

December 2018

Adapting Proteins for Clinical and Industrial Use

Michelle Takacs
Syracuse University

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



Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Takacs, Michelle, "Adapting Proteins for Clinical and Industrial Use" (2018). *Dissertations - ALL*. 975.
<https://surface.syr.edu/etd/975>

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

Abstract

Enzymes are powerful tools that are capable of catalyzing reactions with high specificity and efficiency. Many naturally occurring enzymes are used as tools in industry and medicine, from food additives to pharmaceuticals to biofuel production. Protein engineering is used to make enzymes more readily available, to gain new or improved catalytic function, or optimize properties such as thermostability and enantioselectivity. In this thesis we will discuss the variety of ways proteins can be modified to one day be able to be put to use in industrial and clinical settings.

My goal has been to optimize proteins for industrial and clinical use by either a) understanding how the protein evolved so that the knowledge can be applied to other protein designs, b) finding new ways to express proteins so that they can be isolated and used more easily in industrial and clinical settings, and c) creating new proteins that can be used for industrial and clinical settings. In Chapter 2, through protein engineering, a new protein was made that selectively binds lanthanides over its natural metal, calcium, with high affinity. This provides us with important information into how metalloproteins evolve metal specificity, which can be applied to designing new metalloproteins. In Chapter 3, we investigate a possible application of this protein in therapeutics by attaching an antibody to the protein structure that is specific for Hepatocellular Carcinoma cells and using the protein to bind and deliver radioisotopes. This would allow us to not only create a new treatment option for this variety of cancer, but also allow us to optimize the treatment more effectively and easily through protein evolution. In Chapter 4, we discuss how changing the expression vector of an industrially relevant enzyme, formate

dehydrogenase, allows it to be synthesized and secreted by *Pichia pastoris* and can be an important first step in creating a self-sustaining and environmentally safe method of producing methanol biofuel. Chapter 5 focuses on investigating a new method of evolving designed proteins that uses NMR by analyzing known mutations in an extensively evolved protein. This would one day allow potential enzymes to be optimized readily and cheaply for commercial use. Finally, Chapter 6 will focus on using amino acids as an alternative to traditional chiral ligands in organometallic synthesis. These new ligands would be much cheaper to make, easier to produce, and have the added benefit of forming fibrils that would make them easy to collect and attach to surfaces. These studies contribute to a better understanding of the variety of methods proteins can be optimized for use in our daily lives.

ADAPTING PROTEINS FOR CLINICAL AND INDUSTRIAL USE

By

Michelle P Takacs

B.S. California Polytechnic State University San Luis Obispo 2013

M. Phil. Syracuse University 2015

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry

Syracuse University

December 2018

Copyright © Michelle P Takacs, 2018

All Rights Reserved

Acknowledgments

First, I would like to thank my father, mother, brother, and sister-in-law for their continued support and love. Words cannot express the gratitude I feel towards you and without you all this day would not have been possible. A special thanks to my extended family of aunts, uncles, and cousins as well for their support.

Second, I would like to thank my advisor Dr. Ivan Korendovych and Dr. Olga Makhlynets. Thank you for your support over the years and teaching me how to be a better scientist.

Third, I would like to thank my friends from both graduate and undergraduate for being there for me. A special thanks to my best friend Sarah Brown, whom was always there when I needed someone to lean on. I always enjoyed our time together when I visited and I hope we will get many more opportunities in the future. I would also like to thank my friend Alexa Stathis for being an excellent roommate and confidant over the last four years of graduate school.

Last but not least, I would like to thank my lab mates for their support. I had so much fun learning and growing with you all over the years and you all have a special place in my heart. I wish the best for all of you in your future endeavors.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
1.1. ENZYMES	1
1.1.1. <i>Defined</i>	1
1.1.2. <i>Enzyme kinetics</i>	7
1.1.3. <i>Enzyme structure</i>	8
1.2. PROTEIN ENGINEERING	14
1.2.1. <i>Polymerase chain reaction and mutagenesis</i>	16
1.2.2. <i>Directed evolution</i>	21
1.2.3. <i>Rational design</i>	22
1.2.4. <i>Semi-rational design</i>	23
1.3. METHODS OF PROTEIN PRODUCTION	23
1.4. CONCLUDING REMARKS	32
CHAPTER 2 DESIGN OF A RARE EARTH SELECTIVE CALMODULIN DERIVATIVE....	41
2.0. ABSTRACT	41
2.1. INTRODUCTION	41
2.1.1. <i>Calmodulin</i>	45
2.1.2. <i>AlleyCat</i>	47
2.1.3. <i>Coordinating alternative metals in AlleyCat</i>	52
2.1.4. <i>CuSeCat and previous work</i>	53

2.2. DESIGNING A SELECTIVE ALLEYCAT	56
2.2.1. <i>Position 9 - S101 and N137</i>	56
2.2.2. <i>Position 7- Y99 and Q135</i>	58
2.2.3. <i>Position 3 – D95 and D131</i>	58
2.2.4. <i>MALDI-TOF</i>	60
2.3. CHARACTERIZING HOLLEE WITH DIFFERENT LANTHANIDES	61
2.3.1. <i>Kinetic assays</i>	61
2.3.2. <i>pH profiling</i>	68
2.3.3. <i>Isothermal Titration Calorimetry</i>	68
2.3.4. <i>Circular Dichroism analysis of binding</i>	71
2.4. CONCLUSIONS.....	77
2.5. EXPERIMENTAL	77
2.5.1. <i>Site-directed Mutagenesis</i>	77
2.5.2. <i>Expression of AlleyCat7 mutants</i>	78
2.5.3. <i>Removal of His₆-tag</i>	79
2.5.4. <i>Single state kinetic assays</i>	80
2.5.5. <i>Metal Dependence Assays</i>	80
2.5.6. <i>pH profiles</i>	81
2.5.7. <i>Circular Dichroism</i>	81
2.5.8. <i>MALDI-TOF Mass Spectrometry</i>	82
2.5.9. <i>Isothermal Titration Calorimetry</i>	82

**CHAPTER 3 APPLICATIONS OF HOLLEE: DESIGNING A SELECTIVE HCC-TARGETING
TREATMENT USING MONOCLONAL ANTIBODIES91**

3.0. ABSTRACT 91

3.1. INTRODUCTION 92

 3.1.1. *Hepatocellular Carcinoma* 92

 3.1.2. *Antibodies* 96

 3.1.3. *Radioimmunotherapy* 102

3.2. ISOLATION OF 2E10-HOLLEE CONSTRUCT 104

3.3. CONCLUSIONS..... 107

3.4. EXPERIMENTAL 108

 3.4.1. *Expression and Purification of 2E10-Hollee* 108

 3.4.2. *Removal of His₆-tag* 109

 3.4.3. *Single-state kinetic assays* 110

**CHAPTER 4 FUNCTIONAL FORMATE DEHYDROGENASE SECRETION IN PICHIA
PASTORIS116**

4.0. ABSTRACT 116

4.1. INTRODUCTION 117

 4.1.1. *Common Methods of Biofuel Production* 117

 4.1.2. *Secretory pathway of Pichia pastoris*..... 120

 4.1.3. *Constructing a biofuel producing apparatus* 124

4.2. DESIGN AND SCREENING pPic9 VECTOR OF FDH SECRETION	126
4.3. CHARACTERIZATION OF YEAST-SECRETED FORMATE DEHYDROGENASE.....	129
4.3.1. <i>Time Elapsed Activity of Secreted Yeast</i>	129
4.3.4. <i>Investigating the Decreased Activity of Secreted Formate Dehydrogenase</i>	132
4.3.5. <i>Formaldehyde dehydrogenase and Alcohol dehydrogenase</i>	137
4.4. CONCLUSIONS.....	139
4.5. EXPERIMENTAL	140
4.5.1. <i>Ligation-independent Cloning of FDH-N into pMCSG49</i>	140
4.5.2. <i>Cloning of FDH-C into pET-28a</i>	140
4.5.3. <i>Protein Expression and Purification</i>	141
4.5.4. <i>Cloning of FDH Genes into pPic9</i>	142
4.5.5. <i>Protein Expression and Purification of FDH in P. pastoris</i>	142
4.5.6. <i>Kinetic Assays for FDH</i>	144
4.5.7. <i>Kinetic Assays for FIDH-aeru</i>	145
4.5.8. <i>Kinetic assays for FIDH-puti</i>	146
4.5.9. <i>Kinetic assay for ADH</i>	148
4.5.10. <i>Tryptic digestion of FDH</i>	148
4.5.11. <i>MALDI-TOF of FDH</i>	149
CHAPTER 5 NMR DIRECTED EVOLUTION OF ALLEYCAT SERIES	154
5.0. ABSTRACT	154

5.1. INTRODUCTION	154
5.1.1. <i>Limitations of directed evolution and rational design</i>	154
5.1.2. <i>“Hot” and “Cold” Spots of AlleyCat</i>	158
5.2. ANALYSIS OF THE IMPORTANCE OF A128T MUTATION IN ALLEYCAT	160
5.3. ANALYSIS OF V108 POSITION AS A SITE FOR KEMP ELIMINATION CATALYSIS.....	163
5.4. CONCLUSIONS.....	168
5.5. EXPERIMENTAL	169
5.5.1. <i>Site-Directed Mutagenesis</i>	169
5.5.2. <i>Purification of AlleyCat series</i>	169
5.5.3. <i>Michaelis-Menten for Kemp elimination reaction</i>	170

CHAPTER 6 ASYMMETRIC HYDROGENATION OF KETONES USING CATALYTIC

RUTHENIUM- AND IRIIDIUM-PEPTIDE AGGREGATES.....173

6.0. ABSTRACT	173
6.1. INTRODUCTION	174
6.1.1. <i>Organometallic Catalysts</i>	174
6.1.2. <i>Chiral Ligands</i>	175
6.1.3. <i>Phenylalanine Aggregation</i>	177
6.1.4. <i>Noyori catalysts</i>	178
6.2. DESIGN OF TRIPEPTIDE RUTHENIUM CATALYSTS	180
6.2.1. <i>NMR</i>	180

6.3. DESIGN OF TRIPEPTIDE IRIIDIUM CATALYSTS	187
6.4. CONCLUSIONS.....	188
6.5. EXPERIMENTAL	188
6.5.1. <i>Solid-phase peptide synthesis</i>	188
6.5.2. <i>Preparation of L-Dap(Ts)OH</i>	189
6.5.3. <i>Preparation of L-allo-Thr(Ts)-OH</i>	190
6.5.4. <i>Hydrogenation of acetophenone</i>	190
6.5.4. <i>NMR</i>	191
6.5.5. <i>Gas Chromatography</i>	191
CHAPTER 7 APPENDIX	195
APPENDIX I. COPYRIGHT PERMISSIONS	195
APPENDIX II. SUPPLEMENTAL DATA	195
S.1. SEQUENCING	195
S.2. SUPPLEMENTAL DATA FOR CHAPTER 2	200
S.2.1. <i>Gel Electrophoresis</i>	200
S.2.2. <i>Kemp elimination kinetics</i>	205
S.2.3. <i>pH profiles of HolIEE</i>	212
S.2.4. <i>Isothermal Titration Calorimetry</i>	223
S.2.4. <i>Circular dichroism</i>	224
S.3. SUPPLEMENTAL DATA FOR CHAPTER 4	232

<i>S.3.1. Gel Electrophoresis</i>	232
<i>S.3.2. Tryptic digest</i>	237
S.4. SUPPLEMENTAL DATA FOR CHAPTER 5	239
<i>S.4.1. Gel electrophoresis</i>	239
APPENDIX III. PUBLISHED WORKS	240
APPENDIX IV. CURRICULUM VITAE	240

LIST OF FIGURES

- FIGURE 1.1. TRANSITION STATE DIAGRAM OF AN UNCATALYZED (RED) AND ENZYMATICALLY CATALYZED (BLUE) EXOTHERMIC REACTION.^{1-4, 7} ES REFERS TO THE ENZYME-SUBSTRATE COMPLEX THAT IS FORMED WHEN THE SUBSTRATE BINDS TO THE ENZYME. ES⁺ REFERS TO THE TRANSITION STATE OF THE SUBSTRATE AS IT IS BOUND TO THE ENZYME. EP IS THE ENZYME-PRODUCT COMPLEX THAT IS FORMED BEFORE THE PRODUCT IS RELEASED FROM THE PROTEIN. ΔG^+ IS THE GIBBS FREE ENERGY REQUIRED FOR THE SUBSTRATE TO REACH THE TRANSITION STATE BETWEEN THE SUBSTRATE AND PRODUCT..... 2
- FIGURE 1.2. LOCK-AND-KEY MODEL OF THE ACTIVE SITE HYPOTHESIZES THAT THE ACTIVE SITE (A, B, AND C) OF THE PROTEIN (YELLOW) IS SPECIFIC TO THE SUBSTRATE (GREEN) AND THEY FIT PRECISELY TO FORM THE ENZYME-SUBSTRATE COMPLEX.⁸..... 3
- FIGURE 1.3. INDUCED FIT MODEL SUGGESTS THAT THE PROTEIN (YELLOW) IS ABLE TO CHANGE SHAPE FLUIDLY AND FORM AROUND THE SUBSTRATE (GREEN) SO THAT THE POINTS OF CONTACT IN THE ACTIVE SITE (A, B, AND C) FIT WITH THE SUBSTRATE.¹⁰⁻¹¹..... 4
- FIGURE 1.4. HOW COFACTORS ARE CATEGORIZED. ORGANIC MOLECULES ARE USUALLY DERIVED FROM VITAMINS AND CAN BE CLASSIFIED AS EITHER COENZYMES (IF NON-COVALENTLY BOUND) OR AS PROSTHETIC GROUPS (IF COVALENTLY BOUND).¹³ PROTEINS WITH METAL ION COFACTORS ARE ALSO KNOWN AS METALLOPROTEINS AND MAKEUP APPROXIMATELY 40% OF ALL KNOWN PROTEINS.¹²..... 5
- FIGURE 1.5. AN EXAMPLE OF ENZYME THAT REQUIRES A COFACTOR FOR CATALYSIS.³⁻⁴ IN THIS EXAMPLE, THE ACTIVE SITE IS ON THE COFACTOR (LAVENDER) AND CAN ONLY BE FORMED PROPERLY WHEN BOUND TO THE PROTEIN (LIGHT BLUE). OTHER WAYS COFACTORS ARE USED BY PROTEINS INCLUDE PROVIDING ENERGY FOR THE REACTION, ALTERING CONFORMATIONS, AND ACTING AS A MEANS OF MAINTAINING EQUILIBRIUM..... 5

FIGURE 1.6. A DIAGRAM OF ALLOSTERIC INHIBITION AND ACTIVATION.^{1, 3, 16-18} ALLOSTERIC INHIBITION (A) IS WHEN AN ALLOSTERIC INHIBITOR BINDS TO THE ALLOSTERIC SITE AND PREVENTS THE REACTION FROM OCCURRING BY DISTORTING THE SHAPE OF THE ACTIVE SITE SO THE SUBSTRATE CANNOT BIND. ALLOSTERIC ACTIVATION (B) IS WHEN AN ALLOSTERIC ACTIVATOR BINDS TO THE ALLOSTERIC SITE AND ALLOWS FOR THE ACTIVE SITE TO BIND TO SUBSTRATES. 6

FIGURE 1.7. THE DIFFERENT TYPES OF INHIBITION^{1-4, 20} (A) IS AN EXAMPLE OF A PROTEIN AND SUBSTRATE IN THE ABSENCE OF ANY TYPE OF INHIBITION. (B) IS A DIAGRAM OF COMPETITIVE INHIBITION WHERE THE INHIBITOR OCCUPIES THE ACTIVE SITE. (C) IS UNCOMPETITIVE INHIBITION WHERE THE INHIBITOR BINDS OUTSIDE OF THE ACTIVE SITE TO STOP A REACTION. (D) IS NONCOMPETITIVE INHIBITION WHERE AN INHIBITOR BINDING OUTSIDE OF THE ACTIVE SITE REDUCES THE ACTIVITY OF THE ENZYME. 7

