From Unnatural Amino Acid Incorporation to Artificial Metalloenzymes

Dissertation by

Arwa Abdulaziz Makki

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

© December 2016
Arwa Abdulaziz Makki
All Rights Reserved
ABSTRACT

From Unnatural Amino Acid Incorporation to Artificial Metalloenzymes

Arwa A. Makki

Studies and development of artificial metalloenzymes have developed into vibrant areas of research. It is expected that artificial metalloenzymes will be able to combine the best of enzymatic and homogenous catalysis, that is, a broad catalytic scope, high selectivity and activity under mild, aqueous conditions.

Artificial metalloenzyme consist of a host protein and a newly introduced artificial metal center. The host protein merely functions as ligand controlling selectivity and augmenting reactivity, while the metal center determines the reactivity. Potential applications range from catalytic production of fine chemicals and feedstock to electron transfer utilization (e.g. fuel cells, water splitting) and medical research (e.g. metabolic screening). Particularly modern asymmetric synthesis is expected to benefit from a successful combination of the power of biocatalysis (substrate conversion via multi-step or cascade reactions, potentially immortal catalyst, unparalleled selectivity and optimization by evolutionary methods) with the versatility and mechanism based optimization methods of homogeneous catalysis. However, so far systems are either limited in structural diversity (biotin-avidin technology) or fail to deliver the selectivities expected (covalent approaches).

This thesis explores a novel strategy based on the site-selective incorporation of unnatural, metal binding amino acids into a host protein. The unnatural amino acids can either serve directly as metal binding centers can be used as anchoring points for artificial metallo-cofactors.

The identification expression, purification and modification of a suitable protein scaffolds
is fundamental to successfully develop this field. Chapter 2 and 3 detail a rational approach leading to a highly engineered host protein. Starting with fluorescent proteins, which combine high thermal and pH stability, high expression yields, and fluorescence for ease of quantification and monitoring an efficient and fast purification protocol was developed first. The purification protocol uses a combination of heat precipitation and three-phase-partitioning (TPP). It provides high yield and purity without requiring any tag.

Building on the favourable properties of fluorescent proteins, the non-metal binding, highly stable host-scaffold mTFP* was generated through rational design. The incorporation of artificial metal binding sites, the allowed the selective formation of artificial metalloenzymes, which show catalytic activity and moderate to good chiral induction in the Diels-Alder Cyclization and Friedl-Crafts Acylation.

Chapter 4 of the thesis describes the use of UAA incorporation to generate artificial metal binding sites. Computational studies and homology modelling successfully highlighted several positions in mTFP*, which are particularly suitable for UAA incorporation without any disruption of the protein structure. Application of a functional orthogonal aaRS/tRNA pair developed by P.G. Schultz and co-workers allowed the site-specific incorporation of UAAs in the host protein framework. Changes in fluorescence intensity revealed preferences of various UAAs for specific incorporations sites. The three UAAs, pIF, pAzF, and pEynF were incorporated into mTFP* in good yields, while pBF does only deliver low protein yields. A successfully established on-protein MIYAUCLA borylation reaction allows convert well-incorporated pIF into pBF circumventing the problem of low expression yields.

Chapter 5 details the use of the azide-functionality of pAzF for the bioconjugation of artificial metal-binding cofactors through CuAAC. The triazole ring formed during this reaction serves as an additional moderate σ-donor/π-acceptor ligand of the metal binding site. We demonstrated the potential of site-specific modifications within the protein host with a versatile subset of artificial cofactors. Following transition metal binding, the newly created metal sites show catalytic activities that nature does not provide. The proof of concept study highlights the potential of the present mTFP* based
catalysts in asymmetric Tsuji Trost allylation reactions and Diels-Alder cycloadditions. Dual anchoring of the cofactor lead to increased enantioselectivities, which is a direct result of a better-defined orientation of the catalytic center on the protein surface.

Following the utilization of the CuAAC click reaction for the generation of artificial metalloenzymes, the last chapter of this thesis reports the development of a heterogeneous catalyst system for this reaction, which overcomes limitations of homogenous protocols. The recyclable core-shell structured Cu₂O/Cu-nanowire catalyst is highly active, does not lead to protein precipitation, can be filtered off after the reaction and provides copper free bioconjugation products.
ACKNOWLEDGEMENTS

I would like to begin with my special appreciation and thanks to my advisor Prof. Prof. Jörg Eppinger. It was my honor to be one of the first students to join his group. I would like to thank you for his guidance, advice and suggestions. Besides my advisor, I would like to thank my thesis committee members Prof. Klaus-Victor Peinemann, Prof. Stefan Arold, and Prof. Wael Elmoslimany. I sincerely appreciate the time they have all taken to positively discuss and criticize this dissertation.

I would also take this opportunity to thank my colleagues who have worked with me to accomplish the first part of my thesis (Chapter 3), namely, Dr. S. Abdul Rajjak for his computational studies and homology modeling for mTFP* mutations sites. Dr. Johannes Fischer, Dr. Felix Quitterer, and Dr. Anand Radhakrishnan, for their effort in engineering and characterization mTFP* including Crystallography, tmFRET measurements, Fluorescent lifetimes and Quantification of metal binding constant. Also, Dr. Anna Zernickel for her help in (Chapter 4) performing MIYURA borylation and subsequent SUZUKI cross coupling. I also want to thank Dr. Seema Ghoprade for her optimization of the heterogeneous Cu_2O/Cu-NW catalyzed CuAAC protocol by using fluorogenic coumarin azide assay (Chapter 6). And together Dr. Anna and Dr. Seema for their effort synthesizing and providing the Unnatural amino acid the building blocks for this thesis. A great thanks goes for Prof. Zhiping Lai for providing me the Cu/Cu_2O nanowires catalyst (Chapter 6) and Prof. Kuo-Wei HUANG and Dr. Huaifeng Li for characteristics study of Cu/Cu_2O nanowires catalyst.

Special thanks also to Dr. Dinesh Sawant for his immensely contribution in the optimization click chemistry protocols and catalysis applications in my projects, also for his help in answering all my inquiries. I am extremely grateful to him for conveying me great ideas and developing my problem-solving skills in chemistry. Without his help, it would have been difficult for me to finish some of my projects (Chapter 5-6).

I am also thankful to the Proteomics Core Lab at KAUST, for their help and technical advises for protein analysis by LC/MS.
My appreciation also goes to my friends and colleagues for making my time at KAUST a great experience. I would like to thank all my good friends in KAUST for their kind support. I am grateful for the time I spent with you. Thank you for any advice, support or even a smile that you may have given me (you know who you are, no need to list the names!).

My deepest thanks and appreciation goes to my family for their selflessly support over the past years. Words cannot express my gratitude for their support. Without their help, I could not have completed this work. Words cannot express how grateful I am to my Mom (Feryal Arab) and my Dad (Abdulaziz Makki) for all their support, encouragement, and endless love. I owe all my success to both of you. Also, I like to thank all my sisters (Laila, Najla, and Amani) and brother (Omar). Their continuous love and support cannot be forgotten. I offer special thanks to my big sister Laila and young sister Amani. They have been there to help me whenever I needed, and their generous support has meant a lot to me more than I could ever express. My thanks also go out to the nice help and love I received from Loujain and Saja.

A great thanks go to my dear husband (Osama) and beloved kids (Roaa, Youssef, Ahmad, and Noor), who were always supportive through all those years. Thanks a lot Osama for being the most supportive person in the world, thank you for taking this journey with me for the past five years. I could not have done it without your support and love.

Finally, I would like to express my thanks and gratitude to King Abdullah, may he rest in peace. Thank you for giving me this opportunity.

Context:

Most of this work has been done in King Abdullah University for Science and Technology (KAUST), as a part of my research in my PhD dissertation.

Based on the collaboration between KAUST Catalysis Center (KCC) and Technical University of München [TUM], the crystallization of my purified proteins was achieved in TUM.
1.7 OBJECTIVE AND PRE-CONSIDERATIONS .................................................................................................................. 64
1.8 REFERENCE: .......................................................................................................................................................... 65

2 CHAPTER 2 A PROTOCOL FOR HIGH THROUGHPUT FLUORESCENT PROTEINS PURIFICATION ............ 71
2.1 ABSTRACT: ............................................................................................................................................................ 71
2.2 INTRODUCTION ..................................................................................................................................................... 72
2.3 RESULTS AND DISCUSSION ............................................................................................................................... 75
2.3.1 CLONING AND EXPRESSION OF FLUORESCENT PROTEINS ........................................................................ 75
2.3.2 PURIFICATION OF FLUORESCENT PROTEINS ............................................................................................. 77
2.3.3 PURIFICATION OF FP WITH UAA .................................................................................................................. 83
2.3.4 HIGH THROUGHPUT (HT)-PURIFICATION IN A 96 WELL PLATE FORMED “PARALLEL ASSAY” .................. 85
2.3.5 SPECTRA PROPERTIES OF PURIFIED PROTEINS ....................................................................................... 86
2.4 CONCLUSIONS .................................................................................................................................................... 88
2.5 MATERIAL AND METHODS ............................................................................................................................... 90
2.5.1 CLONING OF CONSTRUCTS ............................................................................................................................ 90
2.5.2 PROTEIN EXPRESSION .................................................................................................................................... 91
2.5.3 PROTEIN EXPRESSION WITH UAA ............................................................................................................... 92
2.5.4 PROTEIN PURIFICATION ................................................................................................................................ 92
2.5.5 PARALLEL HT- (T/3PP) PURIFICATION IN 96 WELL PLATE ........................................................................ 93
2.5.6 UV-VIS AND FLUORESCENCE ANALYSIS ....................................................................................................... 94
2.5.7 MASS SPECTROMETRY: .................................................................................................................................. 94
2.5.8 POLYACRYLAMIDE GEL ELECTROPHORESIS ............................................................................................. 94
2.6 REFERENCE .......................................................................................................................................................... 95
2.7 SUPPLEMENTARY MATERIAL ............................................................................................................................ 97

3 CHAPTER 3 DEVELOPMENT OF A ROBUST AND VERSATILE HOST PROTEIN FOR DESIGN AND EVALUATION OF ARTIFICIAL METAL CENTERS .............................................................................. 99
3.1 ABSTRACT .............................................................................................................................................................. 99
3.2 INTRODUCTION ..................................................................................................................................................... 99
3.3 RESULTS............................................................................................................................................................... 102
3.3.1 DESIGN AND ENGINEERING OF THE HOST PROTEIN MTFP* ......................................................................... 102
3.3.2 MTFP* IS A HIGHLY STABLE HOST SCAFFOLD WITH INTEGRATED FLUORESCENT REPORTER .................. 105
3.3.3 MTFP* SHOWS A VERY LOW BACKGROUND AFFINITY TO TRANSITION METAL IONS ............................... 108
3.3.4 SPECIFIC TRANSITION METAL BINDING SITES CAN BE INCORPORATED INTO MTFP* .............................. 111
3.3.5 tFRET ASSAYS REVEAL THE DYNAMIC NATURE OF THE METAL COORDINATION IN SOLUTION ................ 115

9
3.3.6 THE ARTIFICIAL Cu-PROTEINS SHOW CATALYTIC ACTIVITY AND ENANTIOSELECTIVITY .................................................. 118
3.4 DISCUSSION ........................................................................................................................................................................... 119
3.5 CONCLUSION ........................................................................................................................................................................... 122
3.6 MATERIAL AND METHODS ..................................................................................................................................................... 123
3.6.1 GENERAL REMARKS ....................................................................................................................................................... 123
3.6.2 CLONING AND MUTAGENESIS ....................................................................................................................................... 123
3.6.3 SITE-DIRECTED MUTAGENESIS ..................................................................................................................................... 123
3.6.4 EXPRESSION AND PURIFICATION .................................................................................................................................. 124
3.6.5 HEAT PURIFICATION ....................................................................................................................................................... 124
3.6.6 PROTEIN STORAGE ............................................................................................................................................................ 125
3.6.7 MASS SPECTROMETRY ..................................................................................................................................................... 125
3.6.8 SPECTRAL ANALYSIS ....................................................................................................................................................... 125
3.6.9 STABILITY STUDIES .......................................................................................................................................................... 125
3.6.10 SOLVENT SCREEN: .......................................................................................................................................................... 125
3.6.11 METAL INCUBATION ........................................................................................................................................................ 126
3.6.12 CRYSTALLOGRAPHY ....................................................................................................................................................... 126
3.6.13 TMFRET MEASUREMENTS ........................................................................................................................................... 127
3.6.14 FLUORESCENT LIFETIMES ............................................................................................................................................ 127
3.6.15 QUANTIFICATION OF METAL BINDING CONSTANTS ...................................................................................................... 127
3.6.16 DIELS ALDERASE ACTIVITY .......................................................................................................................................... 132
3.6.17 COMPUTATIONAL DETAILS: ........................................................................................................................................ 132
3.6.18 FORCE FIELD PARAMETERS FOR THE CHROMOPHORE .............................................................................................. 133
3.6.19 MOLECULAR DYNAMICS ............................................................................................................................................ 134
3.7 REFERENCES ........................................................................................................................................................................... 135
3.8 SUPPLEMENTARY MATERIAL ................................................................................................................................................ 138

4 CHAPTER 4 INCORPORATION OF UNNATURAL AMINO ACID FOR BIOCONJUGATION REACTIONS 155
4.1 ABSTRACT .................................................................................................................................................................................. 155
4.2 INTRODUCTION ........................................................................................................................................................................ 156
4.3 RESULT AND DISCUSSION ................................................................................................................................................... 159
4.3.1 IDENTIFYING MUTATION SITES TO CREATE ARTIFICIAL METALLOENZYMES ..................................................... 159
4.3.2 SITE-SPECIFIC INCORPORATION OF VARIOUS P-SUBSTITUTED PHENYLALANINE DERIVATIVES USING AAARS/TRNA PAIR. 160
4.4 CONCLUSION ........................................................................................................................................................................... 172
4.5 MATERIAL AND METHOD ..................................................................................................................................................... 173
4.5.1 Cloning, Mutagenesis, Expression with UAA and Purification .......................................................... 173
4.6 Synthesis of UAA .................................................................................................................................. 173
4.6.2 General Procedure for the Pd Catalyzed MIYURA Borylation Reaction on Protein, Followed by
Suzuki Cross Coupling ......................................................................................................................... 176
4.7 Reference: ......................................................................................................................................... 177
4.8 Supplementary Material ................................................................................................................... 180

5 CHAPTER 5 A SITE-SELECTIVE ANCHORING STRATEGIES TOWARDS FLUORESCENT ARTIFICIAL
METALLOPROTEIN ................................................................................................................................. 183

5.1 Abstract ............................................................................................................................................... 183
5.2 Introduction: ....................................................................................................................................... 185
5.3 Result and Discussion ........................................................................................................................ 187
1. Site Selective Incorporation of UAA and Synthesis of Corresponding Pincer Type Artificial
Metalloenzymes Using CuAAC Approach ............................................................................................... 188
2. Creation of Chiral Pocket by Using Dual Incorporation Concept by Using Cu(I)-Catalyzed Azide-Alkyne
CycloadDITION (CuAAC) ....................................................................................................................... 188
3. Catalytic Testing of the Synthesized Artificial Metalloenzymes ............................................................ 188

5.3.1 Site Selective Incorporation of UAA and Synthesis of Corresponding Pincer Type Artificial
Metalloenzymes Using CuAAC Approach. ............................................................................................... 188
5.3.2 Creation of Chiral Pocket by Using Dual Incorporation Concept by Using Cu(I)-Catalyzed Azide-Alkyne
CycloadDITION (CuAAC) ....................................................................................................................... 196
5.3.3 Catalysis as an Application ............................................................................................................ 200
5.4 Conclusion ......................................................................................................................................... 203
5.5 Materials and Methods ..................................................................................................................... 204
5.5.1 Cloning .......................................................................................................................................... 204
5.5.2 Site-Directed Mutagenesis ........................................................................................................ 204
5.5.3 Expression ................................................................................................................................... 204
5.5.4 Purification .................................................................................................................................... 205
5.5.5 UV-Vis Spectral Analysis ............................................................................................................. 205
5.5.6 Mass Spectrometry ..................................................................................................................... 206
5.5.7 Click Protocol .............................................................................................................................. 206
5.5.8 Catalysis ...................................................................................................................................... 207
5.6 References ......................................................................................................................................... 209
5.7 Supplementary Material ................................................................................................................... 211
5.7.1 Molecular Biology ........................................................................................................................ 211
LIST OF ILLUSTRATIONS

Chapter 1

Figure 1.1. Energetic diagram of a catalysed reaction. $E_1=$ activation energy of uncatalyzed reaction, $E_2=$ activation energy of catalysed reaction. ................................................................. 26

Figure 1.2. General trends in $\sigma$-bonding of early and late transition metals to common ligands. ........... 32

Figure 1.3. Michaelis-Menten Kinetics ................................................................................................... 35

Figure 1.4. Diels-Alder Reaction. Mechanism of [4+2] Cycloaddition of 1,3-butadiene and ethylene...... 37

Figure 1.5. The formation of the alkene with the palladium catalyst leads to a $\pi$ - complex. .................. 39

Figure 1.6. Influence of the $\Delta G^\neq$ on enantiomeric excess (up) and energetic diagram of an asymmetric hydrogenation reaction (down). Curtin - ............................................................... 44

Figure 1.7. Concept of hybrid catalyst .................................................................................................... 47

Chapter 2

Figure 2.1. A) Maturation pathway of Green Fluorescent Protein GFP (Ser-Tyr-Gly)$^{27}$, B) Chemical structures of chromophores of various fluorescent proteins used in this study. ................................. 74

Figure 2.2. pET303/FPs vector. FPs expressed under the control of T7 promoter with XbaI and XhoI restriction site and C-terminal His-tag. .................................................................................. 76

Figure 2.3. Fluorescent intensity screening of FPs expression under different condition using pET303, pET302, and pET45b vectors. T303) under fix conditions (incubation time of 24h at 20°C with 1mM IPTG)................................................................. 77

Figure 2.4. SDS-PAGE analysis, of A) mKate2, B) TagRFP, C) mVenus, D) eCGP123, E) mTAGBFP, F) Sf-GFP, and G) mTFP1; loaded after three different purification protocols, Lane 1: T/3PP, Lane 2: Ni-affinity chromatography and Lane 3: Thermal purification, to compare the purity. ...................... 79

Figure 2.5. Thermal stability: indicate stability measures throughout the quantification of specific spectroscopic features for each protein (absorption: nm, fluorescence: nm). .................................................. 82

Figure 2.6. SDS-PAGE analysis, of mTFP1 loaded after the parallel purification method. Lane 1: cell lysate, Lane 2: 3PP, and Lane 3: T/3PP........................................................................................................ 85

Figure 2.7. Reproducible of mTFP1 by using parallel purification protocol........................................... 86

Figure 2.8. Absorption and emission spectra of purified FPs by T/3PP protocol.................................................. 87

Figure S2.9. ESI-TOF: A) mKate2=27139.9 Da. B) TagRFP=27156.9 Da. C) mVenus=27957.5 Da. D) eCGP123=26913.6 Da. E) mTAGBFP=27871.3 Da. F) Sf-GFP=27477.1 Da. G) mTFP1= 27971.5 Da. H) mTFP1-183 PIF, ($\times$) indicates Mass of 28116.3 Da. ................ 98

Chapter 3

Figure 3.1. A) Prospective cleft in mTFP1 (PDB: 2HQK) H205, Y209; B) Molecular Dynamics study of mTFP1 and mutant1, indicating reasonable stability despite prospective mutagenic conversion..... 105
Figure 3.2. Characterization of mTFP*: .................................................................................. 106
Figure 3.3. X-Ray mTFP*: ...................................................................................................... 108
Figure 3.4. Metal Binding: ........................................................................................................ 111
Figure 3.5. pH Quenching Studies: mbYDCHH titrated with various concentrations of Copper(II)Nitrate over a range of pHs ............................................................................................. 111
Figure 3.6. mTFPCHH: E) B-Factor Comparison of mTFP*, mTFPCHH, mTFPCHH_Cu, showing stabilized coordination throughout Copper coordination by 2xHis-Motive; F) Superposition mTFP*(green), mTFPCHH (orange), mTFPCHH_Cu (blue), visualizing structural rearrangement of cleft flanking loop, especially upon Cu coordination (* indicates shifted residue Y/H200); ............ 114
Figure 3.7. tmFRET: mTFP1 and mTFP* titration against various amounts of Copper(II)Nitrate at pH 6 (A) and pH 7.5 (B). .......................................................................................................................... 116
Figure 3.8. Diels Alder Screen for mTFPXHH mutants ........................................................................ 119
Figure 3.9. tmFRET: mTFP1, mTFP*, mbYDCHH and mbYDEHH; A) pH 6.0, B) pH7.5 ................. 121
Figure 3.10. Anionc form of the chromophore (AYG). For each atom the calculated partial charge was reported together with the amber atom type .............................................................................. 133
Figure S3.11. Sequence mTFP1 (underscore = deleted nucleotides) ........................................... 138
Figure S3.12. Align mTFP1/l.1 ................................................................................................... 139
Figure S3.13. pET303/SUMO_mTFP1.1 .................................................................................... 140
Figure S3.14. Align mTFP1.1/mTFP* ...................................................................................... 141
Figure S3.15. pET303/SUMO_mTFP* ....................................................................................... 143
Figure S3.16. SDS-PAGE: mTFP* purification, (1) Sumo conjugated mTFP*, (2) SUMO-digested mTFP*, (3) pure mTFP* .............................................................................................................. 144
Figure S3.17. ESI-TOF: mTFP*, indicates maturated Mass of 25154 Da ...................................... 144
Figure S3.18. Chromophore Formation (A-Y-G) ......................................................................... 145
Figure S3.19. mTFP*: Concentration dependent Absorbance (A) and Fluorescence (B) (ex: 468nm em: 495nm) ......................................................................................................................... 145
Figure S3.20. Circular Dichroism Scan (mTFP*) indicating structural perturbation at temperatures > 89°C ........................................................................................................................................... 145
Figure S3.21. Urea Stability ...................................................................................................... 146
Figure S3.22. Aprotic and Protic Solvent Screen: ......................................................................... 146
Figure S3.23. ESI-TOF: A, B) mTFP* Heavy Metal Soak with Pd2+. Low concentrations of metal ions show no additional mass peaks .................................................................................................................. 147
Figure S3.24. B-Factor Plot mTFP1 (A, PDB: 2HQK) vs. mTFP* (B, PDB: 4Q9W) ....................... 148
Figure S3.25. mTFP*:Pd, K9, D112 coordinating Pd2+ (PDB: 4Q9X) ........................................ 149
Figure S3.26. B-Factor Comparison of mTFP* (PDB: 4Q9W), mTFPCHH, mTFPCHH_Cu (PDB: 4R6D) ................................................................................................................................. 150

Figure S3.27. Temperature Stability of mTFPCHH/EHH, circular dichroism at 205nm indicates unfolded fraction. ........................................................................................................ 150

Figure S3.28. ESI-TOF: Cu2+ and Pd2+ Soak of mTFPCHH/EHH................................................................. 151

Figure 3.29. Metal Binding: A - F) Fluorescence measurement of mTFP1, mTFP* and mTFPCHH ...... 152

Figure S3.30. Relative fluorescent lifetime of mTFP* and mTFPCHH................................................................. 153

Figure 3.31. Distance Mapping of Coordinating Motifs: ........................................................................ 154

Figure 3.32. Förster Radius Estimations: ........................................................................................................ 154

Chapter 4

Figure 4.1. The Azide–alkyne cycloaddition...................................................................................................... 159

Figure 4.2. Structure analysis was pursued with Pymol by Schrödinger INC............................................... 159

Figure 4.3. Incorporation of para-Azido-(L)-phenylalanine into mTFP*............................................................. 160

Figure 4.4. Fluorescent intensity screening of the expression culture for selecting the best pEVOL plasmid used for incorporation of pEnyF into mTFP*........................................................................ 161

Figure 4.5. Proteins structure. mTFP* (PDB: 4Q9W), representation by for the successful incorporation of UAAs (pAzF, pEynF, pIF, and pBF). ......................................................................................... 162

Figure 4.6. Incorporation screening of different UAAs (pAzF, pEynF, pIF, and pBF) incorporated into mTFP* at different mutations sites (128, 188, 200, and 204)......................................................... 164

Figure 4.7. Crystal structure for a) mTFP*204 pIF at 1Å.................................................................................... 165

Figure 4.8. Paladcycle catalyst (cat-1).............................................................................................................. 169

Figure 4.9. ESI-TOF analysis of MIYAURA borylation, of mTFP*pIF to mTFP*pBF................................. 171

Figure 4.10. Different structures, which could be cross coupled to proteins.................................................. 172

Figure 4.11. Borylation was performed in GC/MS vials, separation of the catalyst was done by centrifugation through NANOSEP 30K OMEGA ................................................................................. 176

Figure S4.12. SDS-gel analysis for pure proteins 1= mTFP*200pAzF, 2=mTFP*204 pEynF and 3=mTFP*188pIF.......................................................... 181

Figure S4.13. ESI-TOF analysis of the incorporation of unnatural amino acids into the TAG position. A) mTFP*Y200 pAzF, B) mTFP*Y204 pEynF, C) mTFP*188pIF, and D) mTFP*188 pBF. .......... 182

Chapter 5

Figure 5.1. Alkynes used for screening and for on protein synthesis of BTP ligands. 1= ethynylpyridine; 2=2-ethynylpyridine and 3=2,6-diethynyl-3,5-dimethoxy pyridine. .......................................................... 188

Figure 5.2. A) ESI-TOF analysis of Cu+2 conjugation to the modified protein mTFP*(Y188) pAzF by click reaction with 2,6-diethynylpyridine2Cu@mTFP*(Y188) pAzF-2-A (red line). B) UV-vis spectra of modified protein mTFP*(Y188) pAzF-2 at 464 nm (green line), and Cu@mTFP*(Y188) pAzF-2-A at 470nm (red dashed line) in 50 mM ammonium acetate buffer pH 7.4. ...... 193
Figure 5.3. A) ESI-TOF analysis of Cu$^{2+}$ and Pd$^{2+}$ coordination to the modified protein mTFP*(128N/21Y)pAzF (P1) by click reaction with ligand 3, Cu@mTFP*(128N/21Y)pAzF-3 (red line), and Pd@mTFP*(128N/21Y)pAzF-3 (blue line). B) UV-Vis spectra of modified protein mTFP*(128N/21Y)pAzF-3 (green line), Cu@mTFP*(128N/21Y)pAzF-3 (red dashed line) and Pd@mTFP*(128N/21Y)pAzF-3 (blue dashed line) in 50 mM ammonium acetate buffer pH 7.4. 199

Figure 5.4. A) ESI-TOF analysis of Cu$^{2+}$ and Pd$^{2+}$ coordination to the modified protein mTFP*(200Y/164E)pAzF (P2) by click reaction with ligand 3, Cu@mTFP*(200Y/164E)pAzF-3 (red line), and Pd@mTFP*(200Y/164E)pAzF-3 (blue line). B) UV-Vis spectra of modified protein mTFP*(200Y/164E)pAzF-3 (green line), Cu@mTFP*(200Y/164E)pAzF-3 (red dashed line) and Pd@mTFP*(200Y/164E)pAzF-3 (blue dashed line) in 50 mM ammonium acetate buffer pH 7.4. 200

Figure S5.5. Nucleotide sequence for mTFP*188TAG, mTFP*204TAG and FtnA79TAG (Red Highlight = mutation to stop codon TAG for pAzF and pEnyF incorporation) 211

Figure S5.6. pET303/SUMO_plasmids 213

Figure S5.7. Selection of Binding UAAs in the Pocket Cavity positions mainly: Y200 for the dual incorporation and Y204 for the single incorporation 214

Figure S5.8. proposed pocket assembly for metal centre complex, by double sites incorporation of UAA 214

Figure S5.9. proteins structure: for single site incorporation A) mTFP* (Y188) pAzF and B) mTFP* (Y204) pAzF. And dual site incorporation C) mTFP*(N128-Y21) pAzF and D) mTFP* (Y200-E164) pAzF. (PDB:4Q9W) 215

Figure S5.10. ESI-TOF analysis of the incorporation of unnatural amino acids pAzF into the TAG position of mTFP*Y188 TAG and mTFP*Y204 TAG. A) mTFP*Y188pAzF 25180 Da, B) mTFP*Y204pAzF 25180 Da. wt = is the wild type mTFP* protein before the incorporation process 216

Figure S5.11. ESI-TOF analysis of the mTFP*188pAzF with 2,6-diethynylpyridine ligand 2. The only product forms from the coupling reaction 216

Figure S5.12. ESI-TOF analysis of the mTFP*Y188pAzF protein modified by click reaction with different azide ligands A-E 217

Figure S5.13. ESI-TOF analysis of the mTFP*Y204pAzF protein modified by click reaction with different azide ligands A-E 218

Figure S5.14. ESI-TOF analysis of the dual incorporation efficiency of 219

Figure S5.15. 2,6-diethynyl-3,5-dimethoxypyridine (Figure 1, compound 3) 221

Chapter 6

Figure 6.1. (BimC4A)3 ligand 226

Figure 6.2. optimization of mTFP*188 pAzF (0.05 mM) conjugations using the heterogenous Cu2O/Cu-NW catalyst 228

Figure 6.3. 1 = Cy3-Alkyne, 2 = Cy5-Alkyne, 3=Sulforhodamine-101-Alkyne, 4= Biotin-PEG4-Alkyne, 5= Cy3-Azide, 6= Cy6-Azide, 7=Sulforhodamine-101-Azide, 8=Biotin-PEF4-Azide 231

Figure S6.4. Nucleotide sequence for mTFP*188TAG, mTFP*204TAG and FtnA79TAG (Red Highlight = mutation to stop codon TAG for pAzF and pEnyF incorporation) 239
Figure S6.5. pET303/SUMO plasmids. A)pET303/SUMO_mTFP*204TAG, .............................................................. 240

Figure S6.6. proteins structure. A)mTFP*Y188TAG and B)mTFP*Y204TAG (PDB:4Q9W), C)FtnA 79TAG (PDB: 3UNO). ........................................................................................................................................... 241

Figure S6.7. ESI-TOF analysis of the incorporation of two unnatural amino acids pAzF and pEynF into the TAG position of mTFP*Y188 TAG, mTFP*Y204 TAG and FtnA E79 TAG................................................................. 242

Figure S6.8. ESI-TOF analysis of the modified mTFP*pAzF after click reaction using homogenous CuAAC protocol. .............................................................................................................................................. 243

Figure S6.9. ESI-TOF analysis of the modified mTFP*pEynF after click reaction using homogenous CuAAC protocol. .............................................................................................................................................. 244

Figure S6.10. ESI-TOF analysis of the modified FtnA E97pAzF after click reaction using homogenous CuAAC protocol. .............................................................................................................................................. 245

Figure S6.11. ESI-TOF analysis of the modified FtnA E97pEynF after click reaction using homogenous CuAAC protocol. .............................................................................................................................................. 246
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Comparison of chemical and biological systems</td>
<td>48</td>
</tr>
<tr>
<td>1.2</td>
<td>Comparison between homogeneous and enzymatic catalysis</td>
<td>49</td>
</tr>
<tr>
<td>2.1</td>
<td>Spectra characteristic and expression yield of FPs.</td>
<td>78</td>
</tr>
<tr>
<td>2.2</td>
<td>Purification comparison of FPs by different purification protocols</td>
<td>80</td>
</tr>
<tr>
<td>2.3</td>
<td>FPs detected mass by ESI-TOF</td>
<td>84</td>
</tr>
<tr>
<td>2.4</td>
<td>Comparison of different purification protocols of mTFP1 by using parallel purification method</td>
<td>86</td>
</tr>
<tr>
<td>2.5</td>
<td>Oligonucleotides for FPs (eCGP123, Sf_gfp, mVenus, TAG RFP, mKate2, mTAG BFP2, mTFP1, and mTFP1K183 PIF)</td>
<td>91</td>
</tr>
<tr>
<td>2.6</td>
<td>Table of pETs vector used for screen FPs expression</td>
<td>91</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary of metal binding characterization for mTFP1, mTFP*, mTFPCHH and mTFPEHH.</td>
<td>115</td>
</tr>
<tr>
<td>3.2</td>
<td>Spectroscopic characterization of the interaction of Copper or Zink with mTFP1, mTFP*, mTFPCHH and mTFPEHH</td>
<td>117</td>
</tr>
<tr>
<td>3.3</td>
<td>Diels-Alderase activity test</td>
<td>118</td>
</tr>
<tr>
<td>3.4</td>
<td>Friedel-craft Alkylase activity test</td>
<td>118</td>
</tr>
<tr>
<td>3.5</td>
<td>Oligonucleotides to establish mTFP*</td>
<td>142</td>
</tr>
<tr>
<td>3.6</td>
<td>Metal Salts for Conjugation Test: The following metal salts did show any influence on mass spectra or fluorescence measurements</td>
<td>147</td>
</tr>
<tr>
<td>3.7</td>
<td>Crystallographic Parameters</td>
<td>148</td>
</tr>
<tr>
<td>3.8</td>
<td>Oligonucleotides to establish mutants of mTFP*</td>
<td>150</td>
</tr>
<tr>
<td>4.1</td>
<td>Expected MS for incorporation of different UAAs into mTFP* at different mutations sites</td>
<td>162</td>
</tr>
<tr>
<td>4.2</td>
<td>Expression yield for all protein positions with different UAAs</td>
<td>163</td>
</tr>
<tr>
<td>4.3</td>
<td>Different borylation conditions were investigated for Below reaction</td>
<td>170</td>
</tr>
<tr>
<td>4.4</td>
<td>Expected MS for different mutants of mTFP*</td>
<td>171</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary of structural and molecular weight of the UAAs used</td>
<td>175</td>
</tr>
<tr>
<td>4.6</td>
<td>Oligonucleotides for mTFP* mutation sites for UAA incorporation</td>
<td>180</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary of result for the Click bioconjugation of fluorescent (mTFP<em>Y188pAzF and mTFP</em>Y204pAzF) proteins with the different azide ligands A-E. To create protein with different chemical moiety</td>
<td>192</td>
</tr>
<tr>
<td>5.2</td>
<td>Metal binding affinity and enantiomeric selectivity by Diels-Alder reaction</td>
<td>194</td>
</tr>
<tr>
<td>5.3</td>
<td>Yield of incorporation efficiency of pAzF to the dual site into mTFP* and the CuAAC reaction between ethynylpyridine 1 or 2,6-diethynyl-3,5 dimethoxypyridine3 and two different variant of mTFP* pAzF.</td>
<td>198</td>
</tr>
</tbody>
</table>
Table 5.4. Conjugation of metal ions (Cu$^{+2}$ and Pd$^{+2}$) to modified proteins with ligand 3 mTFP*(N128-Y21) pAzF-3 and mTFP*(200Y/164E)pAzF-3 and its application in Diels-Alder reactions............ 201

Table 5.5. Conjugation of Pd+2 ion to modified proteins with ligand 2mTFP*(N128-Y21) pAzF-3 and mTFP*(Y200-E164) pAzF-3 and its application in Tsuji trost reactions............................................. 202

Table S5.6. Oligonucleotide, PCR primers used for cloning and mutagenesis ................................................. 212

Table S5.7 Alkynes used for screening and for synthesis of BTP ligands on protein. 1= ethynylpyridine; 2=2-ethynylpyridine and 3=2,6-diethyl-3,5-dimethoxypyridine. ......................................................... 220

Table 6.1. Optimization of CuAAC reaction using Cu2O/Cu-NW catalyst on protein mTFP*188PAzF.... 227

Table 6.2. Summary of result for the Cu2O/Cu-NW at Cu nanowires catalyzed Click bioconjugation of fluorescent (mTFP*Y188pAzF, mTFP*Y204pAzF, mTFP*Y204pEynF and mTFP*Y204pEynF) and non-fluorescent (FtnA 79 pAzF and FtnA 79 pEynF) proteins with the corresponding azide and alkyne ligand 1-8.................................................................................. 232

Table S6.3. PCR primers used for cloning and mutagenesis ..................................................................... 241
LIST OF SCHEMES

Scheme 1.1. Asymmetric synthesis of chiral compounds. ................................................................. 29

Scheme 1.2. The two different biological activities displayed by two enantiomers of propoxyphene: the (R,R)-isomer in an analgesic, while the (S,S)-isomer has antitussive effect. ................................. 30

Scheme 1.3. Chemoenzymatic dynamic kinetic resolution of (DKR). .............................................. 31

Scheme 1.4. Reaction catalyzed by artificial metalloenzyme ................................. 46

Scheme 5.1. Water-soluble click-pincer ligand, which will be attached to the mTFP* by double sites UAA to create a chiral pocket on the mTFP* surface. .................................................................................. 184

Scheme 5.2. Test experiment between the mTFP*188pAzF and 2,6-diethynylpyridine 2, with higher proteins concentration to pyridine dialkyne (2:1). ................................................................. 189

Scheme 5.3. Binding of different chemical moieties to the fluorescent protein by the CUAAC click reaction with 2,6-diethynylpyridine2 (up); R: Different azides used for click on second arm of 2,6-diethynylpyridine 2. Different aromatic, aliphatic and heterocyclic azides represent from A to E (down). .............................................................................................................................. 191

Scheme 5.4. Generation of an artificial metalloproteins pocket by dual site incorporation of pAzF followed by protein modification by click reaction with chelating di-alkyne ligand and further metal ion conjugation. ......................................................................................................................... 197

Scheme 6.1. Fluorogenic CuAAC for optimization ................................................................. 225
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(BimC₄A)₃</td>
<td>Accelerating ligand</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>aaRS/tRNA</td>
<td>aminoacyl-tRNA synthetase</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>ArMs</td>
<td>Artificial metalloenzymes</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>Asparagine</td>
</tr>
<tr>
<td>BAMHI</td>
<td>Restriction enzymes</td>
</tr>
<tr>
<td>BTP</td>
<td>bistriazole pyridine</td>
</tr>
<tr>
<td>Cm</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>Cu</td>
<td>copper</td>
</tr>
<tr>
<td>Cu₂O/Cu</td>
<td>Copper nanowire</td>
</tr>
<tr>
<td>CuAAC</td>
<td>Copper-catalyzed azide-alkyne cycloaddition</td>
</tr>
<tr>
<td>CuSO₄</td>
<td>Copper sulfate</td>
</tr>
<tr>
<td>Diglyme</td>
<td>2-Methoxyethyl ether, Bis(2- methoxyethyl)</td>
</tr>
<tr>
<td>Dioxan</td>
<td>1,4-diethylene dioxide</td>
</tr>
<tr>
<td>diPy</td>
<td>Dipyridylalanine</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide is an organosulfur</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>eCGP123</td>
<td>consensus green protein</td>
</tr>
<tr>
<td>ECKm</td>
<td>expression cassette Kanamycin</td>
</tr>
<tr>
<td>FP</td>
<td>Fluorescent protein</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent protein</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>Glutamine</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>His tag</td>
<td>Histamine purification tag</td>
</tr>
<tr>
<td>HQA</td>
<td>Hydrochinolinalanine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>LC/MS</td>
<td>Liquid chromatography–mass spectrometry</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>Leucine</td>
</tr>
<tr>
<td>Lpp</td>
<td>Lipoprotein</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MEH</td>
<td>Metal Enzyme Hydride</td>
</tr>
<tr>
<td>MES</td>
<td>MES monohydrate</td>
</tr>
<tr>
<td>mKate 2</td>
<td>monomeric Far red fluorescent protein</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mTagBFP</td>
<td>monomeric tag Bright blue fluorescence protein</td>
</tr>
<tr>
<td>mTFP*</td>
<td>genetically modified monomeric Teal fluorescent protein</td>
</tr>
<tr>
<td>mTFP1</td>
<td>monomeric teal fluorescent protein</td>
</tr>
<tr>
<td>mVevus</td>
<td>Monomeric Venus fluorescent protein</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium chloride</td>
</tr>
<tr>
<td>Ni</td>
<td>Nickel</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>OMEH</td>
<td>organometallic enzyme hybrid</td>
</tr>
<tr>
<td>pAZF</td>
<td>para-(L)-Azidophenylalanine</td>
</tr>
<tr>
<td>pBF</td>
<td>para-(L)-Bronophenylalanine</td>
</tr>
<tr>
<td>pCNF</td>
<td>para-(L)-Cyanophenylalanine</td>
</tr>
<tr>
<td>Pd</td>
<td>Palladium</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylenglycol</td>
</tr>
<tr>
<td>pEnyF</td>
<td>para-(L)-Ethenylphenylalanine</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>pIF</td>
<td>para-(L)-Iodophenylalanine</td>
</tr>
<tr>
<td>PTMs</td>
<td>Post Translation Modifications</td>
</tr>
<tr>
<td>Rh</td>
<td>Rhodium</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SDS PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Sf-GFP</td>
<td>Super fold green fluorescent protein</td>
</tr>
<tr>
<td>Term</td>
<td>Definition/Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>SUMO tag</td>
<td>Purification tag</td>
</tr>
<tr>
<td>TAA (UAA)</td>
<td>Ochre stop codon</td>
</tr>
<tr>
<td>TAG (UAG)</td>
<td>Amber stop codon</td>
</tr>
<tr>
<td>TAG (UAG)</td>
<td>Amber stop codon</td>
</tr>
<tr>
<td>TagRFP</td>
<td>Monomeric tag red fluorescent protein</td>
</tr>
<tr>
<td>TFE</td>
<td>Trifluoroethanol</td>
</tr>
<tr>
<td>TGA (UGA)</td>
<td>Opal stop codon</td>
</tr>
<tr>
<td>TGA (UGA)</td>
<td>Opal stop codon</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofuran</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(Hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>Tyr Y</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>UAA</td>
<td>Unnatural Amino Acids</td>
</tr>
<tr>
<td>UV-vis</td>
<td>Ultraviolet–visible spectroscopy</td>
</tr>
<tr>
<td>XhoI</td>
<td>Restriction enzymes</td>
</tr>
</tbody>
</table>
1 Chapter 1 Introduction

1.1 New Avenues in Catalysis

The ever-advancing understanding of catalytic systems is key to overcome limitations that exist in the traditional production of chemicals. It hence represents the main driver for the development of novel catalytic reaction and process\(^1\).

Recent innovations have lead to effective catalysts that can be used to manufacture functional biomolecular compounds or relevant precursors\(^2,3\). These advances have applications in the energy, health and nutrition sectors, industries, changing their processes and output. In turn, catalysts have been further optimized for industrial purposes\(^4\). The progress achieved by research and development of catalysts including heterogeneous and homogeneous forms, is undeniable, yet it appears to have reached a threshold with transition-metal catalysts. Thus far, transition-metal catalysts have been unable to accommodate large-scale application without compromising loss of function or being economically impractical, because their reactivity and/or their selectivity is limiting\(^5\). These shortcomings are highlighted by the range and capability of naturally occurring biocatalysts and their man-made engineered mutant ‘offspring’, which can economically be produced in significance\(^1,6\). With greater understanding into their working mechanisms and bioinformatics development, it has been possible to create de novo devices and optimized mutants\(^2,7\).

The fields of chemistry and biology are becoming more and more connected. Through advancements in each discipline, Scientists are able to gain greater comprehension of life processes. One example of a field that uses elements of both chemistry and biology is bioinorganic chemistry, which builds on aspects of biological science and inorganic chemistry using innovative theoretical and physical approaches\(^8\). Metal sites in biological processes are the main emphasis of the field of bioinorganic chemistry.

The utility of the metal within an organism is the defining aspect of its nature. Sodium, potassium, calcium, and other group I and II metals are found as structural elements or integral aspects of change regulation processes\(^9\). Such transition metals may have a range of uses. Those in single oxidation states can serve as Lewis-acids or structural elements.
Zinc fingers are one example. These protein domains maintain their stability due to zinc chelates residues (such as histidines or cysteines). Other oxidation states lend different utilities to elements: electron transfer, redox chemistry and oxygen transports are among the possible functions of iron or copper. Transition metals typically guarantee the catalytic activity of metalloenzymes. Vitamin B12 includes cobalt, for example; this combination represents a vital cofactor of enzymes used in methionine synthesis or to safeguard the nervous system\textsuperscript{10,11}.

