains (cytochrome b₅)²⁸. Another group of enzymes includes metal inorganic complexes as iron-sulfur clusters present in, for example, ferredoxin²⁹, and Mn₄CaO₅ clusters found in the oxygen-evolving complex of photosystem II³⁰. Proteins containing iron-sulfur clusters exhibit a wide range of reduction potentials that can be tuned by the coordination to ligands, thus they are commonly found in the redox-active proteins.

The characterization of ligands and metals that are found in known metalloproteins gives us an opportunity to identify trends and dependencies that can be used for future protein design efforts.

1.3.2. Metal coordination in proteins: ligand aspects

Amino acids in proteins provide functional groups which can potentially act as ligands for metal cation coordination. Specific chemical properties of metals, such as size, charge, and electronic configuration, define the composition and geometry of the metal-ligand complex. The interactions between metal and ligand have been characterized using the hard and soft acid and base (HSAB) theory. HSAB principle states that hard acids interact with hard bases and soft acids make complexes with soft bases.^{31, 32} Hard acids and bases have a high charge density, meaning these elements have a smaller ionic radius, higher effective nuclear charge and low polarizability. The *d* orbitals of hard acids are often unavailable for π bonding. On contrary, soft acids and bases have larger ionic radii, higher polarizability and lower electronegativity. The *d* orbitals of soft acids are available for π bonding and are filled or nearly filled with electrons.

In biological systems, positively charged metals display the characteristics of Lewis acids, meaning they act as electron acceptors. The amino acids provide ligands for metal coordination, acting as bases and electron donors. Free amino acids can coordinate metals through the amino, the carboxyl and the side chain groups. In polypeptides, the most common ligands for metal coordination are the amino acids with

nitrogen-, sulfur- and oxygen-containing side chains as well as the oxygen from the peptide bond (Table 1.1).

Table 1.1. Metals found in metalloproteins can be classified as hard or soft acids. The amino acids provide ligands for metal coordination that are classified as hard or soft bases.

Hard acids: Li ⁺ , Na ⁺ , K ⁺ , Mg ²⁺ , Ca ²⁺ , Mn ²⁺ , Fe ³⁺ , Co ³⁺	Borderline acids: Fe ²⁺ , Co ²⁺ , Ni ²⁺ , Cu ²⁺ , Zn ²⁺	Soft acids: Cu ⁺ , Ag ⁺ , Hg ²⁺
Hard bases: Oxygen-containing side chains (aspartic acid, glutamic acid, tyrosine, serine, threonine) Oxygen of the peptide bond	Borderline bases: Nitrogen-containing side chains (histidine, arginine)	Soft bases: Sulfur-containing side chains (cysteine, methionine)

HSAB theory provides a good baseline for the understanding of protein-metal interaction. Using these principles, potential metal-binding sites can be predicted or engineered in proteins. The side chains of aspartic and glutamic acids contain carboxylic groups and provide oxygen for metal coordination (Figure 1.3). These ligands are considered to be hard bases and bind to metal ions that are characterized as hard acids. Borderline histidine is the most common nitrogen-containing ligand found in natural proteins. Cysteine and methionine are classified as soft bases and provide sulfur for the binding to metals that are soft acids.³³

Among 20 proteinogenic amino acids, there are several which have been shown to participate in metal coordination but are not as common. In some cases, glutamine

and asparagine have been shown to contribute to metal coordination through the oxygen of the carboxamide group on their side chains.

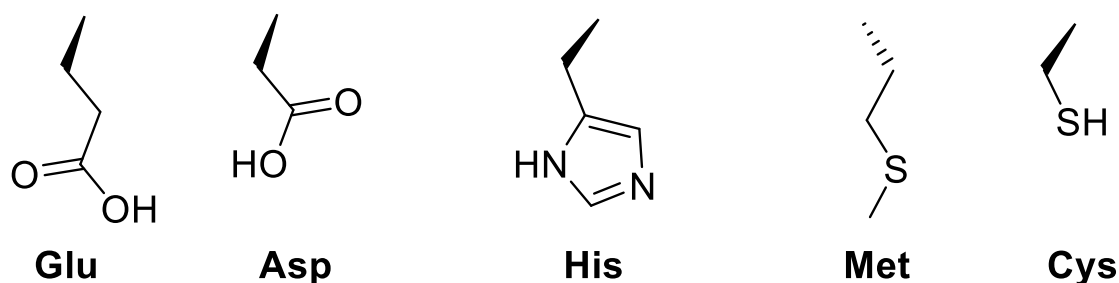


Figure 1.3. The side chains of common amino acids used for metal coordination in proteins. The amino acids are labeled with the three-letter code, where Glu is glutamic acid, Asp is aspartic acid, His is histidine, Met is methionine, Cys is cysteine.

For example, asparagine (Asn462) in isocitrate dehydrogenase kinase/phosphatase (AceK) has been found to coordinate to the divalent metals (Mn^{2+} and Mg^{2+}) through the oxygen atom of the carboxamide group.³⁴ Calcium- and magnesium-binding sites in transglutaminase 3 also contain asparagine residues that provide oxygen as a ligand.³⁵ These amino acids are also often involved in the stabilization of the water molecule coordinated to the metal. Serine, threonine and tyrosine, the amino acids that contain hydroxyl groups in their side chains, can also potentially coordinate metal ions. The tyrosine residue contributes to the binuclear iron-binding site in purple acid phosphatases as uteroferrin³⁶ and serine is found in the copper coordination sphere of CueR, a copper efflux regulator³⁷.

Aspartic and glutamic acids can act as monodentate or bidentate ligands for metal coordination. As those amino acids contain carboxylic groups, both oxygens can participate in metal binding. When metal is coordinated using only one carboxylic

oxygen, the binding is called monodentate. In the case of bidentate interactions, both carboxylic oxygens coordinate to the same metal. If the amino acid is coordinated to two different metals using both oxygens, it represents a bridging ligand (Figure 1.4). In certain cases, a water molecule can act as a bridging ligand or mediate the interaction between metal and ligand.

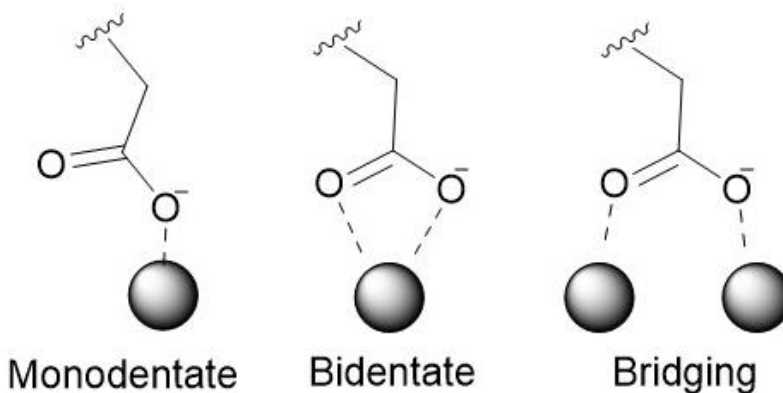
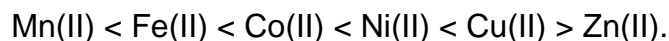


Figure 1.4. Metal coordination modes of carboxylate ligands. The metal ion is depicted as a grey sphere.

1.3.3. Metal coordination in proteins: metal aspects

The stability of transition metal-ligand complexes depends on the metal identity and can be predicted using the Irving-Williams series. This empirical rule refers to the rate of water exchange to any other ligand. The Irving-Williams series proposes that the stability of metal-ligand complexes increases across the period for divalent transition metals:



Most protein-metal complexes follow this trend which suggests that excess concentrations of zinc or copper can outcompete the cognate metal for the binding to

the metal-coordinating site. To ensure the specific metal binding and to guide the metal selectivity of protein, the metal-binding sites require fine-tuning with greater degree of flexibility.³⁸⁻⁴⁰

The coordination geometry is another factor that needs to be considered while describing or designing metal-binding sites. Coordination geometry defines the number and orientation of ligands around the metal and provides stability to the ligand-metal complex. The success of the metalloprotein design is majorly impacted by the structural features responsible for correct metal site geometry and function. To prevent mismetallation and to increase the possibility of specific metal coordination, the metal-binding site needs to be elaborately engineered to reduce the probability of obtaining multiple geometries. The orientation of ligands around metal ion is defined by the nature of this ion.^{41, 42} To assist with the prediction of the metal-binding site organization, many software and databases were developed based on the structural analysis of known metalloproteins.⁴³⁻⁴⁶ Thus, the alkali metal ions as Na⁺ and K⁺ form preferentially six-coordinate octahedral complexes, but distorted tetrahedral with four ligands and bipyramidal with five ligands were described for sodium, while eight-coordinate complexes were identified for potassium ion.⁴⁷⁻⁴⁹ Calcium and magnesium, the alkaline-earth metals, are preferentially coordinated by oxygen-containing ligands in octahedral arrangement (typically, six ligands).⁵⁰ The calcium-binding motif, EF-hand, is found in many natural calcium-binding proteins and coordinates metal in pentagonal bipyramidal geometry by providing seven ligands.⁵¹ The structure of the EF-hand will be discussed in detail in Chapter 4 as well as our work on redesigning the metal selectivity of this motif. Metals of the rare earth metal group (yttrium (III) and lanthanides) possess the

oxidation state of +3 and show tendencies in ligand binding similar to those of calcium. In fact, these metals can efficiently coordinate to calcium-binding sites.⁵²⁻⁵⁴ Rare earth metals can exist in complexes with an array of different coordination numbers of up to 9 within the inner sphere, which provides many opportunities for the design of efficient binders.⁵⁵⁻⁵⁷ Metal coordination geometries adopted by the most common transition metal cofactors found in proteins are summarized in Table 1.2.

Table 1.2. Coordination geometries observed for selected transition metals found in metalloproteins. The number of ligands is specified in parentheses and the most common geometry is in bold.

Metal	Coordination geometry (number of ligands)
Mn ²⁺	Square pyramidal (5), octahedral (6) ⁵⁸⁻⁶⁰
Fe ²⁺	tetrahedral (4), trigonal bipyramidal (5), octahedral (6) ⁶¹⁻⁶³
Fe ³⁺	tetrahedral (4), trigonal bipyramidal (5), octahedral (6) ^{61, 63, 64}
Co ²⁺	tetrahedral (4), square pyramidal (5), octahedral (6) ⁶⁵⁻⁶⁷
Ni ²⁺	square planar (4) , square pyramidal (5), octahedral (6) ⁶⁸⁻⁷⁰
Cu ⁺	Linear (2), trigonal planar (3), or tetrahedral (4) ^{19, 71, 72}
Cu ²⁺	Square planar (4) , square pyramidal (5) , octahedral (6) ^{19, 72, 73}
Zn ²⁺	Tetrahedral (4) , trigonal bipyramidal (5), square pyramidal (5), octahedral (6) ^{58, 59, 74}

The investigation of protein-metal interactions helps us to predict the locations of metal-binding sites and elucidate enzymatic mechanisms of different reactions. The design of new metalloproteins is a challenge that puts our knowledge of how metalloproteins work to test and allows us to reveal previously unknown structural features that may be missing from studies of native metalloproteins and their variants.

Besides the theoretical importance, protein engineering can result in new metalloenzymes for various real-life applications.

1.4. Methods in protein engineering and design of metalloproteins

Protein engineering has become a versatile tool for the exploration of natural enzymes and the development of new catalysts. It implies the modification of a protein's amino acid sequence to reorganize the structure and alter its function. Many successful examples have been reported where natural proteins were modified to obtain an improved or a new function.^{75, 76} In metalloproteins, this task becomes even more challenging as, besides the proper protein fold, the productive positioning of the metal is required. Redesign of natural proteins is a good strategy to obtain proteins with improved functions but the development of proteins with new functionalities might be negatively impacted by the prearrangement of active sites. *De novo* protein design is a field of protein engineering that explores the design of new amino acid sequences from scratch with predetermined properties such as structure and function.^{77, 78} Many productive examples of *de novo* designed proteins demonstrate the advances in our understanding of protein folding with progress in the design of metal-binding proteins. With the development of computational and molecular biology tools, multiple methods have been adopted and successfully used for protein engineering. Rational protein design and directed evolution are the two main approaches that helped to achieve significant breakthroughs in protein design. The combination of these approaches allows to overcome the limitations of the specific method and improve the efficiency of protein engineering efforts.

1.4.1. Rational design

Rational protein design relies on the advanced understanding of protein folding⁷⁹,⁸⁰, as well as the structure and function relationship. The advances in our understanding of kinetics and thermodynamics of protein folding, as well as the principles of protein structure organization, resulted in the development of computational tools. These methods use complex calculations that allow us to predict the effect of a specific mutation in a certain scaffold or to design a protein from scratch. The evolution of natural proteins can be performed using rational design where the researchers would utilize their knowledge of known protein structure-function relationships to incorporate specific changes in a natural scaffold. Another important use of such tools is the design of *de novo* proteins. *De novo* proteins are molecules that were engineered artificially and are not found in Nature.⁸¹ The design of proteins from scratch tests our understanding of protein structural organization and allows us to explore the new functionalities of these molecules. The development of computational programs that aid with protein design requires large sets of information obtained from the investigation of natural proteins.

The process of natural enzyme engineering starts with the analysis of its structure obtained using crystallography or NMR spectroscopy. For proteins the structures of which are not available, the models can be designed using template-based or template-independent modeling.⁸²⁻⁸⁴ In template-based modeling, or homology modeling, the previously determined structure of a related protein is used to model the unknown structure of the target, while template-independent modeling requires specialized software that applies physics-based energy functions to model protein

structure. After the initial analysis is performed, the computational tools are used to predict the positions and identity of potentially productive mutations to achieve a desired characteristic. The library of designed variants needs to be screened in the functional assay that allows for identification of proteins with improved activity. Quantum mechanics calculations, molecular dynamics experiments and deep learning have been successfully used for the design of novel proteins.⁸⁵⁻⁸⁹

One of the simple approaches to rational protein design, the minimalist approach, relies on our understanding of protein structure and function. The minimalist approach simplifies catalysts to small functional units by examining the structure of the target protein and transferring those units into a small, designed scaffold or by introducing them into an existing protein. A great example of the successful application of the minimalist approach is the design of AlleyCat, an allosterically regulated catalyst. AlleyCat is a 74-residue C-terminal domain of calmodulin, a calcium-binding non-enzymatic protein, which was evolved into a Kemp eliminase using the minimalist approach and rational design.⁹⁰ Kemp elimination is a simple unnatural reaction of proton abstraction from the benzisoxazole ring by a base that results in substrate ring opening and color change of the reaction.⁹¹ From the minimalist point of view, the catalyst for this reaction requires a base placed in the hydrophobic environment. The elegant design of this protein takes advantage of the protein structural rearrangement upon calcium binding, resulting in the creation of the hydrophobic pocket that will function as an active site. The computational methods were used to predict the best position and the identity of the active residue in the hydrophobic pocket of calmodulin to obtain the desired function.

The introduction of metal-binding sites in proteins requires the understanding of the coordination sphere arrangement and the role of the outer coordination sphere environment. Yi Lu and coworkers performed a successful introduction of a new metal-binding site into a natural protein. In their work, they have converted oxygen-binding protein myoglobin into a heme-copper oxidase and nitric oxide reductase.^{92, 93} Myoglobin is a non-enzymatic heme-containing protein that participates in oxygen transport and storage.²⁷ Myoglobin is a popular scaffold for protein engineering due to its stability during recombinant protein production, its structure is well-known, and a heme group presents many opportunities for activity manipulation. Heme-copper oxidases are metalloenzymes that contain a heme and Cu_B dinuclear center and catalyze the reduction of molecular oxygen to water, which is linked to proton transfer across the cytoplasmic or mitochondrial membranes.⁹⁴ Nitric oxide reductase (NOR) participates in the denitrification pathway of anaerobic bacteria that is responsible for the two-electron reduction of NO to N₂O. The active site of NOR consists of heme and a nonheme Fe_B center. Using structural comparison and computer modeling, researchers identified the best positions to introduce copper-binding or iron-binding sites into myoglobin. The binding site in both cases consists of three histidine residues that coordinate to copper or iron through the nitrogen atoms of the side chains with another glutamate residue for iron binding.

Rational design has been successfully used for the design of de novo metalloproteins. DeGrado and co-workers were able to create a family of antiparallel four-stranded coiled coils called Due Ferri (DF proteins) with the replication of the first and second metal coordination spheres observed in the non-heme diiron/dimanganese

enzymes such as manganese catalase, ribonucleotide reductase, methane monooxygenase, toluene oxygenase, desaturases, and ferroxidases.⁹⁵ The metal binding site was reconstructed and placed into a four-helix bundle that resembles the active domain of mentioned enzymes. Proteins of DF family show a wide range of enzymatic activities that are observed in natural proteins such as the radical stabilization, N-hydroxylation of arylamines, oxygen reduction to generate peroxide, oxidation of 4-aminophenol, which allowed to use these designs as model proteins to study the mechanisms of multiple reactions.⁹⁶⁻¹⁰¹ This work is an excellent example of *de novo* protein design through the applications of retrostructural analysis, rational design and computational design.^{96, 97, 102, 103}

Many other examples of metalloprotein design have been reported and summarized elsewhere.^{78, 104, 105} Computational methods have been successfully used to create new artificial metalloproteins with specific activities and to engineer new functionalities into existing proteins. The major limitations of this approach are the necessity of protein structural information and the knowledge about the structure-function relationship which are not available for the majority of proteins.

1.4.2. *Directed evolution*

Directed evolution approach unites a set of molecular biology techniques which allow scientists to mimic protein evolution in the laboratory.^{106, 107} The strategy generally involves the mutagenesis or recombination of a starting gene with further screening to isolate new variants with improvements in specific properties and the selected improved genes can be subjected to another round of evolution. Directed evolution allows to

create new protein sequences without using prior structural information as this method relies on the random process for DNA sequence modification. As high-quality structural information is not available for many proteins, it is challenging to predict the positions of productive mutations in such proteins. The solution comes in form of directed evolution, where large libraries of random mutants are created and the effect of those mutations on protein's function is tested.

Mutagenesis is performed using molecular biology tools inspired by Nature. Error-prone PCR (epPCR) is the most commonly used technique that allows to introduce random point mutations in DNA products.¹⁰⁸ The mutations in the sequence occur when Taq polymerase, the enzyme that promotes DNA polymerization, is subjected to imperfect reaction conditions, which, together with the absence of proofreading activity in this enzyme, lead to the introduction of many mutations.¹⁰⁹ In order to find a productive mutation, large libraries of random mutants are created and selected based on the desired activity. For this, proteins need to be produced in the effective host (usually strains of bacteria *E. coli*) and tested in a high-throughput assay. The further identification of mutations that result in evolved proteins is performed using DNA sequencing techniques. DNA shuffling techniques or *in vitro* recombination allow to create new DNA sequences based on the fragments of the template genes.¹¹⁰ The beneficial mutations identified during random mutagenesis can be rapidly combined to expand the sequence diversity derived from a small library of homologous genes. Other techniques based on PCR product staggering and DNA recombination have been developed and successfully used to achieve the improvement in protein's function.¹¹¹ The process of directed evolution is rather laborious as it requires many iterations with

high-throughput screening until the desired changes are obtained or until no further improvement can be achieved. Nonetheless, directed evolution has been successfully used to produce metalloproteins with improved functions.

Metalloprotein design is often challenging as the proper metal and substrate positioning in the protein need to be ensured. Heme-containing proteins have been extensively used for metalloprotein redesign efforts. These proteins present a diverse group of natural catalysts that are involved in many processes from oxidation to small-molecule and electron transfer.¹¹² Heme cofactor in diverse coordination environments provides a wide range of catalytic and non-catalytic activities. Cytochrome P450 enzymes (CYPs) are a group of heme-containing enzymes that utilize molecular oxygen for the oxidative insertion of one oxygen atom into an organic substrate and require the reducing agent for their function. As CYPs are membrane-bound enzymes and are not stable in the solution, the soluble cytochrome P450 BM3 fusion is usually used for protein design efforts. In this protein, a diflavin-containing reductase (BM3) is fused to a heme-containing P450 domain in a single polypeptide chain which allows the cytochrome to receive the reducing equivalents effectively.¹¹³ Cytochrome P540 enzymes catalyze a variety of reactions, including the hydroxylation of long-chain fatty acids, although, the wild-type enzyme is not efficient in recognizing short-chain alkanes. The classic example of metalloprotein engineering is the work led by Dr. Frances Arnold, who received a Nobel Prize in Chemistry in 2018 for her contribution to the field of protein design. Inspired by the functional flexibility of natural cytochrome P450 monooxygenases, the group succeeded in improving enzyme's activity on short-chain alkanes and repurposed the enzymes from classic oxidation to carbon-carbon bond

forming reactions, cyclopropanation, carbene C-H insertions using directed evolution.¹¹⁴⁻

¹¹⁸ Cytochrome p450 is overall a popular scaffold for protein engineering.^{119, 120}

Directed evolution has been also used to enhance the stability of carbonic anhydrase, a Zn-dependent natural enzyme that promotes reversible hydration of carbon dioxide to produce bicarbonate and a proton. This enzyme can become a valuable tool for capturing CO₂ from the atmosphere after the improvement of its thermal and solvent stability.¹²¹ Human carbonic anhydrase II is the most efficient natural enzyme known with the catalytic efficiency close to the diffusion limit. Besides its main function in CO₂ hydration, this enzyme also exhibits promiscuous esterase activity. Using epPCR, Gould and Tawfik were able to adapt the active site for bulkier substrates as nonactivated esters.¹²²

Directed evolution is an effective method for the enhancement of protein properties such as substrate binding or specificity, as well as catalytic activity, although it is not as efficient when the metal-binding site needs to be introduced into the protein. The engineering of metal-binding sites or the modification of metal selectivity are challenging tasks that in many cases cannot be achieved using solely directed evolution methods and usually are combined with computational calculations and rational design. Also, as directed evolution relies on the sampling of large combinatorial libraries, this method requires laborious and time-consuming high-throughput screening.

1.4.3. Semi-rational design

Directed evolution is a productive technique for protein design but often requires many rounds of experiments to achieve a significant result. To improve the efficiency of

this method, scientists have been working to develop methods of guided evolution that would allow to identify promising regions and focus on them, instead of performing evolution on the whole protein sequence. Semi-rational design approach aids in the identification of promising target sites as well as the amino acid diversity for those positions. These methods utilize information about protein sequence, structure and function, in combination with computational predictive algorithms for protein engineering. Thus, semi-rational design is the application of rational design methods to limit the number of variants for the directed evolution.

In recent times, machine learning has been employed to identify the “hot spots”, the positions for potentially productive mutations, important for protein function.^{123, 124} Machine-learning programs learn from examples of protein sequences and the respective functional measurements of the proteins available in the database. After the selection of the appropriate model, the algorithm needs to be trained to recognize the patterns. Finally, after extensive rounds of calculations, the theoretical data needs to be verified in the actual experiment. A great review by Andreini and Rosato discusses the reported strategies and tools developed based on deep learning.¹²⁵ Machine learning is a promising method for guided directed evolution but, because of the current limitations in protein structure-function relationship information available for learning, cannot be used to its full potential.

Semi-rational approach was used by Lippow et. al. to modify the specificity of galactose 6-oxidase and generate a novel enzyme for use in a proposed biosynthetic pathway for D-glucaric acid.¹²⁶ Galactose 6-oxidase is a mononuclear copper enzyme that catalyzes the oxidation of primary alcohols, including glucose. The favorable

positions and amino acid substitutions were predicted by computational design using Rosetta suit and incorporated using site-directed mutagenesis. The docking of galactose into the protein's active site induced conformational change and allowed to identify positions for productive mutations among the metal's second coordination sphere, the amino acid residues used to stabilize metal-coordinating ligands.

Rui et al. performed computational modeling of non-heme enzymes and identified the positions for further directed evolution.¹²⁷ They used 4-hydroxyphenylpyruvate dioxygenase from *Streptomyces avermitilis* (Sav HppD), a non-heme iron(II)-containing enzyme that catalyzes the second reaction in the catabolism of tyrosine, as a scaffold for computational modeling and identified positions for optimization of protein to perform enantioselective azide transfer. All identified positions were located in the active site and in the proximity of the iron metal.

Semi-rational protein design combines multiple approaches to facilitate more productive protein engineering. Computational methods allow to narrow down the number of possible variants produced for screening and directed evolution techniques help to test multiple options at specific positions. Even though multiple successful examples of semi-rational protein engineering have been reported, all those methods heavily rely on the availability of protein structural and functional information, which poses a limitation for this method. Thus, the development of new methods that could allow to predict the positions for productive mutations are in demand. Chapter 3 will discuss the development of a new method that allows to use NMR to predict the positions for productive mutations in the target proteins.

1.5. References

1. Faber, M. S.; Whitehead, T. A., Data-driven engineering of protein therapeutics. *Curr. Opin. Biotechnol.* **2019**, *60*, 104-110.
2. Singh, J.; Vijayan, V.; Ahmedi, S.; Pant, P.; Manzoor, N.; Singh, T. P.; Sharma, P.; Sharma, S., Lactosmart: a novel therapeutic molecule for antimicrobial defense. *Front. Microbiol.* **2021**, *12*, 672589.
3. Kianfar, E., Protein nanoparticles in drug delivery: animal protein, plant proteins and protein cages, albumin nanoparticles. *J. Nanobiotechnology* **2021**, *19* (1), 159.
4. Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009**, *460* (7257), 823-30.
5. Younus, H., Therapeutic potentials of superoxide dismutase. *Int. J. Health Sci. (Qassim)* **2018**, *12* (3), 88-93.
6. Fukai, T.; Ushio-Fukai, M., Superoxide dismutases: role in redox signaling, vascular function, and diseases. *Antioxid. Redox Signal.* **2011**, *15* (6), 1583-606.
7. Sosunov, V.; Sosunova, E.; Mustaev, A.; Bass, I.; Nikiforov, V.; Goldfarb, A., Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.* **2003**, *22* (9), 2234-44.
8. Tsai, M. D., Catalytic mechanism of DNA polymerases – two metal ions or three? *Protein Sci.* **2019**, *28* (2), 288-291.
9. Kimura, E., Dimetallic hydrolases and their models. *Curr. Opin. Chem. Biol.* **2000**, *4* (2), 207-13.
10. Hernick, M.; Fierke, C. A. In *Mechanisms of metal-dependent hydrolases in metabolism*, 2010.

11. van der Werf, M. J.; van den Tweel, W. J.; Kamphuis, J.; Hartmans, S.; de Bont, J. A., The potential of lyases for the industrial production of optically active compounds. *Trends Biotechnol.* **1994**, *12* (3), 95-103.
12. Lindskog, S.; Coleman, J. E., The catalytic mechanism of carbonic anhydrase. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70* (9), 2505-8.
13. Nam, K. H., Glucose Isomerase: Functions, Structures, and Applications. *Appl. Sci.* **2022**, *12* (1), 428.
14. Dock-Bregeon, A.; Sankaranarayanan, R.; Romby, P.; Caillet, J.; Springer, M.; Rees, B.; Francklyn, C. S.; Ehresmann, C.; Moras, D., Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. *Cell* **2000**, *103* (6), 877-84.
15. Kluska, K.; Adamczyk, J.; Krezel, A., Metal binding properties of zinc fingers with a naturally altered metal binding site. *Metallomics* **2018**, *10* (2), 248-263.
16. Kaintz, C.; Mauracher, S. G.; Rompel, A., Type-3 copper proteins: recent advances on polyphenol oxidases. *Adv. Protein Chem. Struct. Biol.* **2014**, *97*, 1-35.
17. Pember, S. O.; Villafranca, J. J.; Benkovic, S. J., Phenylalanine hydroxylase from *Chromobacterium violaceum* is a copper-containing monooxygenase. Kinetics of the reductive activation of the enzyme. *Biochemistry* **1986**, *25* (21), 6611-9.
18. Sakurai, T.; Kataoka, K., Structure and function of type I copper in multicopper oxidases. *Cell Mol. Life Sci.* **2007**, *64* (19-20), 2642-56.
19. Rubino, J. T.; Franz, K. J., Coordination chemistry of copper proteins: how nature handles a toxic cargo for essential function. *J. Inorg. Biochem.* **2012**, *107* (1), 129-43.

20. Cotruvo, J. A., Jr.; Stubbe, J., Metallation and mismetallation of iron and manganese proteins in vitro and in vivo: the class I ribonucleotide reductases as a case study. *Metallomics* **2012**, *4* (10), 1020-36.
21. Foster, A. W.; Young, T. R.; Chivers, P. T.; Robinson, N. J., Protein metalation in biology. *Curr. Opin. Chem. Biol.* **2022**, *66*, 102095.
22. Robinson, N. J.; Glasfeld, A., Metalation: nature's challenge in bioinorganic chemistry. *J. Biol. Inorg. Chem.* **2020**, *25* (4), 543-545.
23. Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N., Mysteries of metals in metalloenzymes. *Acc. Chem. Res.* **2014**, *47* (10), 3110-7.
24. Ryan, M. F.; Barbour, H., Magnesium measurement in routine clinical practice. *Ann. Clin. Biochem.* **1998**, *35* (Pt 4), 449-59.
25. Cowan, J. A., Structural and catalytic chemistry of magnesium-dependent enzymes. *Biometals* **2002**, *15* (3), 225-35.
26. Poulos, T. L., Heme enzyme structure and function. *Chem. Rev.* **2014**, *114* (7), 3919-62.
27. Cole, R. P., Myoglobin function in exercising skeletal muscle. *Science* **1982**, *216* (4545), 523-5.
28. Vergeres, G.; Waskell, L., Cytochrome b5, its functions, structure and membrane topology. *Biochimie* **1995**, *77* (7-8), 604-20.
29. Liu, J.; Chakraborty, S.; Hosseinzadeh, P.; Yu, Y.; Tian, S.; Petrik, I.; Bhagi, A.; Lu, Y., Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **2014**, *114* (8), 4366-469.

30. Narzi, D.; Bovi, D.; Guidoni, L., Pathway for Mn-cluster oxidation by tyrosine-Z in the S2 state of photosystem II. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (24), 8723-8.
31. Pearson, R. G., Hard and Soft Acids and Bases. *J. Am. Chem. Soc.* **1963**, *85* (22), 3533-3539.
32. Ho, T. L.; Ho, H. C.; Hamilton, L. D., Biochemical significance of the hard and soft acids and bases principle. *Chem. Biol. Interact.* **1978**, *23* (1), 65-84.
33. Lu, C. H.; Lin, Y. F.; Lin, J. J.; Yu, C. S., Prediction of metal ion-binding sites in proteins using the fragment transformation method. *PLoS One* **2012**, *7* (6), e39252.
34. Zhang, X.; Shen, Q.; Lei, Z.; Wang, Q.; Zheng, J.; Jia, Z., Characterization of metal binding of bifunctional kinase/phosphatase AceK and implication in activity modulation. *Sci. Rep.* **2019**, *9* (1), 9198.
35. Ahvazi, B.; Kim, H. C.; Kee, S. H.; Nemes, Z.; Steinert, P. M., Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. *EMBO J.* **2002**, *21* (9), 2055-67.
36. Bernhardt, P. V.; Schenk, G.; Wilson, G. J., Direct electrochemistry of porcine purple acid phosphatase (uteroferrin). *Biochemistry* **2004**, *43* (32), 10387-92.
37. Changela, A.; Chen, K.; Xue, Y.; Holschen, J.; Outten, C. E.; O'Halloran, T. V.; Mondragon, A., Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. *Science* **2003**, *301* (5638), 1383-7.
38. Grave, K.; Griese, J. J.; Berggren, G.; Bennett, M. D.; Hogbom, M., The *Bacillus anthracis* class Ib ribonucleotide reductase subunit NrdF intrinsically selects manganese over iron. *J. Biol. Inorg. Chem.* **2020**, *25* (4), 571-582.

39. Kisgeropoulos, E. C.; Griese, J. J.; Smith, Z. R.; Branca, R. M. M.; Schneider, C. R.; Hogbom, M.; Shafaat, H. S., Key structural motifs balance metal binding and oxidative reactivity in a heterobimetallic Mn/Fe protein. *J. Am. Chem. Soc.* **2020**, *142* (11), 5338-5354.
40. Choi, T. S.; Tezcan, F. A., Overcoming universal restrictions on metal selectivity by protein design. *Nature* **2022**, *603* (7901), 522-527.
41. Yao, S.; Flight, R. M.; Rouchka, E. C.; Moseley, H. N., Aberrant coordination geometries discovered in the most abundant metalloproteins. *Proteins* **2017**, *85* (5), 885-907.
42. Barber-Zucker, S.; Shaanan, B.; Zarivach, R., Transition metal binding selectivity in proteins and its correlation with the phylogenomic classification of the cation diffusion facilitator protein family. *Sci. Rep.* **2017**, *7* (1), 16381.
43. Sciortino, G.; Garribba, E.; Rodriguez-Guerra Pedregal, J.; Marechal, J. D., Simple coordination geometry descriptors allow to accurately predict metal-binding sites in proteins. *ACS Omega* **2019**, *4* (2), 3726-3731.
44. Andreini, C.; Cavallaro, G.; Lorenzini, S., FindGeo: a tool for determining metal coordination geometry. *Bioinformatics* **2012**, *28* (12), 1658-60.
45. Andreini, C.; Cavallaro, G.; Lorenzini, S.; Rosato, A., MetalPDB: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.* **2013**, *41*, D312-9.
46. Valasatava, Y.; Rosato, A.; Banci, L.; Andreini, C., MetalPredator: a web server to predict iron-sulfur cluster binding proteomes. *Bioinformatics* **2016**, *32* (18), 2850-2.

47. Noskov, S. Y.; Roux, B., Control of ion selectivity in LeuT: two Na⁺ binding sites with two different mechanisms. *J. Mol. Biol.* **2008**, *377* (3), 804-18.
48. Zheng, H.; Cooper, D. R.; Porebski, P. J.; Shabalin, I. G.; Handing, K. B.; Minor, W., CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr. D Struct. Biol.* **2017**, *73* (Pt 3), 223-233.
49. Thomas, M.; Jayatilaka, D.; Corry, B., The predominant role of coordination number in potassium channel selectivity. *Biophys. J.* **2007**, *93* (8), 2635-43.
50. Kuppuraj, G.; Dudev, M.; Lim, C., Factors governing metal-ligand distances and coordination geometries of metal complexes. *J. Phys. Chem. B* **2009**, *113* (9), 2952-60.
51. Yang, W.; Lee, H. W.; Hellinga, H.; Yang, J. J., Structural analysis, identification, and design of calcium-binding sites in proteins. *Proteins* **2002**, *47* (3), 344-56.
52. Pidcock, E.; Moore, G. R., Structural characteristics of protein binding sites for calcium and lanthanide ions. *J. Biol. Inorg. Chem.* **2001**, *6* (5-6), 479-89.
53. Le Clainche, L.; Plancque, G.; Amekraz, B.; Moulin, C.; Pradines-Lecomte, C.; Peltier, G.; Vita, C., Engineering new metal specificity in EF-hand peptides. *J. Biol. Inorg. Chem.* **2003**, *8* (3), 334-40.
54. Edington, S. C.; Gonzalez, A.; Middendorf, T. R.; Halling, D. B.; Aldrich, R. W.; Baiz, C. R., Coordination to lanthanide ions distorts binding site conformation in calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (14), 3126-3134.
55. Tickner, B. J.; Stasiuk, G. J.; Duckett, S. B.; Angelovski, G., The use of yttrium in medical imaging and therapy: historical background and future perspectives. *Chem. Soc. Rev.* **2020**, *49* (17), 6169-6185.

56. Cotton, S. A., Establishing coordination numbers for the lanthanides in simple complexes. *Comptes Rendus Chimie* **2005**, *8* (2), 129-145.
57. Lim, S.; Franklin, S. J., Lanthanide-binding peptides and the enzymes that might have been. *Cell Mol. Life Sci.* **2004**, *61* (17), 2184-8.
58. Senguen, F. T.; Grabarek, Z., X-ray structures of magnesium and manganese complexes with the N-terminal domain of calmodulin: insights into the mechanism and specificity of metal ion binding to an EF-hand. *Biochemistry* **2012**, *51* (31), 6182-94.
59. Kumar, P.; Dalal, V.; Sharma, N.; Kokane, S.; Ghosh, D. K.; Kumar, P.; Sharma, A. K., Characterization of the heavy metal binding properties of periplasmic metal uptake protein CLas-ZnuA2. *Metallomics* **2020**, *12* (2), 280-289.
60. Udayalaxmi, S.; Gangula, M. R.; Ravikiran, K.; Ettaiah, P., Investigation of manganese metal coordination in proteins: a comprehensive PDB analysis and quantum mechanical study. *Struct. Chem.* **2020**, *31* (3), 1057-1064.
61. Piccioli, M.; Turano, P., Transient iron coordination sites in proteins: Exploiting the dual nature of paramagnetic NMR. *Coord. Chem. Rev.* **2015**, *284*, 313-328.
62. Koropatkin, N.; Randich, A. M.; Bhattacharyya-Pakrasi, M.; Pakrasi, H. B.; Smith, T. J., The structure of the iron-binding protein, FutA1, from *Synechocystis* 6803. *J. Biol. Chem.* **2007**, *282* (37), 27468-27477.
63. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M., Metal ions in biological catalysis: from enzyme databases to general principles. *J. Biol. Inorg. Chem.* **2008**, *13* (8), 1205-18.
64. Bruns, C. M.; Nowalk, A. J.; Arvai, A. S.; McTigue, M. A.; Vaughan, K. G.; Mietzner, T. A.; McRee, D. E., Structure of *Haemophilus influenzae* Fe(+3)-binding

protein reveals convergent evolution within a superfamily. *Nat. Struct. Biol.* **1997**, *4* (11), 919-24.

65. Pennella, M. A.; Shokes, J. E.; Cosper, N. J.; Scott, R. A.; Giedroc, D. P., Structural elements of metal selectivity in metal sensor proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (7), 3713-3718.

66. Sivo, V.; D'Abrosca, G.; Russo, L.; Iacovino, R.; Pedone, P. V.; Fattorusso, R.; Isernia, C.; Malgieri, G., Co(II) Coordination in Prokaryotic Zinc Finger Domains as Revealed by UV-Vis Spectroscopy. *Bioinorg. Chem. Appl.* **2017**, *2017*, 1527247.

67. Smith, A. T.; Majtan, T.; Freeman, K. M.; Su, Y.; Kraus, J. P.; Burstyn, J. N., Cobalt cystathionine beta-synthase: a cobalt-substituted heme protein with a unique thiolate ligation motif. *Inorg. Chem.* **2011**, *50* (10), 4417-27.

68. Selvan, D.; Prasad, P.; Farquhar, E. R.; Shi, Y.; Crane, S.; Zhang, Y.; Chakraborty, S., Redesign of a copper storage protein into an artificial hydrogenase. *ACS Catal.* **2019**, *9* (7), 5847-5859.

69. Higgins, K., Nickel metalloregulators and chaperones. *Inorganics (Basel)* **2019**, *7* (8).

70. Ragsdale, S. W., Nickel-based enzyme systems. *J. Biol. Chem.* **2009**, *284* (28), 18571-5.

71. Gray, H. B.; Malmstrom, B. G.; Williams, R. J., Copper coordination in blue proteins. *J. Biol. Inorg. Chem.* **2000**, *5* (5), 551-9.

72. Hureau, C., Coordination of redox active metal ions to the amyloid precursor protein and to amyloid- β peptides involved in Alzheimer disease. Part 1: An overview. *Coord. Chem. Rev.* **2012**, *256* (19), 2164-2174.

73. Solomon, E. I.; Hare, J. W.; Gray, H. B., Spectroscopic studies and a structural model for blue copper centers in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73* (5), 1389-93.
74. Laitaoja, M.; Valjakka, J.; Janis, J., Zinc coordination spheres in protein structures. *Inorg. Chem.* **2013**, *52* (19), 10983-91.
75. Li, C.; Zhang, R.; Wang, J.; Wilson, L. M.; Yan, Y., Protein engineering for improving and diversifying natural product biosynthesis. *Trends Biotechnol.* **2020**, *38* (7), 729-744.
76. Tobin, P. H.; Richards, D. H.; Callender, R. A.; Wilson, C. J., Protein engineering: a new frontier for biological therapeutics. *Curr. Drug Metab.* **2014**, *15* (7), 743-56.
77. Korendovych, I. V.; DeGrado, W. F., *De novo* protein design, a retrospective. *Q. Rev. Biophys.* **2020**, *53*, e3.
78. Pan, X.; Kortemme, T., Recent advances in *de novo* protein design: principles, methods, and applications. *J. Biol. Chem.* **2021**, *296*, 100558.
79. DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D., Protein design, a minimalist approach. *Science* **1989**, *243* (4891), 622-8.
80. Hellinga, H. W., Rational protein design: combining theory and experiment. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (19), 10015-7.
81. Regan, L.; DeGrado, W. F., Characterization of a helical protein designed from first principles. *Science* **1988**, *241* (4868), 976-8.

82. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10* (6), 845-58.
83. Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T., Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **2009**, *4* (1), 1-13.
84. Kuhlman, B.; Bradley, P., Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (11), 681-697.
85. Ding, W.; Nakai, K.; Gong, H., Protein design via deep learning. *Brief Bioinform.* **2022**, *23* (3).
86. Wang, J.; Ma, C.; Fiorin, G.; Carnevale, V.; Wang, T.; Hu, F.; Lamb, R. A.; Pinto, L. H.; Hong, M.; Klein, M. L.; DeGrado, W. F., Molecular dynamics simulation directed rational design of inhibitors targeting drug-resistant mutants of influenza A virus M2. *J. Am. Chem. Soc.* **2011**, *133* (32), 12834-41.
87. Lin, H. Y.; Chen, X.; Dong, J.; Yang, J. F.; Xiao, H.; Ye, Y.; Li, L. H.; Zhan, C. G.; Yang, W. C.; Yang, G. F., Rational redesign of enzyme via the combination of quantum mechanics/molecular mechanics, molecular dynamics, and structural biology study. *J. Am. Chem. Soc.* **2021**, *143* (38), 15674-15687.
88. Anishchenko, I.; Pellock, S. J.; Chidyausiku, T. M.; Ramelot, T. A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A. K.; DiMaio, F.; Carter, L.; Chow, C. M.; Montelione, G. T.; Baker, D., *De novo* protein design by deep network hallucination. *Nature* **2021**, *600* (7889), 547-552.

89. Feehan, R.; Franklin, M. W.; Slusky, J. S. G., Machine learning differentiates enzymatic and non-enzymatic metals in proteins. *Nat. Commun.* **2021**, *12* (1), 3712.
90. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (17), 6823-7.
91. Kemp, D. S.; Casey, M. L., Physical organic chemistry of benzisoxazoles. II. Linearity of the Bronsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **1973**, *95*, 6670-6680.
92. Sigman, J. A.; Kwok, B. C.; Lu, Y., From myoglobin to heme-copper oxidase: Design and engineering of a Cu_B center into sperm whale myoglobin. *J. Am. Chem. Soc.* **2000**, *122* (34), 8192-8196.
93. Yeung, N.; Lin, Y. W.; Gao, Y. G.; Zhao, X.; Russell, B. S.; Lei, L.; Miner, K. D.; Robinson, H.; Lu, Y., Rational design of a structural and functional nitric oxide reductase. *Nature* **2009**, *462* (7276), 1079-82.
94. Blomberg, M. R. A., Activation of O₂ and NO in heme-copper oxidases – mechanistic insights from computational modelling. *Chem. Soc. Rev.* **2020**, *49* (20), 7301-7330.
95. Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G., Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J. Mol. Biol.* **2003**, *334* (5), 1101-15.
96. Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F., Retrostructural analysis of metalloproteins: application to the design of

a minimal model for diiron proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (12), 6298-305.

97. Di Costanzo, L.; Wade, H.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F.; Lombardi, A., Toward the *de novo* design of a catalytically active helix bundle: a substrate-accessible carboxylate-bridged dinuclear metal center. *J. Am. Chem. Soc.* **2001**, *123* (51), 12749-57.

98. Ulas, G.; Lemmin, T.; Wu, Y.; Gassner, G. T.; DeGrado, W. F., Designed metalloprotein stabilizes a semiquinone radical. *Nat. Chem.* **2016**, *8* (4), 354-9.

99. Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; DeGrado, W. F., Alteration of the oxygen-dependent reactivity of *de novo* Due Ferri proteins. *Nat. Chem.* **2012**, *4* (11), 900-6.

100. Snyder, R. A.; Betzu, J.; Butch, S. E.; Reig, A. J.; DeGrado, W. F.; Solomon, E. I., Systematic perturbations of binuclear non-heme iron sites: structure and dioxygen reactivity of *de novo* Due Ferri proteins. *Biochemistry* **2015**, *54* (30), 4637-51.

101. Faiella, M.; Andreozzi, C.; de Rosales, R. T.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A., An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **2009**, *5* (12), 882-4.

102. Marsh, E. N.; DeGrado, W. F., Noncovalent self-assembly of a heterotetrameric diiron protein. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 5150-4.

103. Pasternak, A.; Kaplan, J.; Lear, J. D.; DeGrado, W. F., Proton and metal ion-dependent assembly of a model diiron protein. *Protein Sci.* **2001**, *10* (5), 958-69.

104. Lin, Y. W., Rational Design of Artificial Metalloproteins and Metalloenzymes with Metal Clusters. *Molecules* **2019**, *24* (15).
105. Floudas, C. A.; Fung, H. K.; McAllister, S. R.; Mönnigmann, M.; Rajgaria, R., Advances in protein structure prediction and de novo protein design: A review. *Chem. Eng. Sci.* **2006**, *61* (3), 966-988.
106. Dalby, P. A., Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 473-80.
107. Jackel, C.; Kast, P.; Hilvert, D., Protein design by directed evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153-73.
108. Chen, K.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (12), 5618-22.
109. Pritchard, L.; Corne, D.; Kell, D.; Rowland, J.; Winson, M., A general model of error-prone PCR. *J. Theor. Biol.* **2005**, *234* (4), 497-509.
110. Stemmer, W. P., Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994**, *370* (6488), 389-91.
111. Sen, S.; Venkata Dasu, V.; Mandal, B., Developments in directed evolution for improving enzyme functions. *Appl. Biochem. Biotechnol.* **2007**, *143* (3), 212-23.
112. Li, T.; Bonkovsky, H. L.; Guo, J. T., Structural analysis of heme proteins: implications for design and prediction. *BMC Struct. Biol.* **2011**, *11*, 13.
113. Munro, A. W.; Girvan, H. M.; McLean, K. J., Cytochrome P450 – redox partner fusion enzymes. *Biochim. Biophys. Acta* **2007**, *1770* (3), 345-59.

114. Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H., Enzymatic assembly of carbon-carbon bonds via iron-catalysed sp(3) C-H functionalization. *Nature* **2019**, *565* (7737), 67-72.
115. Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H., Enantiodivergent alpha-amino C-H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141* (25), 9798-9802.
116. Liu, Z.; Qin, Z. Y.; Zhu, L.; Athavale, S. V.; Sengupta, A.; Jia, Z. J.; Garcia-Borras, M.; Houk, K. N.; Arnold, F. H., An enzymatic platform for primary amination of 1-aryl-2-alkyl alkynes. *J. Am. Chem. Soc.* **2022**, *144* (1), 80-85.
117. Dunham, N. P.; Arnold, F. H., Nature's machinery, repurposed: expanding the repertoire of iron-dependent oxygenases. *ACS Catal.* **2020**, *10* (20), 12239-12255.
118. Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H., Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J. Am. Chem. Soc.* **2003**, *125* (44), 13442-50.
119. Li, Z.; Jiang, Y.; Guengerich, F. P.; Ma, L.; Li, S.; Zhang, W., Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *J. Biol. Chem.* **2020**, *295* (3), 833-849.
120. Di Nardo, G.; Gilardi, G., Optimization of the bacterial cytochrome P450 BM3 system for the production of human drug metabolites. *Int. J. Mol. Sci.* **2012**, *13* (12), 15901-24.
121. Alvizo, O.; Nguyen, L. J.; Savile, C. K.; Bresson, J. A.; Lakhapatri, S. L.; Solis, E. O.; Fox, R. J.; Broering, J. M.; Benoit, M. R.; Zimmerman, S. A.; Novick, S. J.; Liang, J.; Lalonde, J. J., Directed evolution of an ultrastable carbonic anhydrase for

highly efficient carbon capture from flue gas. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (46), 16436-41.

122. Gould, S. M.; Tawfik, D. S., Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* **2005**, *44* (14), 5444-52.

123. Yang, K. K.; Wu, Z.; Arnold, F. H., Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **2019**, *16* (8), 687-694.

124. Wu, Z.; Kan, S. B. J.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H., Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (18), 8852-8858.

125. Andreini, C.; Rosato, A., Structural bioinformatics and deep learning of metalloproteins: recent advances and applications. *Int. J. Mol. Sci.* **2022**, *23* (14), 7684.

126. Lippow, S. M.; Moon, T. S.; Basu, S.; Yoon, S. H.; Li, X.; Chapman, B. A.; Robison, K.; Lipovsek, D.; Prather, K. L., Engineering enzyme specificity using computational design of a defined-sequence library. *Chem. Biol.* **2010**, *17* (12), 1306-15.

127. Rui, J.; Zhao, Q.; Huls, A. J.; Soler, J.; Paris, J. C.; Chen, Z.; Reshetnikov, V.; Yang, Y.; Guo, Y.; Garcia-Borras, M.; Huang, X., Directed evolution of nonheme iron enzymes to access abiological radical-relay C(sp³)-H azidation. *Science* **2022**, *376* (6595), 869-874.

Chapter 2. The origins of tight metal binding to UFsc, a *de novo* designed protein that has extraordinary affinity for metal ions

Work from this chapter has led to the following publications:

Kulesha, A.; Yoon, J. H.; Chester, C.; D'Souza, A.; Costeas, C.; Makhlynets, O. V.

Contributions of primary coordination ligands and importance of outer sphere interactions in UFsc, a *de novo* designed protein with high affinity for metal ions. *J. Inorg. Biochem.* 2020, **212**, 111224. – Text, figures, tables and schemes from this publication were reproduced with the permission from Elsevier Inc., copyright 2020.

Yoon, J. H.*; **Kulesha, A. V.***; Lengyel-Zhand, Z.; Volkov, A. N.; Rempillo, J. J.; D'Souza, A.; Costeas, C.; Chester, C.; Caselle, E. R.; Makhlynets, O. V. Uno Ferro, a *de novo* designed protein, binds transition metals with high affinity and stabilizes semiquinone radical anion. *Chemistry* 2019, **25** (67), 15252-15256. – A.K. performed isothermal titration calorimetry determination of protein-metal binding parameters.

Abstract. *De novo* protein design has been successfully applied for the development of artificial proteins that mimic the properties of natural metalloenzymes. The introduction of one mutation in the metal-binding site and four mutations in the channel leading to the active site allowed us to create 4G-UFsc, a single-metal binding protein with the highest zinc affinity reported to date for designed proteins that do not contain cysteines. In this work, we have shown that metal binding is sensitive to even minor changes in the coordination environment and can be manipulated into binding of various transition metals. Using site-directed mutagenesis, we have also established the location of metal-binding site in UFsc after the disruption of the symmetrical binding site of DFsc.

2.1. Introduction

Metalloenzymes are important in all aspects of physiology as they facilitate many biochemical reactions important for the survival and operation of living organisms. Even with the limited set of available metal ions and proteinogenic amino acids suitable for metal coordination, a diverse range of functions and activities is observed.¹⁻⁶ This suggests that the formation of the protein-metal complex leads to the specific arrangement of the active site which facilitates the specific function. Many natural enzymes have been studied to better understand the relationship between the structure of the metal-binding site and the protein's activity. The characterization of natural proteins brings a lot of challenges with protein isolation, purification and analysis due to the large size of molecules, their often poor stability under *in vitro* conditions and overall complexity. Protein engineering can be applied to simplify the analysis process by transferring the active site from a natural protein to a smaller scaffold. *De novo* protein design not only provides a great opportunity to explore protein functions in small well-controlled scaffolds but also tests our understanding of the protein structure and principles of protein design.

One of the best examples of *de novo* designed proteins created to mimic the properties of naturally existing enzymes is DFsc, single-chain Due Ferri. This protein was designed to model the active site of non-heme diiron/dimanganese enzymes such as manganese catalase, ribonucleotide reductase, methane monooxygenase, toluene oxygenase, desaturases, and ferroxidases.⁷⁻¹¹ These metalloproteins tend to differ in size and folding; however, they all share a similar structure of the active site which consists of a Glu-Xxx-Xxx-His motif inserted in a four-helix bundle. DeGrado and co-

workers created a family of antiparallel four-stranded coiled coils with the replication of the first and second metal coordination spheres observed in the aforementioned enzymes through the applications of retrostructural analysis, rational design and computational design.¹²⁻¹⁵ They successfully created the initial designs (DF1, a homodimer of helix-loop-helix motifs, and DFtet, a heterotetrameric four-helix bundle) which were shown to catalyze the oxidation of 4-aminophenol by dioxygen. Due to the modular composition, it was challenging to correlate the activity of the assembled metalloprotein with the structure of its active site.

A 114-residue protein, DFsc, was created using the helical backbone of DFtet as a template to facilitate further structural analysis of the protein (Figure 2.1). The metal binding site of the designed protein includes two histidine (H77 and H107) and four glutamate (E11, E44, E74 and E104) residues as first-shell ligands which are buried in the core of the four-helix bundle and stabilized by hydrogen bond interactions with the second-shell ligands (D72, D102, Y18 and Y51).

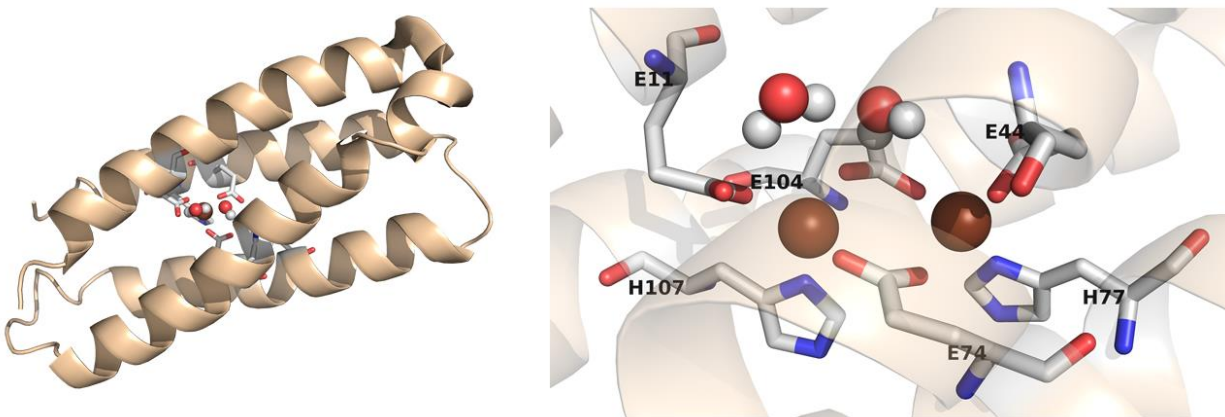


Figure 2.1. DFsc is a four-helix bundle with a divalent metal binding site (left). The metal binding site consists of two histidine and four glutamic acid residues (right). Images generated based on the protein NMR structure (PDB ID 2HZ8).

Most DF designs consist of multiple peptides that are later assembled in the four-helix bundle. The modular composition makes it challenging to introduce point mutations in the active site. Compared to other DF designs, DFsc is a single-chain protein suitable for the introduction of asymmetric changes in this protein. This made DFsc a great scaffold for further protein engineering¹⁶ and opened a possibility for directed evolution. DFsc is also a stable and highly soluble protein that binds divalent metals with 2:1 stoichiometry and can be expressed in bacterial cells allowing for easy manipulation.

As part of our investigation of the role of metal in radical stabilization by DFsc, we have created a series of mutants with one metal-binding site. To evaluate the activity of the one-metal-binding protein, we have mutated a bridging ligand Glu104 to histidine which resulted in UFsc, Uno Ferro, a protein with a single metal-binding site. UFsc was used as a scaffold to create 4G-UFsc by introducing four mutations (A10G, A14G, A43G and A47G) to open the channel that leads to the active site for easier substrate access (Figure 2.2). The elimination of the bridging ligand and the widening of the channel leading to the active site could have a major impact on the metal-binding affinity of the designed proteins. We have analyzed the affinity of the designed proteins for selected transition metals, as well as established the location of the single-metal-binding site. UFsc showed a remarkable affinity to divalent transition metals with the picomolar affinity to zinc (II). To evaluate the contributions of the residues in the metal coordination sphere to metal binding, we have created a library of UFsc mutants and measured their metal-binding affinity using isothermal titration calorimetry and competition titrations with fluorescent dye Mag-Fura 2. Our results show that metal binding in UFsc is sensitive to

even minor changes in the coordination environment and that the metal-coordination sphere can be adapted to binding of different transition metals. The results of this work will guide future efforts in metalloprotein design.