FIGURE 1.8. MICHAELIS-MENTEN EQUATION.¹..... 8

FIGURE 1.9. THE 20 MOST COMMON AMINO ACIDS FOUND IN NATURE. EACH AMINO ACID CAN BE SPECIFIED WITH EITHER A ONE LETTER OR THREE LETTER CODE. THE AMINO ACIDS LINK TOGETHER VIA PEPTIDE BONDS TO FORM LONG POLYPEPTIDE CHAINS THAT CAN BECOME PROTEINS. 10

FIGURE 1.10. A SUMMARY OF THE CATEGORIES OF PROTEIN STRUCTURE: PRIMARY, SECONDARY, TERTIARY, AND QUATERNARY.²²⁻²⁵ THE PRIMARY STRUCTURE IS THE LINEAR AMINO ACID SEQUENCE OF THE PROTEIN. THE SECONDARY STRUCTURE IS THE LOCAL STRUCTURE OF THE LINEAR SEGMENTS DUE TO HYDROGEN BONDS. THE TERTIARY STRUCTURE IS THE THREE-DIMENSIONAL ARRANGEMENT OF ALL THE AMINO ACIDS IN THE POLYPEPTIDE CHAIN, INCLUDING THE SIDE CHAINS. THE QUATERNARY STRUCTURE IS THE ARRANGEMENT OF SINGLE POLYPEPTIDE CHAINS (SUBUNITS) INTO A MULTI-SUBUNIT PROTEIN..... 11

FIGURE 1.11. THE SECONDARY STRUCTURE OF A PROTEIN CAN EITHER BE CATEGORIZED AS AN α -HELIX OR A β -SHEET.^{7, 22, 24, 26} 12

FIGURE 1.12. THE TERTIARY STRUCTURE OF PROTEINS CONSISTS OF INTERMOLECULAR INTERACTIONS SUCH AS IONIC BONDS OR SALT BRIDGES, DISULFIDE BONDS, HYDROPHOBIC INTERACTIONS, AND HYDROGEN BONDING.⁷ 13

FIGURE 1.13. QUATERNARY STRUCTURE OF HEMOGLOBIN.²² THIS PROTEIN CONSISTS OF FOUR SUBUNITS COMPLEXED TOGETHER INTO A FINAL STRUCTURE..... 13

FIGURE 1.14. PROTEIN ENGINEERING STRATEGIES.⁴⁴ DIRECTED EVOLUTION REFERS TO INTRODUCING MUTATIONS RANDOMLY AND SCREENING TO SEE WHICH MUTATIONS ARE SUCCESSFUL. RATIONAL DESIGN USES COMPUTATIONAL ALGORITHMS TO DETERMINE THE MOST STABLE MUTATIONS OF A PROTEIN AND THEN CHARACTERIZES THEM. SEMI-RATIONAL DESIGN IS A COMBINATION OF THE TWO METHODS. 15

FIGURE 1.15. CODON TABLE. EACH THREE-LETTER SEQUENCE OF NUCLEOTIDES ENCODES FOR A PROTEIN. MANY AMINO ACIDS HAVE MULTIPLE CODONS. 17

FIGURE 1.16. DESCRIPTION OF SANGER SEQUENCING.^{2-4, 23} DDNTPS ARE TAGGED AND MIXED WITH THE TEMPLATE STRAND. IF A TAGGED DDNTP BINDS TO THE STRAND, THEN THE SEQUENCE STOPS. IN THIS WAY, ONE CAN EVENTUALLY BUILD A MAP OF WHERE EACH CODON AND NUCLEIC ACID LIE IN RESPECT TO THE GENE. 18

FIGURE 1.17. PROCESS OF POLYMERASE CHAIN REACTION.^{3, 46} PCR CONSISTS OF THREE MAIN STEPS: DENATURATION, ANNEALING, AND REPLICATION. OVER TIME, PCR YIELDS MULTIPLE COPIES OF A DESIRED SEGMENT OF DNA.19

FIGURE 1.18. SITE-DIRECTED MUTAGENESIS BY USING BASE-PAIR SUBSTITUTION, INSERTION, AND DELETION.⁴⁷ 20

FIGURE 1.19. ERROR-PRONE PCR AND DNA RECOMBINATION OR SHUFFLING.⁴⁸ 21

FIGURE 1.20. A DIAGRAM OF DIRECTED EVOLUTION.⁵⁴⁻⁵⁹ THE INNER CIRCLE INDICATES THE THREE STAGES OF DIRECTED EVOLUTION WITH THE NATURAL EQUIVALENT IN PARENTHESIS. THE OUTER CIRCLE INDICATES A TYPICAL EXPERIMENTAL SETUP..... 22

FIGURE 1.21. PROTEIN TRANSFORMATION.^{15, 64-65} PLASMIDS WITH A DESIRED GENE CAN BE INTRODUCED INTO BACTERIA SUCH AS *E. COLI* BY FIRST 'SHOCKING' THE BACTERIUM WITH EITHER HEAT OR AN ELECTRIC SHOCK, THEN ALLOWING THE CELL TO REST SO IT WILL ACCEPT THE PLASMID INTO ITS BODY AND RECOVER. 24

FIGURE 1.22. DIAGRAM OF PROTEIN INDUCTION.³ IN NORMAL CONDITIONS WITHOUT LACTOSE, A REPRESSOR BINDS TO THE OPERON (THE SEQUENCE THAT SIGNALS PROTEIN SYNTHESIS) AND RNA POLYMERASE IS UNABLE TO ATTACH. IN THE PRESENCE OF AN INDUCING AGENT SUCH AS IPTG (BOTTOM), IT WILL BIND TO THE REPRESSOR AND ALLOW PROTEIN SYNTHESIS TO TAKE PLACE..... 26

FIGURE 1.23. NICKEL-NTA RESIN.⁶⁸ NTA CHELATES THE NICKEL ION TO THE RESIN BEAD. WHEN HISTIDINE IS PRESENT, THE NICKEL IONS BIND THE RESIDUES WHILE ALLOWING OTHER IMPURITIES TO PASS BY. TO DETACH THE HISTIDINE, A COMPETING COMPOUND SUCH AS IMIDAZOLE IS ADDED TO REPLACE THE HISTIDINE ON THE NICKEL..... 27

FIGURE 1.24. AFFINITY CHROMATOGRAPHY.^{1, 3} THE PROTEIN BINDS TO THE AFFINITY RESIN WHILE ALL OTHER IMPURITIES ARE WASHED AWAY, THEN THE PROTEIN IS ELUTED OFF USING A COMPETING MOLECULE SUCH AS IMIDAZOLE. TO REMOVE THE ELUTING AGENT, THE PROTEIN CAN THEN BE PUT ON A DESALTING COLUMN WHICH TRAPS SMALLER IONS ON TOP WHILE THE PROTEIN FLOWS THROUGH, ALLOWING THE BUFFER TO BE 'EXCHANGED.'
..... 28

FIGURE 1.25. GEL ELECTROPHORESIS.⁶⁹⁻⁷⁰ PROTEIN SAMPLES ARE DYED AND LOADED ONTO A GEL MADE OF ACRYLAMIDE POLYMER. AN ELECTRICAL CURRENT IS PASSED THROUGH THE GEL, AND SMALLER MACROMOLECULES TRAVEL FURTHER DOWN THE GEL WHILE LARGER ONES STAY NEAR THE TOP..... 29

FIGURE 1.26. UV-VIS SPECTROSCOPY OF PROTEINS AND NUCLEIC ACIDS.⁷⁰ NUCLEIC ACIDS HAVE AN ABSORBANCE MAXIMA AT 260 NM. PROTEINS HAVE ABSORBANCE MAXIMA AT 280 NM DUE TO TYROSINE AND TRYPTOPHAN RESIDUES. 30

FIGURE 1.27. UV-VIS SPECTRA OF AROMATIC AMINO ACIDS.⁷¹ DUE TO THEIR AROMATIC GROUPS, TRYPTOPHAN AND TYROSINE BOTH PEAK AT 280 NM. PHENYLALANINE PEAKS AT 260 NM, THE SAME WAVELENGTH AS NUCLEIC ACIDS, MAKING IT AN UNRELIABLE MEASURE OF PROTEIN CONCENTRATION. 30

FIGURE 1.28. DIAGRAM OF MALDI-TOF.³¹ A LASER RELEASES SAMPLE BOUND TO THE MATRIX, WHICH TRAVELS THROUGH AN ELECTRIC FIELD TO A DETECTOR. THE DETECTOR DETERMINES THE SIZE OF THE SAMPLE BY CALCULATING HOW LONG IT TOOK FOR THE SAMPLE TO REACH THE DETECTOR. 31

FIGURE 2.1. SUMMARY OF THE MOST COMMON METALS USED IN METALLOPROTEINS.^{2, 7} 43

FIGURE 2.2. EXAMPLES OF EVOLVED ORGANISMS AND PROTEINS THAT BIND NON-TYPICAL METALS SPECIFICALLY. LEFT: *CUPRIAVIDAS METALLIDURANS*, A BACTERIUM THAT IS ABLE TO LIVE IN TOXIC WASTE DUE TO ITS ABILITY TO SELECTIVELY BIND Pb^{2+} AND SHIP IT OUT OF THE CELL BEFORE IT DOES DAMAGE.¹¹ RIGHT: THE ACTIVE SITE OF METHANOL DEHYDROGENASE, ISOLATED FROM *METHYLOBACTERIUM EXTORQUENS*, WHICH COORDINATES WITH A La^{3+} ION TO ALLOW IT TO LIVE IN EXTREME TEMPERATURES.¹⁰ 44

FIGURE 2.3. EXAMPLES OF SYSTEMS AND PROTEINS THAT CALMODULIN AND CALCIUM REGULATE.²¹ 45

FIGURE 2.4. DIAGRAM OF EF-HANDS IN CALCIUM-SENSING PROTIENS.²¹ RESIDUES 1, 3, 5, 7, 9, AND 12 COORDINATE DIRECTION WITH THE Ca^{2+} ION TO FORM A PENTAGONAL BIPYRAMIDAL COMPLEX..... 46

FIGURE 2.5. CONFORMATIONAL CHANGE OF CALMODULIN IN THE PRESENCE OF CALCIUM.²⁴ 47

FIGURE 2.6. SCHEMATIC OF CALMODULIN CONVERSION TO C-TERMINAL CALMODULIN. MUTATING THE PHENYLALANINE IN THE 92 POSITION OF C-TERM CAM TO GLUTAMATE YIELDS ALLEYCAT PROTEIN.¹⁻⁴ 48

FIGURE 2.7. A) 3D REPRESENTATION OF CELL SURFACE OF ALLEYCAT. B) COMPARISON OF CATALYTIC ACTIVITY BETWEEN F92E (BLACK CIRCLES) AND NEGATIVE CONTROL MUTANT F92Q (BLACK DIAMONDS).²⁵ 49

FIGURE 2.8. KINETIC ASSAY OF KEMP ELIMINATION SHOWING THE ALLOSTERIC REGULATION OF ALLEYCAT7.²⁷ REMOVING CALCIUM IONS FROM THE REACTION WITH CHELATOR REDUCES KINETIC ACTIVITY AND VICE VERSA. 50

FIGURE 2.9. GENERAL MECHANISM OF AN ENGINEERED PROTEIN METAL SENSOR.²⁶ 51

FIGURE 2.10. CRYSTAL STRUCTURE OF THE EF-HAND METAL BINDING SITE OF CALMODULIN WHEN BOUND TO (A) Ca^{2+} , (B) Pb^{2+} , (C) Sr^{2+} , AND (D) Ba^{2+} .⁴⁵ 52

FIGURE 2.11. MODELING OF S101 AND N137 POSITIONS OF ALLEYCAT.²⁸ THE GREEN RIBBON REPRESENTS THE STRUCTURE OF CUSeCAT, THE BLUE STICKS REPRESENT THE ACTIVE RESIDUES IN THE METAL BINDING SITE, THE RED SPHERES REPRESENT THE F92E MUTATION RESPONSIBLE FOR KEMP ELIMINATION, THE ORANGE SPHERES REPRESENT THE S101E AND N137E MUTATIONS THAT WERE INTRODUCED TO CUSeCAT, AND THE GRAY SPHERES REPRESENT A TRIVALENT METAL ION. 54

FIGURE 2.12. METAL DEPENDENCE ASSAY OF CUSeCATEE.²⁸ IN LOW CONCENTRATIONS OF $YbCl_3$ AND $CaCl_2$, CUSeCATEE IS MORE ACTIVE IN THE PRESENCES OF LANTHANIDE. 55

FIGURE 2.13. METAL DEPENDENCE ASSAY OF CUSeCAT SERIES WITH EXCESS OF $CaCl_2$.²⁸ 56

FIGURE 2.14. THREE-DIMENSIONAL STRUCTURE OF HOLLEE CONSTRUCTED IN PYMOL USING CAM STRUCTURE (PDB CODE 1CLL) AS A TEMPLATE. THE YELLOW SPHERES REPRESENT YTTRIUM IONS.¹ 59

FIGURE 2.15. MALDI-TOF OF CUT HOLLEE. THE CALCULATED MASS OF HOLLEE IS 8530 DA..... 61

FIGURE 2.16. CATALYTIC EFFICIENCY OF KEMP ELIMINATION WITH HOLLEE IN THE PRESENCE OF VARIOUS LANTHANIDES. THE LEFT GRAPH PICTURES ALL THE METALS TESTED WHILE THE RIGHT GRAPH ONLY FEATURES

LUTETIUM, YTTERBIUM, AND CALCIUM. ALL SAMPLES WERE RUN IN 20 mM MOPS, 100 mM NaCl, pH 7.0 WITH 0.08 mM METAL UNLESS OTHERWISE NOTED. THE CONCENTRATION OF HOLLEE WAS 1.0 μ M. ALL K_{CAT}/K_M ARE REPORTED IN TABLE 2.4. (MT7081) 63

FIGURE 2.17. CORRELATION BETWEEN IONIC RADIUS (PM) AND CATALYTIC ACTIVITY WITH 0.08 mM METAL CONCENTRATION. ALL IONIC RADII FOR THE LANTHANIDES ARE CALCULATED FOR SIX-COORDINATE OCTAHEDRAL GEOMETRY.⁵² (MT7081)..... 65

FIGURE 2.18. ACTIVE SITE OF ALLEYCAT (LEFT) AND HOLLEE (RIGHT) IN THE PRESENCE OF CALCIUM (CYAN SPHERE). BOTH FIGURES WERE GENERATED IN PYMOL USING 1CLL AS A TEMPLATE. 66

FIGURE 2.19. CATALYTIC EFFICIENCY OF KEMP ELIMINATION USING 2.5 μ M HOLLEE IN 20MM MOPS AT PH 7.0. KINETIC ACTIVITY WAS MEASURED IN THE PRESENCE OF DIFFERENT METALS AND VARYING CONCENTRATIONS. (MT7047)..... 67

FIGURE 2.20. SCHEMATIC OF BASIC EXPERIMENTAL SETUP FOR ITC (LEFT) AND THE TYPICAL DATA ACQUIRED FROM ITC (RIGHT).⁵⁵..... 69

FIGURE 2.21. ITC DATA OF HOLLEE TITRATED WITH $Ca(NO_3)_2$ (LEFT) AND $Y(NO_3)_3$ (RIGHT). THE SAMPLE CELL CONTAINED 0.1MM HOLLEE IN 20MM MOPS PH 7.0 AT 25°C AND WAS SLOWLY TITRATED WITH EITHER 5MM $Ca(NO_3)_2$ OR $Y(NO_3)_3$ IN 5 μ L INJECTIONS WITH AN EQUILIBRATION TIME OF 240 SECONDS. (MT6155) .. 70

FIGURE 2.22. SECONDARY STRUCTURES VISUALIZED BY CIRCULAR DICHROISM.⁶¹ EACH SECONDARY STRUCTURE HAS A UNIQUE READOUT THAT CAN BE USED TO ANALYZE PROTEIN AND PEPTIDE SAMPLES..... 72