Enzymes are considered to possess qualities that are absent in inorganic catalysts; they incur lower production costs, sustainable, environmentally friendly, show great substrate specificity, are highly efficient despite low energy requirements and involve fewer reaction steps that produce by-products\textsuperscript{2,12,13}. In contrast to non-biological catalysts, biocatalysts act in a narrow manner, unable to compete with the range of reactions of transition metal complexes, for example. The specificity of biocatalysts and their homologues can limit their application, which is emphasized by the ability of transition-metal catalysts to convert a wide range of reagents at high rate. This may be a fundamental limitation of biocatalysts; genetically engineered and enhanced mutant enzymes are unable to catalyse all the synthetically significant conversions that transition metals can achieve in organic solvents or aqueous buffer\textsuperscript{2,14}.

As a result, a ‘best of both’ concept emerges in the form of artificial metalloenzymes (ArMs), which would take the form of a protein scaffold supporting a synthetic metal catalyst. ArMs would combine the versatility of transition metals with the reactivity of enzymes\textsuperscript{6,15,16}.

To achieve this collaborative compound, it needs to be rationally designed, taking into account not only the requirements for catalysis of asymmetric reactions but critically to accommodate the chemo-, enantio-, regio- and substrate selectivity required by enzyme-like structures\textsuperscript{2,17,18}. Finding appropriate protein hosts is proving to be tricky and challenging\textsuperscript{16}.
1.2 Catalysis

1.2.1 Background

Catalysis is the process where a substance known as a catalyst speeds up a chemical reaction without the side effect of enduring chemical alteration. Any substance that accelerates chemical reaction rates without being disbursed or chemically altered is a catalyst. The reaction allows catalyst regeneration, which enables only a minute amount to be utilised\textsuperscript{19}. Reactions are impacted by equilibrium or thermodynamics. Instead, a catalyst alters the transition state’s Gibbs-free energy in that it steadies the transition state that would otherwise occur in a reaction without a catalyst, thereby reducing the greatest energy barrier (Figure 1.1). This indicates that catalysts allow overall thermodynamics and reaction outcomes of chemical reactions to remain the same while impacting the kinetics.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{catalysis_diagram.png}
\captionof{figure}{Energetic diagram of a catalysed reaction. \textit{E1}= activation energy of uncatalyzed reaction, \textit{E2}= activation energy of catalysed reaction.}
\end{figure}

Creation of new energy sources, materials manufacture, creation of effective medicines and environmental protection are among the processes benefitting from catalysis. In
recent years, significant advancement has been achieved in the field of catalysis, in particular regarding the comprehension of the molecular mechanisms that result in successful catalytic reactions. These discoveries have resulted in new catalytic systems development. Because so many chemical industrial processes necessitate successful catalytic reactions, it follows that research in this area is comprehensive.

Catalyst impact on reactions can be elucidated through the following elements:

- Transition state equilibrium
- Reduced reactant entropy through compulsory proximity and complimentary location caused by interactions
- Discerning improvements to particular pathways as opposed to contending pathways

Catalysis is vital both in and out of the lab, as it concerns not only chemical laboratories and industry but is also a naturally occurring phenomenon in chemical and biochemical reactions. Many chemicals cannot be created without catalytic reactions, which utilize less energy and waste as opposed to stoichiometric reactions.

Despite existing long before being named, the term “catalysis” was first used in 1835 by Jöns Jakob Berzelius who observed the ability of some chemicals to speed up reactions. His discoveries were bolstered by Humphry Davy’s realization that platinum catalyses coal gas oxidation. The Nobel Prize committee took notice of catalysis following Leipzig University’s Wilhelm Ostwald’s determination that catalysts only impact the chemical rate of reaction and not thermodynamic equilibrium. He later received the Nobel Prize in Chemistry in 1909\(^2\).

The chemist scientists have divided catalysts into two groups based on depending on their physical point of view;

1.2.2 **Heterogeneous Catalysts**

Heterogeneous catalysts and reactants are found in different phases, as they interact in a cyclical fashion. This interaction results in potentially several atomic or molecular level
reactions, without resulting in the consumption of the catalysts. Heterogeneous reaction rates are influenced by the number of reaction sites available on the catalyst surface to while reactants can adsorb. The binding sites allow for straightforward detachment and reprocessing. Catalysts are often used due to their ability to selectively support specific product creation21.

1.2.3 Homogenous Catalysts

Unlike heterogeneous catalysts, homogenous catalysts and reactants or products interact in the same segment of the reaction compound. Therefore, the beneficial reactant contact allows for a great level of efficiency enabling reactions to occur in mild conditions.

Homogenous catalysts can either be deemed to be organocatalysts or organometallic catalysts in accordance with their individual chemical nature. Enzymatic and biocatalysis are additional potential classifications of homogeneous catalysis.

The past 70 years have seen significant research into homogenous catalysis by transition metal complexes22. G. Wilkinson’s work on alkene hydrogenation using a rhodium-based catalyst triggered the interest in utilizing homogenous catalysis in organic compound creation23.

Despite the fact that the catalytic nature of fermentation processes has been long known, it has only been in the past century that biocatalysis has emerged as a means for imposing chemical alteration on synthetic organic compounds. Employing enzymes to create fine chemicals is a growing area of practice in recent years, particularly within the pharmaceutical industry24.

1.2.3.1 Asymmetric Homogenous Catalysis

When synthetic paths begin in achiral compounds it is known as asymmetric synthesis. Chiral material is an integral aspect of all four major asymmetric synthesis methods (Scheme 1.1)25. Rather than being transferred, chirality is amplified in asymmetric catalysis. This results in significant cost savings during the formulation of enantiomers as compared with stoichiometric approaches.
Homogenous asymmetric catalysts take the following three forms:

- organometallic catalysts
- organocatalysts
- biocatalysts: enzymes or whole cells

Enantioselective synthesis most frequently makes use of organometallic catalysts and organocatalysts\textsuperscript{24,26,27}.

Enantioselective organocatalysis is a parallel approach that has become recently popular. In this approach, a minute chiral molecule is used to mediate the reaction\textsuperscript{28}.

\textbf{Scheme 1.1.} Asymmetric synthesis of chiral compounds.
1.2.3.2 Significance and Production of Chiral Compounds

The geometric property of a rigid object that prevents that object from being superimposed on its reflection, or enantiomer, is known as chirality. The property of chirality is present in all aspects of any biological system, including carbohydrates, lipids, amino acids, and even fully formed organisms. Because all life is built upon molecular chirality, it is thus vital to comprehend and produce chiral processes. This activity is one of the most important and perplexing aspects of organic chemistry.

Thus far, the pharmaceutical industry has paid the most attention to chiral compounds. When viewed in an achiral context, the enantiomers of a chiral molecule are physically and chemically identical. However, the behaviour of those enantiomers can be quite varied in vivo, as living systems are chiral in nature. This results in a range of pharmacologic activity (Scheme 1.2) for an example of two enantiomers with different effects. Other industries also value chiral technology, as it is a part of the production of pesticides, dyes and pigments, nonlinear optical materials, biochemical, aroma and flavour compounds, liquid crystals, polymers, and a range of other matter.

\[
\begin{align*}
\text{C}_6\text{H}_5 & \quad \text{N} & \quad \text{CH}_3 \quad \text{OCOC}_2\text{H}_5 \\
\text{C}_6\text{H}_5 & \quad \text{C}_6\text{H}_5 \\
\text{CH}_3 & \quad \text{CH}_3 \\
\end{align*}
\]

\[
\begin{align*}
\text{C}_6\text{H}_5 & \quad \text{C}_6\text{H}_5 \\
\text{CH}_3 & \quad \text{CH}_3 \\
\text{H}_3\text{C} & \quad \text{C}_2\text{H}_5\text{OCO} \\
\text{C}_6\text{H}_5 & \quad \text{C}_6\text{H}_5 \\
\end{align*}
\]

\[(R,R) - \text{propanol} \quad \text{analgesic} \quad \text{antitussive} \quad (S,S) - \text{propanol} \]

Scheme 1.2. The two different biological activities displayed by two enantiomers of propoxyphene: the (R, R)-isomer in an analgesic, while the (S, S)-isomer has antitussive effect.

A racemic mixture is created when chiral compound enantiomers are combined in equal measure. It is possible to examine the enantiomeric enrichment of a non-racemic mixture with the aid of polarimetric, spectroscopic (NMR) or chromatographic approaches that include a chiral stationary phase. This analysis can be described in terms of enantiomeric excess (ee = ([R] - [S]) / ([R]+[S]) × 100).
A variety of approaches can be used to ready enantiopure compounds. The chemical alteration of “chiral pool” molecules (those chiral molecules organically found in carbohydrates, alkaloids, amino acids and terpenes) is the most widely used of these approaches. Other options include resolution techniques (such as enzymatic kinetic resolution or diastereomeric crystallisation), which are particularly popular for industrial scale use\textsuperscript{29}. Dynamic kinetic resolution techniques have been recently introduced and are notable for their ability to convert a racemic material to a single enantiomer (Scheme 1.3)\textsuperscript{30}.

**Scheme 1.3.** Chemoenzymatic dynamic kinetic resolution of (DKR).

### 1.2.3.3 Transition Metal-Mediated Enantioselective Catalysis

Recent years have seen asymmetric catalysts created from an assortment of chiral ligands and transition metals\textsuperscript{31}. Metal centre permutation is the basis for active and selective catalyst creation. This centre is the cause of activity as it interacts with the substrate. At the same time, selectivity is established by the ligand, which moderates the steric and
electronic characteristics of the catalyst. Some important trends in metal-ligand stabilities are outlined in (Figure 1.2).

![Diagram of metal and ligand bonding](image)

**Figure 1.2.** General styles in $\sigma$-bonding of early and late transition metals to common ligands.

Chemists typically opt to create nature-inspired systems using catalysts when they embark on ligand choice. Contrary to enzyme behaviour, some synthetic catalysts demonstrate high levels of selective and activity for a large panel of substrates and a variety of reactions, as opposed to being restricted to a particular substrate class. These so-called “privileged” ligands are the driving force behind practical catalytic asymmetric process development, and are analogous with pharmaceutical compounds that dynamically interact with a range of biological targets.

First coordination sphere interactions are those that occur directly in the purlieu of the metal. These interactions are typically clearly defined and are the cause of reaction enantioselectivity. However, it continues to be difficult to hypothesize the result of enantioselective catalytic transformation, regardless of the existence of structural, computational and mechanistic research. In addition, anaemic, non-adhesive contacts between the catalytic context and the substrate (such as counter ions, salts or solvent) result in a subsequent harmonisation. Novel and efficacious enantioselective catalysts have been realised as a result of combinatorial approaches invented and sphere implemented in response to this challenge.
1.2.4 Enzymes and Biocatalysts

Enzymes are naturally resourceful energy users; they are highly discerning and organically oriented chemical agents that are frequently used as a model for synthetic catalysts. However, unlike these synthetic agents, enzymes naturally host an energy-efficient second coordination sphere within organisms, in addition to their chiral first coordination sphere. Reaction stereoselectivity of enzymes is bolstered by the catalytic sites that are submerged well within distinct binding pockets, only reached via a constricted channel. The high discernment of enzymes is due to these two characteristics, but they are challenging to reproduce in organometallic catalysis.

The innovation of recombinant DNA technology has long been responsible for enabling the manipulation of biomolecules’ chirality for enantioselective catalysis. This process allows for an efficacious means to create acceptable amounts of laboratory-produced enzymes.

Industrial processes call for a range of biocatalysis uses. This is in spite of the few practical concerns resulting from the production of biocatalysts, which include single enantiomer availability, degree of product/substrate inhibition, high substrate specificity, and instability during particular reaction contexts. Genetic engineering offers today’s practitioners the ability to alter enzyme constructions and purposes in order to build out further industrial uses for enzymes. New applications in the field of organic synthesis have been made possible by recent innovations in biocatalysis promiscuity – the ability of an enzyme to catalyse a reaction that it was not originally intended to accelerate.

It is possible to perform a biocatalytic reaction with a whole cell system or with sequestered enzymes. On occasion, isolated enzymes are favoured despite the fact that whole microorganism alteration technique is able to circumvent the need for protein purification equipment and ensuing co-factor regeneration. This is because isolated enzymes are not associated with membrane permeability or substrate cell toxicity issues, and they result in a reduced number of side products.
Enzymes are considered to possess qualities that are absent in inorganic catalysts; they incur lower production costs, sustainable, environmentally friendly, show great substrate specificity, are highly efficient despite low energy requirements and involve fewer reaction steps that generate by-products\textsuperscript{2,12}. In contrast to non-biological catalysts, biocatalysts act in a narrow manner, unable to compete with the range of reactions of, for example, incorporation of transition-metal complexes.

The specificity of biocatalysts and their homologues can limit their application, which is emphasized by the ability of transition-metal catalysts to convert a wide range of reagents at high rate. This may be a fundamental limitation of biocatalysts; genetically engineered and enhanced mutant enzymes are unable to catalyse all the synthetically significant conversions that transition metals can achieve in organic solvents or aqueous buffer\textsuperscript{2,14}. As a result, a ‘best of both’ concept emerges in the form of artificial metalloenzymes, which would take the form of a protein scaffold supporting a synthetic metal catalyst. Metalloenzymes would combine the versatility of transition metals with the reactivity of enzymes\textsuperscript{5,15}.
1.2.5 Catalyst Performance

Evaluating the quality and performance of a catalyst remains of crucial importance. This may be accomplished by analysing activity and specificity towards substrates and products of a catalyst in a given reaction. As mentioned above, this might be a challenging and inaccurate task. Generally, the activity of a catalyst is defined by the amount of product produced per equivalent of a catalyst in a particular amount of time. Herein several key parameters such as temperature, pH or solvent content are of vast importance and influence the activity greatly. Specificity refers to the bias of a catalyst to convert substrates and/or to form products of a specific stereotype.

Enzyme-catalyzed reactions in which only one substrate is converted, thus e.g. isomerization or cleavage reactions represent the simplest form of enzymatic catalysis. Normally, an irreversible course is assumed. The most enzymes follow the Michaelis-Menten kinetics (Figure 1.3).

\[
\begin{align*}
E + S \quad &\xleftrightarrow{k_1, k_{-1}} \quad ES \quad \xrightarrow{k_2} \quad E + P
\end{align*}
\]

**Figure 1.3. Michaelis-Menten Kinetics**

Herein, the enzyme (E) and the substrates (S) associate in a fast and reversible step to form an enzyme-substrate complex (ES). This complex then undergoes chemical conversion, usually a kinetically slower process, into the desired product (P) and the unreconstructed enzyme (S). Because of its irreversibility the overall rate-determining step is the formation of the product \((k_2 \gg k_1)\). The quantitative relationship between substrate concentration \(\left[S\right]\) and reaction speed \(v\) of a simple irreversible enzyme substrate reaction is represented by the Michaelis-Menten equation, which describes the dependence of the reaction rate, \(k_{\text{cat}}\), of an enzyme reaction on the concentration of enzyme and substrate \((E, [S])\).

\[
v = k_{\text{cat}} [E] \frac{[S]}{K_M + [S]} \quad \text{(Equation 1)}
\]
The so-called Michaelis constant ($K_M$) can be measured experimentally and is related to the dissociation constants. Thus knowing $K_M$ of a certain enzyme allows the study of metabolic reactions. The other constant is $k_{cat}$, the catalytic constant that indicates the conversion rate of the enzyme. However, like every model, the Michaelis–Menten kinetics is limited in this case towards a chiral bias in stereoselective reactions.\textsuperscript{33}
1.2.6 Diels-Alder Catalysis

The Diels-Alder reaction, which can form two carbon-carbon bonds and up to four new stereogenic centers in one step, is a core stone in organic chemistry. Designed stereoselective catalysts for carbon-carbon bond-forming reactions should be broadly useful in synthetic chemistry. Nevertheless, no naturally occurring enzymes have been shown to catalyse a bimolecular Diels-Alder reaction.

The Diels-Alder reaction is a so-called [4+2] cycloaddition. The simplest representative of this type of reaction is the reaction of 1,3-butadiene with ethylene (Figure 1.4).

![Figure 1.4. Diels-Alder Reaction. Mechanism of [4+2] Cycloaddition of 1,3-butadiene and ethylene.](image)

Generally, the mechanism of the Diels-Alder reaction follows pericyclic reactions motifs. The frontier molecular orbital theory dictates that the interaction of the highest occupied molecular orbital (HOMO) of the diene with the lowest unoccupied molecular orbital (LUMO) of the dienophile is the dominant interaction in the transition state. Narrowing the energy gap between the HOMO and LUMO will increase the rate of the Diels-Alder reaction. Lowering the LUMO energy and stabilizing the negative charge accumulating in the transition state by withdrawing substituents can accomplish this. Moreover, if a Lewis Acid, an electron accepting species, is closely oriented a rate enhancement can be observed. In addition to electronic stabilization, binding of the two substrates in a relative orientation optimal for the reaction is expected to produce a large increase in rate through entropy reduction. These observations had lead to a stage for catalyst that could induce asymmetry in this kind of cycloaddition. From synthetic aspects various Diels-Alder variants camp up. An extensive survey of natural [4+2] adducts suggested that more than 400 natural products are potentially bio-synthesized by formal Diels-Alderases. The amount of evidences suggests that nature is able to
provide the mechanistic background for this kind of reaction. But never shows this in an open way. Baker and co-workers designed a Diels-Alderase in 2010 that follows the entropy reduction approach by forcing the two substrates in an energetically favourable orientation\textsuperscript{43}. Following these evidences, a protein with a binding pocket that positions the two substrates in the proper relative orientation in combination with a Lewis Acid is expected to be an effective Diels-Alder catalyst.
1.2.7 The Tsuji-Trost Reaction

The development of asymmetric and symmetric metal-catalyzed reactions has played a crucial role in allowing synthetic access to biologically interesting molecules [61]. Generally speaking, the Tsuji-Trost reaction is a palladium-catalyzed allylation of nucleophiles such as active methylenes, enolates, amines, and phenols with allylic compounds such as allyl acetates and allyl bromides, which has an focus of carbon-carbon as well as carbon-hydrogen/ heteroatom bond formation⁴⁹. This organic reaction is named after Jiro Tsuji, who first reported the method in 1965, and Barry Trost, who further developed it. In 1973 Trost and co-workers introduced phosphine ligands in the reaction and introduced an asymmetric version, which lead to a deep understanding of the reaction mechanism involved⁵⁰. The formation of the alkene with the palladium catalyst leads to a π-complex. In an oxidative additionally step, a η₃-π-allyl complex is formed by inversion and elimination of the leaving group. Depending on the strength of the nucleophile the reaction can take two different pathways: soft nucleophiles (Figure 1.5 [A]) normally add directly to the allyl moiety, whereas hard nucleophiles (Figure 1.5 [B]) first attack the metal center, followed by reductive elimination to give the allylation product.

![Figure 1.5](image)

**Figure 1.5.** The formation of the alkene with the palladium catalyst leads to a π-complex. In an oxidative additionally step a η₃-π-allyl complex is formed by inversion and elimination of the leaving group. Depending on the strength of the nucleophile, the reaction can take two different pathways: A) Soft nucleophiles normally add directly to the allyl moiety; B) hard nucleophiles first attack the metal center, followed by reductive elimination to give the allylation product.
Characterised by its large chemo-, regio- and stereoselectivity the asymmetric Tsuji-Trost reaction, in which metal complexes are used with enantiopure chiral ligands as catalysts, is suitable especially for the construction of stereogenic centers\textsuperscript{50,51}. Currently the aim of present research in palladium-catalyzed enantioselective allylation chemistry regards to the synthesis of natural products\textsuperscript{51}. The proof of this concept took place in 2005 by Trost and co-workers who used this kind of reaction among other reactions as an enantioselective key step in the total synthesis of galantamine and morphine\textsuperscript{52}.
1.3 Metalloenzymes as Catalysts

Metalloenzymes are biological catalysts composed of a proteinic part called enzyme and incorporated metal into a specific site, providing the catalytic activity\(^5^3\). Enzymes are linear polypeptides which, are transformed into complex three-dimensional structures through a process known as folding. This enzymes can catalyze most of the biological reactions that are essential for sustaining life\(^5^4\). From a historical perspective of enzyme activity, Emil Fischer and Koshland proposed the “lock and key” and the “induced fit” models as initial hypotheses as to the way in which enzymes operate\(^5^5\). They postulated that the enzyme’s active site (the lock) allows only a specific substrate (the key) to bind or fit into it, and that the enzyme’s loop movements are responsible for creating a productive enzyme-substrate complex. It is evident that enzymes are well suited to general acid-base and nucleophilic catalysis, but less well suited to electrophilic catalysis and redox reactions.

Enzymes are able to markedly widen their range of reactions by frequently attracting organic and inorganic cofactors to their active site. Organic cofactors, for instance, are typically equipped with functional groups that are able to provide a proteinogenic scaffold with a significant range of useful activities\(^5^4\). The addition of metal ions, by comparison, offers a more adaptable approach than organic cofactors, allowing enzymes to expand their catalytic range\(^5^6\). Furthermore, bioinformatic studies indicate that approximately half of all known proteins carry a metal ion within their scaffold\(^5^7\). A more significant observation is that a third of all enzymes are metalloenzymes; and in general these are biological catalysts containing a proteinic part into which a metal is incorporated, thereby providing the catalytic activity\(^5^3\).

Metalloenzymes bind metal ions into their active site using coordinating side chains or non-proteinogenic, metalated ligands to catalyze a huge diversity of reactions. In particular, metals assign functionality to a protein scaffold by providing strong electrophilic canters, specific ligand affinities (i.e.: O\(_2\), CO\(_2\), or N\(_2\), ATP, and others), multiple stable oxidation states, and various other attributes\(^5^8\). Enzymes employ most of the first row transition-metals, but in addition to these magnesium, calcium, sodium, potassium, tungsten, molybdenum, and even cadmium have all been observed
identified. These metals, with differing impulsions, bond into the entire range of functional groups present in proteins, sometimes with the need for corresponding coordination geometries, therefore the provision of metal ions to the cells is strictly controlled in order to avoid mis-metalation.

Metalloenzymes catalyze a set of reactions that include some of the most remarkable known to chemistry. In nature, Metalloproteins catalyze a wide range of key reactions with high efficiency and selectivity. Some of them are very complex and important such as photosynthesis and water oxidation. Catalyzing metabolic reactions within living organism done under mild condition has attracted much attention and resulted in studies on how metalloproteins work to design new metalloproteins. Although it is much more challenging to design metalloproteins than non-metalloproteins, tremendous improvements have been made, leading to development of novel metalloproteins with tailored properties that cover vast array of functions for biotechnological and pharmaceutical applications.

It is no wonder that, over many years, chemists and biologists have tried to comprehend the reasons behind metalloenzymes’ focused targeting and rapid effectiveness. The unique spectroscopic and magnetic properties of metalloenzymes seen in systematic studies have helped to define some of nature’s approaches for employing metal ions for catalysis. It can generally be stated that all metalloenzymes bind metal ions (or a metal-containing prosthetic group) to their active site with a single labile coordination site. This mechanism specifically defines the location of the catalysis allowing proteins and substrates to combine in the near vicinity of the reactive centre. Metalloenzymes are a feature of all six-enzyme classes, and as discussed earlier, they employ a variety of metals to facilitate transformative reactions.
1.4 Artificial Metalloenzymes
1.4.1 Introduction

In 1978, Wilson and Whitesides presented a pioneering model of artificial metalloenzymes. By introducing a transition metal complex into avidin, a non-catalytic active protein, avidin was transformed into an enzyme with abiotic capability. Since this ground-breaking research, development into artificial metalloenzymes has progressed, but the common aim is to unite the regio- and enantio- selectivity and bioorthogonality of biocatalysts with the range of reactions afforded by transition metals remains constant. The bioscaffold of artificial metalloenzymes delivers the second coordination sphere, which is fundamental to enantioselectivity and metalloenzyme function.

The first coordination sphere is concerned with the chiral ligand coordinated directly to the metal to form a complex that can form two diastereomeric intermediates. In homogenous catalysis, the first coordination sphere is responsible for selectivity and the rate of reaction. The difference in energy between the transition states resulting from diastereomeric intermediates and the enantiomeric products leads to enantiomeric excess. Ligands favor a particular transition state (Figure 1.6). However, in artificial metalloenzymes, it is the second coordination sphere, provided by the bioscaffold, which is of particular interest; it is fundamental to enantioselectivity and metalloenzyme function. It comprises the interactions between the bioscaffold and the substrate, the transition state and the metal complex, thus involves electrostatic, hydrogen and hydrophobic interactions and binding. The amino acids of the bioscaffold can determine the transition state by its interaction with the substrate.

Although artificial metalloenzymes had been realised in the 1970s, it is in the 21st century that their use and research into their applications has vigorously expanded. As the horizons of in vivo chemistry, cascade and catalytic reactions broaden, new challenges emerge.
Figure 1.6. Influence of the $\Delta \Delta G^\neq$ on enantiomeric excess (up) and energetic diagram of an asymmetric hydrogenation reaction (down). Curtin - Hammett Principle states that the product composition is not solely dependent on relative proportions of the conformational isomers in the substrate; it is controlled by the difference in standard Gibbs energies ($\Delta \Delta G^\neq$) of the respective transition states.
Initially artificial metalloenzymes were constructed out of curiosity and their catalytic capabilities were often tested with benchmark reactions. These included simple oxidation reactions, Lewis acid catalyzed Diels-Alder reactions, etc. that proved very effective in proof-of-principle studies.\textsuperscript{49} For an in-depth overview a number of review articles are available.\textsuperscript{3,49} The field is now slowly transforming and more challenging catalytic reactions can be performed, including reaction types that have no equivalent in homogeneous or enzyme catalysis. The next section will describe these advances for artificial metalloenzymes used in enantioselective catalytic transformations of the past few years (Scheme 1.4).
Scheme 1.4. Reaction catalyzed by artificial metalloenzyme⁶⁸-⁷⁷.
1.4.2 Concept of Artificial Metalloenzymes

The combination of transition metal catalysts that are effective in modifying proteins with biomolecular scaffolds, such as protein, that demonstrate high selectivity and activity of enzymes culminates in artificial metalloenzymes\textsuperscript{65,78-82}. By incorporating the transition metal that has catalytic capability into the protein scaffold creates a macromolecular compound, in which the metal is located within a distinct environment. The scaffold provides the second coordination sphere; this discriminates chirality and is central to rate acceleration to enantioselectivity that is observed in numerous catalytic asymmetric reactions\textsuperscript{70,83} (Figure 1.7).

The qualities offered by the combination of transition metal and enzymes have been exploited. Transition metal catalysts are more resilient in extreme temperature and pressure environments than enzymes; they are also easy and cheap to manufacture. Conversely, enzymes function under physiological conditions; they display considerable selectivity and the stability provided by the environment surrounding the active site promotes high product throughput.

![Figure 1.7. Concept of hybrid catalyst](image)
Enzymes and catalysts increase the rate of a chemical reaction\textsuperscript{84} (Table 1.1). Whilst there are advantages and disadvantages for each system, in general enzymes are functional in milder conditions than those demanded of homogeneous or heterogeneous catalysts (Table 1.2). Chirality is presented to catalysts by either chiral metal centres or chiral ligands.

\begin{table}[h]
\centering
\caption{Comparison of chemical and biological systems.}
\begin{tabular}{|c|c|c|}
\hline
\textbf{Reaction} & \textbf{Chemical System} & \textbf{Biological System} \\
\hline
Methane hydroxylation & \( \text{CH}_4 + \text{H}_2\text{O} \xrightarrow{} \text{CO} + 3 \text{H}_2 \) & \( \text{CH}_4 + \text{O}_2 + \text{NADH} + \text{H}^+ \xrightarrow{} \text{CH}_3\text{OH} + \text{H}_2\text{O} \) \\
 & (Ni Catalyst, 700-900 °C, 1-25 bar) & \( + \text{NAD}^\ast \) \\
 & \( \text{CO} + 2 \text{H}_2 \xrightarrow{} \text{CH}_3\text{OH} \) & (Methane Monooxygenase) \\
 & (Cu/Zn catalyst, 250-280 °C, 70-110 bar) & & \\
CO oxidation & \( \text{CO} + \text{H}_2\text{O} \xrightarrow{} \text{CO}_2 + \text{H}_2 \) & \( \text{CO} + \text{H}_2\text{O} \xrightarrow{} \text{CO}_2 + 2 \text{H}^+ + 2\text{e}^- \) \\
 & (Fe/Cu catalyst, >200 °C) & (CO dehydrogenase) & \\
CO insertion & \( \text{CH}_3\text{OH} + \text{CO} \xrightarrow{} \text{CH}_3\text{COOH} \) & \( \text{CH}_3\text{--}[\text{M}] + \text{CO} + \text{HS--CoA} \xrightarrow{} \) \\
 & [Rh(I)(I)(CO)]\text{], 120 °C, 30 atm} & \text{CH}_3\text{(CO)}\text{--S--CoA + H}^\ast + [\text{M}] \text{[a]} & \\
\hline
\end{tabular}
\end{table}

\textsuperscript{[a]} \text{CH}_3\text{--}[\text{M}] \text{ is a corrinoid-iron-sulfur protein that acts in the reaction as a methyl group doner; HS-CoA is the coenzyme}
### Table 1.2. Comparison between homogeneous and enzymatic catalysis.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Homogeneous catalysis</th>
<th>Enzymatic catalysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate scope</td>
<td>Large</td>
<td>Limited</td>
</tr>
<tr>
<td>Enantiomers</td>
<td>2 enantiomers</td>
<td>1 enantiomer</td>
</tr>
<tr>
<td>Tolerance to organic solvents</td>
<td>Excellent</td>
<td>Limited</td>
</tr>
<tr>
<td>Substrate concentration</td>
<td>High</td>
<td>Low to moderate</td>
</tr>
<tr>
<td>TON/TOF</td>
<td>Average</td>
<td>High</td>
</tr>
<tr>
<td>Optimization</td>
<td>Chemical synthesis</td>
<td>Genetic</td>
</tr>
<tr>
<td>Repertoire of reaction</td>
<td>Large</td>
<td>Restricted</td>
</tr>
<tr>
<td>Second sphere coordination</td>
<td>III-defined</td>
<td>Well-defined</td>
</tr>
<tr>
<td>Reactions condition</td>
<td>High temperature and pressure</td>
<td>Room temperature and atmospheric pressure</td>
</tr>
<tr>
<td>Environmental impact</td>
<td>Hazardous</td>
<td>Friendly</td>
</tr>
</tbody>
</table>

### 1.4.3 Design of Artificial Metalloenzymes

Three features are considered in designing artificial metalloenzymes: the biomolecular scaffold, the transition metal catalyst and the method that conjoins the two. In constructing artificial enzymes, the following factors need to be evaluated. The appropriate transition metal needs to be selected according to the reaction that is to be catalysed. Similarly, the bioscaffold that will provide the environment for the transition metal and the second coordination sphere needs to be determined. These need to be brought together; therefore, the factor of anchoring the transition metal catalyst to the bioscaffold to form a stable complex needs to be devised.

#### 1.4.3.1 Transition Metal Catalyst

The activity of an artificial metalloenzyme is determined by the transition metal. The latest development in the discipline of transition metal catalysed reactions has significantly opened up possibilities for the application of technology\(^{85}\). However, it is
not without limitations as the bioscaffold may become unstable under certain temperature and pressure conditions or in the presence of reagents such as surfactants and oxidising agents. Water compatibility is a significant limitation of organometallic catalysts.

1.4.3.2 Biomolecular Scaffold

The bioscaffold is central to the functionality of artificial metalloenzymes as it provides the second coordination sphere. The significance of this function cannot be underestimated and determines the selection of an appropriate bioscaffold. The bioscaffold must be stable within the pH and temperature ranges to which it will be exposed. It must be resilient to the organic solvents used in catalysis. A separate point to consider is whether the scaffold has a suitable pre-existing active site or whether one needs to be constructed.

Creating artificial metalloenzymes using bioscaffolds with a pre-existing active site has the advantage of there being a concomitant second coordination sphere and stability projections can be made based upon the enzyme in an unbound state (i.e. apo form). But one considerable drawback of using such bioscaffolds is that the binding site needs to be of a size that can house the transition metal catalytic complex as well as accommodate the resulting reactants. Although there are bioscaffolds able to undertake enantioselective catalysis, such as avidin/streptavidin\textsuperscript{65,79} bovine serum albumin (BSA)\textsuperscript{86}, (apo) myoglobin\textsuperscript{87} that have been successfully used to create artificial metalloenzyme, typically an enzyme’s active site is suited to only accommodating the substrate and product. Artificial metalloenzymes are most frequently created from bioscaffolds that have an active site that is amenable to modification and can be functionally expanded.

A different approach is to create an active site in its entirety in a bioscaffold where no active site is naturally present. An advantage of this method includes increasing the capability of the bioscaffold so that the second coordination sphere can be tailored to the requirements of the reaction being catalysed. Furthermore, notwithstanding the sizable challenge of creating \textit{de novo} protein scaffolds using \textit{in silico} techniques, creating a new active site allows many more potential scaffolds to be devised. However, to date no artificial metalloenzyme has been created from scratch. Although remodelling of a non-haem iron binding site has given rise to an artificial nitric oxide reductase\textsuperscript{88}. Based on the
findings of an X-ray structure of myoglobin protein, three new residues (one glutamate and two histidines) were introduced into the distal pocket of the protein containing a haem iron complex. X-rays of the predicted and experimental proteins were overlaid to confirm the model’s accuracy\textsuperscript{88}.

An alternative method to creating a new active site where none is present is to use DNA as the scaffold for artificial metalloenzymes as reported by Roelfes. In this instance a Cu(II) complex was supramolecularly anchored in the chiral groove of DNA. Peptides can also be a viable scaffold\textsuperscript{89,90}.

Although proteins are comprised of amino acids, the short chains of amino acids that form peptides, peptides are more flexible than proteins when it comes to designing new active sites. Despite their diminutive size, they are able to create the functional diversity needed for secondary coordination spheres. Proof of concept has been demonstrated by Krämer whereby DNA scission was achieved using a peptide nucleic acid conjugated with a Zr(IV) complex\textsuperscript{91}.

The chiral induction of artificial metalloenzymes is provided by the second coordination sphere interactions of the protein. Therefore, the artificial metalloenzymes target controlling the second coordination sphere that is responsible for chiral discrimination to create transition metal complexes to display enzyme-like selectivities\textsuperscript{82}. Further, co-activation of the substrate by the functionalities of the host protein may result in enzyme like activities. Generally, design of the artificial metalloenzyme uses the native biomolecule scaffold is more simple and favorable rather than the \textit{de novo} design. Duo to the already existing thermodynamic stability and tolerance for mutation required for creation the active transition metal center. Also, attachment or anchoring strategies of the transition metal catalyst to the biomolecule has to be consider as key parameter in this design\textsuperscript{82}.

### 1.4.3.3 Anchoring Strategy

It is essential that the transition metal is not randomly localized within the bioscaffold but anchored appropriately. This can be achieved by several means that are outlined below (Figure 1.8).
Figure 1.8. Representation of concept of artificial metalloenzymes and the different anchoring strategies; A) Supramolecular anchoring, B) Covalent anchoring, C) Direct anchoring and D) Unnatural amino acid anchoring. M the catalytically active transition metal.

A. Artificial Cofactors Anchoring

A series of weak interactions i.e. electrostatic interactions, hydrogen bonds and hydrophobic interactions are non-covalent mechanisms that can anchor the transition metal to the bioscaffold. However, supramolecular interactions can be constraining if binding does not leave adequate space for the reaction’s reactants. To maintain selectivity, catalysis must occur within the scaffold therefore the transition metal complex must bind strongly to localise the activity. Then again, the artificial metalloenzyme is generated through self-assembly and no modification is needed following biosynthesis. The organometallic complex is drafted to the scaffold protein by being attached to a small molecule that has an uncommonly high affinity for the hosting scaffold protein. The key feature that defines the rate of success is the small molecule’s affinity for the bioscaffold.

An often cited example of non-covalent anchoring is that of (strep)avidin that binds biotin with great affinity proteins known so far and specificity, reaching 90% stereoselectivity. The first artificial metalloenzyme dependent on enantioselective hydrogenation was demonstrated by Whitesides, who exploited avidin with a biotinylated Rh-negative bisphosphine compound non-covalently anchored in the active site. However, the ee only reached 41%\(^65\). Research continued with Ward switching to using streptavidin scaffolds, but still using biotinylated metal compounds. This change of scaffold resulted in an improved hybrid catalyst and went on to spawn other streptavidin-based artificial metalloenzymes\(^92\).
Another approach to create ArMs is the use of electrostatic interactions. Replacing ‘natural’ cofactors with synthetic metal complexes is an alternative method of non-covalently attaching a metal compound, whilst utilising the native active site. This has been demonstrated with the haem protein, myoglobin. Research by Watanabe successfully incorporated Cr-salen, Mn-salen and Ru-phebox complexes into apomyoglobin. Gross and co-workers used the high affinity of serum albumins for anionic porphyrins to create ArMs. To overcome the highly specific protein-substrate interaction limitation, Eppinger and co-workers introduced a new concept for ArMs, which uses specially designed metal-conjugated affinity labels to introduce catalytically active transition metal complexes inside the binding pockets of enzymes.

**B. Covalent Anchoring Strategy**

In contrast to supramolecular bonding where no covalent bonds are formed, covalent anchoring is dependent upon the formation of a covalent bond between the transition metal catalyst and one or several amino acids of the bioscaffold. For the chemical bond to form between the two entities the bioscaffold has to be modified post-biosynthetically. For covalent binding to occur an amino acid in the bioscaffold such as cysteine that acts as a nucleophile reacts with an electrophilic counterpart in the ligand. This reaction needs to be bioorthogonal, therefore the reaction partners must be orthogonal to the bioscaffold.

Alkylation of a cysteine residue is the bioconjugation method that is used most frequently to create artificial metalloenzymes. If a cysteine residue is not already present in the appropriate place, one is introduced by mutagenesis. To the ligand, various reactive electrophilic groups may be attached. In constructing the artificial metalloenzyme a suitable electrophilic group is bestowed on the organometallic compound; this is then covalently linked the thiol group of the cysteine. This method of covalent anchoring can be used with a number of diverse protein scaffolds.

For instance, bioconjugation can arise when the nucleophilic cysteine reacts with α-halogenated carbonyl compounds such as iodoacetamide-functionalised phenathroline. Distefano conjugated this compound to adipocyte lipid binding protein (ALBP) and complexed with Cu(II). This compound catalysed amide bond hydrolysis in a number
of unactivated amino acid esters, generating ee of 86%. Reetz et al. used malemide, a
different category of electrophiles that is frequently used for cysteine alkylation14,69,77. In
that study, an artificial metalloenzyme was manufactured using a malemide
functionalised phenanthroline that had been anchored to temperature stabilised tHisF14.

A double-anchoring method was used by Lu et al. to complex Mn(salen) with
apomyoglobin99,100. Site directed mutagenesis was used to incorporate cysteine into the
scaffold then introducing methane thiosulfonate to form a disulfide bridge. When the
resultant artificial metalloenzyme was deployed in an enantioselective and
chemoselective sulfoxidatin of thioanisole the product was not over-oxidised and ee
reached 51%99.

Experimenting with the position of the cysteine residue and the subsequent disulfide
bridges the ee was raised to 66%100. This increase in selectivity highlights the
significance of appropriately locating the metal compound within the bioscaffold. These
artificial metalloenzymes held the record amongst all covalently anchored artificial
metalloenzymes for achieving the highest ee in catalysis.

One drawback of alkylating a selected cysteine residue is that it prevents other cysteine
residues from being incorporated into the scaffold; if present, they cannot contribute to
the rate reaction69,101. The issue can be sidestepped by using mutagenesis to remove
undesired cysteine residues but this can compromise the structure of the bioscaffold.

C. Direct Anchoring

Amino acids of the bioscaffold bearing N, O and S functional groups can act as a ligand
for the metal complex, forming a direct anchor. This method is employed to bind the
catalytically active metal ion in naturally occurring metal-binding enzymes102. A tactic
that has been used to create hybrid catalysts by substituting the active site’s native metal
ion with a non-native metal ion. Hybrid catalyst artificial metalloenzymes have been
evaluated for their hydroformylation103, hydrogenation and sulfoxidation104.

Kazlauskas et al. synthesised a peroxidase by coordinating Mn (II) to the three histidine
residues that normally anchor the native Zn(II) in human carbonic anhydrase (hCA). Up
to 67% ee was achieved in the epoxidation of a range of styrene substrates, indicating fair enantioselectivity\textsuperscript{105}. Yet this approach of using natural metal binding sites is not without its drawbacks. Different metals are coordinated by different residues, which limits which metals can be used and in turn the catalytic capability. This limitation can be overcome by synthesising metal binding sites in proteins that do not naturally bind metal thus increasing the potential of coordination points. Dative anchoring in non-native metal binding proteins has been boosted by the recent development of computational algorithms that can determine metal binding sites in a specified bioscaffold\textsuperscript{106,107}. Indeed, copper, iron and zinc have been successfully fitted to geometries that have been computationally arranged\textsuperscript{88,108}.

A Cu(II) binding site was inserted in an appropriate location in tHisF, a temperature stable protein, by Reetz et al. They created a triad comprising two histidine and one aspartic acid residues. The modified tHisF was catalytically active, able to perform cycloaddition of the Diels-Alder reaction, though the enantioselectivities were up to 46% $ee$\textsuperscript{109}. Rhodium substitution of zinc in carbonic anhydrase has also been achieved, this time by Kazlauskas. The resulting artificial metalloenzyme was hydroformylation and hydrogenation capable\textsuperscript{68,110}. 

D. Non-natural Amino Acids anchoring

Manipulating natural, pre-existing features has advantages, but limitations are present in all the methods. An alternative approach is to generate and integrate unnatural amino acids with metal ion binding capability. This is achieved by using the amber stop codon technique. The resulting artificial metalloenzymes can be applied to enantioselective catalytic reactions in water.

The variety of amino acid side chains has enabled natural proteins to develop a vast range of enzyme reactions that facilitate the multitude of catalytic reactions required for life. Methods that further widen the range of enzymes include raising the number of amino acids from the natural 20 by genetically encoding unnatural amino acids\textsuperscript{111}, by recruiting both inorganic and organic compounds to the reactive center\textsuperscript{54} and creating reactive function by post-translation modification of the enzyme\textsuperscript{112}.

Whether during or after protein biosynthesis, integrating a metal binding group into the bioscaffold has a number of benefits. In combination these make using unnatural amino acids to create and improve artificial metalloenzymes particularly appealing. In contrast to noncovalent anchoring methods, there is no requirement for specific ligand binding interactions. Furthermore, unlike modified anchors that depend on covalent and dative methods, no chemical adjustments or refining stages are required as there can be almost complete control as to the location of the metal complex within the protein. Not least amongst the benefits of using unnatural metal binding amino acids is that compared to natural residues, they confer superior control over the first coordination sphere\textsuperscript{109}.

An example of success with unnatural amino acids comes from Schultz et al\textsuperscript{113}. The unnatural amino acid bipyridyl-alanine (BpyAla) was inserted into a protein with a Z-domain to create a Cu(II) artificial metalloenzyme that could selectively cleave DNA\textsuperscript{113}. Hydroxyquinolinyl-alanine (HQ-Ala) is another unnatural amino acid that has been inserted into a Z-domain protein\textsuperscript{114}. When ligated with zinc ions, as well as being a metal binding site, the HQ-Ala became a fluorescent probe enabling phasing in X-ray structures. Other artificial metalloenzyme manufacturing studies have used 4-Azido-L-phenylalanine to anchor covalently metal binding ligands using a strain-promoted azide–
alkyne cycloaddition (SPAAC) reaction\textsuperscript{115}. However, the metalloenzyme product had limited capability as it demonstrated low rates of conversion and enantioselectivities in a dirhodium-catalysed carbene insertion reaction.

Using the expanded genetic code concept \textsuperscript{16,17} to create unnatural amino acids for use in exclusive anchoring points in the bioscaffold has been validated by Lewis et al. \textsuperscript{15} The unnatural amino acid, 4-Azido-L-phenylalanine was introduced into thermostable tHisF and a number of ligands were incorporated with strain-promoted azide–alkyne cycloaddition (SPAAC). The metalloenzyme end product was capable of catalyzing rhodium cyclopropanation and Si-H insertion reactions, but the metalloenzyme’s activity was below that of the isolated metal complex and enantioselectivity was absent.