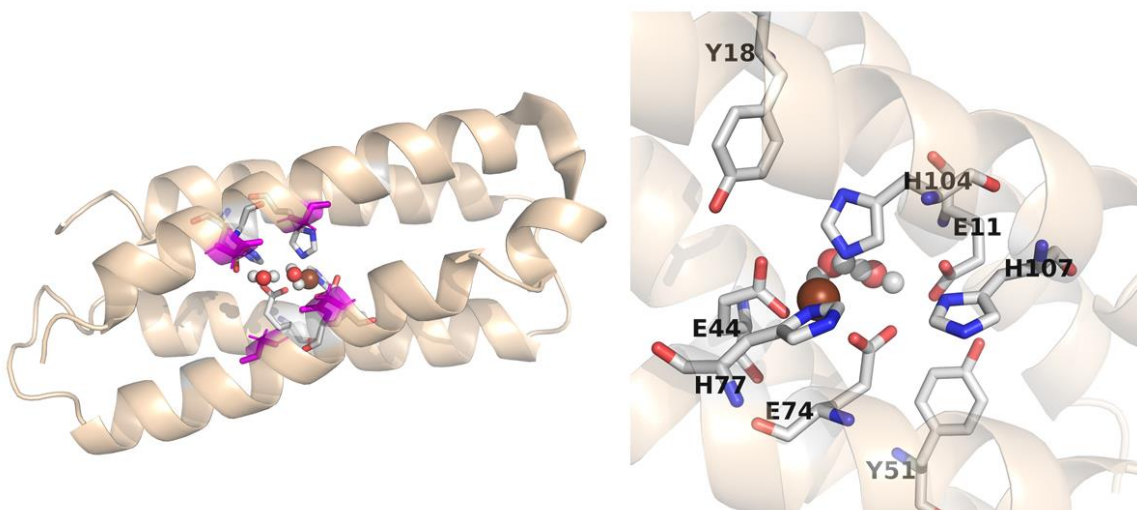


Figure 2.2. UFsc and 4G-UFsc were created by introducing E104H mutation in the metal binding site (right). 4G-UFsc has wider channel (left) due to additional alanine-to-glycine mutations (DFsc PDB ID 2HZ8 was used as scaffold for mutagenesis using PyMol)

2.2. Results and Discussion

2.2.1 *Design and characterization of metal affinity of UFsc, a single-metal binding protein*

DFsc, Due Ferri single chain, is a model protein that mimics the structure of the metal-binding site found in many natural diiron/dimanganese proteins. DFsc was created by placing a Glu-Xxx-Xxx-His motif inside the hydrophobic four-helix bundle using retrostructural analysis, rational design and computational methods. As this protein was intended to model natural enzymes, it showed affinity to many divalent metals as well as diverse activities.

Natural enzymes efficiently utilize organic radicals for catalysis. Even though radicals are highly reactive species and can be harmful to biological molecules, proteins have evolved to control and, stabilize radicals to perform multiple functions. Radical stabilization by proteins is a well-documented strategy, however, the mechanisms underlying this stabilization are poorly understood.¹⁷⁻¹⁹ DFsc mimics the active site of natural metalloproteins that can generate organic radicals and present a great model for further studies. It has been shown previously that zinc (II)-bound DFsc derivatives can stabilize the semiquinone radical anion, the one-electron oxidized intermediate in the redox triad 3,5-di-tert-butylcatechol/semiquinone/o-benzoquinone (QH₂/SQ•/Q).²⁰ Intrigued by the importance of the metal in radical stabilization, we set out to explore the role of metal in radical stabilization by model systems and also test the necessity of two metal ions for this process. As part of our investigation of radical production by DFsc, we have converted this two-metal-binding protein, into UFsc, a single metal-binding protein.

The metal-binding site in DFsc possesses high symmetry and can accommodate two metal ions. Two metal-binding sites are linked with two bridging oxygen-containing ligands E104 and E74 (Figure 2.3 A). The introduction of a E104H mutation resulted in the elimination of one metal-binding site as confirmed by cobalt titration (Figure 2.3 D). Histidine is a nitrogen-containing monodentate ligand, meaning it can coordinate to only one metal. Given that histidine can exist in two tautomeric forms, it is hard to predict the position of the metal-binding site (Figure 2.3 B).

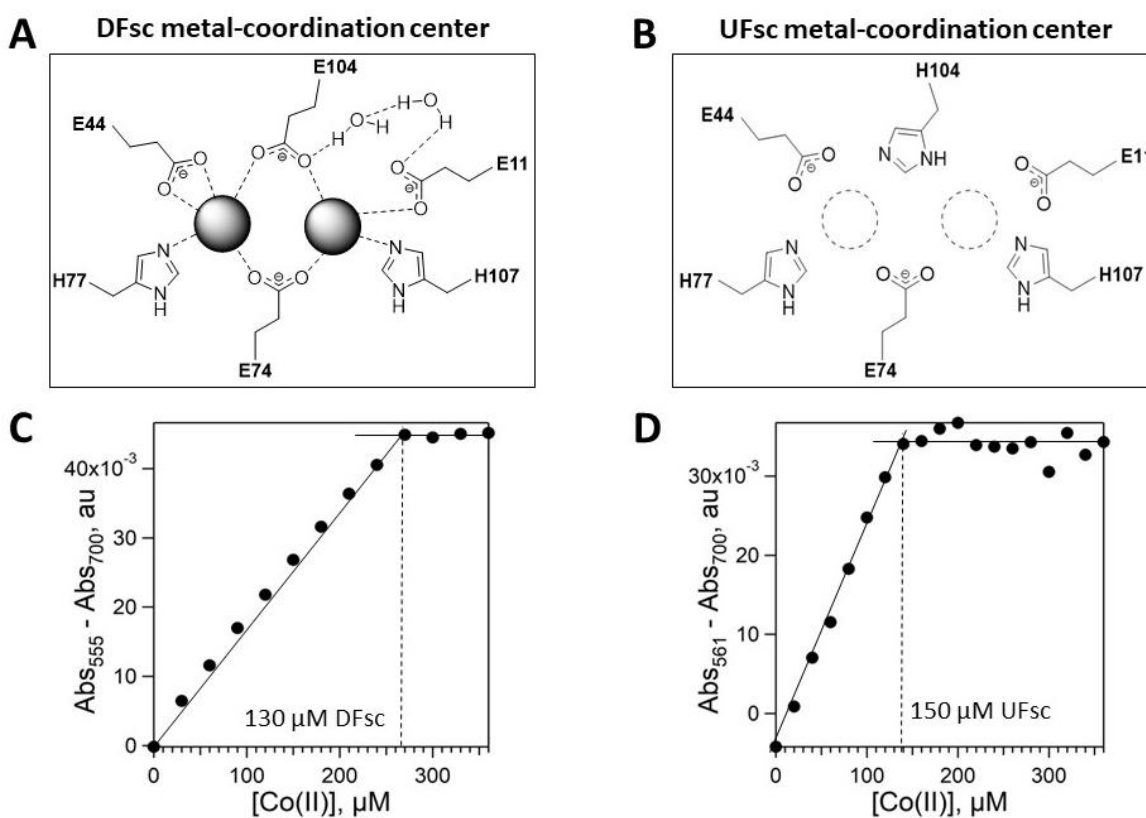


Figure 2.3. Metal binding sites of DFsc (A) and UFsc (B). Cobalt titration experiment showing the metal-to-protein ratio in DFsc (C) and UFsc (D). The 130 μM DFsc (C) and 150 μM UFsc (D) were titrated with aliquots of 1 mM cobalt (II) chloride solution. The solutions were prepared in 25 mM HEPES (pH 7.6), 100 mM NaCl and incubated at 25 $^{\circ}\text{C}$.

The introduction of four mutations in the channel leading to the active site has been shown to improve substrate binding in DF proteins.^{15, 21} We have introduced A10G, A14G, A43G and A47G mutations to facilitate substrate access in our radical stabilization studies (Figure 2.4). As the introduction of a mutation in the metal coordination sphere and in the region leading to the metal-binding site can have a drastic effect on the affinity to selected metals, we evaluated the effect of introduced mutations on metal binding to DFsc.

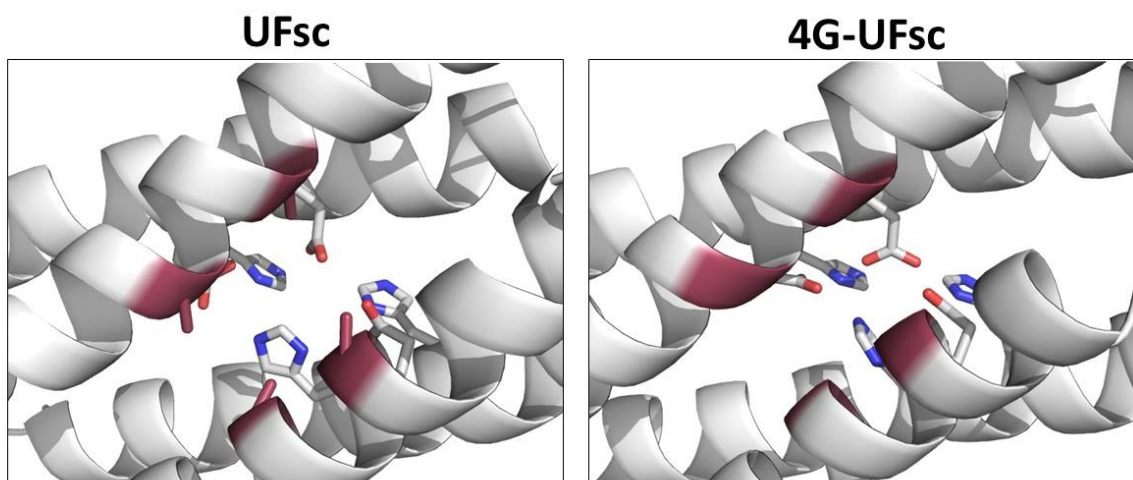


Figure 2.4. The introduction of four Ala-to-Gly mutations widens the channel leading to the active site (PDB 2HZ8 was used as a scaffold for mutagenesis).

Several methods, such as equilibrium dialysis, spectroscopic assays, competition titration and isothermal titration calorimetry, have been used to study the metal affinity of different proteins.²²⁻²⁶ Isothermal titration calorimetry (ITC) has become a versatile tool in protein analysis due to the direct label-free measurement of the enthalpy change for a reaction and the ability to provide thermodynamic parameters from a single experiment. This method also allows us to determine the stoichiometric ratio of molecules to ligands

(number of binding sites), dissociation constant (K_d) and the change in enthalpy upon molecule-ligand binding. As ITC measures the change in heat upon molecule interaction, it is important to ensure low metal binding to the components of the buffer. For this purpose, we performed all experiments in HEPES buffer which was shown to have a negligible metal affinity.²⁷ The ITC analysis confirmed the presence of only one metal-binding site in UFsc. We have also observed the improved binding to nickel (II) and zinc (II) metal ions (Table 2.1).

Table 2.1. Dissociation constants (in nM) determined by ITC for metal complexes of DF family proteins.

		Ni ²⁺	Co ²⁺	Mn ²⁺	Zn ²⁺
DFsc	K _{d1}	2,090 ± 230	56 ± 3	197 ± 37	109 ± 34
	K _{d2}	10,400 ± 1,400	133 ± 10	228 ± 37	213 ± 14
UFsc		112 ± 28	269 ± 116	173 ± 56	0.076 ± 0.013 ^[a]
4G-UFsc		220 ± 37	67 ± 4	9,250 ± 520	0.030 ± 0.004 ^[a]

[a] Dissociation constant determined by ITC competition experiments with triethylenetetramine (TETA).

Zinc (II) complexes are expected to have high stability following the Irving-Williams series. This implies high protein-metal affinity which might be hard to determine using traditional methods. At the same time, zinc ions are unstable at physiological pH and tend to precipitate even in slightly basic buffers. To overcome both limitations, we have tested a few potential zinc chelators. The main idea is that the zinc-chelator complex will be stable under the reaction conditions and, as protein is expected

to have a high affinity to zinc, we could determine the thermodynamic parameters of metal binding from the competition experiment. Triethylenetetramine (TETA) is an organic molecule and a copper chelator used in the treatment of Wilson's disease.²⁸ The structure of TETA contains four nitrogen atoms, which allows this molecule to coordinate to other transition metals, such as zinc (II) (Figure 2.5 A and B).

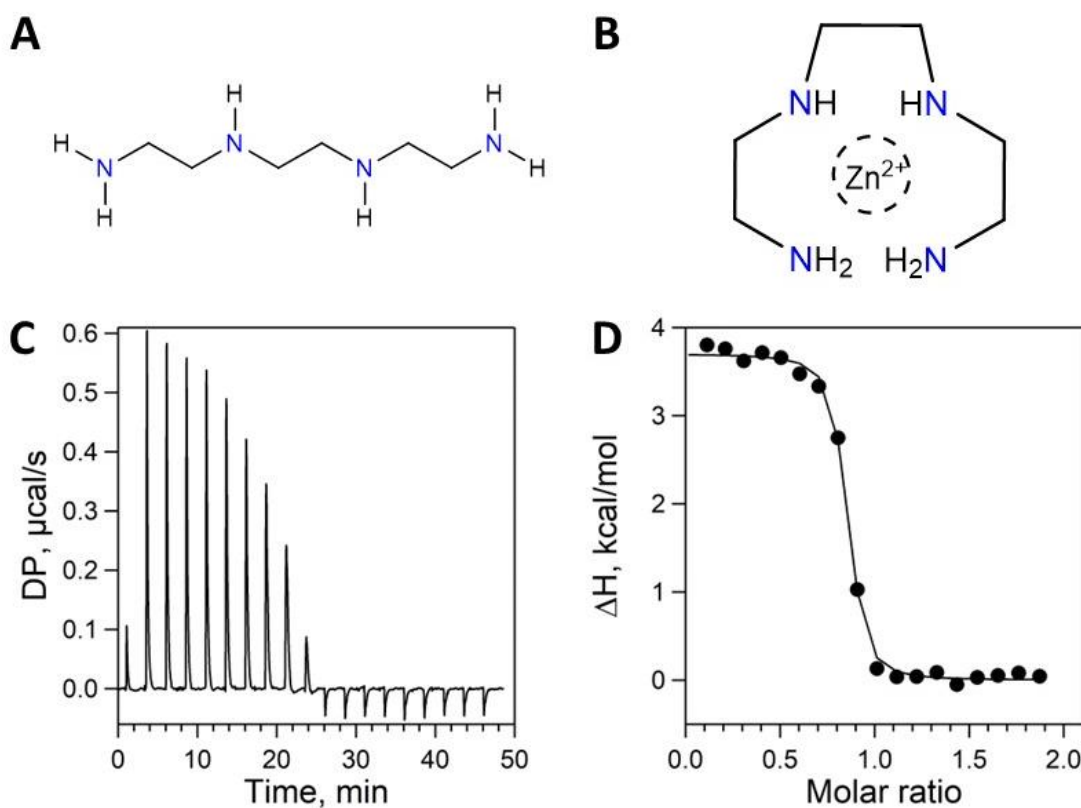


Figure 2.5. **A.** The structure of triethylenetetramine (TETA). The potential ligands for metal coordination are highlighted in blue. **B.** Proposed coordination configuration in TETA-zinc (II) complex. **C and D.** Baseline- and offset-corrected ITC titration curve and binding isotherm, respectively, for the titration of 200 μM zinc (II) with 2 mM TETA at pH 7.6. The dissociation constant for the TETA-zinc (II) complex was used for further determination of protein-zinc (II) dissociation constants in competition titration experiments.

Zinc binding by TETA was characterized using ITC and showed that the chelator binds metal with nanomolar affinity (number of binding sites – 0.777 ± 0.005 , $K_d = 173 \pm 65$ nM, $\Delta H = 3.57 \pm 0.06$ kcal/mol). TETA presented a great candidate for the chelator as (1) it binds metal with 1:1 stoichiometry, (2) it forms a soluble complex with zinc ions and (3) it binds zinc with high affinity. To determine the affinity of UFsc to zinc (II), the zinc (II)-TETA complex was titrated with small aliquots of the protein. As the protein had a higher affinity to metal, it was able to pull out metal ions from its complex with a chelator and this reaction was monitored using ITC. In the competition experiment with TETA, UFsc showed remarkable affinity to zinc (II) ion with a K_d of 76 pM. The widening of the channel leading to the metal-binding site resulted in the improvement of the affinity to zinc and lowered the K_d to 30 pM (Table 2.1). Such tight metal binding is unprecedented for *de novo* designed proteins, the metal coordination spheres of which do not contain cysteine residues.^{29, 30}

Isothermal titration calorimetry is a method sensitive to even minor changes in the reaction mixture and the competition experiments introduce a lot of complexity. As the protein's affinity to zinc (II) was measured to be extraordinarily high, we decided to confirm the results obtained by ITC competition titration using another method. The competition titration with the fluorescent dye Mag-Fura 2 is a simple absorbance-based titration experiment that allows for the determination of dissociation constants. Upon addition of zinc (II) metal, the decrease in the dye's absorbance at 366 nm can be observed and correlated to the concentration of dye-metal complex. The data is later analyzed using the competition titration model, where the protein and the dye compete for the metal ion in the solution, assuming a 1:1 ratio for each complex. The dissociation

constant for 4G-UFsc and zinc (II) was measured to be 37.1 ± 19.3 μM in the competition titration experiment with the dye (Figure 2.6), while the K_d measured by ITC was 30.0 ± 4.0 μM . These results show the reliability of results obtained using the developed in our lab ITC competition method. The affinity to zinc (II) metal to all mutants produced in this work was quantified using both methods.

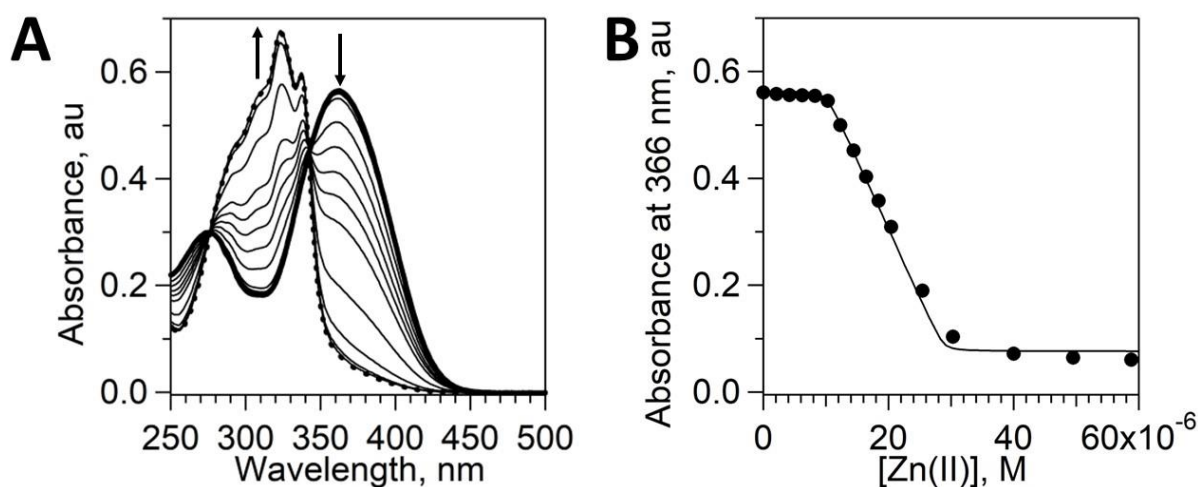


Figure 2.6. Competition titration of 4G-UFsc and Mag-Fura 2 with ZnCl₂. (A) Absorbance spectra of Mag-Fura 2 (18.7 μM) and 4G-UFsc (10.1 μM) titrated with ZnCl₂. The coordination of metal reduces absorbance intensity of the dye at 366 nm. (B) Absorbance at 366 nm as a function of metal concentration was fit using DYNAFIT to determine dissociation constant of zinc (II)-4G-UFsc complex.

2.2.2 The importance of His77 for metal binding by 4G-UFsc

The metal coordination site of DFsc consists of two bridging residues (Glu), two oxygen-containing monodentate ligands (Glu) and two nitrogen-containing ligands (His). The substitution of the bridging Glu104 to a monodentate His ligand resulted in the elimination of one binding site but also improved affinity to transition metals. To

characterize protein-metal binding, we have created a set of mutants with substitutions at each position in the metal coordination sphere and measured their affinity to different transition metals.

To evaluate the importance of nitrogen-containing ligands, we have substituted His77 and His107 with glutamine. Originally, DFsc was meticulously designed to assure proper folding of the four-helix bundle where second shell interactions aided structure stabilization.^{15, 31} H77 and H107 inside the hydrophobic core required proper positioning and the introduction of D73 and D103 stabilized H77 and H107 through the hydrogen bonding (Figure 2.7).

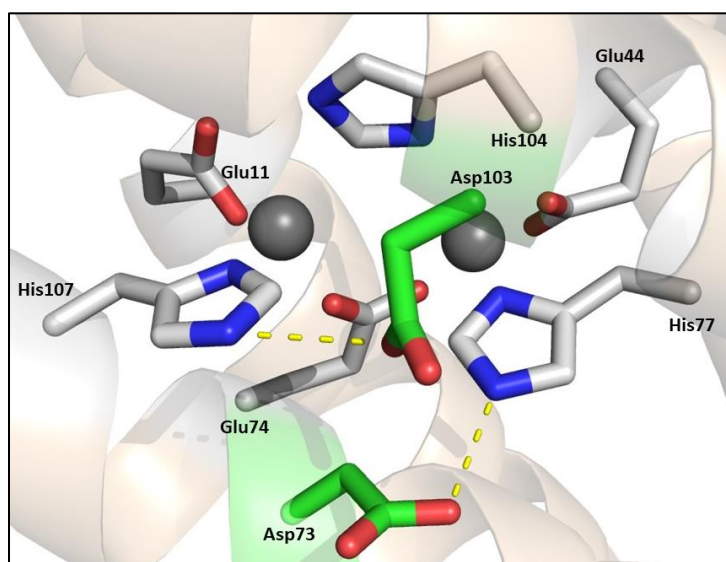


Figure 2.75. The structure of UFsc metal binding site. D73 and D103 (shown in green) provide stabilization and ensure proper orientation of H77 and H107, respectively. The grey spheres represent the locations of possible metal-binding sites in UFsc. The figure was prepared using PDB 2HZ8 as a scaffold.

As the introduction of more hydrophobic residues at positions 77 and 107 would require additional changes to ensure the stability of the four-helix bundle, glutamine

presented a great compromise, allowing to retain the size of the metal coordination sphere, remove the preferred ligand for metal coordination but to provide proper interactions for structure stabilization. First, we have tested the impact of the introduced mutations on the protein structure (Figure 2.8). To evaluate the changes in the structure of tested proteins, we have used circular dichroism (CD) spectroscopy, a method that allows us to evaluate the protein's secondary structure and monitor protein folding.³² DFsc proteins are four-helix bundles and are expected to show prominent helical structures. The CD traces for both mutants showed minor deviations from those of 4G-UFsc confirming that introduced substitutions had little to no effect on protein structure. Similar to the original protein, mutants undergo metal-induced folding observed as the increased helicity upon metal binding.

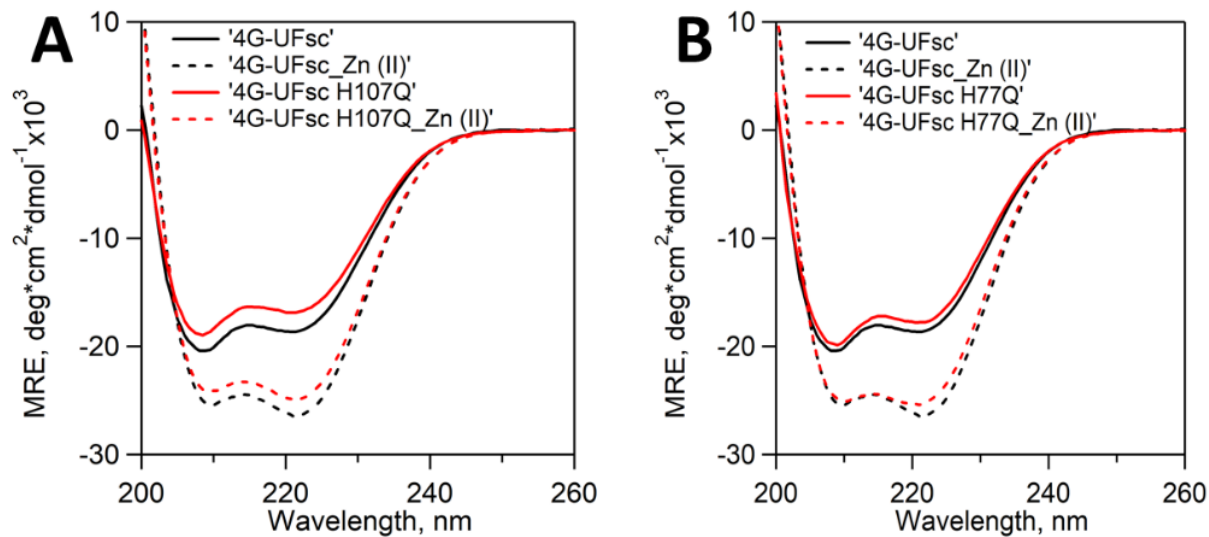


Figure 2.8. CD spectroscopy analysis of 4G-UFsc H107Q (A) and H77Q (B) showed that the mutation did not influence the degree of protein helicity. Proteins were analyzed at 20 μ M in 5 mM HEPES, 20 mM NaCl (pH 7.6). Protein-metal complex was prepared by mixing 20 μ M protein with 200 μ M zinc (II) chloride.

The analysis of the divalent metal affinity of 4G-UFsc H107Q and H77Q provided deeper insight into the localization of the single metal-binding site. H77Q mutant showed no binding to manganese (II) or cobalt (II) and the affinity to nickel (II) reduced drastically indicating the importance of His77 for metal binding (Table 2.2, Figures S 4 and S 5).

Table 2.2. Thermodynamic parameters of divalent metal binding to 4G-UFsc H107Q and H77Q mutants determined by ITC.

Metal	Parameters	4G-UFsc	4G-UFsc H107Q	4G-UFsc H77Q
Mn (II)	N	0.959 ± 0.005	No binding	No binding
	Kd, nM	9,250 ± 521		
	ΔH, kcal/mol	-2.57 ± 0.02		
Co (II)	N	0.836 ± 0.001	0.559 ± 0.002	No binding
	Kd, nM	67 ± 4	6.12 ± 4.19	
	ΔH, kcal/mol	-9.21 ± 0.02	-7.68 ± 0.06	
Ni (II)	N	0.763 ± 0.004	0.775 ± 0.013	
	Kd, nM	220 ± 37	8,240 ± 1,260	> 100 μM
	ΔH, kcal/mol	-7.61 ± 0.08	-3.65 ± 0.12	
Zn (II)	N	0.817 ± 0.000	1.060 ± 0.003	
	Kd, nM	0.030 ± 0.004	0.201 ± 0.025	N/A
	ΔH, kcal/mol	-12.00 ± 0.10	-10.70 ± 0.12	

The H77Q mutant showed binding to zinc (II) ion, although its affinity was the lowest among all tested mutants. The affinity to zinc was measured in the competition experiment where the protein was extracting metal from zinc (II)-TETA complex. In the

case of the H77Q mutant, the protein could not effectively bind the metal, suggesting that its affinity to zinc (II) is lower than that of the chelator (lower than 173 nM). These results suggest the importance of H77 for metal binding in 4G-UFsc.

On the other hand, H107Q substitution improved cobalt (II) binding by decreasing the dissociation constant about 10 times (67 nM vs 6.12 nM). Interestingly, this interaction is also characterized by different stoichiometry: one metal ion links two protein molecules (Figure 2.9). This effect is surprising and needs further investigation.

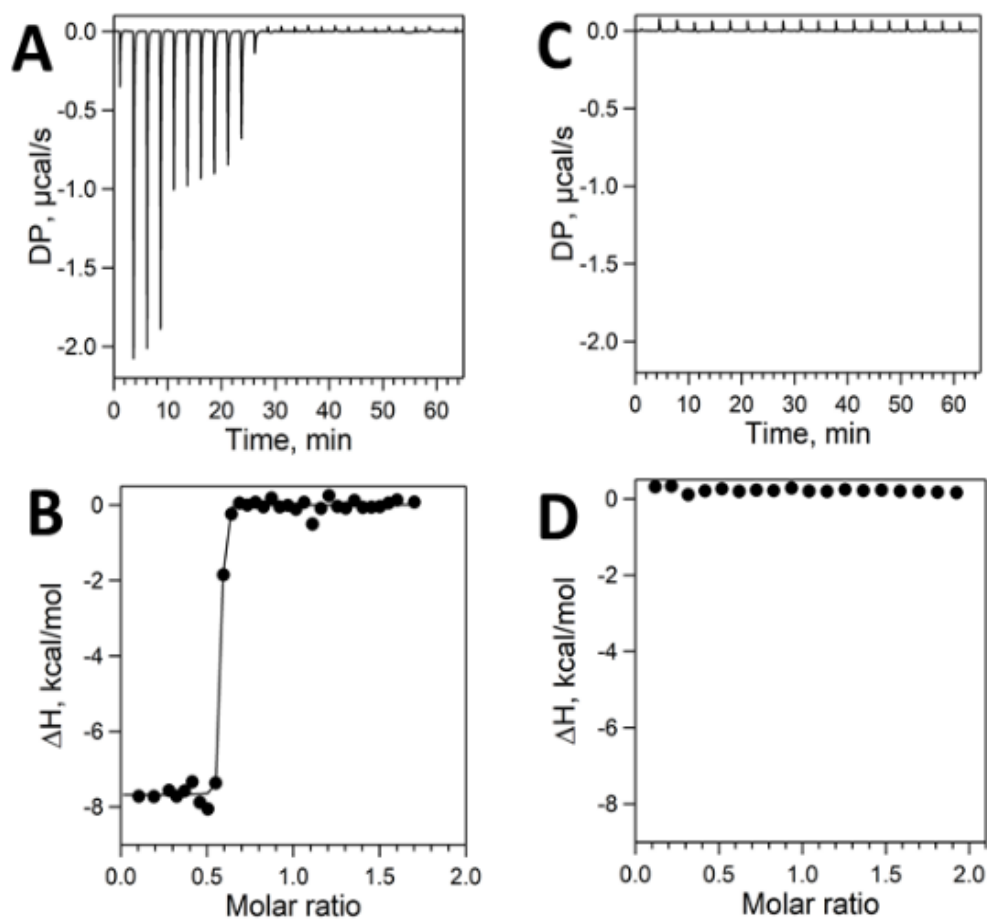


Figure 2.9. ITC traces for the titration of H107Q (A) and H77Q (C) with cobalt (II) and the binding isotherms (B and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 25 mM HEPES, 100 mM NaCl, pH 7.6 at 25 °C.

Overall, the mutation at position 77 significantly reduced protein's affinity to divalent metals, thus we hypothesized that it was one of the crucial residues in the 4G-UFsc binding site. The alteration at position 107 had an ambiguous effect on metal binding, probably due to the changes in protein folding.

2.2.3. Alteration of the bridging Glu74 leads to the decrease in metal binding affinity

The active site of the original DFsc is symmetrical, thus a mutation in the bridging residue which facilitated the creation of UFsc can favor the formation of a single metal-binding site at either position. We have created a set of mutants at position 74, another bidentate ligand in DFsc, to explore the importance of oxygen-containing ligands at that position.

The bridging ligand Glu74 in the original DFsc is coordinated to both metal ions. The mutations at this position will not provide us with the information about the location of the binding site but can shed some light on the role of oxygen-containing ligand in the metal affinity. The introduction of the E74D mutation allowed us to increase the size of the coordination sphere. The E74Q mutation resulted in the removal of charge from the metal-coordinating ligand and Glu-to-His substitution probes the importance of the oxygen-containing ligand at this position (Figure 2.10).

First, we were interested in the effect introduced changes have on the structure of 4G-UFsc. All mutants were analyzed using CD spectroscopy to examine the degree of protein folding (Figure 2.11). 4G-UFsc E74Q and E74H mutants showed higher degree of helicity in both apo- and zinc-bound forms compared to the original protein. For both mutants the ratio of $\theta_{222 \text{ nm}}/\theta_{208 \text{ nm}}$ was higher than one suggesting that those

proteins adopt coiled-coil type of structure regardless of the presence of metal ion.

Unlike all other mutants prepared in this work, E74H did not show metal-induced protein folding upon addition of zinc (II).

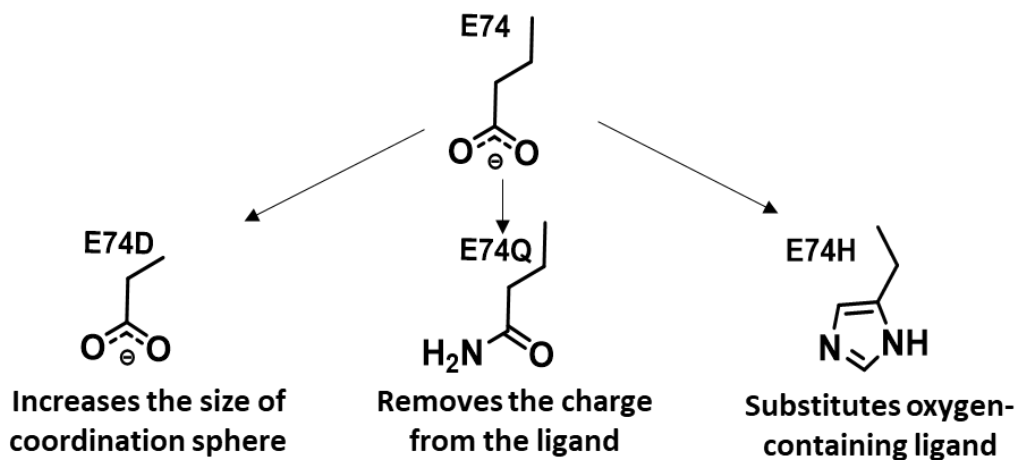


Figure 2.10. The mutations introduced at position 74 allowed to test the importance of the size, charge and identity of this ligand.

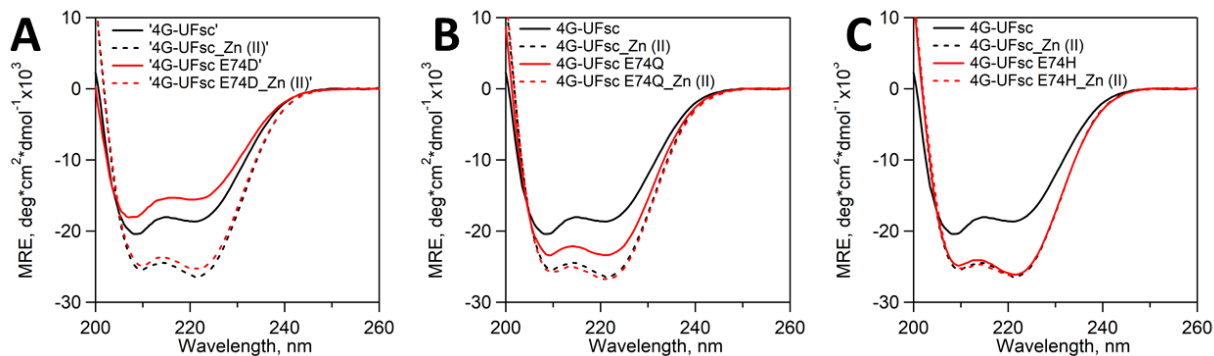


Figure 2.11. CD spectroscopy analysis showed that mutations at the position 74 (A – E74D, B – E74Q, C – E74H) have different effect on protein structure. All apoprotein spectra were recorded at 20 μ M and zinc-bound spectra were recorded at 20 μ M protein with 200 μ M zinc (II) chloride concentration in 5 mM HEPES, 20 mM NaCl (pH 7.6) at room temperature.

Interestingly, Glu-to-Asp mutation at this position had a negative effect on the structure of protein leading to lower degree of helicity. This mutation increases the size of the metal-binding cavity inside the four-helix bundle, which probably results in the destabilization of the protein core.

Using isothermal titration calorimetry, we have tested the effect of the mutations at position 74 on the protein-metal binding. As 4G-UFsc showed high zinc affinity, we have performed competition titrations with zinc (II)-TETA complex. Zinc (II) is known for its high affinity to cysteine (S) and histidine (N) ligands even though it can also adopt a coordination environment containing glutamate or aspartate residues (O).³³ For histidine at position 74, the dissociation constant increased about 10 times but still remained in the picomolar range (Table 2.3, Figure 2.12).

Table 2.3. Thermodynamic parameters of zinc binding to E74 mutants determined by ITC competition titration with triethylenetetramine.

Metal	Parameters	4G-UFsc	4G-UFsc E74D	4G-UFsc E74Q	4G-UFsc E74H
Zn (II)	N	0.817 ± 0.000	1.050 ± 0.007	0.860 ± 0.024	0.634 ± 0.005
	Kd, nM	0.030 ± 0.004	4.19 ± 0.25	21.2 ± 3.2	0.281 ± 0.063
	ΔH, kcal/mol	-12.00 ± 0.10	-6.05 ± 0.12	-0.868 ± 0.212	-5.05 ± 0.13

As protein showed a high degree of folding in its apo form, E74H mutation leads to metal-independent prearrangement of the high-affinity metal-binding site. Even though the dissociation constant increased, it is interesting that the high-affinity site can be prearranged without extensive structural change. E74D and E74Q mutations had a

negative effect on zinc binding increasing dissociation constants to the nanomolar range.

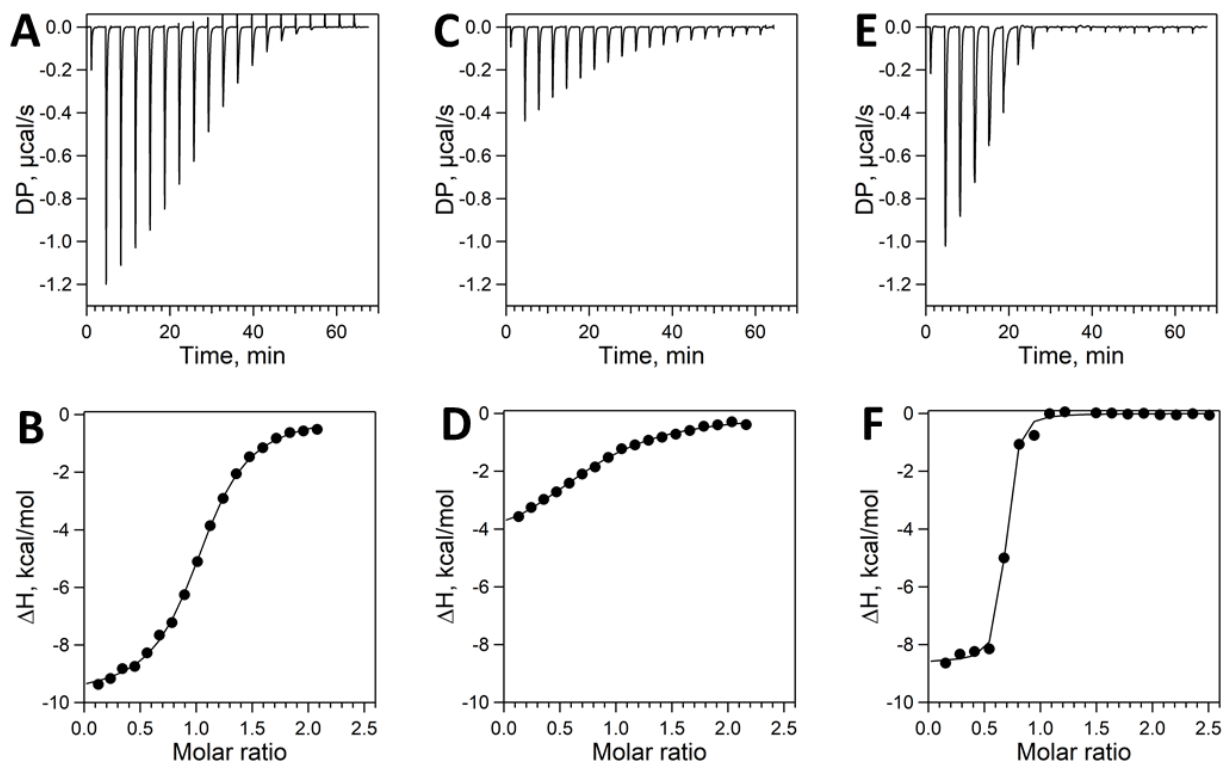


Figure 2.12. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the competition titration of zinc (II) – triethylenetetramine (TETA) complex with Glu74 mutants. The complex of 90 μM zinc (II) with 200 μM TETA was titrated with: A – 1 mM 4G-UFsc E74D (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 210 s); C – 1.04 mM 4G-UFsc E74Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); E – 1.34 mM 4G-UFsc (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 210 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively.

Mutations at position 74 did not show improvement in binding for any of the tested metals (Tables 2.3 and 2.4). Moreover, some mutations appeared to have a detrimental effect on metal coordination. Even though manganese prefers oxygen-containing ligands, both E74D and E74Q mutations led to the loss of metal-binding affinity, while 4G-UFsc E74H showed improved manganese (II) binding. Based on the shape of the isotherm, cobalt (II) binding to 4G-UFsc E74H is accompanied by significant structural changes which interfere with the proper data analysis (Figures S 1 – S 3). All E74 mutations led to the reduction or complete elimination of divalent ion affinity confirming an important role of size, charge and number of oxygen ligands glutamate residue can provide for the metal binding in 4G-UFsc.

Table 2.4. Thermodynamic parameters of divalent metal binding to E74 mutants determined by ITC.

Metal	Parameters	4G-UFsc	4G-UFsc E74D	4G-UFsc E74Q	4G-UFsc E74H
Mn (II)	N	0.959 ± 0.005	No binding	No binding	0.742 ± 0.017
	Kd, nM	9,250 ± 521			43,200 ± 6,410
	ΔH, kcal/mol	-2.57 ± 0.02			4.70 ± 0.32
Co (II)	N	0.836 ± 0.001	0.809 ± 0.005	No binding	
	Kd, nM	67 ± 4	880 ± 142		N/A
	ΔH, kcal/mol	-9.21 ± 0.02	-3.02 ± 0.04		
Ni (II)	N	0.763 ± 0.004	0.892 ± 0.009	0.762 ± 0.009	No binding
	Kd, nM	220 ± 37	7,500 ± 691	6,070 ± 709	
	ΔH, kcal/mol	-7.61 ± 0.08	-5.09 ± 0.08	-3.30 ± 0.08	

2.2.4. Mutations of the non-bridging glutamate ligands have a negative effect on metal binding

Two non-bridging glutamate residues E44 and E11 of DFsc interact with metals by providing oxygen ligands for metal coordination. Similar to the histidine residues in the metal's first coordination sphere, E44 and E11 need proper positioning for productive metal binding and form hydrogen bonds with corresponding tyrosine residues Y18 and Y51 from the second coordination sphere (Figure 2.13.). The introduction of the glutamate-to-glutamine mutation in UFsc and 4G-UFsc does not alter the size of the binding pocket but removes the charge from the ligand.

As the active site of the original DFsc protein is symmetrical, a mutation in the position of the bridging residue (E104H) can lead to the formation of a single metal-binding site at either position. Previously, we attempted to identify the binding site for Fe^{2+} using reported signature phenolate to ferric charge transfer transition at 520 nm between iron (II) and Tyr51.^{34, 35}

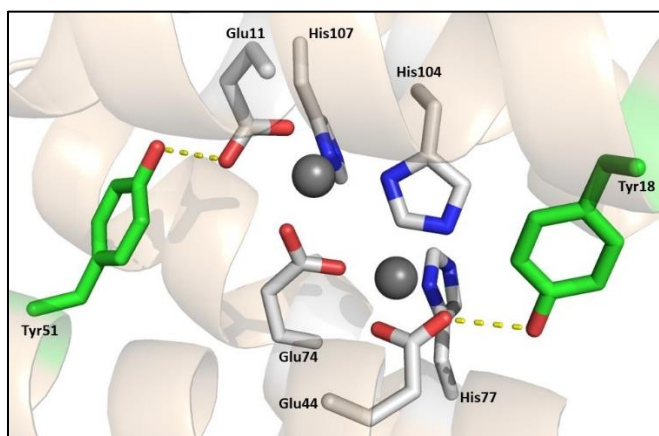


Figure 2.13. The structure of UFsc metal binding site. Y51 and Y18 (shown in green) provide stabilization and ensure proper orientation of E11 and E44, respectively. The grey spheres represent the locations of possible metal binding sites in UFsc. The figure was prepared using PDB 2HZ8 as a scaffold.

The absence of the expected absorbance peak suggested that iron (II) binds to the site formed but not limited to E44, H77 and E74.³⁶ To confirm the location of the metal-binding site, we have also prepared E44L mutant in 4G-UFsc scaffold which allowed to eliminate metal coordination completely. The introduction of this mutation required modification to the second coordination shell, thus we also introduced Y18F substitution to remove the hydrogen bond donor and ensure the stability of the four-helix bundle.

All mutated proteins showed a higher degree of helicity compared to the original design with the further metal-induced formation of the coiled-coil structure (Figure 2.14). Interestingly, the double mutant 4G-UFsc E44L/Y18F showed better folding compared to 4G-UFsc E44Q which might result from the increase in the hydrophobicity of the metal-binding site.

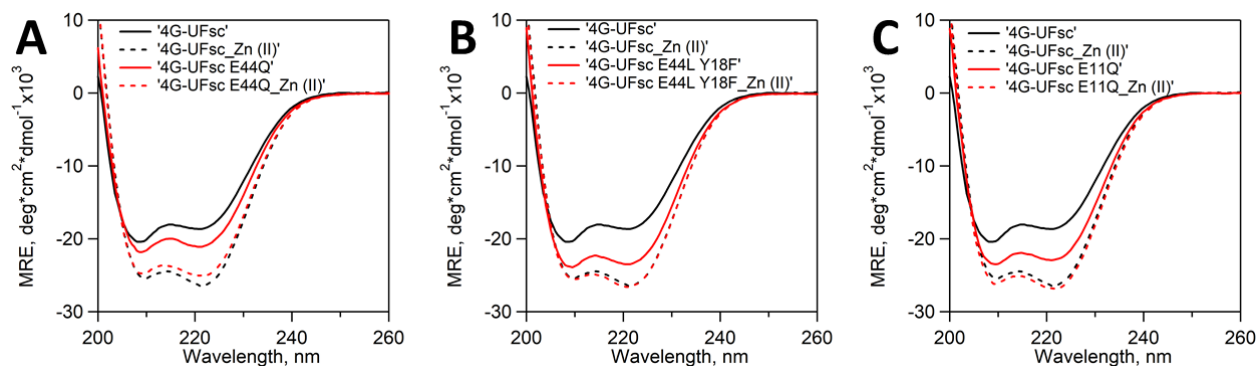


Figure 2.14. CD spectroscopy analysis of 4G-UFsc mutants showed an increase in the degree of helicity upon metal binding. The solution of 20 μM protein in 5 mM HEPES, 20 mM NaCl (pH 7.6) was used as an apoprotein sample. The metal-bound form was prepared by addition of 200 μM zinc (II) chloride.

Mutations at position 44 had a negative impact on protein-metal binding with the most detrimental effect on manganese (II) binding (Table 2.5). Observed effects are consistent with our conclusion that the higher affinity metal-binding site in 4G-UFsc includes residue E44. Cobalt (II) and nickel (II) metal binding events by 4G-UFsc E11Q mutant were accompanied by some additional exothermic events which made the obtained isotherms not suitable for the data analysis and the extraction of dissociation constants (Figure S 7 – S 8).

Table 2.5. Thermodynamic parameters of divalent metal binding to 4G-UFsc E44Q, E44L/Y18F and E11Q mutants determined by ITC (zinc was tested in the competition experiment with TETA).

Metal	Parameters	4G-UFsc	4G-UFsc E44Q	4G-UFsc E44L/Y18F	4G-UFsc E11Q
Mn (II)	N	0.959 ± 0.005	No binding		0.650 ± 0.012
	Kd, nM	9,250 ± 521		>100 µM	4,170 ± 832
	ΔH, kcal/mol	-2.57 ± 0.02			1.6 ± 0.05
Co (II)	N	0.836 ± 0.001	0.730 ± 0.005	0.312 ± 0.021	
	Kd, nM	67 ± 4	275 ± 61	7,650 ± 3,000	N/A
	ΔH, kcal/mol	-9.21 ± 0.02	-6.16 ± 0.11	-5.44 ± 0.61	
Ni (II)	N	0.763 ± 0.004	1.17 ± 0.10	0.737 ± 0.009	
	Kd, nM	220 ± 37	7,180 ± 2,250	1,420 ± 284	N/A
	ΔH, kcal/mol	-7.61 ± 0.08	-5.12 ± 0.86	-3.14 ± 0.09	
Zn (II)	N	0.817 ± 0.000	0.686 ± 0.009	1.31 ± 0.01	0.832 ± 0.003
	Kd, nM	0.030 ± 0.004	1.10 ± 0.21	3.06 ± 0.34	0.076 ± 0.017
	ΔH, kcal/mol	-12.00 ± 0.10	-8.42 ± 0.27	-2.41 ± 0.12	-9.39 ± 0.11

On the other hand, this mutation improved manganese (II) ion binding. Interestingly, the Mn (II) binding to 4G-UFsc E11Q is endothermic and entropically driven, compared to the exothermic reaction between 4G-UFsc and manganese (Figure 2.15). This suggests a major restructure in the water shell around the metal ion and an increase in coordination number, which is consistent with the manganese ion's preference for the larger number of coordinating ligands. According to the Irving-Williams series, manganese (II) forms the least stable complexes among all tested metals. The preference of this metal for the higher number of coordinating ligands makes it challenging to design efficient manganese binding proteins. Manganese binding to 4G-UFsc E11Q ($K_d = 4.2 \mu\text{M}$) is *on par* with some natural enzymes and better than reported designed proteins.^{24, 29, 37}

Notably, mutation at position 11 resulted in the same degree of affinity to zinc (II) as that of UFsc protein but the reaction was more exothermic. This result is consistent with the CD spectroscopy data and suggests that 4G-UFsc E11Q undergoes conformational changes upon metal binding which results in a more organized protein structure while UFsc adopts its fold in the apo form (Figure 2.14 C and Figure 2.16).

In DFsc, both oxygens of E44's carboxylic group participate in the metal coordination. We have established that this residue is important for metal binding in 4G-UFsc. We have observed that the mutations at this position led to an increase in protein helicity, compared to the original scaffold, and metal-induced folding was still observed (Figure 2.14.). As UFsc retains its folding in apo form and did not show metal-induced folding, we decided to investigate the impact of such mutations on its metal-binding activity.

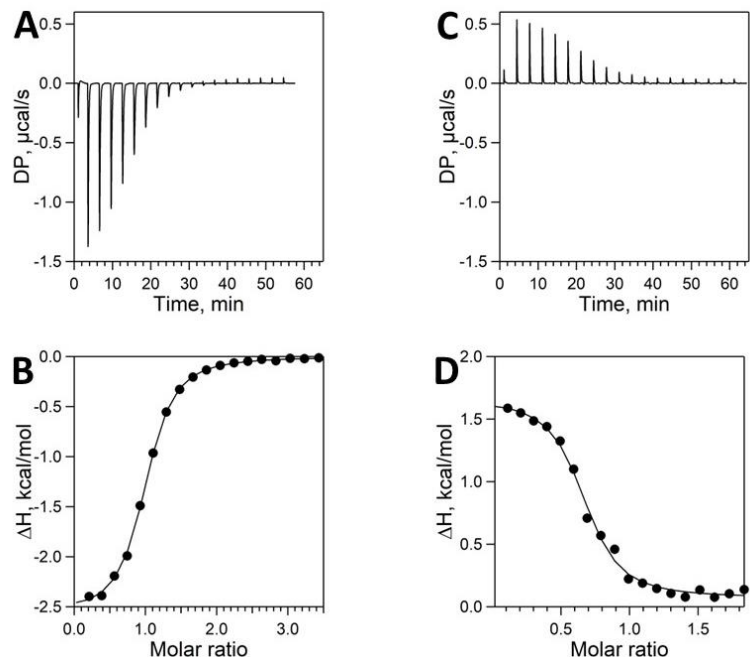


Figure 2.15. ITC traces for the titration of 4G-UFsc (A) and 4G-UFsc E11Q (C) with manganese (II) and the binding isotherms (B and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 25 mM HEPES, 100 mM NaCl, pH 7.6 at 25 °C.

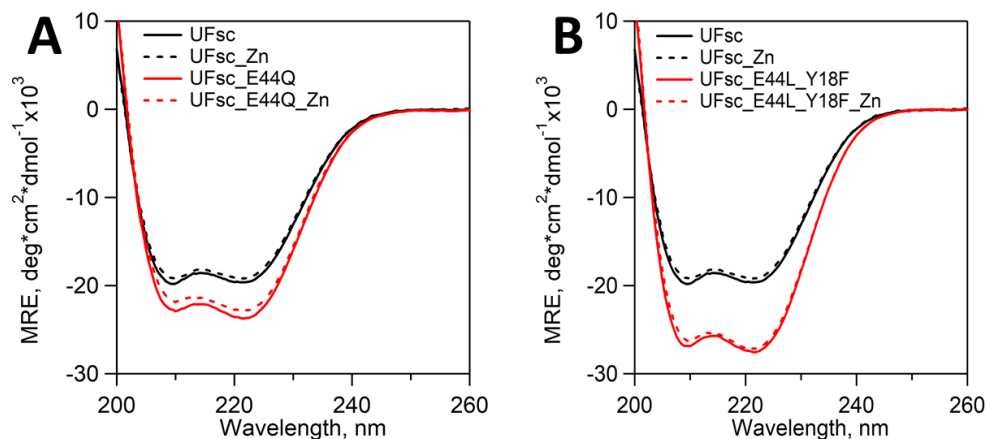


Figure 2.16. CD spectroscopy analysis of UFsc mutants (A – E44Q, B – E44L/Y18F) showed an increase in the degree of helicity upon metal binding. The solution of 20 μM protein in 5 mM HEPES, 20 mM NaCl (pH 7.6) was used as an apoprotein sample. The metal-bound form was prepared by addition of 200 μM zinc (II) chloride.

Similar to 4G-UFsc, E44Q and E44L/Y18F mutations increased the degree of protein helicity but did not affect protein flexibility as no metal-induced folding was observed (Figure 2.15).

The analysis of protein-metal binding by ITC showed, that elimination of E44 in UFsc had a negative impact on manganese (II), nickel (II) and zinc (II) binding. At the same time, both sets of mutants (E44Q and E44L/Y18F) showed improved binding to cobalt (II) ions (Table 2.6).

Table 2.6. Thermodynamic parameters of divalent metal binding to UFsc E44Q and E44L/Y18F mutants determined by ITC (zinc was tested in the competition experiment with TETA).

Metal	Parameters	UFsc	UFsc E44Q	UFsc E44L Y18F
Mn (II)	N	0.755 ± 0.005	0.607 ± 0.003	0.774 ± 0.015
	Kd, nM	173 ± 56.0	560 ± 61	36,600 ± 3,380
	ΔH, kcal/mol	2.39 ± 0.03	4.37 ± 0.04	8.92 ± 0.39
Co (II)	N	0.617 ± 0.005	0.485 ± 0.002	0.606 ± 0.002
	Kd, nM	269 ± 116	51.7 ± 21.7	172 ± 22
	ΔH, kcal/mol	1.46 ± 0.02	5.35 ± 0.07	6.99 ± 0.05
Ni (II)	N	0.632 ± 0.003	No binding	0.716 ± 0.004
	Kd, nM	112 ± 28		824 ± 83
	ΔH, kcal/mol	4.12 ± 0.05		7.50 ± 0.01
Zn (II)	N	0.993 ± 0.003	0.897 ± 0.006	0.760 ± 0.006
	Kd, nM	0.076 ± 0.013	0.373 ± 0.070	0.285 ± 0.077
	ΔH, kcal/mol	-1.85 ± 0.05	-0.309 ± 0.068	0.603 ± 0.048

The UV-Vis spectrum of cobalt (II) is sensitive to its coordination environment. The molar extinction coefficients for d-d bands of five-coordinate, high spin cobalt (II) complexes are below 250 (typically in the range of 100-150 $M^{-1}cm^{-1}$), while the typical six-coordinate complexes will have low extinction coefficients of 10-20 $M^{-1}cm^{-1}$.^{38, 39} The comparison of absorbance spectra collected for E44Q mutants and scaffolds showed the decrease in the extinction coefficient of mutant-cobalt (II) complexes in both cases (Figure 2.17). 4G-UFsc exhibits an extinction coefficient greater than 250 $M^{-1}cm^{-1}$, a distorted tetrahedral coordination geometry can be suggested for this protein (vs 5-coordinate complex for UFsc), although extensive structural studies are required for proper characterization of metal coordination configuration.