FIGURE 2.23. CIRCULAR DICHROISM OF HOLLEE (25 μ M, 4MM MOPS, PH 7.0, 22°C) WITH 100 μ M OF DIFFERENT METALS. (MT7011) 73

FIGURE 2.24. CIRCULAR DICHROISM OF HOLLEE (25 μ M) IN 4MM MOPS PH 7.0 WITH VARYING CONCENTRATIONS OF LuCl_3 , WHICH ARE LISTED IN THE LEGEND (EXCEPT IN THE CASE OF EDTA WHICH ADDED A CONCENTRATION OF 0.1MM EDTA INSTEAD). AN INCREASE IN THE CONCENTRATION OF METAL IONS RESULTS IN INCREASINGLY NEGATIVE MRE AT 207NM AND 220NM. (MT7026) 74

FIGURE 2.25. ANALYSIS OF METAL TITRATION OF FIGURE 2.24 AT 220 NM. THE Δ MRE IS CALCULATED BY SUBTRACTING MEAN RESIDUE ELLIPTICITY AT 220 NM FROM THE MRE AT 220NM OF HOLLEE IN 0.1MM EDTA. (MT7026) 75

FIGURE 2.26. K_D ANALYSIS OF HOLLEE (25 μ M, 4MM MOPS, PH 7.0, 22°C) WITH LuCl_3 TITRATION USING MEAN RESIDUE ELLIPTICITY AT 220NM. (MT7026) 76

FIGURE 3.1. DOCUMENTED INCIDENCES OF HEPATOCELLULAR CARCINOMA FOR ALL AGES AND GENDERS IN 2013 WORLDWIDE.⁶ 92

FIGURE 3.2. ALGORITHM FOR DETERMINING TREATMENT OPTIONS FOR HCC.⁷ 93

FIGURE 3.3. ALGORITHM FOR DIAGNOSIS OF HCC.¹⁰ 94

FIGURE 3.4. TRANSARTERIAL CHEMOEMBOLIZATION OF HCC USING ^{90}Y .¹² 95

FIGURE 3.5. THREE MECHANISMS OF ANTIBODY DEFENSE.¹⁵ THE FIRST MECHANISM IS WHEN THE ANTIBODIES BIND TO BACTERIAL TOXINS SO THEY DO NOT DAMAGE CELLS AND THEN TRIGGER MACROPHAGES TO INGEST THE ANTIBODY-TOXINS. THE SECOND METHOD INVOLVES THE ANTIBODY ATTACKING THE BACTERIA IN THE EXTRACELLULAR SPACE AND TRIGGERING INGESTION BY MACROPHAGES. THE FINAL METHOD IS WHEN ANTIBODIES BIND TO BACTERIUM IN THE PLASMA AND TRIGGER MACROPHAGES TO BREAK APART AND DIGEST THE BACTERIA. 97

FIGURE 3.6. GENERAL STRUCTURE OF ANTIBODY. HEAVY CHAINS ARE FEATURED IN GREEN AND LIGHT CHAINS ARE FEATURED IN YELLOW.¹⁵ DISULFIDE BONDS BETWEEN THE CHAINS ARE USED TO CREATE THE DISTINCTIVE STRUCTURE OF THE ANTIBODY..... 98

FIGURE 3.7. HOW MONOCLONAL ANTIBODIES ARE MADE.¹⁸ MONOCLONAL ANTIBODIES ARE USED BECAUSE THEY ALL BIND TO A SINGLE EPITOPE AND ARE HIGHLY SPECIFIC FOR THE TUMOR CELLS. IF POLYCLONAL ANTIBODIES WERE USED, THEN THERE WOULD BE MUCH LESS SPECIFICITY AND EVEN THE POSSIBILITY THAT THE AB COULD BIND TO A HEALTHY CELL INSTEAD..... 99

FIGURE 3.8. STRATEGIES OF CANCER THERAPEUTIC MONOCLONAL ANTIBODIES.¹⁷ A) THE MAb BINDS TO TARGET AND TRIGGERS ANTIBODY-DEPENDENT CYTOTOXICITY, COMPLEMENT-MEDIATED CYTOTOXICITY, OR DIRECTLY SIGNALS CANCER CELL DEATH. B) THE MAb INHIBITS THE DEVELOPMENT OF NEW BLOOD VESSELS THAT FEED THE TUMOR CELLS. C) THE MAb ELICITS A RESPONSE FROM ANTITUMOR T CELLS BY BLOCKING SIGNALS THAT WOULD NORMALLY PROTECT THE TUMOR FROM THE BODIES OWN IMMUNE SYSTEM. D) THE MAb IS USED TO DELIVER RADIOISOTOPES DIRECTLY TO CANCER CELLS FOR SELECTIVE RADIATION TREATMENT. E) THE MAb IS USED TO DELIVER DRUGS DIRECTLY TO THE CANCER CELLS. F) A SMALL PORTION OF THE MAb IS USED TO BOTH RECOGNIZE CANCER CELLS ON ONE END AND ACTIVATE ANTIGENS TO TRIGGER IMMUNE EFFECTOR CELLS ON THE OTHER END. G) THE MAb VARIABLE REGION IS ATTACHED TO A SIGNALING PEPTIDES THAT IS TRANSFERRED TO T CELLS AND MAKES THEM CHIMERIC ANTIGEN RECEPTOR (CAR) T CELLS THAT ARE SPECIFIC FOR THE TUMOR THE PEPTIDE IS SPECIFIC FOR. 100

FIGURE 3.9. STRUCTURE OF VARIOUS ANTIBODY DERIVATIVES.²⁷ AS LONG AS THE STRUCTURE CONTAINS THE VARIABLE REGION THAT IS RESPONSIBLE FOR BINDING TO ANTIGENS, THE FRAGMENT WILL SUCCESSFULLY BIND. SMALLER BODIES HAVE THE ADVANTAGE OF BEING MORE SOLUBLE AND CELL PERMEABLE..... 101

FIGURE 3.10. DIAGRAM OF RADIOIMMUNOTHERAPY.³⁶ 103

FIGURE 3.11 PERIODIC TABLE HIGHLIGHTING RADIONUCLIDES OF INTEREST FOR CANCER TREATMENT.³² 104

FIGURE 3.12. CATALYTIC EFFICIENCY OF KEMP ELIMINATION USING HOLLEE (0.5 μ M) IN THE PRESENCE OF YTTRIUM (0.08 MM) AND CALCIUM (0.08 MM) IN 20 MM MOPS, 100 MM NaCl, pH 7.0. THE K_{CAT}/K_M OF 2E10-HOLLEE WAS 8.0 $M^{-1}S^{-1}$ IN THE PRESENCE OF YTTRIUM. (MT7021) 105

FIGURE 3.13. CATALYTIC EFFICIENCY OF KEMP ELIMINATION USING 2E10-HOLLEE (0.5 μ M) IN THE PRESENCE OF YTTRIUM (0.08 MM) AND CALCIUM (0.08 MM) IN 20 MM MOPS, 100 MM NaCl, pH 7.0. THE K_{CAT}/K_M OF 2E10-HOLLEE WAS 8.0 $M^{-1}S^{-1}$ IN THE PRESENCE OF YTTRIUM AND THE K_{CAT}/K_M OF HOLLEE WAS 101 $M^{-1}S^{-1}$ IN THE PRESENCE OF YTTRIUM. (MT7021)..... 106

FIGURE 3.14. 10% ACRYLAMIDE SDS-PAGE GEL OF 2E10-HOLLEE PURIFICATION: LANE 1) CELLS BEFORE INDUCTION, 2) CELL PELLET BEFORE PURIFICATION, 3) SUPERNATANT AFTER SONICATION, 4) SUPERNATANT BEFORE LOADING ONTO Ni-NTA COLUMN, 5) Ni-NTA COLUMN FLOW THROUGH, 6) Ni-NTA COLUMN WASH, 7) UNCUT 2E10-HOLLEE BEFORE TEV CLEAVAGE, 8) UNCUT 2E10-HOLLEE AFTER INCUBATION OVERNIGHT WITH TEV, 9) FINAL SAMPLE AFTER REMOVING TEV. THE CALCULATED MASS OF 2E10-HOLLEE WITH HIS6-TAG IS 39,725 DA AND WITHOUT HIS6-TAG IS 37,260 DA. 107

FIGURE 4.1. SUMMARY OF THE PRODUCTION METHODS FOR THE FOUR GENERATIONS OF BIOFUELS.¹ 118

FIGURE 4.2. A SUMMARY OF THE MECHANISMS *PICHIA PASTORIS* USES IN ORDER TO EXPRESS AND SECRETE PROTEIN. 121

FIGURE 4.3. A) THE pPic9 VECTOR USED IN YEAST TRANSFORMATION; B) SUMMARY OF HOMOLOGOUS RECOMBINATION OF pPic9 INTO GS115 GENOME: THE GENE OF INTEREST IS CLONED INTO THE pPic9 VECTOR. UPON TRANSFORMATION, THE pPic9 VECTOR WILL INTEGRATE ITSELF INTO THE YEAST GENOME, WHICH

CONTAINS A MUTATION THAT DOES NOT ALLOW FOR THE PRODUCTION OF HISTIDINE IN YEAST. ADDITION OF THE PPIC9 VECTOR INTO THE GENOME ALLOWS FOR YEAST TO PRODUCE HISTIDINE, WHICH ALLOWS FOR SCREENING OF YEAST COLONIES FOR THE DESIRED GENE. ²⁰	122
FIGURE 4.4. PROTEIN SYNTHESIS IN YEAST FROM THE NUCLEUS TO THE GOLGI APPARATUS (WHERE GLYCOSYLATION MAY OCCUR) TO OUTSIDE THE CELL. ²¹	123
FIGURE 4.5. SECRETORY PATHWAY OF YEAST. ²²	124
FIGURE 4.6. MODEL OF QD-PROTEIN SYSTEM.	125
FIGURE 4.7. SUMMARY OF SELF-SUSTAINING QD-PROTEIN SYSTEM.	126
FIGURE 4.8. CRYSTALLOGRAPHY STRUCTURE OF FORMATE DEHYDROGENASE (PDB 5DNA). ²⁸	127
FIGURE 4.9. CATALYTIC EFFICIENCY OF FDH-C AND FDH-N PROTEINS (0.5 mM) MEASURING THE CONVERSION OF NAD ⁺ (1 mM) TO NADH DURING THE OXIDATION OF SODIUM FORMATE (160 mM) TO CARBON DIOXIDE IN 100 mM SODIUM PHOSPHATE BUFFER, PH 7. THE SPECIFIC ACTIVITY WAS CALCULATED TO BE 0.526 U/MG FOR FDH-C AND 2.77 U/MG FOR FDH-N. ²⁹	128
FIGURE 4.10. CATALYTIC EFFICIENCY OF CONVERSION OF NADH (1.2 mM) TO NAD ⁺ BY FDH-N (0.75 mM) USING SODIUM BICARBONATE (200 mM) IN 100 mM SODIUM PHOSPHATE BUFFER, PH 7.0. THE SPECIFIC ACTIVITY OF FDH-N FOR NAHCO ₃ TO FORMATE CONVERSION IS 0.005 U/MG. (MT2051)	129
FIGURE 4.11. CONTINUOUS SECRETION OF FDH FROM YEAST OVER 120 HOURS. EMPTY PPIC9 VECTOR WAS ALSO CLONED INTO YEAST AS A CONTROL. ²⁹	130
FIGURE 4.12. CATALYTIC EFFICIENCY OF FDH-C-Y (0.5 mM) MEASURING THE CONVERSION OF NAD ⁺ (1.0 mM) TO NADH DURING THE OXIDATION OF SODIUM FORMATE TO CARBON DIOXIDE IN 100 mM SODIUM PHOSPHATE,	

PH 7. THE SPECIFIC ACTIVITY OF FDH AT 24 HOURS IS 0.11 U/MG, AT 48 HOURS IS 0.13 U/MG, AND AT 120 HOURS IS 0.10 U/MG. ²⁹	131
FIGURE 4.13. CATALYTIC EFFICIENCY OF FDH-N (0.5 mM) FROM <i>E. COLI</i> AND <i>P. PASTORIS</i> CONSTRUCTS MEASURING THE CONVERSION OF NAD ⁺ (1 mM) TO NADH DURING THE OXIDATION OF SODIUM FORMATE (160 mM) TO CARBON DIOXIDE IN 100 mM SODIUM PHOSPHATE BUFFER AT PH 7. THE SPECIFIC ACTIVITY OF FDH FROM BACTERIA IS 2.71 U/MG AND THE SPECIFIC ACTIVITY OF FDH FROM YEAST IS 0.4 U/MG. ²⁹	133
FIGURE 4.14. MALDI-TOF OF SECRETED FDH C-TERM. THE MATRIX USED WAS SINAPIC ACID. THE MOLECULAR WEIGHT MATCHES THE WEIGHT OF THE FDH IF IT WERE UNMODIFIED (43.3 KDA). ²⁹	134
FIGURE 4.15. 10% ACRYLAMIDE SDS-PAGE GEL. LANES: 1) CONTROL 48 HR, 2) PMSF 24HR, 3) PMSF 48 HR, 4) EDTA 24 HR, 5) EDTA 48 HR, 6) CASAMINO ACIDS 24 HR, 7) CASAMINO ACIDS 48 HR, 8) ALL THREE 24 HR, 9) ALL THREE 48 HR.	135
FIGURE 4.16. MALDI-TOF OF FDH C-TERM DIGESTED BY TRYPSIN. THE PEAK AT 7624 DA CORRESPONDS TO A DIGESTED α -FACTOR. ²⁹	137
FIGURE 4.17. SDS-PAGE (10% ACRYLAMIDE) OF THE SECRETION OF FLDH-PUTI PROTEIN CLONED INTO YEAST. A BAND IS EXPECTED AT ~35 KDA. LANES 1-6 REPRESENT THE MEDIA SECRETION OF DIFFERENT "POSITIVE" HITS OF FLDH-PUTI CLONED INTO YEAST. EVEN THOUGH THE GENE WAS PRESENT IN THE PLASMIDS, THERE WAS NO EXPRESSION.....	139
FIGURE 5.1. SUMMARY OF THE STEPS OF DIRECTED EVOLUTION. ¹⁰	156
FIGURE 5.2. STRUCTURES OF 5-NITRO-1,2-BENZISOXAZOLE (LEFT) AND KEMP ELIMINATION ANTAGONIST 6-NITRO-1H-BENZOTRIAZOLE (RIGHT), WHICH IS USED DURING NMR TITRATIONS TO DETERMINE THE CHEMICAL SHIFTS OF RESIDUES IN THE ALLEYCAT SERIES.	158

FIGURE 5.3. ^{15}N - ^1H HSQC SHIFT DATA FOR ALL RESIDUES IN ALLEYCAT7. NMR DATA WAS COLLECTED AND ANALYZED BY I. KORENDOVYCH AND O. MAKHLYNETS.	160
FIGURE 5.4. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY THE ALLEYCAT SERIES WITH AND WITHOUT THE A128T MUTATION (0.2 mM) IN 20 mM HEPES BUFFER, PH 7.0 WITH 100 mM NaCl AND 10 mM OF CaCl_2 . THE VALUES FOR $k_{\text{CAT}}/k_{\text{M}}$ ARE LISTED IN TABLE 5.1. AC5 IS THE EQUIVALENT OF Ac4T. (MT4100)	161
FIGURE 5.5. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY ALLEYCAT0 WITH AND WITHOUT THE A128T MUTATION (20 mM) IN 20 mM HEPES BUFFER, PH 7.0 WITH 100 mM NaCl AND 10 mM OF CaCl_2 . THE VALUES FOR $k_{\text{CAT}}/k_{\text{M}}$ ARE LISTED IN TABLE 5.1. (MT4128).....	162
FIGURE 5.6. A DIAGRAM OF THE RELATIVE “HOT” SPOTS OF CALMODULIN AS DETERMINED BY NMR TITRATION. RED, ORANGE, AND YELLOW COLORS SIGNIFY “HOT SPOTS” WITH HIGH CHEMICAL SHIFT PERTURBATIONS AND CYAN AND BLUE COLORS SIGNIFY “COLD SPOTS” WITH SMALL CHEMICAL SHIFT PERTURBATIONS (PDB 1CLL). ¹⁸	164
FIGURE 5.7. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED ALLEYCAT0 (20 mM) WITH AND WITHOUT THE V108E MUTATION IN 20 mM HEPES BUFFER PH 7.0 WITH 100 mM NaCl AND 10 mM OF CaCl_2 . THE VALUES FOR $k_{\text{CAT}}/k_{\text{M}}$ ARE LISTED IN TABLE 5.2. (MT5051).....	166
FIGURE 5.8. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED WILD-TYPE CAM (20 mM) WITH AND WITHOUT THE V108E MUTATION IN 20 mM HEPES BUFFER, PH 7.0 WITH 100 mM NaCl AND 10 mM OF CaCl_2 . THE VALUES FOR $k_{\text{CAT}}/k_{\text{M}}$ ARE LISTED IN TABLE 5.2. (MT5142)	167
FIGURE 6.1. EXAMPLES OF PRIVILEGED LIGANDS. ³	176
FIGURE 6.2. TEM IMAGE OF PHENYLALANINE FIBRILS. ⁵	177
FIGURE 6.3. EXAMPLE OF NEW LIGAND DESIGN. X REPRESENTS VARIABLE AMINO ACID SIDE CHAIN.	178