1.5 An Overview of History and Applications of Site Directed UAA Incorporation
Unnatural amino acids (UAA) are the non-canonical amino acids, which are not encoded in the genetic code and have must be synthesized chemically. These UAA fulfill a variety of roles, they represent an enormous amount of diverse structural elements for the development of new leads in peptide and non-peptide compounds (Fig.5). Also, due to their unlimited structural diversity and functional versatility, they are widely used as chiral building blocks and molecular scaffolds in constructing combinatorial libraries\textsuperscript{116,117}. The ability to introduce unnatural amino acids into proteins beyond the common 20 canonical amino acids enabled the expression of proteins containing amino acids with novel side chains. Including, fluorophores, metal ion chelators, photocaged and photocross-linking moieties, uniquely reactive functional groups, and NMR, IR, and x-ray crystallographic probes\textsuperscript{118}.

Two additional amino acids, selenocysteine (Bock et al., 1991) and pyrrolysine (Srinivasan et al., 2002), are found to be incorporated into proteins cotranslationally and they are thus regarded as natural expansions of the genetic code. These amino acid derivatives have been identified in regular proteins, but none of them are post-translational modifications of the canonical amino acids (Uy and Wold, 1977).
Selenocysteine is incorporated cotranslationally into some oxido-reductions enzymes in eubacteria, archaebacteria and eukaryotes by the opal termination codon UGA. In 1986, the nucleotide sequences were reported for some gene that contain an in-frame UGA codon within their coding sequence at the exact position, which was occupied by a selenocysteine residue in the protein sequence. Since then selenocysteine is considered as the 21st amino acid\textsuperscript{111,119}.

Pyrrolysine is a non-standard amino acid also named as the 22nd amino acid, which is a derivate of lysine. It is encoded by the Amber stop codon TAG. Pyrrolysine is found to be incorporated co-translationally in certain eubacteria and archaea specifically in methanogenic archaea, at the active site in the enzyme methylamine methyltransferase\textsuperscript{120}.

Both non-standard amino acids (selenocysteine and pyrrolysine) are incorporated into a protein by translation of mRNA in same way as amino acids elongation processes occurs in the typical protein biosynthesis. It has been demonstrated that the replacement of a catalytically active cysteine residue by a selenocysteine can successfully improve enzymes kinetic efficiency\textsuperscript{119}. Studies of \textit{Escherichia coli} formate-dehydrogenase showed that the presence of selenocysteine in the preserved active site is essential for formate oxidation. Replacement by cysteine decreased the turnover number of the enzyme by over two orders of magnitude. Additional UAAs represent a wide range of structures and functions not found in the 20 amino acids. They provide new opportunities to generate proteins with enhanced or novel properties and probes of protein structure and function\textsuperscript{121}. Therefore, tremendous efforts have been made regarding the incorporation of UAA into proteins to introduce new functional groups. This can be achieved through chemical or biosynthetic approach, which can be developed for vitro and vivo incorporation\textsuperscript{121}.
1.5.1 Site Directed UAA Incorporation

Specific and directed changes in plasmid DNA is achievable through site-directed mutagenesis (SDM). An extension of this well-established technique is site directed unnatural amino acid (UAA), which integrates UAA at the restriction site. Schultz et al. introduced the amber stop codon suppression method\textsuperscript{118}. This technique, which is alternatively called the expanded genetic code methodology, depends upon orthogonal pairs of tRNA/aminoacyl-tRNA synthetase (aaRS) to integrate unnatural amino acids that do not form proteins at particular locations in response to the amber stop codon. Following this work, \textit{E. coli}, mammalian cells and yeast have had unnatural amino acids introduced into their genetic structures to create new protein structures, functions and uses\textsuperscript{119}.

Situating amino acids to provide the right coordination environment for metal ions is a tricky issue, but can be bypassed by using the expanded genetic code approach that introduces genetically encoded metal ligands\textsuperscript{117}. Artificial nuclease, electron-transfer or biophysical probes, purification tags are some examples of applications that have used proteins with metal binding amino acids\textsuperscript{114,115,122-124}.

1.5.2 Expanding Genetic Code, Orthogonal aaRS/tRNA

The development of \textit{in vivo} methods by re-programming the cellular biosynthetic machinery to introduce novel amino acids eliminated issues related to scalability and protein size and simplifies the study of modified proteins in living cells. This method is considered to be the breakthrough for expanding the genetic code beyond the canonical 20 amino acids. It is applicable for living cells and involves the use of aminoacyl-tRNA synthetase (aaRS) to incorporate UAA that are in close structural analog to the original amino acids at a specific site in the host protein\textsuperscript{116}. The tRNA is constructed orthogonally and suitable for effective \textit{in vivo} incorporation of the respective UAA\textsuperscript{125}. Currently over 80 UAAs were successfully introduced by this technique. The ability to introduce unnatural amino acids site-specifically drives the design of proteins with enhanced properties. For examples incorporating UAA providing a novel functionality can be used to carry out "physical organic chemistry" studies of protein structure and function, like elucidation of crystal structures, NMR spectra, photocrosslinking to study protein-
protein/protein-peptide interaction, allowing the observation of chemical shift changes upon binding with ligands, SUZUKI coupling, protein purification and protein marker as small fluorescent UAAs, redox-active reagents, and probes of hydrogen bonding and packing interactions in proteins\textsuperscript{116,118}.

This directed incorporation relies on the generation of functional aaRS/tRNA pair, which is an orthogonal pair. Therefore, the aminoacyl tRNA synthetase (aaRS) does not cross-react with or load any of the endogenous tRNAs in the host organism. That was achieved by using site-directed mutagenesis technique where the amber TAG stop codon can be reprogrammed to site-specifically introduce a desired UAA\textsuperscript{118,126,127}. The amber stop codon TAG is the least used among the three stop codons in \textit{E. coli} and yeast, and it rarely terminates essential genes and is efficiently translated by amber suppressor tRNAs \textit{in vivo} and \textit{in vitro}. Thus, the use of the TAG codon will not significantly change or alter the growth of the organism. However, other coding techniques have been applied including opal TGA and 4-base AGGA decoding pairs to allow incorporation of multiple UAA into the same protein to facilitate the evolution of new aaRSs specific for UAA with increasing structural diversity\textsuperscript{126,128}.

A suitable aaRS/tRNA pair is typically selected by directed evolution using a double-sieve selection process\textsuperscript{118,126}. This directed evolution approach was developed to change the specificity of the orthogonal synthase to be charged with an UAA. Large libraries of synthetase variants were passed through a series of strict positive and negative selections. Initially the library was generated by mutating five residues of the amino acids randomly at the binding site of TyrRS. Variants were screened to specify UAAs as followed. Positive selection is based on chloramphenicol resistance, which is conferred by the suppression of an amber mutation at a permissive site in the chloramphenicol acetyltransferase gene only in the presence of the unnatural amino acid. Negative selection uses the toxic barnase gene with amber mutations at permissive sites and is carried out in the absence of the unnatural amino acid to eliminate aaRS mutants that aminoacylate endogenous amino acids. Repeating this two-step process of positive and negative selections resulted in a synthetase aaRS that uniquely recognizes the UAA of
interest. This approach should be, in theory, make translations with UAAs accessible in any organism.

**Expanding the genetic code by** the orthogonal approach is the basic method used in this project, through a suppressor tRNA targeting the amber stop codon TAG. The system for UAA incorporation and the modified host protein (mTFP* TAG-mutant) are on two different plasmids. The synthesized UAAs were added to a culture of *E. coli* which contains these two plasmids: 1) the pEVOL-vector coding for an tRNA/aaRS orthogonal pair was optimized for site specific incorporation of the respective amino acid, which has been developed by Peter G. Schultz and his group\textsuperscript{125}, and 2) a vector coding for the modified host protein. After resorption of the UAA, the reprogrammed translation *E. coli* machinery will incorporate the UAA at the site specified by the amber codon TAG. The constructed metal enzyme hybrid MEH can be optimized by directed or rational guided evolution cycles based on a molecular understanding of the catalytic mechanism.

*In vivo* Incorporation of artificial active centers suffers from intrinsic problems, since an uncontrolled new activity may either be harmful to the cell or could be eliminated by the biochemical machinery of a cell. Using expanding genetic code method, a new dimension of expression the host proteins with the new functionalities will be introduced with posttranslational modification can circumvent this problem. Synthesized unnatural amino acids with metal binding or posttranslationally addressable side chains for metal coordination will be directed to a site specific incorporation into host protein.
1.6 Optimization

An advantage of artificial metalloenzymes is that the component bioscaffold and metal complex can be independently optimised; this is the chemogenetic approach\textsuperscript{129}.

The metal complex can be improved through rational design. Synthesis is simplified by not needing a chiral form of the metal complex, which differs to the metal complexes of homogenous catalysts.

The bioscaffold can be optimised by different methodologies including empirical, theoretical and semi-theoretical, though to be effective these do demand some knowledge of the scaffold structure\textsuperscript{130}. The X-ray structures of non-functionalised proteins or homology models that incorporate mutation can be visually inspected. Computer algorithms can determine the impact of active site mutations. The semi-theoretical method is a combination of the above empirical and theoretical methods. Direct evolution of the bioscaffold allows for optimisation dispenses without bias.

Ward et al. used a semi-theoretical method to enhance the streptavidin/biotin artificial metalloenzyme in an asymmetric transfer hydrogenation reaction (ATHse). Ward used computer modelling in this semi-theoretical approach to insert into the active site a histidine residue to coordinate with either iridium or rhodium\textsuperscript{131}. The results were verified using X-ray structures enabling the researchers to refine the model where necessary.

Error prone polymerase chain reaction (PCR) offers an example of enhancement derived through directed evolution. The random mutations inserted into the bioscaffold are then screened for catalytic activity and high-throughput screening is used to find positive hits. Artificial metalloenzymes produced using non-covalent or dative anchors are quicker to enhance as the bioscaffold does not require functionalization, a process that is time burdensome. Nevertheless, to take advantage of random mutation of the bioscaffold a large library of randomly distributed mutations has to be built up. The larger the bioscaffold, the larger the corresponding library has to become; this can be challenging to screen as high throughput gas chromatography or high performance liquid chromatography is needed to screen the catalyzed reaction for enantiomeric excess. This is because no colour assays are available for those optimisations. This drawback led to a
new method being devised, Combinatorial Active-site Saturation Test (CAST), the attention of which is the active site residues, leading to a focused library. CAST methodology has been effectively used to augment a streptavidin/biotin artificial enzyme, raising the $ee$ from 23% to 65% over three rounds of mutagenesis$^{132}$. 
1.7 Objective and Pre-considerations

Artificial metalloenzymes hold the promise of extending the scope of biocatalysis. However, despite the increasing research activity in the field, no application of artificial metalloenzymes outside a research laboratory exists. The aim of this thesis is to advance the field by exploring the creation of metal sites through the incorporation of UAAs. If the resulting artificial metalloenzymes are catalytically activity, the protein will provide a chiral environment, which should lead to asymmetric induction. To achieve this goal, the main objectives of this thesis are:

- To generate a fundamental understanding of strategies that can be used to optimize host proteins for artificial metal binding sites.
- To design novel metalloproteins by site-directed incorporation of UAAs.
- To screen and optimize bioconjugation of cofactors at UAA sites.
- To study the metal binding properties of the created artificial metalloenzymes.
- To investigate the catalytic activity and asymmetric induction of the generated artificial metalloprotein in suitable testreactions (Tsuji-Trost allylic alkylations and Diels-Alder cyclizations).
1.8 Reference:

90. Ball ZT. Designing enzyme-like catalysts: a rhodium (II) metallopeptide case study. Accounts of chemical research 2012;45(2):560-570.
97. áde Vries JG. Merging homogeneous catalysis with biocatalysis; papain as hydrogenation catalyst. Chemical communications 2005(45):5656-5658.
2 Chapter 2 A Protocol for High Throughput Fluorescent Proteins Purification

2.1 Abstract:
Fluorescent proteins (FPs) are fundamental for many biological and biomedical applications. Therefore, considerable efforts have focused on the development of efficient and scalable purification protocols. The majority of FPs purification protocols are, however, time consuming, require extensive lab work and suffer from low yield and poor purity and also require tags or some other earlier modification the desired protein. Here, we established a combined protocol that merges the advantages of heat and 3/Phase partitioning purification methods (T/3PP). We demonstrated the applicability for a broad range of FPs [eCGP123 (green), Sf_GFP (green), mVenus (yellow), Tag RFP (red), mKate2 (purple), mTag BFP2 (blue), mTFP1 (cyan), and mTFP1K183 PIF (cyan)] originating from different sources. All the tested C-terminal His-tagged FPs were expressed in Escherichia coli BL21(DE3) using pET303 expression system. Using our T/3PP method, the yield and purity of proteins are superior to other recently described methods. The T/3PP method requires little specialized equipment and provides a facile and inexpensive procedure for purification of large amounts of very pure FPs. The method can be parallelized and allows purification of FPs in analytical, preparative and large-scale procedures.
2.2 Introduction

Fluorescent proteins (FPs) are among the most widely used proteins in biological and medical research\(^1\). Almost all the FPs are around 25kD in size, consisting of eleven \( \beta \)-sheet parral around central alpha helix\(^2\). Several reports have shown that different fluorescent proteins work as a perfect \textit{in vivo} reporter molecules in biological research\(^1,3\). Green Fluorescent Protein (GFP), a 238 amino acid protein, originally isolated from the jellyfish \textit{Aequorea Victoria}, is a great interest to life scientists. GFP and its variants have remarkable advantages over artificial dyes such as high stability, non-toxicity and renewability. Also, GFP chromophores maturation doesn’t require any additional cofactors or enzymes other than molecular oxygen\(^2\) see (Figure 2.1). FPs are used for a wide range of experimental applications, such as markers for protein localization and folding\(^4\), monitoring of promoter activity and gene expression including analysis of cell type-specific genes, as biosensors\(^5\), detection of the pathogen or disease development, DNA- immunization and gene therapy\(^6\), redox potential and metal/halide ion concentrations, and as probes for protein–protein interactions, utilizing both fluorescence complementation and Fo¨rster resonance energy transfer (FRET) between spectrally-compatible fluorescent proteins (FPs), and other uses\(^7,8\). Therefore, there is an enormous interest and demand for the development of a highly efficient, easy, rapid and convenient approaches to express and purify theses FPs. FPs and its variants have been expressed in a variety of cells and organisms\(^9\), FPs can be introduced into mammalian and other cells using the appropriate vector (usually a plasmid or virus). Biochemical studies and immunological assays require high-purity of fluorescent proteins. The majority of current purification techniques of GFP and other FPs are labor-intensive and time-consuming and require expensive chromatography steps, such as size-exclusion chromatography (SEC), ion-exchange HPLC\(^10\); immobilized metal affinity expanded bed adsorption (IMAEBA) using His-tag proteins and HPLC-SEC\(^11\); hydrophobic interaction chromatography\(^12,13\); expanded bed adsorption column using Q Hyper Z matrix\(^14\); salt-promoted immobilized metal affinity chromatography (IMAC)\(^15\); chromato- focusing with a pH gradient composed of multiple stepwise fronts\(^16\); SEC and concentration by hydrophobic chromatography or ultrafiltration\(^17\); expanded bed adsorption chromatography and protein refolding using a solid phase artificial chaperone\(^18\), and annular chromatography\(^17\).
Recently, some investigators have turned their attention to a relatively new protein purification approach called "three phase partitioning" (TPP)\textsuperscript{19}. TPP is a relatively simple and low-cost purification technique that has been successfully used to purify a range of proteins, including GFPs\textsuperscript{19,20}. However, moderate purity and yield obtained by 3PP require further purification. Basically, this method uses a combination of ammonium sulfate and t-butanol to precipitate proteins from their crude extracts. The binding of t-butanol to hydrophobic GFP regions in the presence of a high concentration of (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} leads to protein precipitation into the interphase between the organic and aqueous phases\textsuperscript{21}. However, in this case, the protein needs to be subjected to further fractionation by HIC—affinity chromatography\textsuperscript{22}. The biggest advantage for protein purification is the production of tag free proteins with high purity and activity on a large scale. However, the majority of present purification procedures include time-consuming multiple chromatography steps with a low yield of the desired product or require tag-containing proteins. Here, we describe a new protocol for FPs purification without affinity extensions by using heat and organic extraction. Eight commercially available FPs namely, eCGP123 (green), Sf_GFP (green), mVenus (yellow), Tag RFP (red), mKate2 (Purple), mTag BFP2 (blue), mTFP1 (cyan), and mTFP1K183 PIF (cyan) were used for testing the efficiency of the described method.

Recently, it has been suggested to use the FPs as photovoltaic and sensitized solar cells\textsuperscript{23}. Pure PFs can be employed in dye-sensitized photoelectrochemical cells (PECs)\textsuperscript{24}. Fluorescent proteins (FPs) are employed as the dye molecules that can easily self-assembled dyes obtained from nature source and used as sensitizers in DSSCs\textsuperscript{25}. Dye-sensitized solar cells (DSSC) have attracted worldwide attention in recent years because of their low production cost and a potential alternative to traditional photovoltaic devices\textsuperscript{26}. It has been reported that the incorporated fluorescent proteins into the DSSCs were able to improve the absorption spectrum of the cells, and clearly they have the potential to improve the standard Dye-Sensitized Solar cell\textsuperscript{23}. FPs researches and experiments will always be the potential applications and identifies for future challenges and opportunities.
Figure 2.1. A) Maturation pathway of Green Fluorescent Protein GFP (Ser-Tyr-Gly). B) Chemical structures of chromophores of various fluorescent proteins used in this study. Chromophores are highlighted according to the color of their emission frequency.
2.3 Results and Discussion

2.3.1 Cloning and Expression of Fluorescent Proteins

Fluorescent proteins (FPs) are relatively small proteins, 230-240 aa (30kDa). FPs lack any post-translational modifications, except for the autocatalytic formation of their chromophore and can be easily expressed in a native form in *E. coli*. We successfully expressed eight FP genes coding for eCGP123, Sf_GFP, mVenus, Tag RFP, mKate2, mTagBFP2, mTFP1, and mTFP1K183 PIF proteins in their mature form under the control of the T7 promoter/lac operator system (Figure 2.2).

To optimize expression conditions, we transformed each FP gene into three different pET plasmids (pET303, pET302, pET45b) as a vector and *E. coli* BL21(DE3) as a host. The transformed *E. coli* BL21(DE3) cells were optimized for growth temperature, incubation time and IPTG concentration and induction temperature (ranging from 20 °C to 30 °C) (Figure 2.3A, B, C). The fluorescent intensity of cells expression product shows that the level of expressed FP is varied over time. This screening study indicates that the pET303 vector and lowering the temperature to 20 °C resulted in a significantly increased amount of fluorescent protein with mature chromophore. Together with the incubation time of 24 h at 20 °C after 1mM IPTG addition proved to be fairly sufficient for high expression and faithfully reflects the expression and growth. These conditions were applied for the expression of all FPs described here. The expression yield of the targeted FPs could be readily quantified using fluorescence of cell suspensions. As shown in (Figure 2.3D), all the FPs expressed with noticeable fluorescence, and PET303 plasmid was found to result in the maximum protein and fluorescence intensity. All colored bacterial pallet were collected and washed with saline buffer, and fluorescent intensity was measured to quantifying the protein yield.
Figure 2.2. pET303/FPs vector. FPs expressed under the control of T7 promoter with XbaI and XhoI restriction site and C-terminal His-tag.
2.3. Fluorescent intensity screening of FPs expression under different conditions using pET303, pET302, and pET45b vectors. A) Effect of the temperature changes on the pET303, pET302, and pET45b expression system, with a fixed IPTG concentration (1mM) and 24h incubation time. B) Effect of the IPTG concentration on the pET303, pET302, and pET45b expression system, with a fixed temperature at 20 °C and 24h incubation time. (C) Effect of the incubation time on the pET303, pET302, and pET45b expression system, with a fixed temperature at 20 °C and IPTG conc. (1mM). (C) Culture fluorescent in different pET vectors (pET 45b, pET302 and pET303) under fixed conditions (incubation time of 24h at 20 °C with 1mM IPTG).

2.3.2 Purification of Fluorescent Proteins

All fluorescent proteins cell pellets were lysed and purified using four different methods, Ni-affinity chromatography column, three phase partitioning, and our new combined protocol which consist of thermal / and 3PP purification protocol (T/3PP). The clear cell lysate of each culture was divided into three equal fractions, one for each purification protocol. FPs were expressed at relatively high levels, although with some variations depending on the particular protein with a total FP-yield between 88mg (mTagBFP) and 370mg (eCGP123) per L culture (Table 2.1).
Table 2.1. Spectra characteristic and expression yield of FPs.

<table>
<thead>
<tr>
<th>FPs</th>
<th>Structure</th>
<th>EX</th>
<th>Em</th>
<th>ε</th>
<th>φ(F)</th>
<th>Chromophore tripeptide</th>
<th>Yield&lt;sup&gt;a&lt;/sup&gt; (mg/L)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>eCGP123</td>
<td>monomer</td>
<td>493</td>
<td>504</td>
<td>109,333</td>
<td>0.69</td>
<td>Gln62-Tyr63-Gly64</td>
<td>370</td>
<td>28</td>
</tr>
<tr>
<td>Sf_gfp</td>
<td>weak dimer</td>
<td>490</td>
<td>510</td>
<td>83,300</td>
<td>0.65</td>
<td>Thr65-Tyr66-Gly67</td>
<td>90</td>
<td>29</td>
</tr>
<tr>
<td>mVenus</td>
<td>monomer</td>
<td>515</td>
<td>528</td>
<td>92,200</td>
<td>0.57</td>
<td>Gly65-Thr66-Gly67</td>
<td>340</td>
<td>30</td>
</tr>
<tr>
<td>TagRFP</td>
<td>monomer</td>
<td>555</td>
<td>584</td>
<td>100,000</td>
<td>0.25</td>
<td>Met63-Tyr64-Gly65</td>
<td>261</td>
<td>31</td>
</tr>
<tr>
<td>mKate2</td>
<td>monomer</td>
<td>588</td>
<td>633</td>
<td>62,500</td>
<td>0.4</td>
<td>Met63-Tyr64-Gly65</td>
<td>224</td>
<td>32</td>
</tr>
<tr>
<td>mTagBFP</td>
<td>monomer</td>
<td>399</td>
<td>456</td>
<td>52,000</td>
<td>0.63</td>
<td>Leu63-Tyr64-Gly65</td>
<td>88</td>
<td>33</td>
</tr>
<tr>
<td>mTFP1</td>
<td>monomer</td>
<td>462</td>
<td>492</td>
<td>64,000</td>
<td>0.85</td>
<td>Ala66-Tyr65-Gly67</td>
<td>420</td>
<td>34</td>
</tr>
<tr>
<td>mTFP1K183</td>
<td>monomer</td>
<td>462</td>
<td>492</td>
<td>64,000</td>
<td>0.85</td>
<td>Ala66-Tyr65-Gly67</td>
<td>280</td>
<td>34</td>
</tr>
</tbody>
</table>

<sup>a</sup> Calculated by Absorption measurements of FPs and lambert-beer equation

A. Ni-affinity chromatography column

Affinity purification is the most common method for protein purification<sup>35</sup>. This method is easy in principle and proves to provide sufficient pure protein for a quick purification. However, if purity greater than 95% are desired, affinity chromatography requires a follow-up step. Most commonly, preparative gel filtration has been used to remove protein aggregates that may form when a few bound ligand leache from the column<sup>36</sup>. Moreover, this method is a relatively expensive method and the protein, which is received contains polyhistidine tags which can be difficult to remove from correctly folded FPs.

To screen the best purification protocol for the FPs, we first employed affinity chromatography on Ni-affinity chromatography column. For yield and purity, both total and eluted proteins were analyzed by SDS–polyacrylamide gel electrophoresis after dialysis (Figure 2.4). We achieved remarkable purity ranging from 99%, for tagRFP, mVenus and Sf-GFP likewise eCGP123 shows noticeable high purity 97%, to 80% for mTAGBFP. The yield was estimated by comparing the initial fluorescent concentration in total protein content sample to the final fluorescent concentration in the pure sample as determined by the lambert-beer equation using relative extinction coefficients (ε) for each protein (Table 2.2). We observed a moderate yield ranging from 88% for mKate2 to 50% for mTagBFP.
The IMAC method generates a highly pure target protein however, a high percentage of protein was lost and diluted while eluting the protein resulting in decreasing the final yield. This method is fast, useful and high-throughput but only suitable for proteins with His-tag. However, many applications like catalysis and crystallization require metal binding which can be affected by His-tag. However, this is not relevant to non-engineered FPs.

**Figure 2.4.** SDS-PAGE analysis, of A) mKate2, B) TagRFP, C) mVenus, D) eCGP123, E) mTAGBFP, F) Sf-GFP, and G) mTFP1; loaded after three different purification protocols, Lane 1: T/3PP, Lane 2: Ni-affinity chromatography and Lane 3: Thermal purification, to compare the purity.
Table 2.2. Purification comparison of FPs by different purification protocols.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Yield % (purity a%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMAC</td>
</tr>
<tr>
<td>mKate 2</td>
<td>88 (80)</td>
</tr>
<tr>
<td>TagRFP</td>
<td>71 (99)</td>
</tr>
<tr>
<td>mVevus</td>
<td>76 (99)</td>
</tr>
<tr>
<td>eCGP123</td>
<td>67 (97)</td>
</tr>
<tr>
<td>mTagBFP</td>
<td>50 (81)</td>
</tr>
<tr>
<td>St-GFP</td>
<td>71 (99)</td>
</tr>
<tr>
<td>mTFP1</td>
<td>64 (94)</td>
</tr>
</tbody>
</table>

*analyzed by the gel band intensity using software ImageJ

---

**B. Three-Phase Partitioning**

A fraction of soluble protein from each FPs was processed using three phase partitioning purification. We achieved a satisfied level of purity by extracting the FPs by ethanol and saturated ammonium sulfate concentration above 60%, used on all proteins fractions. As this concentration is required to create phases in the water/ethanol system which can remove the impurities by preventing the denaturation of the FPs. After ammonium sulfate and NaCl addition, ethanol was added immediately to avoid co-precipitation of the target FPs with other proteins. A high concentration of ammonium sulfate is needed because of its salting-out effect, (NH₄)₂SO₄ decreases the solubility of most of the proteins in the water phase. Compared to water, ethanol is less polar and precipitate proteins by reducing the dielectric constant of the solution. Although the water phase is enriched by ammonium sulfate and the organic phase is enriched by ethanol, it should be noted that both of these components to some extent partition in each phase. In this case, protein solubility is affected by a complex mechanism. For example, as the ethanol phase is also saturated with (NH₄)₂SO₄, the salt promotes solubility of proteins in the organic solvent by preventing their electrostatic aggregation. FPs were extracted into the upper phase ethanol after ethanol addition to the solution of FPs in saturated ammonium sulfate proteins. To recover the FPs in an aqueous phase. A less polar and more hydrophobic organic solvent n-butanol is added. As mentioned above, the ethanol extracts are
saturated with water, which forms a separate phase after the addition of n-butanol. This step provides additional purification, as most of the proteins soluble in the ethanol extract are precipitated at the interphase. The FPs were completely extracted into the aqueous phase, which results in about tenfold concentrated solution of FPs\textsuperscript{19}. Since the aqueous phase also contains organic solvents and residual salts that may cause denaturation of the target protein, prolonged storage of the sample is not recommended, proteins are immediately dialyzed against an appropriate buffer (HEPES, 50 mM, pH 7.0). We noticed with some FPs, for intense, mKate the color changed after adding the organic solvent from its original, presumably due to the different protonation states which is due to the change of the polarity and the pH after adding the organic solvents. The change of the protonation state of mKate upon pH variation influences the electrostatic field in the β-barrel interior, apparently affecting the hydrogen bond system by gain or loss of H-bonds, which might trigger chromophore isomerization, correspond to the change of the color. This change is rapidly coming back to the original after dialysis\textsuperscript{23}.

We obtained samples of fluorescent proteins with the purity range from 65% to 96% as estimated from SDS–PAGE (Figure 2.4), with the yield up to 70%, compare to the initial yield calculated from absorption and fluorescent measurements with the consideration of the protein loss during the purification (Table 2.2). This loss’s correlate with the stability of each protein in the organic solvents, with mVenus being the most and mTag BFP2 the least stable protein in the organic solvent.

\section*{C. Thermal Purification}

Since FPs are relatively stable protein, precipitation of other proteins at high temperature might represent a viable purification strategy. Herein, we first tested the heat stability to identify the suitable temperature for each protein in this purification method. As illustrated in (Figure 2.5), mVenus, mTFP1 and Sf-GFP are stable up to 90 °C, mKate2 and Tag RFP denature above 70 °C. Hence 70 °C showed not to be exceeded for this purification protocol with this group of FPs.

Thermal purification was applied to precipitated all alternative proteins from the lysate at 70 °C. The samples were centrifuged at 5000rpm for 15 min to separate the
impurities from soluble FPs. The purity of proteins was initially estimated by SDS-PAGE analysis and absorption measurements at specific weave length for each protein. This purification worked well for all FPs we are tested with yields ranging from 63% to 89% and purities of 67% to 94% (Table 2.2). However, this purification method was not sufficient to provide the level of purity, which is required in some biological experiment. Additionally, heat purification is not a universal method for all FPs, as denaturation effect and chromophore conformational change could happen for target proteins during the heating step. In most cases this can be reversible similar to the heat effect on the mKate2 as the changing of the color was diminished and the mKate2 protein color retained after cooling the sample. We observed that applying heat to the fluorescent protein could increase the fluorescence intensity of the FP, which may indicate that at elevated temperature some misfolded fluorescent proteins molecules can finally maturated.

**Figure 2.5.** Thermal stability: indicate stability measures throughout the quantification of specific spectroscopic features for each protein (absorption: nm, fluorescence: nm).
D. Combined Thermal – Three Phase Partitioning Purification Protocol (T/3PP)

Since neither 3PP nor thermal purification produced the sufficiently pure protein samples, we decided to combine the thermal and the 3PP purification in one purification sequence. All proteins samples after the heat purification (at 70 °C for 20min) were collected and further purified by the three phase partitioning. The purity of samples was determined by the comparison of absorption spectra, fluorescence profile and SDS analysis of the pure material and cell lysate (Figure 2.4) while the absorption measurements used for conducting the final yield for each FPs. The purity of mTagBFP reached to 86%, Sf-GFP to 88%, TagRFP to 96%, mTFP1 to 98% and mVenus and eCGP123 to 99%. The protein yield ranged from 50 to 95% (50% for mKate2, 75% for Sf-GFP, 80% for mVenus, eCGP123, and mTFP1 and 95% for TagRFP). In most cases, purity and/or yield of the FP samples received by the combined T/3PP protocol excellent chose of all other protocols.

Most of the unwanted proteins were removed from the protein fraction by the heat purification protocol estimated as 70-90% and then the 3PP purification protocol was used for additional purities and remarkable protein concentrating step. The three phase partitioning (high salt buffer, ethanol, butanol) removes water-soluble impurities efficiently. The resulting pure native protein was in aqueous phase, which needs only a simple desalting step. Combination of heat and 3PP purification provides the rapid method for pure protein under native condition as evidenced by absorption and fluorescent spectra which remain similar to the previous studies (Fig 8).

ESI-TOF was used to identify proteins purified by (T/3PP) protocol (Figure S2.1). The results confirmed the identities of all proteins (mKate2=27139.9 Da, TagRFP=27156.9 Da, mVenus=27957.5 Da, eCGP123=26913.6 Da, mTAGBFP=27871.3 Da, Sf-GFP=27477.1 Da, mTFP1= 27971.5 Da, mTFP1-183 PIF = 28116.3 Da) based on molecular weight (Table 2.3).

2.3.3 Purification of FP with UAA

To broaden the application of T/3PP protocol, unnatural amino acid incorporated FP were also purified. The purity level of mTFP_PIF delivered by the T/3PP purification protocol
was notable alike other FPs (see supplementary), it supports the possibility of large-scale production of biocompatible fluorescent tags (FP + clickable uaa, e.g. pAzF). The MS data prove the incorporation was not affected after the T/3PP purification (Figure S2.9H).

### Table 2.3. FPs detected mass by ESI-TOF

<table>
<thead>
<tr>
<th>Protein with 6His-tag</th>
<th>Calculated</th>
<th>Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>mKate 2</td>
<td>27139.9</td>
<td>27142</td>
</tr>
<tr>
<td>TagRFP</td>
<td>27156.9</td>
<td>27159</td>
</tr>
<tr>
<td>mVevus</td>
<td>27957.5</td>
<td>27960</td>
</tr>
<tr>
<td>eCGP123</td>
<td>26913.6</td>
<td>26915</td>
</tr>
<tr>
<td>mTagBFP</td>
<td>27871.3</td>
<td>27872</td>
</tr>
<tr>
<td>Sf-GFP</td>
<td>27477.1</td>
<td>27479</td>
</tr>
<tr>
<td>mTFP1</td>
<td>27971.5</td>
<td>27974</td>
</tr>
<tr>
<td>mTFP1K183PIF</td>
<td>28116.3</td>
<td>28118</td>
</tr>
</tbody>
</table>
2.3.4 High Throughput (HT)-Purification in a 96-Well Plate Formed “Parallel Assay”

Fast purification methods should be parallelizable, therefore, the T/3PP was tested for reproducibility. The lysate fraction were distributed in in the PCR 96 well plate, heat was applied by the thermocycler up to 70 °C for 20min. All samples were centrifuged at 20 rpm for 30min and the supernatant was carefully transferred into the deep 96 well plate for preforming the organic extraction steps. The protein fluorescence and concentration were measured at each step as well as in each well to detect the fluorescent differences between wells and the reproducibility of the method. HP/3PP parallel purification protocol generates high purity of 95% and yields of 70 % (Table 2.4). Figure 2.8 illustrates the well to well reproducibility of the yields, which proofs that this method is well suitable for parallelized expression and purification screens.

Figure 2.6. SDS-PAGE analysis, of mTFP1 loaded after the parallel purification method. Lane 1: cell lysate, Lane 2: 3PP, and Lane 3: T/3PP.
Table 2.4. Comparison of different purification protocols of mTFP1 by using parallel purification method.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Yield % (purity%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heat</td>
</tr>
<tr>
<td>mTFP1</td>
<td>89 (85)</td>
</tr>
</tbody>
</table>

Figure 2.7. Reproducible of mTFP1 by using parallel purification protocol.

2.3.5 Spectra Prosperities of Purified Proteins

The pure protein fractions were identified by comparison of the fluorescence profile of the pure protein with the initial fluorescent of the cell lysate. Comparison of fluorescence profiles before and after purification also indicates that the protein’s chromophore was unaltered by the purification procedure (Figure 2.9). Fluorescence was measured at specific excitation and emission wavelength for each protein. Absorption spectra of purified proteins are indicative of their native state, purity, and concentration of the protein in the pure sample. Determining the concentration of the protein by Lambert beer equation at each step of the protein purification methods implies that T/3PP purification protocol was the most effective protocol in terms of total protein loss during the purification.
Figure 2.8. Absorption and emission spectra of purified FPs by T/3PP protocol. Spectra of indicated proteins in 20 mM Tris–HCl, 150 mM NaCl, pH 7.8 buffer were obtained at a concentration between 0.1 and 0.5 mg/ml and normalized to chromophore maximum absorption A) eCGP123; 504 nm, B) Sf_GFP; 510 nm, C) mVenus; 528 nm, D) TagRFP; 584 nm, E) Kate2; 633 nm, F) mTagBFP; 456 nm, and G) mTFP1; 492 nm.
2.4 Conclusions

In the present study, the T/3PP purification protocol for purifying FPs was developed to provide an outstanding level of purity couple with rapid and easily performing steps using the simple organic extraction procedure\(^1\). In this method, the purification costs constitute a fairly large percentage of the overall production costs of proteins/enzymes. The T/3PP purification protocol represents a fairly economical process since it does not utilize large amounts of salt (always less than used in a ‘salting-out’ protocol during ammonium sulfate precipitation)\(^2\). The availability of this simple purification strategy should help in the large-scale production of FPs in a more efficient and economical way. It will overcome the stock problems the disadvantages from the majority of FPs purification techniques. The (T/3PP) purification method proves to be suitable for HTS (high throughput screening) protein purification, which was developed utilizing the stability of FPs to heat and organic solvents. This method is efficient, simple, rapid and economical. Heat purification (70 °C) can easily be parallelized, gives good purification results by removing the all non-stable proteins. However, heat purification does not get rid of water soluble natural products (DNA, RNA, ATP, etc.). Additional organic 3-phase partitioning (high salt buffer, ethanol, n-butanol) removes water-soluble impurities efficiently instead of the expensive and time-consuming gel-filtration or adsorption chromatography. The T/3PP method allows purification without any requirement of tags, which may interfere with future usage of the protein (e.g., catalytic reactions, crystallography).

In conclusion, this method is a superior alternative to expensive and laborious techniques developed so far. The T/3PP purification method provides the advantages of conventional salting out, isoionic precipitation and co-solvent precipitation of proteins. It can readily be scaled to purify the small or large amounts of protein. Moreover, the method could also be implicated in discovery and screening of novel GFP-like proteins. Herein, the proposed purification technique has proven to be applicable to all the FPs studied. The yield and purity of the resultant product compare favorably with those obtained otherwise. The method allows purification of FPs in analytical, preparative and large-scale procedures.
2.5 Material and Methods

2.5.1 Cloning of Constructs

The fluorescent proteins genes eCGP123, Sf_GFP, mVenus, Tag RFP, mKate2, mTag BFP2 and mTFP1 were ordered from Geneart\textsuperscript{TM} (Life Technologies GmbH) and amplified using respective primers listed in Table 2.4. The purified PCR fragment was doubly digested with Xbal and Xho1 (New England Biolabs Inc.) and ligated into a similarly digested Champion\textsuperscript{TM} pET303/CT-His vector (Invitrogen Life Technologies GmbH) containing a His6-tag construct (T7 promoter) (Figure 2.2). Other pETs vectors details used in this study for screening the best expression system for FPs are listed in (Table 2.5).

Recombinant plasmids were electroporated into \textit{E. coli} XL-1 strain (Stratagene\textsuperscript{TM}, Agilent Technologies). \textit{E. coli} XL-1 strain was grown at 37 °C in Luria-Bertani (LB) medium supplemented with 180 µg/ml ampicillin. The constructs were validated by DNA sequencing.
Table 2.5. Oligonucleotides for FPs (eCGP123, Sf_gfp, mVenus, TAG RFP, mKate2, mTAG BFP2, mTFP1, and mTFP1K183 PIF)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primers sequence</th>
</tr>
</thead>
</table>
| eCGP123 | eCGP123pET303Xbf  
|        | eCGP123pET303Xhr  |
|        | 5'-ataatc tctaga ATGTCTGTTATCAAACC GGAAATG-3'  
|        | 5'-ataatc tctctg TTAGCCTGAGACGGCAG-3' |
| Sf-GFP | sfgfp_pET303_Xbf  
|        | sfgfp_pET303_Xhr  |
|        | 5'-ataatc tctaga ATGC GTAAGATGGAAGACGTTC-3'  
|        | 5'-ataatc tctctg TTTGTACAGTTCGTCCATACCGT-3' |
| mVenus | mVn_pET303_XbaI  
|        | mVn_pET303_XhoI  |
|        | 5'-ataatc tctaga ATGG TTTCTAAGGGTGAAGACTGTT-3'  
|        | 5'-ataatc tctctg TTTGTACAGTTCGTCCATACCCAG-3' |
| mTAG RFP | TRFP_pET303Xbf  
|        | TRFP_pET303Xhr  |
|        | 5'-ataatc tctaga ATGTCG AACTGATCAAAGAAACATG-3'  
|        | 5'-ataatc tctctg TTTGTGACCCAGTTTAGACGG-3' |
| mKate2 | mKt2_pET303Xbf  
|        | mKt2_pET303Xhr  |
|        | 5'-ataatc tctaga ATGGTTTTCTGATCAAGAAAAAC-3'  
|        | 5'-ataatc tctctg ACGGTGACCCAGTTTAGACG-3' |
| mTAG BFP2 | mTBFP2_pET303Xf  
|        | mTBFP2_pET303Xhr  |
|        | 5'-ataatc tctaga ATGTCG AAGATGCAAGAAAAAC-3'  
|        | 5'-ataatc tctctg GTTCAGTTTGTGACCCAGTTTAGA-3' |
| mTFP1 | mTFP1_pET303_Xbf  
|        | mTFP1_pET303Xhr  |
|        | 5'-ataatc tctaga ATGGTTTTCTAGGGTGAAGAACCAC-3'  
|        | 5'-ataatc tctctg TTTGTACAGTTCGTCCATACCGTC-3' |
| mTFP1 K183PIF | mTFP1-K183_pET303_Xbf  
|        | mTFP1-K183_pET303Xhr  |
|        | 5'-ataatc tctaga CTTCCCCCCCTAGGG CCGCGGTG-3'  
|        | 5'-ataatc tctctg CACGG GCCCCCTAGGGGGGAAG-3' |

Table 2.6. Table of pETs vector used for screen FPs expression

<table>
<thead>
<tr>
<th>#</th>
<th>Vector</th>
<th>tag</th>
<th>Vectors links</th>
<th>Resistance</th>
<th>Induction</th>
<th>pro-motor</th>
<th>ori</th>
<th>Cutting site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PET45b</td>
<td>CT-His</td>
<td>user manual</td>
<td>ampicillin</td>
<td>IPTG</td>
<td>T7</td>
<td>pBR322</td>
<td>BAMHI/PstI</td>
</tr>
<tr>
<td>2</td>
<td>PET302</td>
<td>NT-His</td>
<td>user manual</td>
<td>ampicillin</td>
<td>IPTG</td>
<td>T7</td>
<td>pBR322</td>
<td>XhoI/BAMHI</td>
</tr>
<tr>
<td>3</td>
<td>PET303</td>
<td>CT-His</td>
<td>user manual</td>
<td>ampicillin</td>
<td>IPTG</td>
<td>T7</td>
<td>pBR322</td>
<td>XbaI/XhoI</td>
</tr>
</tbody>
</table>

2.5.2 Protein Expression

All fluorescent proteins genes were expressed in *E. coli* BL21(DE3) strain (Stratagene™, Agilent Technologies). The respective strain was grown to OD$_{600nm}$ of 0.6–0.7 at 37 °C with shaking 160rpm in Luria Broth (LB) medium supplemented with 180 µg/ml ampicillin and further incubated overnight at 20 °C after induction by 1mM IPTG (isopropyl-β-D-thio-galactopyranoside). The cells were harvested by centrifugation at 4 °C for 30min at 5,000 rpm. The cell pellets were washed with saline buffer (0.9% NaCl, pH 5.5) and stored at -20 °C until further use.
2.5.3 Protein Expression with UAA

To express mTFP1-uaa, pET303/mTFP1-Lys183TAG vector (Y183TAG mutation was made for this work) was co-transformed into BL21(DE3) with pEvol-pIPheRS vector (Group Schultz, The Scripps Research Institute, USA). Cells were grown in LB media supplemented with 180 µg/ml ampicillin and 50 µg/mL chloramphenicol at 37 °C. The expression of recombinant protein was induced at an OD$_{600nm}$ of 0.6-0.7 by an addition of 1mM IPTG and 0.02% of L-arabinose followed by 1mM of corresponding unnatural amino acid para-(L)-IodoPhenylalnine($p$IF) (synthesized in our lab) (see supplementary material for synthesis details). After Induction, cells were grown at 20 °C for overnight. The cells were harvested by centrifugation at 4 °C for 30 min at 5,000 rpm. The cell pellets were stored at -20°C until further use. The modified mTFP1K183 PIF, were characterized with ESI-TOF.

2.5.4 Protein Purification

The cell pellets collected (5-6 g/L) were re-suspended in lysis buffer [containing 50 mM Tris/HCl pH 7.4, 150 mM NaCl, with addition of DNAase I (Sigma), and Pefabloc (lysozyme)] and disrupted by passage (twice) through a French press (FIRM) at 1.7kBar. Cell debris was removed immediately by centrifugation (25,000 × g, 4 °C, 30 min) to obtain the crude extract. Supernatant was divided into three fractions.