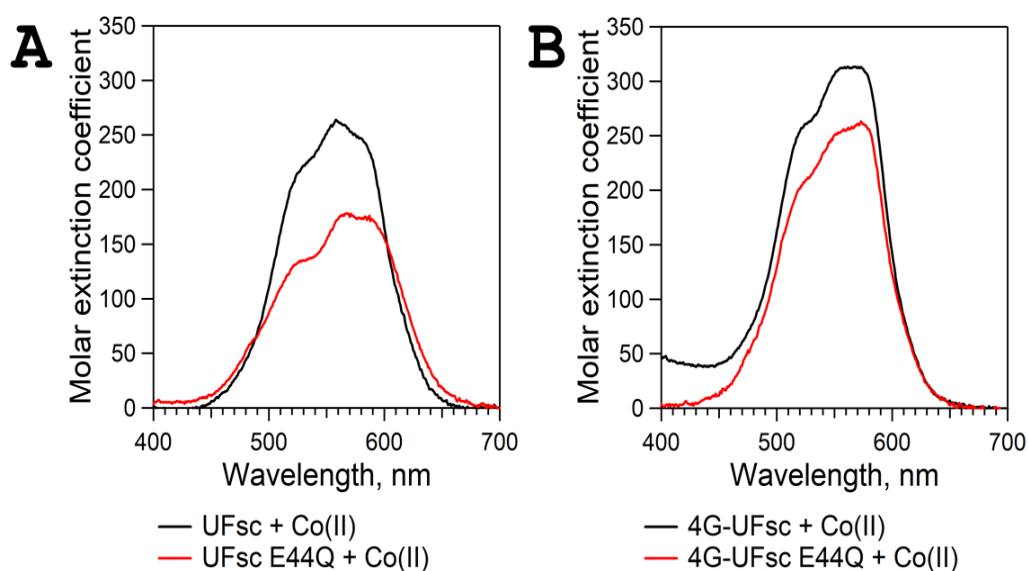


Figure 2.17. Molar extinction coefficients of Co(II)-bound proteins UFsc (A) and 4G-UFsc (B) determined from the absorbance spectra. Samples contained 150 μM protein in buffer (25 mM HEPES, 100 mM NaCl, pH 7.6) with 300 μM $CoCl_2$. The spectra were corrected for the absorbance at 700 nm and shown with the molar extinction coefficient on y-axis.

The competition titration with Mag-Fura 2 was used as an additional method to determine the dissociation constants for protein-zinc (II) binding. In most cases, the values obtained using ITC competition titration and the dye-based absorbance assay were in close agreement (Table 2.7), which validates both methods for the determination of protein metal affinity.

Table 2.7. Dissociation constants for 4G-UFsc and its mutants determined using competition titration assay with Mag-Fura 2.

Protein	Competition titration with Mag-Fura 2 (mean ± SD, M)	ITC competition titration with TETA (mean ± SD, M)
4G-UFsc	$3.71 \pm 1.93 \times 10^{-11}$	$3.00 \pm 0.40 \times 10^{-11}$
4G-UFsc E44Q	0.24×10^{-9}	$1.10 \pm 0.21 \times 10^{-9}$
4G-UFsc E44L Y18F	$23.4 \pm 1.76 \times 10^{-9}$	$3.06 \pm 0.34 \times 10^{-9}$
4G-UFsc E11Q	$32.7 \pm 1.75 \times 10^{-11}$	$7.60 \pm 1.70 \times 10^{-11}$
4G-UFsc E74D	$1.55 \pm 0.80 \times 10^{-9}$	$4.19 \pm 0.25 \times 10^{-9}$
4G-UFsc E74H	$1.56 \pm 0.72 \times 10^{-10}$	$2.81 \pm 0.63 \times 10^{-10}$
4G-UFsc E74Q	$0.339 \pm 1.92 \times 10^{-8}$	$2.12 \pm 0.32 \times 10^{-8}$
4G-UFsc H107Q	$5.14 \pm 3.33 \times 10^{-10}$	$2.01 \pm 0.25 \times 10^{-10}$

Overall, the differences in metal coordination by UFsc and 4G-UFsc indicate a significant influence of alanine-to-glycine mutations not only on the substrate access to the active site but also on protein structure which in its turn affects metal-binding properties.

2.2.5. Copper coordination by UFsc

In our work, we have been focused on the most common transition metals manganese, nickel, cobalt and zinc. Copper is another important transition metal found in many natural metalloenzymes with diverse functions. Copper is a redox-active metal readily switching between Cu^+ and Cu^{2+} which explains its role in electron transfer reactions.⁴⁰ We have shown that the metal coordination sphere in UFsc is flexible and can accommodate multiple metals. Copper (II) coordination in proteins is complex as this ion can be coordinated to only three ligands and usually forms stable complexes. Even though the coordination environment in UFsc is not optimized for copper (II), competition titration experiment with Mag-Fura 2 showed that this protein binds copper ions with a 1:2 stoichiometry (one protein binds two metals) (Figure 2.18).

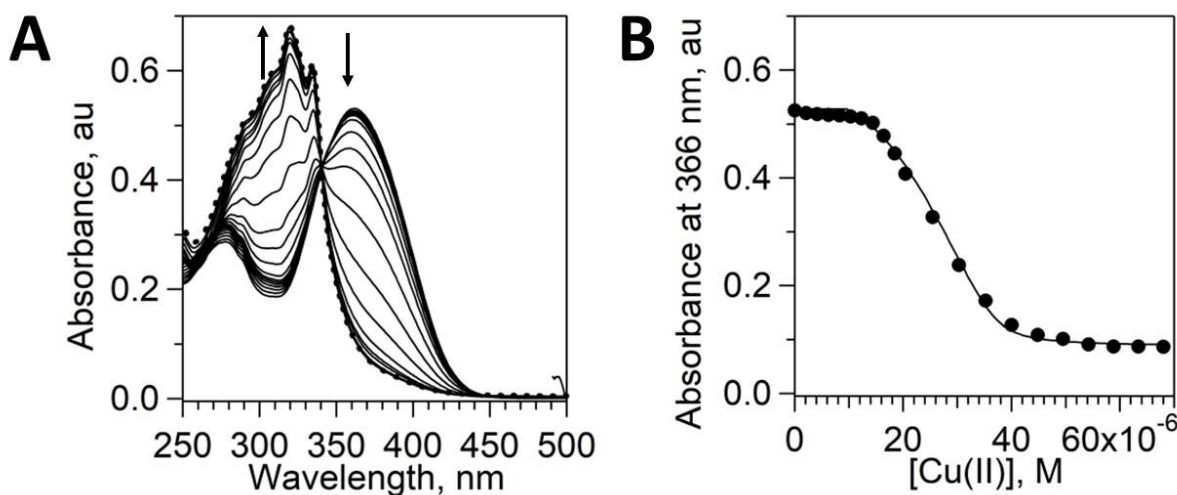


Figure 2.18. Competition titration of UFsc and Mag-Fura 2 with CuCl_2 . (A) Absorbance spectra of Mag-Fura 2 (17.6 μM) and UFsc (9.7 μM) titrated with CuCl_2 . The coordination of metal reduces absorbance intensity of the dye at 366 nm. (B) Absorbance at 366 nm as a function of metal concentration was fit using DYNAFIT to determine dissociation constant of Cu (II)-UFsc complex.

Due to the complexity of the titration curve, the extracted values for dissociation constants need to be verified using another method. To perform the ITC competition titration a chelator with proper characteristics (1:1 metal-chelator binding, low heat of dilution, no interaction with the protein) is required for further studies.

Copper-containing proteins adopt different metal coordination environments and can be divided into three main groups: mononuclear, dinuclear and tetranuclear.^{41, 42} As UFsc binds two copper ions, we hypothesized that this protein could adopt a copper center. The metal-binding site of UFsc consists of histidine and glutamate ligands and this arrangement does not align with the architecture of any described copper centers. Inspired by the *de novo* designed nitrite reductases developed by the Pecoraro group, we decided to test our protein in nitric oxide production. Nitrite reductases represent a group of copper-containing enzymes which catalyze a single electron transfer to produce nitric oxide. These proteins can adopt type 1 and type 2 mononuclear copper centers. The reaction mechanism involves the binding of nitrite to copper (II), reduction of copper (II) to copper (I) and formation of nitric oxide.⁴³ The designed α -helical coiled-coils consist of three peptides and coordinate copper at different oxidation states with three histidine residues. It has been shown that TRIL23H (Ac-G[LKALEEK]3[HKALEEK]G-NH₂) and TRI-EHK24E (Ac-GWKALEEK LWKALEEK LWKALEEE HEALEEKG-NH₂) can produce NO at pH 5.8.⁴⁴⁻⁴⁶ The spectrophotometric assay used to evaluate nitrite reduction activity of the protein is based on the detection of [Fe(NO)EDTA]²⁻ complex: the gas evolved during the reaction is transferred into the detection solution ([FeEDTA]²⁻) and the absorbance of [Fe(NO)EDTA]²⁻ complex is measured at 432 nm. The analysis of nitrite reduction activity of UFsc showed that the

protein facilitates the production of nitric oxide (Figure 2.19), although these results need further investigation to improve data reproducibility for accurate yield calculations.

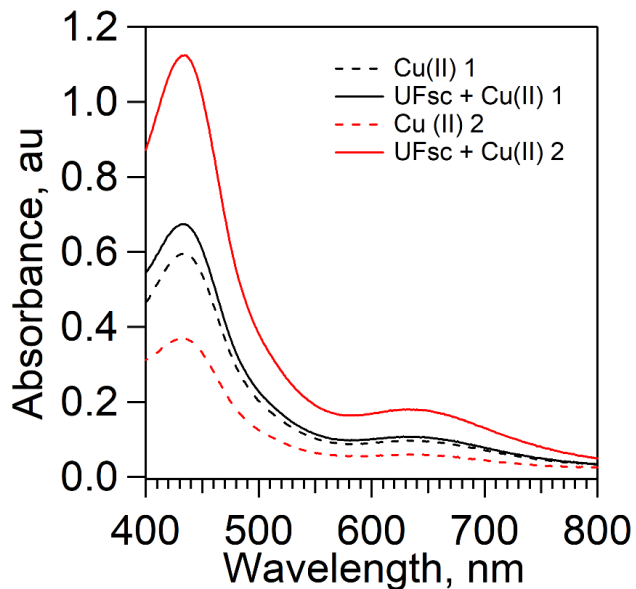


Figure 2.19. NO production by UFsc-Cu(II) complex at pH 7.6. The traces show two experiments where 400 μM copper (II) (dashed lines) or 200 μM UFsc with 400 μM copper (II) (solid lines) were incubated with 20 mM sodium nitrite for 4H at 25 $^{\circ}\text{C}$ in 25 mM HEPES (pH 7.6) with 100 mM NaCl. The presence of produced nitric oxide is correlated with the amount of $[\text{Fe}(\text{NO})\text{EDTA}]^{2-}$ complex, which can be detected at 432 nm.

2.3. Conclusions

In this work, we applied isothermal titration calorimetry for the investigation of the metal-binding sites in *de novo* designed proteins. The introduction of E104H mutation in the highly symmetrical metal-binding site of DFsc introduced asymmetry leading to the formation of one high-affinity metal-binding site. Analysis of the series of mutants allowed us to determine the location of metal-binding site and residues crucial for metal affinity. The devastating effects of mutations at positions 44, 74 and 77 suggest the importance of these residues for the binding of divalent metals in 4G-UFsc, thus they contribute to the primary metal coordination sphere in 4G-UFsc. 4G-UFsc H107Q showed an increased affinity for Co(II) (K_d 6.12 nM), while also affecting the protein-metal stoichiometry (one metal links two protein molecules). This effect is interesting and requires further investigation. 4G-UFsc E11Q demonstrated improved binding to Mn(II) with the K_d of 4.2 μ M. To our knowledge, these are the tightest affinities for Co(II) and Mn(II) reported to date in the *de novo* designed mononuclear metalloproteins. The diverse effects of the mutations outside the primary coordination sphere suggest the necessity of those interactions for maintaining protein structure and achieving the optimal coordination environment for the metal ion.

We have shown the importance of the primary and outer coordination sphere interactions for the metal affinity in *de novo* designed proteins. Fine tuning of the protein structure plays an important role in the productive arrangement of the metal-binding site. Given the importance of model proteins for functional studies of metalloenzymes, our studies will inform future protein design efforts.

2.4. Experimental

Cloning and mutagenesis. DNA sequence that codes for DFsc protein (MDELRELLKAEQQAIKIYKEVLKKAKEGDEQELARLIQEIVKAEKQAVKVYKEAAEKAR NPEKRQVIDKILEDEEKHIEWLKAASKQGNAEQFASLVQQILQDEQRHVVEIEKKN) was synthesized by Bio Basic Inc. and cloned into pMCSG49 expression vector (DNASU) using ligase independent cloning (LIC). Site-directed mutagenesis was performed using Phusion High-Fidelity DNA Polymerase (ThermoScientific) and customized primers (Integrated DNA Technologies, Inc.) following standard protocols.

Scaffold	Mutation	Forward primer	Reverse primer
UFsc	E44Q	CGTCAAAGCACAGAAACAGGC	GCAGTTTCGTGTCTTTGTCCG
	E44L	GAAATCGTCAAAGCACTGAAACAG	CTGTTTCAGTGCTTTGACGATTTTC
	Y18F	ATTAAGATCTTTAAAGAAGTACTGA AGAAAGC	GCTTTCTTCAGTACTTCTTTAAAGA TCTTAAT
4G-UFsc	E44Q	CGTCAAAGGACAGAAACAGGG	CCCTGTTTCTGTCCTTTGACG
	E44L	CGTCAAAGGACTGAAACAGGGG	CCCCTGTTTCAGTCCTTTGACG
	Y18F	ATTAAGATCTTTAAAGAAGTACTGA AGAAAGC	GCTTTCTTCAGTACTTCTTTAAAGA TCTTAAT
	E11Q	CGTGAAGTCTGAAAGGCCAGCAG CAAGGTATTAAGATC	GATCTTAATACCTTGCTGCTGGCC TTTCAGCAGTTTCAGC
	H107Q	GGACCATCAACGTCAGGTGGAAGA GATTGAG	CTCAATCTCTTCCACCTGACGTTG ATGGTCC
	H77Q	GACGAAGAAAACAGATCGAGTGG CTG	CAGCCACTCGATCTGTTTTTCTTC GTC
	E74D	GACAAAATCCTGGAAGACGATGAA AAACATATCGAGTGGC	GCCACTCGATATGTTTTTCATCGT CTTCCAGGATTTTGTC
	E74Q	CGACAAAATCCTGGAAGACCAGGA AAACATATCGAGTGGC	GCCACTCGATATGTTTTTCTGGT CTTCCAGGATTTTGTCG
	E74H	CGACAAAATCCTGGAAGACCATGA AAACATATCGAGTGG	CCACTCGATATGTTTTTCATGGTCT TCCAGGATTTTGTCG

The PCR products (plasmids with mutations) were digested with DpnI (restriction enzyme, which cleaves methylated DNA) at 37 °C for 2–4 h and transformed into

NEB5 α cells (New England Biolabs). The resulting plasmids were sequenced to verify the presence of mutations.

Protein expression and purification. For the expression of DFsc, UFsc, 4G-UFsc and their mutants, the corresponding plasmids were transformed into *E. coli* BL21(DE3) (New England Biolabs) cells. The starter overnight culture (20 mL) was inoculated into Luria Bertani (LB) medium (1.5 L) supplemented with ampicillin (100 μ g/mL). The cell cultures were grown at 37°C until OD₆₀₀ reached 0.6-0.8 and the protein expression was induced with 0.5 mM IPTG (isopropyl β -D-1-thiogalactopyranoside). After the incubation for 4 hours at 37°C, the cells were harvested by centrifugation and lysed by sonication in a lysis buffer (25 mM Tris-HCl (pH 8.0), 100 mM NaCl, 20 mM imidazole) containing 0.5 mM PMSF (phenylmethylsulfonyl fluoride). Soluble protein fraction separated by centrifugation (20,000 x g, 30 min, 4°C) and filtration (0.22 μ m syringe PES filter) was loaded on a Ni-NTA column pre-equilibrated with the lysis buffer. The column was washed with the lysis buffer until no protein came off the column (assessed using the Bradford reagent, Sigma). The proteins were eluted with the elution buffer (25 mM Tris-HCl (pH 8.0), 100 mM NaCl, 250 mM imidazole) and dialyzed into TEV cleavage buffer (50 mM Tris-HCl (pH 8.0), 75 mM NaCl; buffer-to-protein ratio of 100:1) using SnakeSkin dialysis tubing (MWCO 3.5 kDa, Thermo Fisher Scientific) for 2 hours at 4°C.

The cleavage of His-tag was conducted overnight at 34°C in the TEV cleavage buffer supplemented with 1 mM DTT (1,4-dithiothreitol), 2 mM EDTA and TEV protease at a ratio $A_{280 \text{ protein}}/A_{280 \text{ protease}} = 50/1$. The proteins were further purified on C4 preparative reverse-phase HPLC column (Jupiter 15 μ m C4 300Å 250x21.2 mm

column, Phenomenex) using the following solvents and a gradient method: Solvent A (0.1% trifluoroacetic acid (TFA) in water) and Solvent B (90% acetonitrile, 9.9% water with 0.1% TFA); 10% B for 10 min, from 10% B to 50% B over 5 minutes, 50% B to 80% B over 20 minutes, 100% for 5 min and 0% B for 5 min at flow rate of 15 mL/min. The purity (> 95%) of the final samples was checked by SDS-PAGE analysis. The solvents from pure fractions were removed by lyophilization. Lyophilized proteins were dissolved in MilliQ water (~3 mL) and then dialyzed against storage buffer (25 mM HEPES (pH 7.6), 100 mM NaCl) for at least 2 hours at 4°C using Slide-A-Lyzer G2 dialysis cassettes with a molecular weight cutoff of 7 kDa (Thermo Fisher Scientific). Protein concentration was determined using the absorbance at 280 nm and the calculated extinction coefficients of $\epsilon_{280} = 8,480 \text{ M}^{-1}\text{cm}^{-1}$ for DFsc, UFsc, UFsc E44Q, 4G-UFsc, 4G-UFsc E44Q, 4G-UFsc E11Q, 4G-UFsc H107Q, 4G-UFsc H77Q, 4G-UFsc E74D, 4G-UFsc E74Q, 4G-UFsc E74H, and $\epsilon_{280} = 6,990 \text{ M}^{-1}\text{cm}^{-1}$ for UFsc E44L/Y18F and 4G-UFsc E44L/Y18F.

Cobalt titration was used to test for the presence of EDTA. For this, protein samples (150 μM) were incubated overnight with variable amounts of Co(II) (up to 2 eq of metal ion per protein) and the absorbance at 400-700 nm of the mixture was recorded using Agilent Cary 60 UV-Vis spectrophotometer. The protein was considered free of EDTA if the absorbance of the protein-cobalt complex at λ_{max} increased without lag-phase. The amount of metal bound to protein was assessed by ICP-OES (inductively coupled plasma optical emission spectrometry, Perkin Elmer Optima 3300 DV spectrometer) using 25 μM protein samples in 2% HNO_3 . In all samples,

concentration of Mn, Fe, Co, Ni, Cu and Zn did not exceed 5% of the total protein concentration.

CD spectroscopy. Circular dichroism (CD) spectra were collected using Jasco J-715 CD spectrometer in a 0.1-cm pathlength quartz cuvette. Protein samples (20 μM) were prepared in buffer (5 mM HEPES (pH 7.6), 20 mM NaCl) or in buffer containing zinc chloride (200 μM). All samples were scanned in the continuous mode from 260 to 200 nm, with 0.5 nm bandwidth, 0.5 nm data pitch, 50 nm/min scan rate, 4 s averaging time. The final spectra represent a buffer-subtracted average of four scans. The sample absorbance never exceeded 2 at all wavelengths and the mean residue ellipticity (MRE, $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$) values were calculated using the following equations, where θ (mdeg) is the ellipticity, l (cm) is the path length, C (M) is the peptide concentration, N is the number of residues: $\text{MRE} = \theta / (10 \cdot C \cdot l \cdot N)$.

Isothermal titration calorimetry. The thermodynamic parameters of protein metal binding were monitored using a MicroCal PEAQ-ITC instrument (Malvern). Proteins were dialyzed against buffer (25 mM HEPES (pH 7.6), 100 mM NaCl) and filtered using 0.22 μm low protein binding PES filter (Santa Cruz Biotechnology, Inc.); the concentration of each sample was determined by UV-vis spectroscopy using extinction coefficients listed above. Stock solutions of metal salts (50 mM) were prepared in water and filtered. After metal concentrations were established by ICP-OES using appropriate dilutions in 2% HNO_3 , the stock solutions were aliquoted and lyophilized. Before the experiment, an aliquot of lyophilized metal was dissolved in water and then diluted to the appropriate concentrations in working buffer (25 mM HEPES (pH 7.6), 100 mM NaCl).

All titrations were performed at 25°C in the high feedback mode with 750 rpm stir speed and an appropriate equilibration time between injections (150-250 s). For metal-to-protein titrations, the protein sample (100-200 µM) was placed in the calorimeter cell and the syringe was filled with the solution of Mn(II), Co(II) or Ni(II) salts (1-3 mM) in buffer. Zinc-protein binding was established by the competition titration with TETA. For this, triethylenetetramine (TETA) complex with Zn(II) (200 µM TETA and 90 µM ZnCl₂) in the cell was titrated with the protein in the syringe (1-1.5 mM). The analysis of titration data, including baseline correction, peak integration and the correction for heat of dilution observed at the protein saturation with metal, was performed using the Microcal PEAQ-ITC analysis software provided by the manufacturer. The data were fitted to the one set of sites model for UFsc and 4G-UFsc and the two sets of sites model for DFsc to determine the thermodynamic parameters for Mn(II), Co(II) and Ni(II) binding to proteins. The competitive model was used to analyze protein-Zn(II) interactions using independently determined thermodynamic parameters for Zn(II)-TETA binding (number of binding sites – 0.777 ± 0.005 , $K_d = 0.173 \pm 0.065 \mu\text{M}$, $\Delta H = 3.57 \pm 0.06 \text{ kcal/mol}$). For this, 100 mM stock of triethylenetetramine (Sigma) was prepared in water and then dissolved in dialysis buffer to 200 µM. The zinc (II) solution was prepared by dissolving 50 mM stock (in 10 mM HCl in water) in the dialysis buffer. The metal solution was titrated into TETA using the protocol described above.

Competition titration with Mag-Fura 2. Apo proteins were quantified by absorbance at 280 nm (extinction coefficients provided in Protein expression and purification). Mag-Fura 2 (Biotium) was dissolved in buffer (25 mM HEPES (pH 7.6), 100 mM NaCl) to make 1 mM solution and the aliquots were stored at -20°C until

needed. The protein was prepared in buffer to achieve $\sim 10 \mu\text{M}$ concentration (the exact protein concentration was calculated from $A_{280 \text{ nm}}$). Mag-Fura 2 was further added to the reaction at $\sim 20 \mu\text{M}$ and the final concentration was calculated using $\epsilon_{366} = 29,900 \text{ M}^{-1} \text{ cm}^{-1}$ after the baseline-correction.²⁵ The total volume of sample in a 1-mL quartz cuvette was 960 μL . The stock solution of ZnCl_2 (56.7 mM) was prepared by dissolving the appropriate amount of salt (VWR) in 10 mM HCl; CuCl_2 stock (50.4 mM) in 10 mM HCl in water was also prepared from the salt (VWR). The final solutions were analyzed by ICP-OES to ensure the correct concentration of zinc or copper and to check for transition metal contamination (like manganese, iron, cobalt, nickel, zinc and copper). The stocks were diluted with buffer before titrations to make 1 mM solutions. The protein-dye mixture was titrated with aliquots of 1 mM ZnCl_2 or CuCl_2 stock (2 μL to achieve zinc or copper concentration of 2-20 μM and 5 μL to achieve metal concentration of 25-50 μM) and the resulting solution was allowed to equilibrate for 5 min before the measurement. Both zinc and copper tend to precipitate in water-based solutions at physiological pH, thus the fresh metal stocks were prepared every 10 min. Absorbance spectra were recorded from 200 to 500 nm using Agilent Cary 60 UV-Vis spectrophotometer, and binding of metal to the dye was monitored by disappearance of the free Mag-Fura 2 absorbance peak at 366 nm. The data sets of the dye absorbance at 366 nm vs the concentration of added ZnCl_2 were analyzed using a custom DYNAFIT⁴⁷ script that modeled competition between dye and protein for the metal ion (script provided in Appendix I) using zinc-Mag-Fura 2 dissociation constant of 20 nM⁴⁸ and copper-Mag-Fura 2 dissociation constant of 310 nM⁴⁹. Reported dissociation constant values are an average of three individual experiments.

Nitrite reduction by UFsc/Cu(I) complex. All reactions were prepared in reaction buffer consisting of 25 mM HEPES (pH 7.6), 100 mM NaCl. UFsc was purified and quantified as described in *Protein expression and purification* section, 0.5 M L-ascorbic acid (Sigma) solution was prepared by dissolving powder in water, 50 mM CuCl₂ (VWR) was prepared by dissolving salt in 10 mM HCl in water, 0.2 M NaNO₂ solution (Sigma) was prepared by dissolving powder in the buffer. Copper (II) was reduced to copper (I) by mixing 400 μM CuCl₂ and 2 mM L-ascorbic acid in reaction buffer in 25-mL glass Schlenk vessel, then 200 μM UFsc was added to obtain UFsc/Cu(I) complex in total volume of 1 mL and the vessel was covered with septum. This mixture was then degassed using Schlenk system (10 alternate cycles of vacuum and nitrogen flow) and 0.2 M NaNO₂ solution was bubbled with nitrogen to remove oxygen. After degassing, 100 μL of 0.2 M NaNO₂ solution were added to the protein mixture (19 mM final concentration) and the mixture was incubated for 4 H at RT. The control reactions were prepared following the same protocol but substituting the reactants with buffer or water. After the incubation, the produced NO gas was transferred to the detection solution using metallic cannula. The detection solution, 10 mM [FeEDTA]²⁻, was prepared in 1 mL of 100 mM citrate-NaOH buffer (pH 5.06) in the gas tight cuvette. For this, 100 μL of 0.1 M ammonium iron (II) sulfate in 0.9 M citrate buffer were mixed with 100 μL of 0.1 M EDTA in 0.1 M citrate buffer and 800 μL of water in the glove box. The spectrum of the blank was recorded before gas transfer. NO gas was then transferred to the cuvette where it reacted with [FeEDTA]²⁻ forming [Fe(NO)EDTA]²⁻, a compound which absorbs at 432 nm with extinction coefficient of 780 M⁻¹cm⁻¹. To ensure complete transfer of NO, the reaction vessel was additionally purged

with nitrogen. The detection solution became yellow which indicated the formation of the complex and the spectrum of reacted solutions was recorded using Cary 60 spectrometer. For the calculation, $[\text{FeEDTA}]^{2-}$ spectrum was subtracted from $[\text{Fe(NO)EDTA}]^{2-}$, and value at 432 nm was used for the calculation of the amount of nitric oxide produced by reaction.

2.5. References

1. Abreu, I. A.; Cabelli, D. E., Superoxide dismutases-a review of the metal-associated mechanistic variations. *Biochim. Biophys. Acta* **2010**, *1804* (2), 263-74.
2. Boer, J. L.; Mulrooney, S. B.; Hausinger, R. P., Nickel-dependent metalloenzymes. *Arch. Biochem. Biophys.* **2014**, *544*, 142-52.
3. Schmidt, S. B.; Husted, S., The biochemical properties of manganese in plants. *Plants (Basel)* **2019**, *8* (10).
4. McCall, K. A.; Huang, C.; Fierke, C. A., Function and mechanism of zinc metalloenzymes. *J. Nutr.* **2000**, *130*, 1437-46.
5. Ragsdale, S. W., Nickel-based enzyme systems. *J. Biol. Chem.* **2009**, *284* (28), 18571-5.
6. Solomon, E. I.; Decker, A.; Lehnert, N., Non-heme iron enzymes: contrasts to heme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (7), 3589-94.
7. Barynin, V. V.; Whittaker, M. M.; Antonyuk, S. V.; Lamzin, V. S.; Harrison, P. M.; Artymiuk, P. J.; Whittaker, J. W., Crystal structure of manganese catalase from *Lactobacillus plantarum*. *Structure* **2001**, *9* (8), 725-38.

8. Kolberg, M.; Strand, K. R.; Graff, P.; Andersson, K. K., Structure, function, and mechanism of ribonucleotide reductases. *Biochim. Biophys. Acta* **2004**, 1699 (1-2), 1-34.
9. Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J.; Nordlund, P., Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. *Nature* **1993**, 366 (6455), 537-43.
10. Sazinsky, M. H.; Bard, J.; Di Donato, A.; Lippard, S. J., Crystal structure of the toluene/o-xylene monooxygenase hydroxylase from *Pseudomonas stutzeri* OX1. Insight into the substrate specificity, substrate channeling, and active site tuning of multicomponent monooxygenases. *J. Biol. Chem.* **2004**, 279 (29), 30600-10.
11. Lindqvist, Y.; Huang, W.; Schneider, G.; Shanklin, J., Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other diiron proteins. *EMBO J.* **1996**, 15 (16), 4081-92.
12. Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F., Retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97 (12), 6298-305.
13. Marsh, E. N.; DeGrado, W. F., Noncovalent self-assembly of a heterotetrameric diiron protein. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99 (8), 5150-4.
14. Pasternak, A.; Kaplan, J.; Lear, J. D.; Degrado, W. F., Proton and metal ion-dependent assembly of a model diiron protein. *Protein Sci.* **2001**, 10 (5), 958-69.
15. Di Costanzo, L.; Wade, H.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F.; Lombardi, A., Toward the *de novo* design of a catalytically active helix bundle: a

substrate-accessible carboxylate-bridged dinuclear metal center. *J. Am. Chem. Soc.* **2001**, *123* (51), 12749-57.

16. Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; DeGrado, W. F., Alteration of the oxygen-dependent reactivity of *de novo* Dufor proteins. *Nat. Chem.* **2012**, *4* (11), 900-6.

17. Whittaker, J. W., The radical chemistry of galactose oxidase. *Arch. Biochem. Biophys.* **2005**, *433* (1), 227-39.

18. Stubbe, J.; van Der Donk, W. A., Protein radicals in enzyme catalysis. *Chem. Rev.* **1998**, *98* (2), 705-762.

19. Shibata, N.; Toraya, T., Molecular architectures and functions of radical enzymes and their (re)activating proteins. *J. Biochem.* **2015**, *158* (4), 271-92.

20. Ulas, G.; Lemmin, T.; Wu, Y.; Gassner, G. T.; DeGrado, W. F., Designed metalloprotein stabilizes a semiquinone radical. *Nat. Chem.* **2016**, *8* (4), 354-9.

21. Kaplan, J.; DeGrado, W. F., *De novo* design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (32), 11566-70.

22. Wang, M. S.; Hoegler, K. J.; Hecht, M. H., Unevolved *de novo* proteins have innate tendencies to bind transition metals. *Life (Basel)* **2019**, *9* (1).

23. Ghering, A. B.; Shokes, J. E.; Scott, R. A.; Omichinski, J. G.; Godwin, H. A., Spectroscopic determination of the thermodynamics of cobalt and zinc binding to GATA proteins. *Biochemistry* **2004**, *43* (26), 8346-55.

24. Golynskiy, M. V.; Gunderson, W. A.; Hendrich, M. P.; Cohen, S. M., Metal binding studies and EPR spectroscopy of the manganese transport regulator MntR. *Biochemistry* **2006**, *45* (51), 15359-72.
25. Walkup, G. K.; Imperiali, B. J. J. o. t. A. C. S., Fluorescent chemosensors for divalent zinc based on zinc finger domains. Enhanced oxidative stability, metal binding affinity, and structural and functional characterization. *J. Am. Chem. Soc.* **1997**, *119* (15), 3443-3450.
26. Talmard, C.; Bouzan, A.; Faller, P., Zinc binding to amyloid-beta: isothermal titration calorimetry and Zn competition experiments with Zn sensors. *Biochemistry* **2007**, *46* (47), 13658-66.
27. Ferreira, C. M.; Pinto, I. S.; Soares, E. V.; Soares, H. M., (Un) suitability of the use of pH buffers in biological, biochemical and environmental studies and their interaction with metal ions—a review. *RSC Advances* **2015**, *5* (39), 30989-31003.
28. Nurchi, V. M.; Crisponi, G.; Crespo-Alonso, M.; Lachowicz, J. I.; Szewczuk, Z.; Cooper, G. J., Complex formation equilibria of Cu(II) and Zn(II) with triethylenetetramine and its mono- and di-acetyl metabolites. *Dalton Trans.* **2013**, *42* (17), 6161-70.
29. Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A., An efficient, step-economical strategy for the design of functional metalloproteins. *Nat. Chem.* **2019**, *11* (5), 434-441.
30. Zastrow, M. L.; Pecoraro, V. L., Designing hydrolytic zinc metalloenzymes. *Biochemistry* **2014**, *53* (6), 957-78.
31. Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G., Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J. Mol. Biol.* **2003**, *334* (5), 1101-15.

32. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **2006**, 1 (6), 2876-2890.
33. Maret, W., Zinc coordination environments in proteins determine zinc functions. *J. Trace Elem. Med. Bio.* **2005**, 19 (1), 7-12.
34. Calhoun, J. R.; Bell, C. B., 3rd; Smith, T. J.; Thamann, T. J.; DeGrado, W. F.; Solomon, E. I., Oxygen reactivity of the biferrous site in the *de novo* designed four helix bundle peptide DFsc: nature of the "intermediate" and reaction mechanism. *J. Am. Chem. Soc.* **2008**, 130 (29), 9188-9.
35. Bell, C. B.; Calhoun, J. R.; Bobyr, E.; Wei, P. P.; Hedman, B.; Hodgson, K. O.; DeGrado, W. F.; Solomon, E. I., Spectroscopic definition of the biferrous and biferric sites in *de novo* designed four-helix bundle DFsc peptides: implications for O₂ reactivity of binuclear non-heme iron enzymes. *Biochemistry* **2009**, 48 (1), 59-73.
36. Yoon, J. H.; Kulesha, A. V.; Lengyel-Zhand, Z.; Volkov, A. N.; Rempillo, J. J.; D'Souza, A.; Costeas, C.; Chester, C.; Caselle, E. R.; Makhlynets, O. V., Uno Ferro, a *de novo* designed protein, binds transition metals with high affinity and stabilizes semiquinone radical anion. *Chemistry* **2019**, 25 (67), 15252-15256.
37. Cotruvo, J. A., Jr.; Stubbe, J., Metallation and mismetallation of iron and manganese proteins in vitro and in vivo: the class I ribonucleotide reductases as a case study. *Metallomics* **2012**, 4 (10), 1020-36.
38. McMillin, D. R.; Rosenberg, R. C.; Gray, H. B., Preparation and spectroscopic studies of cobalt (II) derivatives of blue copper proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, 71 (12), 4760-2.

39. Bertini, I.; Luchinat, C., High spin cobalt (II) as a probe for the investigation of metalloproteins. *Adv. Inorg. Biochem.* **1984**, *6*, 71-111.
40. Farver, O.; Pecht, I., Electron transfer in blue copper proteins. *Coord. Chem. Rev.* **2011**, *255* (7), 757-773.
41. Rubino, J. T.; Franz, K. J., Coordination chemistry of copper proteins: how nature handles a toxic cargo for essential function. *J. Inorg. Biochem.* **2012**, *107* (1), 129-43.
42. Dell'Acqua, S.; Pauleta, S. R.; Moura, I.; Moura, J. J., The tetranuclear copper active site of nitrous oxide reductase: the Cu₂ center. *J. Biol. Inorg. Chem.* **2011**, *16* (2), 183-94.
43. Castiglione, N.; Rinaldo, S.; Giardina, G.; Stelitano, V.; Cutruzzola, F., Nitrite and nitrite reductases: from molecular mechanisms to significance in human health and disease. *Antioxid. Redox Signal.* **2012**, *17* (4), 684-716.
44. 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.* **2013**, *135* (48), 18096-107.
45. Koebke, K. J.; Pecoraro, V. L., Development of *de novo* copper nitrite reductases: Where we are and where we need to go. *ACS Catal.* **2018**, *8* (9), 8046-8057.
46. Tegoni, M.; Yu, F.; Bersellini, M.; Penner-Hahn, J. E.; Pecoraro, V. L., Designing a functional type 2 copper center that has nitrite reductase activity within alpha-helical coiled coils. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (52), 21234-9.
47. Kuzmic, P., Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **1996**, *237* (2), 260-73.

48. Simons, T. J., Measurement of free Zn^{2+} ion concentration with the fluorescent probe Mag-Fura-2 (furaptra). *J. Biochem. Biophys. Methods* **1993**, 27 (1), 25-37.
49. Petit-Hartlein, I.; Rome, K.; de Rosny, E.; Molton, F.; Duboc, C.; Gueguen, E.; Rodrigue, A.; Coves, J., Biophysical and physiological characterization of ZraP from *Escherichia coli*, the periplasmic accessory protein of the atypical ZraSR two-component system. *Biochem. J.* **2015**, 472 (2), 205-16.

Chapter 3. Design of the efficient Kemp eliminase using NMR-guided directed evolution

Work from this chapter has led to the following manuscript:

Bhattacharya, S., Margheritis, E.* , Takahashi, K.* , **Kulesha, A.***, D'Souza, A., Kim, I., Tame, J., Yoon, J., Volkov, A., Makhlynets, O., Korendovych, I. *NMR-Guided Directed Evolution*. (2022) Under revision. – A.K. performed protein expression, purification and preparation of cofactor-substituted variants, development of the method for stopped-flow kinetics, characterization of myoglobin proteins using Michaelis-Menten kinetics, circular dichroism and spectroelectrochemistry, characterization of inhibitor binding to AlleyCat mutants.

Abstract. Directed evolution is one of the widely used methods for protein design which requires the construction of large combinatorial libraries to find a productive mutation. In our work, we have employed nuclear magnetic resonance (NMR) to identify the positions for potential productive mutations for further directed evolution. As a proof-of-concept experiment, we have evolved myoglobin, a well-studied heme-containing non-enzymatic oxygen transport protein, in an efficient Kemp eliminase, the enzyme which abstracts a proton from a benzisoxazole ring. Using this method, we were able to identify the positions for productive mutations and to create the most effective designed Kemp eliminase reported to date by combining only three mutations. NMR-guided evolution will provide scientists with a tool for the design of new proteins without extensive calculations or structure determination.

3.1. Introduction

Protein engineering has become a valuable tool to create proteins with desired properties as well as for investigating sequence, structure and function relationships. Newly designed proteins become handy in green organic synthesis, food, pharmaceutical and biotechnology industries.¹ Proteins discovered in Nature catalyze a wide range of reactions from water splitting to glucose metabolism. Many of those proteins can be repurposed or employed for the synthesis of small organic molecules later used as dyes, drugs or building blocks for new materials. To satisfy the high demand for “customized” proteins, scientists employ many methods for protein design.

Rational protein design and directed evolution are the two main approaches that helped to achieve significant breakthroughs in protein design. Rational protein design relies on the advanced understanding of protein folding^{2, 3}, as well as the structure and function relationship. This method also requires a significant amount of information such as protein structure dataset along with the information about protein function. One of the simple approaches to protein design, the minimalist approach, relies on our understanding of protein structure and function. A great example of the successful application of the minimalist approach is the design of AlleyCat. AlleyCat is a 74-residue C-terminal domain of calmodulin, calcium-binding non-enzymatic protein, which was evolved into a Kemp eliminase.⁴ The computational methods were used to predict the best position for the active residue in the hydrophobic pocket and obtain the desired function. Rational design is a productive method for protein engineering but requires large sets of structural data, which are not available for many proteins.

Another major approach to protein engineering is directed evolution. Directed evolution unites a range of molecular biology techniques that allow scientists to mimic protein evolution in the laboratory.^{5, 6} The strategy generally involves the mutagenesis of a starting gene with further screening to isolate new variants with improvements in specific properties. The process is rather laborious as it requires many iterations until the desired changes are obtained or until no further change can be achieved.

Mutagenesis is performed using many molecular biology tools inspired by Nature, such as error-prone PCR (epPCR)⁷ which introduces random point mutations in DNA products, and DNA shuffling techniques⁸ which allow the creation of new DNA sequences based on the fragments of the template genes. One of the examples of the application of directed evolution is work led by Dr. Frances Arnold, who received a Nobel Prize in Chemistry in 2018 for her contribution to the field of protein design. Inspired by the functional flexibility of natural cytochrome P450 monooxygenases, her group repurposed the enzymes from classic oxidation to carbon–carbon bond forming reactions, cyclopropanation, carbene C–H insertion and others.⁹⁻¹²

Directed evolution is a productive technique for protein design but often requires many rounds of experiments to achieve a significant result. To improve the efficiency of this method, the methods of guided evolution that would allow us to identify the promising regions in proteins and focus on them, instead of performing evolution on the whole protein are needed. In our work, we proposed a new method for the successful identification of the sites for productive mutations using nuclear magnetic resonance (NMR). Considering the flexible nature of the protein, we hypothesized that the analysis of NMR chemical shifts of the protein residues upon substrate analog binding can

provide us with information about the amino acid residues important for the proper substrate transition state fitting. Experimentally it can be achieved by the calculation of chemical shift perturbations (CSP) for amino acid residues upon substrate binding. After the identification of important positions, the site-directed mutagenesis with further screening can help to identify the improved variants. Using NMR-guided directed evolution, we were able to design FerrE1Cat, an efficient Kemp eliminase, by combining only three mutations.

3.2. Results and Discussion

3.2.1. *Method development for NMR-guided directed evolution*

Nuclear magnetic resonance (NMR) spectroscopy is a method for the structural characterization of proteins and protein complexes with other molecules. This method has been successfully used to determine the interactions between proteins and small molecules in drug research.¹³ The perturbations in chemical shifts have been used to identify the points of contact between a protein and a molecule to uncover the nature of binding. Protein-substrate-transition-state binding during enzyme catalysis has been a topic of research interest for many years.¹⁴⁻¹⁷ Previously it was suggested that to design an effective catalyst the rigid protein scaffold has precisely fit the transition state. Over the years, research showed the importance of protein dynamics for transition state fitting, catalysis and product discharge.¹⁴ Unfortunately, current available methods for protein design struggle to incorporate protein dynamics in the models as it is a complex process. NMR, on the other hand, can detect changes in the local electronic environment provoked by binding events, highlighting the regions of a protein involved in the binding event. During transition state binding, the protein undergoes subtle changes while adopting the most suitable state. We hypothesized that the regions which undergo the most changes are important for the proper substrate transition state coordination and can be identified by NMR. Further directed evolution at specific positions will result in the evolved proteins with improved or completely new activity.

To test our hypothesis, we attempted to evolve a non-enzymatic protein into an effective Kemp eliminase. Kemp elimination is a reaction of proton abstraction with the opening of the benzisoxazole ring.¹⁸ Kemp elimination has become a popular reaction

for the development of the *de novo* designed proteins.^{19, 20} Most designed proteins catalyze this reaction through the acid-base mechanism, where the base (typically, glutamate or aspartate) abstracts the proton from carbon leading to an N–O bond cleavage (Figure 3.1).

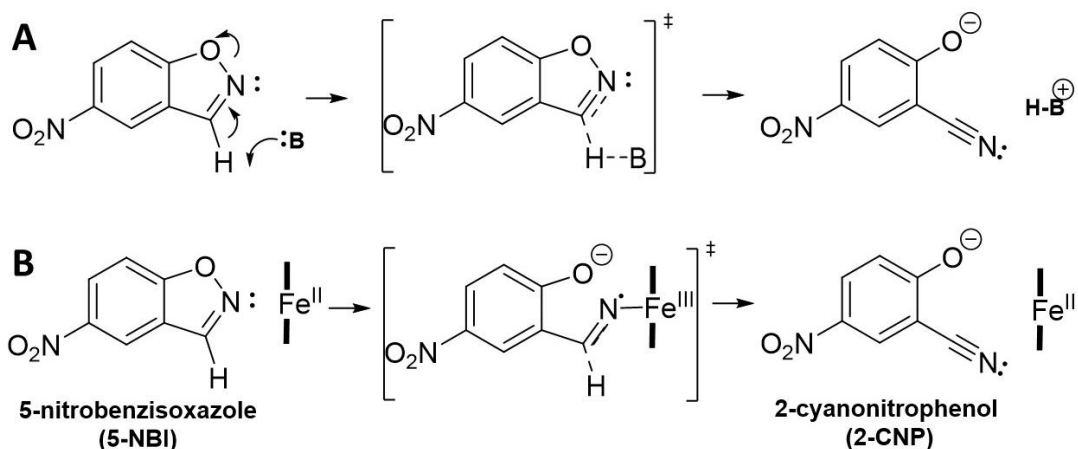


Figure 3.1. The reaction schemes for Kemp elimination promoted by (A) acid-base or (B) redox-mediated mechanisms.

Recently, Li et al. reported that engineered cytochrome P450-BM3 can promote redox-mediated Kemp elimination.²¹ Using QM/MM simulations, Marti et al. showed that the redox-mediated mechanism is the most favorable route for the reaction catalyzed by heme-containing aldoxime dehydratase.²² Inspired by these works, we decided to convert simple non-enzymatic protein myoglobin (Mb) into an efficient Kemp eliminase. Mb has been extensively characterized over the years, it is a small natural non-enzymatic protein without preformed substrate-binding site, which makes it a suitable candidate for method development. Myoglobin, being a heme-containing protein similar to cytochrome p450, might follow a redox-mediated mechanism for Kemp elimination. The redox-mediated mechanism for Kemp elimination suggests that the redox-active

metal in heme can bind to the substrate and induce the ring opening (Figure 3.1. B), thus, from the minimalist point of view, heme is the only necessary condition for the reaction to occur. Keeping in mind that the primary function of myoglobin is oxygen binding and transport utilizing the heme cofactor, we have decided to use Mb-H64V as a starting scaffold. Mb-H64V has a lower affinity to oxygen which improves the stability of the reduced heme (heme with Fe (II)), an important condition for redox-mediated catalysis. Mb-H64V has also been shown useful in other protein design efforts.^{23, 24}

To identify the positions for directed evolution, Mb-H64V was used for the chemical shift perturbation study by NMR. Initially, a series of NMR experiments were performed using a double-labeled protein (¹³C,¹⁵N-Mb-H64V) to obtain the assignments of backbone amide resonances (to correlate the positions of backbone amide NMR peaks with specific amino acids). Then, the 1H-15N HSQC experiments were performed using a ¹⁵N-Mb-H64V protein with or without 6-nitrobenzotriazole (6-NBT), the Kemp substrate transition state analog (Figure 3.2).

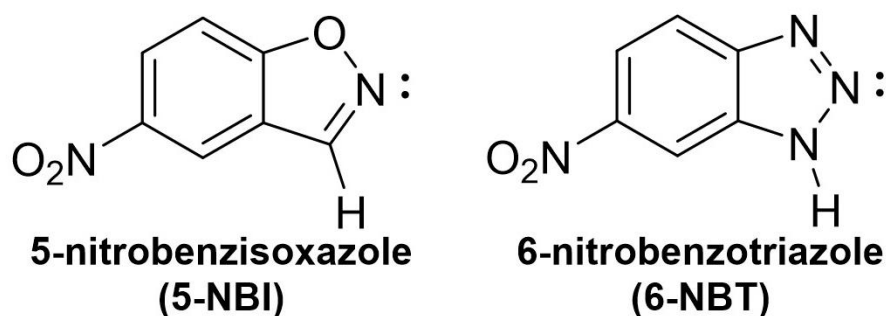


Figure 3.2. The structures of Kemp substrate, 5-nitrobenzisoxazole, and substrate transition state analog, 6-nitrobenzotriazole, used for NMR experiments.

The chemical shift perturbations (CSP) were calculated for the protein residues after the addition of 2 eq of the transitions state analog. The positions with the most

prominent shifts were identified by calculating the Z-score, the value that correlates a given CSP to the average CSP of the experiment. As the protein adopts its conformation for the productive transition state binding, the identified positions were classified as important for the formation of the productive complex. The further protein evolution was performed using saturated mutagenesis, a method that allows probing all 20 proteinogenic amino acids at specific position. For this, the NNK libraries were constructed for each identified position and the activity of the cell lysate containing overexpressed myoglobin mutants was assessed. The colonies which showed the improvement in activity over the original scaffold were sequenced and the identity of the mutant was established. We have tested the NNK libraries at 18 identified positions and the analysis of the purified mutants confirmed the presence of productive mutations in all cases (Figure 3.3 A). For our analysis, we did not include H93 as this residue is essential for the heme binding.²⁵ It should be mentioned that many identified positions are located outside the active site (Figure 3.3 B).

The long-range interactions play an important role in the protein-substrate binding but identification or prediction of such interactions remains a challenge for current methods.²⁶ Unfortunately, the available computation program cannot fully describe the conformational changes in the protein upon substrate binding, thus they cannot predict the positions important for transition state accommodation. The results of our experiments suggest that the NMR-based approach can become a new useful tool for protein engineering, taking into account the productivity of this method. This method is sensitive to the dynamic changes in the protein and can provide more insight into long-range protein dynamics which opens new possibilities in protein design.

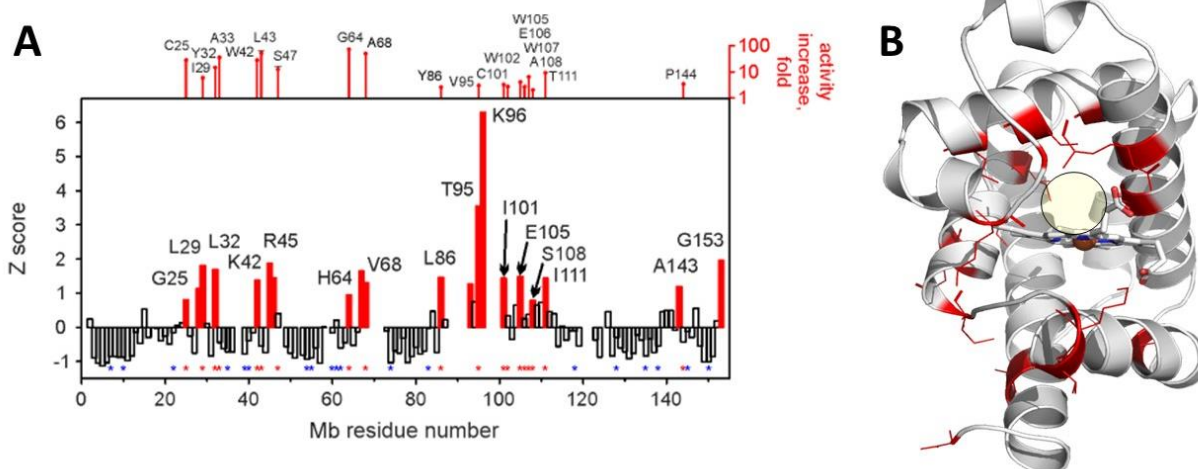


Figure 3.3. A. The positions for productive mutations were identified using a Z score. Out of 18 tested positions with the Z score >1, all produced productive mutations. **B.** The identified hot spots mapped on the structure of myoglobin with the active site highlighted with the yellow sphere (PDB code 6G5T).

3.2.2. The catalytic activity of enzymes designed using NMR-guided directed evolution

We have suggested that myoglobin-based catalysts can promote Kemp elimination reaction following the redox-mediated mechanism. Effective redox reaction requires the presence of redox-active species, and, in the case of myoglobin, it is an iron-containing porphyrin, heme. The redox-mediated mechanism of Kemp elimination requires the presence of the reduced metal (Fe^{2+}) which would transfer the electron to the substrate (Figure 3.1 B). Considering the importance of the reduced state and myoglobin's oxygen sensitivity, we have developed a method for evaluating the activity of reduced proteins under anaerobic conditions.

The initial crude lysate screening of NNK libraries is performed in the 96-well format. This method is essential for high throughput screening but, unfortunately, cannot

be performed in an oxygen-free atmosphere due to the complex multistep procedure. Myoglobin-based catalysts are oxygen-sensitive proteins and require a reduced state for activity. We have prepared the proteins in the reduced forms inside the glove box and have developed a method for the analysis of their catalytic activity using stopped-flow spectroscopy. Stopped-flow UV-Vis spectroscopy is a technique that allows to monitor fast reactions a few milliseconds after the reagent mixing.^{27, 28} The enzyme-containing solution and the substrate are placed in separate gas-tight syringes, injected at the same time in a 1:1 ratio and the accumulation of the product can be monitored over time. The low delay time between the mixing and the start of the reading allows to record the kinetic trace at the very early stage of the reaction, which helps to determine more accurate initial rates.

As our method required the transfer of the reduced protein outside the glove box, we have used a buffer which prolongs the lifetime of the reduced protein species. The combination of ascorbic acid, catalase and superoxide dismutase (SOD) has been previously used in our lab to maintain reduced conditions. Myoglobin mutants used in this work were reduced by the addition of 10 eq of sodium dithionite.²⁹ The addition of ascorbic acid can help to reduce the autooxidation of oxymyoglobin but is coupled with H₂O₂ formation. The further formation of reactive oxygen species (ROS) in the solution is mitigated by the addition of catalase and SOD.^{30, 31} As the stability of reduced proteins can be compromised by oxygen, we have applied the SAC buffer in our experiments. To evaluate the effect of SAC buffer on myoglobin mutant activity, Mb-H64G/V68A's activity was analyzed with and without the addition of ROS-scavenging

buffer and the difference was shown to be within the error of the experiment (Figure 3.4).

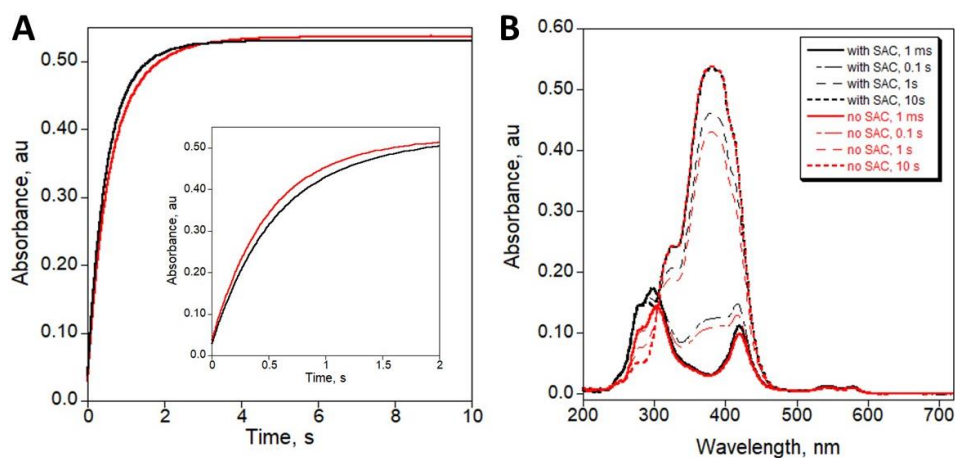


Figure 3.4. A. The kinetic traces for Kemp elimination reaction catalyzed by Mb-H64G/V68A with (black) and without (red) SAC buffer. **B.** The spectra collected for the kinetic analysis of Mb-H64G/V68A with (black) and without (red) SAC buffer.

We have also shown that the Kemp elimination activity of the enzyme-ascorbate mix is negligible (Figure 3.5).

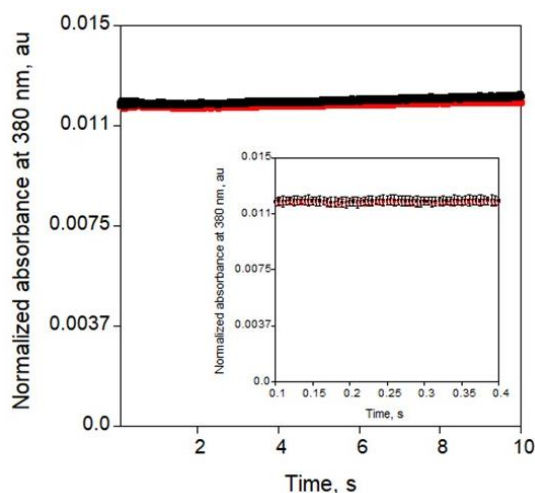


Figure 3.5. Background Kemp elimination rates in 20 mM Tris, pH 8.0 (black) and 20 mM Tris, pH 8.0 with 1 mM ascorbate, 0.1 μM SOD, and 20 nM catalase (red). Presented traces are the average of three runs.

All these results suggest that the SAC buffer does not interfere with the experiment and can be used to maintain reducing conditions in our experiment. As our protocol requires the transfer of proteins between environments, the addition of the SAC buffer could prevent potential oxygen contamination.