FIGURE 6.4. EXAMPLES OF COMMON NOYORI CATALYSTS WITH RUTHENIUM AND IRIIDIUM METAL CENTERS.	179
FIGURE 6.5. A) STRUCTURE OF L-2,3-DIAMINOPROPIONIC ACID. B) SYNTHESIS OF L-DAP(Ts)-OH FROM L-(Boc)DAP-OH.....	180
FIGURE 6.6. MECHANISM OF ACETOPHENONE HYDROGENATION AND SUBSEQUENT CATALYST REGENERATION USING SODIUM FORMATE AS A PROTON SOURCE. IRIIDIUM CAN BE USED IN REPLACE OF RUTHENIUM.....	181
FIGURE 6.7. IDENTIFYING ^1H NMR CHEMICAL SHIFTS OF ACETOPHENONE (LEFT) AND 1-PHENYLETHANOL (RIGHT).	182
FIGURE 6.8. A) SYNTHESIS OF 1-PHENYLETHANOL FROM ACETOPHENONE USING Ru^{2+} COUPLED WITH (R,R)-PTSDPEN. B) ^1H NMR SPECTRUM OF ACETOPHENONE REDUCTION AFTER ONE HOUR (THE AROMATIC REGION IS NOT SHOWN). THERE WAS A 28% CONVERSION OF ACETOPHENONE TO 1-PHENYLETHANOL.....	183
FIGURE 6.9. A) SYNTHESIS OF 1-PHENYLETHANOL FROM ACETOPHENONE USING Ru^{2+} COUPLED WITH L-DAP(Ts)-OH. B) ^1H NMR SPECTRUM OF ACETOPHENONE REDUCTION AFTER ONE HOUR. THERE WAS NO CONVERSION OF ACETOPHENONE TO 1-PHENYLETHANOL.	184
FIGURE 6.10. A) SYNTHESIS OF 1-PHENYLETHANOL FROM ACETOPHENONE USING Ir^{3+} COUPLED WITH (R,R)-PTSDPEN. B) ^1H NMR SPECTRUM OF ACETOPHENONE REDUCTION AFTER ONE HOUR (THE AROMATIC REGION IS NOT SHOWN). THERE WAS A 37% CONVERSION OF ACETOPHENONE TO 1-PHENYLETHANOL.....	185
FIGURE 6.11. A) SYNTHESIS OF 1-PHENYLETHANOL FROM ACETOPHENONE USING Ir^{3+} COUPLED WITH L-DAP(Ts)-OH. B) ^1H NMR SPECTRUM OF ACETOPHENONE REDUCTION AFTER ONE HOUR (THE AROMATIC REGION IS NOT SHOWN). THERE WAS A 5% CONVERSION OF ACETOPHENONE TO 1-PHENYLETHANOL.	186
FIGURE 6.11. ATOMIC FORCE MICROSCOPY IMAGE OF L-ALLO-TFF PEPTIDE. IMAGE WAS TAKEN BY Z. LENGYEL.	187

FIGURE S.1. 15% ACRYLAMIDE SDS-PAGE GEL OF Ac7 PURIFICATION. SAMPLES WERE TAKEN AT VARIOUS STAGES OF THE PURIFICATION AND LABELLED AS FOLLOWS: 2) Ac7 AFTER INDUCTION, 3) Ac7 CELL DEBRIS, 4) Ac7 NI-NTA FLOW THROUGH, 5) Ac7 NI-NTA WASH, 6) Ac7 UNCUT, 7) Ac7 + TEV PROTEASE, 8) Ac7 CUT PURE. 200

FIGURE S.2. 15% ACRYLAMIDE SDS-PAGE GEL OF AcDD (LANES 2-7) AND ACEE (LANES 8-13) PURIFICATION. ALIQUOTS WERE TAKEN AT VARIOUS STAGES OF THE PURIFICATION AND LABELLED AS FOLLOWS: 2) AcDD AFTER INDUCTION, 3) AcDD CELL DEBRIS, 4) AcDD NI-NTA FLOW THROUGH, 5) AcDD UNCUT, 6) AcDD + TEV PROTEASE, 7) AcDD CUT PURE, 8) ACEE AFTER INDUCTION, 9) ACEE CELL DEBRIS, 10) ACEE NI-NTA FLOW THROUGH, 11) ACEE UNCUT, 12) ACEE + TEV PROTEASE, 13) ACEE CUT PURE. 200

FIGURE S.3. A 15% ACRYLAMIDE SDS-PAGE GEL OF PURIFIED Y99E AND Q135E. "UNCUT" REFERS TO PROTEINS WITH SUMO-TAG STILL ATTACHED TO THE PROTEIN WHILE "CUT" PROTEINS HAVE SUMO-TAG REMOVED. NO EXPRESSION WAS SEEN FOR THE Q135E MUTANT. 201

FIGURE S.4. A 15% ACRYLAMIDE SDS-PAGE GEL OF PURIFIED Y99C, S101C, AND N137C. WHILE THE PROTEIN IS NORMALLY EXPECTED AROUND 8-12 kDa, BANDS AT DOUBLE THAT MOLECULAR WEIGHT WERE SEEN. "UNCUT" REFERS TO PROTEINS WITH SUMO-TAG STILL ATTACHED TO THE PROTEIN WHILE "CUT" PROTEINS HAVE SUMO-TAG REMOVED. 202

FIGURE S.5. A 15% ACRYLAMIDE SDS-PAGE GEL OF RECCes PURIFICATION. THE EXPECTED MOLECULAR WEIGHT OF RECCes IS APPROXIMATELY 8 kDa AFTER CLEAVAGE. BEFORE CLEAVAGE OF THE HIS-TAG, RECCes FORMS A DIMER DUE TO DISULFIDE BONDS BETWEEN CYSTEINE RESIDUES. 203

FIGURE S.6. A 15% ACRYLAMIDE SDS-PAGE GEL OF HOLLEE PURIFICATION. THE EXPECTED MW OF HOLLEE IS 8 KDA AFTER CLEAVAGE OF HIS-TAG. DEBRIS REFERS TO THE CELL DEBRIS AFTER SONIFICATION AND CENTRIFUGATION, AND FLOW THROUGH AND WASH REFER TO STEPS DURING NI-NTA PURIFICATION. 204

FIGURE S.7. CATALYTIC EFFICIENCY GRAPH OF KEMP ELIMINATION CATALYZED BY Ac7, AcDD, ACEE, S101C, AND N137C (AFTER HIS-TAG CLEAVAGE) IN 20 MM HEPES BUFFER PH 7.0 WITH ADDITION OF 0.2 MM $Ca(NO_3)_2$ AND WITHOUT REDUCING AGENT (B-ME). (MT3114-8)..... 205

FIGURE S.8. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY Ac7, AcDD, ACEE, S101C, AND N137C IN 20 MM HEPES BUFFER, PH 7.0 WITH ADDITION OF 0.2 MM $Pb(NO_3)_2$ AND WITHOUT REDUCING AGENT (B-ME). (MT3117-8) 206

FIGURE S.9. MICHAELIS-MENTEN GRAPH OF KEMP ELIMINATION CATALYZED BY Y99E AND Y99C IN 20 MM HEPES BUFFER, PH 7.0, WITH ADDITION OF 0.2 MM $Ca(NO_3)_2$ OR 0.2 MM $Pb(NO_3)_2$ AND WITHOUT REDUCING AGENT (B-ME). 207

FIGURE S.10. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY Ac7, AcDD, ACEE, Y99E, Y99C, S101C, AND N137C (2.5 μ M) IN 20 MM HEPES BUFFER PH 7.0 WITH ADDITION OF 0.2 MM $Y(NO_3)_2$. THE K_{CAT}/K_M VALUES ARE 32 $M^{-1}s^{-1}$ (Ac7), 18 $M^{-1}s^{-1}$ (AcDD), 5 $M^{-1}s^{-1}$ (ACEE), N/A (S101E), 6 $M^{-1}s^{-1}$ (N137E), N/A (S101C), AND N/A (N137C). (MT3131) 208

FIGURE S.11. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY D95C D131C MUTANT (5.0 MM) IN 20 MM HEPES BUFFER WITH ADDITION OF 0.1 MM OR 0.25 MM OF EITHER $Ca(NO_3)_2$, $Pb(NO_3)_2$, OR $Y(NO_3)_2$ 209

FIGURE S.12. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY D95C D131C MUTANT (5.0 mM) IN 20 mM HEPES BUFFER AND 0.25 mM OF EITHER $\text{Ca}(\text{NO}_3)_2$, $\text{Pb}(\text{NO}_3)_2$, OR $\text{Y}(\text{NO}_3)_3$, WITH THE ADDITION OR OMISSION OF 1.2 EQUIVALENCE OF B-MERCAPTOETHANOL. (MT5107) 210

FIGURE S.13. CATALYTIC EFFICIENCY OF KEMP ELIMINATION CATALYZED BY D95E D131E MUTANT (5.0 mM) IN 20 mM HEPES BUFFER WITH ADDITION OF 0.1 mM OR 0.25 mM OF EITHER $\text{Ca}(\text{NO}_3)_2$, $\text{Pb}(\text{NO}_3)_2$, OR $\text{Y}(\text{NO}_3)_2$ 211

FIGURE S.14. pH RATE PROFILE OF HOLLEE (1 μM) WITH 80 μM CALCIUM. AT DIFFERENT pH, THE RATE OF 600 μM FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT pH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT pH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT pH 8.0, 20 mM TRIS BUFFER WAS USED. (AG1049) 212

FIGURE S.15. pH PROFILE OF HOLLEE (1 μM) WITH 80 μM LUTETIUM. AT DIFFERENT pH, THE RATE OF 600 μM FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT pH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT pH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT pH 8.0, 20 mM TRIS BUFFER WAS USED. (AG1049) 213

FIGURE S.16. pH PROFILE OF HOLLEE (1 μM) WITH 80 μM YTTERBIUM. AT DIFFERENT pH, THE RATE OF 600 μM FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT pH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT pH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT pH 8.0, 20 mM TRIS BUFFER WAS USED. (AG1049) 214

FIGURE S.17. pH PROFILE OF HOLLEE (1 μM) WITH 80 μM TERBIUM. AT DIFFERENT pH, THE RATE OF 600 μM FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT

PH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT PH 8.0, 20 mM TRIS BUFFER WAS USED. (AG1049) 215

FIGURE S.18. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M GADOLINIUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT PH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT PH 8.0, 20 mM TRIS BUFFER WAS USED. (AG1049) 216

FIGURE S.19. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M SAMARIUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT PH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT PH 8.0, 20 mM TRIS BUFFER WAS USED.(AG1049) 217

FIGURE S.20. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M NEODYMIUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT PH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT PH 8.0, 20 mM TRIS BUFFER WAS USED.(AG1049) 218

FIGURE S.21. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M PRASEODYMIUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT PH 5.5 AND 6.0 20 mM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 mM MOPS BUFFER WAS USED. AT PH 8.0, 20 mM TRIS BUFFER WAS USED.(AG1049) 219

FIGURE S.22. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M CERIUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 mM ACETATE BUFFER WAS USED. AT PH 5.5

AND 6.0 20 MM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 MM MOPS BUFFER WAS USED. AT PH 8.0, 20 MM TRIS BUFFER WAS USED.(AG1049) 220

FIGURE S.23. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M LANTHANUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 MM ACETATE BUFFER WAS USED. AT PH 5.5 AND 6.0 20 MM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 MM MOPS BUFFER WAS USED. AT PH 8.0, 20 MM TRIS BUFFER WAS USED.(AG1049) 221

FIGURE S.24. PH PROFILE OF HOLLEE (1 μ M) WITH 80 μ M YTTRIUM. AT DIFFERENT PH, THE RATE OF 600 μ M FINAL CONCENTRATION OF KEMP WAS FOUND. AT BUFFERS 4.5-5.5, 20 MM ACETATE BUFFER WAS USED. AT PH 5.5 AND 6.0 20 MM MES BUFFER WAS USED. AT PH 7.0 AND 7.5, 20 MM MOPS BUFFER WAS USED. AT PH 8.0, 20 MM TRIS BUFFER WAS USED.(AG1049) 222

FIGURE S.25. ITC DATA OF ALLEYCAT7 TITRATED WITH $\text{Ca}(\text{NO}_3)_2$. THE SAMPLE CELL CONTAINED 0.1MM AC7 IN 20MM MOPS PH 7.0 AT 25°C AND WAS SLOWLY TITRATED WITH 5MM $\text{Ca}(\text{NO}_3)_2$ IN 5 μ L INJECTIONS WITH AN EQUILIBRATION TIME OF 240 SECONDS. 223

FIGURE S.26. CIRCULAR DICHROISM SPECTRUM OF AC7 (4.75 μ M) WITH 4MM HEPES PH 7.0 IN PRESENCE OF 500 MM Ca^{2+} (RED), 25 MM Ca^{2+} (BLUE), Pb^{2+} (GREEN), AND 25 MM EDTA (BLACK). AC7 DEMONSTRATES AN A-HELICAL STRUCTURE IN THE PRESENCE OF ALL THE METAL IONS TESTED. DESPITE NO Ca^{2+} IONS IN THE EDTA SAMPLE, AC7 KEPT ITS STRUCTURE. (MT3045) 224

FIGURE S.27. CIRCULAR DICHROISM SPECTRUM OF ACDD (4.75 μ M) IN 4MM HEPES PH 7.0 IN PRESENCE OF 500 MM Ca^{2+} (RED), 25 MM Ca^{2+} (BLUE), Pb^{2+} (GREEN), AND 25 MM EDTA (BLACK). ACDD DEMONSTRATES AN A-HELICAL STRUCTURE IN THE PRESENCE OF ALL THE METAL IONS TESTED AS WELL AS EDTA. (MT3045)... 225

FIGURE S.28. CIRCULAR DICHROISM SPECTRUM OF ACEE (4.75 μM) WITH 4mM HEPES PH 7.0 IN PRESENCE OF 500 mM Ca^{2+} (RED), 25 mM Ca^{2+} (BLUE), Pb^{2+} (GREEN), AND 25 mM EDTA (BLACK). ACEE DEMONSTRATES AN A-HELICAL STRUCTURE IN THE PRESENCE OF ALL THE METAL IONS TESTED AS WELL AS EDTA. (MT3045)... 226

FIGURE S.29. CIRCULAR DICHROISM OF HOLLEE (25 μM) IN 4 mM MOPS PH 7.0 WITH YCl_3 TITRATION. INCREASING THE CONCENTRATION OF METAL LEADS TO CHANGES AT 207 NM AND 220 NM THAT CAN BE POTENTIALLY EVALUATED FOR K_D . (AG1018)..... 227

FIGURE S.30. CIRCULAR DICHROISM OF HOLLEE (25 μM) IN 4 mM MOPS PH 7.0 WITH PrCl_3 TITRATION. INCREASING THE CONCENTRATION OF METAL LEADS TO CHANGES AT 207 NM AND 220 NM THAT CAN BE POTENTIALLY EVALUATED FOR K_D . (AG1022)..... 228