2.5.4.1 Ni-affinity Chromatography Column

C-terminal (His)$_6$ tagged FPs were purified using a standard IMAC protocol using an Äkta system (GE Healthcare Life Sciences, Piscataway, NJ). The supernatant was loaded at a flow rate of 1.0 ml/min onto a 5-ml HiTrap Ni$^{2+}$ chelating column pre-equilibrated with the binding buffer (100 mM Tris–HCl pH 7.4, 500 mM NaCl, 20 mM imidazole). The elution was carried out with a 0-500 mM imidazole gradient. The eluted fractions were analyzed by SDS–PAGE analysis. The pure fractions were pooled and the imidazole was removed by overnight dialysis at 4 °C against 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.02% sodium azide and 5% glycerol.
2.5.4.2 Three-Phase Partitioning (TPP)

The soluble protein second fraction was purified by Three-phase partitioning (TPP). All procedures for 3PP were performed at RT. For analytical procedure was carried out according to the method of Yakhnin, Alexander V. To soluble protein fraction, 0.3 ml of 5 M NaCl and 2.33 ml of saturated (NH₄)₂SO₄ (pH 7.8) were added to 1 ml of soluble protein portion. The mixture was blended with 1.2 ml of 96% ethanol then shaken for 30s and FP separated into the upper ethanol phase after centrifugation for 7 min at 3000g. N-Butanol (1 volume, 0.4–0.5 ml) was added to the ethanol extract (4 volumes) and separated by centrifugation then the retrieve FP in the lower aqueous phase (0.3–0.4 ml). Finally, pure protein collected and dialyzed against 20 mM Tris–HCl, 150 mM NaCl, pH 7.8 buffer.

2.5.4.3 Thermal Purification

The third fraction of the soluble protein was used for heat purification. Heat stability test was preformed for all proteins to determine the denaturation point for each protein over a temperature range (20 °C–90 °C). In this test, protein sample was diluted to the concentration of 2 mg/ml in 100 mM Tris, 500 mM NaCl and 30 mM imidazole. The protein samples were transferred into a 10 mm quartz cuvette and heated for 2h with a ramp rate of 2 °C/min in a spectrophotometer (JASCO V-670). Absorption and fluorescent measurements at the specific wavelength for each protein were conducted before and after heating.

2.5.5 Parallel HT- (T/3PP) Purification in 96-Well Plate

In order to purify fluorescent protein in a parallel fashion by heat purification followed by organic extraction, mTFP1 crude extract fraction was distributed in the PCR 96 well plate (200µL in each well) and heated by the thermocycler (T100™ Thermal Cycler) to 70 °C for 20min. After centrifugation at 20 rpm for 30min at 4 °C, the supernatant was carefully transferred into the deep 96 well plate (180 µL in each well). Organic extraction (as described above) was applied, and all the solvents were scaled correspondingly to the amount in each well. The protein fluorescence and concentration were measured at each step as well as in each well to detect the fluorescent differences between wells and the producibility of the methods.
2.5.6 UV-vis and Fluorescence Analysis

Absorbance and fluorescence spectra were recorded using a microplate reader (TECAN INFINITE® M1000 PRO) according to standard procedure. Chromophore fluorescence intensity of samples was excited at their corresponding emission maximum: eCGP123; 504 nm, Sf_GFP; 510 nm, mVenus; 528 nm, TagRFP; 584 nm, Kate2; 633 nm, mTagBFP; 456 nm, and 492 nm for mTFP1. Samples were diluted in buffer (20 mM Tris–HCl pH 7.0 containing 150 mM NaCl) to record the fluorescent signal within the calibrated linear range (0.001–0.1g/ml) of FP concentrations (data not shown).

The protein yield and concentration were determined by Lambert beer equation\textsuperscript{38,39} and Bradford method\textsuperscript{40} using bovine serum albumin (BSA) as standard. UV-Visible absorption spectra were recorded with a Jasco V-560 UV/visible spectrophotometer using the molar extinction coefficient with the specific excitation and emission of each FP.

2.5.7 Mass Spectrometry:

Protein mass spectrometry was carried out by BRUKER maXis HD\textsuperscript{TM} ESI-TOF. The pure protein (0.04mM) sample was dialyzed against 5% ACN, 0.1% FA and pure direct injection by (Agilent Technologies), C4 Column (column volume 5ml). Elution occurred with a flow rate of 0.5µl/min and a gradient to 80% ACN, 0.1%FA at 8min. Subsequently, fractions were recorded according to standard procedure.

2.5.8 Polyacrylamide Gel Electrophoresis

To determine the purity of proteins, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the method of Laemmlli\textsuperscript{41,42} using 15% (29:1 acrylamide to bis-acrylamide) cross-linked polyacrylamide gels on a vertical gel electrophoresis unit (Bio-Rad Laboratories). The proteins sample concentrations in all gels were 1mg/ml.
2.6 Reference


2.7 Supplementary Material

A) B)

C) D)

E) F)
Figure 2.9. ESI-TOF: A) mKate2=27139.9 Da. B) TagRFP=27156.9 Da. C) mVenus=27957.5 Da. D) eCGP123=26913.6 Da. E) mTAGBFP=27871.3 Da. F) Sf-GFP=27477.1 Da. G) mTFP1= 27971.5 Da. H) mTFP1-183 PIF, (×) indicates Mass of 28116.3 Da.
Chapter 3 Development of a Robust and Versatile Host Protein for Design and Evaluation of Artificial Metal Centers

3.1 Abstract
The development of catalytically active artificial metalloenzymes (ArMs) requires the combination of tools and concepts from biocatalysis, molecular biology and coordination chemistry. Such systems hold the promise of combining the advantages of homogeneous catalysts and enzymes and have thus artificial metalloenzymes, which comprise synthetic metal catalysts embedded in protein scaffolds, have become a vibrant field of research. Current development strategies aim to complement or even outperform versatilities and specificities of the current catalytic portfolio. Herein we present a novel approach to this field, which is based on the design of a fluorescent protein host (mTFP*), which was engineered to have minimal metal binding capabilities. This very robust host protein remains correctly folded over a broad range of conditions. Its intrinsic fluorescence can be utilized to probe the ArMs stability and metal binding properties. We demonstrate how site-specific mutations within the protein scaffold can be utilized to introduce artificial and biomimetic metal binding sites. Protein X-Ray structures and transition metal FRET studies validate predicted and engineered metal coordinations sites and give insights into metal binding dynamics. The resulting ArMs provide a diverse set of catalytic activities, which otherwise do not exist or are scarce in nature. As an example, we probed the catalytic properties of a mTFP* based Diels-Alderase containing a catalytically active Cu(II) center.

3.2 Introduction
According to a recent estimate more than half of all proteins are metalloproteins. It therefore is of high importance to understand the factors that govern metal binding and reactivity in biological systems. Further, the rich set of functions metalloproteins provide is limited by metabolic needs. Hence, expanding the scope of nature's chemistry toolset through introduction of artificial metal coordination geometries and centers holds the promise to provide novel tool for chemical and biological research. Described
applications include pharmaceuticals\textsuperscript{1}, probes for molecular imaging\textsuperscript{2,3} and contrast agents\textsuperscript{4,5}, tools for biophysical studies targeting metallo-protein functions\textsuperscript{6}, metal-directed protein assembly\textsuperscript{7} electrochemical biosensors\textsuperscript{8} and altered electrochemical potential of electron transporting proteins\textsuperscript{9}. Functional artificial metalloenzymes with artificial catalytic activity hold the promise to greatly expand the range of reactions accessible to biocatalysis and therefore this field currently receives great interest\textsuperscript{10-16}. A variety of methods to generate artificial metal sites were developed including domain-based directed evolution strategies\textsuperscript{17,18}, engineering of transition-metal binding sites through introduction site-directed in vivo incorporation of artificial metal-chelating amino acids\textsuperscript{19,20}, or the covalent\textsuperscript{21-24} or supramolecular\textsuperscript{25-27} introduction of metal-containing artificial cofactors. Some systems achieve impressive enantioselectivities\textsuperscript{28,29} or demonstrate the power of co-activation through the second coordination sphere within a protein pocket\textsuperscript{30}. Despite of the impressive progress in recent years, transfer of the artificial metalloprotein technology to real-world applications in organic synthesis or material science so far suffers from several practical drawbacks. Generation of systems, which are based on artificial cofactors or natural amino acid incorporation, is complex and expensive, while important parameters like solubility and/or stability towards temperature, pH and organic solvents are often insufficient. An alternative approach is the design of artificial metal binding sites through well-designed mutations using only canonical amino acids. Reetz and co-workers elegantly engineered an artificial Copper binding site inside the central cavity of the thermophilic TIM-barrel glycerol phosphate synthase from \textit{Thermotoga maritima}, tHisF\textsuperscript{31}. The thermostable tHisF host protein was later used by Lewis to anchor an artificial co-factor\textsuperscript{32}. Engineering of metal binding sites into host proteins through mutations using only canonical amino acids is highly appealing since generation of a suitable metal coordinating geometry only requires conventional site-specific mutagenesis, eliminating complicated synthesis, bioconjugation and purification steps. The simplicity of this approach facilitates expression and screening protocols, should enable in vivo applications, e.g. for whole cell biocatalysis to overcome the stability and TON problems commonly associated with artificial metalloenzymes, and may even provide access to novel biochemical pathways. However, despite all these advantages and while a number of studies investigate the interaction of metals and short
synthetic peptides\textsuperscript{33}, reports on the strictly canonical amino acids based \textit{de novo} design of artificial metal binding sites in proteins remain scarce\textsuperscript{31,34}. The difficulties of predicting metal binding affinities of protein sites represent a major hurdle for the implementation of for approach. After engineering a metal binding site with canonical amino acids, it is highly uncertain under which conditions the metal ion is bound and to what extent binding at the engineered binding site is selective over the numerous other amino acid side chains on the surface of a protein. An ideal host protein should therefore not only express well, be purified easily, show an intrinsic temperature and pH stability and posses solubility in organic solvents, provide a cavity to harbor the artificial metal site, it should also possess no other metal binding residues on its surface and provide a reporter, which directly indicates both, metal binding as well as changes in protein structure.

In the present study, we selected a fluorescent protein (mTFP1)\textsuperscript{35} and utilized site directed mutagenesis to engineer the host protein mTFP\textsuperscript{*}, which combines the above mentioned list of desired characteristics. The intrinsic fluorescence of the exceptionally stable host facilitates fast quantification of metal binding thermodynamics, reveals subtle changes in coordination geometries for different mutants and even provides insights into dynamic processes of the binding event. X-ray structures reveal highly selective binding of transition metal ions at the engineered binding site. Our concluding study on the Diels-Alderase activity of mTFP\textsuperscript{*} based artificial copper enzymes demonstrates, that potential applications of this host protein extend well beyond structural studies to functional design.
3.3 Results

3.3.1 Design and Engineering of the Host Protein mTFP*

The availability of an appropriate host protein scaffold is fundamental to the design of artificial metalloproteins. This is even more important, if the only canonical amino acids are used to introduce a metal binding sites, since no additional parameters, e.g. a functionalized cofactor can be tuned to create the desired properties. While extremophiles are convenient sources for heat and salt stable proteins, an intrinsic reporter for folding and metal binding is difficult to find. Recently, Zagotta and Taraska introduced transition metal ion FRET to probe structural and conformational movements of proteins. Given a sufficient spectral overlap, the light emission of the chromophore of fluorescent proteins may be quenched when a transition metal is coordinated at a distance smaller than the Förster radius (1.0 – 1.5 nm, depending on the metal ion). The β-barrel fold provides an exceptional stability to many fluorescent proteins, and hence we screened known fluorescent proteins for their suitability as scaffold for the design of artificial metalloproteins.

Using available literature data and structural information or homology modeling, we identified mTFP1 as a suitable parent scaffold for our protein-engineering endeavor (see supporting information for details). mTFP1 combines temperature and pH-stability with a bright fluorescence (ε = 64'000 M⁻¹ cm⁻¹, Φ = 64) and spectral characteristics suitable for transition metal FRET (absorbance: 462 nm, emission: 492 nm). It is monomeric, expresses well using standard E. coli protocols and offers a cleft of about 1.3 nm depth and 1.1 nm width at its upper rim between strands 7, 8 and 10 of the β-barrel (figure 3.1A). The dimensions of this pocket fall into the typical range of metalloenzymes. Hence, the pocket’s second coordination sphere can interact with substrates that bind to the metal center and thus govern substrate orientation or provide co-activation.

Overcoming unspecific metal binding is a major challenge for the design of canonical amino acids based metal-binding sites in proteins. The potential presence of several, structurally different metal sites complicates characterization of the metal environments, and reduces the product selectivity of a resulting artificial metalloenzymes. Introducing
late transition metal ions in low oxidation states, which are not available to nature, is of particular interest. Such metal centers preferentially bind to ligands, which can be classified as intermediate or soft Lewis bases. Therefore, surface-exposed His, Met and Cys side chains may compete with a site engineered for metal binding and should be replaced with amino acids that have less affinity to transition metal ions. Examination of the mTFP1 WT structure (PDB: 2HQK) revealed six surface-exposed amino acid side chains of this category (GenBank: DQ676819; positions: H30, M118, H128, H177, H178 and H209). Analysis of the hydrogen-bonding networks and evolutionary conservation indicated a high structural tolerance of these positions towards mutations. Only H30 and H28 show fairly strong H-bonding interaction. An estimate of the ConSurf server\(^4\) indicated all positions to be fairly variable (amino acid conservation score of 4 and below).

Replacement of six surface exposed residues on a 25 kD polypeptide may affect the protein’s properties and a simple alanine replacement would probably impair the scaffold’s solubility and structural stability. Hence, we implemented a three step in silico design strategy (see supporting information for details). First, we used FoldX\(^4^4\),\(^4^5\) to determine the effect of point mutations on the mTFP1 folding energy at each of the identified positions. To maintain a good solubility, we only considered the seven amino acids (R, K, E, D, Q, N, Y) with a hydrophobicity index lower than that of histidine on the empirical hydrophobicity scale of Janin\(^4^6\), as well as the commonly unproblematic alanine mutation. For position M118 additionally the two similarly sized amino acids Leu and Ile were included. According to the in silico prediction, amino acids with a very similar hydrophobicity index (H: -0.2, Y: -0.4; M: 0.4, L: 0.5) give the best folding energy. The FoldX algorithm gave the highest stabilization energies for Tyr introduction at the histidine positions 30, 128, 117, 178 and 209 as well as for the mutation M118L. Additional stabilizing mutations were predicted for positions H128 (E, Q, K, R), H177 (K, R) and H178 (K). Subsequently, we generated an in silico library of 30 mutants, which include all possible combinations of the predicted stabilizing point mutations and used again FoldX to rank the library members according to the predicted free folding energy changes. Finally, we submitted homology models of the four top scoring mutants (predicted stabilization energies of 4.0 - 4.9 kcal/mol compared to WT mTFP1) to
molecular dynamics simulation at 298 K and 343 K. All mutant model structures remained stable during the simulation time (50 ns) and comparison of the root mean square deviations (RMSD) of atomic positions and B-factors with the calculated trajectory of the WT mTFP1 protein did not reveal any significantly increased thermal movement (compare Figure 3.1B for mutant 1 data). Hence, we proceeded with these four mutants to expression studies (sequence of point mutations in order of residue position: mutant 1: YLYYYY, mutant 2: YLYKYY, mutant 3: YLKYYY, mutant 4: YLRYYY).

Molecular biology work started from the WT mTFP1 gene\textsuperscript{42}. We removed short N- and C-terminal sequences (N-terminus: MVSKEETTM, C-terminus: GMELYK), which showed high flexibility in the MD simulations and correspondingly are not resolved in the published crystal structure. Cloning of the truncated gene into a pET303/CT-His-SUMO vector introduced an N-terminal H\textsubscript{6}-SUMO tag, which facilitates purification and allows post-translational elimination of the obligatory N-terminal methionine. Expression of this pET303/SUMO_mTFP1.1 vector and subsequent digestion with the SUMO-Protease, Ulp1 from \textit{Saccharomyces cerevisiae} yields mTFP1 variants with a nine amino acids shorter N-terminus starting with the serine required by the SUMO-Protease. Subsequent introduction of point mutations lead to vectors containing the genes coding for mutants 1 – 4, of which mutant 1 showed the highest expression yields as indicated by the culture fluorescence measured at 492 nm. Therefore, we selected mutant 1 as our host protein scaffold for further studies. This low metal affinity variant of mTFP1 is called mTFP* in this article.

After optimization of the expression conditions for the pET303/SUMO_mTFP* vector, we achieved yields of up to 500 mg protein per L culture of BL21 (DE3) Gold cells. A Purification sequence of nickel affinity chromatography, digestion with SUMO protease and subsequent heating to 70 °C for 15 min, which precipitates remaining impurities yields highly pure samples of mTFP*.
3.3.2 mTFP* a Highly Stable Host Scaffold with Integrated Fluorescent Reporter

LC/MS (ESI-TOF) spectra of purified mTFP* samples show the expected mass of 25154 Da with an abundance of over 95%. Samples, which were not heated to 70 °C show up to 30% of a protein with a 20 Da higher mass. The mass signal of 25174 Da vanishes completely after the heat purification step. We assume that this mass signal results from the mTFP* protein with an un-maturated A62-Y63-G64 sequence. Chromophore maturation involves oxidation and condensation of these three residues and explains the observed difference of 20 Da. From our observation we estimate a chromophore maturation time of 5 – 10 hrs.

The maturated protein shows an absorbance maximum at 468 nm as well as a distinct fluorescent emission peaking at 495 nm (Figure 3.2A). In comparison to mTFP1, these values are red-shifted by about 3 nm. Since only correctly folded fluorescent protein show fluorescence emission, the matured chromophore serves as an intrinsic probe to quickly estimate the amount of correctly folded mTFP* protein in solution. Hence, the absorption and emission data, the stability of an mTFP* derived artificial (metallo)protein can be determined using standardized fluorescence and absorbance measurements.

mTFP* stability was assessed using fluorescence measurements and confirmed with circular dichroism data (Figure 3.2B). The fluorescence signal at 495 nm drops sharply above 90 °C, indicating protein denaturation. This is in agreement with a marked
decrease in ellipticity at 205 nm between 87 and 90 degrees. The estimated mTFP* melting temperature of 89 °C (in 50 mM Tris/HCl at pH 7.4) (see Figure S3.20) is 3 °C below the value determined for WT mTFP1. At 80 °C, the addition of chaotropic reagents like urea (10 M) did not lead to a decrease in fluorescence signal over a 30 min time period (see Figure S3.21). At 25 °C, CD and fluorescence data indicate that mTFP* remains correctly folded within a pH range of 2.3 to 12.6 (Figures 3.2C and 3.2D) as well as in 50 mM Tris buffer containing up to 60 % of organic solvents such as acetone, 2-butanol, t-butanol, methanol, diglyme, dioxan, dimethylformamide, dimethylsulfoxide, ethanol, hexanol, 2-propanol, tetrafluoroethylene and tetrahydrofuran (see Figure S3.22) with a protein solubility of at least 2 mM/L. An organic solvent content of 80% and above typically leads to precipitation of 50% of the protein within 24 h.

**Figure 3.2.** Characterization of mTFP*: A) Absorbance and Fluorescence scan (ex: 468nm, em: 495nm); B) Temperature Stability: mTFP1, mTFP* (circular dichroism at 205nm indicates unfolded fraction of protein); C) pH Stability: mTFP*, pH < 3 (circular dichroism at 198nm indicates unfolded fraction of protein); D) pH Stability: mTFP*, pH > 12, fluorescence measurement (ex: 468nm em: 495nm) indicates stability of protein and respective unfolding.
The purified mTFP* protein forms bright green crystals over night from HEPES buffer at concentrations around 50 mg/mL. An X-ray structure was determined to a resolution of 1.00 Å (PDB: 4Q9W) by single-wavelength anomalous diffraction methods. Successive rounds of model building and refinement afforded a well-defined electron density for the entire mTFP* protein (Rfree = 15.0 %, Supporting Information, Table S3.13). The monomeric protein shows the expected β-barrel fold comprised of 11 antiparallel β-sheets around the central, maturated A62-Y63-G64 chromophore (Figure 3.3A), which is stabilized by well-resolved hydrogen-bonds to residues T58, W89, R91, S142 and H159 (Figure 3.3B). None of the mutations (mTFP* residues Y21, L109, Y119, Y168, Y169 and Y200) introduces a significant perturbation of the backbone structure relative to mTFP1 (Figure 3.3C). However, B-factor comparison reveals that some mutations influence the thermal movement of the flexible loops, as predicted by the MD simulations (Figure 3.3D and Figure S14). For mTFP*, residues of the neighboring loops D150-G151 and K180-Y188 are nearly stationary at their position, while the mobility of the two loops surrounding the cleft between sheets 7, 8 and 10 (L163-G167 and Y200-Y204) is markedly increased. The higher flexibility should allow mutated residues within the cleft to rearrange and adopt a position favorable for metal binding.
Figure 3.3. X-Ray mTFP*: A) indicates crystal structure (PDB: 4Q9W) at 1.00Å, antiparallel sheet (green), connecting loops (blue), alpha-helix (orange), red residues indicate sites of mutation; B) Chromophore of mTFP* with coordinating residues T58, W89, R91, S142, H159 (mesh indicates electron density); C) Superposition of mTFP1 (orange) and mTFP* (green) visualizing low structural deviation; D) B-Factor Comparison of mTFP1, mTFP* and mutant1, dashed lines indicate differences between mTFP1 and mTFP* and confirm overall low thermal movement and structural integrity.

3.3.3 mTFP* Shows a Very Low Background Affinity to Transition Metal Ions

Since mTFP* was engineered to minimize unspecific metal affinity we undertook several experiments to quantify the background metal binding properties of this scaffold. Initially, we incubated mTFP* protein at a concentration of 0.2 mM with 10 equivalents of sulfates or nitrates of 31 transition metals for up to 24 h. In most cases, neither ESI and MALDI MS data nor fluorescence measurements gave any indication of metal binding (Figures S3.23 and table S3.6). Only addition of at least 2 mM equivalents of Cu^{2+}, Rh^{3+} or Pd^{2+} ions resulted in disappearance of the ESI-MS signal after 1 h. A concomitant decrease of mTFP* fluorescence indicates emergence of a transition metal FRET.
(tmFRET, Figures 3.4A and 3.4B), which only increases for copper after 24 h. In all cases, the fluorescence of mTFP* was fully restored after dialysis, demonstrating that metal coordination of mTFP* is reversible. In contrast, mTFP1 showed a much higher reduction in fluorescence signal when incubated with Cu²⁺ ions and neither for Cu²⁺ nor for Pd²⁺ the tmFRET quenching was found to be fully reversible after dialysis.

Metal-induced changes in protein fluorescence can be used to probe the metal binding properties of natural and engineered protein sites. According to Taraska et al.\textsuperscript{37-39,47}, the change in fluorescence induced by metal binding results either from resonant energy transfer between a colored metal ion and a chromophore (tmFRET), static quenching, electron transfer or by perturbation of the protein’s structure, with a tmFRET being the dominant quenching mechanism. Additionally, collision quenching becomes important at high metal concentrations, which can be described by a Stern-Volmer constant $K_{SV}$. This dynamic quenching process results in a reduction of fluorescence lifetime $\tau$ (see Figure S3.30), which is proportional to the metal concentration. Since the tmFRET intensity depends on the occupation of the binding site, the spectral overlap of fluorophore emission and metal absorption bands and the metal-chromophore distance, tmFRET conveniently allow the quantification of metal binding-site affinities and provide insights into the nature of the binding event.

Cu²⁺ titration experiments are in agreement with a strongly reduced metal affinity for mTFP* compared to mTFP1. As, illustrated by Figure 3.5A and 3.5B mTFP* only shows dynamic quenching at high Cu²⁺ concentrations with a $K_{SV}$ of 0.1 mM\(^{-1}\) at pH = 6 and 1.1 mM\(^{-1}\) at pH = 7.5. This is in stark contrast to the tmFRET data of mTFP1, which require at least two additional specific binding constants $K_{d}^{site}$ 320 µM and 1.2 µM of (pH = 6) or 260 µM and 2.1µM (pH = 7.5), to achieve a reasonable curve fit (R > 0.95) for the underlying TM-FRET model (see supporting information for details on the mathematical analysis of the tmFRET data).

Crystal-soaking experiments and co-crystallizations with high metal salt concentrations > 1 mM resulted in structures, which provide further insights into the nature of the
interaction of transition metal ions with mTFP*. Only for Cu$^{2+}$ or Pd$^{2+}$ ions a coordination to mTFP* side chains is evident in the crystal structures. In a few cases, mTFP* crystals that were soaked with 50 mM PdCl$_2$ solution showed high electron density between surface-exposed Asp or Glu residues and the ε-aminofunctions of lysines, which could be refined as Pd$^{2+}$ centers. The transition metal replaces a proton in inter- or intra-molecular salt-bridges, which indicates that electrostatic interactions play a major role in Pd$^{2+}$ coordination. In most cases, additional acidic residues, lysines and/or chloride anions complete the square planar coordination environment and provide additional stabilization (see Figures 3.4C and S25). Electron density maps recorded from crystals, which were crystallized from solutions containing 35 mM CuSO$_4$ revealed an additional area of high electron density next to Asp140. Coordination of this Cu$^{2+}$ center is further stabilized by interaction with the carbonyl oxygen of Lys160 and the backbone nitrogen. The trigonal planar environment of the latter (sum of angles: 359°) argues for a deprotonation upon copper binding, which is commonly observed in natural copper binding motifs and may also be responsible for copper(II)-induced misfolding of Parkinson’s disease protein$^{47}$. Metal binding opens up H-bonding connection between β-sheets 7 and 8, which allows additional coordination of the otherwise buried His159 side chain. Three effects seem to contribute to this binding mode on mTFP*: (i) increased flexibility of β-sheet 8, and ii) the small radius of Cu$^{2+}$ permitting the cation to slip in-between strands 7 and 8, while iii) the long incubation time of several days provides ample time for the required breaking of several hydrogen-bonds between these β-sheets.
Specific Transition Metal Binding Sites can be Incorporated into mTFP*

The cleft at the upper rim between strands 7, 8 and 10 of the mTFP* β-barrel represents the best location to introduce artificial an metal binding motif. It is about 1.1 nm wide,
1.3 nm deep and 1.7 nm long, which resembles the typical dimensions of pockets found in metal enzymes, and residues from beta-sheets 7, 8, 10, 11 and from the central helix point towards the center of this cavity, providing many possibilities to introduce mutations. The distance between the central chromophore and the cleft (1.05 nm - 2.30 nm) lies within the range which is suitable for *tmFRET* studies\(^{48}\). Residue Y200 sits in the center of this cleft. Since this position corresponds to H209 in *mTFP1*, we chose mutation Y200H as our starting point to engineer an artificial metal binding site. Using a homology model of this mutant, we identified positions D54, Y55, I197 and Y204 as sites for the introduction of additional metal binding His, Met or Cys residues. FoldX predictions rendered mutants I197C-Y200H-Y204H (mTFP\(^{CHH}\)) and I197E-Y200H-Y204H (mTFP\(^{EHH}\)) as the most stable ones. Mutations were introduced using site directed mutagenesis. Following expression, the combined H6-SUMO-tag and heat purification protocol mentioned above yielded > 200 mg/L of highly pure protein which shows the expected molecular masses of the maturated proteins (mTFP\(^{CHH}\): 25’092) and (mTFP\(^{CHH}\): 25’118). Spectroscopic characteristics as well as thermal and \(pH\) stabilities are similar to those of *mTFP*\(^*\). The mTFP\(^{XHH}\) mutants crystalize under conditions similar to those established for *mTFP*\(^*\). X-ray structures reveal that the side chain of the introduced metal chelating amino acid point towards the inside of the cavity resembling a pre-formed metal binding site (Figures 3.6A and 3.6B, “closed” conformation). However, crystals of mTFP\(^{CHH}\), which were grown form slightly acidic conditions (\(pH = 6.0\)) revel a second “open” conformation, in which residue H200 is turned outwards (Figure 3.6C). This conformation, underlines the increased flexibility of residues 197 – 206 of \(\beta\) -strand 10 (Figure 3.6E and 3.6F). It is stabilize by a hydrogen-bond between the protonated H200 and T208 and reduces the repulsive coulombic interaction of two positively charged side chains (H200 and H204). This conformational change reduces the distance between C197 and H204 from 5.7 Å to 5.2 Å.

Co-crystallization and soaking experiments with Pd\(^{2+}\) and the mTFP\(^{XHH}\) variants were unsuccessful. However, Co-crystallization of mTFP\(^{CHH}\) with a slight excess of CuSO\(_4\) yielded crystals of the protein’s closed conformation with one Cu\(^{2+}\)ion in the center of the protein cleft and 1.48 nm from the closest heavy atom of the AYG\(_{62-64}\) chromophore. N-Cu distances of 2.1 Å (H200) and 2.1 Å (H204) indicate a dative bonding interaction
between the metal center and the two histidine side-chains. Cu coordination strongly reduces the thermal flexibility of the residues 195 - 210 (Figure 3.6E). It also induces a slight rearrangement of the flexible part of strand 10, which enhances the Cu-S distance to 5.7 Å and thus prevents a concomitant Cu-coordination of both histidines and C197. This experimental result deviates from force field based predicted metal coordination to all three residues, which underlines the importance of experimental verification, particularly if metal coordination is involved. To gain a more detailed understanding of Cu²⁺ binding, we derived models of the copper complexes of mTFP^{CHH} in the open configuration and mTFP^{EHH} from the respective crystal structures. The geometry of residues C197 and H204 (mTFP^{CHH}, open configuration) as well as E197, H200 and H204 (mTFP^{EHH}), which show a good pre-orientation for metal binding (d(S_{C197-N_{H204}}) = 5.2 Å; d(O_{E197-N_{H200}}) = 4.6 Å, d(O_{C197-N_{H204}}) = 3.0 Å, d(N_{H200-N_{H204}}) = 3.7 Å) makes model building straightforward. Introduction of a copper center induces only a minor re-orientation of these residues. The resulting coordination geometry is stable in MD simulations. While distances and angles of these models should be interpreted with care, comparison of the overall coordination geometry with X-ray structures indicate that coordination of mTFP^{CHH} involves either residues H200 and H204 (closed form) C197 and H204 (open form), while the copper binding motif of mutant mTFP^{EHH} consists of the three residues E197 (η2-coordination in model), H200 and H204.
Figure 3.6. mTFPCHH: E) B-Factor Comparison of mTFP*, mTFPCHH, mTFPCHH_Cu, showing stabilized coordination throughout Copper coordination by 2xHis-Motive; F) Superposition mTFP*(green), mTFPCHH (orange), mTFPCHH_Cu (blue), visualizing structural rearrangement of cleft flanking loop, especially upon Cu coordination (* indicates shifted residue Y/H200);
3.3.5 *tm*FRET Assays Reveal the Dynamic Nature of the Metal Coordination in Solution

ESI/MS data and fluorescence measurements provide insights into the nature of metal coordination in solution. The data reveal a surprisingly stable metal coordination of mTFP<sup>CHH</sup> and mTFP<sup>EHH</sup>. In contrast to mTFP*, the signals indicating 1:1 complexes of Cu<sup>2+</sup> and Pd<sup>2+</sup> are found by LC/MS (Table 3.1). Even after dialysis over night, signals did not decrease below a 50% intensity of the sum of all mass peaks. These findings agree with the observed fluorescence quenching behavior. Quenching is observed for various transition metal ions including Ni<sup>2+</sup>, Cu<sup>2+</sup>, Rh<sup>3+</sup> and Pd<sup>2+</sup>. This effect is more pronounced than for mTFP* and increases in the order Ni < Rh < Pd < Cu. After dialysis for 24 h the fluorescence signals remain at 99, 81, 85 and 67% of the metal free mTFP<sup>CHH</sup> protein (Figure S3.19).

Table 3.1. Summary of metal binding characterization for mTFP1, mTFP*, mTFPCHH and mTFPEHH.

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Metal salt&lt;sup&gt;a)&lt;/sup&gt;</th>
<th>Expected&lt;sup&gt;b)&lt;/sup&gt;</th>
<th>Observed</th>
<th>PDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>Cu(NO&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>27971</td>
<td>27971</td>
<td>2HQK</td>
</tr>
<tr>
<td>mTFP* + I197C, Y200H, Y204H</td>
<td>Cu(NO&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>25156</td>
<td>25155</td>
<td>4R6D</td>
</tr>
<tr>
<td>mTFP* + I197E, Y200H, Y204H</td>
<td>Cu(NO&lt;sub&gt;3&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;</td>
<td>25182</td>
<td>25181</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>PdCl&lt;sub&gt;2&lt;/sub&gt;</td>
<td>25255</td>
<td>25255</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a)</sup> Addition of aqueous solution, 1-10 equivalents. <sup>b)</sup> Mass expected for addition of one M<sup>2+</sup> cation.

As outlined above, *tm*FRET is a powerful tool to study the interaction of metal ions and proteins in solution. The observed quenching of the intrinsic mTFP<sup>XHH</sup> fluorescence depends on metal concentration, pH and temperature. In the presence of 50 mM NaCl, which suppresses weak interactions of copper with negatively charge residues on the
protein surface, copper titration experiments reveal for both mutants a well-defined step in the fluorescence decrease (Figure 3.7).

Curve fitting indicates that these steps arise from a tmFRET process, with a specific $K_d^{\text{site}}$ in the $\mu$M range and a FRET efficiency factor $E$ of 0.19 – 0.51. As expected, tmFRET efficiency and background and site-specific metal affinity drop when the temperature is raised from 25 °C to 40 °C (Table 3.2), while an increase in pH from 6.0 to 7.5 enhances the metal affinity of the engineered binding sites by the factor 10 (mTFP$^{\text{CHH}}$, $K_d^{\text{site}}$ decreases from 0.56 to 0.053 $\mu$M) and 1.8 (mTFP$^{\text{EHH}}$, $K_d^{\text{site}}$ decreases from 8.1 to 4.5 $\mu$M). Interestingly, the FRET efficiency factors $E^{\text{site}}$ of the two mutants show inverted pH profiles. When the pH is raised from 6.0 to 7.5, the FRET efficiency of mTFP$^{\text{CHH}}$ drops by $2/3^{rd}$ from 0.51 to 0.19, while this value increases by $1/3^{rd}$ for mTFP$^{\text{EHH}}$ from 0.22 to 0.33. The interpretation of this behavior is detailed in discussion section. If the copper-loaded mTFP variants are titrated against another metal, the tmFRET can also be used to study the binding process of transition metals, for which the spectral overlap integral is negligible. Titrations of Cu@mTFP$^{\text{CHH}}$ with Zn$^{2+}$ fully restore the fluorescence of the metal-free protein. The experimental data of these metal competition experiments show a characteristic step, which gives access to the dissociation constant of the Zn@mTFP$^{\text{CHH}}$ complex (supporting information, equation 12). Interestingly, the Zn affinity is strongly influenced by pH and drops from 4.0 $\mu$M at pH = 7.5 to 128 $\mu$M at pH = 6.0. The factor of 32 equals the increase of proton concentration ($\Delta$[H$^+$] = 31.6).

Figure 3.7. tmFRET: mTFP1 and mTFP$^*$ titration against various amounts of Copper(II)Nitrate at pH 6 (A) and pH 7.5 (B).
Table 3.2. Spectroscopic characterization of the interaction of Copper or Zink with mTFP1, mTFP*, mTFPCHH and mTFPEHH.

<table>
<thead>
<tr>
<th>Meta</th>
<th>pH</th>
<th>$K_{av}$ b) [mM$^{-1}$]</th>
<th>$K_{d}$ site b) [µM]</th>
<th>$E$ b)</th>
<th>$R$ c) [nm]</th>
<th>$R^2$ d) [nm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>7.5</td>
<td>0.32</td>
<td>2.1; 260</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.26</td>
<td>1.2; 320</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cu</td>
<td>7.5</td>
<td>1.1$^e$ (0.6)$^f$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.01$^e$ (0.05)$^f$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cu</td>
<td>7.5</td>
<td>0.053 (0.1)</td>
<td>0.19 (0.10)</td>
<td>1.48</td>
<td>1.09 (1.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.56 (1.4)</td>
<td>0.51 (0.29)</td>
<td>1.34</td>
<td>1.35 (1.15)</td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>7.5</td>
<td>-</td>
<td>4.0$^g$</td>
<td>0$^i$</td>
<td>1.48</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>-</td>
<td>128$^h$</td>
<td>0$^i$</td>
<td>1.34</td>
<td>0</td>
</tr>
<tr>
<td>Cu</td>
<td>7.5</td>
<td>1.1$^e$ (0.6)$^f$</td>
<td>4.5</td>
<td>0.33</td>
<td>1.33</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.01$^e$ (0.05)$^f$</td>
<td>8.1</td>
<td>0.22</td>
<td>1.33</td>
<td>1.16</td>
</tr>
<tr>
<td>Zn</td>
<td>7.5</td>
<td>-</td>
<td>288$^i$</td>
<td>0$^i$</td>
<td>1.48</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>-</td>
<td>&gt; 10000$^h$</td>
<td>0$^i$</td>
<td>1.34</td>
<td>0</td>
</tr>
</tbody>
</table>

a) Conditions: 1µM Protein, 50mM MES, pH 6.0 or 7.5, containing 50 mM NaCl; values in brackets were determined at 40 °C. b) Determined from tmFRET titration curves in Figure 5 using equations S1-S13. c) Determined from crystal structure (PDB: 4R6D) or models. d) Estimated from equation S14. e) Derived from mTFP*; fitting as free parameter does not lead to a significant change of value. f) T = 40 °C. g) Determined from tmFRET titration curves at [Cu] = 3.6 µM. h) Determined from tmFRET titration curves at [Cu] = 1.5 µM.

i) No energy transfer assumed due to insignificant spectral overlap integral.
3.3.6 The Artificial Cu-Proteins Show Catalytic Activity and Enantioselectivity

Understanding the structural details of metal binding of the mTFPXHH mutants, we tested the catalytic activity and selectivity of these artificial metalloenzymes using a [2+4] Diels-Alder cyclization as well as a Friedel-Crafts Acylation as test reaction. Both mutants show moderate rate acceleration over the background reaction and achieve enantioselectivities of up to 15 % for the Diels-Alder reaction and 92 % in the Friedel-Crafts Acylation (Figure 3.8). Generally, the cysteine containing mutant shows a much lower activity in these Lewis-acid catalyzed reaction than the glutamate containing version of mTFP*. A more detailed study outlining enhanced reactivities and further mutants will be published separately (See Table 3.3), (Table 3.4).

<table>
<thead>
<tr>
<th>Catalyst</th>
<th>pH</th>
<th>Yield [%]</th>
<th>ee [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.0</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>7.5</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Cu(OTf)₂</td>
<td>6.0</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>Cu(OTf)₂</td>
<td>7.5</td>
<td>99</td>
</tr>
<tr>
<td>5</td>
<td>Cu@mTFPCHH</td>
<td>6.0</td>
<td>89</td>
</tr>
<tr>
<td>6</td>
<td>Cu@mTFPCHH</td>
<td>7.5</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Cu@mTFPCHH</td>
<td>6.0</td>
<td>98</td>
</tr>
<tr>
<td>8</td>
<td>Cu@mTFPCHH</td>
<td>7.5</td>
<td>81</td>
</tr>
</tbody>
</table>

Table 3.3: Diels-Alderase activity tests. [a]

[a] Reaction conditions: [Azacalcone] = 1.0 mM, [cyclopentadiene] = 90 mM \& [Cu(NO₃)₂] (0.1 mM), [Protein] (0.12 mM), T = 4 °C, t = 20 h reaction volume = 4 mL. Reactions were run at the give pH in 50 mM HEPES buffer. Yields and ee values were determined by HPLC.

<table>
<thead>
<tr>
<th>Catalyst</th>
<th>pH</th>
<th>Yield [%]</th>
<th>ee [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>7.5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Cu(NO₃)₂</td>
<td>6.0</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>Cu(NO₃)₂</td>
<td>7.5</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>Cu@mTFPCHH</td>
<td>5.0</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>Cu@mTFPCHH</td>
<td>6.5</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>Cu@mTFPCHH</td>
<td>7.5</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Cu@mTFPCHH</td>
<td>5.0</td>
<td>93</td>
</tr>
<tr>
<td>9</td>
<td>Cu@mTFPCHH</td>
<td>6.5</td>
<td>86</td>
</tr>
<tr>
<td>10</td>
<td>Cu@mTFPCHH</td>
<td>7.5</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 3.4: Friedel-Crafts Alkylase activity tests. [a]

[a] Reaction conditions: [imidazole] = 1.0 mM, [Indol] = 5.0 mM [Cu(NO₃)₂] (0.1 mM), [Protein] (0.12 mM), T = 4 °C, t = 480 h reaction volume = 4 mL. Reactions were run in 50 mM MES (pH = 5.0 and 6.5) or HEPES (pH = 6.5) buffer containing 150 mM of NaCl. Yields and ee values were determined by HPLC.
3.4 Discussion
The design of a stable and functional metalloprotein involves two steps: i) the identification of a suitable protein scaffold and ii) the incorporation of the metal binding site. The choice of the right host protein scaffold is fundamental for artificial metalloprotein design. Starting from the highly stable and brightly fluorescent protein mTFP1, we were able to engineer a low metal-affinity variant, mTFP*. Our method is straightforward and involves selection of surface exposed metal binding residues, FoldX-based prediction of changes in folding energy for mutants and stability evaluation of highest scoring mutants by MD. The simplicity of this method should make it applicable to any other host protein scaffold of choice.

Metal binding studies underline the importance of our approach: tmFRET titrations reveal at least two metal binding sites of with Kds of 2.1 µM and 260 µM at pH 7.5 for the parent mTFP1 protein, while within the investigated pH range of 5.0 – 8.0 mTFP* possesses only a low, unspecific metal affinity of roughly 1 mM. mTFP* shows an astonishing stability towards temperature (T_m = 89 °C), pH changes (stability range: 2.3 < pH < 12.6), salt concentrations (e.g. stable in 10 mM urea at 80 °C), and is even stable and soluble in high concentrations of organic solvents. The flappy N- and C-termin of mTFP1 were removed during the protein engineering process. The design principle of creating of maximizing rigidity enhances the crystallization tendency of mTFP*. Correspondingly, mTFP* protein and various mutants crystalize readily and provided electron density maps, which could be refined to outstanding resolutions of up to 1.0 Å.
Rational design of a metal binding site requires more than simply mutating residues in spatial proximity. Stable metal binding that the side-chain heteroatoms, which are supposed to bind the metal, can adopt a strain-free conformation, which i) brings them within a distance of about 4 Å and ii) and which directs the vectors of the metal binding lone pairs to a common point of intersection at a distance of about 2 Å from each heteroatom. Hence, the moderate flexibility which was introduced for residues 196-205 of beta-strand 10 and the following loop will facilitate metal binding in the adjacent cleft of mTFP*, since it permits some adjustment of side-chain orientations at a low strain penalty. However, the spectroscopic and catalytic results presented here illustrate, how fragile this balance of residue flexibility, metal coordination and induced strain is. Both of the mutants tested show stable binding of various metals, with measured single site $K_d$ for Copper and Zink in the µM range. Assuming a constant AYG$_{62-64}$ – Cu$^{2+}$ distance, the calculated Förster radius $R^0$ of this donor/acceptor pair for the HEE facial triad motif does not change significantly over the pH range 6.0 – 7.5 (table 2). This indicates a constant spectral overlap integral $J$ and thus a stable first coordination sphere of the copper center. The situation is different for the CHH binding motif in mTFP$^{\text{CHH}}$, where the transition metal induced fluorescence quenching becomes significantly more pronounced if the pH drops to 6.5 which gives more detailed view on this effect: at pH 5.0, copper addition only induces weak fluorescence quenching. When increasing the pH, quenching reaches a maximum at pH 6.5 and then drops sharply at when reaching pH 7.5. Analysis of these data reveals two underlying effects: between pH 5.0 and 6.5 $K_d$ decreases monotonously and the calculated FRET efficiency coefficient $E$ remains constant. Between pH 7.0 and 7.5 $E$ drops sharply and remains low for pH up to 9, while $K_d$ remains nearly constant. Hence there is over 2-fold increase in $E$ when the pH is lowered from 7.5 to 6.0. Assuming again a constant donor-acceptor distance, this translates into an 32 % increase in $R^0$ and correspondingly also $J$. Therefore, a major rearrangement of the metal's first coordination site needs to be assumed. X-ray structures of the metal free mTFP$^{\text{CHH}}$ reveal two different conformations of the facial triad motif: “closed” at neutral pH and “open” at slightly acidic pH. An animation visualizes that this rearrangement involves the slip of H200 underneath the flexible region of strand 10. The structure of the copper bound protein shows the closed conformation with an increased
Cu-S distance of 5.7 Å, which is too large for a bonding interaction. A model of Cu@mTFP^{CHH} suggests the exchange of H200 for C197 in the “open” conformation, indicating that this triad binding motif only realizes bidentate binding modes. Such a coordination change explains the spectroscopic observations: “open” conformation has a by approx. 1 Å shorter AYG_{62-64} – Cu^{2+} distance, which can account for about ¼ of the observed effect. The major contribution to the observed increase in quenching efficiency in the open coordination geometry should arise from a Cys-Cu ligand to metal charge transfer (LMCT), which is well established e.g. for the blue copper proteins^{49}. The resulting strong absorption band at centred around 600 nm contributes the spectral overlap integral.