During the validation of the NMR-guided method for directed evolution, we produced over 20 mutants and the activity of each protein needed to be tested. To optimize the conditions for the screening, we originally developed a protocol where the protein was prepared under anaerobic conditions with the addition of SAC buffer and then mixed with the substrate prepared outside the glove box. This method allowed us to test proteins faster as only one of the solutions was prepared inside the glove box. The in-depth characterization of myoglobin mutants showed that they were sensitive to oxygen and the activity of these enzymes increased drastically in the oxygen-free environment (Figure 3.6). This finding prompted us to modify our protocol where all solutions used for the reaction were deoxygenated and prepared inside the glove box.

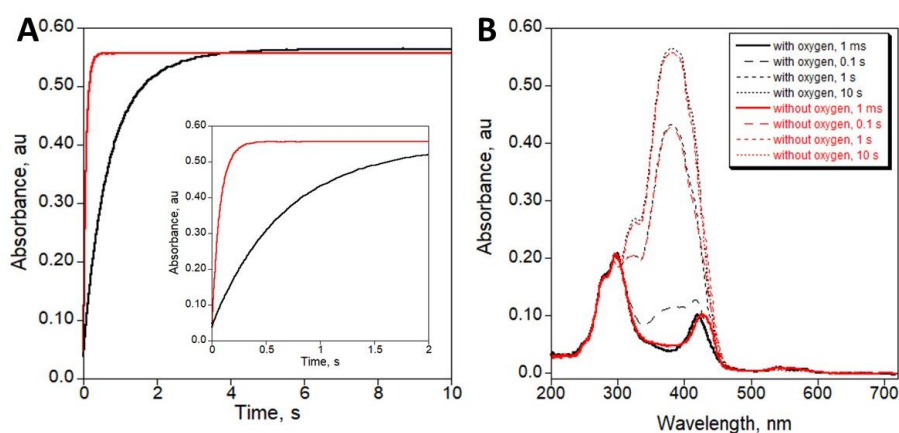


Figure 3.6. A. The kinetic traces for Kemp elimination reaction catalyzed by Mb-H64G/V68A with oxygenated (black) or deoxygenated (red) substrate solution. **B.** The spectra collected for the kinetic analysis of Mb-H64G/V68A with oxygenated (black) or deoxygenated (red) substrate solution.

For further analysis of Kemp elimination activity of designed myoglobin-based proteins was performed using the modified protocol which included oxygen-free solutions and the addition of SAC buffer to provide reducing conditions for myoglobin mutants.

3.2.3. FerrEICat, the most active designed Kemp eliminase

During method development, we probed 18 positions identified by NMR as those that contribute to the transition state analog binding, and at each position, we found a productive mutation. All proteins were purified and characterized using the stopped-flow kinetics experiments to assess the impact of introduced mutations on protein's Kemp elimination activity. The catalytic efficiency of the purified mutants ranged from 2-fold to 53-fold higher compared to the initial scaffold Mb-H64V (Table 3.1, Figure S 13). Such a high rate of hit identification makes NMR-guided directed evolution an effective tool for protein engineering.

The saturation mutagenesis at position 64 produced valine-to-glycine substituted mutant with a 71-fold improvement in the catalytic activity (Table 3.1). Wild-type sperm whale myoglobin has histidine at position 64 which facilitates oxygen binding with further stabilization of gas molecule.³² The substitution of His to Val reduced protein's affinity to oxygen but also increased the size of the binding pocket by removing the bulky side chain of histidine. Further mutagenesis to glycine at that position provided more space for substrate accommodation, which possibly lead to the improvement in catalytic efficiency (Figure 3.7). The introduction of more hydrophobic residues in the binding site could have also contributed to the substrate binding.

Table 3.1. Kinetic parameters for the Kemp elimination reaction catalyzed by myoglobin mutants. Experiments were conducted under anaerobic and reducing conditions at pH 8.0 (25°C).

Protein	k_{cat}, s^{-1}	K_M, mM	k_{cat}/K_M, $M^{-1}s^{-1}$
H64V	-	-	255 ± 8
H64V (oxidized)	-	-	7 ± 1
G25C/H64V	-	-	7,354 ± 143
L29I/H64V	-	-	1,550 ± 55
L32Y/H64V	-	-	3,795 ± 122
F33A/H64V	-	-	9,270 ± 523
K42W/H64V	-	-	7,237 ± 192
F43L/H64V	26.10 ± 3.85	1.94 ± 0.38	13,458 ± 670
K47S/H64V*	2.55 ± 0.39	0.79 ± 0.21	3,240 ± 994
H64V/V68A	-	-	12,939 ± 622
H64V/L86Y	-	-	5,097 ± 355
H64V/T95V	-	-	785 ± 32
H64V/I101C	-	-	3,844 ± 551
H64V/K102W	-	-	710 ± 28
H64V/E105W	-	-	4,786 ± 538
H64V/F106E	-	-	697 ± 22
H64V/I107W	-	-	1,680 ± 135
H64V/S108A	-	-	524 ± 24
H64V/I111T	-	-	2,356 ± 114
H64V/A144P	-	-	900 ± 43
H64G	-	-	18,152 ± 519
H64G/V68A	2,557 ± 372	1.28 ± 0.28	1,992,300 ± 143,420
L29I/H64G/V68A	3,656 ± 667	0.23 ± 0.13	15,721,000 ± 6,035,800

* the value is reported for the initial experiment were the substrate solution contained oxygen

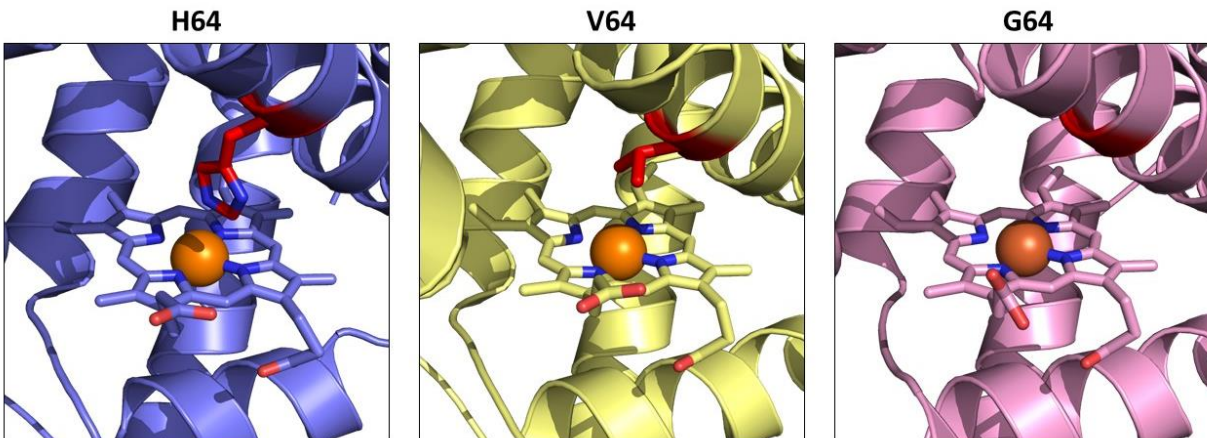


Figure 3.7. The structure of the active site in myoglobin upon mutagenesis at position 64. H64 (left, PDB 1MBN) facilitates oxygen binding, V64 (center, PDB 6G5T) reduced oxygen-binding affinity, G64 (right, the structure was determined as part of this project and deposited to wwPDB) opens the active site for better substrate access. Iron ion in heme is presented as an orange sphere.

Inspired by these results, we have performed a non-exhaustive gene shuffling to test if the identified mutations will have an additive effect. By combining only three mutations identified during initial screening, we were able to create a FerrEICat, Mb-L29I/H64G/V68A, a Kemp eliminase with the remarkable catalytic efficiency of $15,721,000 \pm 6,035,800 \text{ M}^{-1}\text{s}^{-1}$, which corresponds to 62,000-fold improvement over the original scaffold ($255 \text{ M}^{-1}\text{s}^{-1}$). Such high activity is *on par* with the most efficient natural enzymes³³ and is 68-fold higher than that of the most efficient designed Kemp eliminase reported to date.³⁴ Interestingly, the introduction of three mutations in Mb-H64V did not cause major structural rearrangement (RMSD of 0.16 \AA was calculated from the alignment of Mb-H64V and FerrEICat in Pymol), despite a drastic increase in the protein's catalytic activity (Figure 3.8).

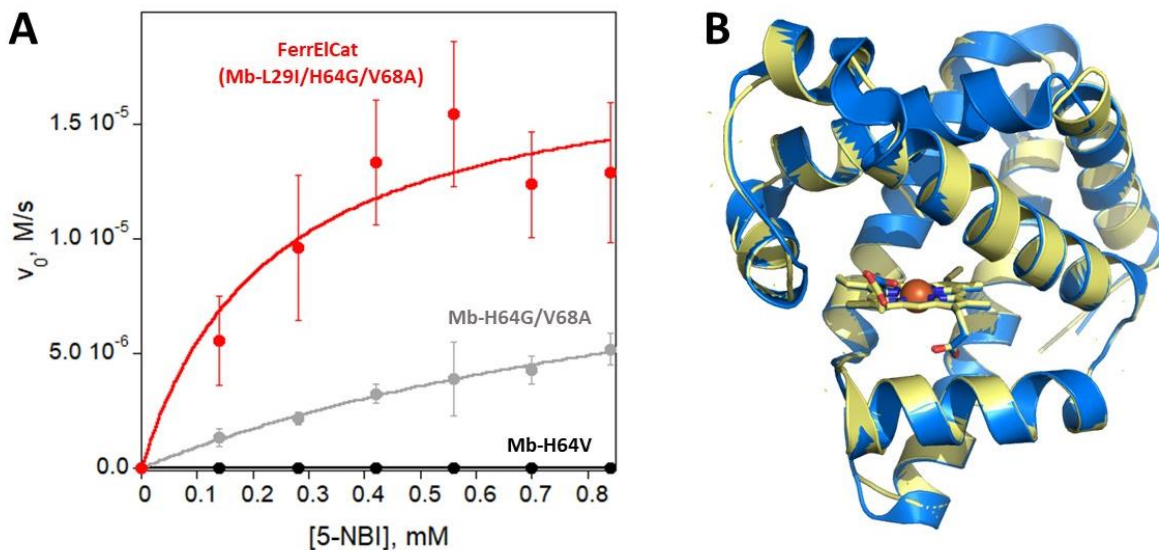


Figure 3.8. A. Michaelis-Menten traces for Mb-H64V and derived mutants showing the improvement in activity after the introduction of three mutations. **B.** The overlay of Mb-H64V (PDB 6CF0) and FerrEICat (the structure was determined by Dr. Jeremy Tame group and deposited to the wwPDB) showing the minor structural changes after mutagenesis.

We have previously suggested that myoglobin variants are following the redox-mediated mechanism of Kemp elimination. This implies that proteins should be able to transfer electrons from the cofactor (heme) to the substrate molecule. The reduction potential of a protein is a measure of the protein's ability to transfer electrons: proteins with more positive reduction potential act as oxidizing agents by donating electrons to the acceptor with lower potential.³⁵ It is known that the environment around heme plays a significant role in protein's ability to transfer electrons,³⁶ so we suggested that the difference in the activity of the proteins designed in this work might originate from the difference in the reduction potential. Surprisingly, the introduced mutations had a minor effect on the reduction potentials of myoglobin mutants (Figure 3.9).

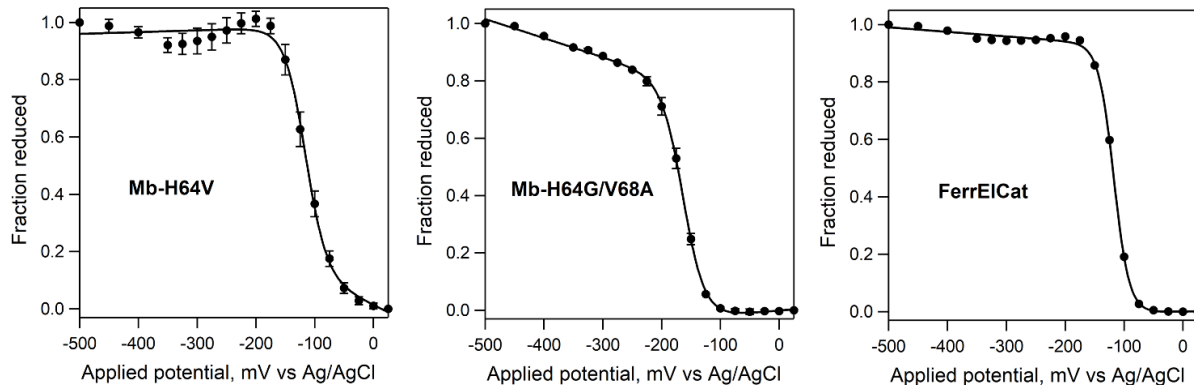


Figure 3.9. Spectroelectrochemical determination of reduction potentials of selected myoglobin mutants. The proteins were analyzed in 20 mM TRIS-HCl (pH 8.0) at 20°C in presence of the mediator (100 μ M phenazine sulfate). The determined reduction potentials (vs Ag/AgCl) for one-electron transfer were as follows: -117.11 ± 4.25 mV (Mb-H64V), -162.33 ± 0.90 mV (Mb-H64G/V68A), -117.32 ± 0.94 mV (FerrEICat).

To catalyze a redox-mediated reaction, protein needs to have a redox-active unit. The heme in myoglobin contains iron, a redox-active metal ion that can transition between oxidized (Fe^{3+}) and reduced (Fe^{2+}) states. As tested proteins did not show major differences in reduction potentials, we decided to confirm the importance of the redox-active metal for Kemp elimination by myoglobin mutants. For this, iron-containing heme in FerrEICat was substituted by Zn (II)-protoporphyrin IX. This molecule resembles heme, except for the central zinc ion which is a redox inactive metal that exists in the form of a divalent ion. Zinc-substituted myoglobin did not show any activity in the Kemp elimination reaction, confirming the importance of the redox metal for the catalysis (Figure 3.10.).

As the formation of a base in a protein's active site is a pH-dependent process, the activity of protein utilizing the acid-base mechanism for Kemp elimination will

change with the concentration of protons, whereas the redox-mediated mechanism should be pH-independent.

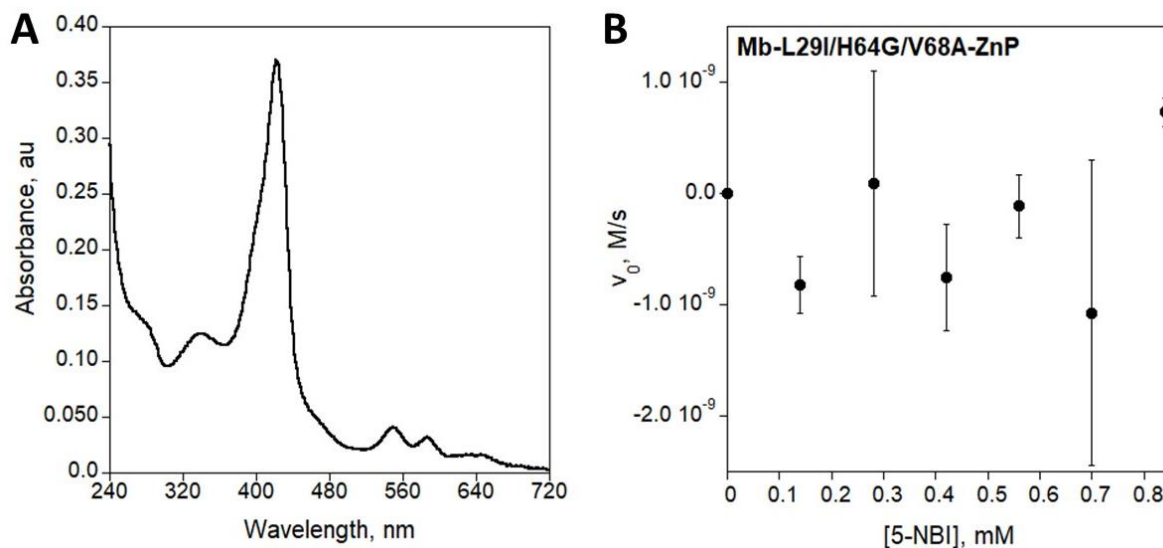


Figure 3.10. A. The absorbance spectrum for zinc-porphyrin-substituted FerrEICat. **B.** Michaelis-Menten kinetic trace for the Kemp elimination reaction catalyzed by FerrEICat-Zn-P at pH 8.0.

Typically, the amino acid residues ionizable at physiological pH can act as a base (glutamate, aspartate and histidine).³⁷ The close examination of the environment around FerrEICat's active site did not reveal the presence of the amino acid residue that can act as a base for Kemp elimination. The analysis of FerrEICat's enzymatic activity did not show significant pH dependence (Figure 3.11), even though we did observe the drop in activity at pH 7.5. A similar result was previously obtained for cytochrome p450 and needs further investigation. Overall, these results provide evidence for the redox-mediated mechanism of Kemp elimination reaction catalyzed by FerrEICat.

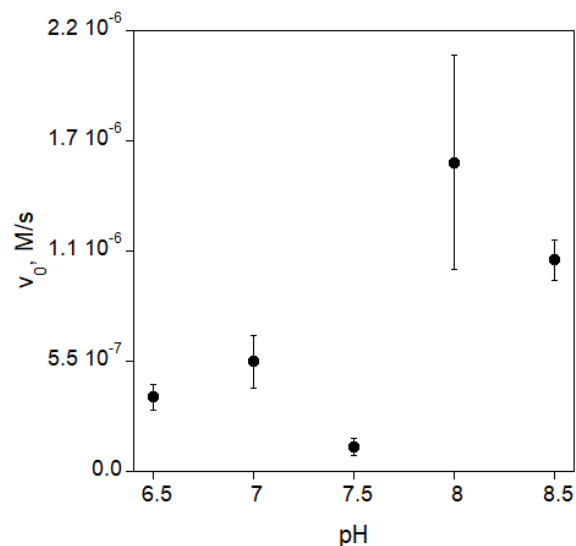


Figure 3.11. pH dependence of FerrEICat's activity in Kemp Elimination. Protein was analyzed at 5 nM with 35 μ M substrate in 1 mM L-ascorbic acid, 0.1 μ M SOD, 20 nM catalase, 1.5% acetonitrile, 20 mM Tris (pH 8.0).

3.2.4. NMR-guided directed evolution allowed to improve affinity of AlleyCat to the inhibitor

NMR-guided directed evolution of myoglobin scaffold allowed us to develop a highly efficient protein by combining only three mutations. To test the capacity of this method, we decided to evolve another scaffold. Our lab has done extensive work studying AlleyCat, a calmodulin-based Kemp eliminase. Using computational methods and a minimalist approach, Korendovych et al. identified the position for a single mutation inside the hydrophobic pocket of the calmodulin C-terminal domain which allowed them to install enzymatic activity in previously non-enzymatic protein.⁴ AlleyCat, a 74-residue C-terminal domain of calmodulin, contains F92E mutation and possesses Kemp elimination activity. Later this protein was subjected to seven rounds of evolution using saturation mutagenesis, error-prone PCR (epPCR) and gene shuffling.³⁸ These

methods allowed to improve protein activity 220 times over the original design.

Considering this previous success and the simplicity of this scaffold, AlleyCat presents a great candidate for further improvements using NMR-guided directed evolution.

The evolution of AlleyCat followed the same protocol as the one used for myoglobin. The identification of hot spots for further mutations had more limitations: as AlleyCat is an allosterically controlled catalyst, no mutations were introduced in the calcium-binding sites (EF-hands). In the first round of NMR-guided directed evolution of AlleyCat7, we were able to identify a productive mutation at position 125 (I125H), which improved protein catalytic efficiency, mainly due to the increase in k_{cat} , and resulted in AlleyCat7-I125H named AlleyCat8 (Table S 1). Then, the productive mutations K115P and T146R were identified and added to the AlleyCat8 sequence, which allowed us to create AlleyCat9 and AlleyCat10 (Table S 2).

Efficient substrate binding facilitates the reaction as it helps to properly position the substrate in the active site by finding the most appropriate orientation.³⁹ Usually, mutations around the active site contribute to the substrate binding as they either directly interact with the substrate or form the environment for sufficient binding. Many research efforts highlighted the importance of protein dynamics in substrate transition state binding and catalysis, suggesting that the structural changes outside the binding pocket have an important contribution to the protein's function. Even though the methods for protein evolution can generate improved variants, the identification of the distal positions crucial for catalysis still remains a challenge.⁴⁰ We have shown that by using NMR to guide protein design we were able to identify productive mutations far from the active site. While analyzing the mutation introduced into AlleyCat using

epPCR, gene shuffling and NMR-guided directed evolution, we can see that all three mutations identified by NMR are located outside the binding pocket (Figure 3.12).

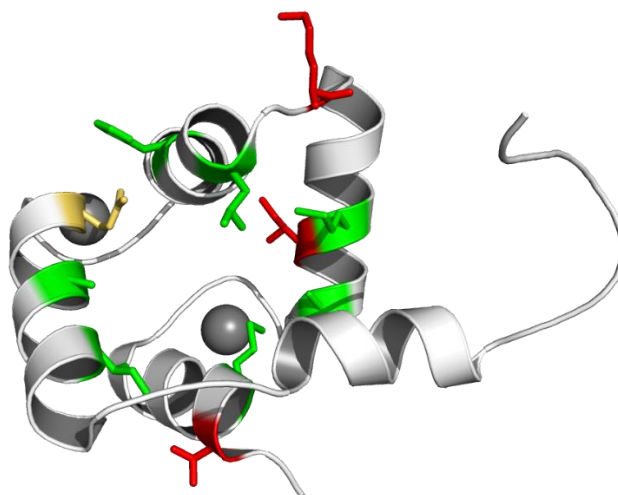


Figure 3.12. The mutations identified in 10 rounds of directed evolution of AlleyCat scaffold. The F92E mutation is colored in yellow, the mutations identified with epPCR and gene shuffling are in green, the positions identified using NMR-guided directed evolution are highlighted in red. The calcium ions are presented as grey spheres. The mutations were mapped on AlleyCat scaffold (PDB 2KZ2).

The design of our method for protein evolution suggests that the introduced mutations will affect the substrate transition state binding as we are altering the amino acid residues important for the transition state analog binding to the protein scaffold. To evaluate the impact of these mutations, we have measured the affinity of AlleyCat mutants to 6-nitrobenzotriazole (6-NBT), the analog of a substrate transition state for Kemp elimination reaction, using isothermal titration calorimetry (ITC). ITC is a method that allows us to determine the thermodynamic parameters of interactions between molecules. AlleyCat proteins, from AlleyCat to AlleyCat10, were obtained using different strategies and each next round of mutagenesis improved enzyme activity mainly by

increasing k_{cat} (Table S 2). ITC analysis showed the improvement of 6-NBT binding to the tested proteins (Table 3.2, Figure S 14).

Table 3.2. Thermodynamic parameters of 6-nitrobenzotriazole binding to AlleyCat mutants determined by ITC. The titrations were performed in 20 mM HEPES, 100 mM NaCl, 10 mM CaCl₂ (pH 7.0), 2% acetonitrile at 25°C. K_d stands for the dissociation constant in micromolar, ΔH is the change in enthalpy in kcal/mol.

Protein	Number of sites	K_d , μM	ΔH , kcal/mol
AlleyCat	-	-	-
AlleyCat 1	-	-	-
AlleyCat 2	-	-	-
AlleyCat 3	0.741 \pm 0.016	64.9 \pm 8.4	-17.4 \pm 1.5
AlleyCat 4	0.783 \pm 0.010	50.1 \pm 4.1	-16.4 \pm 0.8
AlleyCat 5	0.936 \pm 0.01	31.2 \pm 3.7	-12.7 \pm 0.7
AlleyCat 6	0.803 \pm 0.008	26.3 \pm 1.7	-15.2 \pm 0.4
AlleyCat 7	0.812 \pm 0.006	18.8 \pm 0.9	-13.9 \pm 0.2
AlleyCat 8	0.920 \pm 0.008	13.4 \pm 0.7	-13.9 \pm 0.2
AlleyCat 8-T146R	0.759 \pm 0.004	10.8 \pm 0.4	-12.4 \pm 0.1
AlleyCat 9	0.749 \pm 0.004	11.4 \pm 0.5	-13.3 \pm 0.2
AlleyCat 10	0.869 \pm 0.005	9.68 \pm 0.42	-13.9 \pm 0.2

'-' means that the parameters could not be determined using ITC

The contribution of the long-range interactions to protein catalysis has been extensively discussed in the literature.^{41, 42} For years we have imagined protein-substrate complexes as rigid structures. However, many studies have shown the importance of protein dynamics in catalysis. Proteins undergo many microstates upon substrate binding until the most appropriate is found. Conformational flexibility of

proteins is also required for catalytic conversion and product release. These processes involve some degree of movement from the whole protein molecule which explains the connection between the long-distance interactions and substrate binding. Even though their importance has been acknowledged, the protein design toolbox still lacks efficient methods which can pinpoint such interactions. Our studies showed that NMR-guided directed evolution can be used to identify the “hot spots” far from the active site. The amino acids at the positions identified by NMR do not interact directly with the substrate but rather contribute to the efficient protein conformation for further transitions state binding and catalysis.

3.3. Conclusions

With the increasing demand for proteins with new or improved functions, the development of effective methods for protein engineering becomes a priority. We have employed NMR to guide the directed evolution of proteins by identifying the regions important for substrate binding and catalysis. We have shown that NMR-guided directed evolution can be used to identify positions for productive mutations outside the active site without utilizing extensive structural characterization of the protein scaffold. We confirmed that chemical shift perturbation analysis facilitates the identification of positions for effective mutations. We have found productive mutations at all 18 tested hot spots, many of which are located far from the active site. The productive mutations introduced in AlleyCat7 are also located outside the binding pocket but improved the protein's affinity to the substrate analog. As the identification of positions for productive distal mutations still remains a challenge for most methods, NMR-guided directed evolution provides new possibilities for protein design.

By applying gene shuffling, we have created the most active designed Kemp eliminase, FerrEICat, reported to date by combining only three mutations identified using an NMR-guided approach. The introduction of three mutations did not cause major structural changes in the proteins but resulted in the unprecedented activity of $1.6 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$. FerrEICat requires redox-active metal, and its enzymatic activity is pH-independent. These results suggest a redox-mediated mechanism for Kemp elimination promoted by FerrEICat. Our future work will include further elucidation of the factors contributing to the extraordinary enzymatic activity of FerrEICat and the application of the developed method to the evolution of other protein scaffolds.

3.4. Experimental

Protein production and purification of myoglobin variants. A plasmid, pET22b/Mb-WT, encoding recombinant sperm whale myoglobin, was a kind gift from Prof. Rudi Fasan's lab (University of Rochester). The protein sequence contains D122N mutation compared to the wild-type sperm whale myoglobin. The plasmid encoding myoglobin H64V, pET28a/Mb-H64V, was prepared in our lab using SOE-PCR. The plasmids with myoglobin mutants were obtained from the NNK library screenings at specific positions or after the site-directed mutagenesis.

The plasmids (50-100 ng) were transformed into *E. coli* BL21 (DE3) cells (New England Biolabs) using heat shock and cultures were incubated at 37°C with shaking (200 rpm) for 1 h. An aliquot was plated on LB agar plate containing 50 µg/mL kanamycin (BioBasic Inc.) and the plate was incubated at 37°C for 10-16 H. The starter cultures were prepared by inoculating 20 mL of 1x LB medium supplemented with kanamycin with a single bacterial colony and grown for 5-12 H at 37°C with shaking (200 rpm). A starter culture (10 mL) was diluted in 1 L of 1x LB (with 50 µg/mL kanamycin) and grown at 37°C until OD₆₀₀ reached 0.6-0.8 (this optical density range was important for the successful protein expression). It was also important to provide proper aeration for the culture, so the 2.8-L glass Fernbach flasks were used to fit 1 L of bacterial culture and the culture was incubated with shaking at 180 rpm in New Brunswick Excella 25R shaker. After the cultures reached the density, the temperature was set to 25°C and the cultures were incubated for 15 min at new conditions. δ-aminolevulinic acid hydrochloride (0.3 mM, TCI) was added to the culture as heme precursor. After 5-10 min, the protein production was induced by adding 0.25 mM IPTG

(BioBasic Inc.) and the cultures were incubated at 25°C for 20 H. Cells were harvested by centrifugation at 4,000 xg for 20 min. The typical yield of the cell paste was 3-5 g per 1 L of culture. The cell paste was flash frozen in liquid nitrogen and preserved at -80°C until further use.

The cells were resuspended in buffer containing 25 mM Tris-HCl (pH 8.0) and sonicated at 20-23 W on ice for 10 min (20 s pulse, 20 s rest) on ice. The soluble protein fraction was separated by centrifugation at 20,000 xg for 30 min at 4°C and applied on 5-mL preequilibrated with resuspension buffer Ni-NTA agarose column (Clonthech). The resin was then washed with the resuspension buffer until no protein was eluting from the column (the presence of proteins was checked using Bradford reagent). The protein was then eluted with buffer containing 25 mM Tris-HCl (pH 8.0) and 50 mM imidazole. Fractions containing the protein were combined, concentrated down to ~ 3 mL using 5K MWCO spin concentrator (Corning) and exchanged into 20 mM HEPES (pH 7.0) using an Econo-Pac 10 DG desalting column (BioRad). The final purification was performed using HiTrap SP-sepharose column (GE Healthcare) and NGC Quest 10 Plus Chromatography System (Bio-Rad). Protein was eluted with a linear gradient of NaCl (0-600 mM) in 20 mM HEPES (pH 7.0). In many cases the protein of interest was eluted with higher salt concentrations (the right shoulder of the peak) and presented the minor fraction compared to the total heme-containing protein mixture. Protein fractions were examined using UV-Vis spectroscopy and the fractions which contained proteins with proper Soret band were combined, concentrated and buffer exchanged into 20 mM HEPES (pH 7.0) as described previously. The final buffer exchange acted also as an additional purification step and only fractions with proper

Soret bands were combined and used for further experiments. Protein concentration was determined using the extinction coefficient of $157,000 \text{ M}^{-1}\text{cm}^{-1}$. The SDS-PAGE analysis (10%) confirmed the purity (> 95%) of proteins.

Substitution of heme cofactor to Zn-porphyrin in FerrEICat. Purified FerrEICat in 25 mM HEPES (pH 7.0) was used for the heme extraction. Protein sample (~1mL) was chilled on ice and then the pH of solution was adjusted to 2 by addition of small aliquots of 1 M HCl (the pH was monitored using indicator paper). After the addition of the equal volume of ice-cold butanone (Sigma) to the protein solution, the solution was vortexed for 10 sec and incubated on ice for 15 min. After the incubation, the solution had two layers: the bottom layer contained protein without heme in buffer and the top layer contained heme in butanone. The bottom layer was carefully removed using a syringe with thin needle, placed in the Slide-A-Lyzer 7K MWCO dialysis cassette (Thermo Scientific) and dialyzed against 3 L of water for 1 H at 4°C. The sample was then dialyzed again for 1 H and the third time over night at the same conditions. Zn (II)-protoporphyrin IX (Zn-P) was prepared by dissolving the powder in DMF. After dialysis, the apo-protein was incubated with the excess (3-5 eq based on the concentration of the original protein sample used for the experiment) of Zn-porphyrin for 15 min on ice. Then, the equal volume of 20 mM Tris-HCl (pH 7.0) was added to the protein-Zn-P mixture. To remove free Zn-P, the mixture was applied on Sephadex G-25 (BioRad) gravity column and Zn-FerrEICat was eluted with 20 Tris-HCl (pH 8.0). Zn-FerrEICat was characterized by the Soret band at 424 nm, and Q-bands at 552 nm and 589 nm. Protein concentration was determined using the absorbance at 424 nm and the extinction coefficient of $157,000 \text{ M}^{-1}\text{cm}^{-1}$.

Expression and purification of AlleyCat variants. Plasmids encoding AlleyCat variants were prepared in our lab by cloning the C-terminal domain of calmodulin (AlleyCat) sequence into pET-SUMO vector using pET-SUMO cloning kit (Invitrogen). Plasmids were transformed into *E. coli* BL21 (DE3) cells using heat shock method. The aliquot was plated on LB agar plate containing kanamycin (50 µg/mL) and incubated at 37°C for 10-16 h. A single colony obtained from the plate was then inoculated into 20 mL of 1x LB media supplemented with kanamycin (50 µg/mL) and grown at 37°C for 12-16 H to prepare the starter culture. 10 mL of the starter culture were added to 1 L of 1x LB media (with 50 µg/mL kanamycin) and the culture was grown at 37°C with shaking at 200 rpm. After the OD₆₀₀ reached 0.6-0.8, the temperature was lowered to 18°C and protein expression was induced by adding 0.5 mM IPTG. After incubation at 18°C at 200 rpm for 10-12 H, cells were harvested by centrifugation at 4,000 xg for 20 min at 4°C. The typical yield of the cell paste was 3-5 g per 1 L of cell culture. The cell paste was flash frozen in liquid nitrogen and stored at -80°C.

Cell pellets were resuspended in 25 mM Tris-HCl (pH 8.0), 20 mM imidazole, 10 mM CaCl₂, 300 mM NaCl and sonicated (Microson) on ice for 10 min (20 s pulse, 20 s rest) at 20-23 W on ice. The soluble protein fraction was separated by centrifugation at 20,000 xg for 30 min at 4°C and applied on the preequilibrated 5-mL Ni-NTA column. The column was washed with the resuspension buffer until no protein was eluted from the column. The protein of interest was eluted with 25 mM Tris-HCl (pH 8.0), 10 mM CaCl₂, 300 mM NaCl and 250 mM imidazole. Fractions containing the protein (verified by the SDS-PAGE analysis in 15% gel) were combined and exchanged into the SUMO

cleavage buffer (50 mM Tris-HCl (pH 8.0), 75 mM NaCl) using Econo-Pac 10 DG desalting column (Bio-Rad). The eluted proteins contained a His6 tag and SUMO tag (~11 kDa) attached to the N-termini of the proteins. SUMO tag was cleaved off using SUMO protease in presence of 1 mM DTT (Sigma) and 0.5 mM EDTA (Gibco) while the mixture was incubated at 30°C for 4-14 H. The required amount of the SUMO protease was calculated as following: $A_{280}(\text{SUMO-AlleyCat}):A_{280}(\text{SUMO protease}) = 200:1$. After digestion, the mixture was exchanged into the storage buffer (20 mM HEPES (pH 7.0), 10 mM CaCl_2 , 100 mM NaCl) to remove DTT and EDTA and the obtained solution was passed through Ni-NTA column pre-equilibrated with the same buffer. Collected fractions were analyzed using SDS-PAGE to confirm the presence of the protein without tag. The fractions were further combined and concentrated using 5K MWCO spin concentrator (Corning). The final purification was performed using HiTrap Q-sepharose column (GE Healthcare) and NGC Quest 10 Plus Chromatography System (Bio-Rad). The protein was eluted with the gradient of 100-800 mM NaCl in 20 mM HEPES (pH 7.0), 10 mM CaCl_2 . Fractions containing protein were combined, concentrated and exchanged into storage buffer (20 mM HEPES (pH 7.0), 10 mM CaCl_2 , 100 mM NaCl). Protein concentration was determined by measuring the absorbance at 280 nm using the extinction coefficient of $2980 \text{ M}^{-1}\text{cm}^{-1}$ calculated using Expasy (<http://expasy.org/tools/protparam.html>). Finally, the purity (> 95%) of the final samples was checked on SDS-PAGE (15%).

Expression and purification of SUMO protease. pTB145 plasmid encoding SUMO protease sequence was a gift from the Prof. Bradley Pentelute's lab (Massachusetts Institute of Technology). The plasmid was transformed into *E. coli*

BL21-CodonPlus (DE3)-RP (Agilent) using heat shock and aliquot was plated on LB agar plate with 100 µg/mL of ampicillin and 34 µg/mL of chloramphenicol. The starter culture was prepared by transferring a colony in 15 mL of 1x LB supplemented with 100 µg/mL of ampicillin and 34 µg/mL of chloramphenicol and grown at 37°C for 10-16 H. A starter culture (10 mL) was diluted in 1 L of 1x LB supplemented with ampicillin and chloramphenicol and grown at 30°C until OD₆₀₀ 0.6-0.8. The protein expression was induced by addition of 0.5 mM IPTG and the culture was further incubated at 30°C for 4-5 H. The cells were harvested by centrifugation at 4,000 xg for 20 min at 4°C. The typical yield of cell paste was 3 g per 1 L of culture.

Cell paste was resuspended in the buffer containing 50 mM Tris-HCl (pH 8.0), 5% glycerol, 10 mM 2-mercaptoethanol and 25 u/mL benzonase (Millipore). Cells were lysed by sonication on ice for 10 min (20 s pulse, 20 s rest) at 20-23 W and then sample was incubated at 37°C for 30 min to ensure proper DNA digestion by benzonase. To separate the soluble fraction, the cell lysate was spun down at 20,000 xg for 30 min at 4°C. The DNA precipitation was performed by adding the streptomycin sulfate solution to a final concentration of 0.8% (streptomycin sulfate salt was dissolved in buffer and added dropwise to the protein solution). After 30 min of stirring, the lysate was spun down at 20,000 xg for 30 min at 4°C. The supernatant was then loaded onto a 5-mL Ni-NTA column (Clontech) pre-equilibrated with resuspension buffer without 2-mercaptoethanol. The column was washed with 50 mM Tris-HCl (pH 8.0), 5% glycerol, 10 mM 2-mercaptoethanol, 40 mM imidazole until no protein was elute from the column (the fractions were checked for protein presence with Bradford reagent). The protein was eluted using the resuspension buffer with 250 mM imidazole. Fractions containing

SUMO protease identified by Bradford assay were pooled and loaded onto a 5-mL Q-sepharose column preequilibrated with the resuspension buffer. The column was washed with the resuspension buffer and protein was eluted with 50 mM Tris-HCl (pH 8.0), 5% glycerol, 10 mM 2-mercaptoethanol and 200 mM NaCl, Fractions containing protein with A_{280}/A_{260} ratio of >1 were combined and concentrated using 10K MWCO centrifugal concentrator (Corning). The protein was flash frozen in liquid nitrogen and stored at -80°C .

Reduction of myoglobin mutants. The reduction of myoglobin-based proteins was performed inside the glove box (MBraun) which maintained the low oxygen atmosphere (<1 ppm). The proteins were degassed by placing the solutions in the Schlenk vessels and applying 10 cycles of nitrogen refill and evacuation using the Schlenk line. After the proteins were transferred inside the glove box, 10 equivalents of the freshly prepared dithionite were added to reduce the proteins. The reduced proteins were characterized by the sharp Soret band at 432 ± 2 nm and the concentration was estimated using the extinction coefficient of $157,000 \text{ M}^{-1}\text{cm}^{-1}$ or the experimentally determined values using standard pyridine hemochromagen assay.

To determine the concentration of dithionite, 20-30 mg of solid sodium dithionite (Riedel-de Haen) as well as potassium ferricyanide (Sigma) were brought inside the glove box. Each solid reagent was dissolved in 1 mL of degassed milliQ water to prepare stock solutions. Dithionite stock was further diluted 10- or 30-fold, depending on the protein stock concentration, making working solution of dithionite. Next, two 1-mL solutions were prepared: A - potassium ferricyanide stock was diluted 100 times in water; B - the mixture of 100-fold diluted ferricyanide stock and 100-fold diluted working

solution of dithionite was prepared in water. Both solutions were removed from the glove box and the absorbances were measured at 420 nm using UV-Vis spectrometer (Agilent 8453 or Cary 60). The reducing equivalence of the working solution of dithionite was calculated from the difference in absorbances of the solutions using extinction coefficient of $1,020 \text{ M}^{-1}\text{cm}^{-1}$ at 420 nm.

Kinetic characterization of myoglobin mutants. Extinction coefficients of $15,800 \text{ M}^{-1}\text{cm}^{-1}$ at 380 nm was used for the product of Kemp elimination, 2-hydroxybenzoxazole. Catalytic efficiency (k_{cat}/K_M) as well as individual kinetic parameters (k_{cat} and K_M , where possible), were determined by fitting the dependence of initial rates on substrate concentration (final concentration of 140-840 μM) to the Michaelis-Menten equation $v_0 = k_{\text{cat}}[E][S]/(K_M+[S])$. For proteins with high K_M values, where substrate saturation could not be achieved due to solubility limits, k_{cat}/K_M was determined by fitting data to $v_0 = (k_{\text{cat}}/K_M)[E][S]$. In all cases the rates are corrected for the background rates in the appropriate buffers without the enzymes. The substrate, 5-nitrobenzisoxazole, was prepared as 100 mM stock in acetonitrile inside the glovebox (MBraun). Degassed protein samples were reduced by adding ca. 10 equivalents of sodium dithionite (Riedel-de Haen) inside a glove box. Concentrations of reduced proteins were determined using a Soret band (typically around 434 nm) and the extinction coefficient of $157,000 \text{ M}^{-1}\text{cm}^{-1}$. The quality of reduction was assessed by examining the position and shape of Soret and Q-bands. Protein solutions containing reduced myoglobin mutants (10-200 nM) in 40 mM Tris-HCl (pH 8.0) with ascorbate (2 mM), SOD (0.2 μM), catalase (40 nM), were prepared inside the glove box and transferred to gas-tight syringes. The substrate was prepared as 2x solution in water containing 3% acetonitrile

inside the glove box and transferred in gas-tight syringes. The syringes with reaction solutions were removed from the glove box one set a time and attached to the stopped-flow spectrometer (Applied Photophysics SX20). Reactions were initiated by mixing reduced enzyme in buffer and solutions of 5-nitrobenzisoxazole in water in 1:1 ratio. The final reaction mixtures contained reduced protein (5-100 nM) in 20 mM Tris-HCl (pH 8.0) with ascorbate (1 mM), SOD (0.1 μ M), catalase (20 nM), 1.5 % acetonitrile and variable concentrations of the substrate. The reaction mixtures for each substrate concentration were prepared and analyzed in triplicates. Product formation was monitored at 25°C (temperature was controlled by Thermo Scientific water bath) for the time interval of 0.1-10 s. The slopes were calculated for the linear portion of the kinetic traces. The initial rates were calculated by dividing the slopes by the product extinction coefficient and correcting for the pathlength (the pathlength of the cell was 5 mm).

The activity of zinc-substituted FerrEICat was analyzed in oxygen-containing environment as the protein does not need reducing conditions. The protein solution was prepared in 40 mM Tris (pH 8.0) at the concentration of 100 nM. The substrate solutions were prepared as described above but outside the glove box using oxygenated water. The final concentrations of the reactions were: 50 nM FerrEICat-Zn-P, 140-840 μ M substrate, 20 mM Tris (pH 8.0) and 1.5% acetonitrile. The concentration of zinc-substituted protein was determined using the extinction coefficient of 157,000 $M^{-1}cm^{-1}$ at 424 nm.

The pH dependence of FerrEICat enzymatic activity was assessed using the anaerobic protocol described above, except different buffers were used (all at 20 mM; MES at pH 6.5, HEPES at pH 7.0 and 7.5, TRIS at pH 8.0 and pH 8.5).

Kinetic assay for uncatalyzed Kemp elimination. The rate constant for uncatalyzed Kemp elimination (k_{uncat}) was determined using several different concentrations (five different concentrations ranging from 0-100 mM) of buffer containing Tris-HCl (pH 8.0) at 25°C and extrapolated to the zero buffer concentration. The product formation was monitored at 380 nm after the addition of 0.2 mM 5-nitrobenzisoazole to the buffer solutions, maintaining acetonitrile concentration at 1.5%, using UV-Vis spectrophotometer (Agilent Cary 60). For each different buffer concentration, aliquots of the reaction mixture were transferred to the quartz cuvette (Starna cells, Inc.) and the change in absorbance was monitored for 2 days. The kinetic traces were fit to the pseudo first order reaction equation ($A = c*(1-e^{-kt})$), where A is the absorbance of the product at 380 nm, c is the proportionality constant, k is the rate constant for uncatalyzed reaction (k_{uncat}) and t is time in min. The k_{uncat} at zero buffer concentration was determined by extrapolation of linear trendline to the zero buffer concentration (Figure S 15.).

Spectroelectrochemical analysis of reduction potential for myoglobin mutants. The reduction potentials of myoglobin variants were measured using a platinum honeycomb spectrochemical electrode (0.17 cm) with a Ag/AgCl reference electrode connected to the WaveNow potentiostat (all purchased from Pine Research Instrumentation). All reduction potential mentioned here were determined against Ag/AgCl reference electrode (electrode potential reported as +199 mV vs NHE by manufacturer). The redox titration was performed using 1.2 mL of working solution containing ~ 20 μM protein and 100 μM phenazine methosulfate (Sigma) as a redox mediator in 20 mM HEPES (pH 7.0) or 20 mM Tris-HCl (pH 8.0). A highly positive

potential of +100 mV vs Ag/AgCl reference electrode was initially applied to ensure complete oxidation of protein. Next, the potential was applied with an increment of 25 mV from +25 mV to -500 mV (vs Ag/AgCl) with the equilibration time of 5 min at 20°C for each step. The absorbance spectra of the protein containing mixtures at each potential were recorded using UV-Vis spectrophotometer (Agilent 8453). The absorbance of the protein solution was corrected for the contribution of the mediator. The absorbance at 434 nm was normalized to 800 nm and used to determine the fraction of reduced protein in the sample using the following equation: Fraction reduced = $(A_{434} - A_{434, \text{min}})/(A_{434, \text{max}} - A_{434, \text{min}})$. Fraction reduced was plotted as a function of the applied potential and the midpoint reduction potential (E_m) was determined using the following equation: Fraction reduced = $(A + g_1 \cdot E_{\text{app}}) + (B + g_2 \cdot E_{\text{app}}) / (1 + \exp((E_m - E_{\text{app}})/b))$, where A represents the upper asymptote, B represents the difference between the upper and lower asymptotes, g_1 and g_2 represent the slopes of the upper and the lower asymptotes, respectively, b is the growth rate, E_{app} is applied potential, E_m is midpoint reduction potential..

Circular dichroism (CD) spectroscopy of myoglobin mutants. All CD spectra of myoglobin variants were recorded using Jasco J-715 CD spectrometer in continuous mode with 1 nm bandwidth, 2 nm data pitch, scan rate of 50 nm/min with 8 s averaging time. The final spectra represent a buffer-subtracted average of three runs. The CD spectra of non-reduced proteins in the far-UV region (200-260 nm) were collected using quartz cuvette with 1 mm pathlength while for the Soret band region (390-470 nm) quartz cuvette of 1 cm pathlength was used. The spectra at Soret band region (390-470 nm) were obtained to determine mean residue ellipticity values (MRE) assuming protein

binds heme in a 1:1 ratio. Proteins were degassed and transferred inside the glove box as described above. Protein stocks were diluted in 2 mM Tris-HCl (pH 8.0) to 5 μ M, transferred outside the glove box and the spectra were recorded for oxidized protein. For the analysis of the reduced protein, the stock was diluted in the same way as for the oxidized sample and ten equivalents of sodium dithionite were added to the protein inside the glove box. The concentrations of the reduced proteins were calculated based on the Soret band maxima using the corresponding extinction coefficients. Sample absorbance never exceeded 2 at all wavelengths. The mean residue ellipticity (MRE, $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$) values were calculated using the following equation ($\text{MRE} = \theta/(10\cdot c\cdot l\cdot N)$), where θ (mdeg) is ellipticity, l (cm) is the pathlength of the cuvette, c (M) is the protein concentration and N is the number of residues.

Determination of the dissociation constants for AlleyCats and 6-nitrobenzotriazole (inhibitor). The thermodynamic parameters of 6-nitrobenzotriazole (6-NBT) binding to AlleyCat proteins were measured using a MicroCal PEAQ-ITC instrument (Malvern Panalytical). Proteins were dialyzed against buffer (20 mM HEPES, 100 mM NaCl, 10 mM CaCl_2 , pH 7.0) containing 2% acetonitrile, filtered using 0.22 μ m low protein binding PES filter (Santa Cruz Biotechnology, Inc.) and the concentration of each sample was determined by UV-vis spectroscopy using extinction coefficients of $2,980\text{ M}^{-1}\text{cm}^{-1}$ at 280 nm. 6-NBT was prepared by dissolving the powder (AK Scientific) in the protein dialysis buffer to obtain 1 mM solution. All titrations were performed at 25°C in the high feedback mode with 750 rpm stir speed and an appropriate equilibration time between injections (150 s). The protein sample ($\sim 100\ \mu\text{M}$) was placed in the calorimeter cell and 1 mM solution of 6-NBT (in the syringe) was added to the

protein in eighteen 2- μ L aliquots. As a control, 1 mM 6-NBT solution was titrated into the dialysis buffer.

The analysis, including baseline correction, peak integration and correction for heat of dilution observed at the protein saturation with inhibitor, was performed using the Microcal PEAQ-ITC analysis software provided by the manufacturer. To obtain binding parameters for each reaction, the data were fitted to the one set of sites model. Each titration was repeated at least two times.

3.5. References

1. Faber, M. S.; Whitehead, T. A., Data-driven engineering of protein therapeutics. *Curr. Opin. Biotechnol.* **2019**, *60*, 104-110.
2. DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D., Protein design, a minimalist approach. *Science* **1989**, *243* (4891), 622-8.
3. Hellinga, H. W., Rational protein design: combining theory and experiment. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (19), 10015-7.
4. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (17), 6823-7.
5. Dalby, P. A., Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 473-80.
6. Jackel, C.; Kast, P.; Hilvert, D., Protein design by directed evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153-73.

7. Chen, K.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (12), 5618-22.
8. Stemmer, W. P., Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **1994**, *370* (6488), 389-91.
9. Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H., Enzymatic assembly of carbon-carbon bonds via iron-catalysed sp(3) C-H functionalization. *Nature* **2019**, *565* (7737), 67-72.
10. Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H., Enantiodivergent alpha-amino C-H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141* (25), 9798-9802.
11. Liu, Z.; Qin, Z. Y.; Zhu, L.; Athavale, S. V.; Sengupta, A.; Jia, Z. J.; Garcia-Borras, M.; Houk, K. N.; Arnold, F. H., An Enzymatic platform for primary amination of 1-aryl-2-alkyl alkynes. *J. Am. Chem. Soc.* **2022**, *144* (1), 80-85.
12. Dunham, N. P.; Arnold, F. H., Nature's Machinery, Repurposed: Expanding the repertoire of iron-dependent oxygenases. *ACS Catal.* **2020**, *10* (20), 12239-12255.
13. Dias, D. M.; Ciulli, A., NMR approaches in structure-based lead discovery: recent developments and new frontiers for targeting multi-protein complexes. *Prog. Biophys. Mol. Biol.* **2014**, *116* (2-3), 101-12.
14. Kohen, A., Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Acc. Chem. Res.* **2015**, *48* (2), 466-73.
15. Rajakumara, E.; Abhishek, S.; Nitin, K.; Saniya, D.; Bajaj, P.; Schwaneberg, U.; Davari, M. D., Structure and cooperativity in substrate-enzyme interactions:

perspectives on enzyme engineering and inhibitor design. *ACS Chem. Biol.* **2022**, *17* (2), 266-280.

16. Bhabha, G.; Biel, J. T.; Fraser, J. S., Keep on moving: discovering and perturbing the conformational dynamics of enzymes. *Acc. Chem. Res.* **2015**, *48* (2), 423-30.

17. Callender, R.; Dyer, R. B., The dynamical nature of enzymatic catalysis. *Acc. Chem. Res.* **2015**, *48* (2), 407-13.

18. Kemp, D. S.; Casey, M. L., Physical organic chemistry of benzisoxazoles. II. Linearity of the Bronsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **1973**, *95*, 6670-6680.

19. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453* (7192), 190-195.

20. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (10), 3790-5.

21. Li, A.; Wang, B.; Ilie, A.; Dubey, K. D.; Bange, G.; Korendovych, I. V.; Shaik, S.; Reetz, M. T., A redox-mediated Kemp eliminase. *Nat. Commun.* **2017**, *8*, 14876.

22. Martí, S.; Tuñón, I.; Moliner, V.; Bertran, J., Are heme-dependent enzymes always using a redox mechanism? A theoretical study of the Kemp elimination catalyzed by a promiscuous aldoxime dehydratase. *ACS Catal.* **2020**, *10* (19), 11110-11119.

23. Bordeaux, M.; Singh, R.; Fasan, R., Intramolecular C(sp³)H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **2014**, *22* (20), 5697-704.
24. Brantley, R. E., Jr.; Smerdon, S. J.; Wilkinson, A. J.; Singleton, E. W.; Olson, J. S., The mechanism of autooxidation of myoglobin. *J. Biol. Chem.* **1993**, *268* (10), 6995-7010.
25. Liong, E. C.; Dou, Y.; Scott, E. E.; Olson, J. S.; Phillips, G. N., Waterproofing the Heme Pocket: Role of proximal amino acid side chains in preventing hemin loss from myoglobin*. *J. Biol. Chem.* **2001**, *276* (12), 9093-9100.
26. Abeysinghe, T.; Kohen, A., Role of long-range protein dynamics in different thymidylate synthase catalyzed reactions. *Int. J. Mol. Sci.* **2015**, *16* (4), 7304-19.
27. Wilson, M. T.; Torres, J., Stopped-flow spectroscopy. In *Stopped-flow spectroscopy*, Oxford University Press: 2000.
28. Bagshaw, C. R., A beginner's guide to flow kinetics. *The Biochemist* **2020**, *42* (2), 40-44.
29. Cox, R. P.; Hollaway, M. R., The reduction by dithionite of Fe(III) myoglobin derivatives with different ligands attached to the iron atom. A study by rapid-wavelength-scanning stopped-flow spectrophotometry. *Eur. J. Biochem.* **1977**, *74* (3), 575-87.
30. Tsukahara, K.; Okazawa, T.; Takahashi, H.; Yamamoto, Y., Kinetics of reduction of metmyoglobins by ascorbate - effect of the modification of the heme distal side, heme propionates, and 2,4-substituents of deuterohemin. *Inorg. Chem.* **1986**, *25* (26), 4756-4760.

31. Giulivi, C.; Cadenas, E., The reaction of ascorbic acid with different heme iron redox states of myoglobin. Antioxidant and prooxidant aspects. *FEBS Lett.* **1993**, 332 (3), 287-90.
32. Olson, J. S.; Mathews, A. J.; Rohlfs, R. J.; Springer, B. A.; Egeberg, K. D.; Sligar, S. G.; Tame, J.; Renaud, J. P.; Nagai, K., The role of the distal histidine in myoglobin and haemoglobin. *Nature* **1988**, 336 (6196), 265-6.
33. Boone, C. D.; Gill, S.; Habibzadegan, A.; McKenna, R., Carbonic Anhydrase: An efficient enzyme with possible global implications. *Int. J. Chem. Eng.* **2013**, 2013, 813931.
34. Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013**, 503 (7476), 418-+.
35. Perrin, B. S., Jr.; Miller, B. T.; Schalk, V.; Woodcock, H. L.; Brooks, B. R.; Ichiye, T., Web-based computational chemistry education with CHARMMing III: Reduction potentials of electron transfer proteins. *PLoS Comput. Biol.* **2014**, 10 (7), e1003739.
36. Olson, T. L.; Williams, J. C.; Allen, J. P., Influence of protein interactions on oxidation/reduction midpoint potentials of cofactors in natural and de novo metalloproteins. *Biochim. Biophys. Acta* **2013**, 1827 (8-9), 914-22.
37. Khersonsky, O.; Kiss, G.; Rothlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109 (26), 10358-63.

38. Moroz, O. V.; Moroz, Y. S.; Wu, Y.; Olsen, A. B.; Cheng, H.; Mack, K. L.; McLaughlin, J. M.; Raymond, E. A.; Zhezherya, K.; Roder, H.; Korendovych, I. V., A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew. Chem. Int. Ed.* **2013**, *52* (24), 6246-9.
39. Verma, R.; Mitchell-Koch, K., *In silico* studies of small molecule interactions with enzymes reveal aspects of catalytic function. *Catalysts* **2017**, *7* (7).
40. Osuna, S., The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Comp. Mol. Sci.* **2021**, *11* (3), e1502.
41. Joughin, B. A.; Green, D. F.; Tidor, B., Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.* **2005**, *14* (5), 1363-9.
42. Kihara, D., The effect of long-range interactions on the secondary structure formation of proteins. *Protein Sci.* **2005**, *14* (8), 1955-63.

Chapter 4. Developing a Probe for Noninvasive Detection of Hepatocellular Carcinoma by Positron Emission Tomography

This chapter was co-authored with Dr. Michelle Takas, who engineered HolLEE, and Dr. Z. Lengyel-Zhand, who performed initial scFv expression experiments

Abstract. Liver cancer is one of the most common cancers in the world. Diagnosis of hepatocellular carcinoma (HCC) at its early stages is difficult due to the small tumor size and lack of specific detection procedures. Positron emission tomography (PET) with ^{18}F -fluorodeoxyglucose (FDG) is a widely used method for the detection of liver cancer but due to challenges with its specificity and stability, more specific PET tracers are needed. In this work, we attempt to create a metalloprotein-based radiotracer for the detection and radiotherapy of HCC. For this, we have rationally designed a yttrium-binding protein, HolLEE, by introducing mutations into the C-terminal domain of calmodulin, a calcium-binding protein. We have shown that HolLEE has a preference for rare earth metals and no affinity for calcium, which makes this protein suitable for intravenous administration. To guide the tracer to a specific site, we attempted to fuse a glypican-3-specific single-chain variable antibody fragment (scFv) with HolLEE and establish the purification process for this protein.