FIGURE S.31. CIRCULAR DICHROISM OF HOLLEE (25 μM) IN 4 mM MOPS PH 7.0 WITH LaCl_3 TITRATION. INCREASING THE CONCENTRATION OF METAL LEADS TO CHANGES AT 207 NM AND 220 NM THAT CAN BE POTENTIALLY EVALUATED FOR K_D . (AG1020)..... 229

FIGURE S.32. CIRCULAR DICHROISM OF HOLLEE (25 μM) IN 4 mM MOPS PH 7.0 WITH CaCl_2 TITRATION. (AG1025)..... 230

FIGURE S.33. CIRCULAR DICHROISM OF HOLLEE (25 μM) IN 4 mM MOPS PH 7.0 WITH CeCl_3 TITRATION. INCREASING THE CONCENTRATION OF METAL LEADS TO CHANGES AT 207 NM AND 220 NM THAT CAN BE POTENTIALLY EVALUATED FOR K_D . (MT7012) 231

FIGURE S.34. A 10% ACRYLAMIDE SDS-PAGE GEL OF PURIFICATION OF FDH-N EXPRESSED IN BL21(DE3) *E. COLI* CELLS. ALIQUOTS WERE TAKEN BEFORE IPTG INDUCTION, AFTER IPTG INDUCTION, FROM THE PELLET AFTER SONICATION AND CENTRIFUGATION (LABELED DEBRIS 1 ON GEL), FROM THE PELLET AFTER STREPTOMYCIN SULFATE ADDITION AND CENTRIFUGATION (LABELED DEBRIS 2), FROM THE Ni-NTA COLUMN FLOW THROUGH,

FROM THE NI-NTA COLUMN WASH, AND BEFORE AND AFTER Q-SEPHAROSE COLUMN. THE PROTEIN BAND WAS
~45 kDa, WHERE IT WAS EXPECTED TO APPEAR..... 232

FIGURE S.35. CATALYTIC EFFICIENCY OF FLDH-AERU (10 mM) MEASURING THE CONVERSION OF NAD⁺ (1 mM) TO
NADH DURING THE OXIDATION OF FORMALDEHYDE(100 mM) TO FORMATE IN 100 mM SODIUM PHOSPHATE
BUFFER AT PH 7. THE SPECIFIC ACTIVITY IS 2.6 U/MG. (MT1102)..... 233

FIGURE S.36. CATALYTIC EFFICIENCY OF FLDH-AERU (1 mM) MEASURING THE CONVERSION OF NADH (0.2 mM) TO
NAD⁺ DURING THE REDUCTION OF FORMATE (160 mM) TO FORMALDEHYDE IN 100 mM SODIUM PHOSPHATE
BUFFER AT PH 7. THE SPECIFIC ACTIVITY IS 0.26 U/MG. (MT1102)..... 234

FIGURE S.37. CATALYTIC EFFICIENCY OF FLDH-PUTI (0.1 mM) MEASURING THE CONVERSION OF NAD⁺ (1 mM) TO
NADH DURING THE OXIDATION OF FORMALDEHYDE (100 mM) TO FORMATE IN 100 mM SODIUM PHOSPHATE
BUFFER AT PH 7. THE SPECIFIC ACTIVITY IS 3.13 U/MG. (MT1146)..... 235

FIGURE S.38. CATALYTIC EFFICIENCY OF FLDH-PUTI (12 mM) MEASURING THE CONVERSION OF NADH (0.1 mM)
TO NAD⁺ DURING THE REDUCTION OF FORMATE (100 mM) TO FORMALDEHYDE IN 100 mM SODIUM
PHOSPHATE BUFFER AT PH 7. THE SPECIFIC ACTIVITY IS 0.0003 U/MG. (MT1146) 236

FIGURE S.39. CATALYTIC EFFICIENCY OF ADH (16 nM) MEASURING THE CONVERSION OF NADH (0.2 mM) TO NAD⁺
DURING THE REDUCTION OF FORMALDHYDE (100 mM) TO METHANOL IN 100 mM SODIUM PHOSPHATE BUFFER
AT PH 7. THE SPECIFIC ACTIVITY IS 84 U/MG. (MT1064) 237

FIGURE S.40. 15% ACRYLAMIDE SDS-PAGE GEL OF Ac0 AND Ac0-V108E PURIFICATION. THE EXPECTED MW OF
Ac0 AND Ac0-V108E UNCUT IS ~12 kDa AND THE EXPECTED MW OF THE CUT PROTEINS IS ~8 kDa. 239

FIGURE S.41. 15% ACRYLAMIDE SDS-PAGE GEL OF WILD-TYPE (WT; cCAM) AND WT V108E (CV108E). DEBRIS REFERS TO THE PELLETT LEFT BEHIND AFTER SONICATION AND CENTRIFUGATION, AND FLOW AND WASH REFER TO STEPS IN NI-NTA PURIFICATION. THE EXPECTED MASSES OF THE UNCUT PROTEINS IS ~12 kDA..... 240

LIST OF SCHEMES

SCHEME 1. KEMP ELIMINATION REACTION.....	48
SCHEME 2. OVERVIEW OF ENZYMATIC CONVERSION OF CARBON DIOXIDE TO METHANOL.	120

LIST OF TABLES

TABLE 2.1. SUMMARY OF ACTIVITY OF MUTATIONS TO POSITION 9 AFTER THE ADDITION OF 0.1 mM $\text{Ca}(\text{NO}_3)_2$, 0.1 mM $\text{Pb}(\text{NO}_3)_2$, OR 0.1 mM $\text{Y}(\text{NO}_3)_3$. ALL ASSAYS WERE RUN IN 20MM HEPES AT PH 7.0 WITH 5.0 μM PROTEIN AND CAN BE FOUND IN TABLE S.7-S.12.....	57
TABLE 2.2. SUMMARY OF ACTIVITY OF POSITION 7 MUTANTS AFTER ADDITION OF 0.1 mM $\text{Ca}(\text{NO}_3)_2$, 0.1 mM $\text{Pb}(\text{NO}_3)_2$, OR 0.1 mM $\text{Y}(\text{NO}_3)_3$. ALL ASSAYS WERE RUN IN 20MM HEPES AT PH 7.0 WITH 5.0 μM PROTEIN.	58
TABLE 2.3. ACTIVITY OF HOLLEE AND RECCes IN PRESENCE OF CALCIUM, LEAD, AND YTTRIUM. ALL ASSAYS WERE RUN IN 20MM HEPES AT PH 7.0 WITH 5.0 μM PROTEIN.	60
TABLE 2.4. SUMMARY OF HOLLEE ACTIVITIES IN 20 mM MOPS, 100 mM NaCl AT PH 7.0. THE CONCENTRATION OF ALL METALS WAS 0.08 mM UNLESS OTHERWISE NOTED. (MT7081).....	64
TABLE 4.1. SUMMARY OF SPECIFIC ACTIVITIES OF ENZYMES FOR FDH.....	131
TABLE 4.2. COMPARISON OF SPECIFIC ACTIVITY OF SECRETED FDH IN THE PRESENCE OF PROTEASE INHIBITORS. ..	136
TABLE 4.3. SUMMARY OF SPECIFIC ACTIVITIES OF FLDH PROTEINS AND ADH IN THE CONVERSION OF NADH. KINETIC ASSAYS CAN BE FOUND IN S.34 TO S.38.	138
TABLE 5.1. SUMMARIZED DATA OF ALLEYCAT MUTANTS WITH AND WITHOUT T IN 128 POSITION. Ac0 AND Ac0T WERE AT A CONCENTRATION OF 20 μM ; Ac6, Ac6A, Ac7, AND Ac7A WERE AT A CONCENTRATION OF 0.2 μM ; ALL OTHER PROTEINS WERE AT A CONCENTRATION OF 4.0 μM	163
TABLE 5.2. SUMMARIZED DATA OF V108E MUTANTS WITH AND WITHOUT GLU IN 108 POSITION. ALL PROTEINS WERE AT 20 mM CONCENTRATION.	168
TABLE S.1.1. SEQUENCING FOR CHAPTER 2	195

TABLE S.1.2. SEQUENCING FOR CHAPTER 3	196
TABLE S.1.3. SEQUENCING FOR CHAPTER 4	196
TABLE S.1.4. SEQUENCING DATA FOR CHAPTER 5	198
TABLE S.3.2.1. LIST OF PEPTIDE FRAGMENTS OF FDH-C-Y AFTER TRYPSIN DIGESTION.	237

LIST OF EQUATIONS

EQUATION 1.1. MICHAELIS-MENTEN EQUATION.....	8
EQUATION 5.1. EQUATION USED TO DETERMINE THE CHANGE IN CHEMICAL SHIFT BETWEEN ANALOG BOUND AND NONBOUND ALLEYCAT7 WHEN DETERMINING HOT AND COLD SPOTS.	159

List of Abbreviations

ADH	Alcohol dehydrogenase
Amp	Ampicillin
APS	Ammonium persulfate
BMMY	Buffered methanol-complex media
BSA	Bovine Serum Albumin
Boc	tert-Butyloxycarbonyl
CD	Circular dichroism
Chl	Chloramphenicol
Dap	Diaminopropionic acid
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
FDH	Formate dehydrogenase
FIDH	Formaldehyde dehydrogenase
Fmoc	Fluorenylmethyloxycarbonyl
His-tag	Hexahistidine tag
HCTU	1,1,3,3-tetramethyluronium hexafluorophosphate
Hepes	(4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid

HCC	Hepatocellular carcinoma
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kan	Kanamycin
Kemp	5-nitro-1,2-benzisoxazole
LB	Luria-Bertani broth
mAb	Monoclonal antibodies
MALDI-TOF	Matrix-assisted laser desorption/ionization Time of Flight
Mes	2-(N-morpholino)ethane sulfonic acid
MM	Minimal media
Mops	3-(N-morpholino)propane sulfonic acid
NAD ⁺	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
Ni-NTA	Nickel Nitrilotriacetic acid resin
NMR	Nuclear magnetic resonance
<i>P. aeru</i>	<i>Pseudomonas aeruginosa</i>
<i>P. putida</i>	<i>Psuedomonas putida</i>
PAGE	Polyacrylamide gel electrophoresis
PES	Poly ether sulfone
RDB-	Regeneration dextrose base (no histidine added)
<i>P. pastoris</i>	<i>Pichia pastoris</i>

PMSF	Phenylmethylsulfonyl fluoride
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	Sodium dodecyl sulfate
TACE	Transarterial chemoembolization
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TIS	Triisopropylsilane
Tris	Tris(hydroxymethyl)amino methane
Ts	Tosyl
UV-Vis	Ultraviolet-visible spectroscopy/spectrometer
YNB	Yeast nitrogen base
Zym	Zym-5052 autoinduction media

Chapter 1 Introduction

1.1. Enzymes

1.1.1. Defined

Catalysts are capable of enabling otherwise kinetically unfavorable reactions to occur more quickly and readily.¹⁻⁴ They are characterized by their ability to perform this function without changing the thermodynamic properties of the reaction or being consumed during the catalytic process. They are one of the most powerful available tools in both the laboratory and in the natural world. Many natural catalysts can be classified as enzymes, which are made up of either amino acids or RNA (these are known as ribozymes). For the purposes of this dissertation, we will only be discussing enzymes that are composed of amino acids or peptides, which are shorter sequences of amino acids.

Enzymes and other catalysts function by lowering the energy required for a reaction to occur.^{1-2, 5} Every chemical reaction carries with it an activation energy threshold that must be overcome in order for the product to be formed (Figure 1.1).¹⁻⁴ When a reaction is occurring in favorable conditions, such as in a heated, acidic, or basic environment, or the reactant is highly unstable, the reaction can often proceed forward without intervention. However, the presence of an enzyme can help the reaction along by binding the reactant (or substrate) and changing the conditions of the reaction to make it occur more easily.²

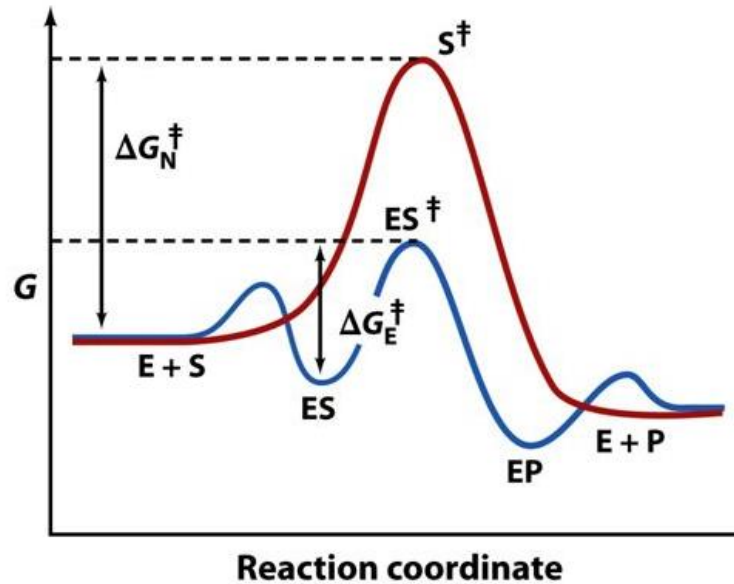


Figure 1.1. Transition state diagram of an uncatalyzed (red) and enzymatically catalyzed (blue) exothermic reaction.^{1-4, 7} ES refers to the enzyme-substrate complex that is formed when the substrate binds to the enzyme. ES^\ddagger refers to the transition state of the substrate as it is bound to the enzyme. EP is the enzyme-product complex that is formed before the product is released from the protein. ΔG^\ddagger is the Gibbs free energy required for the substrate to reach the transition state between the substrate and product.

Enzymes are capable of performing this task in a number of ways. All enzymes have what is known as an 'active site' where a substrate can bind.^{1-4, 7} Binding to this active site can change the environment the substrate is in, can change its conformation, or can weaken bonds. The active site also serves the purpose of making an enzyme selective; many enzymes are only capable of catalyzing one reaction with one specific substrate to give one specific product. This selectivity serves to allow the enzyme to regulate how much of a substance is in the environment and when the catalyzed reaction will take place.

There are two popular theories about how a substrate binds in the active site (Figure 1.2). The first is known as the lock-and-key theory, theorized by Emil Fischer in 1894.⁸⁻⁹ It hypothesizes that each substrate fits perfectly with the enzyme like a key will fit into only one lock and vice

versa. While this theory is able to explain the specificity of enzymes for substrates, it does not explain how an enzyme can have multiple substrates, nor does it explain how an enzyme is able to stabilize the transition state of the substrate.^{8, 10}

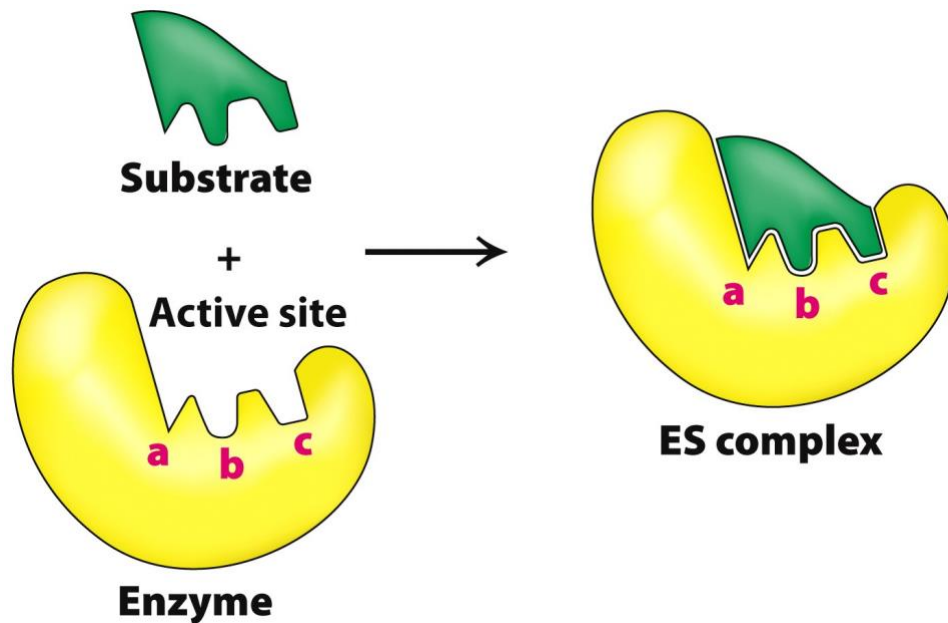


Figure 1.2. Lock-and-key model of the active site hypothesizes that the active site (a, b, and c) of the protein (yellow) is specific to the substrate (green) and they fit precisely to form the enzyme-substrate complex.⁸

The second theory is called “induced fit” proposed by Daniel Koshland in 1958 (Figure 1.3).¹ The theory suggests that the enzyme and its active site are able to change conformation when needed. This allows the enzyme to fit around different substrates as needed and also to change the conformation of substrate to allow the reaction to proceed.