Creating a catalytic function form a non-catalytic scaffold is exceedingly more challenging than the introduction of a mere metal binding motif. Catalytic activity of the coordinated metal center requires a well-defined geometry of the first coordination sphere with at least one free coordination site and a suitable electronic structure. This was the rational behind the introduction of facial triad motifs known from enzymes.

Catalytic selectivity requires a firm orientation of the bound substrate through interaction with the second coordination sphere in the binding pocket, which may also provide substrate co-activation. The high enantioselectivites achieved for the Friedl-Crafts alkylation of indole underline that the pocket size of the mTFP^{XHH} variants is of the right size to induce sufficient energy differences of the diastereotopic transition states. A directed evolution approach should hence provide access to a selective Diels-Alderase utilizing only naturally occurring amino acids for metal binding.

**Figure 3.9.** tmFRET: mTFP1, mTFP*, mbYDCHH and mbYDEHH; A) pH 6.0, B) pH 7.5
3.5 Conclusion

In summary, mTFP* is readily expressed in good yields and shows a high temperature and salt stability as well as decent solubility in organic solvents. It also provides an intrinsic fluorescent reporter that directly gives information about the folding state and the concentration of the protein. Moreover, this reporter can be scrutinized to follow upon metalation and the dynamics that come along with this process. mTFP* and respective mutants crystallize well which allow X-ray structure analysis to follow on structural changes and metal binding.

Further mutants of mTFP*, mTFP$^{XHH}$, provide insights regarding the establishment of artificial metalloenzymes on the basis of canonical binding motifs. Generally speaking this field suffers from significant drawbacks that are only rarely addressed in literature because of the prominence of supramolecular and covalent anchoring methodologies. Namely stable coordination of metal motifs, suitable host protein stabilities, mutagenic tolerance of the host and lastly stereoselective conversion are described. Reetz and co-workers suggested a Darwinian approach to optimize catalysts, and argued the necessity of an efficient expression system, a thermostable nature of the host protein that suffices easy purification methodologies and lastly quantitative conjugation of transition metal, to establish powerful artificial metalloenzymes. In the context of biocatalysis artificial co-factors are greatly incompatible and cause for significant drawbacks during optimization. Moreover, envisioning metabolic engineering on the basis of artificial metalloenzymes suggest further advances on the basis of canonical binding motifs that allow quantitative conjugation and directed evolution optimization under physiological conditions. Research here is of particular importance.
3.6 Material and Methods

3.6.1 General Remarks

Unless otherwise noted, all chemicals were purchased from Aldrich or VWR and were used without further purification. LB-Media was obtained from Luria Broth Ready Mix (Sigma Aldrich). Structure analysis were pursued with Pymol by Schrödinger INC. or YASARA (Yet Another Scientific Artificial Reality Application) by YASARA Biosciences GmbH.

3.6.2 Cloning and Mutagenesis

The gene of mTFP142 (monomeric Teal Fluorescent Protein) was ordered from Geneart™ (Life Technologies GmbH) and amplified using the following Oligonucleotides (Eurofins MWG Operon): 5’-ATTACAGGATCCGGCGTAATCAAGCCCGACATGAAG-3’ and 5’-ATTACACTCGAGTTAGTCGGTGAGTTGCGGGCCAC-3’. The amplification product was digested with BamH1 and Xho1 (New England Biolabs Inc.) and subsequently fused to a modified version of Champion™ pET303/CT-His, (Invitrogen Life Technologies GmbH) that had been supplemented with a SUMO protein coding region (Smt3p from Saccharomyces cerevisiae). The resulting plasmid pET303/SUMO_mTFP1.1 was then electrotransformed into E. Coli BL21 (DE3) Gold (Stratagene™, Agilent Technologies) for further plasmid purification and mutagenesis.

3.6.3 Site-Directed Mutagenesis

Site-directed mutagenesis was performed on pET303/SUMO_mTFP1.1 (denoted as wild-type) to obtain our designed mutant mTFP* - the basis to all relevant mutants of this study. A detailed outline of all Mutants and Primers is shown in the Supplementary section. Primers were designed according to the guidelines of the QuikChange Kit (Stratagene™, Agilent Technologies) and subsequently obtained from Eurofins MWG Operon. Mutant plasmids were transformed into E. Coli BL21 (DE3) Gold (Stratagene™, Agilent Technologies) and purified using peqGOLD Plasmid MiniPrep Kit II (peqlab Biotechnologie GmbH). Purities and successful mutagenesis were checked by 1% Agarose Electrophoresis and Sanger Sequencing (KAUST Core Facilities).
3.6.4 Expression and Purification

For production of mTFP*, LB Medium that contained 180mg L⁻¹ ampicillin was inoculated with E. Coli BL21 (DE3) Gold cells, purchased from Agilent Technologies, that had been transformed with pET303/SUMO_mTFP*. Expression yields were improved by picking fluorescent colonies, from LB Agar containing 180 mg L⁻¹ ampicillin and 1 mM IPTG, for pre-culture. Pre-cultures were left to grow overnight at 37 °C and were further used to inoculate expression cultures under standard protocol. Induction occurred at OD 0.5 – 0.6, using 1 mM IPTG. Cells were left to express for 48 hours at 20 °C. Cultures were then centrifuged at 4500 rpm for 30 minutes and cooled to 4 °C. Pellets were washed with Lysis Buffer A, containing 100 mM Tris/HCl pH 7.4, 500 mM NaCl and 20mM Imidazole. Before lysis complete protease Inhibitor cocktail (Roche), was added together with DNase. Lysis occurred using Frenchpress (FIRM). Lysates were immediately centrifuged at 25000 rpm and also cooled to 4 °C. Respective supernatant was prawn to Ni-Affinity Chromatographic principles and eluted from the column between 20 and 35% of Elution Buffer B containing 100 mM Tris/HCl pH 7.4, 500 mM NaCl and 500mM Imidazole. SUMO protease was added to the eluted protein and the mixture then dialysed overnight, at 4v°C, against Buffer A. Further Ni-Affinity Chromatography followed to remove the SUMO-Tag. Pure Protein was collected from the flow-through whereas elution with Buffer B showed a distinct elution of residual SUMO-Tag and Protease. If necessary, Batches of almost pure protein were further purified using Anion Exchange Chromatography. Anion Exchange Chromatography occurred using Q-Sepherose Colum. Protein samples were dialysed against 100 mM Tris/HCl pH 7.4 and loaded onto the column. Elution occurred with a smooth gradient up 20 % of the same Buffer containing additional 1 M NaCl. Purity was determined by SDS-PAGE gel electrophoresis at 12.5% acrylamide.

3.6.5 Heat Purification

Alternatively, to Anion Exchange Chromatography, Heat Purification principles were applied to obtain purest protein. Up to 5% of Glycerol added to almost pure protein and subsequent heating to 70 °C for 15 minutes caused protein based impurities to precipitate.
Centrifugation at 7100 rpm allowed the separation of impurities and condensed pure samples.

### 3.6.6 Protein Storage

After protein purification pure protein samples were dialysed against 50 mM Tris/HCl pH 7.4 or 50 mM HEPES pH 7.9 depending on further usage. Samples that were to be stored for a longer period were frozen at -80 °C in either Buffer.

### 3.6.7 Mass Spectrometry

Protein mass spectrometry was either carried out on a BRUKER maXis HD™ ESI-TOF. For ESI-TOF 0.04mM of pure protein sample was dialysed against 5% ACN, 0.1% FA and prone to either direct injection or HPLC (Agilent Technologies), C4 Column (Column volume 5ml). Elution occurred with a flowrate of 0.5µl/min and a gradient to 80% ACN, 0.1%FA at 8min. Subsequently fractions were recorded according to standard procedure.

### 3.6.8 Spectral Analysis

Absorbance and Fluorescence measurements were performed with pure protein samples, 0.04mM mTFP* or respective mutant in 50mM Tris/HCl, pH 7.4 using a TECAN INFINITE M1000 according to standard procedure.

### 3.6.9 Stability Studies

Melting curves of mTFP* or respective mutant were obtained from circular dichroism measurements using a JASCO J-815 Spectropolarimeter. 400 µL of protein solution, 4 µM mTFP* or respective mutant in 50mM Tris/HCl pH 7.4, was prepared and given into a Hellma Precision Cell (Type no. 110 QS, 1mm light path). Temperature resolved (50 – 98 °C) far UV Spectra (210 – 200 nm) were recorded.

### 3.6.10 Solvent Screen:

2 mM mTFP* in 50mM Tris/HCl, pH 7.4 was diluted to 0.05mM with a mixture of water and organic solvent (acetone, 2-butanol, t-butanol, methanol, diglyme, dioxan, dimethylformamide, dimethylsulfoxide, ethanol, hexanol, 2-propanol, tetrafluoroethylene...
and tetrahydrofuran). Concentrations varied from 10 to 90% organic solvent. Fluorescence readings were accomplished with a TECAN INFINITE M1000.

3.6.11 Metal Incubation

50 µl of protein solution, 0.32 mM mTFP* or respective mutant in 50mM Tris/HCl pH 7.4, was mixed with the according metal at a ratio of 1:1, 1:5, 1:10 and 1:25. Incubation occurred at 25 °C or 60 °C for 1 hour. Metal salts were primarily dissolved in 50mM Tris/HCl pH 7.4 ± DMSO to ensure solution. Metal addition was tested with ESI-TOF measurements as described above. Regarding affinity studies using Fluorescence, samples of mTFP1, mTFP* and mTFP\textsuperscript{CHH} were dissolved in 50 mM HEPES pH 7.9, 50 mM NaCl and diluted with respective metal solutions to a final concentration of 0.2 mM. Metal binding was monitored throughout the reduction of fluorescence over a period of 18h and compared with readings after samples had been dialysed. Signal recovery was assumed to indicate affinities towards metals. Fluorescence readings were accomplished with a TECAN INFINITE M1000

3.6.12 Crystallography

mTFP* and the respective mutant were concentrated to 50 mg/mL (in 50 mM HEPES, pH 7.9) using a 10 kDa NMWL Amicon Ultra Centrifugal Filter Device (Millipore, Billerica, MA). Crystals of mTFP and respective mutant were grown at 20 °C, overnight to their final size of about 0.5 x 0.1 x 0.1 mm\textsuperscript{3}, using the sitting drop vapour diffusion method. Drops contained 0.2 µL mTFP and 0.2 µL of reservoir buffer (100 mM MES, pH 6.5 and 21.5% PEG 3000). Co-crystallisation were performed by adding respective metal salts to a final concentration of 1.25mM. For the mTFP*:Pd dataset, mTFP* crystals were soaked for 24 h in 1 mM PdCl\textsubscript{2}. For X-ray measurements, crystals were soaked for 1 min in a mixture of mother liquor and 50% glycerol (1:1. v/v) and subsequently flash-frozen in liquid nitrogen. Datasets were recorded using synchrotron radiation at $\lambda = 1.0$ Å at the beam line X06SA, Swiss Light Source (SLS), Villigen, Switzerland. X-ray intensities were evaluated with XDS.\textsuperscript{50} The space group of mTFP* was P2\textsubscript{1} (monoclinic) with unit cell dimensions of $a = 38.5$ Å, $b = 85.2$ Å, and $c = 63.1$ Å for the native dataset and $a = 34.5$ Å, $b = 83.6$ Å, and $c = 37.8$ Å for the mTFP*:Pd
dataset. Phases were obtained by molecular replacement using PHASER and the coordinates of mTFP1 (PDB code 2HQK)\(^{42,51}\). Model building was performed with Coot. Rigid body, TLS (Translation/Libration/Screw) and positional refinements with REFMAC5 resulted in the final models\(^{52}\). The quality of the stereochemistry was confirmed by the Ramachandran plot determined with PROCHECK\(^{53}\).

### 3.6.13 tmFRET Measurements

All measurements were recorded on a Spectrofluorometer (PTI, New Jersey). Protein Samples were diluted to the final concentration of 1\(\mu\)M using 50mM MES \(\pm\) 100mM NaCl pH 6 and 7.5 and heated to 25 °C or 40 °C. Copper and Zinc titrations were accomplished with a series of Cu(NO\(_3\))\(_2\) and Zn(NO\(_3\))\(_2\) stock solutions, ranging from 100 to 0.01mM. Indirect titrations were fitted with 1.5\(\mu\)M and 3\(\mu\)M Cu(NO\(_3\))\(_2\). Spectra were recorded from 480 to 510nm with a bandwidth filter of 5nm.

### 3.6.14 Fluorescent Lifetimes

Lifetimes measurements were recorded on a Spectrofluorometer (PTI, New Jersey) with white laser illumination (WhiteLase) at 20 MHz (\(\lambda_{\text{ex}} = 462\) nm; \(\lambda_{\text{em}} = 495\) nm). Intensity measurements were done on the same instrument with Xenon light illumination.

### 3.6.15 Quantification of Metal Binding Constants

Data analysis was performed similar to the method described by Taraska and co-workers\(^{37,39}\):

Fluorescence values were averaged over an 10 nm window surrounding the fluorescence emission peak at 492 nm. All spectra were blank corrected and normalized to the fluorescence before the addition of metals. Fluorescence was corrected for the inner filter effect of metals\(^{54}\).

Fluorescence quenching through \(tmFRET\) arises from both, collisional quenching (dynamic quenching) and metal ions coordinated at a binding sites (static quenching, \(tmFRET\)). Equation (1)
\[
\frac{F_M}{F_0} = \frac{1}{1 + [M] \cdot K_{SV}} \quad (1)
\]

accounts for dynamic quenching through collisions and equation (2)

\[
\frac{F_M}{F_0} = \prod_{i=1}^{n} \left( 1 - E_i \cdot f_i \right) \quad (2)
\]

describes the tmFRET quenching of metals bound to a binding site. \( F_M \) is the fluorescence of the donor in the presence of metal, \( F_0 \) is the fluorescence of the donor without metal, [M] is the metal concentration, \( K_{SV} \) is the Stern-Volmer constant for collisional quenching (dynamic quenching), \( E_i \) are the FRET efficiencies between protein fluorophore and coordinated metal ion and \( f_i \) are the occupation factors of the respective binding sites. Assuming that static and dynamic quenching are occurring independently, fluorescence in the presence of metal ions can be calculated by equation (3)

\[
\frac{F_M}{F_0} = \frac{1}{1 + [M] \cdot K_{SV}} \cdot \prod_{i=1}^{n} \left( 1 - E_i \cdot f_i \right) \quad (3)
\]

The individual binding site occupation factors are given by

\[
f_i = \frac{[FP_iM]}{[FP_0]} \quad (4)
\]

where \([FP_iM]\) stands for the concentration of fluorescent protein with metal bound at binding site \( i \) (note that the same fluorescent protein molecule may have more than one metal ion bound at a given time, each of them in a different binding site) and \([FP_0]\) is the total concentration of fluorescent protein.

The FPiM can be derived from the dissociation constants. For the case of only one binding site \( i = 1 \), the dissociation constant is:

\[
K_{d,1} = \frac{[FP] \cdot [M]}{[FP_1M]} = \frac{([FP_0] - [FP_1M]) \cdot ([M_0] - [FP_1M])}{[FP_1M]} \quad (5)
\]

resulting in
\[ [FP_iM] = \frac{1}{2} \left( [M_0] + [FP_0] + K_{d,i} - \sqrt{([M_0] + [FP_0] + K_{d,i})^2 - 4 \cdot [M_0] \cdot [FP_0]} \right) \]  
(6)

and

\[ f_i = \frac{\left( [M_0] + [FP_0] + K_{d,i} - \sqrt{([M_0] + [FP_0] + K_{d,i})^2 - 4 \cdot [M_0] \cdot [FP_0]} \right)}{2 \cdot [FP_0]} \]  
(7)

with \([M_0]\) being the total metal concentration. The depletion of available binding sites and metal in solution is only noticeable from the data if \(K_d\) is smaller than both, \([FP_0]\) and \([M_0]\). For experiments, where the binding does not reduce the concentration of unbound metal in solution (\(K_d > [FP_0]/5\)), \([M_0]\) can be assumed to be constant and equation (7) can be simplified to equation (8), which was previously used by Taraska et al.\(^{37,38}\) (note that \(K_d\) values, which are smaller than the protein concentration can not be determined reliably).

\[ f_i = \frac{1}{1 + \frac{K_{d,i}}{[M_0]}} \]  
(8)

For two binding sites, two cases exist:

i) two weak binding sites (\(K_{d,i} > [FP_0]\)). In this case equation (8) is sufficient to describe the metal binding. This case is typical for binding on non-engineered protein surfaces. This approximation was therefore used to determine any background binding, which does not arise from an engineered site and the respective dissociation are symbolized as \(K_{d,i}^{\text{back}}\) with the corresponding FRET efficiencies \(E_i^{\text{back}}\).

ii) combination of a weak and a high affinity binding site. In this case, the weak binding site is described as in case i), while the occupation factor for the high affinity binding site is calculated from equation (7). Since equation (7) is typically required for engineered (or natural) metal binding sites, the respective symbolized as \(K_{d}^{\text{site}}\) with the corresponding FRET efficiencies \(E^{\text{site}}\).
Theoretically, the case of two strong binding sites could occur, too, but it is not treated here. The equations for this case can be solved analytically, however experimentally it will be difficult to distinguish two binding sites with both $K_d$ within two orders of magnitude from one binding site with an increased Fret efficiency. $K_d$'s below 10 nM are at the experimental limit, since they will require protein concentrations in the nM range to determine binding constants reliably.

Taking the discussion above into account, all fluorescent data where fitted to equation (3).

Fluorescence data mTFP1 and mTFP*, which do not contain an engineered metal sites were fitted the using equation (8) for the binding site occupation. The number of binding sites was increased until no significant improvement in the RMSD value (less than 3 % improvement) was achieved. For the same $pH$, the same $K_{SV}$ was used for both proteins. The number of binding sites increases with more basic $pH$ ($pH = 6.0$, one binding site for mTFP1, $pH = 7.5$ two binding sites for mTFP1 and one weak binding sites for mTFP*). The resulting constants for mTFP* ($pH = 6.0$ only $K_{SV}$, $pH = 7.5$ $K_{SV}$, $K_{d}^{\text{back}}$ and $E^{\text{back}}$) were used to describe the background for proteins with engineered metal binding site (mTFP$^{\text{CHH}}$ and mTFP$^{\text{CHH}}$). Equation (10) was used to determine the dissociation constant of this binding site $K_d^{\text{site}}$ and the corresponding FRET efficiencies $E^{\text{site}}$ (binding site occupation factor $f_i$ from equation (7)).

$$\frac{F_M}{F_0} = \frac{1}{1+[M] \cdot K_{SV}} \cdot (1 - E^{\text{back}} \cdot f^{\text{back}}) \cdot (1 - E^{\text{site}} \cdot f^{\text{site}})$$

(10)

with

$$f^{\text{back}} = \frac{1}{1 + \frac{K_d^{\text{back}}}{[M]}}$$

(11)

Where needed, distances between the fluorophore and the metal were taken from X-ray structures and the respective FRET efficiencies were calculated using the
Förster equation: \( R = R_0 / (E - 1)^{1/6} \). \( R \) is the distance between the fluorophore and the metal and \( R_0 \) is the Förster distance. \( R_0 \) strongly depends on transition metal quencher and was taken from literature. \[\text{Error! Bookmark not defined.}\]

Fluorescence data from competition titration experiments were fitted to equation (10) using the occupation factors given in equations (12) and (13). Both equations assume, that binding does not reduce the free metal concentrations in solution, which is valid in case of \((K_d > [FP_0]/5)\), which is valid for the experimental set-up used here.

\[
f_{\text{comp,back}}^{\text{site}} = \frac{1}{1 + \frac{K_{d,Cu}^{\text{back}}}{[Cu]} + \frac{K_{d,Zn}^{\text{back}}}{[Zn]} + \frac{[Zn]}{[Cu]}}
\]

\[
f_{\text{comp,site}}^{\text{back}} = \frac{1}{1 + \frac{K_{d,Cu}^{\text{site}}}{[Cu]} + \frac{K_{d,Zn}^{\text{site}}}{[Zn]} + \frac{[Zn]}{[Cu]}}
\]

The Förster-Radius \( R_0 \) of a metal-chromophore \( tm \)FRET pair was estimated from the FRET efficiency parameters using equation 14:

\[
R_0 = \frac{R}{\left(\frac{1}{E^{\text{site}}} - 1\right)^{1/6}}
\]

The distances \( R \) between the fluorophore and the metal were taken from X-ray structures.
3.6.16 Diels Alderase Activity

Cyclopentadiene (2) was freshly prepared from dicyclopentadiene and stored at -80 °C. Azachalcone 1 was prepared according to literature\(^55\) hybrid catalysts were prepared by mixing a solution of the desired mTFP* variant (900 \(\mu\) l, 4.9 mg/ml in 10 mL MES-HCl pH 5.5) with CuSO\(_4\) solution (100 \(\mu\) L 1.2mM CuSO\(_4\) in 10 mM MES-HCl pH 5.5) and incubation for 2 h (400 rpm, 20 ° C, Thermomixer comfort, Eppendorf). These hybrid catalysts were immediately used for Diels-Alder reactions: a solution of 1 (42 mM in Acetone, 25 \(\mu\) l) were added to the catalyst solution and mixed immediately by inverting the tube 3 times. Then, a solution of 2 (210 mM in Acetone, 25 \(\mu\) l) was added and mixed immediately as in the last step. The tube was placed immediately in a pre-cooled shaking incubator (12 °C, 400 rpm, Thermomixer comfort, Eppendorf). The standard reaction time was 24 h. Reactions were extracted with diethyl ether (2 \(\times\) 1 ml). The combined organic phase was dried by filtering over and the solvent was evaporated. The samples were analyzed by HPLC using anisole as internal standard (column: 250 mm Chiralcel OD-H, 4.6 mm inner diameter; n-heptane/2-propanol = 99:1; 0.5 ml/min, 25 °C; UV-detector 220 nm; retention times: 10.4 min (anisole), 14.2 min (exo-3), 16.1 min (exo-3'), 20.1 min (endo-3), 25.4 min (endo-3'), 31.7 min (3)). The absolute configuration of the enantiomers 3 has not been established in literature\(^56\).

3.6.17 Computational Details:

**FOLDX Calculation**

Crystal structure of a monomeric cyan fluorescent protein (mTFP, Pdb Id 2HQK) was used in our study. First, surface exposed HIS, MET, CYS were selected from mTFP protein. Hydrogen bonding interaction of these residues was analyzed. Then we have used ConSurf to identify degree of conservation of the amino-acid sites among there close sequence homologues at those positions.\(^57\) We observed that H21, M109, H119, H168, H169, and H204 (Numbering in 2HQK is H25, M113, H123, H172, H173, and H204) were exposed to the surface. We have used YASARA program (“Yet Another Scientific Artificial Reality Application,” http://www.yasara.com) with FoldX plugin to scan the appropriate position.\(^44\) The relative free energy of folding (\(\Delta\Delta G\)) for mTFP and
mutants was predicted using FoldX, version 3.0. First, each amino acid was mutated with other remaining 17 amino acids and their change in free energy was calculated for each mutation. After this, amino acid which shows favourable energy was shortlisted. Then different combinations were carried out to get stability of the overall protein. The mutant which shows good favourable energy was selected for molecular dynamics simulation and also to design a library of mutants. Homology model for mTFP* (H21Y, M109L, H119Y, H168Y, H169Y, and H204Y) was built using YASARA.

3.6.18 Force Field Parameters for the Chromophore

The chromophore in the protein was built using combination of A62, Y63, and G64. The parameters for the chromophore were partially derived from reference 58 and 59. DFT calculation was carried out using Gaussian 09 using B3LYP functional and 6-31+G* basis set. The equilibrium bond lengths, angles, and dihedrals were taken from ab initio DFT optimized structures of an anionic model chromophore, consisting of the two aromatic rings linked by the bridging carbon and two methyl groups substituting the linkage of the chromophore to the protein. The same models were used to derive the partial charges (reported in Fig. 1), by means of a standard procedure consisting in the fit of the electrostatic potential computed at the HF/6-31+G* level.61.

Figure S3.10. Anionic form of the chromophore (AYG). For each atom the calculated partial charge was reported together with the amber atom type.
3.6.19 Molecular Dynamics

Minimization and molecular dynamics simulation was carried out using Amber 11 software package\textsuperscript{62} with the AMBER99SB force field for amino acid residues. Wild type protein and mutant was solvated using TIP3P water molecules. We followed the standard protocol for a MD simulation, which consists of an initial minimization, followed by gradual heating of the system. The solute is initially minimized by 1000 steps of steepest descent and conjugate gradient using constraint on protein atoms. Then full system minimization was carried out for 1000 steps. Temperature of system was increased upto 298K in 3 steps using NVT dynamics. The heating procedure starts with 10 ps of MD at 50K, followed by 20 ps at 150K and 30 ps at 298K. Next, the systems were allowed to equilibrate in a 5 ns NPT ensemble at constant temperature of 298K and pressure of 1 atm (Berendsen algorithm). Finally, 50 ns long constant pressure-constant temperature (NPT) production dynamics was performed at 298 K and 1 atm pressure; with temperature regulation achieved using the Berendsen weak coupling method (1 ps time constant for heat bath coupling and 0.5 ps for pressure relaxation time). Equations of motion were solved using the Verlet leapfrog algorithm, with an integration step of 2 fs and the trajectory snapshots were saved after 2 ps. SHAKE constraints were imposed on all the heavy-atom-hydrogen covalent bonds. The non-bonded cutoff was kept at 9 Å, and long-range electrostatic interactions were treated by the Particle-Mesh Ewald (PME) method.

In order to see the effect of temperature on stability of protein, further 50 ns simulation was carried out at 343K. The amber utility program ptraj (AmberTools) was used to calculate root mean square deviation (RMSD) and B-factor. Visualization was carried out using VMD 1.8 program.\textsuperscript{63}
3.7 References


3.8 Supplementary Material

Molecular Biology

> mTFP1 711 bp (GenBank: DQ676819)

ATGGTGAGCAAGGGCGAGGAGACCAACATGGGCCTATCAAGGCCGACATGAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCCACGCCTTCGTGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACACCCACACATCAACCTGGAGGTGAAGGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGGGAAGAGGGAGCCCCCCTGCCCTTGCTACGACATTCTGGAACCGCAGG GAACCTCTTCTACGAGATACACCTCAAGGGCGAGAATTCCTCCCAACGCGCCCTCG ATGCAAGAAGAAGACCGAGCCTGGGAGCGCTCCACCCGAGAGGATGTACGTGCGCGACGC GTGCTGAAGGGCGACGTCAAGCAAGTGCTGCTGGGAGGCAGGGCAGGCCACACCGCGTT GACTTCAAGACCATCCTACAGGGGCAAGAGGCGGTGAAGGTGAGGTGAAGGTGACTGCACGCCTACCACTTGTG GACCAACCCGCTCGAGATCTGAAACCACACGACAAGGACTACAACAAAGGGCTACGCTTTACGAG AGCGCGTGGCCCGAACCCTCCCGACCGAGATGACGTGCAAGGTAA

Figure 3.11. Sequence mTFP1 (underscore = deleted nucleotides)
Figure S3.13. pET303/SUMO_mTFP1.1
Figure S3.14. Align mTFP1.1/mTFP*
Table S3.5. Oligonucleotides to establish mTFP*

<table>
<thead>
<tr>
<th>MUTATION (mTFP* → mTFP*)</th>
<th>POSITION (2HQK)</th>
<th>PRIMER</th>
</tr>
</thead>
<tbody>
<tr>
<td>H30Y</td>
<td>H25</td>
<td>5'-GGGCAACGTGAATGGCTATGCTTCTCGTGATCGGAGG-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-CCTCGATCAGAGGATAGGGATCCTACGTTGCCC-3'</td>
</tr>
<tr>
<td>M118L</td>
<td>M113</td>
<td>5'-AGTCCGACATCTCCTGGAGGGAGGACTCC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GGAGTCCTCGCTTCAAGGAGGACTGTTCC-3'</td>
</tr>
<tr>
<td>H128Y</td>
<td>H123</td>
<td>5'-CTCCTTCATCTACGAGATATATCTCAAGGGGCGAGAAGCTTC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GAAGTTCCTCCCTGAGATATATCTCAAGGAGGAGG-3'</td>
</tr>
<tr>
<td>H177Y, H178Y</td>
<td>H172, H173</td>
<td>5'-TGCTGGAGGGCGGCGCTATTATCGCGTTAAGATCAAGAC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GTCTTGAAGTCAGCGGATAATAGCCGCGCGCCCTCCAGCA-3'</td>
</tr>
<tr>
<td>H209Y</td>
<td>H204</td>
<td>5'-CCGCACTCAGATCTCCTGAAGATCAAGGACTACAAAGG-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-CCTTGGTGAGTTCTTCAGTTAGTTCAGATCGGAGG-3'</td>
</tr>
</tbody>
</table>
Figure S3.15. pET303/SUMO_mTFP*

mTFP1       MVSKGEETTMGVIKPDMKIKLMGKGEVNGHAFIEGEGERGKPYDGFNTTINLEVEKAPLP
mTFP*       -----------SVIKPDMKIKLMEGNGVNGYAFIEGEGERGKPYDGFNTTINLEVEKAPLP
mTFPCHH     -----------SVIKPDMKIKLMEGNGVNGYAFIEGEGERGKPYDGFNTTINLEVEKAPLP
mTFPEHH     -----------SVIKPDMKIKLMEGNGVNGYAFIEGEGERGKPYDGFNTTINLEVEKAPLP

mTFP1       FSYDILTTAFAYGNRAFTKYPDDIPNYFQSFPEGYSWERTMTFEKIGVKSDDIMEE
mTFP*       FSYDILTTAFAYGNRAFTKYPDDIPNYFQSFPEGYSWERTMTFEKIGVKSDDISLEE
mTFPCHH     FSYDILTTAFAYGNRAFTKYPDDIPNYFQSFPEGYSWERTMTFEKIGVKSDDISLEE
mTFPEHH     FSYDILTTAFAYGNRAFTKYPDDIPNYFQSFPEGYSWERTMTFEKIGVKSDDISLEE

mTFP1       DSFIYEIHLKGENFPFPVQKKTGWDASERTMYVRDGVLKGDVHKLLLEGIDHRV
mTFP*       DSFIYEIHLKGENFPFPVQKKTGWDASERTMYVRDGVLKGDVHKLLLEGIDYRV
mTFPCHH     DSFIYEIHLKGENFPFPVQKKTGWDASERTMYVRDGVLKGDVHKLLLEGIDYRV
mTFPEHH     DSFIYEIHLKGENFPFPVQKKTGWDASERTMYVRDGVLKGDVHKLLLEGIDYRV

mTFP1       DFKTIYRAKKAVKLIPDHYFVDRIEILHNHDKDYMKTIVYESAVARNSTDGMELYK
mTFP*       DFKTIYRAKKAVKLIPDHYFVDRIEILNYDKDKHMKIVYESAVARNSTD-------
mTFPCHH     DFKTIYRAKKAVKLIPDHYFVDRIEILNYDKDKHMKIVYESAVARNSTD-------
mTFPEHH     DFKTIYRAKKAVKLIPDHYFVDRIEILNYDKDKHMKIVYESAVARNSTD-------
Characterization of mTFP*

Figure S3.16. SDS-PAGE: mTFP* purification, (1) Sumo conjugated mTFP*, (2) SUMO-digested mTFP*, (3) pure mTFP*

Figure S3.17. ESI-TOF: mTFP*, indicates matured Mass of 25154 Da
Figure S3.18. Chromophore Formation (A-Y-G)

Figure S3.19. mTFP*: Concentration dependent Absorbance (A) and Fluorescence (B) (ex: 468nm em: 495nm)

Figure S3.20. Circular Dichroism Scan (mTFP*) indicating structural perturbation at temperatures > 89°C
Figure S3.21. Urea Stability: 0.05 mM mTFP* (50mM Tris/HCl, pH 7.4) incubated for 1h at room temperature with 6, 8 and 10 M Urea, and subsequent temperature resolved melt. A, B and C indicate stability measures throughout the quantification of specific spectroscopic features (absorption: 462nm, fluorescence: 492nm)

Figure 3.22. Aprotic and Protic Solvent Screen: 0.05mM mTFP* (50mM Tris/HCl, pH 7.4) incubated with a subset of aprotic and protic solvents for 24h at room temperature. Integrity of mTFP* is indicated throughout overall fluorescence preservation – unfolding significantly depletes fluorescence.
Table S3.6. Metal Salts for Conjugation Test: The following metal salts did show any influence on mass spectra or fluorescence measurements:

<table>
<thead>
<tr>
<th>Metal Salt</th>
<th>Adduction as judged by Mass</th>
<th>Adduction as judged by Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>before dialysis</td>
<td>after dialysis</td>
</tr>
<tr>
<td>Palladium(II) chloride</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>Nickel(II) chloride</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Copper(II) chloride</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Rhodium(III) chloride</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Iridium(III) chloride</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ruthenium(III) chloride</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.23. ESI-TOF: A, B) mTFP* Heavy Metal Soak with Pd2+. Low concentrations of metal ions show no additional mass peaks (A). Palladium adduct to mTFP*, 25255 Da at 4 and more eq. After dialysis, originally Pd2+ containing sample lacks adduct species.

Structural Characterization
**Table S3.7. Crystallographic Parameters**

<table>
<thead>
<tr>
<th>Crystallographic Parameters</th>
<th>mTFP*</th>
<th>mTFPM</th>
<th>mTFP-Cu</th>
<th>mYDCHI_ext</th>
<th>mYDCHI_catal</th>
<th>mYDCHI_Cu</th>
<th>mYDCHI_ext</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td><strong>Unit cell parameters (Å)</strong></td>
<td>a=38.5, b=38.5, c=67.3</td>
<td>a=38.5, b=38.5, c=67.3</td>
<td>a=38.5, b=38.5, c=67.3</td>
<td>a=38.5, b=38.5, c=67.3</td>
<td>a=38.5, b=38.5, c=67.3</td>
<td>a=38.5, b=38.5, c=67.3</td>
<td>a=38.5, b=38.5, c=67.3</td>
</tr>
<tr>
<td><strong>Resolution range (Å)</strong></td>
<td>10.6-3.0</td>
<td>10.6-3.0</td>
<td>10.6-3.0</td>
<td>10.6-3.0</td>
<td>10.6-3.0</td>
<td>10.6-3.0</td>
<td>10.6-3.0</td>
</tr>
<tr>
<td><strong>R work,</strong></td>
<td>52.1, Rfree</td>
<td>52.1, Rfree</td>
<td>52.1, Rfree</td>
<td>52.1, Rfree</td>
<td>52.1, Rfree</td>
<td>52.1, Rfree</td>
<td>52.1, Rfree</td>
</tr>
<tr>
<td><strong>Rmerge</strong></td>
<td>2.02</td>
<td>2.02</td>
<td>2.02</td>
<td>2.02</td>
<td>2.02</td>
<td>2.02</td>
<td>2.02</td>
</tr>
<tr>
<td><strong>No. of reflections</strong></td>
<td>27,800</td>
<td>27,800</td>
<td>27,800</td>
<td>27,800</td>
<td>27,800</td>
<td>27,800</td>
<td>27,800</td>
</tr>
<tr>
<td><strong>Rmerge (%)</strong></td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>I, I</strong></td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Data collection**

- Wavelength: 0.65 Å
- Resolution range: 20.0-4.0 Å
- Rmerge: 0.06
- Completeness: 99.99%
- I / I (AVG): 1.00
- Diffracted region: -1.21-2.20

**Refinement**

- Resolution range: 20.0-4.0 Å
- No. of reflections: 27,800
- No. of water molecules: 8
- No. of hydrogen atoms: 111
- Rmerge: 0.18
- I / I (AVG): 1.00
- RMSD bond lengths (Å): 0.06
- Average B-factor (Å²): 24.0

**Figure S3.24. B-Factor Plot mTFP1 (A, PDB: 2HQK) vs. mTFP* (B, PDB: 4Q9W)**
Figure 3.25. mTFP*: Pd, K9, D112 coordinating Pd2+ (PDB: 4Q9X)
Modification and Conjugation

Table S3.8. Oligonucleotides to establish mutants of mTFP*

<table>
<thead>
<tr>
<th>MUTATION (mTFP*)</th>
<th>POSITION (2HQK)</th>
<th>PRIMER</th>
</tr>
</thead>
<tbody>
<tr>
<td>I197C, Y200H</td>
<td>I201, Y204</td>
<td>5'-TGGACCACCGCATCGAGTGCCTGAACCATGACAAGGACTACAAC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GTTGTAGTCCTTTGTCATGGTTCAGGCACTCGATGCGGTGGTCCA-3'</td>
</tr>
<tr>
<td>Y204H</td>
<td>Y208</td>
<td>5'-CTGAACATATGACAAAGGACACAACAAAGGTGACCCTT-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-AAACGGTACGGTTGTCATGTTGTGCATAGTTTCA-3'</td>
</tr>
<tr>
<td>I197E</td>
<td>I201</td>
<td>5'-TGGACCACCGCATCGAGGAGCTGAACCATGACAAGGAC-3'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5'-GTCTCGTCATGGTTCAGCTCGATGCGGTGGTCCA-3'</td>
</tr>
</tbody>
</table>

Figure 3.26. B-Factor Comparison of mTFP* (PDB: 4Q9W), mTFPCHH, mTFPCHH_Cu (PDB: 4R6D)

Figure S3.27. Temperature Stability of mTFPCHH/EHH, circular dichroism at 205nm indicates unfolded fraction.
Figure S3.28. ESI-TOF: Cu2+ and Pd2+ Soak of mTFPCHH/EHH
Metal Affinity

**Figure 3.29.** Metal Binding: A - F) Fluorescence measurement of mTFP1, mTFP* and mTFPCHH at 0.2mM (50mM HEPES pH 7.9, 50mM NaCl). Samples were soaked with 1, 5, 10 and 25 eq. of Cu2+, Pd2+, Rh2+ and Ni2+ for 0h (left column) and over night (right column). Binding reversibility confirmed throughout signal recovery after dialysis.
Figure S3.30. Relative fluorescent lifetime of mTFP* and mTFPCHH
Figure 3.31. Distance Mapping of Coordinating Motifs: A) mbYDCHH (open); B) mbYDCHH (closed); C) mbYDCHH_Cu; D) mbYDEHH; Data derived from crystal structure (Table S 3.3)

Figure 3.32. Förster Radius Estimations: A) mbYDCHH_Cu; B) mbYDCHH_open_Cu; C) mbYDEHH_Cu; D) mTFP*C204_phen_Cu; Closest distances between transition metal and chromophore are indicated. Förster Radii are approximated accordingly. Distances in A derived from crystal structure; Distances in B-D derived from Model.
4 Chapter 4 Incorporation of Unnatural Amino Acid for Bioconjugation Reactions

4.1 Abstract
Protein engineering through unnatural amino acid (UAA) incorporation has advanced significantly during the last decade. The exploitation of unnatural amino acids (UAAs) in bioconjugations is ultimate due to their ability to present a degree of bioorthogonality and specificity. In order to elucidate optimal conditions for the bioconjugation of UAAs containing proteins, we synthesized four UAAs own an azide, alkyne, or halide functional group. All four UAAs were then incorporated into mTFP* fluorescent protein using a promiscuous aminoacyl-tRNA synthetase. Incorporation sites were selected through computational studies and homology modeling. The highlighted positions in mTFP* proves to be suitable for incorporation without any disruption or missfolding of the protein. The incorporation efficiency was then analyzed for optimal yield via mass spectroscopy and fluorescent intensity. Interestingly, the Engineered protein mTFP* allows incorporation at four different positions and was found to be dependent on both, the type of UAAs and selected incorporation site. Overall, these findings provide a better understanding of various parameters that can be optimized for the efficient preparation of bioconjugates. The description of the site-specific incorporation common features is presented in this chapter, but its focus mainly on the successful application of the site-specific integration and implementation. Furthermore, the emphasis is laid on work that involves the integration of UAAs containing reactive chemical handles for click chemistry applications.

In the chapter, we address successful incorporation of \textit{para-}(L)-iodophenylalanine(pIF) and its posttranslational MIYAUARA borylation on mTFP*. This posttranslational modification utilizing the iodine functionality which can be easily transformed into a boronic acid by MIYAUARA borylation on mTFP* host protein.
4.2 Introduction

Utilizing chemical reactions to alter biological systems is vibrant research area. Conjugation reactions involving bioactive molecules and either molecular surfaces or probes are of a particular interest. The synthesized biconjugates can be used as diagnostic tools or therapeutic agents, and are critical to fundamental research\(^1\)-\(^6\). One of the important strategies for producing biconjugates involves the creation of strong covalent bonds between bioactive molecules and their partners (contrary to the processes involving encapsulation or adsorption)\(^7\),\(^8\). Because of their wide areas of applications, researchers are always interested in the development and optimization of methods that can generate such molecules. In order for such reactions be successful, specific conditions have to be maintained based on the chemical reactions involved. Essentially, the chemical reactions used must be compatible with the existing biological setting. This implies that reactions should proceed at fast rates under normal physiological environment (neutral pH and temperature at \(37^\circ\) C). The reactions have to be inert to a plethora of chemical functionalities that are found in biological systems in order to provide chemoselectivity\(^9\). If the above conditions are met, the chemistry is regarded as biorthogonal. Although a number of biorthogonal reactions have been described, there are a number of issues that impede their general applications\(^10\). Examples of such biorthogonal reactions are the Sonogashira and the SUZUKI cross coupling reactions, 1,3-dipolar cycloadditions, photo-cross-linking, oxime ligations, and the recently developed Glaser-Hay couplings\(^11\)-\(^18\). The major hurdle for bioconjugations is the introduction of a new functionality into the protein/molecule; however, this can be accomplished efficiently using site-specific unnatural amino acids (UAAs) inclusion\(^19\).

This novel chemical functionality can serve as a reactive handle to facilitate the generation of a stable bioconjugate in a well-defined fashion. As a matter of fact, the application of UAAs in the process of bioconjugation has grown to become a field, which is capable of incorporating functional groups that are chemically reactive and which are absent in biological systems. There are a number of techniques that have been devised for UAAs introduction. These techniques range from purely synthetic to the utilization of endogenous translational mechanism\(^20\)-\(^22\). A decision will then be made on which approach to adopt in order to incorporate the right unnatural amino acid for a specific use.
The advent of orthogonal aminoacyl-tRNA synthetase/tRNA pairs resulted in a number of successes in making biconjugates and in various applications\textsuperscript{22}. This approach has enabled incorporation of a number of chemical utilities into the proteins like amino acids with fluoro\textsuperscript{23,24}, bromo\textsuperscript{25}, chloro\textsuperscript{25}, iodo\textsuperscript{26}, azido\textsuperscript{27,28}, alkyne\textsuperscript{28}, and nitro\textsuperscript{21} side chains. The approach has also broadened the acquisition of unnatural amino acids, which can be introduced into the proteins translationally.

This study illustrates that each of the UAAs are incorporated into proteins employing the orthogonal pair technique that was reported by P.SCHULTZ and others\textsuperscript{29}. This technique is very useful since it reveals the capabilities of the site-specific method to generate larger quantities of a recombinant protein having a number of functionalities and physical characteristics, which suit various functions. As a matter of fact, the application of UAAs in the process of bioconjugation has grown to become a field, which is capable of incorporating functional groups that are chemically reactive and which are absent in biological systems.