4.1. Introduction

4.1.1. Detection of liver cancer and positron emission tomography

Liver cancer is considered one of the emerging malignant cancers in the world. Statistics on this disease vary depending on gender, age and country, but current data suggest that it ranks as the sixth most common cancer and the third leading cause of cancer-related deaths worldwide (Figure 4.1). Hepatocellular carcinoma (HCC) contributes to about 75% of liver cancers.^{1, 2} Risk factors for hepatocellular carcinoma include a wide range of genetic and environmental factors such as hepatitis B or C infection, autoimmune hepatitis, liver cirrhosis, alcohol abuse and metabolic syndrome.³

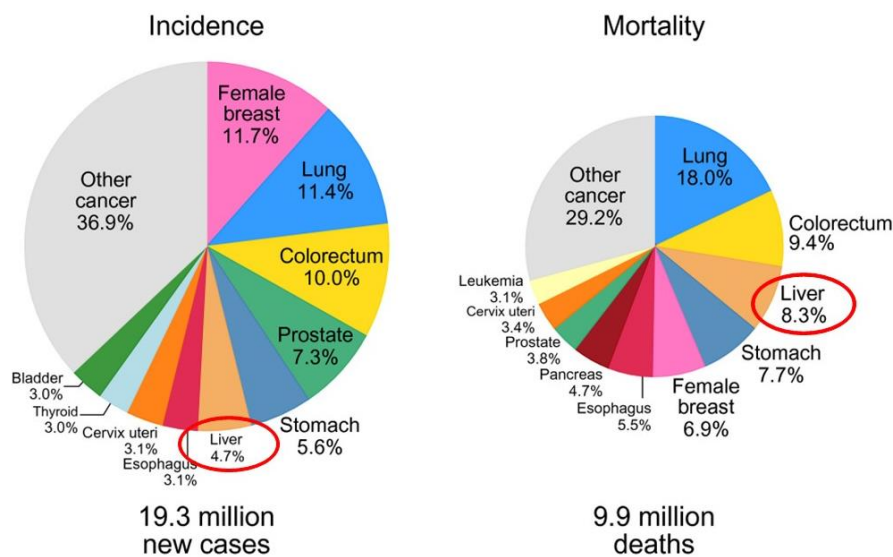


Figure 4.1. Distribution of cases and deaths for the top 10 most common cancers in 2020. Reproduced from ref 1 with the permission from American Cancer Society, copyright 2021.

Current methods for liver cancer detection include abdominal ultrasound with either a non-invasive confirmation by testing for the presence of liver-specific proteins in

blood or invasive biopsy procedure, contrast imaging using computed tomography (CT) or magnetic resonance imaging (MRI), and positron emission tomography (PET) by using a radiotracer, ^{18}F -fluorodeoxyglucose (FDG).⁴ Ultrasound on its own cannot provide definite results and further medical testing is needed to check for liver cancer susceptibility. Biopsy with histological studies is a reliable method to identify HCC, however, the result of this invasive procedure depends on the quality of collected tissue samples, and the lack of validated early-stage cancer markers makes it challenging to detect the disease at the onset.⁴

Among all contrast imaging techniques, ^{18}F -FDG PET had shown the highest sensitivity for detecting the disease due to high glucose uptake in cancer cells.⁵ Positron emission tomography (PET) is an *in vivo* imaging technique based on the detection of 511 keV photons emitted during the annihilation of positrons with electrons almost perpendicularly to the positron's trajectory. Radionuclides introduced in the body undergo positron emission decay which results in the emission of positrons. These positrons travel a small distance and then interact with the electrons in the tissues. During the collision, particles annihilate and produce two photons that are traveling in opposite directions.⁶ The ^{18}F -fluorodeoxyglucose (^{18}F -FDG) is a positron-emitting sugar radiolabeled with fluorine-18.⁷ Being an analog of glucose, FDG is easily incorporated into the cell through the natural mechanisms and accumulated in the cells with high metabolism, such as cancer cells (Figure 4.2).^{8,9} Even though FDG is successfully used for many types of cancer, the high metabolic activity of liver makes it challenging to differentiate small early stage neoplasms from the background, which leads to a high rate of false-positive results. Besides ^{18}F -FDG, ^{11}C -based tracers showed promising

results in the detection of liver cancer. The amino acids and cell metabolites labeled with ^{11}C were designed as probes for PET but as they also rely on the cell metabolism and have a short half-life (20 min vs 110 min for FDG), their use in cancer detection is limited.¹⁰⁻¹²

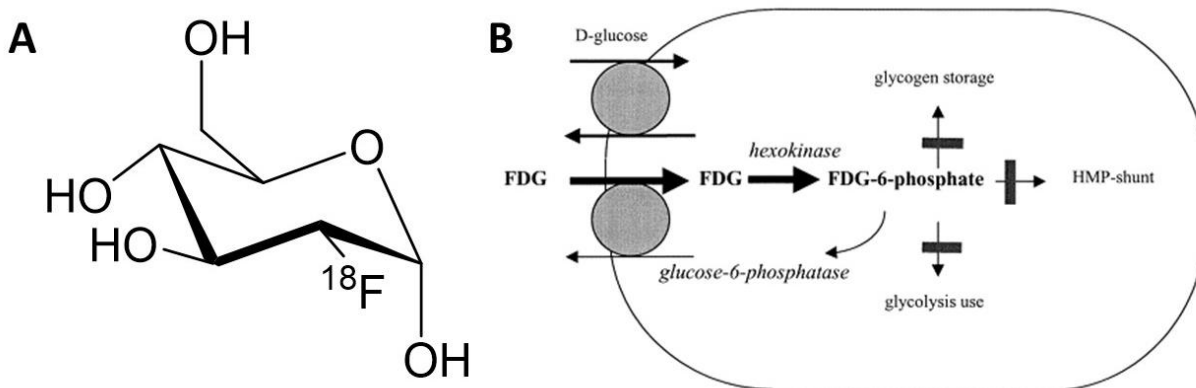


Figure 4.2. **A.** The structure of ^{18}F -fluorodeoxyglucose (^{18}F -FDG). **B.** The uptake of FDG by cells (reproduced from ref. 9 with permission from Elsevier Ireland Ltd., copyright 2003).

4.1.2. Metal-based PET tracers

The utilization of modified metabolites as PET tracers takes advantage of natural mechanisms which aid with the delivery of those molecules to the sites of interest. The analogs of metabolites usually contain four isotopes, ^{11}C , ^{13}N , ^{15}O and ^{18}F , which have been proven effective but also are characterized by the short half-life (carbon-11 – 20 min, nitrogen-13 – 10 min, oxygen-15 – 2 min and fluorine-18 – 110 min).¹³⁻¹⁵ This puts the time restrictions on the production of tracer, which requires the proximity of cyclotron to the PET facility, and also limits the duration of analysis. To overcome these restrictions, other radionuclides have been tested for the development of PET agents.

Metal-based PET tracers comprise another group of promising detecting agents. Such metals ^{82}Rb , ^{68}Ga , ^{64}Cu , ^{86}Y and ^{89}Zr are widely used to create positron-emitting probes with relatively long half-lives. Yttrium is considered a highly promising theranostic tool: while ^{86}Y emits positrons with a half-life of 14.7 hours, ^{90}Y is a high beta emitter with 64-hours half-life.¹⁶ Yttrium-90 is a known radionuclide for targeted irradiation of cancer cells. As this isotope does not emit positrons, it is usually hard to calculate the dose for proper radio treatment. Yttrium-86, being a positron-emitter, is often used as a surrogate of ^{90}Y for its quantification and visualization. The dual activity of this isotope pair can be productively used to develop tools for simultaneous therapy and diagnostics.

Yttrium-90 is used for radioembolization, the radioactive treatment protocol for hepatocellular carcinoma. This method includes the immobilization of radioactive isotope in the glass microbeads with further delivery to the liver through the invasive procedure.^{17, 18} Even though radioembolization is an effective treatment, it requires the accurate delivery of the sample to the tumor, which might be challenging at the early cancer stages.

4.1.3. Antibody-based delivery system for radiometals

For most diagnostic studies, the radioactive tracer is administered to a patient by intravenous injection; however, it may also be delivered by inhalation, oral ingestion, or direct injection into an organ. After intravenous injection, the commonly used PET tracer ^{18}F -FDG will distribute around the body and the cells with higher glucose uptake will be

visualized. While talking about metal-based agents, their distribution is not specific, thus tracers capable of delivering the metal to a specific site need to be developed.

One of the best-described and widely used methods for site-specific delivery is the use of antigen-specific antibodies. Antibodies are natural proteins, produced by the immune system in response to the antigen. As antibodies can be engineered to recognize certain antigens, they are often used for the detection of proteins, the sorting of cells and for the delivery of small cargo, including radioactive metals.^{19, 20} One of the successful examples of an antibody-based delivery system for metal is ⁹⁰Y-ibritumomab tiuxetan, an FDA approved drug for the treatment of B-cell non-Hodgkin lymphoma.²¹ This drug is based on the CD-20-specific monoclonal antibody conjugated to yttrium-90. The advantage of this method is that the metal is delivered in proximity to the cancer cell which decreases the side effects. Another example is ⁹⁰Y-tacatuzumab tetraxetan, a DOTA-conjugated monoclonal antibody specific to the alpha-fetoprotein, a tumor marker found in many cancers, including hepatocellular carcinoma.²² This agent contains a metal chelator, DOTA, widely used for the delivery of radioactive metals.

Direct metal conjugation to the protein and the use of metal chelator are two main strategies for metal attachment to the antibody. Direct conjugation requires the engineering of the high-affinity metal binding site, and the use of chelators is limited by the properties of the chelator (if metal requires a different number of ligands for coordination or a specific coordination environment, the binding may not be productive). One way to improve the metal binding is to use proteins with metal affinity. Proteins provide a great framework as they can be modified to adjust for the binding of a specific

metal. In our work, we attempted to create a metalloprotein that can act as a selective chelator for yttrium.

Precise targeting of liver tumors can be achieved by taking advantage of tissue-specific markers. Glypican-3 (GPC3) is a member of the glypican family of glycosyl-phosphatidylinositol-anchored cell-surface heparan sulfate proteoglycans. It is expressed at high levels in most HCCs which makes it a promising candidate for tumor diagnosis at early stages.^{23, 24} To take advantage of this marker, multiple GPC3-specific antibodies were developed for therapeutic or diagnostic purposes.^{25, 26}

Protein engineering has been successfully used to modify protein structure with the preservation of its function. The development of single-chain antibody fragments (scFv) allowed to simplify the production of specific antibodies without sacrificing the reactivity.²⁷ ScFv is a simplified version of the antibody: a fusion protein of the variable regions of heavy (V_H) and light chains (V_L) of immunoglobulins, connected with a short peptide linker (Figure 4.3).

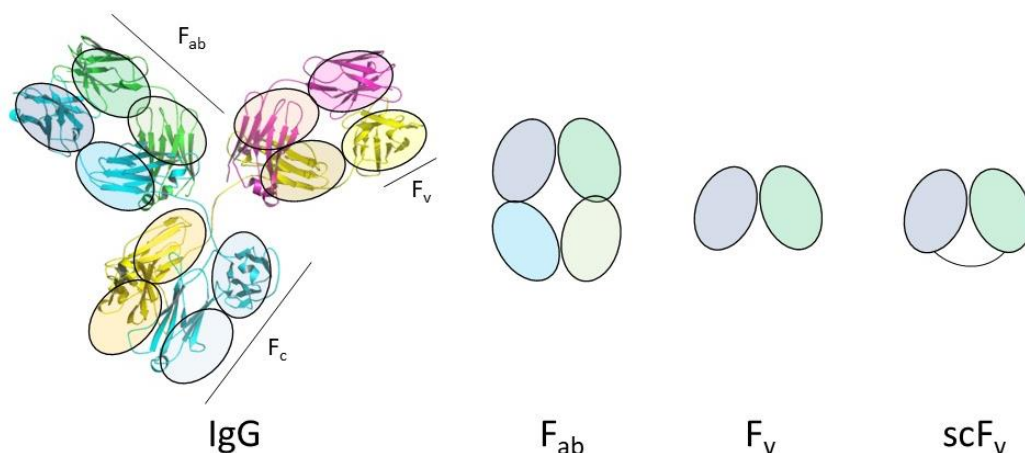


Figure 4.3. The schematic representation of different types of engineered antibodies. ScFv stands for single-chain variable domain. The antibody structure was visualized by Pymol (PDB 1IGY).

These proteins can be further engineered using phage or yeast display techniques.^{28, 29} Kaplan and co-workers isolated and validated glypican-3-specific scFv from a yeast display library.³⁰ This antibody fragment is small enough to be expressed in bacteria or yeast in high yields and at the same time displays reactivity of the full antibody. In our work, we hypothesize that a combination of an anti-glypican-3 antibody and an yttrium-binding metalloprotein can result in a more effective PET probe for the detection of the early stage HCC.

4.2. Results and Discussion

4.2.1. Design of yttrium-binding protein

Our idea for the novel PET contrast agents is based on the fusion of glypican-3-specific antibody fragment and yttrium-binding protein. For this purpose, we attempted to modify the metal affinity of a known metalloprotein. AlleyCat7 is a C-terminal domain of calcium-binding protein calmodulin, where mutations were introduced to create a novel catalyst for Kemp elimination reaction.³¹ This protein, being the derivative of calcium-binding protein, contains two EF-hands, a well-known metal-binding motif found in many proteins that coordinate to calcium. Calcium binding induces the change in protein structure leading to a more compact fold and the creation of a hydrophobic pocket necessary for the substrate binding.³² AlleyCat7 exhibits enzymatic activity upon proper metal-induced folding.

The rare earth metals are rarely used as cofactors for natural proteins with only a few reports of lanthanide-containing proteins. The methanotrophic bacteria, found in the crater of a volcano, use methanol as a source of energy, which is converted to formaldehyde by Ca^{2+} - and lanthanide-dependent methanol dehydrogenases (Ln-MDH).³³ Lanmodulin, the lanthanide-modulating protein, was co-purified with Ln-MDH and showed high affinity to lanthanides while utilizing the canonical calcium-binding motif.³⁴ Lanthanides have been successfully used to characterize calcium-binding sites in various proteins due to the similarity in size and their preference for carboxylates as coordination ligands.³⁵ Y(III) ion (medium-hard to hard Lewis acid) and a carboxylic acid oxygen atom (hard base) are expected to form a hard-hard acid-base pair, as predicted by HSAB theory. Mack et al. have previously reported the design of calmodulin-based

lanthanide-binding proteins by introducing the mutations in the EF-hands to modulate lanthanide affinity.³⁶ The designed proteins of CuseCat family showed improved affinity to the trivalent metals over calcium. As we attempted to design a metalloprotein-based radiotracer for intravenous delivery, our goal was not to only develop an efficient yttrium binder, but also to eliminate protein's affinity to calcium. Human blood contains low millimolar concentrations of calcium (a typical range of 2-2.5 mM³⁷), which can outcompete yttrium from the metal binding site. By removing calcium affinity, we expect to ensure the precise delivery of radionuclide to the site.

The characterization of protein-metal binding is often challenging due to the instability of the metal or metal-protein complex under experimental conditions. The complex interactions, that include structural reorganization of protein scaffold or the water shell around metal ions, can also interfere with the accurate determination of binding parameters. Moreover, many methods that are used for the determination of protein-metal binding are low-throughput which puts restrictions on the situations when large libraries of mutants need to be tested. Previous efforts in our lab were focused on the development of the simple assay that allows to detect protein-metal binding. As it was shown for CuseCat proteins, the coordination of metal results in the change in protein folding that leads to the formation of the active site. Metal binding induced CuseCat's catalytic activity which can be tested using simple absorbance assays. For our design efforts we decided to use colorimetric assay to probe protein-metal binding.

Calcium ions bind to the EF-hands in a pentagonal bipyramidal geometry with seven coordination ligands. The EF-hands of calmodulins consist of 12 amino acids arranged in a canonical helix-loop-helix structure (Figure 4.4).

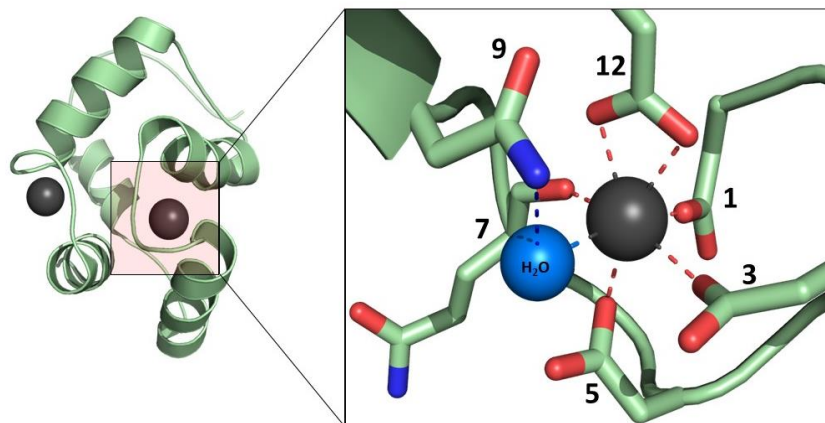


Figure 4.4. The structure of the EF-hand, a metal binding site present in AlleyCat7. The EF-hand consists of 12 canonical amino acids. The protein structure was prepared using Pymol and PDB 2KZ2.

In all described binding sites, a coordinating water (through residue 9), a backbone amide C=O group (residue 7), a bidentate glutamate carboxylate group (residue 12), and two monodentate aspartate carboxylate groups (position 1 and 5) provide six coordination ligands for metal binding. The last position is occupied by a third monodentate aspartate carboxylate group or by an asparagine side-chain C=O (position 3).

We have started our design effort from the examination of the EF-hand motif of AlleyCat7. Aspartate at position 1 is important for proper protein folding.³⁸ Position 12 requires bidentate ligand and the options are limited to Asp and Glu. The amino acid at position 7 coordinates to the metal through the carboxyl oxygen which is present in all amino acids. As the mutations at positions 3, 5 and 9 modulated the metal affinity of CuseCat, we have tested them in the AlleyCat7 scaffold. The introduction of D95E and D131E mutations at position 3 in AlleyCat7 resulted in the elimination of affinity to calcium while preserving affinity to yttrium. As the Kemp elimination activity of this

enzyme is linked to proper metal-induced folding, we used the colorimetric assay to assess the binding of metals to the newly designed protein HolIEE (AlleyCat7 D95E/D131E) (Figure 4.5).

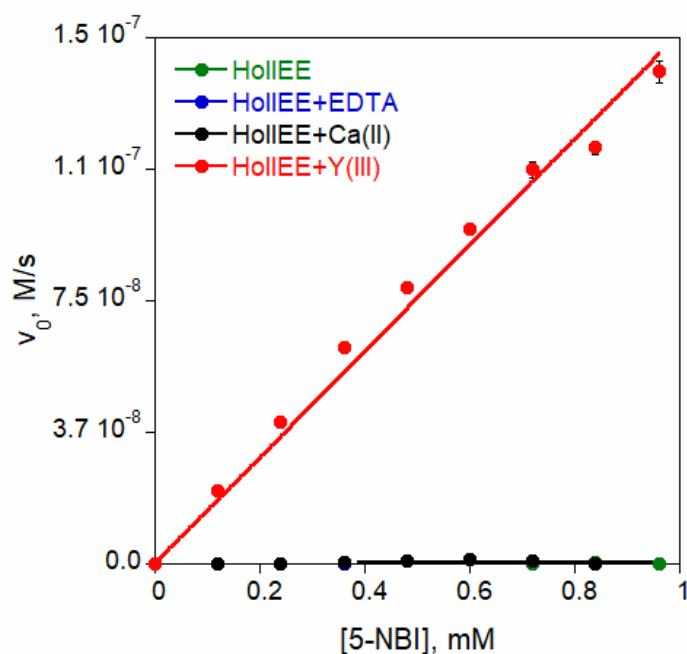


Figure 4.5. Kemp elimination activity of HolIEE. The protein's activity was tested at 1 μM in the buffer containing 20 mM MOPS (pH 7.0), 100 mM NaCl and 80 μM EDTA, Ca (II) or Y (III)). The catalytic efficiency of HolIEE in the presence of yttrium was measured as $161 \pm 14 \text{ M}^{-1}\text{s}^{-1}$.

Based on the protein's catalytic efficiency, HolIEE adopted the efficient fold in the complex with yttrium while the addition of calcium did not improve protein's activity. These results suggested that the designed protein met out initial requirements (preserved binding to yttrium but showed no calcium affinity), thus we proceeded to further investigation of metal binding to HolIEE.

4.2.2. Determination of the binding affinity of HolIEE to rare earth metals

AlleyCat7 is an allosterically regulated protein that adopts catalytically active conformation upon calcium binding. The introduction of two mutations in the EF-hands of AlleyCat7 (one mutation in each coordination sphere) resulted in the loss of catalytic activity in the presence of calcium. As metal-induced folding is essential for protein's Kemp elimination activity, we suggested that HolIEE lost the affinity to calcium but preserved the ability to coordinate yttrium. To confirm our hypothesis, we analyzed the secondary structure of the protein using circular dichroism (CD) spectroscopy. CD spectroscopy allows to evaluate the degree of helicity of analyzed protein and can be used to predict the secondary structure motifs and the degree of folding. HolIEE, being a derivative of AlleyCat7, is expected to undergo major restructuring upon metal binding. In its apo form, HolIEE is a protein with a prominent alpha-helical character (Figure 4.6.).

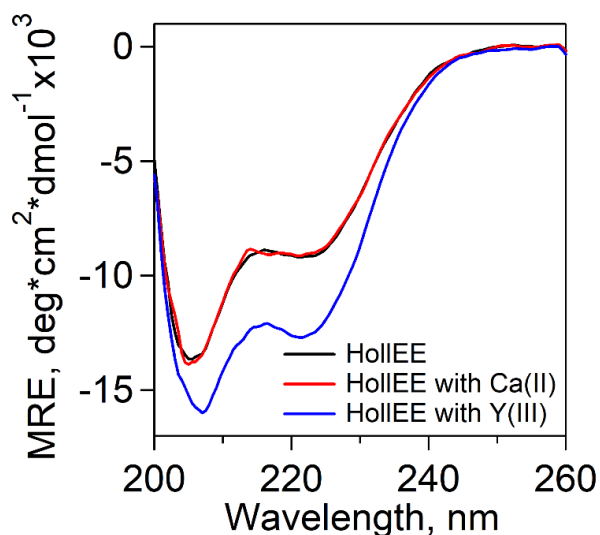


Figure 4.6. Circular dichroism spectra for the analysis of HolIEE upon metal binding. All samples were analyzed using 20 μM protein and 80 μM metal (Ca(II) or Y(III)) in 30 mM MOPS (pH 6.5), 100 mM KCl at 25°C.

The incubation of HolIEE with 4 eq of calcium (II) did not affect protein's helicity suggesting that the protein does not bind Ca (II). On contrary, protein folding improved after the incubation with yttrium (III) indicating the interaction between protein and metal. To further confirm protein's inability to bind calcium, we analyzed protein-metal interactions using isothermal titration calorimetry.

Isothermal titration calorimetry (ITC) is a routine technique for the determination of thermodynamic parameters of binding, such as the number of binding sites, change in enthalpy (ΔH) and dissociation constant (K_d). The C-terminal domain of calmodulin, which was used as a scaffold for the development of AlleyCat and HolIEE, binds two metal ions in two non-identical binding sites with possible cooperativity.³⁹ The titration of both proteins with calcium (II) chloride solution showed that HolIEE has no affinity to calcium while AlleyCat7 binds metal ion (Figure 4.7.).

The dissociation constants obtained for AlleyCat7 ($K_{d,1} = 4.7 \mu\text{M}$, $K_{d,2} = 49.6 \mu\text{M}$) were derived using the "sequential binding sites" model, which is applicable to proteins that have more than two ligand binding sites that might be identical or non-identical, independent or cooperative. This model is more general and has fewer constraints compared to the "two binding sites" approach which is used to analyze two sets (groups) of non-identical and independent (non-cooperative) ligand binding sites. Even though further confirmation of AlleyCat7's calcium affinity using a different method is needed, the obtained results align with the predicted mode of action of metal-binding sites in calmodulin derivatives. The titration of HolIEE with calcium (II) did not produce any heat indicating that HolIEE does not interact with this metal confirming the results of the calorimetric assay used to test protein's catalytic activity.

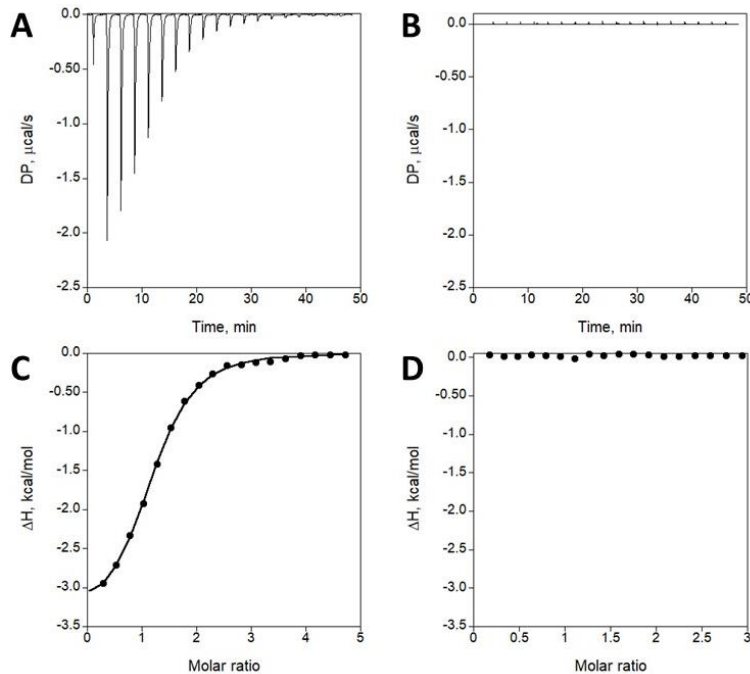


Figure 4.7. ITC traces for the titration of 198 μM AlleCat7 (A) and 195 μM HolIEE (B) with 3 mM calcium (II) and the binding isotherms (C and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 100 mM MES, 100 mM KCl, pH 6.5 at 25 $^{\circ}\text{C}$.

The analysis of protein affinity to yttrium (III) ions brings some challenges. Yttrium is known to lower the pH of a solution and precipitate in the buffers at high pH. To mitigate the impact of acidification of the solution by yttrium, we performed the experiments using high buffer concentrations (100 mM MES) and to ensure metal stability in the solution we lowered pH of the buffer to 6.0 (compared to 6.5 used for calcium). Even though these conditions are different, we hoped to obtain any preliminary results which could provide any insight into the protein-metal interactions. The ITC analysis of metal binding to AlleyCat and HolIEE produced complex isotherms that could not be reliably analyzed. (Figure 4.8).

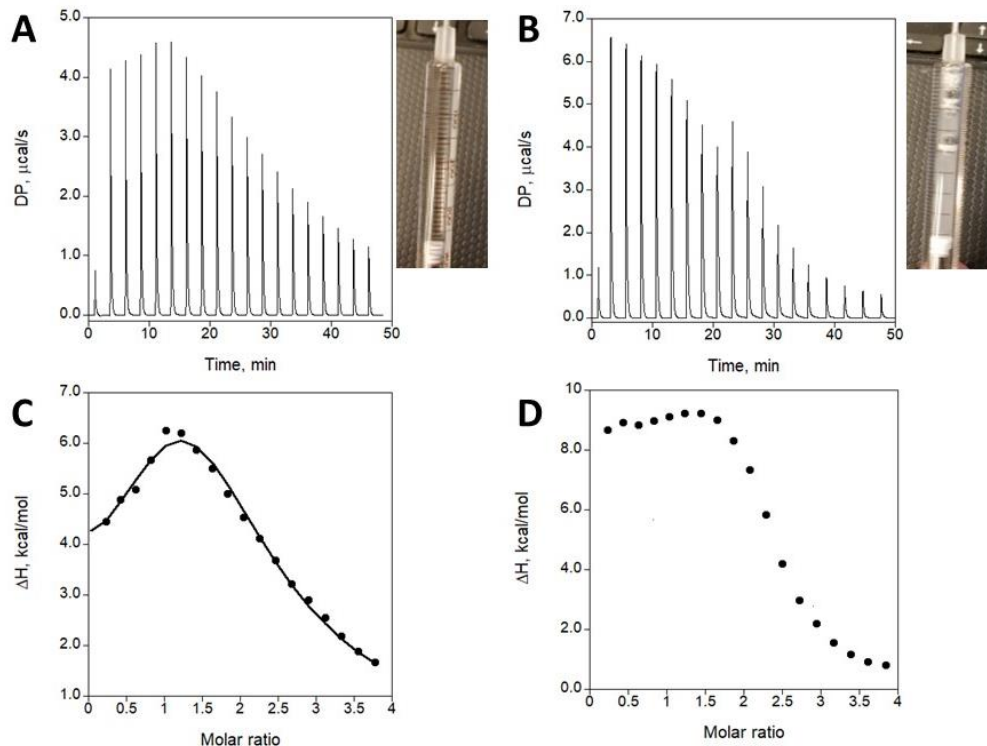


Figure 4.8. ITC traces for the titration of 198 μM AlleCat7 (A) and 195 μM HolIEE (B) with 4 mM yttrium (III) chloride and the binding isotherms (C and D, respectively) used for the derivation of thermodynamic parameters. The titrations were performed in 100 mM MES, 100 mM KCl, pH 6.5 at 25 °C. The photos attached to panels A and B show the precipitation of HolIEE upon addition of yttrium (B).

Moreover, we have observed the precipitation of HolIEE upon the addition of yttrium. Yttrium (III) is a highly positively charged metal ion with an ionic radius similar to that of calcium (II). As HolIEE has a high negative net charge (Figure 4.9.), the addition of trivalent metal induces protein precipitation. Lanthanides have been previously utilized for the precipitation of phosphoproteins⁴⁰, which, similar to HolIEE, have a high negative charge and abundance of oxygen ligands. This effect is pronounced in the experiments that require high protein concentrations. Thus, we decided to employ other

techniques that could allow us to mitigate this effect by using lower protein concentrations.

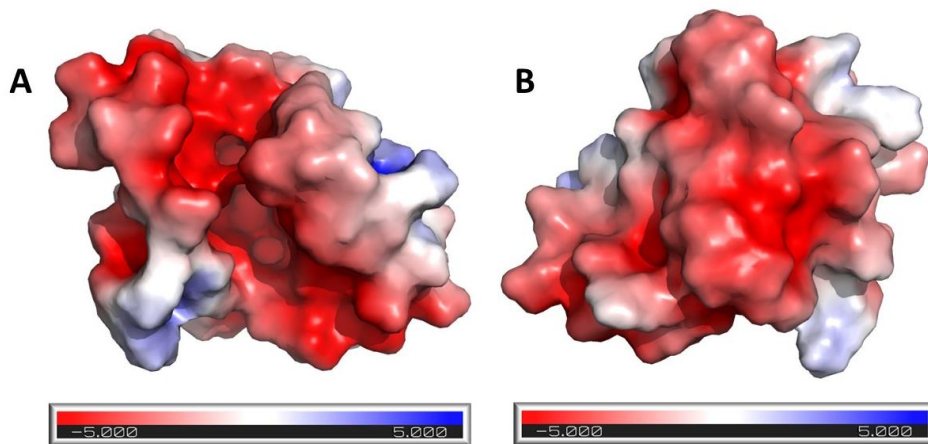
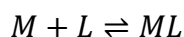


Figure 4.9. The model of HolIEE created using the structure of AlleyCat as a scaffold (PDB 2KZ2). A. The view from the side of substrate binding site. B. The view from the side opposite to the binding site. The calculations were performed using APBS electrostatics plugin in Pymol for solvent excluded surface.

As we have observed before, the presence of metal ions in the solution influences the degree of protein helicity suggesting metal binding. As the circular dichroism experiment requires much lower protein concentrations, we suggested that this method can be used to determine the binding parameters of yttrium (III)-HolIEE binding. Unfortunately, the direct titration of protein with metal lead to metal precipitation which interfered with the analysis (the experiment was performed previously in our lab). We decided to employ metal-chelator buffers that are used to control metal concentration in the solution.⁴¹

The equilibrium equation suggests that the ratio between the metal-chelator complex and a free chelator defines the free concentration of the metal:



$$K_d = \frac{[M]_{free}[L]_{free}}{[ML]}$$

$$[M]_{free} = K_d \frac{[ML]}{[L]_{free}}$$

We started with establishing the protocol for analysis of HolIEE in yttrium-EDDS (ethylenediamine-N, N'-disuccinic acid) buffer, which was previously used for the determination of lanthanide affinity to proteins.³⁴ The chelation of metal ions by EDDS results in a change in its absorption of circular light which allows for accurate determination of stoichiometry and concentrations necessary to prepare 1:1 EDDS:metal complex (Figure 4.10.). The EDDS:Y(III) complex is mixed with free EDDS at different ratios to produce precisely controlled concentrations of free yttrium in the solution.

As the designed system is very sensitive, the preparation of protein included EDTA treatment to remove all bound metals. The dilution/concentration method and dialysis were used to ensure the removal of EDTA from the protein solution. The stability constant of EDDS for yttrium (III) is 13.55 which indicates very tight binding. As the procedure of titration includes the mixing of two solutions at different ratios, the highest free metal concentration which can be achieved using this approach is 0.1 μM . This method works well for proteins with the dissociation constant in the low nanomolar or picomolar range. The results of CD analysis showed no significant change in the protein's helicity which suggests that HolIEE binds yttrium (III) in the micromolar range (Figure 4.11).

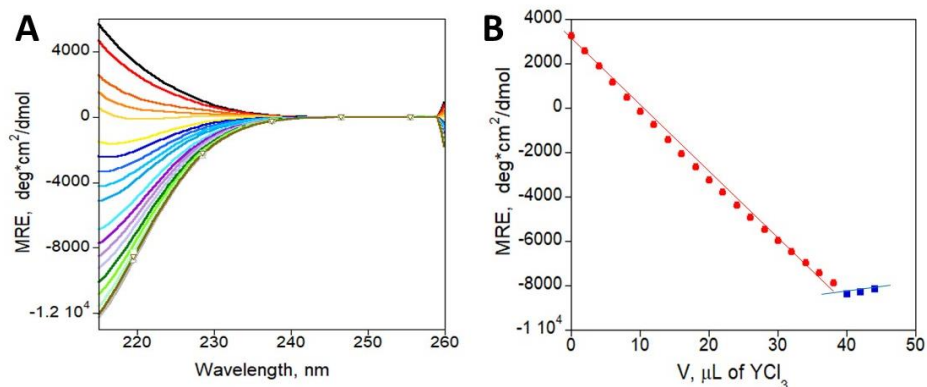


Figure 4.10. A. CD spectra for the titration of 10 mM EDDS with the aliquots of 100 mM yttrium (III) chloride. **B.** The determination of metal binding stoichiometry and the concentrations necessary to prepare one-to-one EDDS:Y(III) complex.

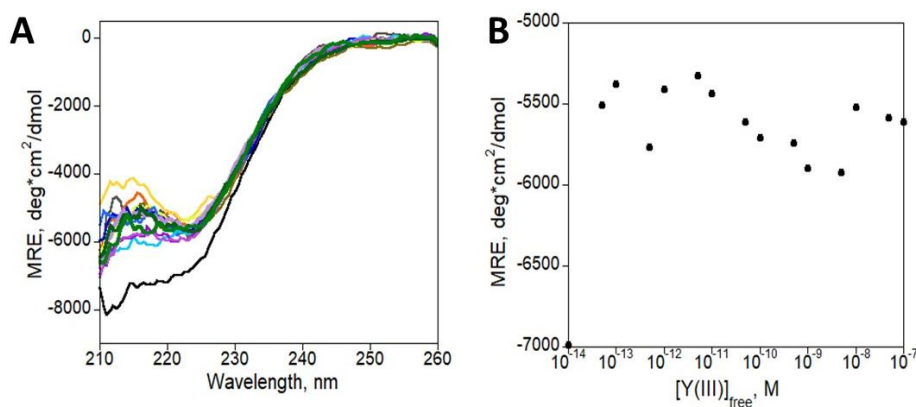


Figure 4.11. A. CD spectra of the titration of 10 μM HolIEE with yttrium (III) in yttrium-EDDS. **B.** The plot presenting MRE values at 220 nm plotted versus the concentration of free yttrium (III) did not show any dependence.

We have shown that the introduction of two mutations into AlleyCat7 successfully created a yttrium-binding protein, HolIEE, with no affinity to calcium. The analysis of metal binding to HolIEE showed that the dissociation constant for HolIEE-Y(III) binding is, presumably, in the micromolar range, although further experiments are required for the reliable determination of protein-metal binding parameters.

4.2.3. Development of 2E10-HolIIEE radiotracer

To continue our efforts in the development of the novel metalloprotein-based PET tracer, we proceeded to engineer the fusion of HolIIEE and the glypican-3-specific antibody fragment (scFv). To establish a protocol for the expression of scFv fused with yttrium-binding protein HolIIEE, we started with the bacterial expression system in *Escherichia coli* as it is a well-studied, simple and highly productive. Test expression of pMCSG49/2E10 plasmid in bacterial cells *E. coli* Rosetta pLysS showed a good level of protein expression, but the protein was localized in the inclusion bodies. To improve the solubility of the protein, we decided to fuse 2E10 with eGFP, an enhanced green fluorescent protein. This can help with the solubilization of scFv and also provides an opportunity to test 2E10 binding to the antigen on the cell surface by fluorescence microscopy. For further activity and binding studies, scFv 2E10 was also fused to HolIIEE to make our target probe. As certain movements and interactions are required for the activity of each domain in the fusion protein, a flexible linker with the sequence of GSAGSAAGSGGF⁴² was inserted between proteins.

The protein expression was tested in two different bacterial strains. Origami 2 *E. coli* strain contains genes that encode for modified thioredoxin reductase and glutathione reductase, proteins that help to maintain reducing conditions in the cell. These modifications were shown to greatly enhance disulfide bond formation in the *E. coli* cytoplasm, which makes Origami 2 a suitable host for expression of disulfide-rich antibody fragments.⁴³ As scFv was derived from yeast display library, Rosetta pLysS *E. coli* strain could enhance the expression of eukaryotic genes that contain codons rarely used by bacteria.⁴⁴ Test expression in Origami 2 cells showed neither high protein

expression level nor target protein in the soluble fraction (Figure 4.12). The 2E10 antibody sequence was derived from yeast and contains rare codons. As Origami 2 bacterial strain does not facilitate the expression of rare codons, this could lead to the complication during protein expression which results in the very low yield.

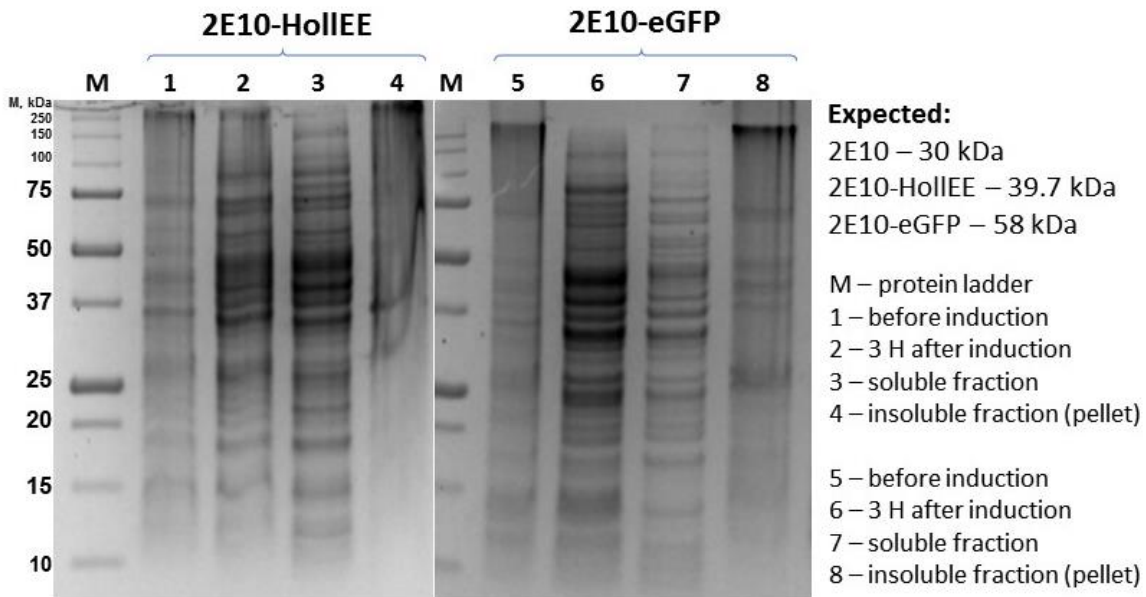


Figure 4.12. SDS-PAGE analysis of the fractions from test protein expression of 2E10-eGFP and 2E10-HolIIEE in Origami 2 *E. coli* cells after IPTG induction in LB medium.

Test expression of target proteins in Rosetta pLysS cells with IPTG induction was more successful and showed high level of protein production (Figure 4.13). However, both proteins were found in the insoluble fractions. The localization of the overexpressed protein the inclusion bodies is usually caused by the improper folding of scFv in the bacterial cytoplasm that leads to protein aggregation. Antibody fragments have been shown to localize in the inclusion bodies due to the misfolding in the bacterial

cells, but it is interesting that scFv fusion with soluble proteins does not improve protein solubility.

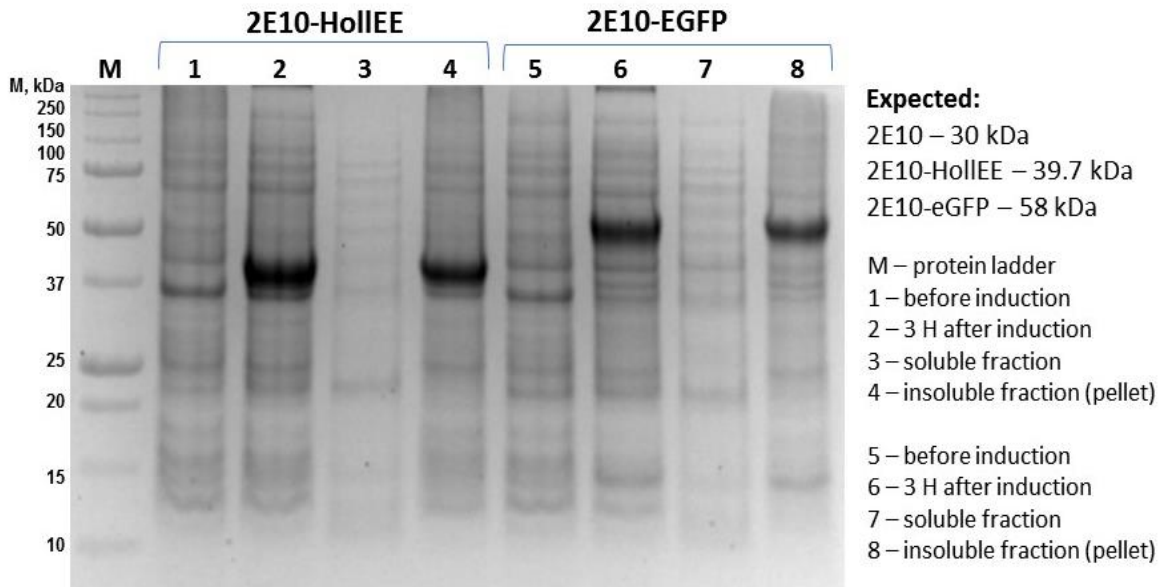


Figure 4.13. SDS-PAGE analysis of the fractions from test protein expression of 2E10-eGFP and 2E10-HolIEE in Rosetta pLysS *E. coli* cells after IPTG induction in LB medium.

IPTG induction is a reliable method for expression of soluble non-toxic proteins in bacterial cells.⁴⁵ When IPTG is added to the medium, protein expression occurs at much higher rate. Certain proteins require longer time for proper folding and a high rate of protein expression alters this process, leading to protein aggregation into inclusion bodies. To mitigate these effects, we tested an autoinduction medium where instead of synthetic inducers like IPTG, sugars such as lactose and glucose are used. This method allows proteins to express under natural conditions as the inducers of protein overexpression are produced during cell metabolism. The decrease in the protein synthesis rate could facilitate proper folding.

Test expression of 2E10 and 2E10-HolIEE in Rosetta pLysS cells grown in autoinduction medium showed that the proteins were still aggregated in the inclusion bodies (Figure 4.14).

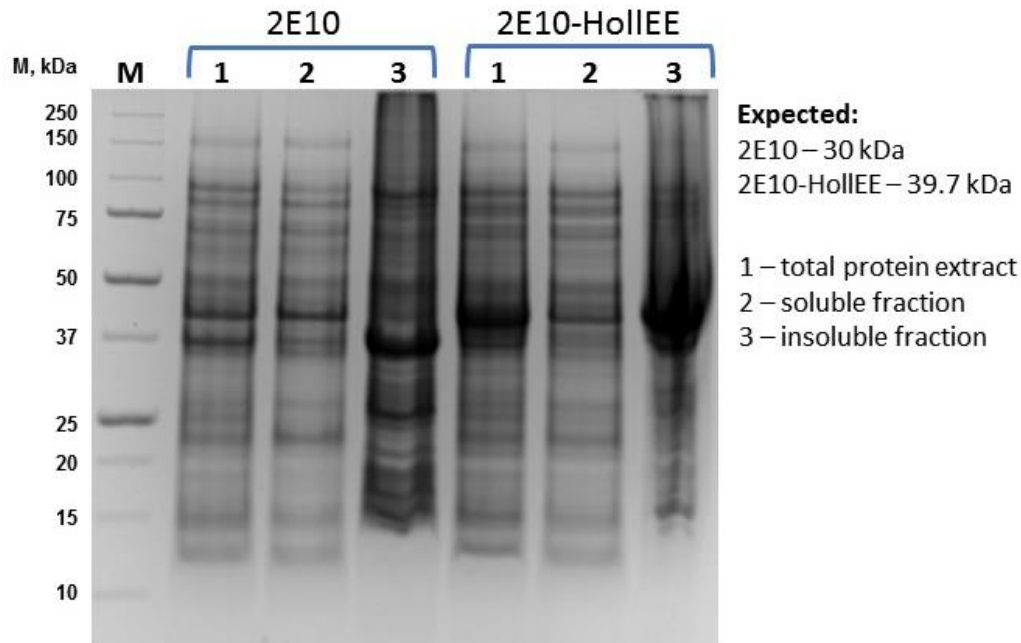


Figure 4.14. SDS-PAGE analysis of the fractions from test protein expression of 2E10 and 2E10-HolIEE in Rosetta pLysS *E. coli* cells in autoinduction medium.

Considering high expression level of the 2E10-HolIEE, we purified the protein fusion from the inclusion bodies formed in Rosetta pLysS cells. This involved protein unfolding using chaotropic agents as guanidine hydrochloride or urea, followed by renaturation by decreasing the concentration of those agents. After 2E10-HolIEE was refolded on Ni-NTA agarose column, it was eluted by increasing the concentration of imidazole in the buffer (Figure 4.15). This method is often used for the production of the antibody fragment in bacterial cells. The final yield was 0.2 mg of protein from 1 L of bacterial culture.

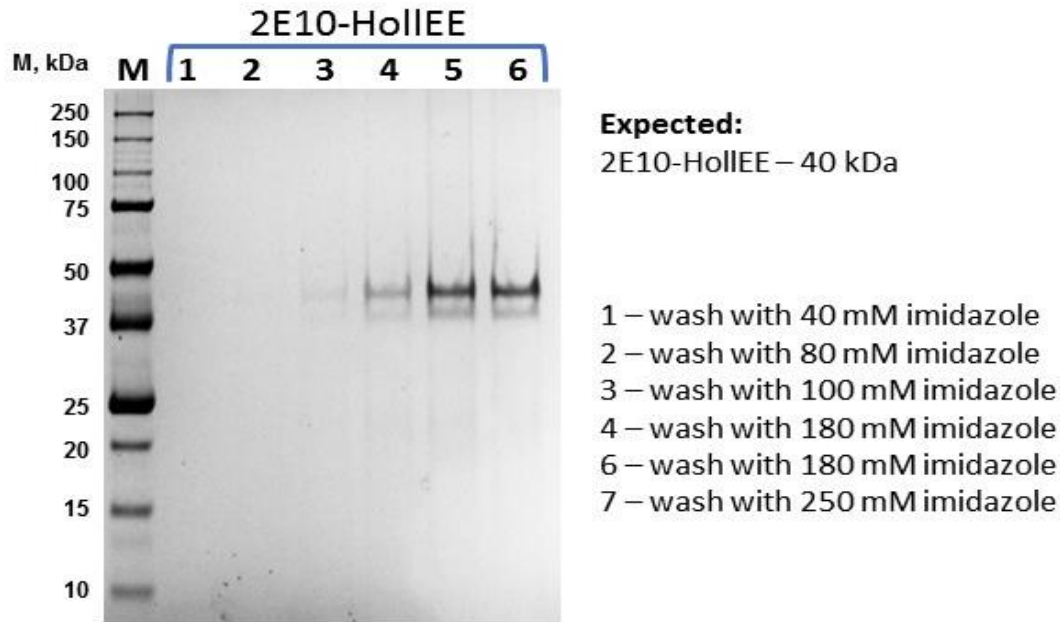


Figure 4.15. SDS-PAGE analysis of the elution fractions which contain 2E10-HolIIEE after purification from inclusion bodies.

Even though we were able to successfully obtain scFv-HolIIEE fusion, protein purification from inclusion bodies did not provide sufficient amount of the properly folded protein for further studies. Thus, we decided to test other protein expression systems. As 2E10 was derived from yeast display library in *Saccharomyces cerevisiae*, a yeast protein expression system was suggested to solve the problem of protein solubility. Previously, our lab has established protein expression system in *Pichia pastoris*. The similarity between *S. cerevisiae*, used in yeast display, and *P. pastoris* allows us to use those hosts interchangeably which eliminates the problems with codon usage between species. The redox potential of the yeast cytoplasm is suitable for the formation of the disulfide bonds in the scFv.⁴⁶ *P. pastoris* is easier to manipulate, and often gives higher protein expression levels compared to other yeast-based systems. Also, yeast

expression system gives an option of secreted protein expression. This method was proposed to improve protein yield, simplify purification process and reduce protein toxicity to the host cell. After yeast cell transformation with pPIC9 vector containing the target sequence, yeast clones were selected based on methylotrophy, the ability to use methanol as a sole carbon source. This is the characteristic of transformed yeast as the vector introduction into yeast genome completes the sequence of alcohol oxidase that allows growth on methanol. Positive clones were tested for secreted protein expression during a 5-day period. Unfortunately, the analysis of medium showed no traces of target proteins (Figure 4.16).

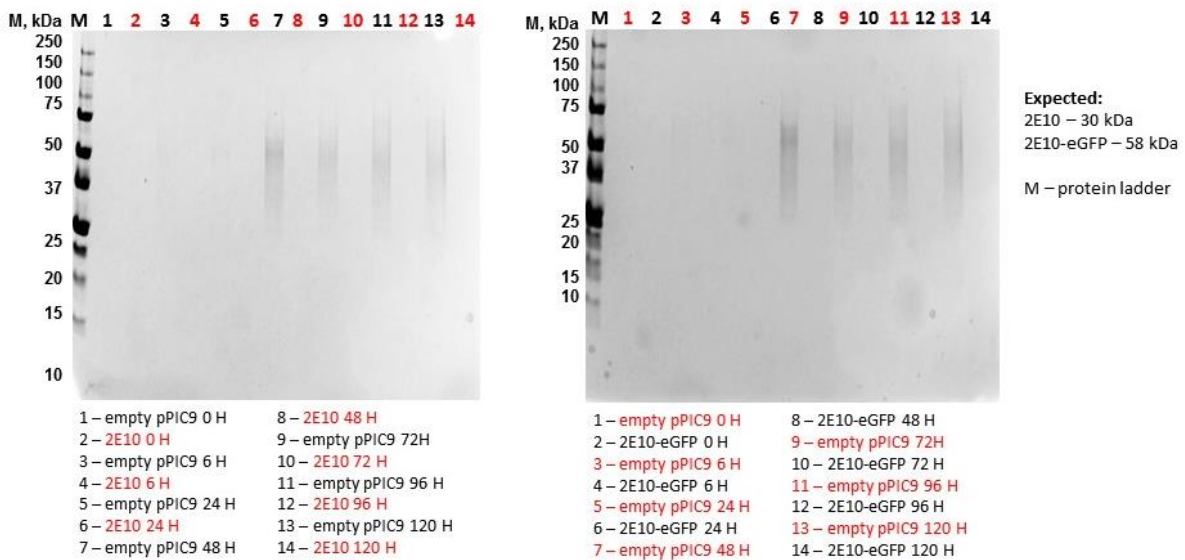


Figure 4.16. SDS-PAGE analysis of the medium samples collected during the test secreted protein expression of 2E10 and 2E10-eGFP in *Pichia pastoris*.

The system we utilized should produce secreted proteins, since 2E10 and 2E10-eGFP were fused to α -factor, the peptide sequence that labels protein for secretion.⁴⁷ In case of α -factor malfunction but successful expression, the protein should accumulate

inside the yeast cell. However, yeast cell extract analysis did not show the presence of target proteins (Figure 4.17) which suggests that scFv and its fusions cannot be produced by *Pichia pastoris*.

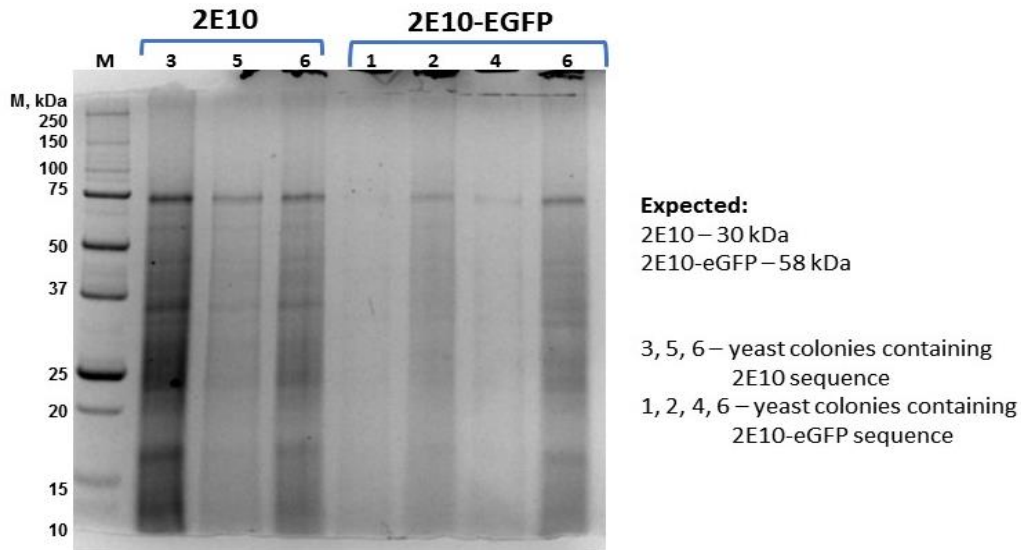


Figure 4.17. SDS-PAGE analysis of the yeast cell extracts collected during the test secreted protein expression of scFv 2E10 and 2E10-eGFP in *Pichia pastoris*.

Meanwhile, we decided to test a baculovirus protein expression system for scFv production. Baculoviruses belong to a diverse group of large double stranded DNA viruses that specifically infect insect cells. After infection, the cells start producing viral proteins and, in case of recombinant expression system, the protein of interest.⁴⁸ As insects are more advanced eukaryotes than yeast, their cells are more suitable for producing such complex proteins as antibodies and scFv. After 96-hours infection, cells were collected by centrifugation and the cell pellet showed a distinguishable red tint of the reporter red fluorescent protein (RFP) under visible light indicating a successful infection by baculovirus. When exposed to UV light, a sample which contained the gene

of 2E10-eGFP, glowed green (Figure 4.18) suggesting the expression and proper folding of the fluorescent tag. However, the analysis of the total protein extracts did not show prominent bands indicating the low expression level of our target proteins (Figure 4.19).