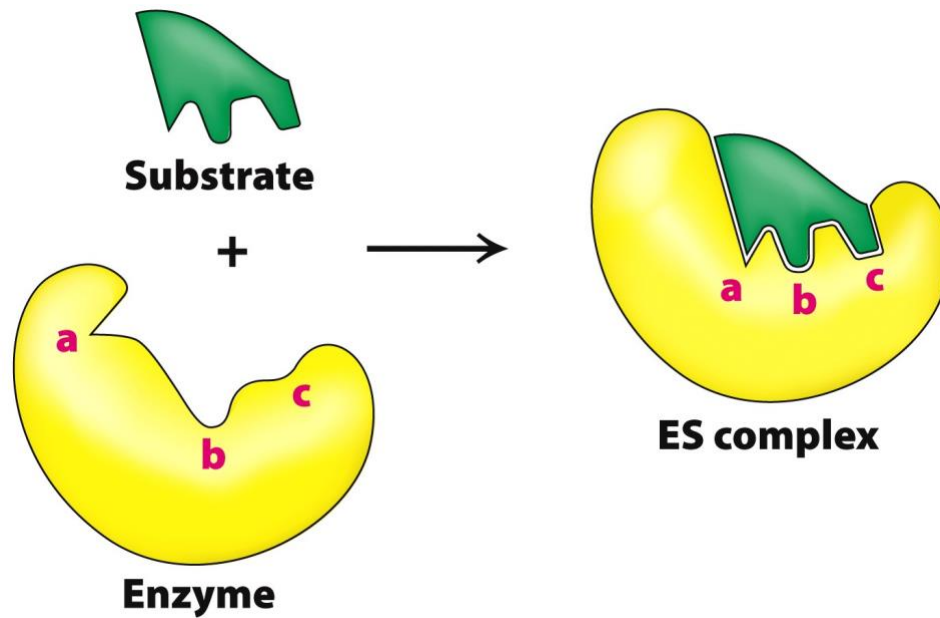


Figure 1.3. Induced fit model suggests that the protein (yellow) is able to change shape fluidly and form around the substrate (green) so that the points of contact in the active site (a, b, and c) fit with the substrate.¹⁰⁻¹¹

Still others have proposed that both theories are at least partially true.¹⁰⁻¹¹ A flexible lock-and-key motif suggests that the active site is rigid and fits substrates specifically. However, the substrate is also able to alter the active site during the reaction and the substrate can even be flexible.

The active site is not the only part of the enzyme that plays an important role in catalysis. Proteins also utilize cofactors, which are either non-protein chemical compound or metallic ions (Figure 1.4).¹²⁻¹³ Non-protein chemical compounds can refer to complex organic molecules such as vitamins (known as coenzymes) or inorganic ions such as phosphate or iron-sulfur clusters.

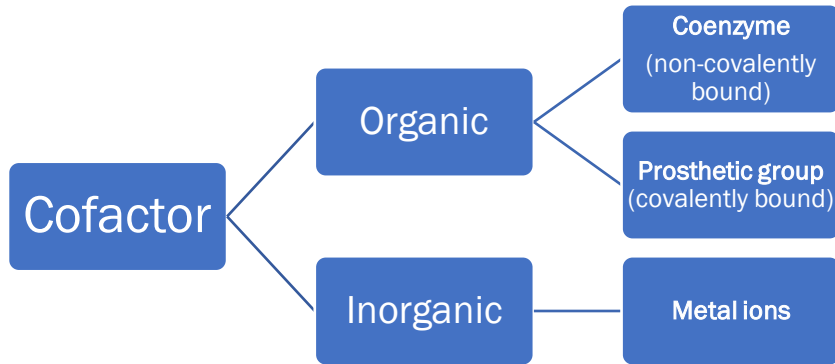


Figure 1.4. How cofactors are categorized. Organic molecules are usually derived from vitamins and can be classified as either coenzymes (if non-covalently bound) or as prosthetic groups (if covalently bound).¹³ Proteins with metal ion cofactors are also known as metalloproteins and makeup approximately 40% of all known proteins.¹²

Often an enzyme will not even be considered ‘active’ unless the cofactor is present (Figure 1.5).^{12, 14} Cofactors can help a reaction proceed by allowing a protein to attain the right conformation for catalysis, providing energy, or storing and releasing electrons. Cofactors are also another way that the organism is able to regulate the activity of enzymes within itself.^{12, 14}

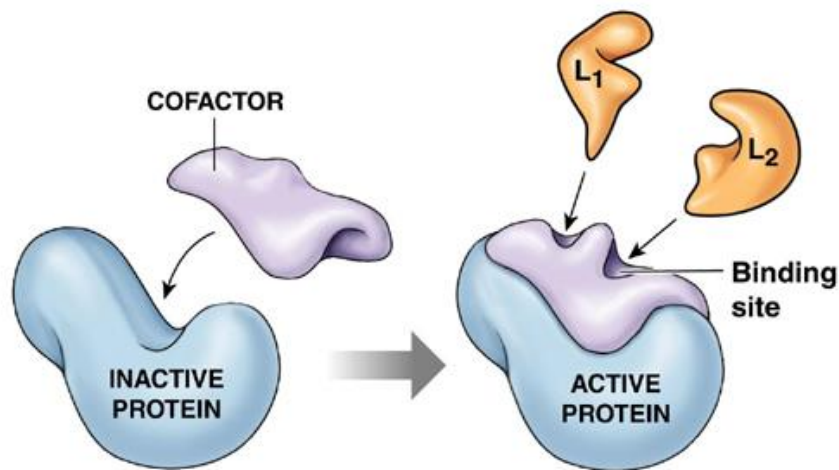


Figure 1.5. An example of enzyme that requires a cofactor for catalysis.³⁻⁴ In this example, the active site is on the cofactor (lavender) and can only be formed properly when bound to the protein (light blue). Other ways cofactors are used by proteins include providing energy for the reaction, altering conformations, and acting as a means of maintaining equilibrium

Many proteins also contain an allosteric domain.³⁻⁴ This is a domain separate from the active site where cofactors and coenzymes can bind (Figure 1.6). Once the cofactor or coenzyme are bound, the structure of the protein is changes. Allosteric domains are used to either activate or inhibit an enzyme depending upon how the structure changes and if the structure allows for catalysis.^{3-4, 12, 15}

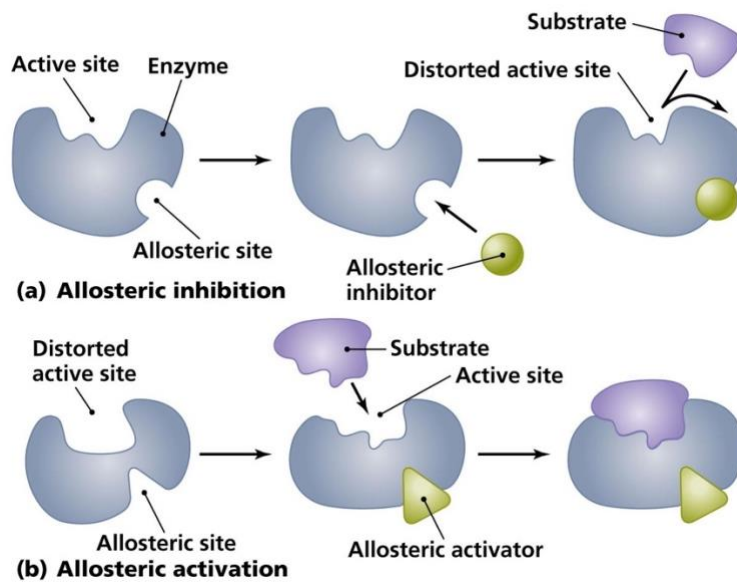


Figure 1.6. A diagram of allosteric inhibition and activation.^{1, 3, 16-18} Allosteric inhibition (a) is when an allosteric inhibitor binds to the allosteric site and prevents the reaction from occurring by distorting the shape of the active site so the substrate cannot bind. Allosteric activation (b) is when an allosteric activator binds to the allosteric site and allows for the active site to bind to substrates.

Finally, competing substrate analogs can reduce or eliminate activity of enzymes.^{1, 3, 16-18} These analogs are known as inhibitors and can either bind directly to the active site (competitive or irreversible) or allosterically (uncompetitive or noncompetitive) (Figure 1.7).^{1, 18} An inhibitor that can be removed from an active site is known as competitive and one that cannot be removed from the active site is an irreversible inhibitor.^{4, 16, 18-19} Uncompetitive inhibition refers to when an inhibitor binds to the enzyme-substrate (ES) complex and cannot be reversed by increasing

the concentration of substrate.^{4, 17-18} Noncompetitive inhibition reduces the activity of an enzyme and is able to bind to a protein both in the presence and absence of substrate.^{3, 18} Usually, the noncompetitive inhibitor will prefer either the enzyme state or the ES complex, which is a phenomenon known as mixed inhibition.¹

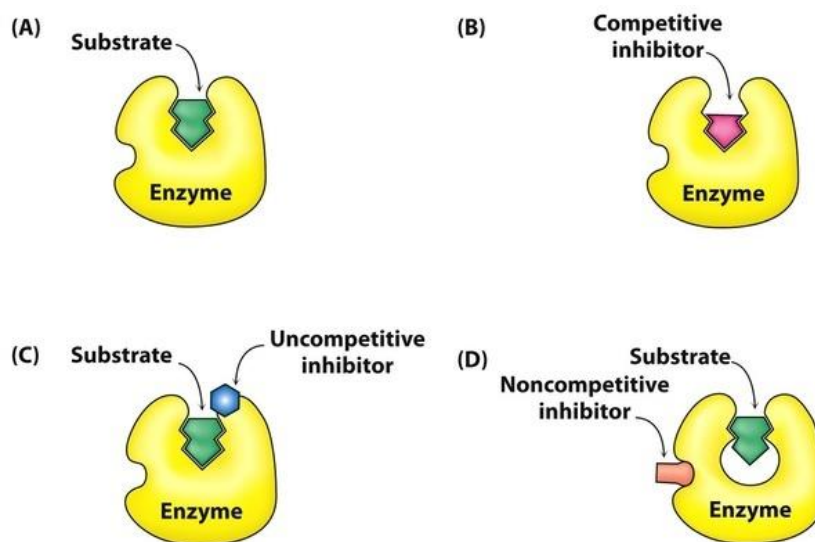


Figure 1.7. The different types of inhibition^{1-4, 20} (A) is an example of a protein and substrate in the absence of any type of inhibition. (B) is a diagram of competitive inhibition where the inhibitor occupies the active site. (C) is uncompetitive inhibition where the inhibitor binds outside of the active site to stop a reaction. (D) is noncompetitive inhibition where an inhibitor binding outside of the active site reduces the activity of the enzyme.

1.1.2. Enzyme kinetics

Enzyme catalysis can be analyzed in a number of ways. The most popular method is through the Michaelis-Menten equation (Equation 1.1 and Figure 1.8).^{1-4, 20} This equation relates the reaction velocity (v_0) to the product turnover per unit time (k_{cat}), the initial concentration of enzyme ($[E]_0$), the concentration of substrate ($[S]$), and the concentration of substrate when the

reaction is at half of its maximal velocity (K_M). The product turnover multiplied by the concentration of enzyme also gives the maximal velocity (v_{max}) that the protein can accomplish.

Equation 1.1. Michaelis-Menten equation

$$v_0 = \frac{k_{cat}[E]_0[S]}{K_M + [S]} = \frac{v_{max}[S]}{K_M + [S]}$$

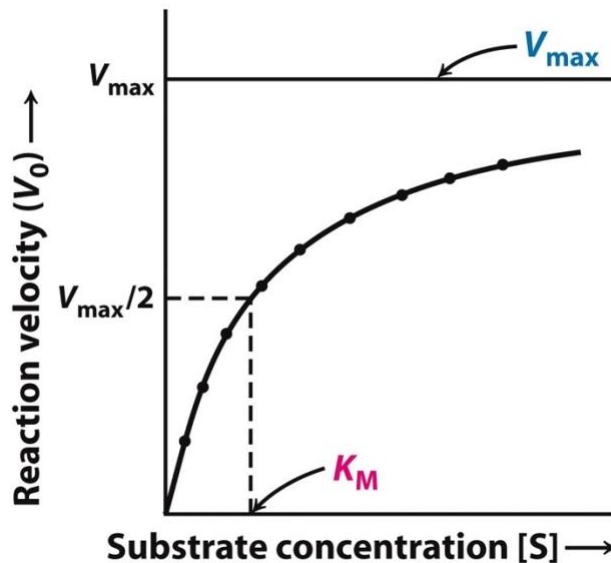


Figure 1.8. Michaelis-Menten equation.¹

1.1.3. Enzyme structure

As mentioned previously, proteins and enzymes are made up of 20 different amino acids (Figure 1.9) that bind together in long polypeptide chains. Each amino acid is indicated by either a three letter or one letter code. In addition to the common 20 amino acids, there are also several amino acids that occur much less frequently in nature. Examples include selenocysteine (a

cysteine containing a selenium atom instead of a sulfur atom) and ornithine (a lysine with one less carbon in its side chain).

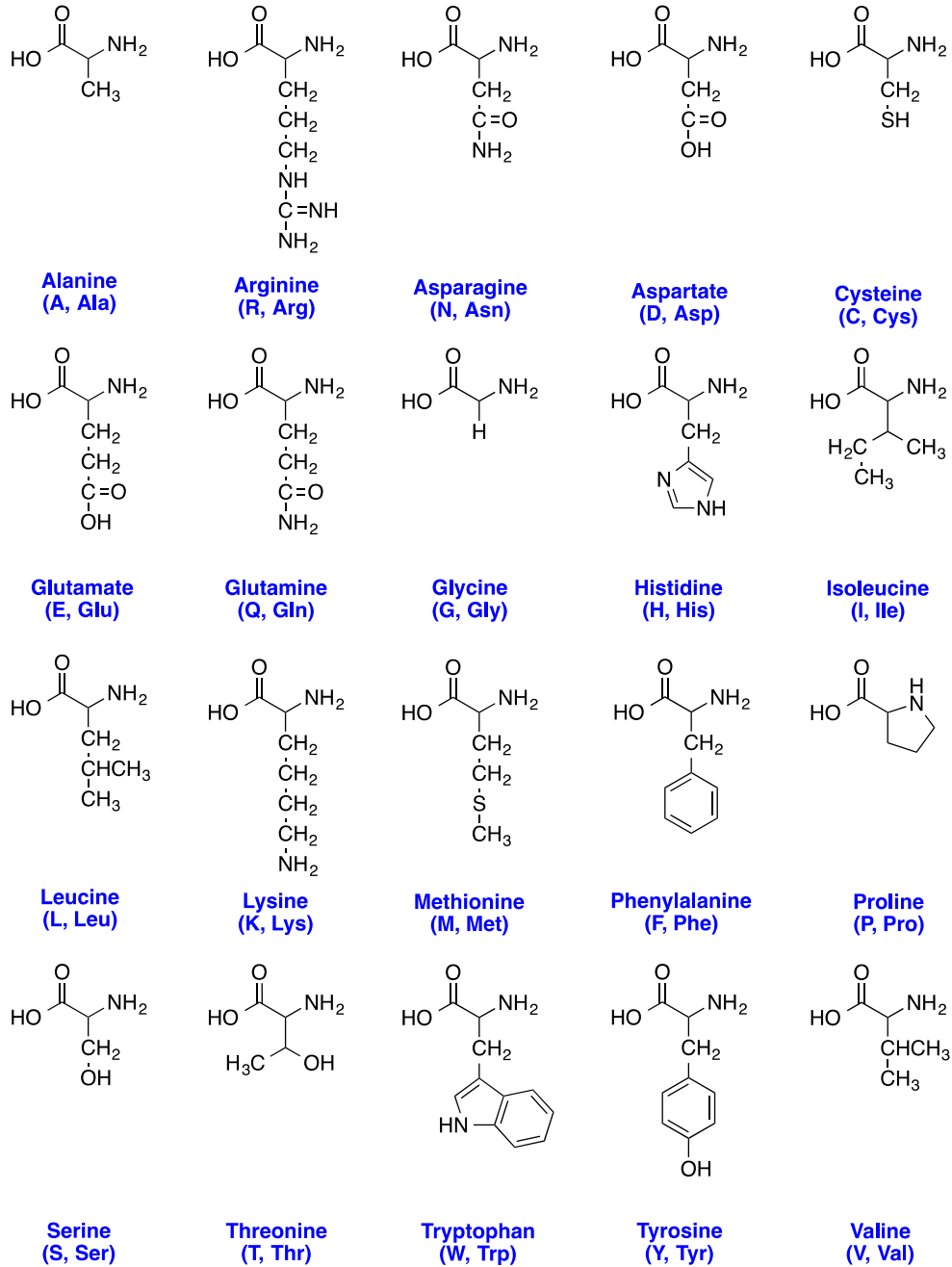


Figure 1.9. The 20 most common amino acids found in nature. Each amino acid can be specified with either a one letter or three letter code. The amino acids link together via peptide bonds to form long polypeptide chains that can become proteins.