In particular, there are a number of examples of the applications of UAAs in the generation of well-defined biconjugates, including the reaction of nanoparticles with calmodulin-containing acids onto carbon nanoparticles and nanotubes through the Staudinger-Bertozzi ligation\textsuperscript{30,31}. Latest contributions from Ben. G. Davis and team report a highly efficient and robust Suzuki-Miyaura\textsuperscript{16,32,33} as well as Sonogashira\textsuperscript{17} cross-coupling on the alkyne group or halide bearing protein substrates under ambient environment both on the cell surface\textsuperscript{33} and in vitro\textsuperscript{16,34}. Others addressed the phase-problem that complicates crystal structure determination through incorporaton of phenylalanine analogs heavy halogen atoms, e.g. para-bromophenylalanine and para-iodophenylalanine\textsuperscript{35}. The other use of para-bromophenylalanine and para-iodophenylalanine has been elaborated at the end of this section i.e. the SUZUKI coupling reaction by way of palladium-catalyzed coupling\textsuperscript{35,36}.

The compilation of these investigations shows the use of site-specific biconjugates and the stabilizing advantage of solid support. Other investigations have employed similar techniques in the labelling of proteins using molecular probes or in the generation of new therapeutic bioconjugates\textsuperscript{4,37}. With the potential of this technology, our objective was to
expand the UAA bioconjugate utility via the modification of the reaction UAAs in order to optimize the reaction parameters. Initial investigations aimed at further utilizing UAAs in order to optimize bioconjugation reactions via a number of different biorthogonal reactions. Bioconjugation by click reaction is the other application that utilizes the site-specific incorporation technique. The most effective click chemistry reaction is the azide-alkyne cycloaddition because of its effectiveness and biocompatibility (Figure 4.1)\textsuperscript{38}. The research outcomes from several areas such as organic chemistry\textsuperscript{39}, bioconjugation\textsuperscript{40}, drug discovery\textsuperscript{41} and polymer chemistry\textsuperscript{42} have been prosperous after the findings on azide-alkyne ligation. The azide moiety is termed as “bioorthogonal” because it is absent in biological systems and does not react with naturally occurring biomolecules. Generally, its functionality only experiences ligation with a very small group of chemical partners. These aspects are very beneficial in the process of incorporating the unnatural amino acids mainly because the two common amino acids used for chemical bioconjugations are lysine and cysteine and they have their own disadvantages such as their adequacy in biological systems and the changeability of disulfide bond\textsuperscript{43}. Other bioconjugation reactions such as isothio-cyanateamine, thiol-maleimide, and amine-carboxylic acid couplings\textsuperscript{44-46} are not compatible with \textit{in vivo} labeling since the nucleophiles in the DNA, RNA and protein molecules are in competition. Still, such reactions cannot be used \textit{in vivo} since they are not biorthogonal, and they can also be reversed at physiological pH\textsuperscript{46}.

Therefore, click chemistry via azide-alkyne cycloaddition is an optimal bioconjugation reaction that has been facilitated by UAA incorporation. This chapter focuses mainly on the click reaction and the application of this reaction. Because of the unmatched selectivity and modularity of bioconjugation via click chemistry make it a convenient, and now invaluable, tool for a multitude of applications. While these studies did not use UAAs for protein engineering purposes, they were significant because of their contribution on broadening the field of artificial metalloenzymes.

\textbf{Figure 4.1.} The Azide–alkyne cycloaddition

\[
\begin{array}{c}
\text{Azide} \\
R_1 \cdots \text{N} \cdots \text{N} \cdots R_1
\end{array} + \begin{array}{c}
\text{Alkyne} \\
\cdots \text{Cu(I)} \cdots \text{N} \cdots \text{N} \cdots \text{N}
\end{array} \rightarrow \begin{array}{c}
\text{Triazole} \\
R_2 \cdots \text{N} \cdots \text{N} \cdots R_2
\end{array}
\]
4.3 Result and discussion

4.3.1 Identifying Mutation Sites to Create Artificial Metalloenzymes

The position for the site-specific incorporation of a specific UAA has a huge influence on the catalytic selectivity of the enzyme. Therefore, designing mutation sites to create a transition-metal center in the chiral protein environment rely on information about the target catalytic reaction that determines the catalyst reactivity and selectivity.

In this study we used the expanded genetic code methodology introduced by Peter Schultz, which has been reported by several research groups for the establishment of metalloproteins or artificial metalloenzymes containing non-canonical amino acids with metal binding moiety or functionalization capabilities\(^47\). Typically, a TAG (amber stop codon) is reassigned to code for the incorporation of the UAA, because it is the least used stop codon in the \textit{E. coli} genome. This mutation site will later serve as site-specific anchoring points for the active catalytic center (Figure 4.2). Computational studies and homology modeling successfully highlighted several positions in mTFP*. As a result, we evaluated potential sites of modification on the surface of our tailored mTFP* (Asn128, Tyr188, Tyr200 and Tyr204), which are especially suitable for incorporation without any disruption or misfolding of the protein (as described in previously in Chapter 3). Two of them represent exposed positions (Asn128, Tyr188), whereas the other two are in a cavity (Tyr200, Tyr204).

![Figure 4.2](image.png)

**Figure 4.2.** Structure analysis was pursued with Pymol by Schrödinger INC.
4.3.2 Site-Specific Incorporation of Various P-Substituted Phenylalanine Derivatives Using aaRS/tRNA pair.

A functional aaRS/tRNA pair is required that acts independently of the endogenous aminoacylation machinery of the cell. Orthogonality must be maintained, so that the suppressor tRNA is not a substrate for any endogenous aaRS, and the orthogonal aaRS does not aminoacylate any endogenous tRNAs. The active site of the aaRS must activate and aminoacylate its cognate tRNA with only the unnatural amino acid of interest and no endogenous host amino acids. A suitable aaRS/tRNA pair is typically selected by directed evolution using a double-selection process\textsuperscript{48-50}. An aaRS/tRNA pair optimized for site specific incorporation of the respective amino acid is best introduced into the cell using a pEVOL vector, which was developed by Schultz and coworkers. The functional aaRS/tRNA pair is orthogonal to the E. coli translational machinery and suitable for effective in vivo incorporation of the respective UAA into the host proteins\textsuperscript{29}.(Figure 4.3)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.3}
\caption{Incorporation of para-Azido-(L)-phenylalanine into mTFP\textsuperscript{*}.}
\end{figure}
Tests for of the polyspecific aaRS were successful and four different UAAs were incorporated into mTFP*, albeit with differential expression levels (Table 4.1). Using the respective pEVOL plasmids (pEVOL pAzF, pEVOL pBF pEVOL pIF) we successfully incorporated the four different UAAs para-(L)Azidophenylalanine (pAzF), para-(L)-ethenylphenylalanine (pEnyF), para-iodophenylalanine (pIF) and para-boronophenylalanine (pBF). pEynF incorporation was been tested using three pEvol plasmids (pCNF, pPIF and pAzF pEvol). As expected the pCNF synthetase had also been evolved to recognize pEynF, incorporated this amino acid with much better yields than the other two. This is possible due to the structural similarity of the cyano group and alkyne group with respect to in size and polarizability (Figure 4.4).

![Figure 4.4](image_url)

**Figure 4.4.** Fluorescent intensity screening of the expression culture for selecting the best pEVOL plasmid used for incorporation of pEnyF into mTFP*.

To test if the position and the type of UAA has any influence on the mTFP* expression yield, complexation to the metal center and at the end the enantioselectivity of the catalytic reactions product, we first, incorporated a variety of conjugation UAA; para-(L)-azidophenylalanine (pAzF), para-(L)-ethenylphenylalanine (pEnyF), para-iodophenylalanine (pIF) and para-boronophenylalanine (pBF). The result of the UAA
incorporation was visualized both by fluorescence intensity (Figure 4.6) and SDS-PAGE analysis for each protein (see Figure S4.12). LC/MS analysis confirmed the incorporation of the respective UAA into mTFP* (Table 4.1). We could successfully show that the most favorable position for pIF incorporation is 188 with yield reaching up to 350mg/L, which is the highest among all UAA incorporation (Table 4.2 entry 10), whereas pAzF prefers position 200 with yield up to 285mg/L (Table 4.2 entry 3). The position 204 shows the highest yield for pEynF and reaches up to 200mg/L (Table 4.2 entry 8). This confirmed the efficient and robust expression of FPs with site-specific incorporation of UAAs by the orthogonal pair approach. However, the incorporation of pBF shows a very low expression yield, which indicates a bad recognition of pBF by the translation system (Table 4.2 entry 13 and 14).

![Figure 4.5.](image)

*Figure 4.5.* Proteins structure. mTFP* (PDB:4Q9W), representation by for the successful incorporation of UAAs (pAzF, pEynF, pIF, and pBF) respectively at 188 position. Illustration and prediction by the PyMOL Molecular Graphics System, Version1.8.2.3 Schrödinger, LLC.

<table>
<thead>
<tr>
<th>mTFP*wt.</th>
<th>mTFP*-UAA incorporated</th>
<th>Calculated</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>128N</td>
<td>128N</td>
</tr>
<tr>
<td>25154.4</td>
<td></td>
<td>25228.5</td>
<td>25229.6</td>
</tr>
<tr>
<td>pAzF</td>
<td></td>
<td>25212.4</td>
<td>25215.7</td>
</tr>
<tr>
<td>pEynF</td>
<td></td>
<td>25313.3</td>
<td>25316.2</td>
</tr>
<tr>
<td>pIF</td>
<td></td>
<td>25231.3</td>
<td>25233.9</td>
</tr>
<tr>
<td>pBF</td>
<td></td>
<td>25182.2</td>
<td>25184.9</td>
</tr>
<tr>
<td>Calculated</td>
<td>188, 200, 204 Y</td>
<td>25179.4</td>
<td>25267.5</td>
</tr>
<tr>
<td>Observed</td>
<td>188, 200, 204 Y</td>
<td>25164.4</td>
<td>25180.3</td>
</tr>
<tr>
<td>M[Da]</td>
<td>188, 200, 204 Y</td>
<td>25264.2</td>
<td>25167.8</td>
</tr>
<tr>
<td>M[Da]</td>
<td>188, 200, 204 Y</td>
<td>25231.3</td>
<td>25233.9</td>
</tr>
</tbody>
</table>

Table 4.1. Expected MS for incorporation of different UAAs into mTFP* at different mutations sites.
Table 4.2. Expression yield for all protein positions with different UAAs.

<table>
<thead>
<tr>
<th>Entry</th>
<th>UAA</th>
<th>pEVOL plasmid</th>
<th>Protein-Mutation sites</th>
<th>Yield mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>pAZF</td>
<td>pEVO-pAZF</td>
<td>mTFP*128</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>mTFP*188</td>
<td>214</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>mTFP*200</td>
<td>285</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>mTFP*204</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>pEynF</td>
<td>pEVOL-pCNF</td>
<td>mTFP*128</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>mTFP*188</td>
<td>110</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>mTFP*200</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>mTFP*204</td>
<td>200</td>
</tr>
<tr>
<td>9</td>
<td>pIF</td>
<td>pEVOL-pIF</td>
<td>mTFP*128</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>mTFP*188</td>
<td>350</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>mTFP*200</td>
<td>70</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>mTFP*204</td>
<td>150</td>
</tr>
<tr>
<td>13</td>
<td>pBF</td>
<td>pEVOL-pBF</td>
<td>mTFP*128</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>mTFP*188</td>
<td>22</td>
</tr>
</tbody>
</table>
Figure 4.6. Incorporation screening of different UAs (pAzF, pEynF, pIF, and pBF) incorporated into mTFP* at different mutations sites (128, 188, 200, and 204)

As shown in scheme 4.1, the exposed position will allow and simplify future functionalization on the protein, e.g. using copper-catalyzed azide-alkyne cycloaddition (which will be discussed in chapter 5) and palladium-catalyzed cross coupling reactions (see below). These kinds of reactions may find applications in FRET measurement, protein labeling, protein detection, protein bioconjugation, and catalysis application.

Scheme 4.1. Overview of incorporation catalytically active functionalized UAA into a Fluorescent protein to generate an active biocatalyst. Relevant motives will be coupled to a host protein through the Catalytic Reaction.

For pIF incorporation, a crystal structure resolved clearly the unnatural iodine moiety embedded in the protein scaffold (Figure 4.7). Besides the visual evidence of a successful incorporation by a full-length protein with maturated chromophore, LC/MS determination was conducted to prove correct incorporation (Figure S4.13).
A Direct Borylation Strategy Through Selective Modification of Biomolecules

Boronic acids are important chemical building blocks for synthetic applications such as SUZUKI cross couplings\textsuperscript{51,52}. These functionalities provide ideal handles for chemical posttranslational modification yet, they are so far difficult to introduce into proteins and typically only available in for synthetic ligands. For couplings, the haloarene counterpart can be incorporated into a peptide, sugar, protein or DNA. SUZUKI cross couplings on proteins are known and presented in literature to introduce a variety of biological markers, sugars and dyes or new functionalities\textsuperscript{16,34,53} (Scheme 4.2). Commonly those cross couplings are transition palladium catalyzed (Scheme 4.3)
Scheme 4.2. SUZUKI cross coupling to introduce new functionalities on proteins.

Scheme 4.3. SUZUKI coupling reaction on mTFP*, which has $p$I$F$ and $p$BF incorporated on a single position.
To broaden the range of applications it would be beneficial to introduce the boronic acid functionality into the biomolecule. However, as described above the introduction of boronic acid functionalities through UAAs such as $p$BF into the biological systems suffers from very low yields (about 50-fold reduction in protein yield compared to WT). Further, $p$BF suffers from cellular degradation reactions. Oxidation leads to (L)-tyrosine, whereas reduction leads to (L)-phenylalanine$^{54}$ (Scheme 4.4).

Scheme 4.4. Cellular degradation limits incorporation of para-borono phenylalanine.

In contrast to these limitations, in vivo incorporation of the very stable para-iodophenylalanine ($p$IF) gives excellent incorporation yields of up to 350 mg/L and is therefore well suited for various applications including phasing for crystal structure determination,$^{55}$ posttranslational modifications or as simple precursor for other unnatural amino acids. Hence, the question arouse, if the iodine functionality can efficiently be transformed on the protein into a boronic acid by MIYURA borylation (Scheme 4.5).

To confirm the applicability of the selective MIYURA borylation on the mTFP* we first incorporated $p$IF into mTFP* as described above (surface exposed sites: Asn128 $p$IF, Tyr188$p$IF in cavity sites: Tyr200 $p$IF and Tyr204 $p$IF).
Scheme 4.5. Borylation reaction on mTFP*, which has pIF incorporated on a single position.
4.3.2.1 MIYAURA Borylation

Investigation of the palladium-catalyzed posttranslational modification, started form conditions of an aqueous protocol, which had been developed in our lab recently. (See material and methods). Different parameters were investigated to optimize the borylation yields, including the concentration of protein, borylation agents such as bisoronic acid (BBA) and catalyst loading. As recommended by the protocol conducted in our lab, an optimum of pH occurs at around neutral pH, not higher then 8, with a NOAc buffer giving the best results. pH higher than 8 leading to decomposition of BBA under hydrogen evolution. As good catalyst, which allows acceleration of the reaction under very small catalyst loadings under aqueous conditions, we used the palladacyle catalyst (cat-1), which was developed in our group51 (Figure 4.8).

![Figure 4.8. Paladacyle catalyst (cat-1).](image)

As expected mTFP1-183pIF with HisTag was not borylated, since the HisTag immediately complex the Pd-catalyst and resulting in catalyst deactivation (Table 4.3 entry 1).

Initial tests at a protein concentration of 4.8 mg/mL in 100 mM acetate buffer achieved full conversion (Table 4.3 entry 2). LC/MS determination does not indicate any remaining mTFP*pIF. If the pH is adjusted to pH 10 or the reactions were conducted under oxidative conditions, degradation of the protein takes place resulting in a complete absence of protein MS signals. Degradation depends very much on the stirring or shaking techniques, buffer pH and concentration as well as the final work up of the protein. Results were confirmed by LC/MS determination and showed signals at given MS ranges (Table 4.4, Figure 4.9).
Table 4.3. Different borylation conditions were investigated for Below reaction.

![Diagram of molecules](image)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein</th>
<th>Protein conc.</th>
<th>Acetate buffer</th>
<th>conversion</th>
<th>Estimated yielda</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mTFP*1-183pIF with HisTag</td>
<td>4.7 mg/mL</td>
<td>1 M</td>
<td>no conversion</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>mTFP*-*128pIF</td>
<td>4.8 mg/mL</td>
<td>1 M</td>
<td>full</td>
<td>14-26%</td>
</tr>
<tr>
<td>3</td>
<td>mTFP*.*-204pIF</td>
<td>2.0 mg/mL</td>
<td>100 mM</td>
<td>full</td>
<td>24% Protein degradation</td>
</tr>
<tr>
<td>4</td>
<td>mTFP*-*204pIF</td>
<td>1.0 mg/mL</td>
<td>100 mM</td>
<td>full</td>
<td>11-20% Protein degradation</td>
</tr>
<tr>
<td>5</td>
<td>mTFP*-*204pIF</td>
<td>0.5 mg/mL</td>
<td>100 mM</td>
<td>approx. 50% conversion</td>
<td>22% Protein degradation</td>
</tr>
</tbody>
</table>

Estimated from relative LC-MS signal intensities (relative abundance).
Table 4.4. Expected MS for different mutants of mTFP*.

<table>
<thead>
<tr>
<th>UAA position</th>
<th>M[Da] $p$IF</th>
<th>M[Da] borylation</th>
<th>M[Da] Suzuki*</th>
</tr>
</thead>
<tbody>
<tr>
<td>mTFP*wt</td>
<td>25154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mTFP*128</td>
<td>25313</td>
<td>25305</td>
<td>25305</td>
</tr>
<tr>
<td>mTFP*188, 200, 204</td>
<td>25264</td>
<td>25256</td>
<td>25256</td>
</tr>
</tbody>
</table>

Figure 4.9. ESI-TOF analysis of MIYURA borylation, of mTFP*$p$IF to mTFP*$p$BF.

4.3.2.2 Subsequent SUZUKI Cross Coupling

Subsequent SUZUKI cross coupling would allow introduction of further functionalities. As described previously, subsequent cross coupling can be initiated by a change of base from sodium acetate to sodium carbonate, which is accompanied by a rise of pH. In the case of the protein this can be conducted by simple dialysis against a new carbonate-containing buffer. Addition of new catalyst must be considered.

A variety of different halide containing building blocks can be coupled to mTFP*$p$BF in that way (Figure 4.10).
In our primary attempt, different biological markers, such as sugars or dyes are considered. More challenging are transition metal chelating ligands. Different positions of the initial incorporated $p$IF will give new insights into possible secondary interactions and therefore might have different influences on enantioselectivities in the executed catalysis. To broaden this modification’s application, borylation can be conducted on different proteins. Proteins of medical interests can be investigated to determine structural change, for instant by introducing dyes acting as FRET partners.

![Figure 4.10. Different structures, which could be cross coupled to proteins.](image)

### 4.4 Conclusion

The borylation reaction has been used to convert the iodo group of $p$IF on the modified protein into a boronic acid functionality. For that purpose, a mild aqueous protocol for the MIYURA borylation of aryl iodides, bromides and chlorides using the water-soluble boron source tetrahydroxydiboron (BBA) has been developed in our lab. This protocol for MIYURA borylation was successfully transferred to the protein and bypasses the problematic direct incorporation of borylated building blocks. This simple strategy allows now chemical modification of different biomolecules.
4.5 Material and Method

4.5.1 Cloning, Mutagenesis, Expression with UAA and Purification

4.5.1.1 Cloning

mTFP* fluorescent protein was cloned into the pET303 vector. mTFP* gene was digested by BAMHI and XhoI restriction enzymes and fused to a N-terminal SUMO Tag to generate the construct (pET303_SUMO_mTFP*). Mutant variants of mTFP*TAG (mTFP*N128 TAG, mTFP*Y188 TAG, mTFP*Y200 TAG and mTFP*Y204 TAG) were generated via QuickChange® Site-Directed Mutagenesis KitII, for UAA site incorporation. The resulting plasmids were Co-transformed with pEvol-aaRS plasmids (from Peter G. Schultz) (pEVOL pAzF, pEVOL pBF pEVOL pIF and pEVOL pCNF).

4.5.1.2 Expression and Purification

The bacterial culture was grown in LB-containing 180mg/ml ampicillin and 25mg/L Chloramphenicol at 37 °C. Induction by 1mM IPTG and 0.2% of L-arabinose followed by 1mM of corresponding unnatural amino acid was applied when the OD reached 0.6-0.7. Cells were left to express for 48 hours at 20 °C. Cultures were then centrifuged at 4500 rpm for 30 minutes and cooled to 4 °C. The purification perpetration and procedure were applied exactly as the purification of mTFP* in the last chapter.

Purity was determined by SDS-PAGE gel electrophoresis at 12.5% acrylamide. All incorporation confirmation was done using LC/MS to detect the mass differences before (wt protein) and after the incorporation (modified protein).

4.6 Synthesis of UAA

All synthesized UAA are listed in (Table 4.5)

4.6.1.1 Synthesis para-Iodophenylalanine (pIF)

To a solution of (L)-Phenylalanine (42.6 g, 258 mmol) in sulphuric acid (31 mL) and acetic acid (235 mL) iodine (26.2 g, 103 mmol) and sodium iodate (10.2 g, 52 mmol) is added. The mixture is heated to 70 °C and stirred vigorously over night (16-20 h). Sodium periodate is added (2x 1 g). After 4 h the reaction is completed, indicated by the solution becoming orange. Acetic acid was removed under reduced pressure and the
crude mixture was diluted with water (80 mL) and washed with Et2O (250 mL) and DCM (250 mL). The aqueous layer is basified to pH 5.5 with NaOH 10%. The precipitate formed over night at 4 °C, is filtered under vacuum and washed with water (250 mL) and ethanol (100 mL) to afford 48.8 g (65 %) of 4-Iodophenylalanine as a white solid.

4.6.1.2 Synthesis para-(L)Azidophenylalanine (pAzF)
4-Amino-(L)-phenylalanine (0.828 g, 4 mmol) are dissolved in half concentrated hydrochloric acid (32 mL) and Sodium nitrite (0.304 g, 4.4 mmol) dissolved in water (8 mL) is added dropwise at 0 °C. After 10 minutes half concentrated hydrochloric acid (38 mL) and Sodium azide (0.408, 4.2 mmol) in water (8 mL) is added dropwise over 15 minutes at 0 °C. The solution keeps stirring at 0 °C for 15 minutes and then 4.5 -7 h at room temperature until TLC shows full conversion. After evaporation in vacuo the residue is resolved in water, adjusted to pH = 5 with NaOH. The aqueous phase is washed with Diethylether (1x), Ethylacetate (1x) and DCM (1x). The water is evaporated, the solid residue is redissolved in MeOH (HPLC) and all salts are filtered off. Evaporating of the organic solvents give 4-Azido-(L)-phenylalanine (0.458 g, 56%) as a white solid.

4.6.1.3 Synthesis para-(L)-Ethenylphenylalanine (pEnyF)
To (S)-2-((tert-butoxycarbonyl)amino)-3-(4-ethynylphenyl)propanoic acid (82 mg, 0.283 mmol) HCl in Dioxane (4.0 M, 1 mL) is added and stirred at room temperature until TLC shows complete conversion (app. 2 h). The solids are filtered off and washed with Et2O to give (S)-2-Amino-3-(4-ethynylphenyl)propanoic acid (27 mg, 50%) as a white solid.
### Table 4.5. summary of structure and molecular weight of the UAAs used.

<table>
<thead>
<tr>
<th>UAAs</th>
<th>pIF</th>
<th>pBF</th>
<th>pAzF</th>
<th>pEynF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structure</strong></td>
<td><img src="image1" alt="Structure of pIF" /></td>
<td><img src="image2" alt="Structure of pBF" /></td>
<td><img src="image3" alt="Structure of pAzF" /></td>
<td><img src="image4" alt="Structure of pEynF" /></td>
</tr>
<tr>
<td><strong>M.wt</strong></td>
<td>291</td>
<td>209</td>
<td>206.2</td>
<td>189</td>
</tr>
</tbody>
</table>
4.6.2 General Procedure for the Pd Catalyzed MIYURA Borylation Reaction on Protein, Followed by SUZUKI Cross Coupling

To achieve small catalyst loadings, the Paladacycle cat (cat-01) (1 mol-% Pd, 0.5 mol-% cat-1) is added to the reaction vial (2 mL Eppendorf tube) as a 1M stock solution in DCM. Prior the reaction the solvent is evaporated. Tetrahydroxydiboron BBA 5 eq. is added to the vial, followed by the protein in degassed acetate buffer 100 mM, pH=10 (NaOAc/NaOH/AcOH). The solution vigorously shakes for 24 h at 40 °C in Eppendorf Thermomixer comfort® at 1400rpm or in a closed GC/MS vial with atmosphere exchanged to Argon and a small stirring bar stirring at 1500 rpm. After 24 h the catalyst is removed by filtration through a centrifugal NANOSEP 30K OMEGA from Life Sciences. Samples for LC-MS analysis are dialyzed against carbonate buffer in Slide ALyzer ® MINI Dialysis Units 2.000MWCO, concentrated and diluted prior the measurement in water/acetonitril 5%. For subsequent SUZUKI cross coupling the reaction mixture is dialyzed against carbonate buffer 100 mM, pH=11.5 (Na2CO3) (for 20 h using Slide-A-Lyzer® MINI Dialysis Units 2.000MWCO). The protein solution is transferred into a new reaction vial containing fresh catalyst (cat-1) (1 mol% Pd, 0.5 mol-% cat-1) and the compound, which is supposed to be cross coupled. The reaction mixture stirs for 24 h at 40 °C in Eppendorf Thermomixer comfort® at 1400 rpm. After 24 h the catalyst is removed by filtration through a centrifugal NANOSEP 30K OMEGA from Life Sciences and dialysed for LC-MS (Figure 4.11). 3-mercaptopropionic acid (3-MPrAc) as Pd scavenger:140 Prior to MS analysis a scavenger concentration of 4.4 mmol mL-1 (3 equiv. w.r.t. Pd) was added as a dilute solution (5 mL mL-1) in water to the sample to ensure accurate titration.

Figure 4.11. Borylation was performed in GC/MS vials, separation of the catalyst was done by centrifugation through NANOSEP 30K OMEGA.
4.7 Reference:


4.8 Supplementary Material

Incorporation Sites and respective Primers

1. mTFP*6hism-N128  a379t_c381g
2. mTFP*6hism-Y188  t561g
3. mTFP*6hism-Y200  t597g
4. mTFP*6hism-Y204  c609g

Table S4.6. Oligonucleotides for mTFP* mutation sites for UAA incorporation.

<table>
<thead>
<tr>
<th>Protein-Mutation position</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>mTFP*128 f</td>
<td>5'-GAACTTCCCCCCCCTAGGGGCCGCTGATGC-3'</td>
</tr>
<tr>
<td>mTFP*128 r</td>
<td>5'-GCATCACGGGGCCCTAGGGGGGAAGTTC-3'</td>
</tr>
<tr>
<td>mTFP*188 f</td>
<td>5'-GTGAAGCTGCCCGACTAGCCTTGTGGAGCCA-3'</td>
</tr>
<tr>
<td>mTFP*188 r</td>
<td>5'-TGGTCCACAAAGTGTGCTGAGGCGAGCCCTC-3'</td>
</tr>
<tr>
<td>mTFP*200 f</td>
<td>5'-GACATGGAGATTCTCTAAGGACTACAAAGTGGA-3'</td>
</tr>
<tr>
<td>mTFP*200 r</td>
<td>5'-TGGTCCACAAAGTGTGCTGAGGCGAGCCCTC-3'</td>
</tr>
<tr>
<td>mTFP*204 f</td>
<td>5'-TCCTGAACATGGAGAAGGACTACAAAGTGGA-3'</td>
</tr>
<tr>
<td>mTFP*204 r</td>
<td>5'-GGTCACCTTGTCTAGTCATGTGCAG-3'</td>
</tr>
</tbody>
</table>

mTFP*128

GGCGTAATCAAGCCGCCACATGAAGATCAAGCTGAAGATGGAGAGGAGAACGTGAATGGTCTGTCCTCCTCCTAGGGGCCGCTGATGCAGAAGAAGACCACCCGGCTGGGACGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGAACGTCAAGCACAAGCTGCTGAAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGG

mTFP*188

GGCGTAATCAAGCCGCCACATGAAGATCAAGCTGAAGATGGAGAGGAGAACGTGAATGGTCTGTCCTCCTAGGGGCCGCTGATGCAGAAGAAGACCACCCGGCTGGGACGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGAACGTCAAGCACAAGCTGCTGAAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGG

mTFP*128

GGCGTAATCAAGCCGCCACATGAAGATCAAGCTGAAGATGGAGAGGAGAACGTGAATGGTCTGTCCTCCTAGGGGCCGCTGATGCAGAAGAAGACCACCCGGCTGGGACGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGAACGTCAAGCACAAGCTGCTGAAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGG
Figure S4.12. SDS-gel analysis for pure proteins 1= mTFP*200pAzF, 2=mTFP*204 pEynF and 3=mTFP*188PIF.
**Figure 4.13.** ESI-TOF analysis of the incorporation of unnatural amino acids into the TAG position. A) mTFP*Y200 pAzF, B) mTFP*Y204 pEynF, C) mTFP*188pIF, and D) mTFP*188 pBF.
5 Chapter 5  A Site-Selective Anchoring Strategies Towards Fluorescent Artificial Metalloprotein

5.1 Abstract

Bioconjugation utilizing Unnatural Amino Acids (UAAs) and click chemistry is a growing field of research at the interface chemistry and biology. Formation of catalytically active metal center through introduction of metal ions into a protein can noticeably expand the range of biocatalysis’ functionalities and applications. The control of the strics and electronics of metal-binding and accordingly reaction selectivity such as enantioselectivity is particularly challenging. In this chapter, an unnatural cofactor is introduced into the host protein mTFP* through conjugation to UAA. This cofactor provides a chelateing moiety, which is supposed to bind transition metal ions well. To evaluate the importance of the rigidity of the metal binding site, single as well as dual anchouring of the cofactor on the protein scaffold is investigated. The UAAs pAzF provides the anchoring points, while corresponding water–soluble motives are conjugated through Copper Catalyzed Alkyne-Azide Cycloaddition (CuAAC) click chemistry. The small library of single site UAA containing model proteins described in chapter 4 is used to study CuAAC-bioconjugation. Conversion with pyridyl-alkyne allows the on-protein formation of pincer-type ligands. In the artificial metalloenzymes, which are formed after metal coordination, the host protein serves as ligand controlling selectivity, while the metal site determines the reactivity. The generated artificial metalloenzymes are tested in the asymmetric Diels-Alder cyclisation and Pd-catalyzed Tsuji-Trost cross-coupling reactions.

The double anchouring strategy requires two UAA sites and delivers significantly increased rigidity, which should translate into enhanced selectivity for the newly introduced active metal center in the protein’s chiral environment. On-protein synthesis of pincer motives of from clickable ligands lead to creation of a novel artificial metalloenzymes that possess a highly selective robust metal center that is facing towards the protein surface exposed to the target reactions. The two triazole rings formed during the click reaction will contribute to the metal binding site. Click reaction CuAAC will
offer the functionalization with pincer-type complexes. (Scheme 5.1). The catalytic performance of the resulting artificial metalloenzymes is significantly improved.

Here we report a novel site-selective dual UAAs incorporation strategy to introduce a di-pyridyl-alkyne (which can coordinate metal) into mTFP* protein with bioconjugation yield reaching up to 99%. The enantioselective excess increases from 0% for single site attachment to 32% for a dual conjugation in the Diels-Alder cyclisation and from 0% to 21% for Tuji-Trost reaction. The dual anchoring method proves to have the advantage of limiting the conformational freedom of the metal within the vicinity of the protein surface and cavity.

![Scheme 5.1. Water-soluble click-pincer ligand, which will be attached to the mTFP* by double sites UAA to create a chiral pocket on the mTFP* surface.](image-url)
5.2 Introduction

The first time in 1978 when Wilson and Whitesides introduced the concept of Artificial Metalloenzymes (ArMs) in their pioneering work on asymmetric hydrogenation catalysts. Since then, construction of artificial metalloenzymes became a new research area and one of the most important subjects in bioinorganic chemistry. Artificial metalloenzymes have particularly attracted the attention of biologists as well as chemists. Therefore, many researchers from biology and chemistry are interested to work on artificial metalloenzymes, mainly because it is an interface area between chemistry and biology with a lot of potential for chemistry applications. The main target are artificial metalloenzymes that can catalyze chemical transformations with high selectivity and reactivity under mild reaction conditions in the universal green solvent water. However, natural protein scaffolds have not been evolved sufficiently to tightly coordinate the metal at the artificial metal site in a single conformation and one of the major challenges for biologists is to make artificial metalloenzymes which should have the ability to control the chemo and enantioselectivity of chemical reactions. Therefore, introducing nonnative metal ions or metal-containing prosthetic groups into a protein is a branch of metalloprotein design which has considerable impact, as it can dramatically expand the repertoire of protein functionalities and, thus, their range of applications in the uses of chemistry. Moreover, with the recent advance in biotechnology and available techniques, a protein can be designed in several ways to synthesize artificial metalloenzymes. Artificial metalloenzyme synthesis can be achieved by modification of natural proteins using various approaches such as the introduction of metal binding sites, chemical modification of prosthetic groups, design of substrate binding cavities, and covalent attachment of metal cofactors. Among the several approaches to modify the protein, one of the most promising approaches is the incorporation of unnatural amino acid (UAA) into the protein scaffold. This way, a new chemical functional group can be introduced at a selected specific protein site. One of the most simple and widely used approaches for doing modification on protein is introducing azide or alkyne functional groups using incorporated unnatural amino acid. Ultimately, this incorporated azide or alkyne site can selectively transformend through Copper Catalyzed Azide Alkyne Cycloaddition (CuAAC) reactions. There are many reports on the application of CuAAC
reaction for bioconjugation, but to the best of our knowledge use of CuAAC for creation of artificial chiral pocket/sites on a protein i.e. for the synthesis of artificial metal binding sites has not been studied.

Herein, we described a new strategy for the preparation of a library of artificial metalloenzymes (mTFP*) by using the widely accepted copper catalyzed Azide-Alkyne Cycloaddition reaction (CuAAc), a click bioconjugation protocol. First, we introduce an unnatural amino acid (UAA) containing an azide (Az) functional group selectively into the selected site of the mTFP* host protein. In the next step, corresponding ethynylnyridine (Figure 5.1) was clicked using CuAAC reaction, resulting in formation of triazole pyridine or bis-triazole pyridine (BTP) formation into the protein. In the next step, a transition metal such as palladium or copper is coordinated to the selective site of bis-triazole pyridine moiety resulting in a pincer type metal complex on the protein. The reason for selecting this approach the fact that pincer complexes are known to catalyze various coupling reactions such as hydrogenation reactions and organic transformations15,16. However, most of the reported protocols required harsh reaction conditions, longer reaction time and there are only a few reports in the application of asymmetric catalysis. Therefore, with our new approach, a variety of protein based BTP ligands or pincer type artificial metalloenzyme (mTFP*) can be synthesized and chemical transformations with high selectivity and reactivity under mild conditions can be achieved. Moreover, the single-point incorporation approach makes it possible to attach metal cofactors and different chemical motifs at a defined location, but it does not necessarily fix the conformation of the cofactor inside the protein. Therefore, multiple sites of incorporation of UAAs, which can coordinate the co-factor will improve the enantioselective of the resulting catalytically active organometallic hybrid due to the defined orientation of the transition metal catalyst on the chiral surface of the proteins. In this way, a potent metal center of interest can be introduced targeting a particular reaction. For example, a single covalent attachment of a 4-(2-methanesulphonyl-thioethoxy)salicylidene-1,2-ethanediarnino-manganese(iii) (Mnsalen) co-factor to apomyoglobin through a single Cys-selective linker resulted in only 12% enantiomeric excess towards sulphoxidation of thioanisole17 suggesting that the Mn-salen conformation may be too flexible inside myoglobin to facilitate high enantioselectivity. The addition of
a second linker to the cofactor, together with selective cavity positioning led to a significant increase in catalytic efficiency (more than a sevenfold increase) and enantiomeric excess (increased from 12% to 88%)\textsuperscript{17,18}. Till now, the multiple site incorporation of unnatural amino acid has not been used for creating a chiral environment on the protein cavity. It has been used mainly for site selective cell surface and proteins manipulation\textsuperscript{19}, glycosylated\textsuperscript{20,21}, labeling\textsuperscript{22}, and PEGylation\textsuperscript{23}. In this study, we demonstrate the effectiveness of the dual anchoring approach to markedly improve the enantioselectivity of a fluorescent protein with a minimal structural modification. We have explored the dual UAA incorporation approach, which could be effective for designing a chiral pocket in the protein by applying different clickable metal chelating centers using copper-catalyzed azide-alkyne cycloaddition (CuAAC). Application of the synthetic metal containing artificial metalloenzyme was tested in asymmetric Diels-Alder as well as Tsuji-Trost reactions.

### 5.3 Result and Discussion

Based on the previous results on protein screening studies (mTFP* stability chapter 3), mTFP* was chosen as a chiral scaffold for synthesis of ArMS. The initial biochemical characterization revealed very distinctive properties of mTFP*, thus allowing a good visualization and estimation of its capabilities to act as scaffold to create ArMs. The main purpose for selecting mTFP* as host protein of choice is its intrinsic fluorescent, non-toxic nature, high expression yields (up to 1 g/L), highly solubility, temperature and pH stability and the fact that mTFP* can even be dissolved in a variety of organic solvents, which might be useful when it comes to catalytic reactions and the need for an organic co-solvent. The most obvious advantage is the quantification of protein amount and successful UAA incorporation through fluorescence measurements, since after incorporation of UAA only a correctly folded full length folded protein chain will show fluorescence. So it’s easy to judge if UAA incorporation happened or not at the preliminary stage by visualization. The following steps are required to create ArMs from mTFP* through UAA/bis-ethynylpyridine conjugation:
1. Site selective incorporation of UAA and synthesis of corresponding Pincer type artificial metalloenzymes using CuAAC approach.

2. Creation of chiral pocket by using dual incorporation concept by using Cu(I)-Catalyzed Azide-Alkyne Cycloaddition (CuAAC)

3. Catalytic testing of the synthesized artificial metalloenzymes

5.3.1 Site-Selective Incorporation of UAA and Synthesis of Corresponding Pincer Type Artificial Metalloenzymes Using CuAAC Approach.

In continuation of our studies on single site mutation that we address in chapter 4, we selected protein mTFP*188pAzF and mTFP*204pAzF for the synthesis of artificial metalloenzymes (ArMs) (see supplementary for the LC-MS data Figure S5.10). An Azide group in protein mTFP*188pAzF is located on the surface of mTFP* and, in case of mTFP*204pAzF, azide group is located in a protein pocket. Therefore, it would be easy to compare both positions for reactivity and enantioselectivity. Next, to create a metal binding-site on the selected mutants, we used the CuAAC reaction approach for achieving our goal. The CuAAC reaction generates a triazole group, which has ability to coordinate to a transition metal. Therefore, we selected derivatives of ethynylpyridines (Figure 5.1), which will form triazole and bistriazole pyridine (BTP) on selected protein. Therefore, a BTP ligand should be the selective site for directing the metal ion and forming a pincer type complex with selected protein site ultimately resulting in formation of artificial metalloenzyme (ArMs).

![Figure 5.1](image-url) Alkynes used for screening and for on protein synthesis of BTP ligands. 1= ethynylpyridine; 2=2-ethynylpyridine and 3=2,6-diethyl-3,5-dimethoxypyridine.

At first, we checked the test experiment between 2,6-diethynylpyridine with excess of azide containing single site mutation protein mTFP* (mTFP*188pAzF and
mTFP*204pAzF, Scheme 5.2). The purpose of this experiment was to optimize the CuAAC reaction of a protein azide with both alkynes at 2 and 6 position of 2,6-diethynylypyridine 2 (Scheme 5.2).

The reaction sample was analyzed by ESI-MS analysis (see Figure S5.11), which shows the a mass matching only one protein clicked on one side of 2,6-diethynylypyridine 2 (Figure 5.2), while the second alkyne remains unaltered (Table 1, entry 2 and entry 9). This observation suggests that because of the large size of azide incorporated protein molecules mTFP* (mTFP*188pAzF or mTFP*204pAzF) the possibility of clicking protein mTFP* on both sides of 2,6-diethynylypyridine is low and will require to add a second conversion step, which should be slow due to steric hindrance (Scheme 5.2). Ultimately, this observation also confirms that there will be no side reaction between two proteins mTFP* which will help to get desired artificial metalloenzymes selectively.

Scheme 5.2. Test experiment between the mTFP*188pAzF and 2,6-diethynylypyridine 2, with higher proteins concentration to pyridine dialkyne (2:1).
Next we tested the possibility to do CuAAC reaction on azide-incorporated protein with alkyne selectively at 2 position of 2,6-diethynylpyridine 2 and a second CuAAC reaction at 6 position with a small azide to generate protein based BTP ligands (Scheme 5.3).

Therefore, reaction with selected protein mTFP*Y188pAzF or mTFP*Y204pAzF is carried out with excess of 2,6-diethynylpyridine (1:10 equivalence) to a corresponding protein on 2 position of 2,6-diethynylpyridine. After 2 hours, to click on 6 positions of 2,6-diethynylpyridine, in situ, subsequently the second azide group containing chemical compound (Scheme 5.3, corresponding compounds A to E) is added in excess (1:100 equivalence). Different aromatic, aliphatic and heterocyclic azides (Scheme 2, azides A to E) were chosen for the second CuAAC click conversion, to generate a library of bis-triaziole pyridine (BTP) based protein ligands.

After completion of reaction, ESI-MS mass show the conversion of azide containing proteins mTFP*Y188pAzF or mTFP*Y204pAzF into the expected BTP type protein ligand due to observation of the expected molecular (Table 5.1). This confirms that it is possible to synthesize a BTP containing protein library using the CuAAC approach (Table 5.1). All new modified BTP proteins were received in excellent conversion of up to 99% (Table 5.1). Theses result clearly indicated the successful establishment of the envisioned library. We also verified all expected motifs by ESI-TOF mass spectrometry (supporting information Figure S5.12 and S5.13).
Scheme 5.3. Binding of different chemical moieties to the fluorescent protein by the CUAAC click reaction with 2,6-diethynylpyridine (up); R: Different azides used for click on second arm of 2,6-diethynylpyridine 2. Different aromatic, aliphatic and heterocyclic azides represent from A to E (down).
Table 5.1. Summary of result for the Click bioconjugation of fluorescent (mTFP*Y188pAzF and mTFP*Y204pAzF) proteins with the different azide ligands A-E. To create protein with different chemical moiety.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein</th>
<th>Azides</th>
<th>Mass Spectrometry [Da]</th>
<th>Fraction BTP Complex* / %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Calculated</td>
<td>Observed</td>
</tr>
<tr>
<td>1</td>
<td>mTFP*Y188pAzF</td>
<td>non</td>
<td>25179.4</td>
<td>25180.0</td>
</tr>
<tr>
<td>2</td>
<td>mTFP*188 pAzF -2</td>
<td>non</td>
<td>25306.4</td>
<td>25306.0</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td></td>
<td>25509.4</td>
<td>25509.0</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td></td>
<td>25439.4</td>
<td>25439.0</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td></td>
<td>25469.4</td>
<td>25469.3</td>
</tr>
<tr>
<td>6</td>
<td>D</td>
<td></td>
<td>25512.4</td>
<td>25512.7</td>
</tr>
<tr>
<td>7</td>
<td>E</td>
<td></td>
<td>25421.4</td>
<td>25422.9</td>
</tr>
<tr>
<td>8</td>
<td>mTFP*Y204pAzF</td>
<td>non</td>
<td>25179.4</td>
<td>25163.0</td>
</tr>
<tr>
<td>9</td>
<td>mTFP*204 pAzF -2</td>
<td>non</td>
<td>25306.4</td>
<td>25306.0</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td></td>
<td>25509.4</td>
<td>25509.7</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td></td>
<td>25493.4</td>
<td>25493.1</td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td></td>
<td>25469.4</td>
<td>25468.5</td>
</tr>
<tr>
<td>13</td>
<td>D</td>
<td></td>
<td>25512.4</td>
<td>25514.1</td>
</tr>
<tr>
<td>14</td>
<td>E</td>
<td></td>
<td>25421.4</td>
<td>25423.9</td>
</tr>
</tbody>
</table>

*Estimated from relative LC-MS signal intensities (relative abundance).