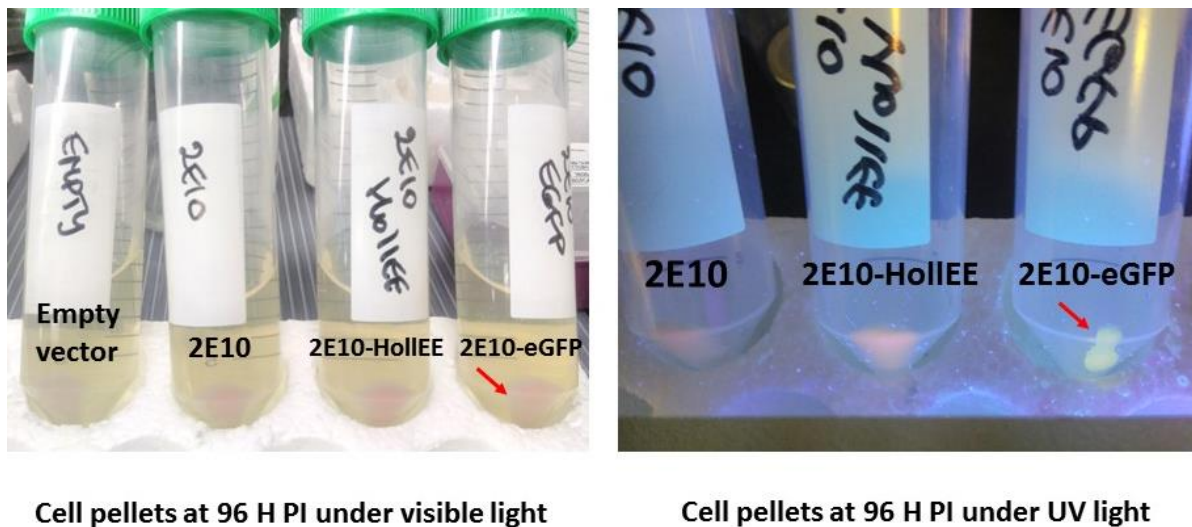


Figure 4.18. Photos of the insect cells infected with the viruses containing 2E10, 2E10-HolIIEE and 2E10-eGFP coding regions. Sample containing 2E10-eGFP emitted green light indicating functional eGFP.

As the color of culture indicated the presence of protein, we performed test purification of the protein. Unfortunately, no protein was identified in the elution fraction or flow-through, and the insoluble protein fraction was still emitting a green light under UV light. These results suggest that 2E10 scFv forms insoluble aggregates in insect cells, similar to the result we obtained in bacterial expression system.

The production of soluble antibodies or their forms is a challenge due to the complex structure of these proteins. The protein expression protocols need to be developed for every new antibody or their fusions. In our work we focused on the

expression of one glypican-3-specific scFv that showed the most promising results.³⁰ Even though we did not succeed with the production of the 2E10-based protein fusion, our future efforts will be focused on developing the expression and purification protocols for other candidates.

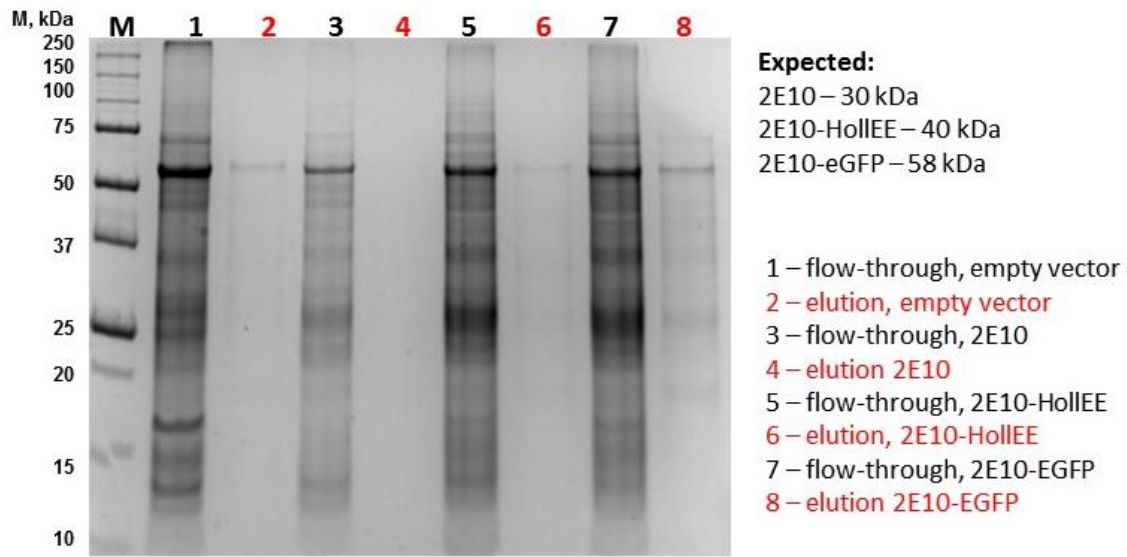


Figure 4.19. SDS-PAGE analysis of flow-through and elution fractions collected during test purification of 2E10, 2E10-HolLEE and 2E10-eGFP from infected insect cells.

4.3. Conclusions and Outlook

We have successfully shown that the introduction of two mutations into the EF-hand motif of calcium-binding protein AlleyCat7 could create a selective yttrium-binding protein HolIEE. The introduction of mutations at position 3 of the EF-hand completely eliminated affinity to calcium as shown by the colorimetric kinetic activity assay, circular dichroism spectroscopy and isothermal titration calorimetry. These results highlight the importance of position 3 for calcium binding. HolIEE preserved the affinity to rare earth metals, specifically yttrium (III). Unfortunately, we were not able to characterize HolIEE-yttrium (III) binding due to the protein precipitation issues but the estimation suggests that the dissociation constant is in the micromolar range.

We have used three different expression systems to produce metalloprotein-antibody fusion. The fusion was expressed in the bacterial and insect cell but in both cases it was insoluble which raises challenges with successful protein purification. We were able to obtain a soluble protein after the protein purification from the inclusion bodies and our future efforts will be focused on the development of the protocol for the large-scale purification of protein fusion and its further testing in the cell-based assay.

4.4. Experimental

HolIEE and AlleyCat7 protein expression and purification. The recombinant plasmids containing SUMO-HolIEE sequence (pET-SUMO/HolIEE) or SUMO-AlleyCat7 (pET-SUMO/AlleyCat7) were transformed into *E. coli* BL21 [DE3] cells. A single colony was inoculated into 20 mL of LB medium supplemented with 50 µg/mL kanamycin (Kan, Bio Basic Inc.) and grown over night at 37°C with shaking. Next morning, 20 mL of starter culture were diluted in 1 L of medium with Kan and incubated (with shaking) at 30°C until OD₆₀₀ of 0.5–0.7. The protein synthesis was induced using 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside, Bio Basic Inc.) and the culture was incubated for additional 5 h at 30°C (with shaking). Later, the cells were collected by centrifugation (4,000xg, 4°C, 20 min), flash-frozen in liquid nitrogen and stored at -80°C until needed.

Target protein was purified as a His-fusion protein using Ni-NTA agarose (Genetech). The cell pellet was resuspended in 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole and sonicated for 10 min (20s on/20s off cycles) at 20 W on ice. To aid with DNA removal, 25 U/mL of benzonase (Millipore) were added and protein solution was incubated at 37°C for 30 min. After the incubation, the soluble protein fraction was separated by centrifugation at 20,000xg for 30 min (at 4°C). The supernatant (contains soluble proteins) was decanted and applied on Ni-NTA column, pre-equilibrated with resuspension buffer. The column was later washed with resuspension buffer until no protein was eluted. For the elution of protein of interest, the column was washed with 25 mM Tris-HCl (pH 8.0), 300 mM NaCl and 250 mM imidazole, and the eluate was collected as 1 mL fractions. The fractions were analyzed using UV-vis spectrometry, the fractions which contained target protein were combined

and concentrated using spin concentrators (10K MWCO, Corning). After concentration, the protein was exchanged in SUMO cleavage buffer (50 mM Tris-HCl (pH 8.0), 75 mM NaCl) using 10DG buffer exchange columns (BioRad).

The His-SUMO tag cleavage reaction was set up by adding 1 mM DTT and 0.5 mM EDTA to the protein in SUMO cleavage buffer; the protein solution was sterilized using 0.2 μm syringe filter (Santa Cruz) and incubated overnight at 30°C. The efficiency of SUMO cleavage reaction was verified using SDS-PAGE (10% gel, Tris-MES buffer). After the incubation, 1 mM EDTA was added to the protein solution and incubated for 5 min at room temperature. Then, the cleavage reaction was buffer exchanged into 100 mM MES (pH 6.0), 100 mM KCl and collected fractions were applied on Ni-NTA column. As both SUMO tag and SUMO protease had His-tag, they attached to the column and pure HolIEE or AlleyCat7 were collected as flow-through. Collected fractions were subjected to dilution-concentration cycles using 5K MWCO spin concentrator (Corning) with further dialysis into 100 mM MES (pH 6.0), 100 mM KCl (3 L of buffer, at least 3 H, 4°C). The protein concentration was assessed by UV-vis spectroscopy (extinction coefficient at 280 nm is 2,980 $\text{M}^{-1}\text{cm}^{-1}$).

Preparation of yttrium-EDDS buffer. The procedure was described previously by Cotruvo et al.³⁴ Ethylenediamine N,N'-disuccinic acid (EDDS) was obtained as a ~35% (~0.977 M) trisodium salt solution in water (Sigma). For the purpose of metal-protein K_d determination, 1:1 metal:EDDS stock solutions were prepared by taking advantage of the change in the CD signal of the free ligand upon metal binding. A solution of 30 mM MOPS (pH 7.2), 100 mM KCl, and 10 mM EDDS was prepared in a 50-mL tube. Yttrium (II) chloride stock was prepared as 1 M stock by dissolving salt in

water and later 100 mM working solution was prepared by diluting 1 M stock in 30 mM MOPS (pH 7.2), 100 mM KCl. CD spectra were acquired at 260–215 nm, 0.5 nm data pitch, continuous mode, 50 nm/min speed, 1 nm band width, 4 sec average time at 25°C. Total of three scans were acquired and averaged for each experiment. The solution of 10 mM EDDS in buffer (300 μ L) was placed in a 1-mm pathlength CD cuvette and titrated with 2- μ L aliquots of 100 mM YCl_3 , until the CD signal at 220 nm, adjusted for volume change during the titration, was no longer altered by metal addition (1:1 metal:ligand complex). The volume of the metal stock solution required to reach the equivalence point was calculated from the intersection of the linear regression line fitted to points in the first half of the titration and the average ellipticity from the first 3 points after signal was constant. The 1:1 yttrium-EDDS buffer consisted of 30 mM MOPS (pH 7.2), 100 mM NaCl, 10 mM EDDS and 11.45 mM YCl_3 . The solution was prepared fresh before each experiment.

Determination of K_d of protein-metal binding by CD spectroscopy. A concentrated solution of HolIEE was diluted to 10 μ M separately in yttrium-EDDS buffer (final composition: 30 mM MOPS (pH 6.5), 100 mM KCl, 10 mM EDDS, 11.72 mM YCl_3 , 10 μ M HolIEE) and in EDDS buffer (final composition: 30 mM MOPS (pH 6.5), 100 mM KCl, 10 mM EDDS, 10 μ M HolIEE). These two protein solutions were mixed at different ratios to give various free metal concentrations. The same ratios of buffers without protein were mixed to yield the blank samples for CD experiments. The samples were incubated at room temperature for ~1 h to allow for equilibration prior to collection of CD spectra at 25 °C. The blank spectra at each ratio were subtracted from the protein containing spectra, and the $[\theta]_{222nm}$ was plotted vs. free metal concentration.

Determination of protein-metal binding parameters using isothermal titration calorimetry. The thermodynamic parameters of the metal binding to HolIEE were monitored by isothermal titration calorimetry using a MicroCal PEAQ-ITC (Malvern). A concentrated protein stock solution (3 mL) was dialyzed against 1 L of 100 mM MES (pH 6.0), 100 mM KCl. A diluted solution of apoprotein was made, and the exact concentration of each dilution was determined by the absorbance at 280 nm using an extinction coefficient of $2,980 \text{ M}^{-1}\text{cm}^{-1}$. The metal salts were prepared as 100 mM stocks in water and then diluted to 250 μM in the buffer collected after protein dialysis (100 mM MES (pH 6.0), 100 mM KCl). Protein solution in the calorimeter cell was titrated with metals by automatic 2- μL injections at 25°C with 750 rpm stir speed and the intervals of 3-5 min between injections. Baseline correction and integration of the peaks corresponding to each injection were performed using the Microcal PEAQ-ITC Analysis software provided by the manufacturer. The data were fitted to the two sets of sites model to determine such parameters as the stoichiometry (n), enthalpy of complex formation (ΔH), and equilibrium binding constant (K). The control experiment consisting of injections of the protein solution into the buffer was performed to evaluate the heat of dilution.

Cloning of 2E10-HolIEE and 2E10-eGFP into pMCSG49 vector. Cloning was carried out using In-Fusion HD Cloning Kit (Takara) according to the manufacturer's standard protocols. The linker GGATCAGCAGGTTCCGCTGCTGGTTCTGGCGAATTC, coding for GSAGSAAGSGEF, was introduced using primers with extensions.

Test expression of 2E10, 2E10-HoIIIEE and 2E10-eGFP in bacterial cells.

Purified plasmid (1 μ L) was mixed gently with 50 μ L of Rosetta pLysS cells, kindly provided by Dr. Carlos Castaneda, or Origami 2 cells (Novagen). Cells were incubated on ice for 15 mins followed by heat shock treatment at 42°C for 45 seconds in the water-bath. Then, cells were cooled down on ice for 2 min before the addition of 950 μ L of Luria Bertani (1x LB). Transformed cells were incubated at 37°C for 1 hour with continuous shaking at 250 rpm. 100 μ L of the transformed culture were plated on an ampicillin-containing plate (100 μ g/mL) and colonies were allowed to grow overnight at 37°C. A single colony containing plasmid was transferred into 20 mL 1x LB containing 100 μ g/mL of ampicillin or 20 mL of autoinduction medium and the culture was incubated at 37°C under continuous shaking at 250 rpm until OD₆₀₀ ~ 0.6. Then, 1 mL of culture was saved as an uninduced sample and protein expression was induced with 0.5 mM IPTG with further incubation at 37°C for 3 hours. Cells were harvested by centrifugation (4,000xg, 10 min, 4 °C), resuspended in 1 ml of Lysis buffer (25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole, 0.5 mM PMSF) and sonicated for 3 min (20 sec on/20 sec off) at 15 W. Total protein extract sample was saved and the rest was centrifuged at 14,000xg for 20 min at RT. The supernatant was treated as a soluble protein sample and pellet was assigned to insoluble protein fraction. All samples were analyzed using SDS-PAGE (10% gel).

Purification of 2E10-HoIIIEE from bacterial inclusion bodies. The cell paste was resuspended in the lysis buffer (10 mL per 1 g of pellet) containing 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 20 mM imidazole, 0.4 mM PMSF and 1 μ L/1g of pellet of benzonase (Millipore). Solution was sonicated for 10 min (20 sec on/20 sec off) at 19 W

followed by centrifugation at 20,000xg for 20 min at 4°C. Then, pellet was subjected to two rounds of washing with 30 ml of ice-cold Pellet wash 2UT buffer (25 mM Tris-HCl (pH 8.0), 500mM NaCl, 2 M Urea and 2 % Triton X-100) with further 6-min sonication and centrifugation at 20,000xg for 10 min at 4°C. Inclusion bodies were washed with 30 ml of ice-cold Pellet wash buffer (25 mM Tris-HCl (pH 8.0), 500mM NaCl and 20 mM imidazole), sonicated for 6 min and pelleted at 20,000xg for 10 min at 4°C. Pure inclusion bodies were resuspended in Pellet lysis 6Gu buffer (50 ml of buffer per 1 g of inclusion bodies) containing 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 5 mM imidazole and 6 M guanidine hydrochloride by stirring for 60 min at RT until the pellet is completely dissolved. To remove insoluble particles, the solution was centrifugated for 15 min at 20,000xg at 4°C and passed through the 0.22 µm or 0.45 µm syringe filter.

Filtered solution was applied on the Ni-NTA agarose column (Takara, 3 mL) equilibrated with Pellet lysis 6Gu buffer, and incubated for 30 min at RT. After incubation, column was washed with 10 ml of Pellet lysis 6Gu buffer and then with 10 ml of Refolding 6U buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 20 mM imidazole and 6 M urea). Protein refolding on column was performed by manual step-ramp gradient wash from 6M urea to 0M urea by mixing Refolding 6U buffer and Refolding buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl and 20 mM imidazole) to decrease the concentration of urea in the buffer. After refolding, protein was eluted with Elution buffer containing 20 mM Tris-HCl (pH 8.0), 500 mM NaCl and 250 mM imidazole. Concentration of the protein was deduced by checking its absorbance at 280 nm using a UV-Vis spectrometer and the calculated extinction coefficient of 64,540 M⁻¹cm⁻¹.

Subcloning of 2E10 and 2E10-eGFP into pPIC9 vector for protein

expression in yeast cells. All target sequences were subcloned from pMCSG49-based constructs into pPIC9 using ligation cloning. For this, all inserts were amplified using primers containing EcoRI and NotI restriction sites. Both inserts and vector were digested with EcoRI and NotI (both – New England Biolabs) and then ligated with T4 DNA ligase (New England Biolabs) according to the manufacture's protocols.

Expression of 2E10 and 2E10-eGFP in *Pichia pastoris* GS115 cells. pPIC9-based plasmids containing 2E10 and 2E10-eGFP (2 µg of each) were linearized by digestion with Sall restriction enzyme (New England Biolabs, following manufacturer's protocol) and purified using DNA purification kit (Bio Basic Inc). The overnight culture (2 mL) of GS115 cells in YTD medium (1% yeast extract, 2% tryptone, 2% dextrose) was transferred into 100 mL of YTD and grown at 28-30°C with vigorous shaking (~175 rpm) until OD₆₀₀ ~1.5. Then, culture was pelleted by centrifugation at 3,000xg for 2 min at RT, resuspended in 8 ml of ice cold LiAc-DTT buffer (10 mM Tris-HCl (pH 7.5), 100 mM lithium acetate, 10 mM DTT, 0.6 M sorbitol) and kept at RT for 30-40 min. After incubation, cells were spun down at 3,000xg for 2 min at RT, resuspended in 1.5 mL of ice-cold 1 M sorbitol, transferred into sterile 1.5-mL tube, pelleted again (3,000xg, 2 min, RT) and additionally washed with 1.5 mL of 1 M sorbitol twice. After wash, cell pellet was resuspended in 600 µL of 1 M ice-cold sorbitol. To quantify the cells, the 1,000x dilution was prepared and OD₆₀₀ was measured (should be below 0.2). Then, 1.5 µg of linearized DNA was added to 200 µL of yeast suspension. Mix was incubated on ice for 10 min, electroporated at 1.5 kV for 6.2 msec, 1 mL of ice-cold 1 M sorbitol was added immediately and cells were kept on ice. Transformation mix was plated on the RDB

plates (2% dextrose, 1.34% yeast nitrogen base with ammonium sulfate, 0.00004% biotin, 0.005% L-glutamic acid, 0.005% L-methionine, 0.005% L-lysine, 0.005% L-leucine, and 0.005% L-isoleucine, 18.6% sorbitol, 2% agar) which were incubated at 28-30°C for 3 days.

After incubation, colonies were picked up and grown in 2 mL of YTD medium overnight. For these small cultures, 5 µL of each culture were plated on MM (1.34% yeast nitrogen base with ammonium sulfate, 0.00004 % biotin, 0.5% methanol, 1.5% agar) and MD (1.34% yeast nitrogen base with ammonium sulfate, 0.00004 % biotin, 2% dextrose, 1.5% agar) plates with further incubation at 28-30°C for 3 days to check for methylotrophy. The methanol (100 µL) was added to MM plates every 24 hours to provide the source of carbon. After 3 days, a few colonies were picked from MM plates for PCR screening with 5'AOX1 and 3'AOX1 primers for the presence of the target inserts. The colonies which contained the insert of interest were inoculated in 2 mL of YTD medium and grown over night 28-30°C with good shaking. Next morning, the cells were collected by centrifugation (3,000xg for 2 min at RT), resuspended in YT medium with 40% glycerol, frozen in liquid nitrogen and stored at -80°C for further use.

To test the protein expression, the remaining culture (~2 mL) was resuspended in 20 mL of BMMY medium (1% yeast extract, 2% tryptone, 100 mM potassium phosphate (pH 6.0), 1.34% yeast nitrogen base with ammonium sulfate, 0.00004% biotin, 0.5% methanol) and grown at 28-30°C for 5 days to check for protein expression. The methanol (100 µL) was added to cultures every 24 hours to provide the source of carbon. The 0.5-mL culture aliquots were taken at 0, 6, 24, 48, 72, 96, 120 hours and analyzed using SDS-PAGE (10% gel).

Subcloning of 2E10, 2E10-HolIEE and 2E10-eGFP into pFastBac Dual vector for baculovirus protein expression system. All target sequences were subcloned from pMCSG49-based constructs into pFastBac Dual vector (Invitrogen) under polyhedrin promoter. 2E10 and 2E10-eGFP coding regions were amplified using primers containing EcoRI and NotI restriction sites and 2E10-HolIEE was amplified using primers for In-Fusion HD Cloning Kit (Takara). 2E10 and 2E10-eGFP inserts and vector were digested with EcoRI and NotI (both – New England Biolabs) and then ligated with T4 DNA ligase (New England Biolabs) according to the manufacturer's protocols.

Bacmid preparation. To prepare competent cells for bacmid production, a single colony of *E. coli* DH10BAC cells (Invitrogen) was inoculated into 5 mL of LB medium containing 10 µg/mL tetracycline (Tet) and 50 µg/mL kanamycin (Kan), and incubated overnight at 37°C with shaking (250 rpm). Then, 1 mL of the night culture was transferred into 100 mL of LB medium containing Tet and Kan, and incubated at 37°C with shaking (250rpm) to OD₅₉₀ of 0.375. After the culture reached the desired optical density, it was transferred into pre-chilled tubes and left on ice for 10 min. After centrifugation at 4,000xg at 4°C for 15 min, the cells were gently resuspended in 10 mL of ice-cold CaCl₂ solution (10 mM PIPES (pH 7.0), 60 mM CaCl₂, 15% glycerol) and pelleted at 4,000xg for 15 min at 4°C. Then, cells were resuspended in 10 mL of ice-cold CaCl₂ solution again and kept on ice for 30 min. After cells were spun down at 4,000xg for 15 min at 4°C, they were resuspended in 1 mL of ice-cold CaCl₂ solution and 100 µL of competent cells were used for heat-shock transformation with 1-2 µg of DNA. For cell recovery, 900 µL of SOC medium were added and cells were incubated at

37°C during 4-6 H with vigorous shaking (250 rpm). A 200- μ L aliquot of the transformed culture was plated on the blue-white selection plate containing 1 % tryptone, 1 % NaCl, 0.5 % yeast extract, 1.5 % agar, 50 μ g/mL kanamycin (Kan), 7 μ g/mL gentamycin (Gen), 12.5 μ g/mL tetracycline (Tet), 40 μ g/mL X-gal, 210 mM IPTG, and colonies were allowed to grow for 2 days at 37°C. A few white colonies (blue colonies did not have an insert as β -galactosidase gene is working) were selected and the presence of the target inserts was verified using PCR screening with polyhedrin forward primer and reverse primers specific to the inserts. The colonies containing target inserts were chosen for further work. The colonies were grown in 5 mL of LB medium supplemented with Kan, Tet and Gen at 37°C over night, collected by centrifugation and saved as flash-frozen pellets (stored at -80°C).

The bacmids for further insect cells transfection were purified from frozen pellets obtained from white colonies using BioBasic Plasmid Purification Kit. The pellets were resuspended in 200 μ L of Solution I and lysed after the addition of 400 μ L of Solution II (lysis buffer). The tubes were gently inverted for proper mixing and kept on ice for 5 min. Then, 700 μ L of Solution III (neutralization buffer) were added, the tubes were gently inverted and the formation of white cloudy pellet was observed. After the samples were centrifuged for 10 min at 14,000xg at RT, the supernatant was carefully transferred in 850 μ L of 100% isopropanol (in 1.5-mL tube). The contents were gently mixed by inversion and kept on ice for 10 min. At this point DNA was collected by centrifugation at 14,000xg for 15 min and the supernatant was carefully removed. The pelleted DNA then washed twice with pre-chilled at -20°C 70% ethanol (the ethanol was added to DNA, the tube was inverted a few times and the DNA was collected by centrifugation at 14,000xg

for 15 min). After the final wash, the supernatant was carefully removed and DNA was dried at 42°C for 10 min (or until it does not smell with ethanol). The bacmids were rehydrated by adding 50 µL of elution buffer and incubating DNA at 4°C overnight.

Baculovirus production and amplification Transfection of SF9 adherent cell culture (maintained in SF-900 III SFM medium (Gibco) at 28°C with gentle shaking) was performed using Trans-IT-Insect Transfection Reagent (Mirus) according to the manufacturer's protocol. Briefly, 2.7×10^6 cells were plated on 60-mm cell-culture treated dish in SF-900 III SFM medium and allowed to attach for 1 H at 28°C. Then, 7.5 µg of bacmid DNA were resuspended in 500 µL of medium, and 15 µL of transfection reagent were added. The solution was mixed by pipetting and allowed to incubate at RT for 30 min. After incubation, the mixture was added to cell culture drop-wise to the different areas of the plate and gently rocked for even distribution of complexes. Cells were incubated at 28°C until the signs of transfection were prominent. After 70-H infection, the primary virus stock was collected (the cells were scraped from the plate, the medium was collected and clarified by centrifugation and filtration). For virus amplification, 20×10^6 cells in 20 ml of medium (in 125-mL glass flasks) were infected with 300 µL of primary virus stock. Cell cultures were incubated at 28°C with continuous shaking (150 rpm). Each 24 H the aliquot of cells was taken to check the infection progress and cell viability. The amplified virus stock was collected after 96 H of infection. Cells were spun down at 3,000xg for 5 min at 4°C, supernatant was collected and filtered through 0.22 µm sterile syringe filter. Virus stocks are stored protected from light at 4°C.

Expression of 2E10 and 2E10-eGFP in Sf9 insect cells. For test protein expression 20×10^6 cells in 20 mL of growth medium were infected with 200 µL of

amplified virus stock and incubated at 28°C with continuous shaking. Cell samples were collected each 24 H until 70 H time-point. All samples were analyzed using SDS-PAGE (10% gel).

4.5. References

1. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71* (3), 209-249.
2. Xu, J., Trends in liver cancer mortality among adults aged 25 and over in the United States, 2000–2016. *NCHS Data Brief.* **2018**, 314,1-8.
3. Llovet, J. M.; Ducreux, M.; Lencioni, R.; Di Bisceglie, A. M.; Galle, P. R.; Dufour, J. F.; Greten, T. F.; Raymond, E.; Roskams, T.; De Baere, T.; Ducreux, M.; Mazzaferro, V.; Bernardi, M.; Bruix, J.; Colombo, M.; Zhu, A., EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J. Hepatol.* **2012**, *56*, 908-943.
4. Llovet, J. M.; Zucman-Rossi, J.; Pikarsky, E.; Sangro, B.; Schwartz, M.; Sherman, M.; Gores, G., Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* **2016**, *2*, 16018.
5. Hofman, M. S.; Hicks, R. J., How we read oncologic FDG PET/CT. *Cancer Imaging* **2016**, *16*, 1-14.
6. Shukla, A. K.; Kumar, U., Positron emission tomography: An overview. *J. Med. Phys.* **2006**, *31* (1), 13-21.

7. Almuhaideb, A.; Papathanasiou, N.; Bomanji, J., 18F-FDG PET/CT imaging in oncology. *Ann. Saudi Med.* **2011**, 31 (1), 3-13.
8. Kawada, K.; Iwamoto, M.; Sakai, Y., Mechanisms underlying (18)F-fluorodeoxyglucose accumulation in colorectal cancer. *World J. Radiol.* **2016**, 8 (11), 880-886.
9. Vansteenkiste, J. F., PET scan in the staging of non-small cell lung cancer. *Lung Cancer* **2003**, 42, 27-37.
10. Ho, C. L.; Yu, S. C.; Yeung, D. W., 11C-acetate PET imaging in hepatocellular carcinoma and other liver masses. *J. Nucl. Med.* **2003**, 44 (2), 213-21.
11. D'Souza M, M.; Sharma, R.; Jaimini, A.; Saw, S. K.; Singh, D.; Mondal, A., Combined (18)F-FDG and (11)C-Methionine PET/CT scans in a case of metastatic hepatocellular carcinoma. *Indian J. Nucl. Med.* **2014**, 29 (3), 171-4.
12. Sun, A.; Liu, X.; Tang, G., Carbon-11 and Fluorine-18 labeled amino acid tracers for positron emission tomography imaging of tumors. *Front. Chem.* **2017**, 5, 124.
13. Moreau, A.; Febvey, O.; Mognetti, T.; Frappaz, D.; Kryza, D., Contribution of different positron emission tomography tracers in glioma management: focus on glioblastoma. *Front. Oncol.* **2019**, 9, 1134.
14. Honka, M. J.; Rebelos, E.; Malaspina, S.; Nuutila, P., Hepatic positron emission tomography: Applications in metabolism, haemodynamics and cancer. *Metabolites* **2022**, 12 (4).
15. Nesterov, S. V.; Sciagra, R.; Orozco, L. E. J.; Prior, J. O.; Settimo, L.; Han, C.; Deshayes, E.; deKemp, R. A.; Ryzhkova, D. V.; Gwet, K. L.; Knuuti, J. M., One-tissue compartment model for myocardial perfusion quantification with N-13 ammonia PET

provides matching results: A cross-comparison between Carimas, FlowQuant, and PMOD. *J. Nucl. Cardiol.* **2021**, online ahead of print.

16. Wright, C. L.; Zhang, J.; Tweedle, M. F.; Knopp, M. V.; Hall, N. C., Theranostic Imaging of Yttrium-90. *Biomed. Res. Int.* **2015**, 2015, 481279.

17. Saini, A.; Wallace, A.; Alzubaidi, S.; Knuttinen, M. G.; Naidu, S.; Sheth, R.; Albadawi, H.; Oklu, R., History and evolution of yttrium-90 radioembolization for hepatocellular carcinoma. *J. Clin. Med.* **2019**, 8 (1), 55.

18. Kulik, L. M.; Atassi, B.; van Holsbeeck, L.; Souman, T.; Lewandowski, R. J.; Mulcahy, M. F.; Hunter, R. D.; Nemcek, A. A., Jr.; Abecassis, M. M.; Haines, K. G., 3rd; Salem, R., Yttrium-90 microspheres (TheraSphere) treatment of unresectable hepatocellular carcinoma: downstaging to resection, RFA and bridge to transplantation. *J. Surg. Oncol.* **2006**, 94 (7), 572-86.

19. Khaw, B., Antibodies as delivery systems for diagnostic functions. *Adv. Drug Deliv. Rev.* **1999**, 37 (1-3), 63-80.

20. Stoeckius, M.; Zheng, S.; Houck-Loomis, B.; Hao, S.; Yeung, B. Z.; Mauck, W. M.; Smibert, P.; Satija, R., Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology* **2018**, 19 (1), 224.

21. Mondello, P.; Cuzzocrea, S.; Navarra, M.; Mian, M., 90 Y-ibritumomab tiuxetan: a nearly forgotten opportunity. *Oncotarget.* **2016**, 7 (7), 7597-609.

22. Mohammadinejad, A.; Kazemi Oskuee, R.; Eivazzadeh-Keihan, R.; Rezayi, M.; Baradaran, B.; Maleki, A.; Hashemzaei, M.; Mokhtarzadeh, A.; de la Guardia, M.,

Development of biosensors for detection of alpha-fetoprotein: As a major biomarker for hepatocellular carcinoma. *Trends Analyt. Chem.* **2020**, 130, 115961.

23. Filmus, J.; Capurro, M., Glypican-3: A marker and a therapeutic target in hepatocellular carcinoma. *FEBS J.* **2013**, 280, 2471-2476.

24. Capurro, M.; Wanless, I. R.; Sherman, M.; Deboer, G.; Shi, W.; Miyoshi, E.; Filmus, J., Glypican-3: A novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* **2003**, 125, 89-97.

25. Li, Z.; Zeng, Y.; Zhang, D.; Wu, M.; Wu, L.; Huang, A.; Yang, H.; Liu, X.; Liu, J.; Schlessinger, D., Glypican-3 antibody functionalized Prussian blue nanoparticles for targeted MR imaging and photothermal therapy of hepatocellular carcinoma. *J. Mater. Chem. B* **2014**, 2, 3686-3696.

26. Zhang, Y. F.; Ho, M., Humanization of high-affinity antibodies targeting glypican-3 in hepatocellular carcinoma. *Sci. Rep.* **2016**, 6, 1-11.

27. Ahmad, Z. A.; Yeap, S. K.; Ali, A. M.; Ho, W. Y.; Alitheen, N. B.; Hamid, M., scFv antibody: principles and clinical application. *Clin. Dev. Immunol.* **2012**, 2012, 980250.

28. Gai, S. A.; Wittrup, K. D., Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **2007**, 17 (4), 467-73.

29. Eisenhardt, S. U.; Schwarz, M.; Bassler, N.; Peter, K., Subtractive single-chain antibody (scFv) phage-display: tailoring phage-display for high specificity against function-specific conformations of cell membrane molecules. *Nat. Protoc.* **2007**, 2 (12), 3063-73.

30. Li, Y.; Siegel, D. L.; Scholler, N.; Kaplan, D. E., Validation of glypican-3-specific scFv isolated from paired display/secretory yeast display library. *BMC Biotechnol.* **2012**, 12, 23.
31. Moroz, O. V.; Moroz, Y. S.; Wu, Y. B.; Olsen, A. B.; Cheng, H.; Mack, K. L.; McLaughlin, J. M.; Raymond, E. A.; Zhezherya, K.; Roder, H.; Korendovych, I. V., A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew. Chem. Int. Ed.* **2013**, 52 (24), 6246-6249.
32. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108 (17), 6823-7.
33. Pol, A.; Barends, T. R.; Dietl, A.; Khadem, A. F.; Eygensteyn, J.; Jetten, M. S.; Op den Camp, H. J., Rare earth metals are essential for methanotrophic life in volcanic mudpots. *Environ. Microbiol.* **2014**, 16 (1), 255-64.
34. Cotruvo, J. A., Jr.; Featherston, E. R.; Mattocks, J. A.; Ho, J. V.; Laremore, T. N., Lanmodulin: A highly selective lanthanide-binding protein from a lanthanide-utilizing bacterium. *J. Am. Chem. Soc.* **2018**, 140 (44), 15056-15061.
35. Edington, S. C.; Gonzalez, A.; Middendorf, T. R.; Halling, D. B.; Aldrich, R. W.; Baiz, C. R., Coordination to lanthanide ions distorts binding site conformation in calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, 115 (14), 3126-E3134.
36. Mack, K. L.; Moroz, O. V.; Moroz, Y. S.; Olsen, A. B.; McLaughlin, J. M.; Korendovych, I. V., Reprogramming EF-hands for design of catalytically amplified lanthanide sensors. *J. Biol. Inorg. Chem.* **2013**, 18 (4), 411-8.

37. Yee, J., Hypercalcemia. In xPharm: The Comprehensive Pharmacology Reference, Enna, S. J.; Bylund, D. B., Eds. Elsevier: New York, 2007; pp 1-6.
38. Xiong, L. W.; Kleerekoper, Q. K.; Wang, X.; Putkey, J. A., Intra- and interdomain effects due to mutation of calcium-binding sites in calmodulin. *J. Biol. Chem.* **2010**, 285 (11), 8094-103.
39. Ouyang, H.; Vogel, H. J., Metal ion binding to calmodulin: NMR and fluorescence studies. *Biometals* **1998**, 11 (3), 213-22.
40. Guzel, Y.; Rainer, M.; Mirza, M. R.; Bonn, G. K., Highly efficient precipitation of phosphoproteins using trivalent europium, terbium, and erbium ions. *Anal. Bioanal. Chem.* **2012**, 403 (5), 1323-31.
41. Tsien, R.; Pozzan, T., Measurement of cytosolic free Ca²⁺ with quin 2. *Methods Enzymol.* **1989**, 172, 230-62.
42. Martin-Garcia, J. M.; Hansen, D. T.; Zook, J.; Loskutov, A. V.; Robida, M. D.; Craciunescu, F. M.; Sykes, K. F.; Wachter, R. M.; Fromme, P.; Allen, J. P., Purification and biophysical characterization of the CapA membrane protein FTT0807 from *Francisella tularensis*. *Biochemistry* **2014**, 53 (12), 1958-70.
43. Bessette, P. H.; Aslund, F.; Beckwith, J.; Georgiou, G., Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96 (24), 13703-8.
44. Baca, A. M.; Hol, W. G., Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol.* **2000**, 30 (2), 113-8.

45. Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q., Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* **2009**, 18 (5), 936-48.
46. Verma, R.; Boleti, E.; George, A. J., Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems. *J. Immunol. Methods* **1998**, 216 (1-2), 165-81.
47. Lin-Cereghino, G. P.; Stark, C. M.; Kim, D.; Chang, J.; Shaheen, N.; Poerwanto, H.; Agari, K.; Moua, P.; Low, L. K.; Tran, N.; Huang, A. D.; Nattestad, M.; Oshiro, K. T.; Chang, J. W.; Chavan, A.; Tsai, J. W.; Lin-Cereghino, J., The effect of alpha-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene* **2013**, 519 (2), 311-7.
48. Chambers, A. C.; Aksular, M.; Graves, L. P.; Irons, S. L.; Possee, R. D.; King, L. A., Overview of the Baculovirus Expression System. *Curr. Protoc. Protein Sci.* **2018**, 91, 541-546.

Chapter 5. Design of peptides to block association of HIV virus with human cells

This chapter was co-authored with Dr. Pallavi Gosavi, who conducted the initial studies with SEVI fibrils, and Dr. Megha Jayachandran, who tested the binding of selected peptides to SEM1 and SEM2 peptides

Abstract. HIV and AIDS remain a persistent problem in many countries around the world. Amyloid fibril-forming peptides from human semen have been previously shown to enhance HIV infection by facilitating the attachment of virions to cells. These peptides fall into two main groups: derivatives of prostatic acidic phosphatase (PAP), which form a semen-induced enhancer of HIV infection (SEVI), and derivatives of the semenogelins (SEM). Developing a strategy to target amyloid fibrils would provide an effective way to reduce the rate of the sexual transmission of HIV. Using a minimalist approach, we have designed peptide sequences that selectively bind to the amyloid forms of SEVI or SEM, but not the non-aggregated peptides. We also have used a selected peptide as a starting point for the phage display library to enhance peptide's binding and specificity. The identified peptides could potentially reduce the infectivity of HIV virions by preventing their attachment to the target cells.

5.1. Introduction

5.1.1 *HIV and AIDS*

HIV and AIDS remain a persistent problem in many countries around the world. Since it was first discovered in 1981, HIV has affected the lives of millions of people across the globe. According to UNAIDS, there are approximately 38 million people currently living with HIV, with about 2 million being children aged 0-14 years, and tens of millions of people have died of AIDS-related causes since the beginning of the epidemic.¹ The tremendous efforts were put into the prevention and treatment of the disease, as well as the education and scientific research. Despite advances in our scientific understanding of HIV, its prevention and treatment, many people with HIV or at risk for HIV still do not have access to prevention, care, and treatment due to many socioeconomic factors. As of today, AIDS has no successful cure strategy, and the patients receive supporting treatment which allows them to suppress the infection. Even though a few cases of HIV cure were reported, they require further investigation for successful general applicaiton.^{2, 3}

The human immunodeficiency virus (HIV) belongs to the class of retroviruses, the RNA-containing viruses. These viruses carry a reverse transcriptase, the enzyme that converts a single-stranded RNA to double-stranded DNA. Another enzyme, integrase, facilitates the incorporation of the virus into the host cell genome which allows further virus replication. After the viral RNA synthesis, the new virion particle will be assembled near the host cell membrane.^{4, 5} As a result, HIV virions are coated in the lipid bilayer derived from the cell membrane, meaning that viral particles will possess similar properties, including the charge (Figure 5.1). The HIV virions target CD4+ helper

T cells, the key regulators of the humoral and cellular immune responses.⁶ When the virus infects an activated CD4+ T cell, it hijacks and manipulates its transcriptional and translational machinery to reproduce itself leading to the cell damage. The destruction and depletion of T cells result in the weakening of the immune system and impede body's ability to defend itself against opportunistic pathogens.

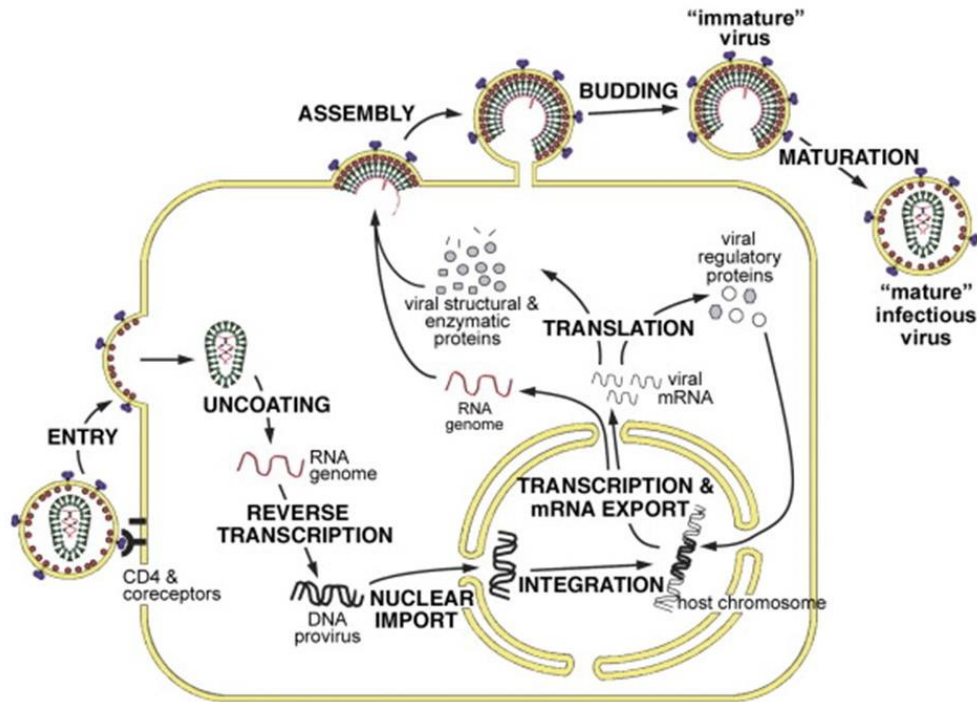


Figure 5.1. HIV virion life cycle. The figure is reproduced from ref. 3 with the permission from Elsevier Ltd., copyright, 2008.

Among all transmission routes, sexual transmission contributes to the HIV epidemic the most, despite the low risk of infection. Approximately 65% of HIV infections recorded in 2020 globally were detected in key populations, which include sex workers and their clients, gay men and other men who have sex with men, people who inject drugs, transgender people and their sexual partners.¹ Thus, the development and improvement of cheap, convenient and accessible prevention strategies are necessary.

The treatment strategies for HIV infection have advanced over the years as our understanding of the infection process progressed. Over 30 drugs have been approved by FDA for the use as part of antiretroviral therapy for patients with AIDS.⁷ These drugs target either viral proteins by blocking their functions or cellular processes involved in the virus life cycle. The current therapies allow patients to reduce the titer of the virus and improve their quality of life. Even though the success in the development of treatment for the infection has been significant, prevention methods are in demand. The current prevention methods for the sexual transmission of HIV include the use of physical contraception methods (condoms), voluntary medical male circumcision (VMMC) and pre-exposure prophylaxis.⁸ The challenges are associated with VMMC, which is a surgical procedure, and prophylaxis, the administration of antiretroviral drugs before infection, as these methods are not widespread and remain accessible only to certain groups of the population.⁹ The use of condoms remains the most accessible method for the prevention of HIV infection. To increase the reliability of this method, strategies aiming to reduce the infectivity of HIV virions need to be developed.

5.1.2. Amyloid fibrils promote HIV infection

The HIV virions can be detected in all body fluids, including human semen. The role of semen in the sexual transmission of HIV is well established as it majorly enhances the infectivity of the virions.¹⁰ To identify natural factors that might play a role in sexual transmission of HIV/AIDS, Munch et. al. screened a library of peptides and proteins derived from human semen to identify the inhibitors and enhancers of HIV infection. They have discovered that proteolytic fragments of prostatic acid phosphatase

(PAP) aggregate to form amyloid fibrils that enhance the infectivity of HIV virions.¹¹ PAP is a dimeric natural enzyme found in human semen and is acting as a tumor suppressor.¹² The proteolytic cleavage of this enzyme produces a pool of 34-40 amino acid residue peptides which can be mapped on the same region of PAP. The predominant fragment, PAP₂₄₈₋₂₈₆, aggregates forming cationic amyloid fibrils named semen-derived enhancer of virus infection (SEVI), which boosts HIV infectivity (Figure 5.2).

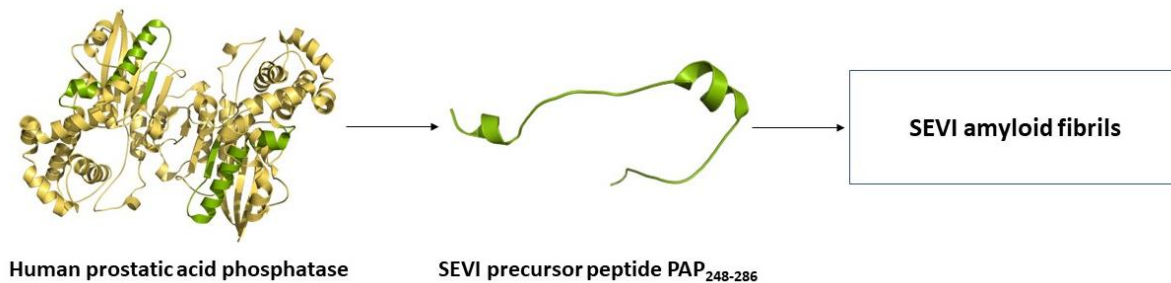


Figure 5.2. The SEVI amyloid fibrils are formed by the peptides derived from the prostatic acid phosphatase (PAP). The structure of PAP is modeled based on PDB 1CVI and the structure of PAP₂₄₆₋₂₈₆ is from PDB 2L3H.

Interestingly, the monomeric form of this peptide or the full-length enzyme were unable to promote viral infection suggesting the exclusive role of amyloid fibrils in this process.¹¹ Further investigations revealed that PAP₂₄₈₋₂₈₆ is highly disordered but contains two helical regions.¹³ Upon fibrillation and formation of SEVI, the aggregates show prominent cross- β structures with significant thioflavin T fluorescence, which are the markers of typical amyloid fibrils. The sequence analysis revealed a highly cationic nature of this peptide (contains six lysine and two arginine residues) suggesting that the positively charged fibrils help to decrease the charge repulsion between the cells and

the virions which both have negatively charged surfaces.¹⁴ Interestingly, further studies revealed the presence of another HIV enhancer derived from PAP – the peptide from the N-terminal region corresponding to residues 85 to 120 (PAP₈₅₋₁₂₀), which also contains many positively charged residues.¹⁵

Another group of amyloidogenic peptides with HIV-enhancing activity originates from semenogelins 1 and 2 upon their proteolysis. Semenogelins, similar to PAP, are found in the human semen and participate in sperm immobilization in the seminal coagulum, which is important for fertilization.¹⁶ In the search for cationic factors contributing to the semen HIV infection enhancement activity, Roan et. al. identified a new group of positively charged amyloid-forming peptides.¹⁷ Similar to SEVI, the cationic nature of these peptides is playing an important role in the enhancement of the infection by increasing the probability of virion binding to the cell.

Even though all experiments for the identification of HIV enhancers were conducted *in vitro*, it was important to mention that the amyloid fibrils were detected in the semen of donors suggesting the biological relevance of these fibrils.¹⁸ The importance of the charge for the infectivity was confirmed by removing the positively charged residues from PAP₂₄₈₋₂₈₆: such manipulation resulted in the peptide being susceptible to fibrillation but unable to facilitate virion binding to human cells.¹⁴ As amyloid fibrils play a significant role in HIV infection, several inhibitors of SEVI and semen-mediated enhancement of HIV infection have been developed.

It was demonstrated that the intrinsic positive charges of SEVI facilitate virion attachment to and fusion with target cells. Anionic polymers were proposed to disrupt the electrostatic interactions between fibrils, virions and cells. Heparin, dextran sulfate,

oversulfated heparin, and oversulfated chondroitin sulfate were shown to bind to the amyloid fibrils and block the binding of virions.¹⁴ Unfortunately, the addition of various polyanionic compounds to the monomeric forms of amyloidogenic peptides caused a rapid aggregation of PAP₂₄₈₋₂₈₆ into the amyloid fibrils and facilitated the infection, making these compounds unsuitable for use as microbicides.¹⁹ Additionally, anionic compounds have been shown to induce inflammation in the reproductive organs of women.²⁰

The observed interaction between SEVI and heparan sulfate suggested that these amyloid fibrils could potentially bind to the heparan sulfate proteoglycans (HSPG), naturally occurring anionic carbohydrate polymers found on human cells, and this interaction could be disrupted using heparan sulfate proteoglycan antagonists. Roan et al. showed that the pretreatment of the cells with surfen, HSPG antagonist with anti-inflammatory and anti-bacterial activity, reduced the HIV infection enhancement activity of SEVI 3-fold by inhibiting the binding of virions to fibrils.²¹ Other small molecules, including BTA-EG₆ (the hexa(ethylene glycol) derivative of benzothiazole aniline), ADS-J1 (an HIV virion fusion inhibitor), EGCG (epigallocatechin-3-gallate, polyphenol found in green tea), myricetin (a common dietary flavonoid), tolcapone (an FDA-approved agent for Parkinson's disease), have been tested for their ability to prevent the virion binding to the fibrils or to break down the fibrils, but their mechanisms of action need further investigation.²²⁻²⁶ Additionally, the molecular tweezer CLR01 was designed by Sievers et. al. to interact with the positively charged SEVI fibrils and remodel them to disrupt the effective binding to virions.²⁷

All mentioned methods present a great opportunity for the development of effective microbicides and provide evidence that SEVI is a promising target for such agents. In our work, we attempted to develop the peptides capable of specific binding to cationic fibrils and interfering with their HIV-virion-binding activity using rational design and minimalist approach. We have created a series of peptides, by modifying the known antimicrobial sequences, and tested their binding to amyloid fibrils. We have shown that short anionic peptides can bind to amyloid fibrils and distinguish between monomeric and aggregated forms. We have identified the peptides that selectively bind to SEVI and SEM1 fibrils. To further expand the library of specific binders and analyze the importance of amino acids in the lead peptide, we have employed the phage display. The produced peptides present promising candidates for further examination of their potential use as microbicides for prevention of HIV infection.

5.2. Results and Discussion

5.2.1. Design of amyloid binding peptides

SEVI and SEM amyloid fibrils found in human semen contain many positively charged residues and their cationic nature is important for the enhancement of HIV infection.^{11, 18} Many polyanionic compounds were shown to effectively bind SEVI fibrils and interfere with the virion binding to fibrils. Although those molecules interacted with the fibrils, many of them also had side effects which lead to the cytotoxic or pro-inflammatory effects. In our work, we hypothesized that natural peptides can be used as scaffolds for the further development of amyloid-specific binders. We aimed to create a short negatively charged peptide that could bind to semen-derived amyloid fibrils and disrupt the interaction between amyloid and the virion (Figure 5.3).

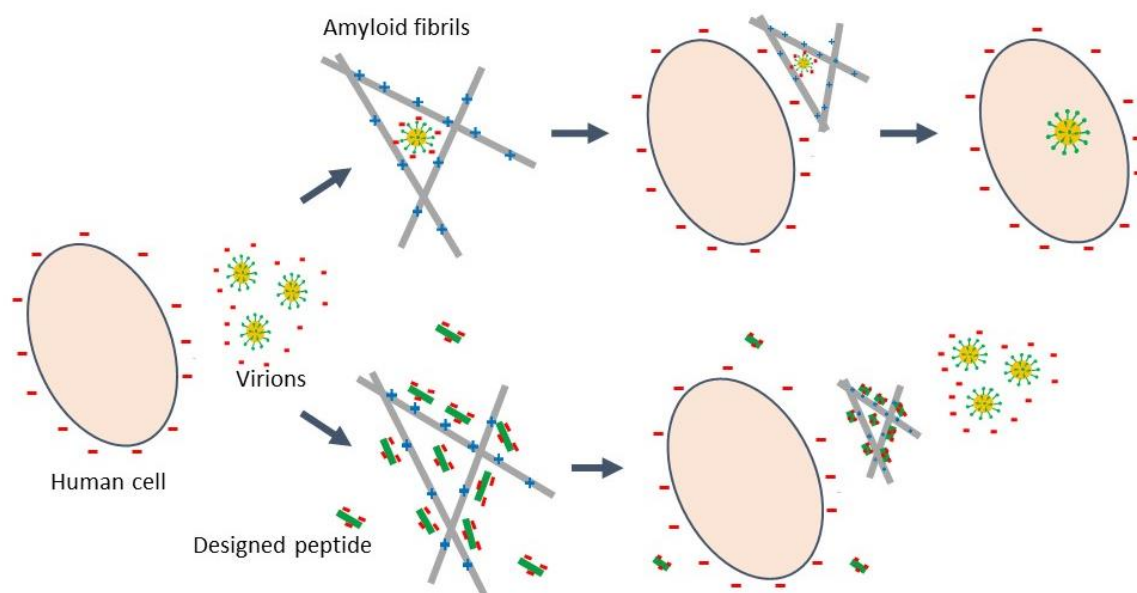


Figure 5.3. The schematic representation of the activity of designed peptides.

Upon binding to the cationic amyloid fibrils, the designed peptides will block the association of fibrils with HIV virions and reduce the rate of HIV infection.

Besides charge neutralization, the binding of peptides to amyloid fibrils, but not the monomers, will provide the opportunity for specific targeting of the active species which would simplify the further development of the microbicide.

We decided to focus our efforts on the semen-derived enhancer of virus infection (SEVI) and the most common SEM1 and SEM2 fibrils (Table 5.1). The selected sequences differ in length and net charge, they come from different precursors, but all are amyloidogenic and were shown to promote HIV infection. The use of diverse target amyloids also provides a possibility for testing the specificity of the designed peptides.

Table 5.1. The sequences of peptides found in human semen and shown to form cationic amyloid fibrils that promote HIV infection. The residues that possess positive charge at physiological pH are colored in blue and those that have negative charge are shown in red.

Name	Sequence	pI
PAP ₂₄₈₋₂₈₆	GIHKQKEKSR LQGGVLVNEILNHMKRATQIPSYKKLIMY	10.21
SEM1 ₈₆₋₁₀₇	DLNALHKTTKSQRHLGGSQQLL	9.99
SEM2 ₆₈₋₁₀₇	TYHVDINDHDWTRKSQQYDLNALHKATKSKQHLGGSQQLL	8.12

Antimicrobial peptides are the natural agents acting as a part of the innate immune system. They are found across different species, usually 10-40 amino acids long, generally have a positive net charge and low toxicity to the host organism.^{28, 29} Antimicrobial peptides are good starting points for our design as these molecules were extensively studied in different conditions which provides us with useful information about their properties. We suggested that the replacement of positively charged amino

acid residues in known antimicrobial peptides with negatively charged ones can lead to the design of effective and non-toxic amyloid fibril binders. We have selected a set of known antimicrobial peptides and exchanged all arginine and lysine residues for glutamic acid residues. This allowed us to create a library of negatively charged peptides and test their ability to bind to amyloid fibrils. The initial experiments, previously conducted in our lab, allowed us to select a group of the most promising candidates (Table 5.2) which were further tested for binding to amyloid fibrils.

Table 5.2. Peptides used in this study and their theoretical pI values calculated using ExPASy ProtParam. The positively charged residues are highlighted in blue and the negatively charged residues are colored in red.

Peptide	Sequence	pI	Design	Sequence	pI
FKV	FKRIVQR IKDFLR NLV	11.72	FEV	F EE IVQ EIED FL EN LV	3.40
ILR	ILPW KWPWWP W RR	12.01	ILE	ILPW EWPWWP W EE	3.67
GKV	G K PR P YS P RS P TS H PR P IRV	12.01	GEV	G E PE P YS P E P T S HP E PI E V	3.98

5.2.2. Designed peptides selectively bind cationic amyloid fibrils

We applied fluorescence spectroscopy to monitor the binding of the designed peptides to fibrils. Fluorescence assays have been widely used for the characterization of protein-protein and protein-ligand interactions.³⁰ The fluorescence emission maximum of 5-(dimethylamino)naphthalene-1-sulfonyl or dansyl is known to be strongly dependent on the polarity of its environment.³¹⁻³³ The increase in the fluorescence quantum yield, as well as the blue shift, of the dansyl group is observed when the fluorophore is placed in the hydrophobic environment.³⁴ As amyloid fibrils possess a

high degree of hydrophobicity, the increase in fluorescence would indicate that the labeled peptide is bound to the fibril.