Enzyme structure consists of four structural elements: primary, secondary, tertiary, and quaternary (Figure 1.10).^{1-4, 7, 21} The primary structure is the linear amino acid sequence of the polypeptide chain, including any posttranslational modifications the sequence might have undergone.³ The amino acids join together to form amide bonds that link together from the N-terminus to the C-terminus. The primary sequence of the protein dictates the kinds of secondary and tertiary folds that will ultimately form.

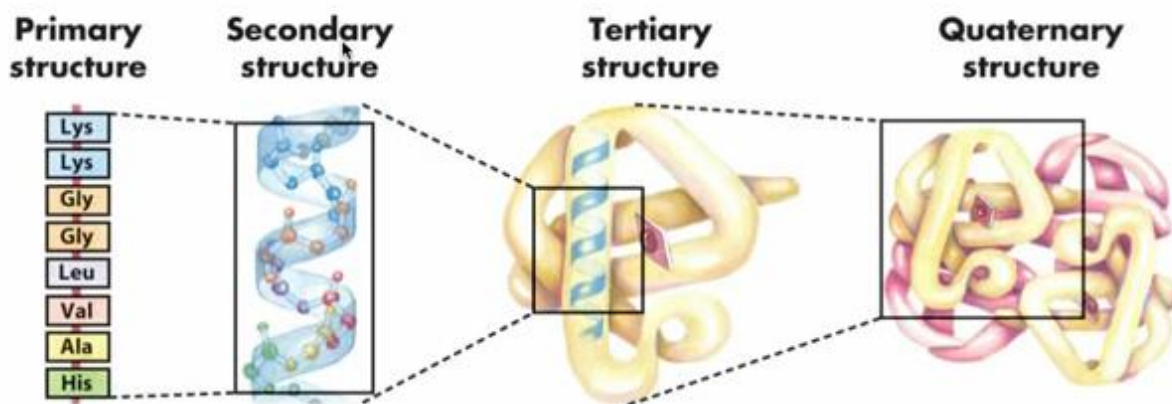


Figure 1.10. A summary of the categories of protein structure: primary, secondary, tertiary, and quaternary.²²⁻²⁵ The primary structure is the linear amino acid sequence of the protein. The secondary structure is the local structure of the linear segments due to hydrogen bonds. The tertiary structure is the three-dimensional arrangement of all the amino acids in the polypeptide chain, including the side chains. The quaternary structure is the arrangement of single polypeptide chains (subunits) into a multi-subunit protein.

The secondary structure is the three dimensional fold that the sequence undergoes due to hydrogen bonds and without regard for the conformation of the side chains (Figure 1.11).²²⁻²⁵ These structures can usually be classified as either an α -helix or a β -sheet and often a protein can contain multiple of each.²² If the structure does not fit into either of these classifications or falls between them then it is called a random coil.²²

The first structure, α -helix, is a spiral conformation where the NH and C=O bonds of the main chain or backbone form hydrogen bonds.²² A single turn of the helix consists of 3.6 amino acids and an α -helix usually contains at least three turns total. In addition to the hydrogen bond, there are also van der Waals interactions along the helical axis. The α -helix is the most stable of the secondary structures and the most common.

The β -sheet occurs when the polypeptide chains arrange themselves in parallel ribbons that connect via hydrogen bonds between the NH and C=O backbone.⁷ The amino acid side chains lie perpendicular to the main chains where the hydrogen bonds occur. Each strand is usually five to ten amino acids long.

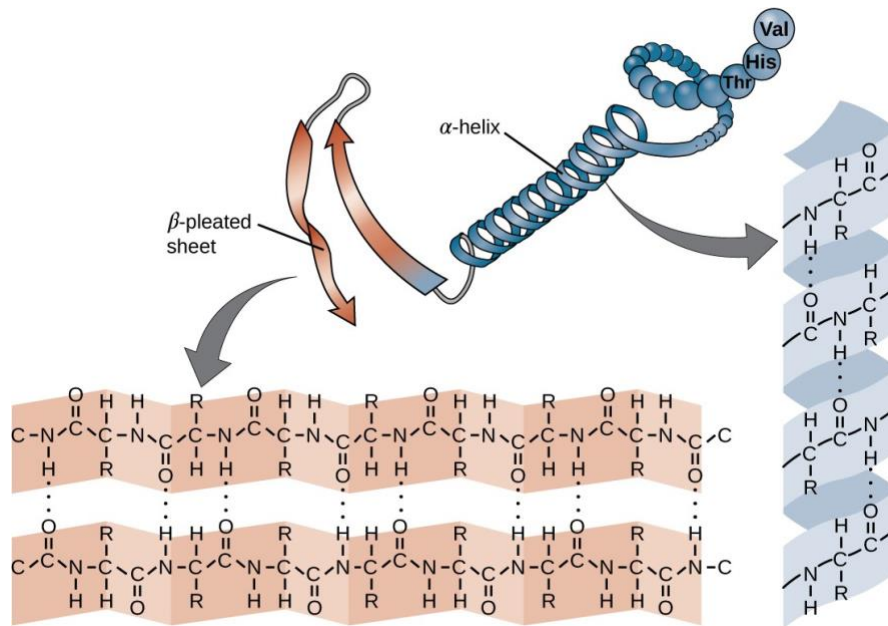


Figure 1.11. The secondary structure of a protein can either be categorized as an α -helix or a β -sheet.^{7, 22, 24, 26}

The tertiary structure of a protein refers to the intermolecular interactions in the protein between its side chains (Figure 1.12).^{7, 22, 24, 26} This includes disulfide bonds between cysteine residues, hydrogen bonds between polar amino acids, and van der Waals interactions. The

primary sequence is the sole dictator of tertiary structure.^{7, 22} The quaternary structure of a protein is the result of multiple subunits of the same protein coming together via intramolecular interactions (Figure 1.13).^{7, 22} These subunits can either be identical or have varying structures.

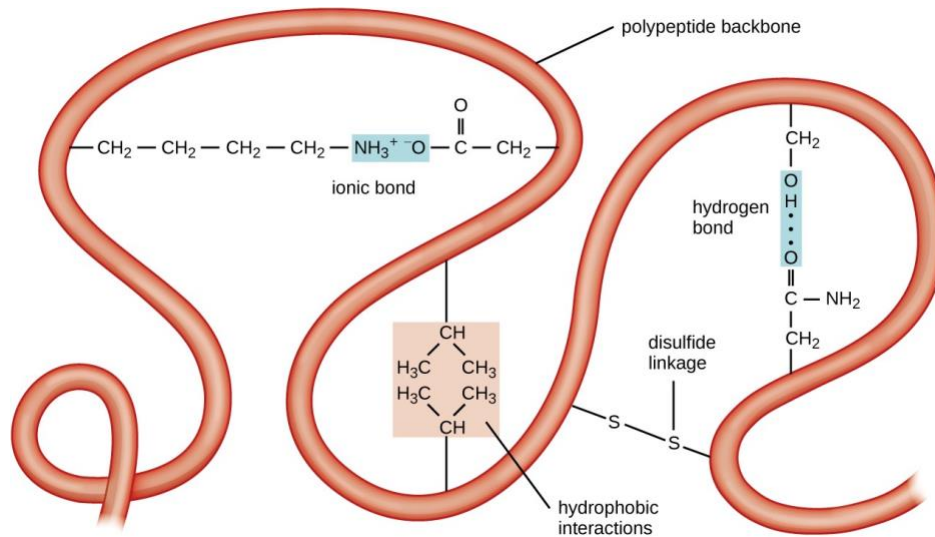


Figure 1.12. The tertiary structure of proteins consists of intermolecular interactions such as ionic bonds or salt bridges, disulfide bonds, hydrophobic interactions, and hydrogen bonding.⁷

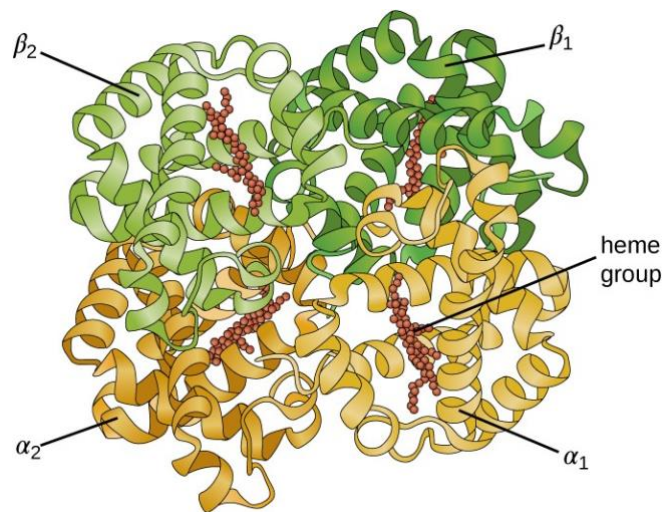


Figure 1.13. Quaternary structure of hemoglobin.²² This protein consists of four subunits complexed together into a final structure.

Two proteins are considered similar, or homologous, if they share over 40% of the same sequence.²⁸ They tend to have similar secondary, tertiary, and quaternary structures despite changes in the primary sequence. It is common to have homologous proteins between species; for example, calmodulin, a calcium sensing protein, is present in all eukaryotes and functions similarly between them despite varying primary structures.²² However, some mutations can be more favorable than the parent sequence, which can result in new structures that are more thermostable and/or catalytic.²⁹⁻³⁴ This is known as divergent evolution.

1.2. Protein Engineering

Enzymes have many advantages in comparison to organic and inorganic catalysts. They are environmentally friendly because they can function in water at low temperature and are biodegradable.²⁹⁻³⁴ In addition, enzymes are also stereoselective, have greater substrate specificity, can have their activity regulated, and can often perform in neutral conditions.³⁵⁻³⁷ Furthermore, enzymatic catalysts are easy to obtain through by introducing the genetic code into a variety of organism (this is discussed in Chapter 4).³⁸⁻⁴¹ However, while enzymes are an excellent alternative to organic and inorganic catalysts, there are a limited number of proteins that are actually capable of doing reactions that we need.²⁹⁻³⁴ Protein design strives to not only create new enzymes for these purposes, but also optimize natural enzymes for greater stability, catalytic activity, stereoselectivity, or even metal binding (which is detailed in Chapter 2 and 3).²⁹⁻³⁴ Protein engineers also have additional goals, such as creating new structural motifs and understanding how evolution of enzymes occurred.^{33, 42-43}

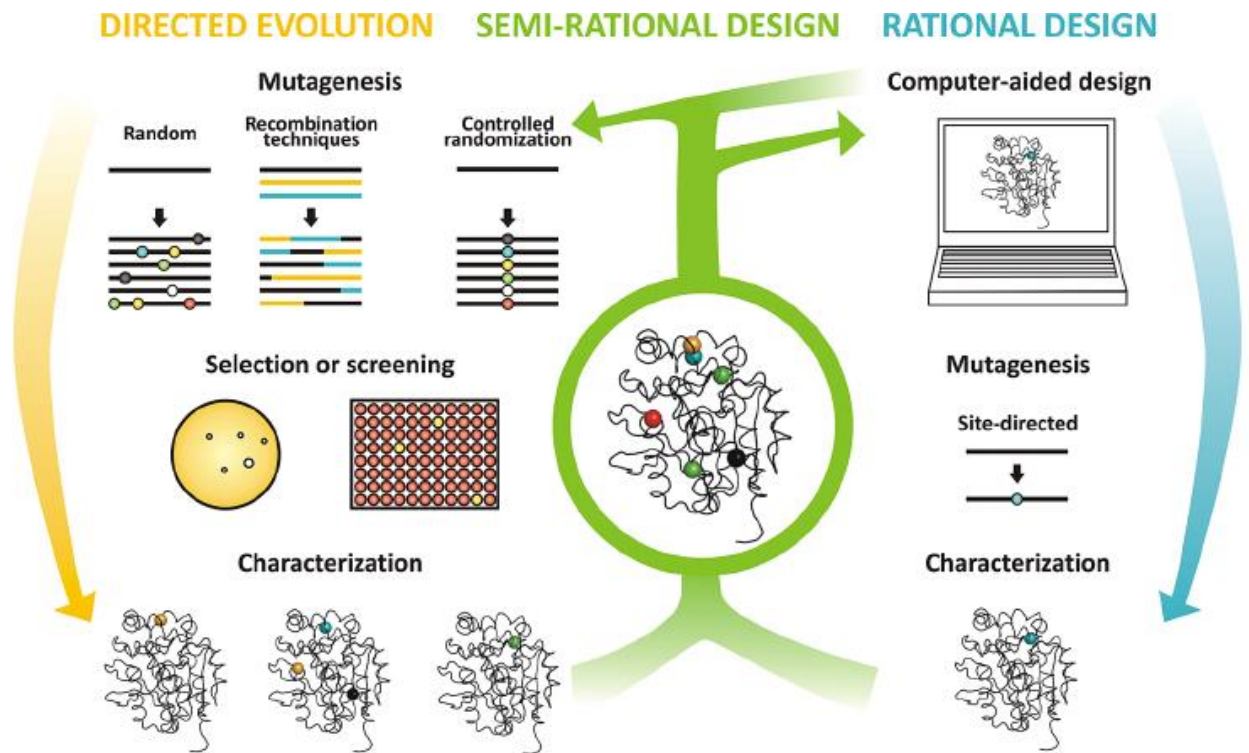


Figure 1.14. Protein engineering strategies.⁴⁴ Directed evolution refers to introducing mutations randomly and screening to see which mutations are successful. Rational design uses computational algorithms to determine the most stable mutations of a protein and then characterizes them. Semi-rational design is a combination of the two methods.

Protein engineering typically falls into one of three categories: directed evolution, rational design, and semi-rational design (Figure 1.14).³ Each of these approaches requires the use of mutagenesis to alter the genetic code and thus alter the protein sequence. The next section details how one is able to alter the amino acid sequence of a protein and the various methods of mutating said sequence.

1.2.1. Polymerase chain reaction and mutagenesis

Proteins are encoded into DNA using codons. DNA is made up of four different nucleic acids: adenine (A), guanine (G), cytosine (C), and thymine (T). A codon is a group of three nucleotides that is read by the cell and transcribed into a specific amino acid (Figure 1.15). Some amino acids, such as leucine, have multiple codons for the same amino acid, while others like tryptophan only have one.⁴⁵ Methionine is known as the 'start' codon and is the codon that triggers protein synthesis, or translation. Stop codons signal the end of the protein and stop translation.