The goal of this thesis is creating artificial metal complexes on a protein. Therefore, we next tested the binding affinities between the BTP based modified mTFP*proteins and Cu metal ion (Figure 5.2). For this study we selected the mTFP*188 pAzF-2-A BTP ligand (Table 5.1 entry 3). As outlined in the experimental section, 0.04 mM protein from the metal binding library were incubated with 2 equivalents of copper (II) nitrate (Cu(NO₃)₂) for 1h. Measurements occurred using a C4 column prior to ESI-TOF principles. We could show successful metal binding through an additional signal (∆M = 64Da) respective to the expected modified Cu-mTFP*188pAzF-2-A BTP ligand (Figure 5.2 A) masses. This result confirms the formation copper containing pincer type artificial metalloenzyme as it shows a single peak corresponding to the protein metal complex i.e. Cu@mTFP*188pAzF-2-A. Due to the well-described difficulty to determine transition metal conjugation to proteins from mass spectrometry data only, we continued further with spectroscopic evaluations. The UV-vis spectrum of protein copper complex shows a shift of the absorption peaks to 470 nm, suggesting the formation of a protein-Cu adduct Cu@mTFP*188pAzF-2-A causes this shifting (Figure 5.2 B). These result strongly
validate Cu binding to the modified protein forming ArMs with complete conversion from mTFP* to mTFP*modified artificial metalloenzyme. The observed a shoulder at about 530 nm was formed in case of Cu@mTFP*188pAzF-2-A (red dashed line), which stems from the coumarin dye of the ligand A.

A similar study has been done for mTFP*Y204pAzF-2-A (Table 5.1, entry 10) and formation of corresponding artificial metalloenzyme is also confirmed i.e. Cu@mTFP*Y204pAzF.

Figure 5.2. A) ESI-TOF analysis of Cu+2 conjugation to the modified protein mTFP*(Y188) pAzF by click reaction with 2,6-diethynylpyridine Cu@mTFP*(Y188) pAzF-2-A (red line). Sign (black arrows) indicate modified protein mTFP*(Y188) pAzF-2 with no metal bind. (P) expected product. B) UV-vis spectra of modified protein mTFP*(Y188) pAzF-2 at 464 nm (green line), and Cu@mTFP*(Y188) pAzF-2-A at 470nm (red dashed line) in 50 mM ammonium acetate buffer pH 7.4.
Catalytic Application of Cu-mTFP*188pAzF-2-A and Cu-mTFP*Y204pAzF for Diels Alder reaction

The catalytic activity and selectivity of the successfully synthesized pincer-type copper artificial copperenzymes (Cu-ArMs i.e. Cu-mTFP*188pAzF-2-A and Cu-mTFP*Y204pAzF-2-A), was established for the Diels Alder [4+2] cycloaddition reaction (Table 5.2).

Table 5.2. Metal binding affinity and enantioselectivity by Diels-Alder reaction.

<table>
<thead>
<tr>
<th>No</th>
<th>Modified protein</th>
<th>L</th>
<th>M^2+</th>
<th>ESI-MS [Da]</th>
<th>Δm</th>
<th>yield%</th>
<th>endo:exo</th>
<th>ee%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Calculated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>mTFP*188pAzF-2-A</td>
<td>A</td>
<td>Cu^{2+}</td>
<td>25572.4</td>
<td>2.9</td>
<td>99</td>
<td>80:20</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>mTFP*204pAzF-2-A</td>
<td>A</td>
<td>Cu^{2+}</td>
<td>25572.4</td>
<td>0.5</td>
<td>94</td>
<td>87:13</td>
<td>0</td>
</tr>
</tbody>
</table>

*a Reaction Conditions: Diels-Alder: Azacalcone (1), Cyclopentadiene (2), Cu(OTf)_2, in MES Buffer at 4 ºC, 20 hours. *b Enantiomeric excesses were determined by chiral HPLC.

It has been reported that Cu^{2+}-complexes catalyze a variety of Diels-Alder reactions in organic solvents and high enantomeric excesses were reported for chiral copper(II) catalysts. Recently, aqueous Diels-Alder protocols were used to test the enantioselectivity and catalytic ability of amino acids, nucleic acids and proteins based Cu(II)-catalyst. We started our screening with our previous optimized reaction conditions for Diels-Alder reaction using azachalcone (1) and cyclopentadiene (2) as substrates and Cu-mTFP*188 pAzF-2-A and mTFP*Y204pAzF-2-A as catalysts (Table...
5.2). After completion of the reaction crude product was analyzed by chiral HPLC to establish the yield and enantiomeric excess of the product. It was observed that reaction gives a complete conversion but no enantioselectivity was observed (Table 5.2, entry 1 and 2). These results suggest that these artificial metalloenzyme show catalytic activity but the environment of the Cu(II) center is not defined enough to generate asymmetric induction. This issue is addressed in the next section describing the approach to create an artificial pocket/chiral site by dual incorporation concept.
5.3.2 Creation of Chiral Pocket by Using Dual Incorporation Concept by Using Cu(I)-Catalyzed Azide-Alkyne Cycloaddition (CuAAC)

To introduce enantioselectivity into the protein, we planned to create artificial pockets within the protein through introduction of two azide groups by dual incorporation of pAzF into mTFP*. The two azide group incorporated should possess the appropriate distance (1.2 nm) and orientation (approx. 130° angle between the azide vectors) to allow linkage by a dietyyne cofactor. Further, they should be accessible to Cu(II) and the diylkyne to allow cyclization (see supplementary for more design details Figure S5.7 and S5.8). To enhance the electron donating effect of the BTP ligand, we introduced methoxysubstituents in the 3- and 5-positions of 2,6-diethynylpyridine (2,6-diethynyl-3,5-dimethoxypyridine 3, Figure 5.1). The on protein formation of the BTP pincer is illustrated in (Scheme 5.4).

We identified two combinations of mutation sites that fulfill the criteria above and should hence allow the BTP synthesis: positions 128 and 21 at the exposed position as well as positions 200 and 164 at the rim of the cavity. The dual site incorporation of pAzF into mTFP* was tested by producing two different variant of the protein. The two pAzF were successfully incorporated into mTFP*(128N/21Y)pAzF (P1) and mTFP*(200Y/164E)pAzF (P2) and the incorporation efficiency was confirmed by LC-MS analysis (Table 5.3 entry 1 and 4) (see supplementary material for the LC-MS data Figure S5.14 A and B). Protein yields were 35 mg/L for mTFP*(128N/21Y)pAzF (P1) and 88 mg/L for mTFP*(200Y/164E)pAzF (P2).
First, to test the applicability of click protocol in the dual incorporation sites, we tested the click reaction between the mTFP* variants (P1 and P2) with the monoalkyne 2-ethynylpyridine 1 (Table 5.3 entry 2 and entry 5) at a protein to ligand ratio of 1:100. As depicted in Table 5.3 the products mTFP*(128N/21Y)pAzF-1 and mTFP*(200Y/164E)pAzF-1 were detected with their exact mass by LC-MS analysis with 81% and 86% purity respectively (Table 5.3 entry 2 and entry 5). Interestingly, it was notable that the major product was the one 2-ethynylpyridine 1 clicked to each protein with a good conversion P1 (81%) and P2 (86%) in presence of two azide groups incorporated in close vicinity. While the minor product detected corresponded to 2 molecule of mono-alkyne bound to two azide groups groups incorporated in close vicinity in P1 and P2 with conversions ranging between 22% and 18% respectively (supplementary Figure S 5.14 C and D blue arrow). These results demonstrated the successful strategy to introduce only one pyridine alkyne by the click reaction. This selectivity results presumably from the steric interaction and the space boundary of the surrounding amino acid in each pocket.
Secondly, to make actual bridge the pockets on mTFP* we examined the linkage using the diazido incorporated protein mTFP*(128N/21Y)pAzF or mTFP*(200Y/164E)pAzF with the 2,6-diethynyl-3,5dimethoxypyridine ligands, i.e. 3 (Scheme 5.4). The mass spectrum analysis confirms the good purity obtained: 88% for mTFP*(128N/21Y)pAzF-3 (Table 5.3, entry 3) and 91% for mTFP*(200Y/164E)pAzF-3 (Table 5.3, entry 6) (supplementary Fig S5.14 E and F) (According to this MS analysis, the remaining 9 – 12% are small amount of other masses may be degraded protein formed during the reaction). Converting the reaction mixture once more with newly added copper catalyst and and 100-fold excess of the coumarin azide (Scheme 5.3, ligand A) did not lead to any noticeable change in the MS spectra, hence no free alkyne group is accessible for conversion with additional external azide. These results validate the choice of mutation sites for dual azide group incorporation yielding proteins P1 and P2. At the selected positions, both groups are close enough and oriented properly to click with the same diethynyl pyridine 3 resulting in formation of on-protein BTP ligands (Scheme 5.1).

Under the chosen reaction conditions, the intramolecular cyclization is favored over the addition of a second external biyalkyne. This can be attributed to the dilution effect of a surface, which is well-established for SOMC (surface organometallic chemistry) and/or the entropic enhancement of an intramolecular reaction.

**Table 5.3.** Incorporation efficiency of pAzF to the dual site into mTFP* and the CuAAC reaction between ethynylpyridine 1 or 2,6-diethynyl-3,5 dimethoxypyridine3 and two different variants of mTFP* pAzF. a

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein</th>
<th>L</th>
<th>Mass Spectrometry [Da]</th>
<th>Δm</th>
<th>Fraction BTP Complex b/ %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mTFP*(128N/21Y)pAzF (P1)</td>
<td>non</td>
<td>25253.8</td>
<td>25250.2</td>
<td>3.6</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td>25356.9</td>
<td>25357.1</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
<td>25441.0</td>
<td>25439.7</td>
<td>1.3</td>
</tr>
<tr>
<td>4</td>
<td>mTFP*(200Y/164E)pAzF (P2)</td>
<td>non</td>
<td>25238.8</td>
<td>25236.5</td>
<td>2.3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td></td>
<td>25341.9</td>
<td>25340.3</td>
<td>1.6</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td></td>
<td>25426.0</td>
<td>25423.1</td>
<td>2.9</td>
</tr>
</tbody>
</table>

*a Reaction condition: protein mTFP* (0.05mM), 0.5 mL HEBES buffer pH7.5, (ethynylpyridine 1 or 2,6-diethynyl-3,5 dimethoxypyridine 3 100 equiv) in 10% acetone, CuSO4 (1 equiv), THPTA (5 equiv), Na- ascorbate (50 equiv) at 30 °C, 2h.

*b Estimated from relative LC-MS signal intensities (relative abundance).
To convert the BMP-containing precursors into artificial metalloenzymes we added equimolar amounts of copper(II) and palladium(II) since copper and palladium can be applied widely in to catalyze organic transformations\(^{15,16}\). Among the different metal salts tested, Cu(NO\(_3\))\(_2\) and Pd(OAc)\(_2\) showed the best binding efficiency for synthesis for copper or palladium based artificial metalloenzymes (Cu or Pd ArMs). The ESI-MS analysis of all tested proteins suggested a quantitative binding of the metal to the BMP-moiety forming Cu@mTFP*(128N/21Y)pAzF-3, Pd@mTFP*(128N/21Y)pAzF-3 (Figure 5.3 A), Cu@mTFP*(200Y/164E)pAzF-3, and Pd@mTFP*(200Y/164E)pAzF-3 (Figure 5.4 A), (Table 5.4 and Table 5.5). Further confirmation of the metal binding comes from the UV-Vis spectrum, which shows a shift of the absorption peaks to 470nm for the modified proteins with Cu\(^{2+}\) (Figure 5.3 B) and 474 nm for modified proteins with Pd\(^{2+}\) (Figure 5.4 B).

---

**Figure 5.3.** A) ESI-TOF analysis of Cu\(^{2+}\) and Pd\(^{2+}\) coordination to the modified protein mTFP*(128N/21Y)pAzF (P1) by click reaction with ligand 3, Cu@mTFP*(128N/21Y)pAzF-3 (red line), and Pd@mTFP*(128N/21Y)pAzF-3 (blue line). B) UV-Vis spectra of modified protein mTFP*(128N/21Y)pAzF-3 (P1) (green line), Cu@mTFP*(128N/21Y)pAzF-3 (red dashed line) and Pd@mTFP*(128N/21Y)pAzF-3 (blue dashed line) in 50 mM ammonium acetate buffer pH 7.4.
Figure 5.4. A) ESI-TOF analysis of Cu$^{2+}$ and Pd$^{2+}$ coordination to the modified protein mTFP*(200Y/164E)$_p$AzF (P2) by click reaction with ligand 3, Cu@mTFP*(200Y/164E)$_p$AzF-3 (red line), and Pd@mTFP*(200Y/164E)$_p$AzF-3 (blue line). B) UV-vis spectra of modified protein mTFP*(200Y/164E)$_p$AzF-3 (P2) (green line), Cu@mTFP*(200Y/164E)$_p$AzF-3 (red dashed line) and Pd@mTFP*(200Y/164E)$_p$AzF-3 (blue dashed line) in 50 mM ammonium acetate buffer pH 7.4.

5.3.3 Catalysis as an Application

After successful synthesis of copper and palladium based artificial metalloenzymes (Cu or PdArMs), we tested their catalytic potential. Recently, various reports studied the suitability of artificial metalloenzymes to catalyze Diels-Alder and Tsuji-Trost reactions. These reactions are well-suited, since both, Diels-Alder reactions and Tsuji-Trost reactions can lead to chiral products and reactions can be carried out in aqueous conditions.
5.3.3.1 **Diels Alder Reactions**

The copper bound artificial metalloenzymes (Cu-ArMs) Cu@ mTFP*(128N/21Y)pAzF-3 (Table 5.4, entry 1) and Cu@mTFP*(200Y/164E)pAzF-3 (Table 5.4, entry 2) were screened in the Diels-Alder reactions using our previous optimized conditions\textsuperscript{29}.

Table 5.4. Conjugation of metal ions (Cu\textsuperscript{2+} and Pd\textsuperscript{2+}) to modified proteins with ligand 3 mTFP*(N128-Y21)pAzF-3 and mTFP*(200Y/164E)pAzF-3 and its application in Diels-Alder reactions.

<table>
<thead>
<tr>
<th>No</th>
<th>protein-ligand</th>
<th>M\textsuperscript{2+}</th>
<th>ESI-MS[Da]</th>
<th>Δm</th>
<th>Yield\textsuperscript{a}</th>
<th>endo:exo</th>
<th>ee\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Calculated</td>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>mTFP*(128N/21Y)pAzF-3</td>
<td>Cu\textsuperscript{2+}</td>
<td>25504.54</td>
<td>25502.3</td>
<td>3.7</td>
<td>98</td>
<td>83:17</td>
</tr>
<tr>
<td>2</td>
<td>mTFP*(200Y/164E)pAzF-3</td>
<td>Cu\textsuperscript{2+}</td>
<td>25489.54</td>
<td>25487.2</td>
<td>2.6</td>
<td>88</td>
<td>88:12</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Reaction conditions: Diels-Alder: Azacalcone (1), Cyclopentadiene (2), Cu(OTf)\textsubscript{2}, in MES Buffer at 4 \degree C, 20 hours. \textsuperscript{b} Enantiomeric excesses were determined by chiral HPLC.

Almost full conversion was obtained in both reactions with 98% yield and a moderate ee, 27% for Cu- mTFP*(128N/21Y)pAzF-3 (Table 5.4, entry 1), and 88% yield with 32% ee for Cu-mTFP*(Y200-E164) pAzF-3, (Table 5.4, entry 2). As expected, these results confirm that these Cu-ArMs possess catalytic activity as well as enantioselectivity.
### 5.3.3.2 Tsuji Trost Reactions

Following the tests in the Diels-Alder reactions with dual anchored catalysts (Cu-ArMs), we check the applicability of palladium containing artificial metalloenzymes (Pd-ArMs) as Tsuji Trost catalysts.

**Table 5.5.** Conjugation of Pd+2 ion to modified proteins with ligand 2mTFP*(N128-Y21) $p$AzF-3 and mTFP*(Y200-E164) $p$AzF-3 and its application in Tsuji trost reactions.\(^a\)

<table>
<thead>
<tr>
<th>No</th>
<th>protein-ligand</th>
<th>$M^{2+}$</th>
<th>ESI-MS [Da]</th>
<th>$\Delta m$</th>
<th>Yield (%)</th>
<th>ee(^b) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mTFP*(128N/21Y)$p$AzF-3</td>
<td>Pd(^{2+})</td>
<td>25547.42</td>
<td>25545.7</td>
<td>2.2</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>mTFP*(200Y/164E)$p$AzF-3</td>
<td>Pd(^{2+})</td>
<td>25532.42</td>
<td>25530.7</td>
<td>1.7</td>
<td>81</td>
</tr>
</tbody>
</table>

\(^a\) **Reaction conditions:** 1,3-diphenylpropenylacetate (1.892 mg, 0.0075 mmol, 1 equiv), acetylacetone (3 equiv), (ArMs, 0.050 equiv, HEPES buffer pH 7.5 0.5 mL) allylpalladium(II) chloride (0.025 equiv), 4°C, 24 h. \(^b\) Enantiomeric excesses were determined by chiral HPLC.

Tsuji Trost reactions resulted in 95% yield and 18% ee for Pd@mTFP*(128N/21Y)$p$AzF-3 (Table 5.5, entry 1) and 81% yield with 21% ee for Pd@mTFP*(200Y/164E)$p$AzF 3 (Table 5.5, entry 2). It should be noted that an Pd-ArM published by Kamer et al. did not achieve any enantioselectivity in this reaction\(^{31}\). Therefore, these results are promising for the development of artificial metalloenzymes with enhanced activity and selectivity.
5.4 Conclusion

Overall, the comparison of the dual anchor catalysts (Table 5.4 and Table 5.5) with single site attached catalysts (Table 5.2) demonstrates the advantages of the dual anchoring strategy, which in achieves higher catalyst activity as well as enantioselectivity. These results indicate that a dual attachments strategy is a more effective alternative to the common single attachment to control selectivities. During the catalytic application test on Diels-Alder and Tsuji-Trost reactions, we notice a higher ee was obtained from the catalysts contained the metals at the cavity pocket on the dual anchoring protein version Cu@ mTFP*(200Y/164E)pAzF-3 and Pd@mTFP*(200Y/164E)pAzF-3. These results strongly indicate that the dual anchoring strategy limits the number of conformational states available to the metalloprotein complex inside the protein, which improves reaction ee. In general, the catalytic application study shows the effectiveness and scope for designing artificial pockets into the protein for catalytic application.
5.5 Materials and Methods

5.5.1 Cloning

The gene of monomeric Teal Fluorescent Protein mTFP* (described in Chapter 3) was engineered in our lab. The mTFP* gene cloned into the ChampionTM pET303/CT-His vector. This vector is supplemented with a N-terminal SUMO-Tag. mTFP* gene was digested by BAMHI and XhoI restriction enzymes (New England Biolabs Inc.) and subsequently fused to a N-terminal SUMO Tag to generate the construct pET303/SUMO_mTFP*. The engineered vector has mTFP* linked directly to SUMO-Tag. FtnA gene was fused to the same vector to generate pET303/SUMO_FtnA. Post expression cleavage of tag allows expression of the proteins in a native form with no residual linker.

5.5.2 Site-Directed Mutagenesis

Site-directed mutagenesis was performed on pET303/SUMO_mTFP*. detailed of all Mutants nucleotide sequence and Primers is shown in the Supplementary section (Table S5.6). Mutant variants of mTFP*TAG (TAG, mTFP*Y188 TAG and mTFP*Y204 TAG) were generated via QuickChange® Site-Directed Mutagenesis Kit II (StratageneTM, Agilent Technologies). All primers were designed according to the user guidelines of the QuikChange Kit and subsequently obtained from Eurofins MWG Operon. Mutant plasmids were transformed into E. Coli BL21 (DE3) Gold (StratageneTM, Agilent Technologies) and purified using peqGOLD Plasmid MiniPrep Kit II (peqlabBiotechnologie GmbH). Purities and successful mutagenesis were checked by 1% Agarose Electrophoresis and Sanger Sequencing (KAUST Core Facilities). For Uaa site incorporation the resulting plasmids were Co-transformed with respective pEvol-aaRS plasmids (from Peter G. Schultz) (pEVOLpAzF).

5.5.3 Expression

Pre-cultures that contain 180mg L\(^{-1}\) ampicillin and 25mg L\(^{-1}\) were grown overnight at 37°C and were further used to inoculate expression cultures under standard protocol. For protein production the LB Medium containing 180mg L\(^{-1}\) ampicillin and 25mg L\(^{-1}\)Chloramphenicol was inoculated with E. Coli BL21 (DE3) Gold cells, (Agilent
Technologies), that had been transformed with pET303/SUMO_mTFP* and pEVOLpAzF at 37 °C. The media conc. started at OD 0.05 at wavelength 600nm, and incubate overnight at 20 °C, induction by 1mM IPTG and 0.2% of L-arabinose followed by 1mM of corresponding unnatural amino acid was applied when the OD reached 0.6-0.7. Cultures were then centrifuged at 4500 rpm for 30 minutes and cooled to 4 °C. Cell pellets were washed with Lysis Buffer A, containing 100 mMTris/HCl pH 7.4, 500 mMNaCl and 20mM Imidazole. Before lysis complete protease Inhibitor cocktail (Roche), was added together with DNase. Lysis occurred using French press (Constant Systems Ltd) at 1.7KBar. Lysates were immediately centrifuged at 25000 rpm for 30min and also cooled to 4 °C.

5.5.4 Purification

All protein were purified by using Ni-Affinity Chromatographic principles. The respective supernatant was transferred to a pre-washed nickel column and eluted from the column between 20 and 35% of Elution Buffer B containing 100 mMTris/HCl pH 7.4, 500 mMNaCl and 500mM Imidazole. The eluted protein was digested with SUMO protease overnight in the shaker at 30 °C and the mixture then dialysed overnight, at 4 °C, against Buffer A. Further Ni-Affinity Chromatography followed to remove the SUMO-Tag. Elution occurred with a smooth gradient up 20 % of the same Buffer containing additional 1 M NaCl. Pure Protein was collected from the flow-through before elution with Buffer B which showed a distinct elution of residual SUMO-Tag and Protease. The purity of mTFP* was analyzed by SDS–PAGE gel and its fluorescence was measured at excitation at 462 nm and an emission at 492 nm. All incorporation confirmation was done using LC/MS to detect the mass differences before (wt protein) and after the incorporation (modified protein).

5.5.5 UV-Vis Spectral Analysis

Absorbance and Fluorescence measurements were performed with pure protein samples, 0.04mM mTFP* or respective mutant in 50mM Tris/HCl, pH 7.4 using a TECAN INFINITE M1000 according to standard procedure.
5.5.6 Mass Spectrometry

Protein mass spectrometry was carried out by BRUKER maXis HD™ ESI-TOF. the purified pure proteins were dialyzed against (50 mMTris/HCl pH 7.4) and concentrated to 8-10 mg/mL. The protein sample was diluted in buffer containing 5% ACN, 0.1% formic acid, and analyzed using maXis UHR-TOF mass spectrometer (Bruker) coupled to (Agilent1260) HPLC system. Briefly, 2µL of the sample was loaded onto a C4 column (Advantage 300+ C4, 5µm particles, 50 x 2.1 mm, Analytical sales), and desalted using a divert valve at 500µl/min flow rate for 15 min. The proteins were eluted at 200µl/min flow rate using a linear gradient of 5-80% Acetonitrile (in 0.1% formic acid and H2O) for 10 min, followed by isocratic gradient of 80% for 5 min, and then isocratic gradient at 5% for 5 min. MS analysis was carried out using (Bruker) ESI source at 4200 v potential, 1.4 bar nebulizer gas and 8.0 l/min dry gas at a temperature of 200 °C. The MS data was collected in positive mode, and m/z of ions was determined at rolling average of 2. The peak list was extracted using Compass Data Analysis software ver 4.0 (Bruker) at signal to noise ratio of 5. The mass of protein was deconvoluted where minimum 5 peaks for the specific large molecule were identified.

5.5.7 Click Protocol

5.5.7.1 Protocol for CuAAC Reaction on Single Site Azide Group Incorporate Protein mTFP*188 pAzF or mTFP*204 pAzF

In a 2 mL glass vial equipped with magnetic stirring bar, first the single site azide group incorporated protein mTFP*188pAzF or mTFP*204pAzF (0.05 mM, 0.5 mL HEPES buffer pH 7.5) was added. Then next ligand THPTA (5 equiv) was added and protein solution was stirred for 2 minutes. To the protein reaction mixture, a solution of CuSO₄ (1 equiv) was added, followed by addition of 2,6-diethynylpyridine (2, 10 equiv in 10% acetone). Finally, reaction was started by addition of sodium ascorbate (50 equiv). Reaction tube was closed with the rubber septum and reaction mixture flush with nitrogen (to prevent more oxygen from diffusing in) and stirred the reaction mixture for 2 hours. After 2 hours one arm of 2,6-dialkyne pyridine were clicked with protein, then for second click on the another arm of 2,6-diethynylpyridine a azide group containing
corresponding chemical compound is added in excess (100 equivalence). Reaction tube was closed with the rubber septum and reaction mixture flush with nitrogen (to prevent more oxygen from diffusing in), After the reaction protein were purified by using Ni-Affinity Chromatographic principles.

5.5.7.2 Protocol for CuAAC Reaction on Duel Site Azide Group Incorporate Protein mTFP*(N128-Y21) pAzF or mTFP*(Y200-E164)

In a 2 mL glass vial equipped with magnetic stirring bar, first duel site azide group incorporate protein mTFP*(N128-Y21) pAzF or mTFP*(Y200-E164) (0.05 mM, 0.5 mL HEPES buffer pH 7.5) was added. Then next ligand THPTA (5 equiv) was added and protein solution was stirred for 2 minutes. To the protein reaction mixture a solution of CuSO₄ (1 equiv) was added, followed by addition of corresponding 2-ethynylpyridine or 2,6-diethynyl-3,5-dimethoxypyridine (1 or 3,100 equiv in 10% acetone). Finally, reaction was started by addition of sodium ascorbate (50 equiv). Reaction tube was closed with the rubber septum and reaction mixtures flush with nitrogen (to prevent more oxygen from diffusing in) and stirred the reaction mixture for 2 hours. After the reaction protein were purified by using Ni-Affinity Chromatographic principles.

5.5.8 Catalysis

5.5.8.1 Diels-Alder Activity

Cyclopentadiene (2) was freshly prepared from dicyclopentadiene and stored at -80 °C. Azachalcone (1) was prepared according to literature. Hybrid catalysts on the basis of modified protein were prepared by weighing appropriate amounts of lyophilized catalyst and dissolving in appropriate MES Buffer, or mixing a solution of the desired variant (970 µL, 3 mg/ml of Protein in MES Buffer, pH as anticipated) with Cu(OTf)₂ solution (10 µL 10 mM Cu(OTf)₂) and incubation for 1 h (400 rpm, 20 °C, Thermomixer comfort, Eppendorf). These hybrid catalysts were immediately used for Diels-Alder reactions: a solution of 1 (0.1 M in Acetone, 10 µL) were added to the catalyst solution and mixed immediately by inverting the tube 3 times. Then, a solution of 2 (10 M in Acetone, 9 µL) was added and mixed immediately as in the last step. The tube was placed immediately in
a pre-cooled shaking incubator (4 °C, 400 rpm, Thermomixer comfort, Eppendorf). The standard reaction time was 20 h. Reactions were extracted with diethyl ether (2 × 1 ml). The combined organic phase was dried by filtering over MgSO₄ and the solvent was evaporated. The samples were analyzed by HPLC using anisole as internal standard (column: 250 mm Chiralcel OD-H, 4.6 mm inner diameter; n-heptane/2-propanol = 99:1; 0.5 ml/min, 25 °C; UV-detector 220 nm; retention times: 10.4 min (anisole), 14.2 min (exo-4), 16.1 min (exo-4'), 20.1 min (endo-3), 25.4 min (endo-3'), 31.7 min (3)). The absolute configuration of the enantiomers 3 has not been established in literature²⁵,³³,³⁴.

5.5.8.2 Tsuji Trost Protocol
In a 2 mL glass vial equipped with magnetic stirring bar first solution of proteinmTFP*(N128-Y21) pAzF -2 or mTFP*(Y200-E164) pAzF -2 (0.050 equiv, HEPES buffer pH 7.5 0.5 mL) a allylpalladium(II)chlorie (0.025 equiv) were added and mixed in a vial for 30 minutes at room temperature. Then 1,3-diphenylpropenylacetate (1.892 mg, 0.0075 mmol, 1 equiv) was added and stir reaction mixture at 4 °C for 30 minutes, finally, sodium acetylacetonate (3 equiv) were added to the catalysis mixture. The resulting cloudy mixture was stirred at 4 °C for 24 h. After completion, the reaction was extracted with Et₂O (2 x 500 microL), filtered through a short silicagel. The samples were analyzed by HPLC using diphenylether as internal standard (column: 250 mm Chiralcel OD-H, 4.6 mm inner diameter; n-heptane/2-propanol = 99:1; 0.5 ml/min, 25 °C; UV-detector 220 nm; The absolute configuration of the enantiomers 3 has not been established in literature. The product yields were determined by HPLC using calibration curves of as internal standard and an authentic sample of the product.
5.6 References

29. Fischer JPN. A Pathway to Artificial Metalloenzymes [PhD Dissertation]: King Abdullah University of Science and Technology Thuwal, Kingdom of Saudi Arabia; 2015.
5.7 Supplementary Material

5.7.1 Molecular Biology

mTFP*188
GGCGTAATCAAGCCGCACATGAAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCTATGCCTTCGTGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACCAACACCATCAACCTGGAGGTGAAGGAGGGAGCCCCCCTGCCCTTCTCCTACGACATTCTGACCACCGCGTTCGCCTACGGCAACAGGGCCTTCACCAAGTACCCCGACGACATCCCCAACTACTTCAAGCAGTCCTTCCCCGAGGGCTACTCTTGGGAGCGCACCATGACCTTCGAGGACAAGGGCATCGTGAAGGTGAAGTCCGACATCTCCCTGGAGGAGGACTCCTTCATCTACGAGATATATCTCAAGGGCGAGAACTTCCCCCCC AACGGCCCCGTGATGCAGAAGAAGACCACCGGCTGGGA
CGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGACGTCAAGCACAAGCTGCTGGAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGGGCCAAGAAGGCGGTGAAGCTGCCCGACTATCACTTTGTGGACCACCGCATCGAGATCCTGAAC TATGACAAGGAC

mTFP*204
GGCGTAATCAAGCCGCACATGAAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCTATGCCTTCGTGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACCAACACCATCAACCTGGAGGTGAAGGAGGGAGCCCCCCTGCCCTTCTCCTACGACATTCTGACCACCGCGTTCGCCTACGGCAACAGGGCCTTCACCAAGTACCCCGACGACATCCCCAACTACTTCAAGCAGTCCTTCCCCGAGGGCTACTCTTGGGAGCGCACCATGACCTTCGAGGACAAGGGCATCGTGAAGGTGAAGTCCGACATCTCCCTGGAGGAGGACTCCTTCATCTACGAGATATATCTCAAGGGCGAGAACTTCCCCCCC AACGGCCCCGTGATGCAGAAGAAGACCACCGGCTGGGA
CGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGACGTCAAGCACAAGCTGCTGGAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGGGCCAAGAAGGCGGTGAAGCTGCCCGACTATCACTTTGTGGACCACCGCATCGAGATCCTGAAC TATGACAAGGAC

mTFP*di128_21
GGCGTAATCAAGCCGCACATGAAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCTATGCCTTCGTGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACCAACACCATCAACCTGGAGGTGAAGGAGGGAGCCCCCCTGCCCTTCTCCTACGACATTCTGACCACCGCGTTCGCCTACGGCAACAGGGCCTTCACCAAGTACCCCGACGACATCCCCAACTACTTCAAGCAGTCCTTCCCCGAGGGCTACTCTTGGGAGCGCACCATGACCTTCGAGGACAAGGGCATCGTGAAGGTGAAGTCCGACATCTCCCTGGAGGAGGACTCCTTCATCTACGAGATATATCTCAAGGGCGAGAACTTCCCCCCC AACGGCCCCGTGATGCAGAAGAAGACCACCGGCTGGGA
CGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGACGTCAAGCACAAGCTGCTGGAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGGGCCAAGAAGGCGGTGAAGCTGCCCGACTATCACTTTGTGGACCACCGCATCGAGATCCTGAAC TATGACAAGGAC

mTFP*200_164
GGCGTAATCAAGCCGCACATGAAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCTATGCCTTCGTGATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACCAACACCATCAACCTGGAGGTGAAGGAGGGAGCCCCCCTGCCCTTCTCCTACGACATTCTGACCACCGCGTTCGCCTACGGCAACAGGGCCTTCACCAAGTACCCCGACGACATCCCCAACTACTTCAAGCAGTCCTTCCCCGAGGGCTACTCTTGGGAGCGCACCATGACCTTCGAGGACAAGGGCATCGTGAAGGTGAAGTCCGACATCTCCCTGGAGGAGGACTCCTTCATCTACGAGATATATCTCAAGGGCGAGAACTTCCCCCCC AACGGCCCCGTGATGCAGAAGAAGACCACCGGCTGGGA
CGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGACGTCAAGCACAAGCTGCTGGAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGGGCCAAGAAGGCGGTGAAGCTGCCCGACTATCACTTTGTGGACCACCGCATCGAGATCCTGAAC TATGACAAGGAC

Figure 5.5. Nucleotide sequence for mTFP*188TAG, mTFP*204TAG and FtnA79TAG (Red Highlight = mutation to stop codon TAG for pAzF and pEnyF incorporation).
<table>
<thead>
<tr>
<th>Protein-Mutation position</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>mTFP*-Y204TAG(^f)</td>
<td>5'-GAAATCCTGAACTACGACAAAGACTAGAAACAAAGTTACCCTG-3'</td>
</tr>
<tr>
<td>mTFP*-Y204TAG(^r)</td>
<td>5'-ACGGTAACTTGTCTAGTCTTTGTCGTTACTCAAGGATTTC-3'</td>
</tr>
<tr>
<td>mTFP*-Y188TAG(^f)</td>
<td>5'-TGTTAAACTGCCGACTAGCAGTCTCTCAGTGACCAC-3'</td>
</tr>
<tr>
<td>mTFP*-Y188TAG(^r)</td>
<td>5'-GTGGTCAACGAAAGTGCTACGTCCGGCAAGTTAAACA-3'</td>
</tr>
<tr>
<td>mTFP*di128-21 TAG(^f)</td>
<td>5'-GGCAACGATGAAATGCTAGGCACTGTTTGATCG-3'</td>
</tr>
<tr>
<td>mTFP*di128-21 TAG(^r)</td>
<td>5'-CGATCACGAAGCGCTAGGCATTACGTGGCC-3'</td>
</tr>
<tr>
<td>mTFP*di200-164 TAG(^f)</td>
<td>5'-AGCTGCTGCTAGGCTGGCGCGGC-3'</td>
</tr>
<tr>
<td>mTFP*di200-164 TAG(^r)</td>
<td>5'-GCCGCGCCCTACAGACAGCAGCT-3'</td>
</tr>
</tbody>
</table>
Figure S5.6. pET303/SUMO_plasmids.
5.7.2 Structural Identification for Mutation Site

The crystal structure of mTFP* (PDB: 4Q9W) suggested an ideal spacing of neighboring residues Y200 and Y204 (2HQKY204 = 4Q9WY200), having roughly 4 Å distance in three-dimensional space. Further, the surrounding loops also prompted residues I197 (non-polar aliphatic), D55 (polar negative) and Y54 (non-plar-aromatic) to be taken into consideration contribute in the effectiveness of the metal binding affinity to the special ligand at this pocket. Hence, we chose the Y200-E164 positions as suitable sites for pAzF attachment on the mTFP* pocket. Therefore, we anticipated the successful coupling of the special synthesized di-alkyne ligand by click reaction to this positions, and subsequent proper metal-binding sites for creation the metal hybrid catalyst. we anticipated the limited space present in this pocket give more fix coordination to the metal center in this pocket which raise the possibility and the percentage of the provided enantioselectivity.

![Figure 5.7](image1.png)

**Figure 5.7.** Selection of Binding UAAs in the Pocket Cavity positions mainly: Y200 for the dual incorporation and Y204 for the single incorporation.

![Figure 5.8](image2.png)

**Figure 5.8.** proposed pocket assembly for metal centre complex, by double sites incorporation of UAA.
Figure 5.9. Proteins structure: for single site incorporation A) mTFP* (Y188) pAzF and B) mTFP* (Y204) pAzF. And dual site incorporation C) mTFP*(N128-Y21) pAzF and D) mTFP* (Y200-E164) pAzF. (PDB:4Q9W).
MS analysis for protein Modification and Conjugation

**Figure 5.10.** ESI-TOF analysis of the incorporation of unnatural amino acids pAzF into the TAG position of mTFP*Y188 TAG and mTFP*Y204 TAG. A) mTFP*Y188pAzF 25180 Da, B) mTFP*Y204pAzF 25180 Da. wt = is the wild type mTFP* protein before the incorporation process.

**Figure S5.11.** ESI-TOF analysis of the mTFP*188pAzF with 2,6-diethynlypyridine ligand 2. The only product forms from the coupling reaction.
Figure S5.12. ESI-TOF analysis of the mTFP*Y188pAzF protein modified by click reaction with different azide ligands A-E. A) mTFP*Y188 pAzF -2 A. B) mTFP*Y188 pAzF-2 B. C) mTFP*Y188 pAzF-2 C. D) mTFP*Y188 pAzF -2 D. E) mTFP*Y188 pAzF -2 E. Sign (black arrows) indicate mTFP*wt, unmodified mTFP*Y pAzF, intermediate product (protein with dialkyne ligand) and Cu binding to the ligand forming coordination complex proteins (25154, 25180, 25306 and product Mwt +63Da) respectively. (P) expected product. wt = is the wild type mTFP** protein before the incorporation process.
Figure S5.13. ESI-TOF analysis of the mTFP*Y204pAzF protein modified by click reaction with different aizde ligands A-E. A) mTFP*Y202 pAzF-2 A. B) mTFP*Y204 pAzF-2 B. C) mTFP*Y204 pAzF-2 C. D) mTFP*Y204 pAzF-2 D. E) mTFP*Y204 pAzF-2 E. Sign (black arrows) indicate mTFP*wt, unmodified mTFP*Y pAzF, intermediate product (protein with dialkyne ligand) and Cu binding to the ligand forming coordination complex proteins (25154, 25180, 25306 and product Mwt +63Da) respectively. (P) expected product. wt = is the wild type mTFP* protein before the incorporation process.
Figure 5.14 ESI-TOF analysis of the dual incorporation efficiency of A) mTFP*(N128-Y21) pAzF (P1), B) mTFP*(Y200-E164) pAzF (P2) and their modification by the CuAAC reaction with the mono alkyne ligand 2-ethynylpyridine 1, C) mTFP*(N128-Y21) pAzF-1, D) mTFP*(Y200-E164) pAzF-1. And dialkyne ligand 2,6-diethynyl-3,5 dimethoxypyridine 3, E) mTFP*(N128-Y21) pAzF-3, D) mTFP*(Y200-E164) pAzF-3.

Synthesis of ethynylpyridines and its derivatives
Compound 1 was purchased from Sigma Aldrich, while compound 2 is reported compounds and synthesized by reported methods\textsuperscript{36}. However, compound 3 is synthesized in our laboratory and characterized by NMR.

Table S5.7. Alkynes used for screening and for synthesis of BTP ligands on protein. 1= ethynylpyridine; 2=2-ethynylpyridine and 3=2,6-diethynyl-3,5-dimethoxypyridine.

<table>
<thead>
<tr>
<th>Ligands</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td><img src="image1" alt="Structure1" /></td>
<td><img src="image2" alt="Structure2" /></td>
<td><img src="image3" alt="Structure3" /></td>
</tr>
<tr>
<td>Chemical name</td>
<td>2-ethynylpyridine</td>
<td>2,6-diethynylpyridine</td>
<td>2,6-diethynyl-3,5-dimethoxypyridine</td>
</tr>
<tr>
<td>M.w</td>
<td>103.12</td>
<td>127.15</td>
<td>187.2</td>
</tr>
</tbody>
</table>
NMR Data:

Figure S5.15 2,6-diethynyl-3,5-dimethoxypyridine (Figure 1, compound 3):

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 6.71 (S, 1H, A), 3.94 (s, 6H, B), 3.29 (s, 2H, C)

$^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 158.76 (s, 1H), 124.65 (s, 1H), 101.56 (s, 2H), 80.70 (s, 4H), 78.96 (s, 1H), 56.42 (s, 3H).
6 CHAPTER 6 Heterogeneous Protocol for Click Labeling of Biomolecules Using a Recyclable Cu/Cu$_2$O Nano Wire-Catalyst

6.1 Abstract
Copper catalyzed azide–alkyne cyclo-addition (CuAAC) reactions have found widespread applications to study biological molecules. However, for most of the reported bio-conjugation CuAAC protocols, an excess of copper salt, ligands and Na-ascorbate is applied. This is particularly problematic, since, homogeneous copper and Na-ascorbate contribute to adverse side-reactions and formation of toxic radicals, which cause damage to bio-molecules (Protein, DNA, RNA). Also, high contaminations of copper in the formed biomolecule-conjugate can be a major issue.

Herein, we report the application of 5 nm thin Cu$_2$O layers supported on Cu(0) nanowires (Cu$_2$O/Cu-NW) as heterogeneous catalyst for CuAAC bio-conjugation. Our protocol circumvents the formation of radicals that lead to damage of biomolecules, while the catalyst is easily separated from the solution by filtration to eliminate copper contamination. We demonstrate the use of this protocol for the efficient site-selective conjugation of ligands on biomolecules containing unnatural amino acid (UAA) with azido- or ethynyl-side chains.
6.2 Introduction

In 2002, Sharpless and Meldal independently reported the homogeneous Copper(I)-Catalyzed Alkyne-Azide Cycloaddition (CuAAC) reaction\(^1\). The success of these CuAAC reactions results from an excellent efficiency, selectivity, and orthogonality that allows straightforward covalent connection of two different molecules. Due to the water-compatibility of the reaction, it has become a highly protocol in medicine and (molecular) biology\(^3\). Finn et al., 2009 first reported optimized reaction conditions for CuAAC bio-conjugations\(^4\). Since then, implementation of CuAAC based labeling of biomolecules has been found in studies on proliferation, DNA repair, cell imaging and quantification of RNA transcription\(^5\). Furthermore, it is used for in-depth or \textit{in vivo} proteomics as azide- or alkyne-tagged biomolecules can be detected after conjugation with a corresponding alkyne- or azide-containing fluorescent dye\(^6\).

Interestingly, nearly all reported bio-conjugation CuAAC protocols are over-stoichiometric rather than catalytic in copper as for chemical synthesis\(^4\) and until today most reported bio-conjugation CuAAC are based on the same strategy with some variation in the ligands applied\(^6\). Described methods for CuAAC use excess of a copper (II) salt, which is then reduced to copper (I) by an even larger excess of sodium ascorbate, which prevents oxidation to copper (II) during the reaction in presence of air. Copper (I) is further stabilized by a multidentate ligand forming the active catalyst and the CuAAC reaction connects an azide-alkyne pair in a 2+3 cycloaddition reaction leading to triazole forming\(^4\). The excess of reagents goes along with major limitations for the efficient applications in bio-conjugation reactions. Particularly problematic is commonly observed oxidative damage to the biomolecules caused by the hydrogen peroxide or hydroxyl and superoxide radical species formed during CuAAC reaction\(^7\). Moreover, sodium ascorbate may lead to covalent modification and potential aggregation of proteins as dehydroascorbate and other ascorbate byproducts can react with amine (lysine) and guanidine (arginine) functionalities\(^7\). Ligands such as THPTA were reported to be toxicity for living cells limiting \textit{in vitro} CuAAC reaction\(^8\). To overcome this drawback, Finn recently published a low copper catalytic protocols using a chelating azide pyridine ligand\(^9\). However, this protocol is limited to alkyne tagged proteins. It still in requires a homogeneously dissolved copper salt the use of Na-ascorbate, thus side reactions are
reduced but not absent. In summary, major drawbacks of homogeneous CuAAC protocols are the formation of active radicals, toxic homogeneous Cu contaminates, tedious catalyst separation the concomitant impossibility to recover the copper catalyst for reuse. Therefore, homogenous CuAAC reactions still have limits in their application for the labeling of biomolecules. During our previous study on the synthesis of artificial metalloenzymes using CuAAC reactions (Chapter 5), we also observed that standard CuAAC labeling results in partial precipitation of an otherwise highly stable protein and deactivation of some others, particularly under prolonged reaction times of over 4 h.