To evaluate the binding of the designed peptides to amyloid fibrils, we have synthesized all sequences using solid phase peptide synthesis. We have functionalized the designed peptides and the parent antimicrobial sequences with the N-terminal dansyl fluorophore. This allowed us to monitor the binding of these peptides to the hydrophobic amyloid fibrils by observing the change in the fluorescence of the functional group.

The binding of the anionic peptide named GEV to different amyloid fibrils was monitored using a fluorescence spectroscopy assay. The fluorescence spectra were recorded after the mixing of the dansylated peptide with either fibrillated or non-fibrillated forms of amyloidogenic peptides to visualize the initial interactions between molecules. Then, the reaction mixtures were incubated at room temperature for 24 hours to evaluate the impact of prolonged incubation on the stability of the complex. The binding between the dansylated peptide and the hydrophobic amyloid fibril is characterized by the emission peak around 510 nm and by the increase in the intensity of the fluorescence.³⁵ The obtained results suggested that GEV does not bind to any of the tested amyloid fibrils (Figure 5.4).

The peptide named ILE was also tested for binding to the amyloid fibrils and has been shown to bind to both the monomeric peptide PAP₂₄₈₋₂₈₆ and its fibrillated form SEVI (Figure 5.5. A). Interestingly, after 24 hours the intensity of the emission spectrum for the sample containing non-fibrillated PAP₂₄₈₋₂₈₆ and ILE dropped (Figure 5.5 D), which might be associated with the peptide complex rearrangement. The incubation of

the dansylated ILE with the fibrillated SEM1 resulted in an increase in dansyl fluorescence, compared to the sample containing its non-fibrillated form. Even though the fluorescence intensity was low, we have observed the peak shift of the emission spectrum which suggests the positioning of the dansyl group in the hydrophobic environment upon peptide-amyloid binding (Figure 5.5 B and E).

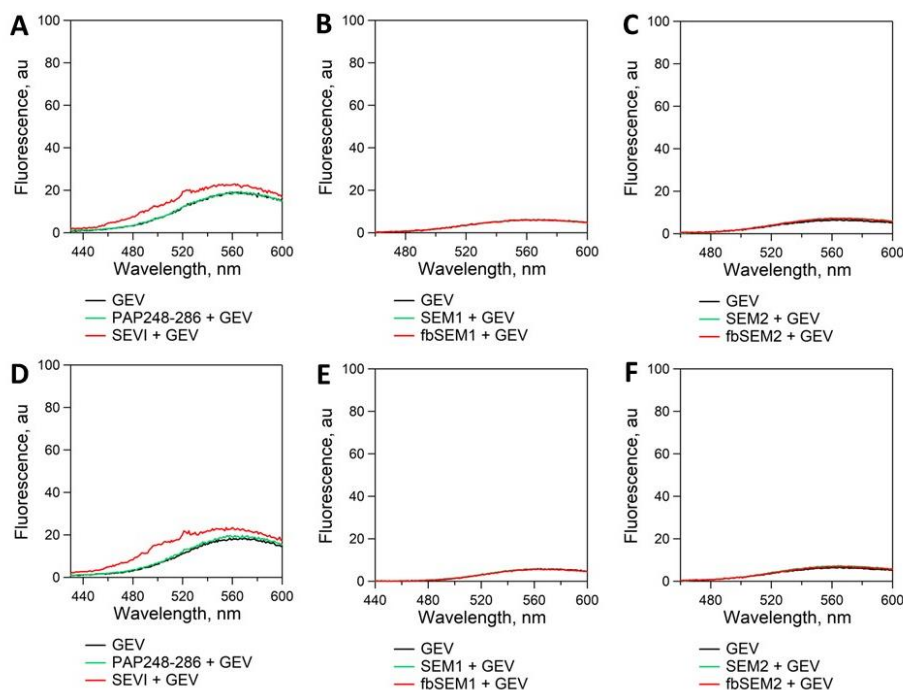


Figure 5.4. The fluorescence spectra for the reaction containing dansylated GEV peptide and amyloid fibrils. The spectra were collected right after the mixing (A-C) and after the 24-hour incubation of the reactions at room temperature (D-F).

As the fluorescence intensity for ILE binding to fibrillated SEM1 was low, we decided to confirm the binding by conducting additional experiments. Amyloid fibrils are relatively big molecules that can be precipitated after centrifugation. Per our experimental design, the dansylated ILE should bind to the amyloid fibrils and after the

fibrils are removed from the solution, we should not be able to observe the fluorescence of the dansyl group.

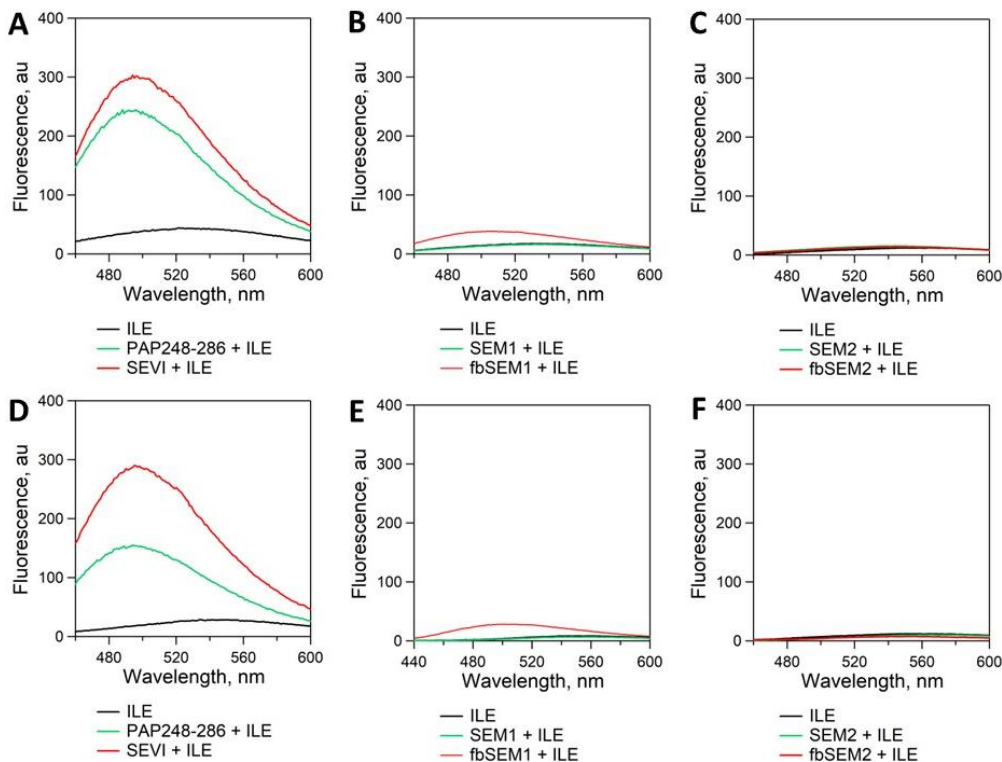


Figure 5.5. The fluorescence spectra for the reaction containing dansylated ILE peptide and amyloid fibrils. The spectra were collected right after the mixing (A-C) and after the 24-hour incubation of the reactions at room temperature (D-F).

The analysis of the supernatant obtained after the centrifugation of the reaction showed no emission at 510 nm confirming peptide binding to the amyloid fibrils and their co-precipitation (Figure 5.6). We have also tested the binding of ILE to the fibrillated SEM1 peptide at different ratios (Figure 5.7). For this, the dansylated peptide was added to the fibrils in small aliquots and the fluorescence was recorded. The titration showed that more than 10 equivalents of the dansylated peptide can bind to the

amyloid fibrils formed by the SEM1₈₆₋₁₀₇ peptide, which suggests that short peptides have multiple binding sites on the amyloid fibrils.

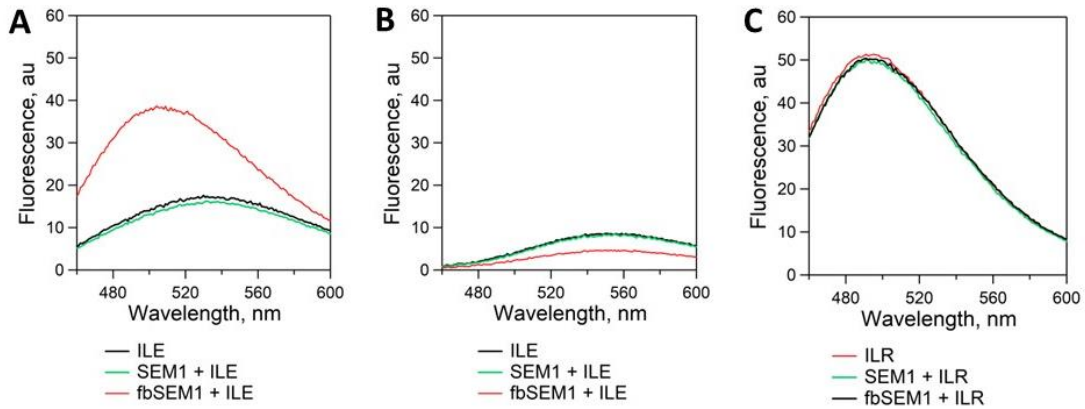


Figure 5.6. The fluorescence spectra for the reaction containing dansylated ILE and amyloid fibrils before (A, zoomed Figure 5.5. B) and after (B) centrifugation. The binding of positively charged ILR (C) to the amyloid peptides showed no interaction. The fluorescence spectra were recorded at 25° in PBS buffer (pH 7.3).

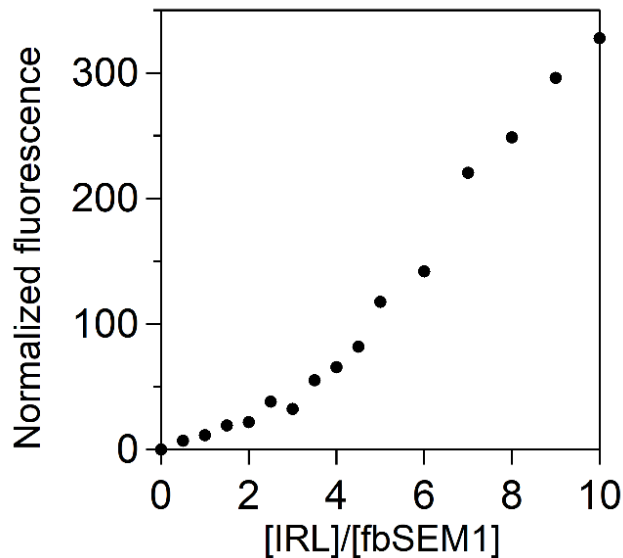


Figure 5.7. The titration of fibrillated SEM1 (5 μM) with dansylated ILE peptide at pH 7.3. The fluorescence at 510 nm was recorded and plotted versus the peptide/fibril ratio.

The designed anionic peptides are expected to bind to the cationic amyloid fibrils based on the electrostatic interactions. To confirm the importance of the negative charge in ILE peptide, we have tested the original positively charged antimicrobial peptide ILR for the binding to amyloid fibrils. Unlike other peptides tested in this work, ILR had low solubility in the PBS buffer we used for the studies. To increase peptide solubility, the stock was prepared in a mixture of 50% isopropanol in the buffer. During the analysis of ILR binding to the amyloidogenic peptide, we observed a peak at 510 nm which is usually associated with the placement of the dansyl group in the hydrophobic environment. As the emission spectra for the sample that contained only dansylated peptide are the same as those for the samples with amyloidogenic peptides, we suggested that there was no binding of ILR to SEM1 (Figure 5.7), which confirms the importance of the negative charge for the function of the designed peptides.

Similar to ILE, FEV anionic peptide showed diverse binding activities towards different amyloid fibrils. No interaction was observed between FEV and SEM1-derived fibrils or monomeric peptides (Figure 5.8. B and E). The increase in fluorescence and the peak shift were observed upon mixing fibrillated and non-fibrillated forms of SEM2 with FEV peptide suggesting that the peptide is not specific to amyloid fibrils (Figure 5.8 C and F). On contrary, this peptide selectively binds to the SEVI fibrils (Figure 5.8 A and D). Interestingly, PAP₂₄₈₋₂₈₆ is the most cationic amyloidogenic peptide tested (pI 10.21) and FEV peptide is the most negatively charged binder designed in this study (pI 3.40).

Our results showed that rationally designed peptides can bind to the preformed amyloid fibrils and differentiate between monomeric and aggregated forms. This suggests that our peptides bind not only using the electrostatic interactions, even

though the charge is essential for binding, but also exhibit sequence-specific binding. Such selective interactions could be potentially utilized to block the association of HIV particles and amyloid fibrils which promote virion binding to human cells.

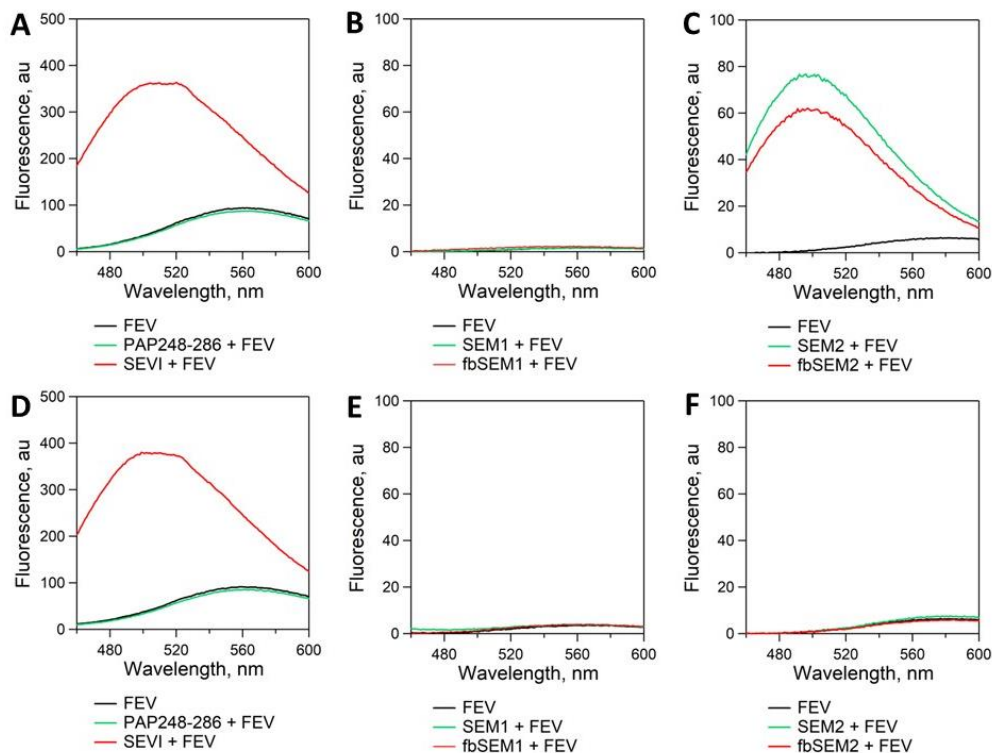


Figure 5.8. The fluorescence spectra for the reaction containing dansylated FEV peptide and amyloid fibrils. The spectra were collected right after the mixing (A-C) and after the 24-hour incubation of the reactions at room temperature (D-F).

5.2.3. FEV as a promising scaffold for the HIV microbicide development

We have shown that rationally designed anionic peptides can selectively bind to the amyloid fibrils. Dansylated ILE peptide could differentiate between monomeric and fibrillated forms of SEM1₈₆₋₁₀₇ peptide, while FEV binds specifically to SEVI but not monomeric PAP₂₄₈₋₂₈₆. As SEVI is the predominant form of amyloid fibrils found in

human semen, we decided to focus our further efforts on the improvement and characterization of the SEVI binder, the FEV peptide.

As our final goal is to use the designed peptide to prevent HIV infection *in vivo*, we have tested the toxicity of this peptide towards mammalian cells. We have analyzed the effect of 1, 5, 10, 25 and 50 μM FEV on the viability of HEK-293 cells using a resazurin-based assay (Figure 5.9). Resazurin is a non-fluorescent dye that undergoes reduction inside metabolically active cells to the fluorescent product resorufin. The production of resorufin is then correlated to the cell viability in the experiment.³⁶ These initial results showed that the peptide did not have a cytotoxic effect at the highest tested concentration and, thus, could be used for further testing in cell-based assays. To evaluate the HIV infection rate reduction activity of the designed peptides, they will be further analyzed in the luciferase assay using TZM-bl cells, a HeLa-derived cell line used to evaluate HIV-1 virion binding to CD4/CCR5-positive cells.³⁷

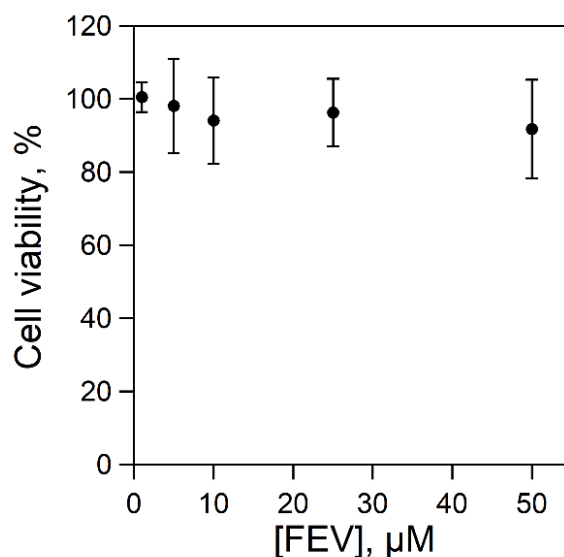


Figure 5.9. The test of FEV cytotoxicity towards HEK-293 mammalian cells. The viability of the cells treated with the peptide was normalized to the untreated cells.

Additionally, we decided to create a phage display library based on the sequence of FEV peptide. Phage display is a versatile *in vitro* selection technique that allows to generate millions of sequences and test their binding to the specific target. Phage display has been successfully employed for the identification of monoclonal antibodies and their modifications, B-cell and T-cell epitope mapping, and selection of peptides for specific binding to enzymes or receptors.^{38, 39} The technique requires the preparation of a targeted library of sequences, production of phages that express and manifest those peptide sequences, and a further selection of the best binders. The analysis of the phage display library will allow us to identify the sequences with possibly improved binding to the target amyloid, but also could provide information about the role of specific amino acids in the peptide-amyloid interactions.

In our work, we have created a library of peptides using FEV sequence as a scaffold. The introduction of degenerate codons (Table S 3)⁴⁰ allowed us to introduce selected alternative bases at the specific positions in the codons and create a library of amino acid substitutions for further testing. Previously, we have shown the importance of the negatively charged residues for the peptide binding to amyloid fibrils, thus we decided to preserve the negatively charged amino acids at their original positions by testing only glutamic and aspartic acids, two amino acids that bear negative charges at neutral pH. All other positions contained the combination of hydrophobic amino acids and those that can participate in hydrogen bonding. As a result, we have created a library of 3.6×10^6 peptide sequences fused to the N-terminus of the pIII phage capsid protein using PhD peptide cloning system from New England Biolabs.

The selection of peptides from the phage display library involves panning, the incubation of the phage pool with the target molecule (in our case – SEVI amyloid fibrils) and further sequence identification. After the first round of panning, the collected phages undergo the next round of panning in order to enrich the library with the most productive sequences. After three rounds of selection against the SEVI fibrils, we could identify a set of peptides with possibly improved binding to SEVI (Table 5.3).

Table 5.3. Amino acid sequences obtained after three rounds of panning using FEV-based phage library.

Original sequence	pI
FEEIVQEIEDFLENLV	3.40
Hits from round 2	
FDDVLEEVDSELENFV	3.26
FEEVVQDLEDSVENVL	3.34
SDDLQDIEEYLDNLV (SDDV)	3.25
SDDMLQDLDDYVENFL	3.21
SDEILEELDEYVENFV	3.30
SEDFVQDVDEYMENVF	3.29
YDDLLEEVEDSFENII (YDDI)	3.26
YDDLQEFEDSFENVL	3.29
YDDMVEEVDDFVENMM	3.22
YDELFQEVDDYEMENIM	3.29
YDEVIQDIEEYVDNIV	3.29
YEDVLQDLDEYVDNIF	3.25
Hits from round 3	
SEDFIEDFEEYIENMI	3.30
YDDMIQDMEEYVDNVL	3.25
YDDVIQEFEDSIENVV	3.29
YDEIVQEVEESFENVI	3.40
YDEVFEDIDEYIENVM	3.26
YDEVVQDVDDFLENIL	3.25
YEDMIQDLDEYVENMI	3.29

Interestingly, most sequences contained tyrosine instead of phenylalanine at position 1, the aspartic acid in place of glutamic acid at positions 2 and 3, whereas, similar to the original peptide, glutamic acid was prevalent at position 13 (Figure 5.10). It is also interesting to note that glutamine at position 6 was more prevalent than glutamic acid, suggesting that the addition of the extra negative charge does not necessarily improve peptide affinity to the amyloid fibrils. We have analyzed a set of peptides identified from the phage display to confirm that these sequences preserve binding affinity to SEVI.



Figure 5.10. A logo plot (a graphical representation of an amino acid multiple sequence alignment) for the peptides identified from the phage display library after the three rounds of panning. The plot was created using WebLogo.⁴¹ The height of single-letter amino acid notation denotes the relative frequency of each amino acid in that position.

We have synthesized a few peptides identified from the phage display library. The selected sequences showed binding to SEVI amyloid fibrils but not to the monomeric form of this peptide (Figure 5.11), similar to the parent sequence.

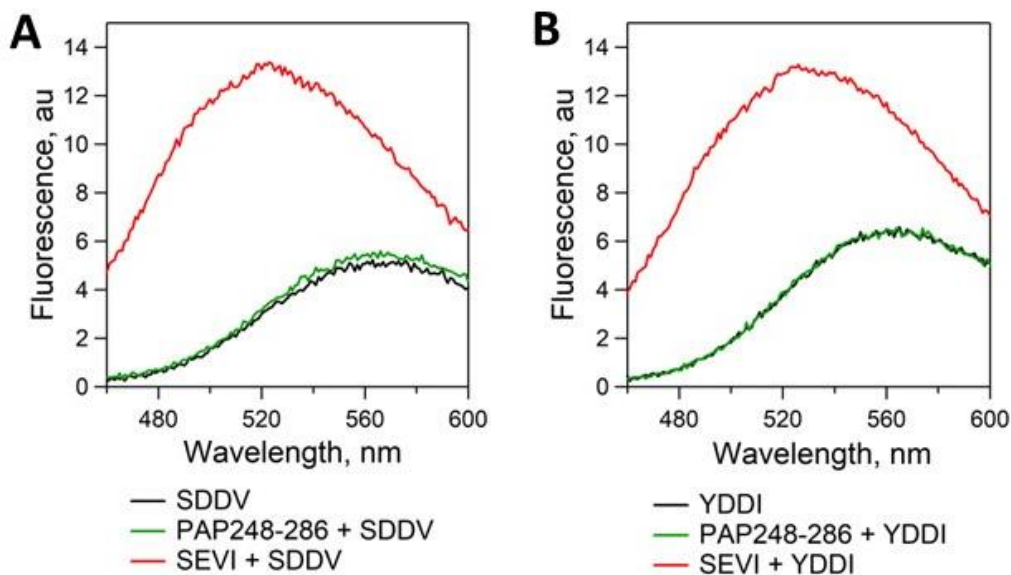


Figure 5.11. The fluorescence spectra for the reaction containing dansylated peptides, identified from the phage display library, and amyloid fibrils.

Overall, the phage display allowed us to identify peptides with potentially improved binding affinity to the SEVI amyloid fibrils. Further characterization will include the preparation of the larger selection of peptides and quantification of their affinity to the amyloid fibrils. Such experiments will allow us to evaluate the contribution of each amino acid in the sequence and potentially find some dependencies between the structure of the peptide and its binding activity.

5.3. Conclusions and Outlook

Despite the low infectivity rate, sexually transmitted HIV contributes to more than half of all cases of AIDS. Today, many strategies have been developed for the treatment of HIV infection, but cheap, convenient and accessible prevention methods still need to be developed. With the discovery of the amyloid fibrils, it has been shown that human semen plays a major role in HIV transmission by increasing the infectivity rate. The new prevention strategies are focused on the development of agents that can bind to the cationic amyloid fibrils and disrupt their interaction with the virions.

In this work, we have successfully demonstrated that rationally designed peptides can selectively bind to semen-derived amyloid fibrils. FEV peptide showed selective binding to SEVI fibrils while ILE interacts with the fibrillated form of SEM1. While the charge-charge interactions are important for the activity of the peptides, we have shown that the identity of the amino acid sequence also plays a role in selective binding. Using phage display, we have identified a library of FEV derivatives that were shown to bind to SEVI. The majority of hits identified during phage panning contained the substitution which led to the introduction of hydrogen bond donors into the sequence, while at some positions the parent amino acids were preserved. These observations confirm the importance of the sequence identity of the peptide for the successful binding to the amyloid fibrils.

The future work will include the verification of peptide's ability to block the association of HIV virions with cells in the *in vitro* cell-based experiments. The specific targeting of seminal amyloid fibrils provides an opportunity for the development of new agents that reduce HIV transmission.

5.4. Experimental

Peptide synthesis and purification. All peptides were produced by manual fluorenylmethyloxycarbonyl (Fmoc) solid phase peptide synthesis at elevated temperature (70 °C) using Rink Amid resin (ChemImpex International). Briefly, the resin was swelled in DMF for 30-40 min at RT. Each amino acid addition cycle started with Fmoc-deprotection with 5% piperazine in DMF for 3 minutes (after 10 AA the times was increased twice) and followed by the flow-wash with DMF for 2 min. Coupling of the F-moc-protected amino acid in the presence of HCTU [2-(6-chloro-1H-benzotriazol-1-yl)-1,1,3,3-tetramethylamminiumhexafluorophosphate] and DIEA (N,N-diisopropylethylamine) (AA:HCTU:DIEA:resin 3:2.8:6:1) was carried for 2 min (after 10 AA the time was increased twice) with subsequent flow-wash with DMF for 2 min twice. The peptides were capped using acetic anhydride (acetic anhydride:resin 6:1) at RT for 2 min. The peptides for the fluorescent analysis were dansylated by incubation of resin with 1 mL of dansyl chloride in DMF (DC:DIEA:resin 1.5:6:1) for 2 H at RT. Cleavage of the peptides from resin and simultaneous side chain deprotection was carried in a mixture of trifluoroacetic acid (TFA)/H₂O/triisopropylsilane (TIS) (95:2.5:2.5, v/v) for 2 hours at room temperature with stirring. The crude peptide samples were separated from the resin using a glass wool filter and concentrated to remove TFA excess using a nitrogen flow. Peptide solutions were further precipitated in the cold tert-butyl methyl ether, the precipitate was collected by centrifugation and the procedure was repeated for 3 times. After the ether was evaporated, peptides were dissolved in TFA, diluted with HPLC solvents, filtered using 0.22 µm PES filter and injected on the column. Preparative reverse phase HPLC purification was performed on Shimadzu using C4 preparative

column (Phenomenex) using a custom linear gradient of solvent A (0.1% TFA in water) and solvent B (0.1% TFA, 10% water in 90% acetonitrile). The purified peptide fractions were analyzed by MALDI-TOF on Bruker Autoflex III Smartbeam mass spectrometer. The purity of the peptides was determined on Agilent 1260 Infinity II HPLC instrument with an analytical Zorbax Eclipse XDB-C18 column (4.6 mm x 150 mm).

Fibril preparation. SEM1 peptide stock was prepared at 2.5 mg/mL (or 0.7 mM) in PBS1x (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na₂HPO₄, 1.47 mM KH₂PO₄, pH 7.3). The concentration was confirmed by UV-vis spectroscopy using the extinction coefficient at 214 nm of 31, 809 M⁻¹ cm⁻¹. The 500 µL aliquots were placed in the glass vials and incubated for 24 hours at 37°C with 1400 rpm shaking. SEVI fibrils were prepared at 2 mg/mL by dissolving lyophilized PAP₂₄₈₋₂₈₆ peptide in PBS and incubating the peptide solution at 37°C with 1400 rpm agitation for 30 hours in glass vials. The concentration of PAP peptide stock was also verified by UV-vis spectroscopy using the extinction coefficient of $\epsilon_{280\text{nm}}$ 2,980 M⁻¹ cm⁻¹.

Dansylated peptide stock preparation. Lyophilized peptides were dissolved in the appropriate solvent and the concentration was determined based on the UV-vis absorbance using the extinction coefficient of dansyl at 330 nm of 3,430 M⁻¹ cm⁻¹ (see the table below). Dansyl-FEV-CONH₂, Dansyl-FKV-CONH₂, Dansyl-ILE-CONH₂ were dissolved in PBS1x buffer and showed fluorescence at 550 nm. Dansyl-ILR-CONH₂ was dissolved in PBS1x and isopropanol 1:1 mix. Dansyl-GEV-CONH₂ and Dansyl-GKV-CONH₂ were dissolved in PBS1x.

Fluorescence binding assay. Fibrillated and non-fibrillated peptide samples (final concentration of 5 µM) were mixed with 5 µM dansylated peptides in PBS buffer

making the final reaction volume of 500 μ L. The fluorescence spectra were collected immediately and after 24 h incubation at RT on a Cary Eclipse (Agilent) fluorescence spectrophotometer in a rectangular SpectroSil quartz cuvette (Starna Cells, Inc.) with a pathlength of 0.21 cm. The dansyl group was excited at 330 nm and the emission spectra were recorded in the range of 440-600 nm. The other settings: emission and excitation band width of 5 nm, the scan control at medium, auto excitation filter, open emission filter, PMT detector voltage at 600. The fluorescence is reported as the average of three scans for each sample. The samples were further incubated at RT and the fluorescence was recorded again after 24 H.

Cell cytotoxicity assay. HEK-293 cells were cultured in DMEM high glucose medium (Corning) supplemented with 10% FBS (Gibco) and 1% penicillin-streptomycin mix (Gibco) at 37°C in high humidity environment. After cell detachment by trypsinization, cells were stained using trypan blue and counted using hemocytometer. The cells were transferred in 96-well plate (Corning) in 160 μ L per well resulting in 50,000 cells per well and incubated overnight under cell culture conditions. The peptides were added as 9x stocks (20 μ L) and the cells were incubated for 24 hours. Resazurin was added to the final concentration of 0.7 mM, incubated for 3 hours at 37°C and the fluorescence was measured using BioTek Synergy H1 Hybrid instrument (excitation – 550 nm, emission – 585 nm).

Thioflavin T Assay. A solution of 0.5 mM thioflavin T was prepared in water. Samples were prepared by mixing 5 μ M ThT and 5 μ M fibril solution. Fluorescence spectra were collected on Cary Eclipse (Agilent) fluorescence spectrophotometer with

excitation at 440 nm and slit width of 5 nm using a quartz cuvette. The emission spectra were recorded from 460 nm to 600 nm with the slit width of 10 nm.

Phage display library preparation. The phage display library was constructed using Ph.D Peptide Display Cloning System (New England Biolabs Cat E8101S) based on M13KE vector. The following designed primer was synthesized by Integrated DNA technologies:

Library oligonucleotide: 5'- CAT GTT TCG GCC GAG CCG CCG CCM ANM
ANA TTK TCM ANA DAK TCK TCM ANK TCT TSM ANM ANK TCK TCA DAA GAG
TGA GAA TAG AAA GGT ACC CGG G-3'

Universal extension primer was provided with the Cloning System: 5'- CAT GCC
CGG GTA CCT TTC TAT TCT C-3'.

The library oligonucleotide primer (5 µg, 165 pmol) was mixed with extension primer (495 pmol, 3 molar equivalence) in total volume of 50 µL of TE buffer containing 10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl (pH 8.0). The mixture was incubated at 95 °C for 1 min (Primer Annealing) and slowly cooled down to 25 °C for 30 min (1°C/26 sec) in a thermocycler (BioRad). The duplex was then extended by mixing 50 µL of the annealing reaction with Klenow fragment (30 units), 0.4 mM dNTPs, 10X Klenow buffer and water to final volume of 200 µL. This reaction mixture was incubated at 37 °C for 10 min and with further enzyme deactivation at 65 °C for 15 min. The extended duplex reaction and 10 µg of M13KE vector were then digested with 50 units of EagI-HF and 50 units of Acc65I (New England Biolabs) in NEBuffer 3.1 by incubating at 37 °C for 3 h. After the incubation digested duplex (insert) was purified using PCR purification kit (Qiagen). The size of the insert was assessed on 8% non-denaturing polyacrylamide gel

comparing to pBR322 DNA-MspI DNA molecular weight marker. The digested vector was purified from 0.8% agarose gel using gel purification kit using EZ-10 spin column DNA gel extraction kit (Bio basic Inc.).

The ligation reaction containing the purified insert and digested M13KE phage vector in 3:1 ratio was mixed with 200 units of T4 DNA ligase (Thermo Scientific) in T4 DNA ligase buffer (the total volume of reaction was 40 μ L) and incubated overnight at 16 °C with further enzyme heat inactivation at 65 °C for 15 min. The ligated plasmid was then incorporated into ER2738 cells (New England Biolabs) by electroporation using BioRad MicroPulser Electroporation Unit. The 100- μ L aliquot of electrocompetent ER2738 cells were mixed with 2 μ L (20 electroporations total) of ligation reaction on ice, carefully transferred to 0.2-cm electroporation cuvette and electroporated at 2.5 kV (program Ec2). Immediately, 1 mL of room temperature SOC medium was added to the cells and the cell mixture was transferred to 15-mL PP tube. The SOC outgrowths from 5 electroporation were combined in one tube and incubated at 37°C for 30-45 min. Then, total of 20 mL of SOC outgrowth were transferred to 1 L of phage LB medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) and incubated at 37°C for 4.5-5 hours. The cells were further collected by centrifugation at 5,000xg for 20 min at 4°C. The phages were recovered from the supernatant by adding 1/6 volume of 20% PEG-8000/2.5 M NaCl and incubating overnight at 4 °C. The phages were pelleted by centrifugation at 5,000xg for 20 min at 4 °C to obtain white pellet. The pellet was resuspended in 100 mL of TBS buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) by gentle shaking at 4°C. The residual cells were removed by centrifugation at 5,000xg for 20 min at 4°C. The phages were again precipitated from the supernatant with PEG/NaCl

solution at 4°C for 1 hour. The solution was then centrifuged at 5,000xg for 20 min at 4°C, the pellet was resuspended in TBS buffer (25 mL) by rocking at 4°C for 48 h and stored at 4°C until panning.

Phage surface panning. Prepare SEVI fibrils as described above, add 1.5 mL of fibrils onto the 6-cm tissue culture treated dish and incubate overnight at 4°C with gentle rocking in a humidified environment. Remove non-bound peptides, add 1.5 mL of blocking buffer (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na₂HPO₄, 1.47 mM KH₂PO₄, 5 mg/mL BSA, 0.02% sodium azide) and incubate for 2 hours at 4°C. The control plate was prepared without fibril binding step. After the incubation remove the blocking buffer and wash the plate with TBS/0.1% Tween-20 for 6 times. Add 1 mL of the phage library together with 0.5 mL of 3x TBS/0.3% Tween-20 and incubated for 2 hours at RT on the rocker. After the incubation discard non-bound phages and wash plates with TBS/0.1% Tween-20 for 6 times. Elute the phages with 1 mL of elution buffer (0.2 M glycine, 1 mg/mL BSA, pH 2.2) for 30 min at RT and neutralize with 150 µL of 1M Tris-HCl (pH 9.17). The elutions were kept at 4°C over night. The 20-mL of phage LB medium were inoculated with ER2738 cells and grown for 3 hours at 37°C in a glass flask. After the eluate was added the culture was incubated for 4.5 more hours, cells were pelleted at 5,000xg for 20 min at 4°C and then at 12,000xg for 10 min. The supernatant was transferred to a fresh tube and phages were precipitated by adding 1/6 volume of 20% PEG-8000/2.5 M NaCl) over night at 4°C. The phages were collected by centrifugation (12,000xg, 15 min, 4°C), resuspended in 1 mL of TBS buffer and transferred in 1.5-mL PP tube to further remove all residual cells by centrifugation. The phages were PEG/NaCl precipitated again and resuspended in 200 µL of TBS buffer. The panning

was repeated twice using the phages from round 1 and the obtained phages (round 2 and 3) were sequenced.

5.5. References

1. UNAIDS, D., Report. *Retrieved February 2021*.
2. Hutter, G.; Nowak, D.; Mossner, M.; Ganepola, S.; Mussig, A.; Allers, K.; Schneider, T.; Hofmann, J.; Kucherer, C.; Blau, O.; Blau, I. W.; Hofmann, W. K.; Thiel, E., Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N. Engl. J. Med.* **2009**, *360* (7), 692-8.
3. Gupta, R. K.; Abdul-Jawad, S.; McCoy, L. E.; Mok, H. P.; Peppas, D.; Salgado, M.; Martinez-Picado, J.; Nijhuis, M.; Wensing, A. M. J.; Lee, H.; Grant, P.; Nastouli, E.; Lambert, J.; Pace, M.; Salasc, F.; Monit, C.; Innes, A. J.; Muir, L.; Waters, L.; Frater, J.; Lever, A. M. L.; Edwards, S. G.; Gabriel, I. H.; Olavarria, E., HIV-1 remission following CCR5Delta32/Delta32 haematopoietic stem-cell transplantation. *Nature* **2019**, *568* (7751), 244-248.
4. Goodsell, D. S., Illustrations of the HIV Life Cycle. In *The Future of HIV-1 Therapeutics: Resistance Is Futile?*, Torbett, B. E.; Goodsell, D. S.; Richman, D. D., Eds. Springer International Publishing: Cham, 2015; pp 243-252.
5. Ganser-Pornillos, B. K.; Yeager, M.; Sundquist, W. I., The structural biology of HIV assembly. *Curr. Opin. Struc. Biol.* **2008**, *18* (2), 203-217.
6. Martin, N.; Sattentau, Q., Cell-to-cell HIV-1 spread and its implications for immune evasion. *Curr. Opin. HIV AIDS* **2009**, *4* (2), 143-9.

7. Tompa, D. R.; Immanuel, A.; Srikanth, S.; Kadirvel, S., Trends and strategies to combat viral infections: A review on FDA approved antiviral drugs. *Int. J. Biol. Macromol.* **2021**, *172*, 524-541.
8. Schaefer, R.; Thomas, R.; Nyamukapa, C.; Maswera, R.; Kadzura, N.; Gregson, S., Accuracy of HIV risk perception in East Zimbabwe 2003-2013. *AIDS Behav.* **2019**, *23* (8), 2199-2209.
9. Bavinton, B. R.; Grulich, A. E., HIV pre-exposure prophylaxis: scaling up for impact now and in the future. *Lancet Public Health* **2021**, *6* (7), e528-e533.
10. Coombs, R. W.; Krieger, J. N., Semen as the way forward to understand HIV-1 transmission. *J. Infect. Dis.* **2016**, *214* (10), 1473-1474.
11. Munch, J.; Rucker, E.; Standker, L.; Adermann, K.; Goffinet, C.; Schindler, M.; Wildum, S.; Chinnadurai, R.; Rajan, D.; Specht, A.; Gimenez-Gallego, G.; Sanchez, P. C.; Fowler, D. M.; Koulov, A.; Kelly, J. W.; Mothes, W.; Grivel, J. C.; Margolis, L.; Keppler, O. T.; Forssmann, W. G.; Kirchhoff, F., Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **2007**, *131* (6), 1059-71.
12. Muniyan, S.; Chaturvedi, N. K.; Dwyer, J. G.; Lagrange, C. A.; Chaney, W. G.; Lin, M. F., Human prostatic acid phosphatase: structure, function and regulation. *Int. J. Mol. Sci.* **2013**, *14* (5), 10438-64.
13. Lee, Y. H.; Ramamoorthy, A., Semen-derived amyloidogenic peptides - key players of HIV infection. *Protein Sci.* **2018**, *27* (7), 1151-1165.
14. Roan, N. R.; Munch, J.; Arhel, N.; Mothes, W.; Neidleman, J.; Kobayashi, A.; Smith-McCune, K.; Kirchhoff, F.; Greene, W. C., The cationic properties of SEVI

underlie its ability to enhance human immunodeficiency virus infection. *J. Virol.* **2009**, *83* (1), 73-80.

15. Arnold, F.; Schnell, J.; Zirafi, O.; Sturzel, C.; Meier, C.; Weil, T.; Standker, L.; Forssmann, W. G.; Roan, N. R.; Greene, W. C.; Kirchhoff, F.; Munch, J., Naturally occurring fragments from two distinct regions of the prostatic acid phosphatase form amyloidogenic enhancers of HIV infection. *J. Virol.* **2012**, *86* (2), 1244-9.

16. de Lamirande, E., Semenogelin, the main protein of the human semen coagulum, regulates sperm function. *Semin. Thromb. Hemost.* **2007**, *33* (1), 60-8.

17. Roan, N. R.; Muller, J. A.; Liu, H.; Chu, S.; Arnold, F.; Sturzel, C. M.; Walther, P.; Dong, M.; Witkowska, H. E.; Kirchhoff, F.; Munch, J.; Greene, W. C., Peptides released by physiological cleavage of semen coagulum proteins form amyloids that enhance HIV infection. *Cell Host Microbe* **2011**, *10* (6), 541-50.

18. Usmani, S. M.; Zirafi, O.; Muller, J. A.; Sandi-Monroy, N. L.; Yadav, J. K.; Meier, C.; Weil, T.; Roan, N. R.; Greene, W. C.; Walther, P.; Nilsson, K. P.; Hammarstrom, P.; Wetzel, R.; Pilcher, C. D.; Gagsteiger, F.; Fandrich, M.; Kirchhoff, F.; Munch, J., Direct visualization of HIV-enhancing endogenous amyloid fibrils in human semen. *Nat. Commun.* **2014**, *5*, 3508.

19. Tan, S.; Lu, L.; Li, L.; Liu, J.; Oksov, Y.; Lu, H.; Jiang, S.; Liu, S., Polyanionic candidate microbicides accelerate the formation of semen-derived amyloid fibrils to enhance HIV-1 infection. *PLoS One* **2013**, *8* (3), e59777.

20. van de Wijgert, J. H.; Shattock, R. J., Vaginal microbicides: moving ahead after an unexpected setback. *AIDS* **2007**, *21* (18), 2369-76.

21. Roan, N. R.; Sowinski, S.; Munch, J.; Kirchhoff, F.; Greene, W. C., Aminoquinoline surfen inhibits the action of SEVI (semen-derived enhancer of viral infection). *J. Biol. Chem.* **2010**, *285* (3), 1861-9.
22. Olsen, J. S.; Brown, C.; Capule, C. C.; Rubinshtein, M.; Doran, T. M.; Srivastava, R. K.; Feng, C.; Nilsson, B. L.; Yang, J.; Dewhurst, S., Amyloid-binding small molecules efficiently block SEVI (semen-derived enhancer of virus infection)- and semen-mediated enhancement of HIV-1 infection. *J. Biol. Chem.* **2010**, *285* (46), 35488-96.
23. Hauber, I.; Hohenberg, H.; Holstermann, B.; Hunstein, W.; Hauber, J., The main green tea polyphenol epigallocatechin-3-gallate counteracts semen-mediated enhancement of HIV infection. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (22), 9033-8.
24. Ren, R.; Yin, S.; Lai, B.; Ma, L.; Wen, J.; Zhang, X.; Lai, F.; Liu, S.; Li, L., Myricetin antagonizes semen-derived enhancer of viral infection (SEVI) formation and influences its infection-enhancing activity. *Retrovirology* **2018**, *15* (1), 49.
25. Xun, T.; Li, W.; Chen, J.; Yu, F.; Xu, W.; Wang, Q.; Yu, R.; Li, X.; Zhou, X.; Lu, L.; Jiang, S.; Li, L.; Tan, S.; Liu, S., ADS-J1 inhibits semen-derived amyloid fibril formation and blocks fibril-mediated enhancement of HIV-1 infection. *Antimicrob. Agents Chemother.* **2015**, *59* (9), 5123-34.
26. Qiu, M.; Li, Z.; Chen, Y.; Guo, J.; Xu, W.; Qi, T.; Qiu, Y.; Pang, J.; Li, L.; Liu, S.; Tan, S., Tolcapone potently inhibits seminal amyloid fibrils formation and blocks entry of ebola pseudoviruses. *Front. Microbiol.* **2020**, *11*, 504.
27. Sievers, S. A.; Karanicolas, J.; Chang, H. W.; Zhao, A.; Jiang, L.; Zirafi, O.; Stevens, J. T.; Munch, J.; Baker, D.; Eisenberg, D., Structure-based design of non-

- natural amino-acid inhibitors of amyloid fibril formation. *Nature* **2011**, 475 (7354), 96-100.
28. Malmsten, M., Antimicrobial peptides. *Ups. J. Med. Sci.* **2014**, 119 (2), 199-204.
29. Lei, J.; Sun, L.; Huang, S.; Zhu, C.; Li, P.; He, J.; Mackey, V.; Coy, D. H.; He, Q., The antimicrobial peptides and their potential clinical applications. *Am. J. Transl. Res.* **2019**, 11 (7), 3919-3931.
30. Mocz, G.; Ross, J. A., Fluorescence techniques in analysis of protein-ligand interactions. *Methods Mol. Biol.* **2013**, 1008, 169-210.
31. Gomes, A. V.; Barnes, J. A.; Vogel, H. J., Spectroscopic characterization of the interaction between calmodulin-dependent protein kinase I and calmodulin. *Arch. Biochem. Biophys.* **2000**, 379 (1), 28-36.
32. Niemann, A.; Takatsuki, A.; Elsasser, H. P., The lysosomotropic agent monodansylcadaverine also acts as a solvent polarity probe. *J. Histochem. Cytochem.* **2000**, 48 (2), 251-8.
33. Berezowska, I.; Lemieux, C.; Chung, N. N.; Zelent, B.; Schiller, P. W., Dansylated analogues of the opioid peptide [Dmt1]DALDA: in vitro activity profiles and fluorescence parameters. *Acta Biochim. Pol.* **2004**, 51 (1), 107-13.
34. Niemann, A.; Baltes, J.; Elsasser, H. P., Fluorescence properties and staining behavior of monodansylpentane, a structural homologue of the lysosomotropic agent monodansylcadaverine. *J. Histochem. Cytochem.* **2001**, 49 (2), 177-85.
35. Sha, C.; Chen, Y.; Chen, Y.; Xu, D., An easily prepared fluorescent pH probe based on dansyl. *J. Fluoresc.* **2016**, 26 (5), 1709-14.

36. Chen, J. L.; Steele, T. W. J.; Stuckey, D. C., Metabolic reduction of resazurin; location within the cell for cytotoxicity assays. *Biotechnol. Bioeng.* **2018**, *115* (2), 351-358.
37. Ozaki, D. A.; Gao, H.; Todd, C. A.; Greene, K. M.; Montefiori, D. C.; Sarzotti-Kelsoe, M., International technology transfer of a GCLP-compliant HIV-1 neutralizing antibody assay for human clinical trials. *PLoS One* **2012**, *7* (1), e30963.
38. Bazan, J.; Calkosinski, I.; Gamian, A., Phage display – a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Hum. Vaccin. Immunother.* **2012**, *8* (12), 1817-28.
39. Wu, C. H.; Liu, I. J.; Lu, R. M.; Wu, H. C., Advancement and applications of peptide phage display technology in biomedical science. *J. Biomed. Sci.* **2016**, *23*, 8.
40. Acevedo-Rocha, C. G.; Reetz, M. T.; Nov, Y., Economical analysis of saturation mutagenesis experiments. *Sci. Rep.* **2015**, *5*, 10654.
41. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14* (6), 1188-90.

Appendix I

Custom DYNAFIT script for the analysis of Mag-Fura 2 vs protein competition titration.

This script was used to fit the competition of Mag-Fura 2 and 4G-UFsc (or mutants) for a zinc metal ion that forms a 1:1 complex with the protein monomer. Specifically, Mag-Fura 2 dye (L) was mixed with protein (P) and metal (M) was titrated into the solution; formation of the Zn-Mag-Fura 2 (LM) complex resulted in the decrease in the dye absorbance signal at 366 nm.

```
=====
=====
```

```
[task]
task = fit
data = equilibria
; L ... Mag-Fura-2
; P ... protein
; M ... Zn - competing ligand
[mechanism]
L + M <==> LM : Kd1 dissociation in M
P + M <==> PM : Kd2 dissociation in M
[constants]
Kd1 = 20e-9
Kd2 = 1e-10 ?
[concentrations]
L = (concentration of the dye determined for each reaction in M)
P = (concentration of the protein determined for each reaction in M)
[responses]
LM = 3600 ?
L = 29900 ?
[data]
directory ./data/
extension txt
variable M
file name.txt
[output]
directory ./output/fit-name
```

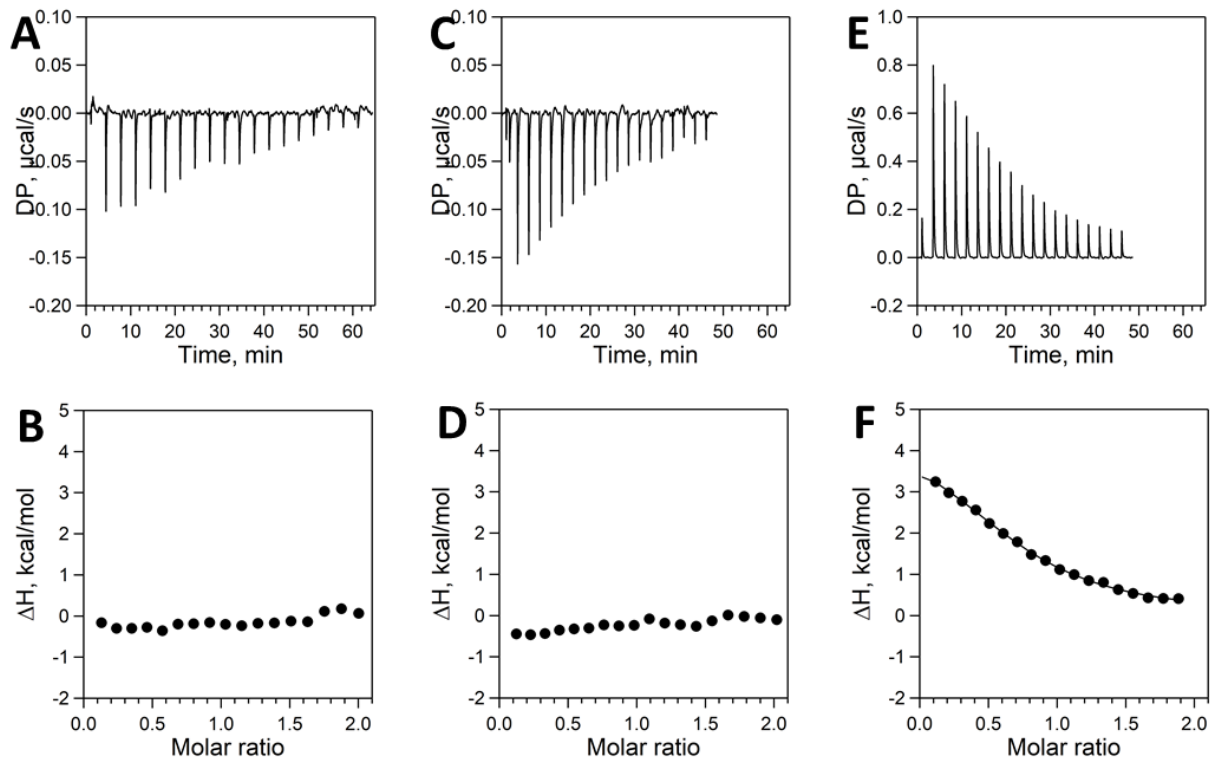


Figure S 1. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of E74 mutants with manganese (II) chloride. Solution of 1.5 mM MnCl_2 was titrated into: A – 132 μM 4G-UFsc E74D (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); C – 139 4G-UFsc E74Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); E – 149 μM 4G-UFsc E74H (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively

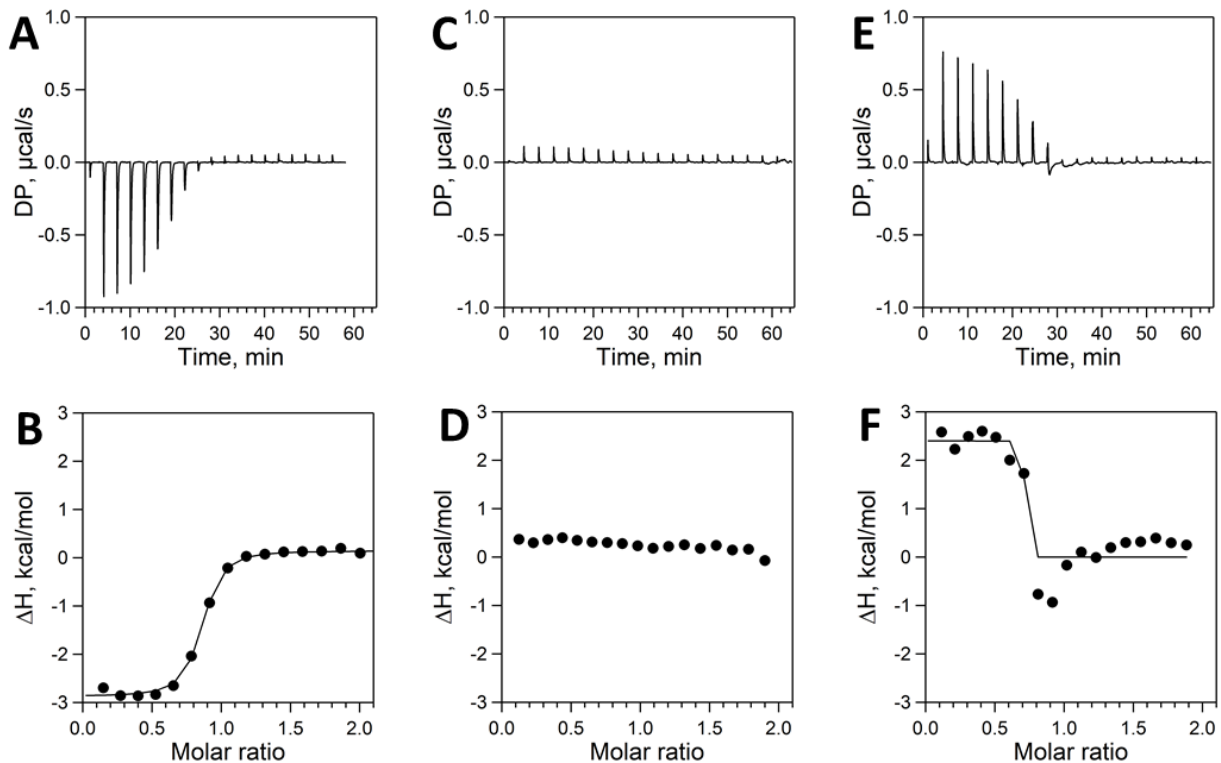


Figure S 2. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of E74 mutants with cobalt (II) chloride. A – 2 mM CoCl₂ was titrated into 154 µM 4G-UFsc E74D (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 180 s); C – 1.5 mM CoCl₂ was titrated into 139 µM 4G-UFsc E74Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s); E – 1.5 mM CoCl₂ was titrated into 149 µM 4G-UFsc E74H (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively

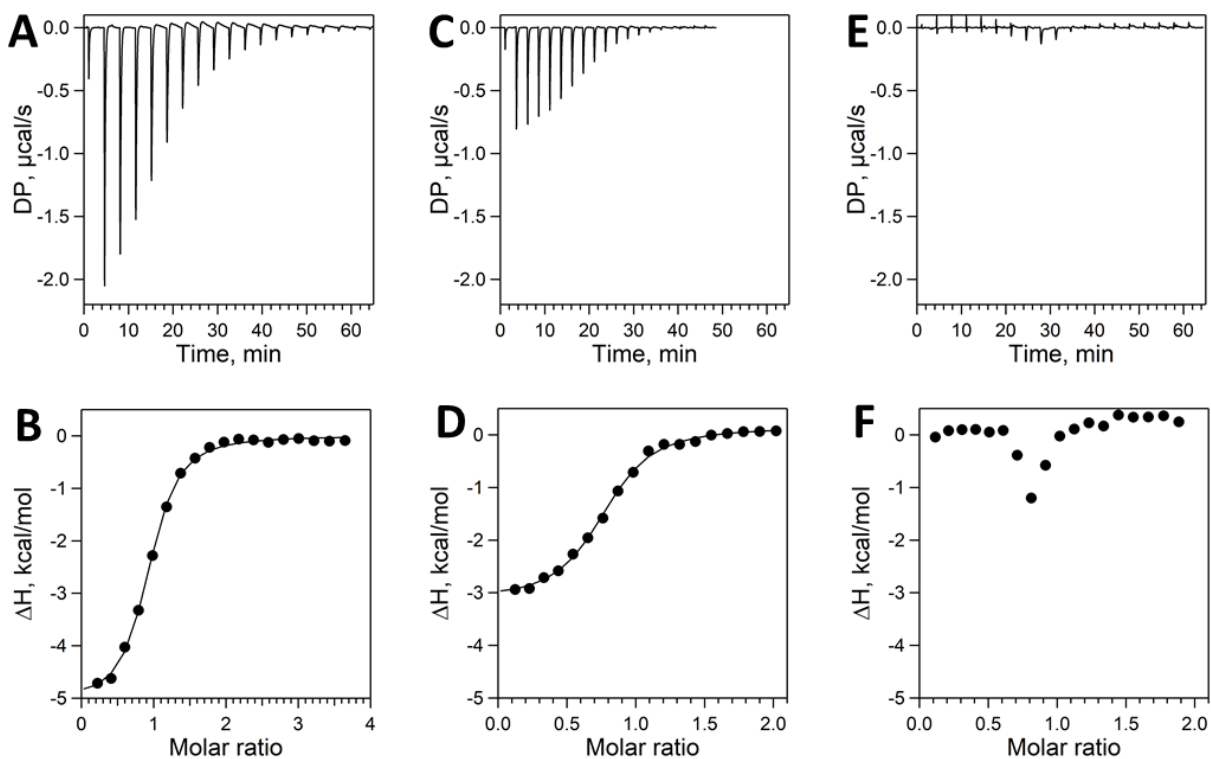


Figure S 3. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of E74 mutants with cobalt (II) chloride. A – 3 mM NiCl₂ was titrated into 154 μM 4G-UFsc E74D (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); C – 1.5 mM NiCl₂ was titrated into 139 μM 4G-UFsc E74Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); E – 1.5 mM NiCl₂ was titrated into 149 μM 4G-UFsc E74H (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s). B, D, F – binding isotherms derived from the integration of A, C and E, respectively

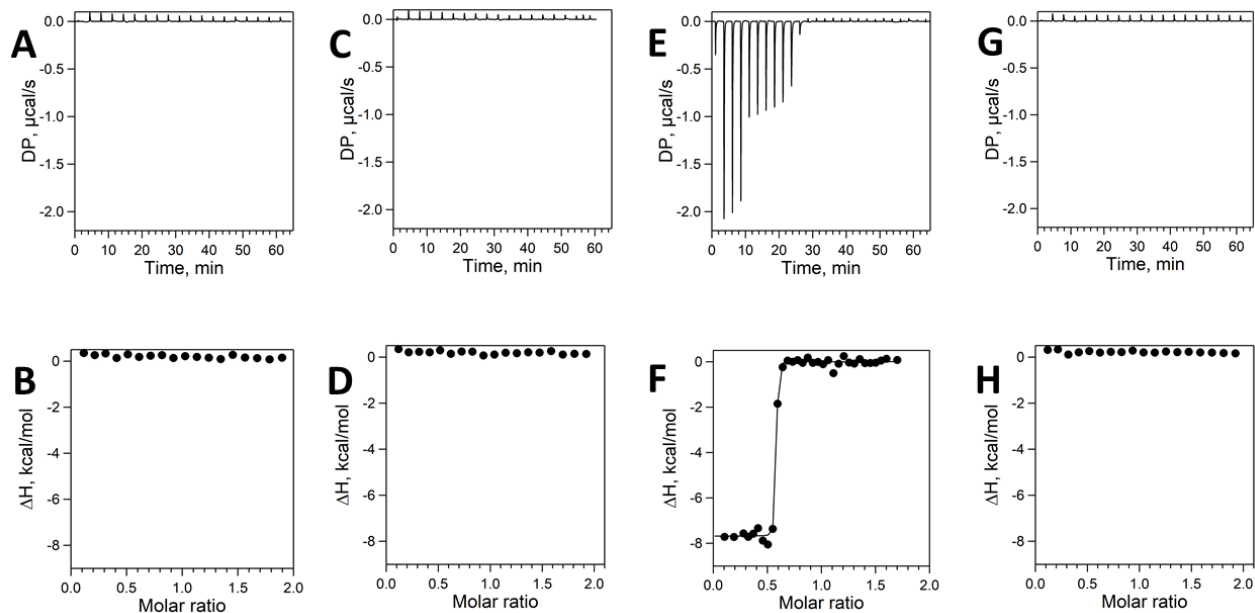


Figure S 4. Baseline- and offset-corrected ITC titration curves (A, C, E, G) and binding isotherms (B, D, F, H) for the titration of H107Q and H77Q mutants with manganese (II) (A-D) and cobalt (II) (E-H) chloride. A – 1.5 mM MnCl₂ was titrated into 148 µM 4G-UFsc H107Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s); C – 1.5 mM MnCl₂ was titrated into 146 µM 4G-UFsc H77Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s); E – 1.3 mM CoCl₂ was titrated into 143 µM 4G-UFsc H107Q (first injection of 0.4 µL with further 3 injections of 2 µL, 28 injections of 1 µL and 2 injections of 2 µL with the equilibration time of 150 s); G – 1.5 mM CoCl₂ was titrated into 146 µM 4G-UFsc H77Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F, H – binding isotherms derived from the integration of A, C, E and G, respectively.