We reasoned that a high surface nanostructured material could be utilized as a heterogeneous catalytic systems and thus overcome some of the limitations described above. A recent mechanistic study by of Scaiano et al. confirms that CuAAC can be catalyzed on Cu(I)-containing surfaces. This encourages us to test a Copper(I) nanocatalyst as heterogeneous catalyst for CuAAC reactions for bioconjugation.

Herein, we report for the first time, a heterogeneous CuAAC protocol for click labeling of biomolecules using a heterogeneous recyclable 5 nm thin Cu$_2$O layers supported on Cu(0) nanowires (Cu$_2$O/Cu-NW) as heterogeneous catalyst as catalyst. It is well established that unsupported metallic nanoparticles (NPs) usually undergo aggregation and suffer from deactivation and loss of reusability. The special architecture of our catalyst circumvents this problem. The Cu$_2$O layer is negatively charged at the neutral reaction pH leading to coulombic repulsion of the nanoparticles and therefore suppresses aggregation tendencies. It possesses Cu(I) centers in a near-ideal of 0.32 nm and thus can act in a activation mechanism similar to the on recently established by Fokin and co-workers further is insoluble in water and leaching occurs only in the sub-ppm range. The very high surface area of the nanostructured material leads accessibility and reaction rate while the Cu(0) core stabilizes the structure and serves as internal reducing agent to regenerate the catalytic Cu(I) species after oxidation by oxygen. Expensive and or toxic ligands are further not required.
6.3 Result and Discussion:
A collaborator had established the catalytic activity of the well characterize Cu$_2$O/Cu-NW catalyst for in CuAAC under typical organic synthesis conditions. It was therefore of interest, if this catalyst would be active in an aqueous bioconjugation protocol. Initial activity screens used a fluorogenic coumarin azide assay developed by Wang et al.$^{15}$ (Scheme 6.1). The heterogeneous protocol was found to provide good activity in various aqueous buffers, an important finding, which suggested that a move to larger biomolecules like proteins should be feasible.

![Scheme 6.1](image)

**Scheme 6.1.** Fluorogenic CuAAC for optimization

6.3.1 Optimization of The Heterogeneous Cu$_2$O/Cu-NW Catalyzed CuAAC Protocol for Bioconjugation

Optimization studies were carried out on the surface of the fluorescent protein mTFP* with $p$AzF in position 188 (mTFP*188$p$AzF). Conversion was followed by LC-MS (Table 6.1). In most of the reported homogenous CuAAC protocol, an excess of copper precursor CuSO$_4$, Na-Ascorbate and THPTA or (BimC$_4$A)$_3$ ligand is used. Therefore, to compare the heterogeneous protocol CuAAC with the standard homogenous method we first applied an excess of propargyl alcohol to the azide incorporated protein mTFP*188$p$AzF under the homogenous CuAAC reaction conditions$^4$ using most active ligands (BimC$_4$A)$_3$.$^{16}$ As expected, a reaction with homogeneous CuAAC protocol, which has an excess of CuSO$_4$ (10 mM) resulting within 1 h, in 99% conversion of the soluble protein (Table 6.1, entry 1) with a small amount of protein precipitate. To test if the homogeneous protocol indeed requires overstoichometric amounts of reagents, we repeated the reaction with catalytic amount of copper (5 mol%) and a catalytic amount of...
additives. Surprisingly this reaction failed to produce any detectable amount of cycloaddition product, even after 24 hours (Table 6.1, entry 2). These results confirmed that homogeneous protocols for the labeling of biomolecules are not catalytic at the concentrations typically used in bio-conjugation reactions, which is possible a result of the second order influence of the Cu(I)-concentration. On the other hand, the same reaction with Cu₂O/Cu-NW was completed within one hour in the presence of ligands such as THPTA and (BimC₄A)₃ as well as without ligands (Table 6.1, entries 3-5). Copper leaching was less than one ppm, a concentration, at which no conversion is observed in the homogeneous protocol. This observation confirms the heterogeneous nature of protocol and that no requirement of ligand or any other additives were needed for the outcome of reaction. Since the surface shows a fixed spacing of Cu(I) centers of 0.35 nm, it shows a pre-orientation similar to the geometry of the recently published transition state¹³. On the surface the reaction becomes 1ˢᵗ order in Cu-Cu units, and hence slows down less at low concentrations.

Figure 6.1. (BimC₄A)₃ ligand¹⁶
Table 6.1. Optimization of CuAAC reaction using Cu$_2$O/Cu-NW catalyst on protein mTFP*188pAzF

<table>
<thead>
<tr>
<th>Entry</th>
<th>mTFP*188pAzF (mM)</th>
<th>Catalyst</th>
<th>Ligand</th>
<th>Na-ascorbate (mM)</th>
<th>Conv. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>CuSO$_4$ (10 mM)</td>
<td>(BimC$_4$A)$_3$(200 mM)</td>
<td>500</td>
<td>&gt;99</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>CuSO$_4$ (0.05 mM)</td>
<td>(BimC$_4$A)$_3$(0.10 mM)</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Cu$_2$O/Cu nano (3 mol%)</td>
<td>(BimC$_4$A)$_3$(6 mol%)</td>
<td>----</td>
<td>&gt;99</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Cu$_2$O/Cu nano (3 mol%)</td>
<td>THPTA (6 mol%)</td>
<td>----</td>
<td>&gt;99</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Cu$_2$O/Cu nano (3 mol%)</td>
<td>-------</td>
<td>----</td>
<td>&gt;99</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>Cu$_2$O/Cu nano (3 mol%)</td>
<td>-------</td>
<td>----</td>
<td>&gt;99</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>Cu$_2$O/Cu nano (30 mol%)</td>
<td>-------</td>
<td>----</td>
<td>&gt;99</td>
</tr>
</tbody>
</table>

**Reaction conditions:** a) mTFP*188pAzF (1 mM), Propargyl alcohol (100 equiv), Copper catalyst, ligand, Na-ascorbate, 30°C, 4 h. b) LC-MS conversion.

Next, to lower the protein concentration further, we conducted a CuAAC reaction of propargyl alcohol with 0.5 mM and 0.05 mM mTFP*188pAzF protein without any additive and still achieved >99% conversion within 4h without any detectable precipitation (Table 6.1 entries, 6 and 7). This result confirms that heterogeneous Cu$_2$O/Cu-NW can catalyze CuAAC at very low protein concentrations. While catalysis by the Cu$_2$O/Cu-NW resulted in very clean conjugation products, homogeneous protocols with an excess of CuSO$_4$ generally resulted in the partial precipitation of protein after 4 hrs. along with formation of some small amount of side products (Table 6.1 entry, 1).

An additional benefit of our heterogeneous protocol is the possibility to separate the Cu$_2$O/Cu-NW after the reaction by straight-forward micro-filtration. (Table 6.1 entries, 5-
Since various biological buffers like PBS contain up to 300 mM of NaCl, we tested the influence of chloride ions on outcome of Cu$_2$O/Cu-NW catalyzed bio-conjugations (Figure 6.2 down). It is well known, that chloride ions may act as ligands for Cu(I)$^{17}$, and hence might lead to both, leaching and catalyst deactivation. For this study we kept the mTFP*188pAzF concentration at 0.05 mM and used our optimized reaction conditions but switched to a biotinylated alkyne ligand (Figure 6.2 up).

**Figure 6.2.** optimization of mTFP*188 pAzF (0.05 mM) conjugations using the heterogenous Cu$_2$O/Cu-NW catalyst.

**Reaction conditions:** a) mTFP*188 pAzF (0.05 mM), propargyl alcohol (100 equiv), Cu$_2$O/Cu (30 mol%), 30oC. Conversions based on LC-MS analysis.

The Cu$_2$O/Cu-NW catalyzed CuAACreaction in buffer without salt and ligand was completed in four hours (Figure 6.2). Addition of ligand only slightly accelerated the
reaction. However, after addition of 100 mM sodium chloride conversion was completed within two hours and addition of ligand had no effect on the reaction rate. This observation confirms that addition of salt or ligands may enhance the leaching of the Cu$_2$O/Cu-NW catalyst and may lead to homogenous copper species. Also after the reaction, precipitation of some of the protein was observed therefore the use of such additives to buffers should be avoided. Based on our optimization results for bioconjugation using Cu$_2$O/Cu-NW catalyst optimized conditions are: azide-tagged protein (0.05 mM), alkyne (5 mM), Cu$_2$O/Cu-NW (30 mol% based on Cu), 50 mM HEPES-buffer (pH= 7.4), 30 C, 4 h.
6.3.2 Application of the Heterogeneous Cu$_2$O/Cu-NW Catalyzed CuAAC Protocol for The Labeling of Biomolecules

To check the scope of developed Cu$_2$O/Cu-NW catalyzed CuAAC reaction we selected fluorescent proteins mTFP*188UAA, mTFP*204UAA and non-fluorescent protein FtnA79UAA. Incorporation of azide and alkyne amino acids into these proteins was done by the method published by Schultz and co-workers$^{18}$ (supporting information and chapter number 4). The results of the conversion with various alkyne and azide dyes (Figure 6.3) are presented in Table 6.2. all six proteins are stable under the optimized reaction conditions with added Cu$_2$O/Cu-NW catalyst since no precipitation, loss in fluorescence were observed and also LC-MS data show no changes in mass (Table 6.2, entries 1, 6, 8, 13, 16 and 21). This once more confirms the biocompatibility of the, Cu$_2$O/Cu-NW catalyst. Data for the azide tagged protein pAzF-mTFP*188 reveal complete conversion after four hrs. for all dye-alkynes except for sulforhodamine-101 (3), which gives only 65% of labeled protein (Table 6.2, entries 2 to 5). An azide group, which is incorporated 204 mTFP* is much less accessible than the one in position 188. Accordingly, conversions are slower for mTFP*Y204pAzF (Table 6.2, entry 5 vs. entry 7) and a reaction time of 8 hrs. is required to achieve complete labeling. Alkyne-tagged proteins generally react sluggish, resulting in 4 hrs. labeling efficiencies between 20 and 48% for the mTFP*UAA mutants (Table 6.2, entries 9-12, entries 14 and 15). This is in agreement with the recently established mechanism$^{11}$, which assumes that the alkyne activation is rate determining and hence the alkyne should be used in excess. Labeling studies with the non-fluorescent protein FtnA79UAA gave similar results, with over 95% labeling efficiencies for FtnA79pAzF (Table 6.2 entries 17-20) and lower conversion for FtnA79pEynF of 33% - 65% (Table 6.2 entries 22-25). In all cases, no protein precipitation was observed and the separation of the catalyst at the end of reaction was done by micro-filtration.
Figure 6.3. 1 = Cy3-Alkyne, 2 = Cy5-Alkyne, 3=Sulforhodamine-101-Alkyne, 4= Biotin-PEG4-Alkyne, 5= Cy3-Azide, 6= Cy6-Azide, 7=Sulforhodamine-101-Azide, 8=Biotin-PEF4-Azide.
Table 6.2. Summary of result for the Cu$_2$O/Cu-NW at Cu nanowires catalyzed Click bioconjugation of fluorescent (mTFP*Y188pAzF, mTFP*Y204pAzF, mTFP*Y204pEynF and mTFP*Y204pEynF) and non-fluorescent (FtnA 79 pAzF and FtnA 79 pEynF) proteins with the corresponding azide and alkyne ligand 1-8.\(^a\)

<table>
<thead>
<tr>
<th>Entry</th>
<th>Protein</th>
<th>Ligand</th>
<th>Mass [Da]</th>
<th>Conv.(^a) %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Calculated</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>mTFP*Y188pAzF</td>
<td>non</td>
<td>25179.4</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>25942.0</td>
<td>65</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2</td>
<td>25969.0</td>
<td>88</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td>26001.0</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>4</td>
<td>25637.6</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>mTFP*Y204pAzF</td>
<td>non</td>
<td>25179.4</td>
<td>70</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>4</td>
<td>25637.6</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>mTFP*Y188pEynF</td>
<td>non</td>
<td>25164.0</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>5</td>
<td>25971.0</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>6</td>
<td>25998.0</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>7</td>
<td>25972.0</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>8</td>
<td>25608.6</td>
<td>30</td>
</tr>
<tr>
<td>13</td>
<td>mTFP*Y204pEynF</td>
<td>non</td>
<td>25164.0</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>5</td>
<td>25971.0</td>
<td>45</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>8</td>
<td>25607.5</td>
<td>30</td>
</tr>
<tr>
<td>16</td>
<td>FtnA79pAzF</td>
<td>non</td>
<td>19480.7</td>
<td>30</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>1</td>
<td>20242.6</td>
<td>99</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>2</td>
<td>20269.6</td>
<td>99</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>3</td>
<td>20301.6</td>
<td>95</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>4</td>
<td>19938.3</td>
<td>95</td>
</tr>
<tr>
<td>21</td>
<td>FtnA79pEynF</td>
<td>non</td>
<td>19466.7</td>
<td>99</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>5</td>
<td>20309.7</td>
<td>50</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>6</td>
<td>20273.6</td>
<td>45</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>7</td>
<td>20288.7</td>
<td>33</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>8</td>
<td>19911.2</td>
<td>62</td>
</tr>
</tbody>
</table>

\(^a\)`Reaction conditions:` Protein with azide or alkyne group (0.05 mM, 0.5 mL, HEPES buffer with pH 7.4, Alkyne or alkyne dye (5 mM), Cu-NW (30 mol%), 30°C, 4 h. \(^b\)`Enantiomeric excesses were determined by chiral HPLC."
6.4 Conclusion

The core-shell structured Cu₂O/Cu-nanowire catalyst described in this chapter is well suited for bioconjugation applications. It provides several benefits over traditional homogeneous systems, which are summarized below:

1. We have developed a heterogeneous CuAAC bio-conjugation protocol using a nano-structured Cu₂O/Cu-nanowire catalyst.
2. This protocol does not require the addition of any toxic ligands or reducing agents like Na-ascorbate, which have adverse effect on biomolecules.
3. The reaction works in aqueous buffers, no need of toxic co-solvents as used in traditional protocols.
4. Catalyst loadings can be low compared to conventional protocols and leaching is limited to less than 1 ppm Cu.
5. Separation of the catalyst from the product is very easy (filtration) compared to traditional homogeneous protocols.
6. Bio-conjugate products are very pure and high yield.
6.5 Material and Methods

6.5.1 Unnatural Amino Acid Synthesis and Incorporation

\textit{para-Azido-(L)-phenylalanine (pAzF)} and \textit{para-(L)-Ethenylphenylalanine (pEnyF)}

This Part has been described in details in Chapter 4. I.

6.5.2 Synthesis of Cu/Cu$_2$O nanowires catalyst$^{14}$

A well characterized \textit{Cu/Cu$_2$O} nanowires catalyst was obtained from Prof. Ziping Lai’s research group and used as it is for the all reactions.

6.5.3 Clickable Labeling Reagents

All clickable azide or alkyne ligands listed in Figure 6.3, were purchased from’ click chemistry tools’ from Bioconjugate Technology Company, and were used without further purification.

6.5.4 Labeling of Purified Proteins by Click Reaction

The corresponding azide or alkyne incorporated protein mTFP* or FtnA79 protein sample was concentrated to the concentration for catalytic reaction 0.05mM. The CuAAC reaction on the proteins with heterogeneous Cu$_2$O/Cu-NW catalyst was performed in a glove box under nitrogen atmosphere in a glass vial (2 mL) equipped with appropriate size magnetic stir bar and was charged with copper-nano catalyst (30 mol%), followed by addition of corresponding azide or alkyne incorporated protein mTFP* or FtnA79 protein (0.5 mL, 0.05mM in HEPES buffer pH=7.4), next corresponding azide or alkyne ligand is added (Figure 6.3, 100 equiv). Then the glass vial with reaction mixture was closed with rubber septum and kept for stirring at at 30 °C temperature for 4 h. After completion of reaction, the copper nanowires are simply separated using syringe with nano-filter. All samples have been measured by LC/MS analysis after performing the click to verify proteins modification with corresponding ligand.
For final analysis to make sure that all samples are free of heterogeneous impurities (copper nanowires), the protein samples were dialyzed against a more or less salt free buffer (50mM HEPES pH 7.9 or 100mM Tris/HCl pH 7.4). Dialysis should at least last for 4 hours. After dialysis, samples were prepared for LC/MS analysis. The 5µl of the protein sample was mixed with 95µl of Acetonitrile Buffer then Spin down at 10000xg for 20min afterword 20-50 µl has been taken of the supernatant and transfer them to a sample tube for LC/MS.

Conditions for working with Cu₂O/Cu NW catalyst for CuAAC reactions:

1. Use of NaCl or any other corrosive salts should be avoided as it reacts with copper.
2. Use of trisbuffer, which form complex with copper, should be avoided as it binds to the catalytic site.
3. Reaction should be performed in inert atmosphere.
4. Stirring technique has influence on rate of reaction proper stirring is necessary.

6.5.5 Cloning

The gene of mTFP*[citation mTFP* paper] (monomeric Teal Fluorescent Protein) was engineered in our lab. The mTFP* gene cloned into the ChampionTM pET303/CT-His vector. This vector is supplemented with a N-terminal SUMO-Tag. mTFP* gene was digested by BAMHI and XhoI restriction enzymes (New England Biolabs Inc.) and subsequently fused to a N-terminal SUMO Tag to generate the construct pET303/SUMO_mTFP*. The engineered vector has mTFP* linked directly to SUMO-Tag. FtnA gene was fused to the same vector to generate pET303/SUMO_FtnA. Post expressionial cleavage of tag allows expression of the proteins in a native form with no residual linker.

6.5.6 Site-Directed Mutagenesis

Site-directed mutagenesis was performed on pET303/SUMO_mTFP* and pET303/SUMO_FtnA. Detailed of all Mutants nucleotide sequence and Primers is shown in the Supplementary section. Mutant variants of mTFP*TAG (TAG, mTFP*Y188 TAG, mTFP*Y204 TAG and FtnA E97 TAG) were generated via QuickChange® Site-Directed
Mutagenesis Kit II (StratageneTM, Agilent Technologies). All primers were designed according to the user guidelines of the QuikChange Kit and subsequently obtained from Eurofins MWG Operon. Mutant plasmids were transformed into E. Coli BL21 (DE3) Gold (StratageneTM, Agilent Technologies) and purified using peqGOLD Plasmid MiniPrep Kit II (peqlabBiotechnologie GmbH). Purities and successful mutagenesis were checked by 1% Agarose Electrophoresis and Sanger Sequencing (KAUST Core Facilities). For Uaa site incorporation the resulting plasmids were Co-transformed with respective pEvol-aaRS plasmids (from Peter G. Schultz) (pEVOLpAzF and pEVOLpCNF).

6.5.7 Expression

Pre-cultures that contain 180mg L⁻¹ampicillin and 25mg L⁻¹ were grown overnight at 37 °C and were further used to inoculate expression cultures under standard protocol. For protein production the LB Medium containing 180mg L⁻¹ampicillin and 25mg L⁻¹ Chloramphenicol was inoculated with E. Coli BL21 (DE3) Gold cells, (Agilent Technologies), that had been transformed with pET303/SUMO_mTFP* and pEVOLpAzF or pEVOLpCNF at 37 °C. The media conc. started at OD 0.05 at wavelength 600nm, and incubate overnight at 20 °C, induction by 1mM IPTG and 0.2% of L-arabinose followed by 1mM of corresponding unnatural amino acid was applied when the OD reached 0.6-0.7. Cultures were then centrifuged at 4500 rpm for 30 minutes and cooled to 4 °C. Cell pellets were washed with Lysis Buffer A, containing 100 mM Tris/HCl pH 7.4, 500 mM NaCl and 20mM Imidazole. Before lysis complete protease Inhibitor cocktail (Roche), was added together with DNase. Lysis occurred using French press (Constant Systems Ltd) at 1.7KBar. Lysates were immediately centrifuged at 25000 rpm for 30min and also cooled to 4 °C.

6.5.8 Purification

All proteins were purified by using Ni-Affinity Chromatographic principles. The respective supernatant was transferred to a pre-washed nickel column and eluted from the
column between 20 and 35% of Elution Buffer B containing 100 mMTris/HCl pH 7.4, 500 mMNaCl and 500 mM Imidazole. The eluted protein was digested with SUMO protease overnight in the shaker at 30 °C and the mixture then dialysed overnight, at 4 °C, against Buffer A. Further Ni-Affinity Chromatography followed to remove the SUMO-Tag. Elution occurred with a smooth gradient up 20% of the same Buffer containing additional 1 M NaCl. Pure Protein was collected from the flow-through before elution with Buffer B which showed a distinct elution of residual SUMO-Tag and Protease. The purity of mTFP* was analyzed by SDS–PAGE gel and its fluorescence was measured at excitation at 462 nm and an emission at 492 nm. All incorporation confirmation was done using LC/MS to detect the mass differences before (wt protein) and after the incorporation (modified protein).

6.5.9 LC/MS Analysis

To acquire mass spectra of the intact proteins, the purified proteins were dialyzed against (50 mMTris/HCl pH 7.4) and concentrated to 8-10 mg/mL. The protein sample was diluted in buffer containing 5% ACN, 0.1% formic acid, and analyzed using maXis UHR-TOF mass spectrometer (Bruker) coupled to (Agilent1260) HPLC system. Briefly, 2µL of the sample was loaded onto a C4 column (Advantage 300+ C4, 5µm particles, 50 x 2.1 mm, Analytical sales), and desalted using a divert valve at 500µl/min flow rate for 15 min. The proteins were eluted at 200µl/min flow rate using a linear gradient of 5-80% Acetonitrile (in 0.1% formic acid and H₂O) for 10 min, followed by isocratic gradient of 80% for 5 min, and then isocratic gradient at 5% for 5 min. MS analysis was carried out using (Bruker) ESI source at 4200 v potential, 1.4 bar nebulizer gas and 8.0 l/min dry gas at a temperature of 200°C. The MS data was collected in positive mode, and m/z of ions was determined at rolling average of 2. The peak list was extracted using Compass Data Analysis software ver 4.0 (Bruker) at signal to noise ratio of 5. The mass of protein was deconvoluted where minimum 5 peaks for the specific large molecule were identified.
6.6 References


14. Li H. Development of Copper-Catalyzed Electrophilic Trifluoromethylation and Exploiting Cu/Cu2O Nanowires with Novel Catalytic Reactivity [ PhD Dissertation]: King Abdullah University of Science and Technology Thuwal, Kingdom of Saudi Arabia; 2014.


6.7

Supplementary Material

Molecular Biology
mTFP*Y188TAG
GGCGTAATCAAGCCCGACATGAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCTATGCCTTCGTG
ATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACCAACACCATCAACCTGGAGGTGAAGGAGGG
AGCCCCCCTGCCCTTCTCCTACGACATTCTGACCACCGCGTTCGCCTACGGCAACAGGGCCTTCACCAAG
TACCCCGACGACATCCCCAACTACTTCAAGCAGTCCTTCCCCGAGGGCTACTCTTGGGAGCGCACCATGA
CCTTCGAGGACAAGGGCATCGTGAAGGTGAAGTCCGACATCTCCCTGGAGGAGGACTCCTTCATCTACGA
GATATATCTCAAGGGCGAGAACTTCCCCCCCAACGGCCCCGTGATGCAGAAGAAGACCACCGGCTGGGA
CGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGACGTCAAGCACAAGCTGCTGCT
GGAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGGGCCAAGAAGGCGGTGAAGCTGCC
CGACTAGCACTTTGTGGACCACCGCATCGAGATCCTGAACTATGACAAGGACTACAACAAGGTGACCGTT
TACGAGAGCGCCGTGGCCCGCAACTCCACCGACTAA

mTFP*Y204TAG
GGCGTAATCAAGCCCGACATGAAGATCAAGCTGAAGATGGAGGGCAACGTGAATGGCTATGCCTTCGTG
ATCGAGGGCGAGGGCGAGGGCAAGCCCTACGACGGCACCAACACCATCAACCTGGAGGTGAAGGAGGG
AGCCCCCCTGCCCTTCTCCTACGACATTCTGACCACCGCGTTCGCCTACGGCAACAGGGCCTTCACCAAG
TACCCCGACGACATCCCCAACTACTTCAAGCAGTCCTTCCCCGAGGGCTACTCTTGGGAGCGCACCATGA
CCTTCGAGGACAAGGGCATCGTGAAGGTGAAGTCCGACATCTCCCTGGAGGAGGACTCCTTCATCTACGA
GATATATCTCAAGGGCGAGAACTTCCCCCCCAACGGCCCCGTGATGCAGAAGAAGACCACCGGCTGGGA
CGCCTCCACCGAGAGGATGTACGTGCGCGACGGCGTGCTGAAGGGCGACGTCAAGCACAAGCTGCTGCT
GGAGGGCGGCGGCTATTATCGCGTTGACTTCAAGACCATCTACAGGGCCAAGAAGGCGGTGAAGCTGCC
CGACTATCACTTTGTGGACCACCGCATCGAGATCCTGAACTATGACAAGGACTAGAACAAGGTGACCGTT
TACGAGAGCGCCGTGGCCCGCAACTCCACCGACTAA

FtnA E79TAG
ATGCTGAAACCAGAAATGATTGAAAAACTTAATGAGCAGATGAACCTGGAACTGTACTCTTCACTGCTTT
ATCAGCAAATGAGCGCCTGGTGCAGCTATCATACCTTCGAAGGTGCTGCCGCGTTCCTGCGCCGCCACGC
CCAGGAAGAGATGACACATATGCAGCGTCTGTTTGATTACCTGACTGATACCGGAAACTTACCACGTATT
AATACCGTTGAATCTCCGTTTGCTTAGTATTCTTCACTTGATGAATTATTCCAGGAAACCTATAAACACGA
ACAATTAATCACGCAGAAAATTAACGAACTGGCTCATGCTGCAATGACCAATCAGGACTACCCAACATTT
AATTTCCTGCAATGGTATGTTTCTGAGCAGCATGAAGAAGAGAAACTGTTCAAATCGATTATTGATAAAT
TAAGCCTGGCAGGCAAAAGCGGCGAAGGTCTGTATTTTATCGACAAAGAACTCTCTACCCTCGACACACA
AAACTAA

Figure S6.4. Nucleotide sequence for mTFP*188TAG, mTFP*204TAG and FtnA79TAG(Red Highlight =
mutation to stop codon TAG for pAzF and pEnyF incorporation ) .


Figure S6.5. pET303/SUMO_plasmids. A) pET303/SUMO_mTFP*204 TAG, B) pET303/SUMO_mTFP*188 TAG, C) pET303/SUMO_FtnA79TAG.
Table S6.3. PCR primers used for cloning and mutagenesis

<table>
<thead>
<tr>
<th>Protein-mutation position</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>mTFP*Y204TAG&lt;sup&gt;f&lt;/sup&gt;</td>
<td>5'-GAAATCCTGAACTACGACAAAGACTAGAACAAAAGTTACCCTG-3'</td>
</tr>
<tr>
<td>mTFP*Y204TAG&lt;sup&gt;r&lt;/sup&gt;</td>
<td>5'-ACCGTAACCTTTGTTCTAGTCTTTGTCGTAGTTCAGGATTTT-3'</td>
</tr>
<tr>
<td>mTFP*Y188TAG&lt;sup&gt;f&lt;/sup&gt;</td>
<td>5'-TGTTAAACTGCCGGACTAGCACTTGTTGACCAC-3'</td>
</tr>
<tr>
<td>mTFP*Y188TAG&lt;sup&gt;r&lt;/sup&gt;</td>
<td>5'-GTGGTCAACGAAGTTGCTAGTCCGGCAATTTAACA-3'</td>
</tr>
<tr>
<td>FtnA-97TAG&lt;sup&gt;f&lt;/sup&gt;</td>
<td>5'-ACAATTAATCAGCGAAGAAATTAAGTCGGGAACAAAGT-3'</td>
</tr>
<tr>
<td>FtnA-97TAG&lt;sup&gt;r&lt;/sup&gt;</td>
<td>5'-AGCGGCAAGTTGCTGTATTTATCGAAAGACTCTCTAC-3'</td>
</tr>
</tbody>
</table>

Figure S6.6. proteins structure. A) mTFP*Y188TAG and B) mTFP*Y204TAG (PDB:4Q9W), C) FtnA 79TAG (PDB: 3UNO).
Figure S6.7. ESI-TOF analysis of the incorporation of two unnatural amino acids \( p\text{AzF} \) and \( p\text{EynF} \) into the TAG position of mTFP*Y188 TAG, mTFP*Y204 TAG and FtnA E79 TAG. A) \( m\text{TFP}*Y188p\text{AzF} \) 25180 Da, B) \( m\text{TFP}*Y204p\text{AzF} \) 25180 Da, C) \( m\text{TFP}*Y188p\text{EynF} \) 25163 Da and D) \( m\text{TFP}*Y204p\text{EynF} \) 25163 Da. Sign (black and dashed arrows) indicate \( m\text{TFP}^\ast \text{wt} \) maturated and unmaturated 25154 and 25174 Da respectively. E) FtnA97pAzF 19479.9 Da and F) FtnA97pEynF 19465.7 Da. Sign (black arrows) indicate the FtnAwt 19425 Da. \( \text{wt} = \) is the wild type of the protein before the incorporation process.
Figure S6.8. ESI-TOF analysis of the modified mTFP*pAzF after click reaction using homogenous CuAAC protocol. A) mTFP*Y188 pAzF with ligand 1. B) mTFP*Y188 pAzF with ligand 2. C) mTFP*Y188 pAzF with ligand 3. D) mTFP*Y188 pAzF with ligand 4. E) mTFP*Y204 pAzF with ligand 4. Sign (black arrows) indicate unmodified mTFP*Y pAzF and mTFP*wt proteins (25180 and 25154 Da) respectively. (P) expected product. wt = is the wild type of the protein before the incorporation process.
Figure S6.9. ESI-TOF analysis of the modified mTFP*pEynF after click reaction using homogenous CuAAC protocol. A) mTFP*Y188 pEynF with ligand 5. B) mTFP*Y188 pEynF with ligand 6. C) mTFP*Y188 pEynF with ligand 7. D) mTFP*Y188 pEynF with ligand 8. E) mTFP*Y204 pEynF with ligand 5. F) mTFP*Y204 pEynF with ligand 8. Sign (black arrows) indicate unmodified mTFP*Y pEynF and mTFP*wt proteins (25163 and 25154 Da) respectively. (dashed arrows) indicate unmaturred mTFP* 25174 Da. (P) expected product. wt = is the wild type of the protein before the incorporation process.
Figure S6.10. ESI-TOF analysis of the modified FtnA E97pAzF after click reaction using homogenous CuAAC protocol. A) FtnA E97pAzF with ligand 1. B) mTFP*Y188 pEynF with ligand 2. C) mTFP*Y188 pEynF with ligand 3. D) mTFP*Y188 pEynF with ligand 4. Sign (black arrows) indicate unmodified FtnA E97pAzF and FtnAwt proteins (19480 and 19424 Da) respectively. (P) expected product. wt = is the wild type of the protein before the incorporation process.
Figure S6.11. ESI-TOF analysis of the modified FtnA E97pEynF after click reaction using homogenous CuAAC protocol. A) FtnA E97pEynF with ligand 5. B) FtnA E97pEynF with ligand 6. C) FtnA E97pEynF with ligand 7. D) FtnA E97pEynF with ligand 8. Sign (black arrows) indicate unmodified FtnA E97pEynF and FtnAwt proteins (19466 and 19424 Da) respectively. (P) expected product. wt = is the wild type of the protein before the incorporation process.
Chapter 7       Discussion and Conclusion

One of the most extensively utilized proteins in the important biological and biocatalysts research is the green fluorescent protein, derived from the jellyfish Aequorea Victoria (avGFP).

In the beginning of the thesis, we studied different fluorescent proteins and compared their physical, chemical, and fluorescent properties in order to select the best scaffold for the design of artificial metalloproteins. Fluorescent combine several advantages, which are beneficial for this task, e.g. high thermal and pH stability a, high solubility and expression yields and and particularly an intrinsic fluorescence, which allows for rapid quantification and detection of structural perturbations. The metalloprotein design project was expected to require many screening rounds. Hence, a rapid and parallelizable purification protocol was established, which allows the successful purification of a broad range of fluorescent proteins. The combination of thermal and three phase partitioning protocol (T-3PP) is an excellent alternative to expensive and laborious techniques developed recently. This protocol significantly removes impurities in the first thermal purification and then builds on salting out, isoionic precipitation, and co-solvent precipitation of proteins by the organic extraction. It can readily be scaled to purify small or large amounts of protein and delivers in some cases near quantitative yields and over 99% purity. Importantly, this very economic protocol does not require any tag for the purification\(^1\) and therefore gives access to pure WT proteins. This protocol meets the needs of scientific researchers as well as industrialists in providing a high degree of purity of recombinant proteins. Though the results of parallelization experiments demonstrate that this protocol is a promising process allowing automation purification and screening.

The third chapter describes the engineering of metal binding proteins based on the mTFP1 scaffold and the application of those to generate artificial metalloenzymes, which show activity in several Lewis-acid catalysed reactions. mTFP1 is a monomeric derivative of the tetrameric CFP (cyan FP), which currently ranks among the brightest and most
photostable FPs reported. Mutating seven surface-exposed His and Met residues of mTFP1 resulted in a protein with low binding affinity to transition metals, which we chose as host protein for the engineering of metal binding sites. This protein was termed mTFP*. It combines high expression, thermal stability (melting temperature 88 °C), solubility in buffers containing up to 90% organic solvents and a broad pH range. Crystallographic studies further confirmed the low affinity to transition metals. And the evaluation of all crystal structures in this study, but most importantly of Copper co-crystallized proved the ideal spacing of the projected binding pocket for our mutation approach which is corresponding to the ideal spacing discussed and reported by Reetz and Ward. The establishment of a library of metalloproteins prepared by incorporation of artificial metal binding motifs followed by the treatment with low concentrations of transition metal salts led to the formation of artificial metalloenzymes that showed moderate to good asymmetric induction in Diels-Alder cyclizations and Friedel-Crafts-Acylations. Generally, this work strongly build a design principles targeting dative Anchoring.

The chapter 4 describes, the successful incorporation of the unnatural amino acids para-azido-(L)-phenylalanine (pAzF), para-ethynyl-(L)-phenylalanine (pEynF), para-iodophenylalanine (pIF), and para-boronophenylalanine (pBF) using an orthogonal tRNA/synthetase pair approach developed by P.G. Schultz and co-workers. The high expression levels and stability of the beta-barrel protein make mTFP* a suitable host for the incorporation of UAAAs. Co-transfection of plasmid harbouring a mTFP* gene with a TAG-codon together with a pEVo1 plasmid containing the orthogonal tRNA/synthetase pair for UAA incorporation allows the high-yielding expression of mTFP* mutants with site-specific incorporated UAAAs. Computational studies and homology modelling were used to identify the best positions in mTFP* for site-specific UAA incorporation. Asn128, Tyr188, Tyr200, and Tyr204 are particularly suitable for sites, since they are at the protein surface and seem to tolerate a broad variety of UAAAs without disruption of the protein structure. All four sites are surface exposed, however Tyr200, and Tyr204 are in a cavity and therefore less accessible, particularly for conjugations with larger ligands. The best incorporation efficiency varies with the nature of the UAA: pIF incorporation prefers position 188 (expression yield 350 mgL⁻¹), whereas pAzF is best incorporated in
position Tyr200 (expression yield 280 mgL⁻¹) and pEynF is incorporated with up to 195 mgL⁻¹ in position 204. Incorporation of the para-boronophenylalanine (pBF) generally suffers from low yields (< 20 mgL⁻¹). We approached this problem through an on-protein borylation of pIF. A MIYURA borylation protocol was successfully established for mTFP* using a palladycyclic catalyst. This simple strategy allows the chemical modification of various biomolecules in aqueous buffers. And it broadens the application of different halide containing building blocks to be coupled to mTFP*pBF for subsequent SUZUKI cross coupling. Eppinger and co-workers recently developed an aqueous protocol for Pd-catalyzed cross-coupling reactions which can be used for this application⁴⁻⁶.

pAzF and pEynF were incorporated into mTFP* and acted as the anchoring point for bioconjugation reactions through CuAAC.

Chapter 5 demonstrates the efficient biosynthesis of proteins with site-specific incorporation of UAAs via the orthogonal pair approach. Incorporation of azide (pAzF) and alkyne (pEynF) containing UAAs facilitate conversion through the CuAAC click to introduce ligands that provide a high affinity to transition metals (metal-chelating ligands). The required BTPs ligands were successfully synthesized and fully characterized and further coupled to the proteins. We showed that the genetically encoded clickable UAA pAzF can bind be converted with various ligands, which easily bind to Cu²⁺ ions with high affinity. Since the metal binding modifications route consists of only one step, without organic reagents or column purification, metals ions soacks can be easily prepared in any laboratory and mutant proteins bearing pAzF at any site can be easily obtained and purified in milligram quantities by site-directed mutagenesis and recombinant gene expression. The UAA pAzF was recently synthesized in large quantity in our lab by a newly developed protocol, it can also serve as an anchoring point to introduce ligands via the CuAAC click reaction, which subsequently bind to transition metal, Therefore, we believe that the technological barriers that have limited the application of genetic incorporation of UAAs bearing transition metal ions directly or through ligands have now been overcome. This new method may find applications in protein engineering, metalloprotein design and protein NMR spectroscopy using paramagnetic metal ions. While the resulting ArMs show some asymmetric induction in a
Diels-Alder cycloaddition, the selectivity still require optimization to be applicable.

The improvement of the low ee using the single a attachment approach was improved by the multiple attachment of the introduced ligand forming metalloprotein with a more precise relative orientation of metal and protein surface. We established a method for the selective dual-site on-protein synthesis of BTP ligands through bioconjugation to UAAs and used this method to synthesize artificial metalloenzymes. The two UAAs mutant sites are strategically placed on the protein surface in order to be close enough to react be converted in a CuAAC, linking surface-exposed UAAs on the same protein. This technique is analogous to “peptide stapling”, which is currently only achievable through UAA introduction in solid-phase peptide synthesis. This work is unique because it co-incorporates two UAAs in vivo for the purpose of forming a “staple” that will endow the protein with desirable thermostable properties. From these results, we hypothesized that an improved catalyst could be engineered using a dual anchoring strategy, which allows a precise control of the placement of the artificial metallo-complex with specific orientation and limited rotational freedom. The mTFP*s dual anchoring pAzF mutants were successfully modified by coupling dialkyne ligand (2,6-diethynyl-3,5 dimethoxypyridine 3) in near quantitative yield and high purity as indicated by ESI-MS data. Subsequently, metalloproteins were formed with Cu$^{2+}$ and Pd$^{2+}$ salts. The formation of metalloprotein complexes was validated by ESI-MS UV-Vis spectroscopy. All different mTFP*-based metalloproteins were evaluated in Diels-Alder and Tsuji-Trost reactions. All metalloproteins gave good and reproducible conversions yielding the expected product in both reactions. As anticipated, the enantiomeric excess improved over the single anchoring strategy. Hence, the dual covalent anchoring strategy of the metal complex to the protein effects catalyst with strongly improved chiral induction compared to the single-point covalent attachment strategies. Higher enantioselectivities were observed for mutants that place the transitional metals located in the cavity position, which should provide better interaction of the substrate and the protein surface and result in a more defined metal/substrate orientation. Word and co-worker reported a double-anchoring method to complex biotinylated rhodium within streptavidin which significantly improves the catalytic activity and selectivity in the transfer hydrogenation of imines. Here we established conjugation method can be applied to the coupling of any
metal complexes, by using the CuAAC click reaction. Thus mTFP*(pAzF)₂ can provide a platform to create artificial metalloproteins for with a plethora of transition metals for a diverse set of catalytic reactions which exhibitions significantly improved catalytic performance in both activity and selectivity. Generally, to achieve higher ee's it will be necessary to purify the protein mixture, which may still contain non-cyclized triazolopyridyl moieties; yet, in our study we did not find any indication of these. Second, catalytic reaction conditions should be optimized (substrate, buffer, solvent, and pH). Finally, protein precipitation may be reduced with the heterogeneous CuAAC protocol described below.

A protocol for this click reaction was established using a Cu₂O/Cu-nanowire catalyst design in the Huang lab. In this regard, the advances in nanotechnology offer promising opportunities for the design of more effective, efficient, recyclable, and sustainable nanostructured heterogeneous (insoluble) catalytic systems. This studies shows, that heterogenous nano-structured catalysts can minimize the damage to biomolecules and may advance the field of CuAAC applications in bioconjugation by minimizing other drawbacks of the homogeneous catalyst system⁹. A recently reported mechanistic study by Decan et al. confirms the hereterogenous nature of CuAAC on using solid Cu-sources¹⁰. It shows that the reaction occurs at the copper nanopartical surface and not at ions leaching from the nanoparticle. This encouraged us to examine the Cu₂O/Cu-nanowire catalyst in bioconjugation.

This catalyst gives access to a heterogeneous CuAAC protocol for click labeling of biomolecules using a recyclable core-shell structured nanoparticle. The catalysts architecture prevents side effects commonly associated with nanoparticles, e.g. aggregation of unsupported metal NPs or poisoning under reaction conditions, which results in deactivation and prevents catalyst recycling. The catalyst used here has a special architecture: a Cu(0) nano-wires core supports a thin Cu₂O layer of 5-10 nm with a high surface area. Our experimental results confirm that the Cu₂O layer is virtually insoluble in water. Cu leaching is limited to less than 1 ppm copper during catalysis. Application of core-shell structured the Cu₂O/Cu-nanowire catalyst avoids the use of toxic ligands and Na-ascorbate and thus indirectly avoids the formation of peroxide and
radicals using anaerobic conditions. This also avoids the toxic effects of ligands and Na-
ascorbate. To the best of our knowledge, this is the first example of the use of a
heterogeneous, nanostructured copper catalyst used in CuAAC for bioconjugation.

The work described in this thesis combines advantages of the tools developed in the last
decade including UAA incorporation and click chemistry. These tools are invaluable in
this work and in the expanding field chemical biology as a whole. It opens up new
avenues for the generation of artificial metalloenzymes and their application as selective
catalysts, which may become valuable tools for organic synthesis.
7.1 Referanse

PUBLICATIONS

Manuscripts


- Johannes Fischer, Felix Quitterer, Arwa A. Makki, Anand Radhakrishnan, Meina Liu, Seema Ghorpade, S. AbdulRajjak, Michael Groll, Jörg Eppinger, Development of a robust and versatile Host Protein for Design and Evaluation of Artificial Metal Centers.

- Arwa A. Makki, Dinesh N. Sawant, Seema Ghorpade, Huaifeng Li, Kuo-Wei Huang, Jörg Eppinger, Heterogeneous Protocol For Click Labeling Of Biomolecules Using A Recyclable Cu/Cu2o Nano Wire-Catalyst.

- Arwa A. Makki, Dinesh N. Sawant, Seema Ghorpade, Jörg Eppinger, A Site-Selective Anchoring Strategy Towards Fluorescent Artificial Metalloprotein.
Conference Contributions

- Arwa A. Makki, Dinesh Sawant, Anna Zernickel, Johannes Fischer, Kuo-Wei Huang, Jörg Eppinger, ICCC-41, Singapore, Singapore, July 21st, 2014, Poster: Highly Efficient Protocol for Click Labeling of Biomolecules Using a Recyclable Cu$_2$O nanowire-catalyst

- Jörg Eppinger, Arwa Makki, Johannes Fischer, Anna Zernickel, 12$^{th}$ EUROBIC, Zürich, Switzerland, August 24$^{th}$ – 28$^{th}$ 2014, Lecture: Expanding Nature’s Toolbox with Artificial Metalloenzymes.


- Jörg Eppinger, Johannes Fischer, Susan Lauw, Arwa Makki, Paul Harris, 245th ACS National Meeting, New Orleans, USA, April 6-11, 2013. Lecture: Catalytically active organometallic enzyme hybrids.