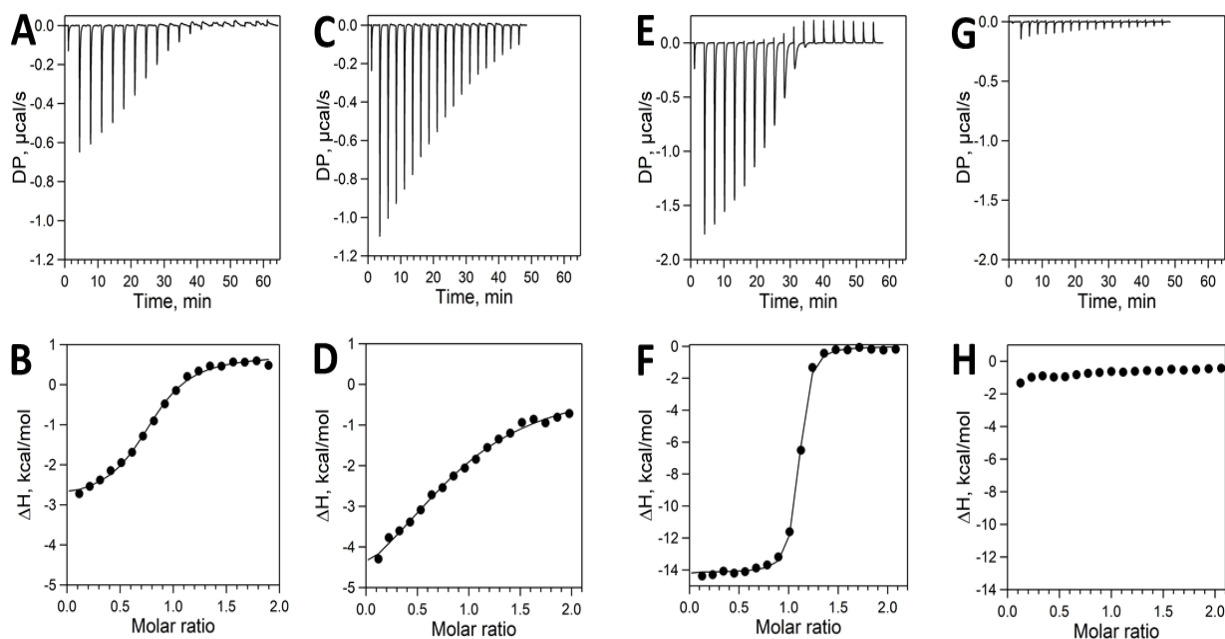


Figure S 5. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of H107Q and H77Q mutants with nickel (II) chloride (A-D) and competition titration with zinc-TETA complex (E-H). A – NiCl₂ (1.5 mM) was titrated into 4G-UFsc H107Q (148 µM, first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s); C – NiCl₂ (1.5 mM) was titrated into 4G-UFsc H77Q (142 µM, first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s); E – Zn(II)-TETA complex (90 µM Zn(II) and 200 µM TETA) was titrated with 4G-UFsc H107Q (1 mM, first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 180 s); G – Zn(II)-TETA complex (90 µM Zn(II) and 200 µM TETA) was titrated with 4G-UFsc H77Q (0.99 mM, first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s) in 25 mM HEPES, 100 mM NaCl, pH 7.6. B, D, F, H – binding isotherms derived from the integration of A, C, E and G, respectively.

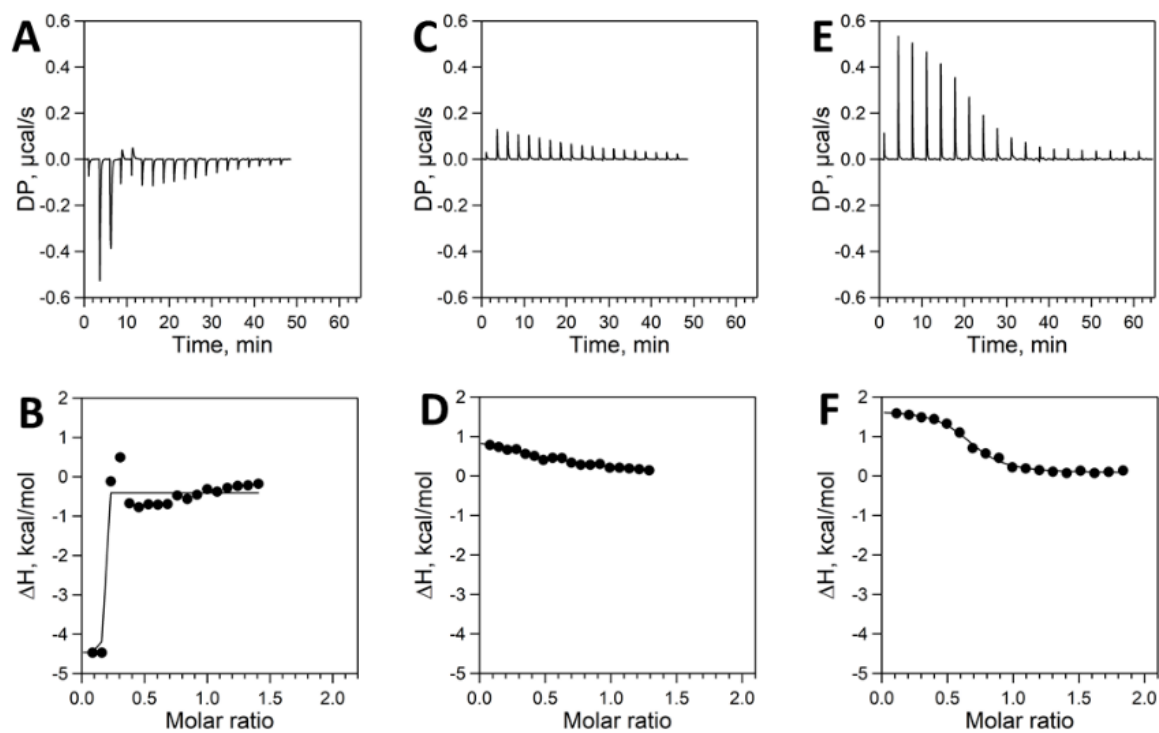


Figure S 6. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of E44 and E11 4G-UFsc mutants with manganese (II) chloride. A – 1 mM MnCl₂ was titrated into 133 µM 4G-UFsc E44Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s); C – 1 mM MnCl₂ was titrated into 133 4G-UFsc E44L Y18F (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s); E – 1.5 mM MnCl₂ was titrated into 153 µM 4G-UFsc E11Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively.

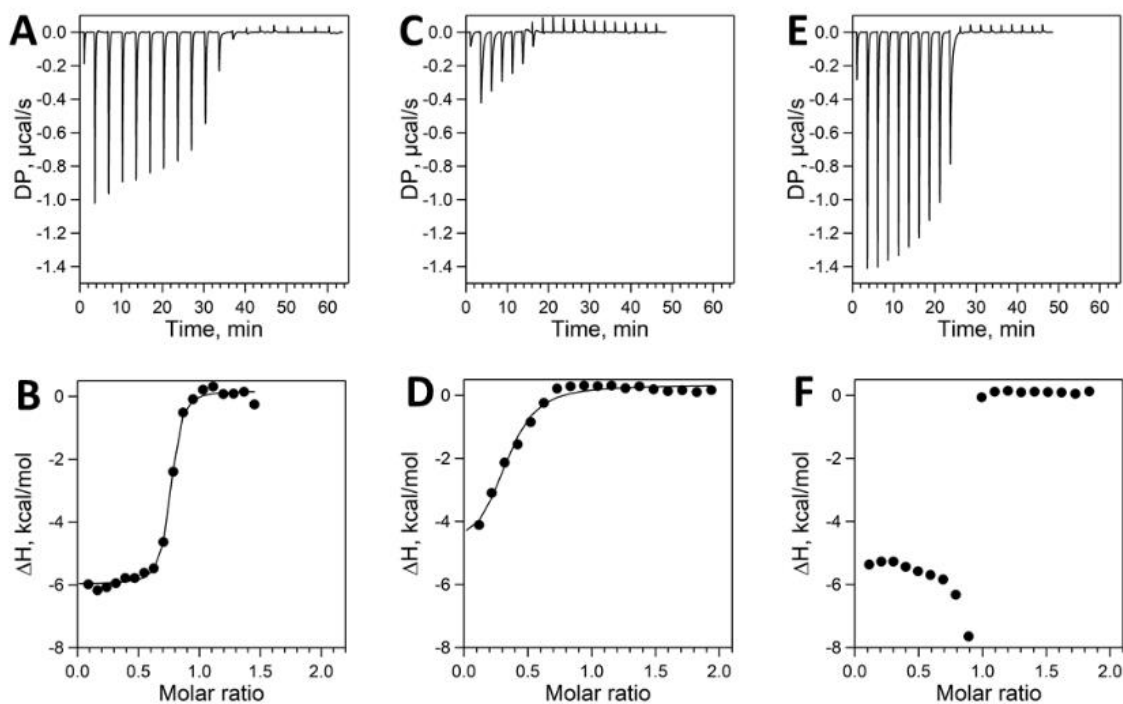


Figure S 7. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of E44 and E11 4G-UFsc mutants with cobalt (II) chloride. A – 1 mM CoCl₂ was titrated into 129 μM 4G-UFsc E44Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); C – 1.5 mM CoCl₂ was titrated into 145 4G-UFsc E44L Y18F (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 150 s); E – 1.5 mM CoCl₂ was titrated into 153 μM 4G-UFsc E11Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively.

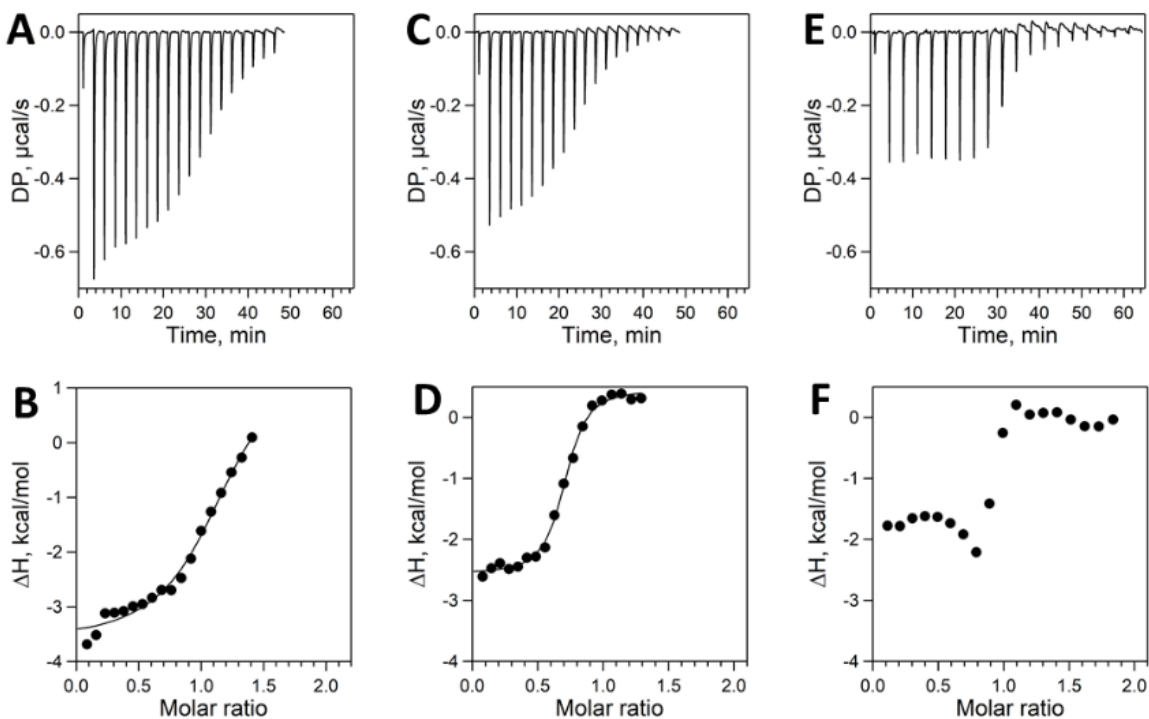


Figure S 8. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the titration of E44 and E11 4G-UFsc mutants with nickel (II) chloride. A – 1 mM NiCl₂ was titrated into 133 µM 4G-UFsc E44Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s); C – 1 mM NiCl₂ was titrated into 145 4G-UFsc E44L Y18F (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s); E – 1.5 mM NiCl₂ was titrated into 153 µM 4G-UFsc E11Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively.

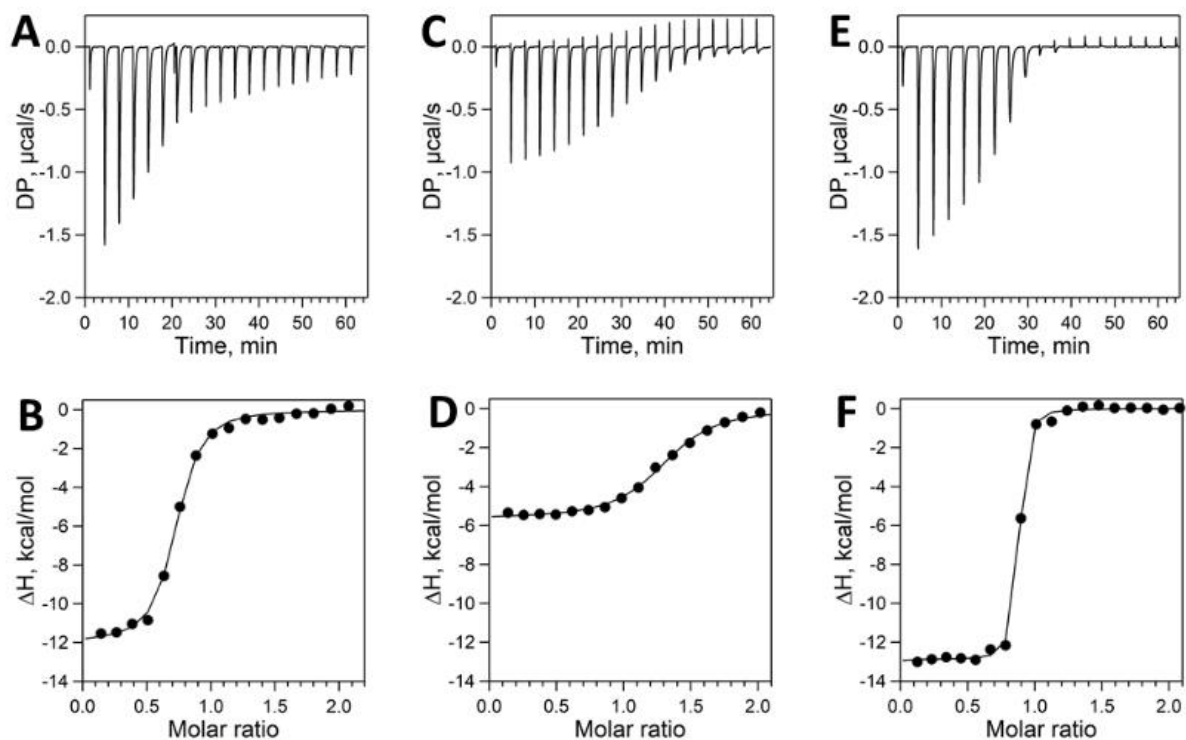


Figure S 9. Baseline- and offset-corrected ITC titration curves (A, C, E) and binding isotherms (B, D, F) for the competition titration of zinc (II) – TETA complex with 4G-UFsc mutants. The complex of 90 μM zinc (II) with 200 μM TETA was titrated with: A – 1.13 mM 4G-UFsc E44Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); C – 1.1 mM 4G-UFsc E44L Y18F (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); E – 1 mM 4G-UFsc E11Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F – binding isotherms derived from the integration of A, C and E, respectively.

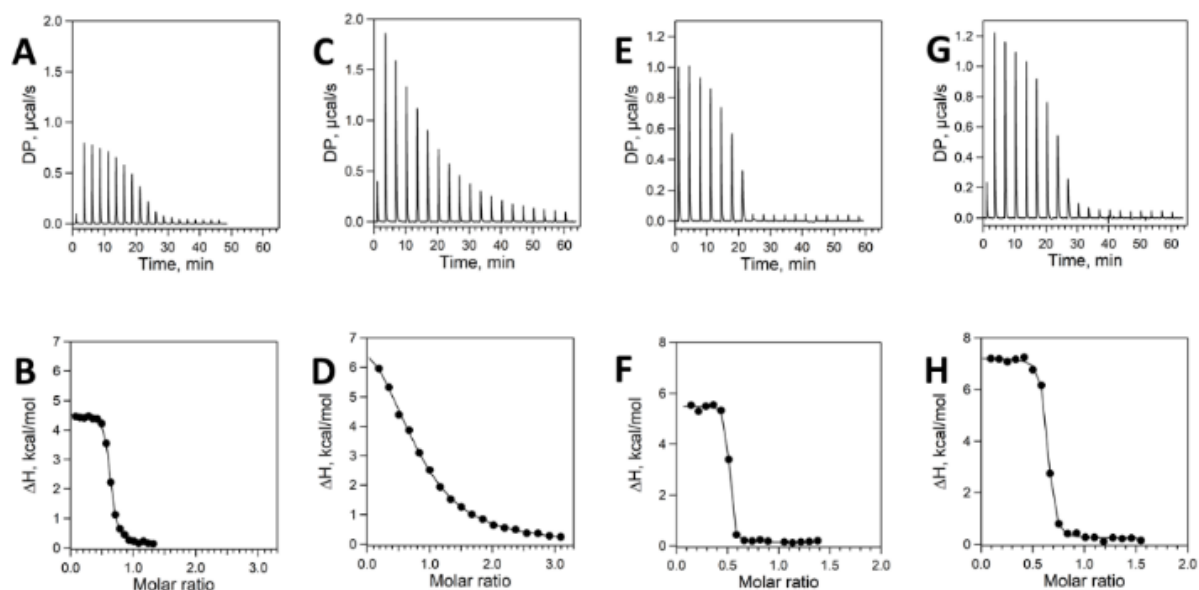


Figure S 10. Baseline- and offset-corrected ITC titration curves (A, C, E, G) and binding isotherms (B, D, F, H) for the titration of UFsc E44Q and UFsc E44L Y18F mutants with manganese (II) (A-D) and cobalt (II) (E-H) chloride. A – 1 mM MnCl_2 was titrated into 142 μM UFsc E44Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 150 s); C – 2 mM MnCl_2 was titrated into 121 μM UFsc E44L Y18F (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); E – 1 mM CoCl_2 was titrated into 134 μM UFsc E44Q (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s); G – 1 mM CoCl_2 was titrated into 121 μM UFsc E44L Y18F (first injection of 0.4 μL with further 18 injections of 2 μL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F, H – binding isotherms derived from the integration of A, C, E and G, respectively.

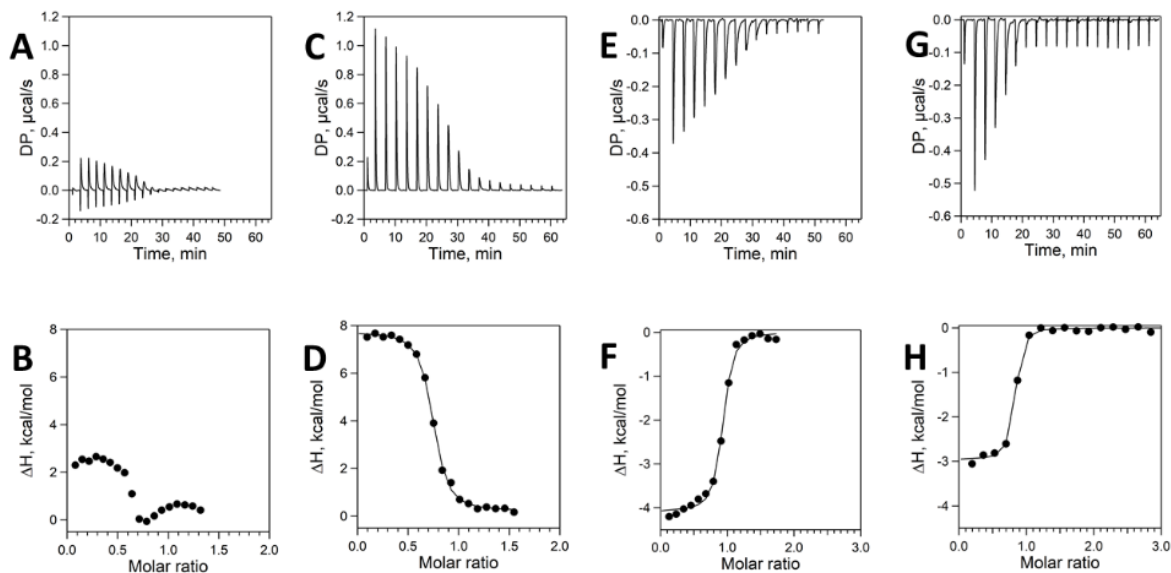


Figure S 11. Baseline- and offset-corrected ITC titration curves (A, C, E, G) and binding isotherms (B, D, F, H) for the titration of UFsc E44Q and UFsc E44L Y18F mutants with nickel (II) chloride (A-D) and competition titration of zinc – TETA complex with protein. A – 1 mM NiCl₂ was titrated into 142 µM UFsc E44Q (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 150 s); C – 1 mM NiCl₂ was titrated into 121 µM UFsc E44L Y18F (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s); E – 1.01 mM UFsc E44Q was titrated into 90 µM zinc – 200 µM TETA solution (first injection of 0.4 µL with further 3 injections of 2 µL, 28 injections of 1 µL and 2 injections of 2 µL with the equilibration time of 200 s); G – 1.55 mM UFsc E44L Y18F was titrated into 90 µM zinc (II) – 200 µM TETA solution (first injection of 0.4 µL with further 18 injections of 2 µL with the equilibration time of 200 s) in 25mM HEPES, 100 mM NaCl (pH 7.6). B, D, F, H – binding isotherms derived from the integration of A, C, E and G, respectively.

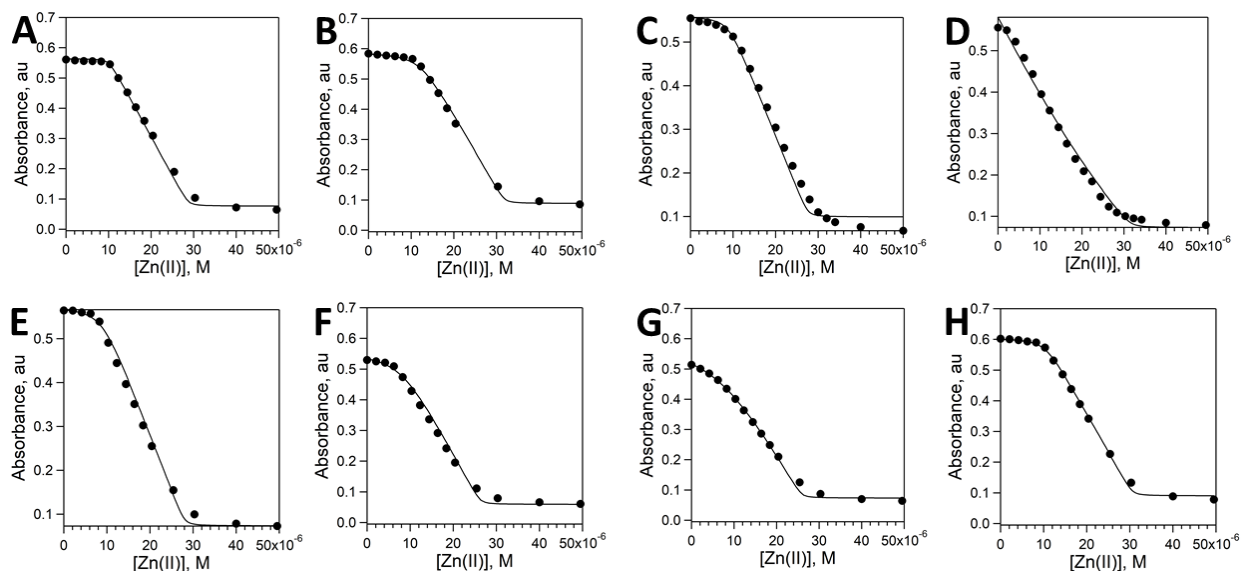
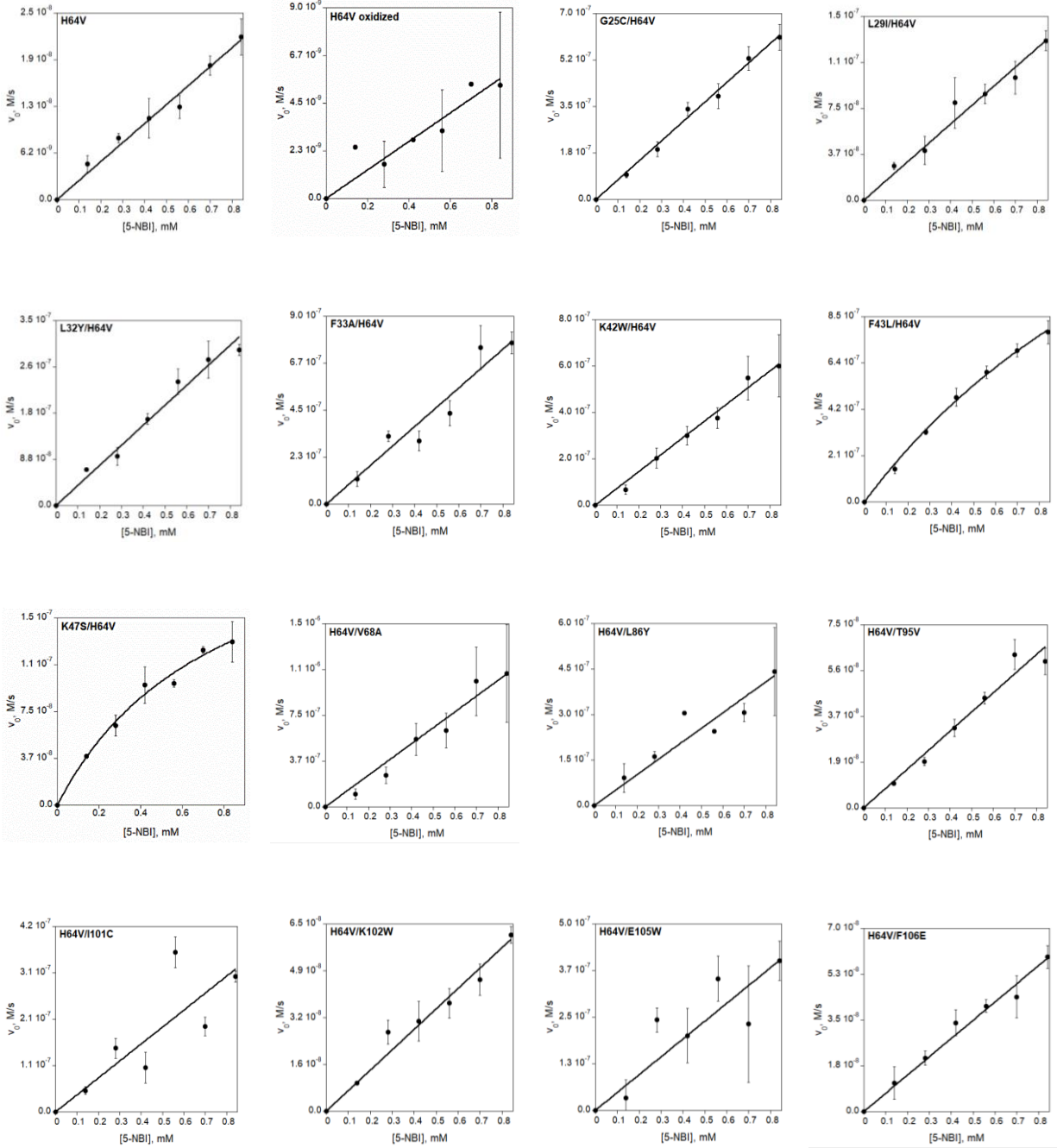


Figure S 12. Fit of the absorbance intensity at 366 nm of Mag-Fura 2 upon titration with Zn(II). Small aliquots of 1 mM ZnCl₂ were titrated into: A – 10.08 μM 4G-UFsc with 18.72 μM Mag-Fura 2 (representative titration out of three repeats); B – 12.36 μM 4G-UFsc H107Q with 18.98 μM Mag-Fura 2 (representative titration out of three repeats); C – 9.00 μM 4G-UFsc E44Q with 18.60 μM Mag-Fura 2; D – 12.70 μM 4G-UFsc E44L/Y18F with 19.00 μM Mag-Fura 2 (representative titration out of two repeats); E – 9.17 μM 4G-UFsc E11Q with 18.90 μM Mag-Fura 2 (representative titration out of two repeats); F – 8.57 μM 4G-UFsc E74D with 17.73 μM Mag-Fura 2 (representative titration out of three repeats); G – 9.90 μM 4G-UFsc E74Q with 17.78 μM Mag-Fura 2; H – 11.33 μM 4G-UFsc E74H with 19.66 μM Mag-Fura 2 (representative titration out of three repeats). All titrations were performed in buffer (25 mM HEPES, 100 mM NaCl, pH 7.6) at room temperature.

Appendix II



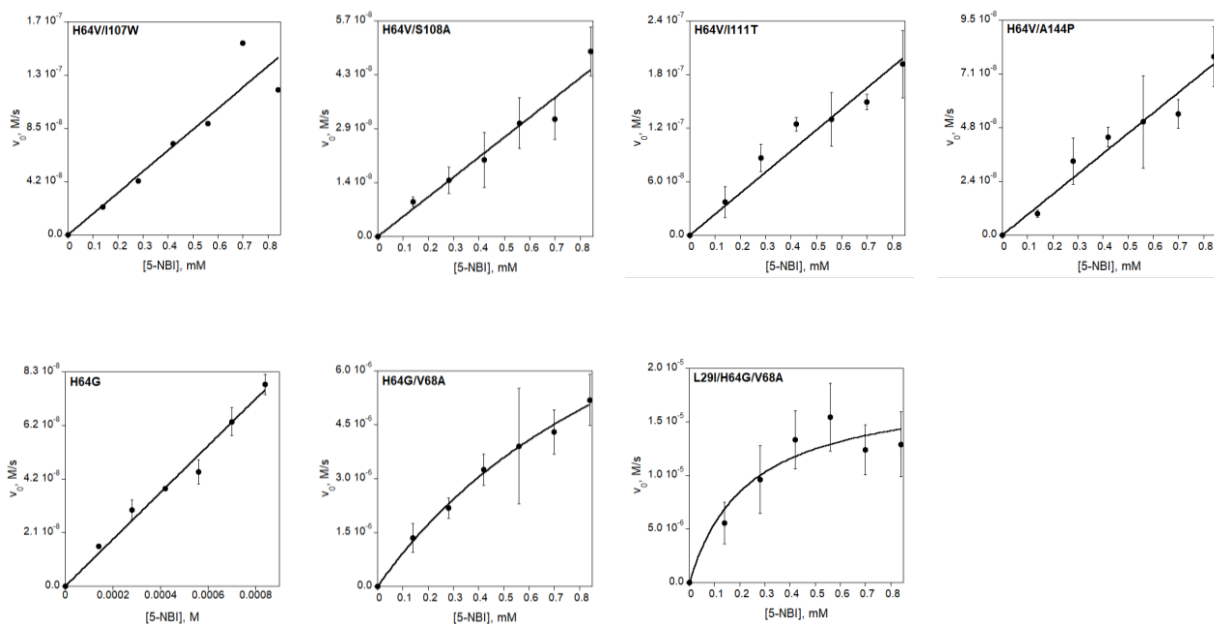


Figure S 13. Michaelis-Menten plots of Kemp elimination catalyzed by reduced (unless otherwise stated) myoglobin mutants. Final reaction mixtures for myoglobin mutant analyses contained 1 mM L-ascorbic acid, 0.1 μ M SOD, 20 nM catalase, 140-840 μ M substrate, 1.5% acetonitrile in 20 mM Tris (pH 8.0). The protein concentration was 1 μ M for Mb-H64V, 0.1 μ M for Mb-H64V-based double variants, 5 nM for Mb-H64G or Mb-H64G-based double or triple variants. Kinetic parameters are summarized in Table 3.1.

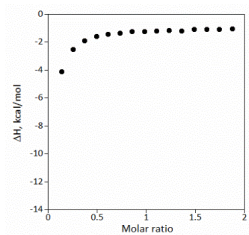
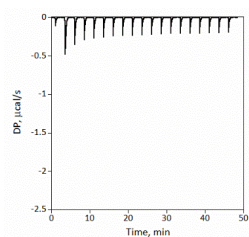
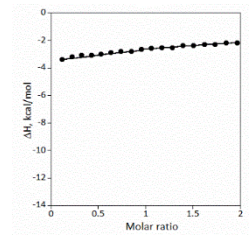
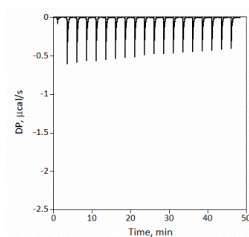
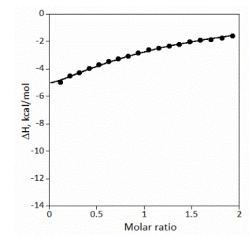
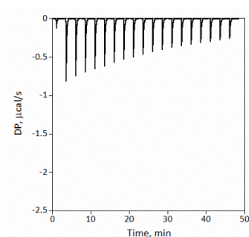
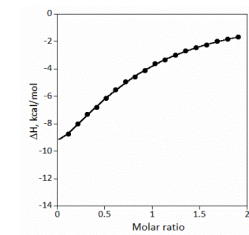
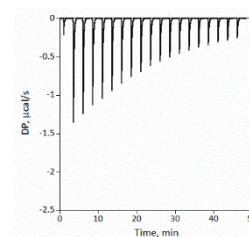
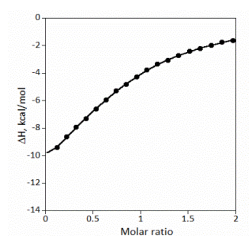
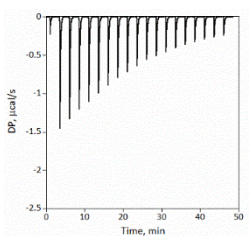
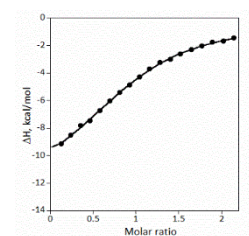
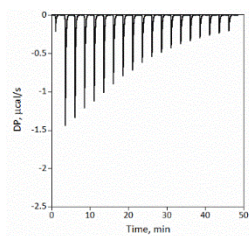
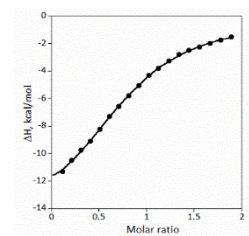
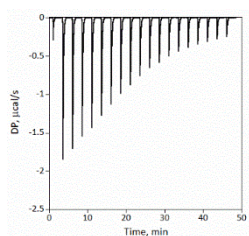
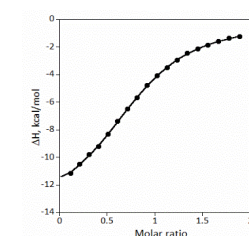
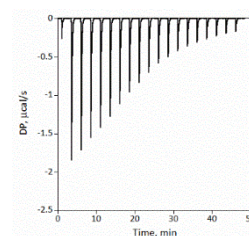
Table S 1. Protein sequences of AlleyCat mutants.

Proteins	Mutations introduced	Sequences
AlleyCat	F92E	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RHVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMM TAK*
AlleyCat1	F92E, M144R	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RHVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMR TAK*
AlleyCat2	F92E, H107I, M144R	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMR TAK*
AlleyCat3	F92E, H107I, L112R, M144R	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMR TAK*
AlleyCat4	I85L, F92E, H107I, L112R, M144R	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMR TAK*
AlleyCat5	I85L, F92E, H107I, L112R, A128T, M144R	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDEMI RETDIDGDGQ VNYEEFVQMR TAK*
AlleyCat6	I85L, F92E, H107I, L112R, M124L, A128T, M144R	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQMR TAK*
AlleyCat7	I85L, A88Q, F92E, H107I, L112R, M124L, A128T, M144R	MKDTDSEEEI REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK*
AlleyCat8	I85L, A88Q, F92E, H107I, L112R, M124L, I125H, A128T, M144R	MKDTDSEEEI REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELH RETDIDGDGQ VNYEEFVQRM TAK*
AlleyCat8-T146R	I85L, A88Q, F92E, H107I, L112R, M124L, I125H, A128T, M144R, T146R	MKDTDSEEEI REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELH RETDIDGDGQ VNYEEFVQRM RAK*
AlleyCat9	I85L, A88Q, F92E, H107I, L112R, K115P, M124L, I125H, A128T, M144R	MKDTDSEEEI REQFRVEDKD GNGYISAAEL RIVMTNRGEP LTDEEVDELH RETDIDGDGQ VNYEEFVQRM TAK*
AlleyCat10	I85L, A88Q, F92E, H107I, L112R, K115P, M124L, I125H, A128T, M144R, T146R	MKDTDSEEEI REQFRVEDKD GNGYISAAEL RIVMTNRGEP LTDEEVDELH RETDIDGDGQ VNYEEFVQRM RAK*

Table S 2. Kinetic parameters for the Kemp elimination reaction catalyzed by AlleyCat mutants.

Protein	k_{cat} , s^{-1}	K_M , mM	k_{cat}/K_M , $M^{-1}s^{-1}$
AlleyCat	-	-	5.8 ± 0.3 ^{a,b}
AlleyCat7	3.2 ± 0.2 ^c	2.4 ± 0.2 ^c	1283 ± 13 ^c
AlleyCat8	10.1 ± 1.5	4.1 ± 0.7	2451 ± 15 2369 ± 99 ^c
AlleyCat8-T146R	5.8 ± 0.7	1.6 ± 0.3	3563 ± 39
AlleyCat9	18.9 ± 3.9	4.9 ± 1.2	3857 ± 27 3894 ± 61 ^c
AlleyCat10	21.2 ± 2.8	4.8 ± 0.7	4378 ± 20 4392 ± 83 ^c

^a Individual k_{cat} and K_M values could not be determined due to substrate solubility. ^b from Korendovych, I. V. et al. *Proceedings of the National Academy of Sciences* 108.17, 6823-6827 (2011). ^c $(k_{cat}/K_M)_{max}$ values were obtained from the pH activity profiles.

AlleyCat**AlleyCat 1****AlleyCat 2****AlleyCat 3****AlleyCat 4****AlleyCat 5****AlleyCat 6****AlleyCat 7**

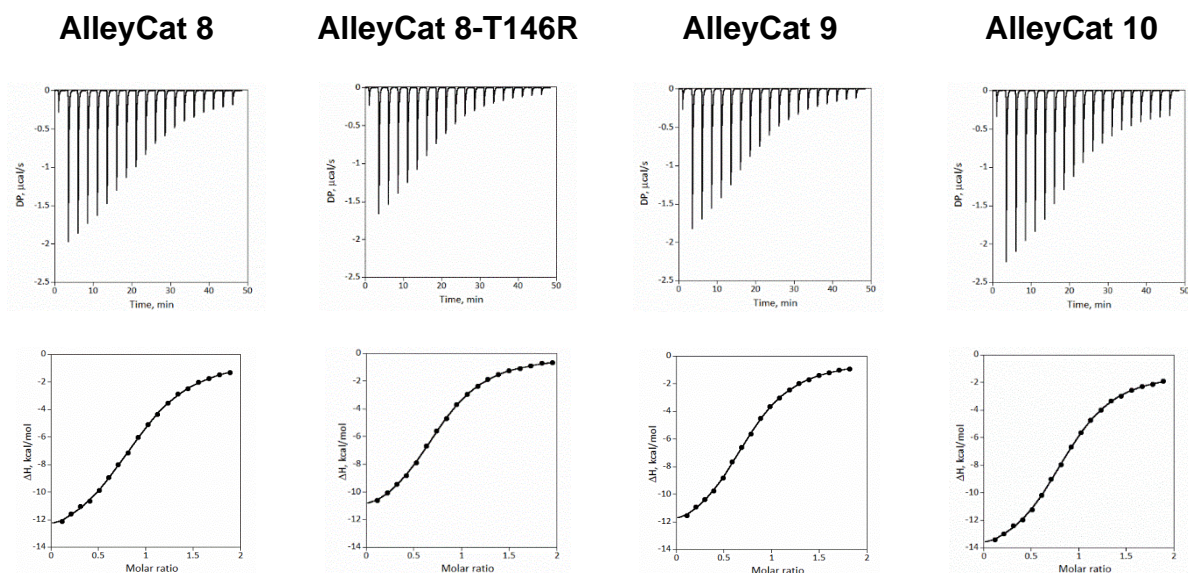


Figure S 14. ITC titration curves (top panels) and binding isotherms (bottom panels) for the titration of 1 mM 5-nitrobenzotriazole in 82.2 μM AlleyCat, 95.4 μM AlleyCat 1, 96.9 μM AlleyCat 2, 98.2 μM AlleyCat 3, 954.9 μM AlleyCat 4, 87.3 μM AlleyCat 5, 99.2 μM AlleyCat 6, 99.1 μM AlleyCat 7, 99.2 μM AlleyCat 8, 95.7 μM AlleyCat 8-T146R, 103 μM AlleyCat 9, and 99.2 μM AlleyCat 10 in 20 mM HEPES, 100 mM NaCl, 10 mM CaCl_2 (pH 7.0), 2% acetonitrile at 25°C. The baseline and offset-corrected titration curves were analyzed in MicroCal PEAQ-ITC Analysis software provided by Malvern Panalytical and the thermodynamic parameters were summarized in the Table 3.2.

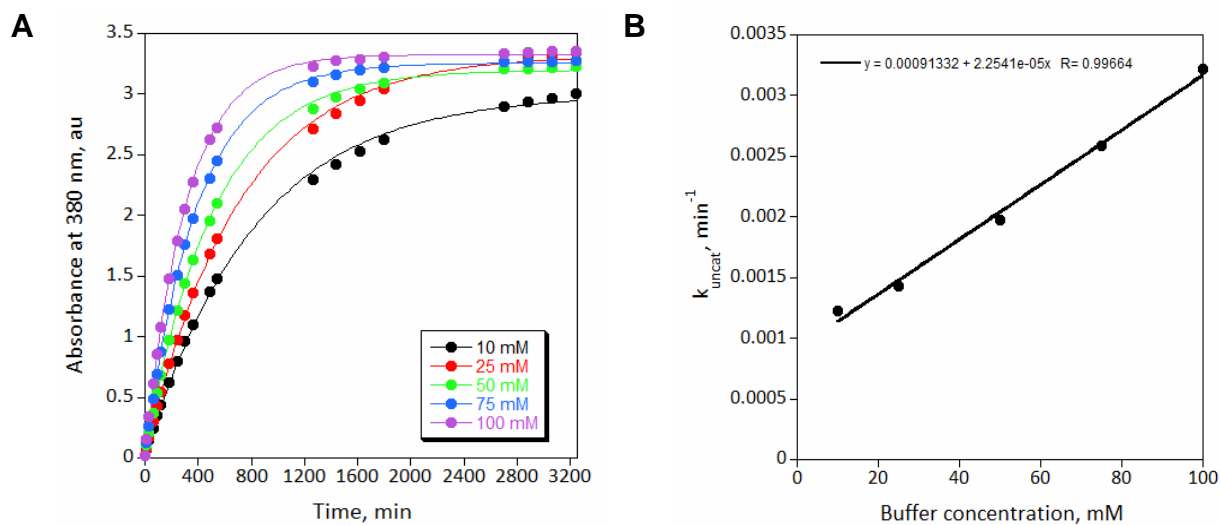


Figure S 15. Determination of k_{uncat} at no buffer concentration from k_{uncat} values at various concentrations of Tris-HCl buffer at pH 8.0 (25°C). The obtained k_{uncat} is $(1.52 \pm 0.11) \times 10^{-5} \text{ s}^{-1}$ at pH 8.0. **A.** The kinetic spectra at various buffer concentrations. **B.** The extrapolation of linear trendline to the zero buffer concentration to obtain k_{uncat} at zero buffer concentration.

Appendix III

Table S 3. The IUPAC degenerate base codes used to create the phage display peptide library.

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

Curriculum Vitae

Alona Kulesha

Syracuse, NY 13210 | akulesha@syr.edu |
linkedin.com/in/alona-kulesha/

EDUCATION

Syracuse University, Syracuse, NY

Ph.D. in Chemistry

August 2022

Dissertation: Design of novel metalloenzymes and investigation of protein-metal interactions

Advisor: Dr. Ivan Korendovych

Syracuse University, Syracuse, NY

M.Phil. in Chemistry

August 2019

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

M.S. in Biochemistry (specialization in Molecular Biology)

June 2017

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

B.S. in Biology (specialization in Biochemistry)

June 2015

ACADEMIC RESEARCH EXPERIENCE

Graduate Research and Teaching Assistant, Syracuse University, Syracuse, NY

December 2017 – Present, Syracuse, NY

- Investigated the contributions of coordination ligands to the extraordinary metal affinity in the de novo designed protein using mutagenesis and isothermal titration calorimetry (ITC).
- Analyzed properties of protein catalysts designed during the validation of NMR-guided directed evolution, the novel method for evolution of proteins to gain specific activities.
- Expanded the library of peptides rationally designed to block the association of HIV virus with human cells
- Worked on the development of the radiotracer for PET imaging using a de novo designed metalloprotein
- Studied the activity of rationally designed antimicrobial peptides

Visiting Undergraduate Research Assistant, French National Center for Scientific Research, Gif-sur-Yvette, France

July 2014 – September 2014

Research Advisor: Dr. Marc Mirande

Research project: Identification of protein-protein interactions between NPRL2 and components of MARS complex

- Mastered techniques in mammalian cell culture and analysis of protein-protein interactions by Flag-tagged immunochromatography and Western Blot

Visiting Undergraduate Research Assistant, French National Center for Scientific Research, Gif-sur-Yvette, France

September 2013 – December 2013

Research Advisor: Dr. Marc Mirande

Research project: Investigation of interactions between the components of multienzyme complex MARS subcomplex I

- Gained hands-on experience using baculovirus expression vector system

Undergraduate Research Assistant, Institute of Molecular Biology and Genetics of National Academy of Sciences of Ukraine, Kyiv, Ukraine

March 2013 – June 2015

Research Advisor: Dr. Vyacheslav Shalak

Research project: Investigation of the protein interactions in eEF1B complex

- Learned and practiced molecular biology techniques
- Learned to organize research data and lab notebook
- Assisted with preparation of truncated forms of eEF1B beta subunit
- Performed molecular subcloning of human valyl-tRNA-synthetase ORF for baculovirus expression system

INDUSTRIAL RESEARCH EXPERIENCE

Research Intern, Bienta, Enamine Ltd., Kyiv, Ukraine

June 2015 – August 2017

Research Advisors: Dr. Yurii Moroz and Dr. Sergey Zozulya

Research project: Bacterial expression of disulfide-rich peptide Mambalgin-1, a potent blocker of acid-sensing ion channel 1 (ASIC 1)

- Developed a protocol for expression and purification of disulfide-rich therapeutic peptide in bacterial protein expression system
- Managed large-scale protein purification for high-throughput screening of small molecules
- Established protocols and performed activity screening of small molecules (antibacterial test on *Staphylococcus aureus* and *Escherichia coli*; cytotoxicity test on mammalian cell cultures, in vitro bacterial reverse mutation (Ames) test)
- Prepared standard operating procedures (SOP) for new assays
- Prepared written reports for internal and external use

TEACHING EXPERIENCE

Teaching Assistant, Syracuse University, Syracuse, NY

September 2017 – present

- Facilitated recitations for the introductory chemistry course (General Chemistry Lecture), prepared teaching materials, assisted students in the learning process
- Supervised undergraduate lab sections (General Chemistry Laboratory I and II) and facilitated group discussions
- Designed experiments for the undergraduate research lab, educated students regarding good biochemistry lab practices and safety (General Chemistry Laboratory for Honors)
- Facilitated recitations for the graduate-level biochemistry courses (Physical and Structural Biochemistry, Biochemistry II), prepared teaching materials, led discussions
- Provided feedback regarding class assignments, evaluated students' performance and progress, graded reports, assignments and exams for different courses, advised students on the individual basis
- Mentored and supervised graduate and undergraduate students involved in research

SKILLS

Molecular Biology: primer design, PCR, cloning, mutagenesis, sequence analysis, codon optimization, DNA preparation, gel electrophoresis

Cell culture: bacterial cell culture (*E. coli*, *S. aureus*), yeast cell culture (*P. pastoris*), insect cell culture (Sf9, Sf21, High-Five cells), mammalian cell culture (adherent and suspension cultures)

Protein and peptide techniques: protein and peptide expression in bacterial expression system, baculovirus expression vector system, mammalian expression system; phage display; affinity chromatography, IEC, HIC, SEC and RP-HPLC; protein analysis with UV-Vis spectroscopy, SDS-PAGE, Western-Blot, isothermal titration calorimetry (ITC), circular dichroism (CD); F-moc solid-phase peptide synthesis and RP-HPLC purification; MALDI-TOF; Michaelis-Menten kinetics, stopped-flow spectrometry; working in anaerobic conditions (Schlenk line, glove box)

Programs: Microsoft Office package, SnapGene, PyMol, GraphPad Prism, ChemDraw

PUBLICATIONS

- Bhattacharya, S., Margheritis, E.*, Takahashi, K.*, **Kulesha, A.***, D'Souza, A., Kim, I., Tame, J., Yoon, J., Volkov, A., Makhlynets, O., and Korendovych, I. (2022). NMR-guided directed evolution. (Under revision).
- D'Souza, A. R., Necelis, M. R., **Kulesha, A.**, Caputo, G. A., and Makhlynets, O. V. (2021). Beneficial impacts of incorporating the non-natural amino acid azulenyl-alanine into the trp-rich antimicrobial peptide buCATHL4B. *Biomolecules*, 11(3), 421.
- **Kulesha, A.***, Yoon, J. H., Chester, C., D'Souza, A., Costeas, C., and Makhlynets, O. V. (2020). Contributions of primary coordination ligands and importance of outer sphere interactions in UFsc, a *de novo* designed protein with high affinity for metal ions. *J. Inorg. Biochem.*, 212, 111224.
- Yoon, J. H.*, **Kulesha, A. V.***, Lengyel-Zhand, Z., Volkov, A. N., Rempillo, J. J., D'Souza, A., Costeas, C., Chester, C., Caselle, E. R., and Makhlynets, O. V. (2019) Uno

Ferro, a *de novo* designed protein, binds transition metals with high affinity and stabilizes semiquinone radical anion. *Chemistry*, 25 (67), 15252-15256.

CONFERENCE PRESENTATIONS

Design of peptides to block association of HIV virus with human cells

Poster presentation at ACS National Meeting and Exposition: Chemistry for New Frontiers

Orlando, FL, April 2019

AWARDS, CERTIFICATIONS AND AFFILIATIONS

Outstanding Teaching Assistant Award, <i>Syracuse University</i>	April 2022
Certified Associate in Project Management (CAPM), <i>Project Management Institute</i>	February 2022 – present
Facilitator for Peer Chats organized by WISE-FPP Program	2021-2022
American Association for the Advancement of Science (AAAS)	2019-present
American Peptide Society	2019-present
Women in Science and Engineering Future Professionals Program (WiSE)	2019-2021
Active member of the volunteering team at the Milton J. Rubenstein Museum of Science & Technology	2019- 2021
American Chemical Society	2018-present
Biochemistry Journal Club, President – Syracuse University	2018-2020