		Second Position									
		T		C		A		G			
		code	Amino acid	code	Amino acid	code	Amino acid	code	Amino acid		
FIRST POSITION	T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T	THIRD POSITION
		TCT		TCC		TAC		TGC		C	
		TAT	Leu	TCA		TAA	stop	TGA	stop	A	
		TGT		TCG		TAG	stop	TGG	Trp	G	
	C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T	
		CTC		CCC		CAC		CGC		C	
		CTA		CCA		CAA	CGA	A			
		CTG		CCG		CAG	CGG	G			
	A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T	
		ATC		ACC		AAC		AGC		C	
		ATA		ACA		AAA	AGA	A			
		ATG	Met	ACG		AAG	Lys	AGG	Arg	G	
	G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T	
		GTC		GCC		GAC		GGC		C	
		GTA		GCA		GAA	GGA	A			
		GTG		GCG		GAG	GGG	G			

Figure 1.15. Codon Table. Each three-letter sequence of nucleotides encodes for a protein. Many amino acids have multiple codons.

The most common method of genetic sequencing is Sanger sequencing (Figure 1.16).⁴⁵ A single stranded DNA is incubated with labelled nucleotides (dNTPs), ddNTPs (nucleotides lacking the 3'-OH group necessary for chain elongation), and a DNA polymerase. The ddNTPs are usually labelled with a fluorescent dye or a radioactive phosphorous. By looking at where the DNA fragments have stopped synthesizing and the size of the DNA, one is able to piece together the DNA sequence of the gene.

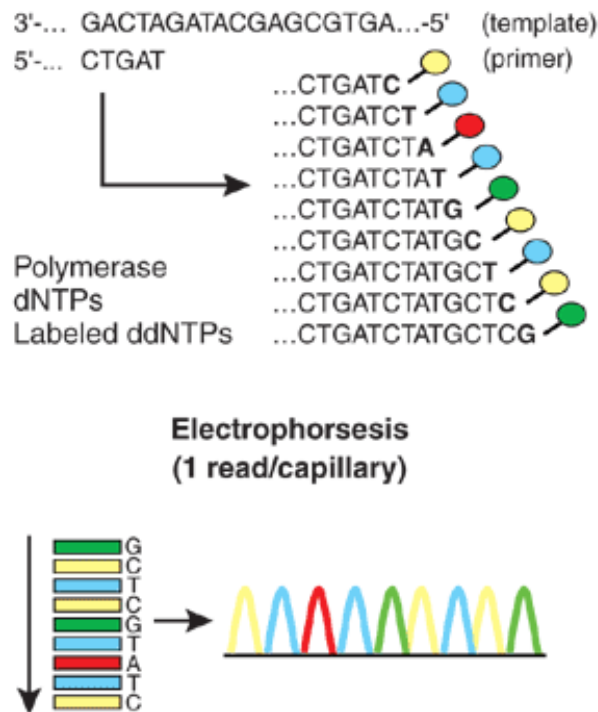


Figure 1.16. Description of Sanger sequencing.^{2-4, 23} ddNTPs are tagged and mixed with the template strand. If a tagged ddNTP binds to the strand, then the sequence stops. In this way, one can eventually build a map of where each codon and nucleic acid lie in respect to the gene.

In 1985, the polymerase chain reaction (PCR) was published as a way to create many copies of a segment of DNA (Figure 1.17).^{2-4, 23} PCR consists of three main steps that are repeated multiple times. The first step, denaturation, applies heat to the sample to separate the double stranded DNA into single strands. The next step, annealing, allows for custom made primers to bind to complementary sequences on the two single strands. The temperature is usually set to be a few degrees lower than the melting temperature of the primers, or the temperature at which the primer would dissociate into two single strands of DNA. The final step of the process, replication, allows for the complementary strands of new DNA to be polymerized starting from the primers. Repetition of these three steps allows one to obtain a lot of copies of a gene in a short amount of time.

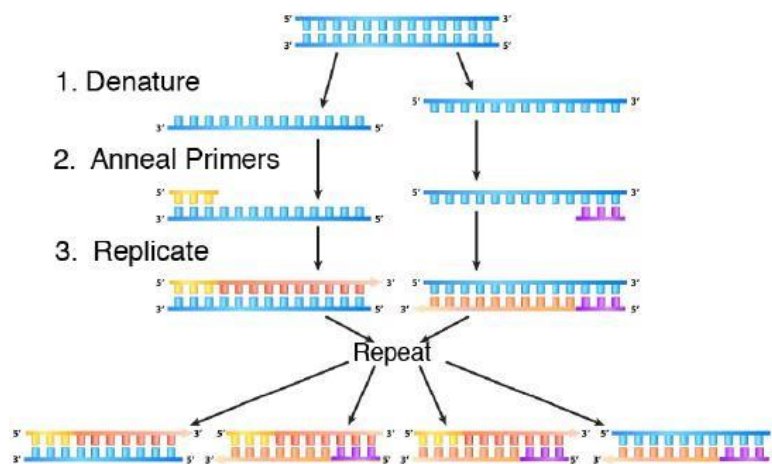


Figure 1.17. Process of polymerase chain reaction.^{3, 46} PCR consists of three main steps: denaturation, annealing, and replication. Over time, PCR yields multiple copies of a desired segment of DNA.

PCR allows for specific mutations to be made to a gene to alter protein sequence. By customizing the primers to contain different codons than the template strand, multiple copies of

new DNA are created. When done purposefully, this process is known as site-directed mutagenesis.^{3,46} Site-directed mutagenesis allows for base pair substitution and/or introduction of a new amino acid, insertion of new base pairs, or the deletion of specific base pairs (Figure 1.18). Site-directed mutagenesis can also randomize an amino acid in a specific position by using a primer solution that contains a mixture of all dNTPs at that codon position, which is known as random mutagenesis.

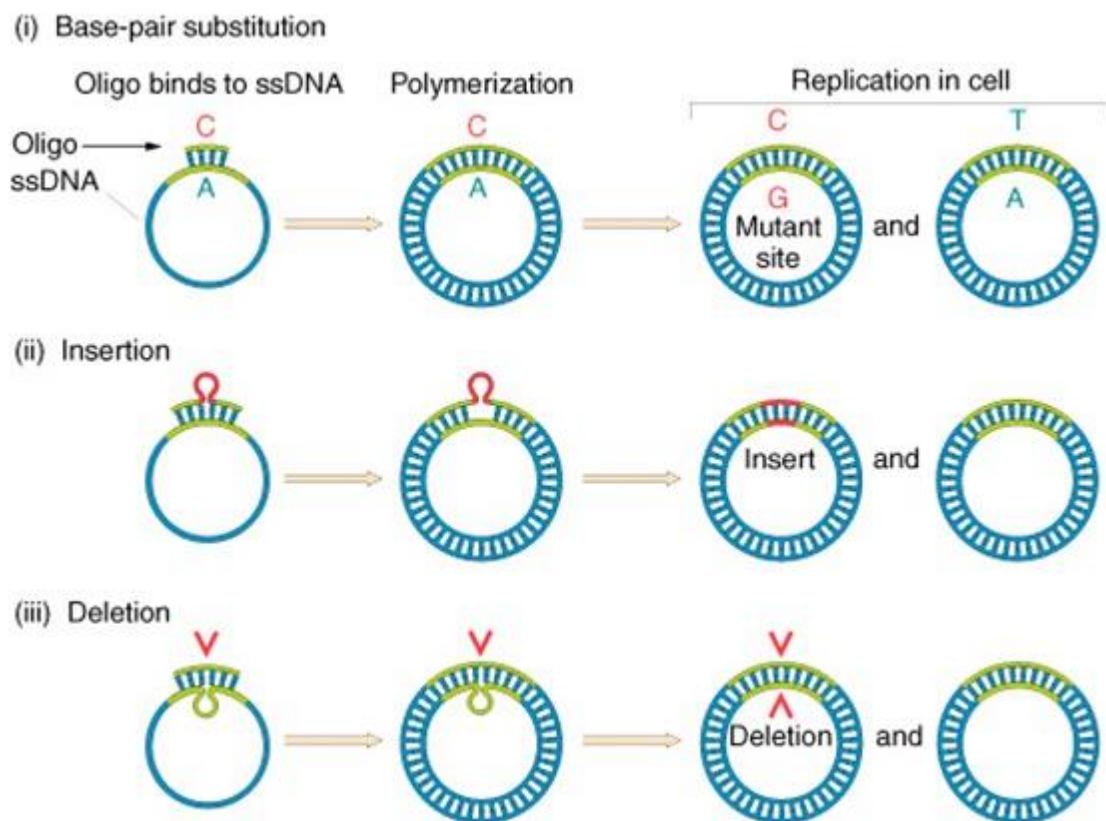


Figure 1.18. Site-directed mutagenesis by using base-pair substitution, insertion, and deletion.⁴⁷

Error-prone PCR is used as a non-specific method of directed evolution.⁴⁷ Error-prone PCR introduces random mutations into the plasmid using DNA polymerase that is prone to errors (Figure 1.19). Unfortunately, while error-prone PCR is good for creating a large library in multiple

positions of the plasmid quickly, since some amino acids only have one or two codons encoding them, there is not an equal distribution of amino acids within the mutants.

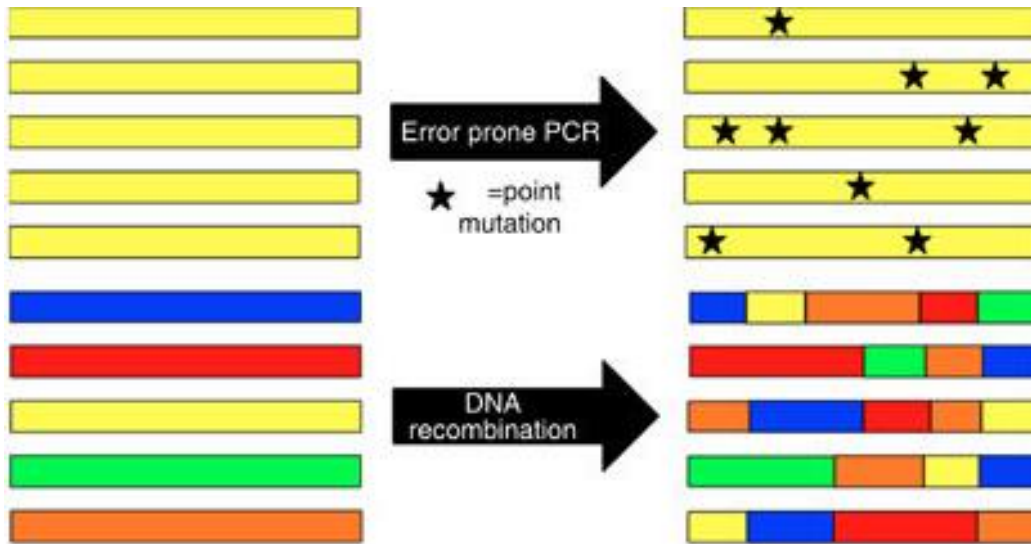


Figure 1.19. Error-prone PCR and DNA recombination or shuffling.⁴⁸

Finally, DNA recombination or shuffling is when separate DNA sequences randomly recombine to create a single gene (Figure 1.19).⁴⁹⁻⁵¹ This technique is similar to homologous recombination, which is detailed in Chapter 4. The parent DNA is fragmented and purified, then used as templates for PCR. This allows advantageous mutations between genes to be amplified and combined while also removing detrimental mutations.

1.2.2. Directed evolution

Directed evolution is a method of protein engineering that is analogous to natural selection (Figure 1.20).⁴⁹⁻⁵¹ Directed evolution is used to create vast libraries of different proteins and genes that contain variations in sequence. Once these vast libraries are created, they can be screened for activity or stability.^{10, 52} There are many methods of screening for hits such as

fluorescent signaling, cell surface display, and activity assays.^{48, 53} Positive hits are sequenced to determine sequence and mutations.⁵¹ A single protein usually undergoes several rounds of directed evolution to find the optimal sequence. Directed evolution, while powerful, carries distinct challenges and limitations that will be discussed in Chapter 5.

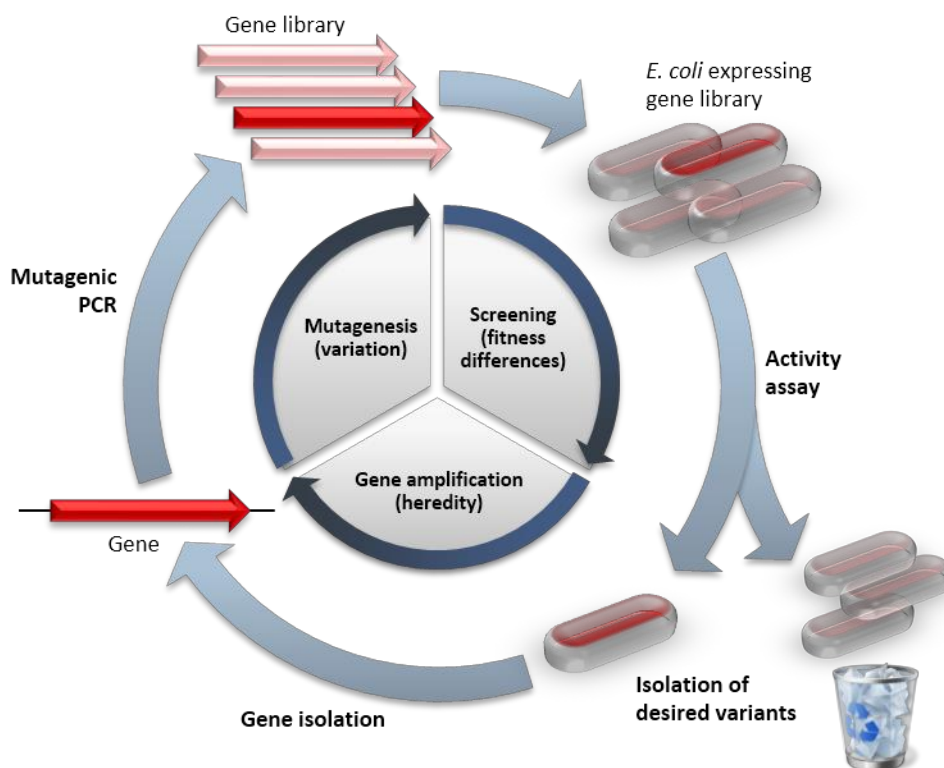


Figure 1.20. A diagram of directed evolution.⁵⁴⁻⁵⁹ The inner circle indicates the three stages of directed evolution with the natural equivalent in parenthesis. The outer circle indicates a typical experimental setup.

1.2.3. Rational design

Rational design involves determining the optimal mutation for a protein using physical models.⁵⁴⁻⁵⁹ Rational design uses computational programs to determine a protein's optimal sequence that gives the lowest energy structure. These sequences can either be designed from scratch using known peptide sequences or those inspired by nature (*de novo*) design.^{34, 60}

Computational design that uses the protein's backbone as a scaffold and then calculates a theoretical substrate transition state is known as theozyme-based computational design.³⁷ Iterative based computation design takes the same technique and includes optimizing the active site as well.⁶² While this is a powerful method, the proteins usually developed often have low catalytic activity because of a missing component in the process; enzymatic catalysis frequently relies on destabilizing substrates, which is not taken into account in computational methods since they only calculate the lowest energy transition states.⁴⁴ Furthermore, the reactions are highly specific to a single substrate.

1.2.4. Semi-rational design

Semi-rational design manages to merge the best qualities of directed evolution and computational design.⁶⁰ Usually an amino acid or group of amino acids are identified using computational algorithms and then a site-directed or random mutagenesis is done on the amino acid(s).³⁸⁻⁴¹ Then the protein is screened to determine if the desired traits have increased. A round of directed mutagenesis will then follow that, or more computational design.

1.3. Methods of Protein Production

Once a gene is created, organisms can be used to produce the protein in bulk.³⁸⁻⁴¹³⁸⁻⁴¹³⁸⁻
⁴¹ *Escherichia coli* is the typical organism used to carry out protein expression since they grow easily, multiply quickly, and are easy to mutate.⁵² First, the gene is spliced into a plasmid, which is circular DNA specific for bacteria. Then the plasmid is introduced into the bacterium in a

process known as transformation whereby the cell is coaxed into taking the plasmid in.⁶³ Transformation can be accomplished either by the use of heat or by the use of a gentle electrical current (Figure 1.21).⁶³

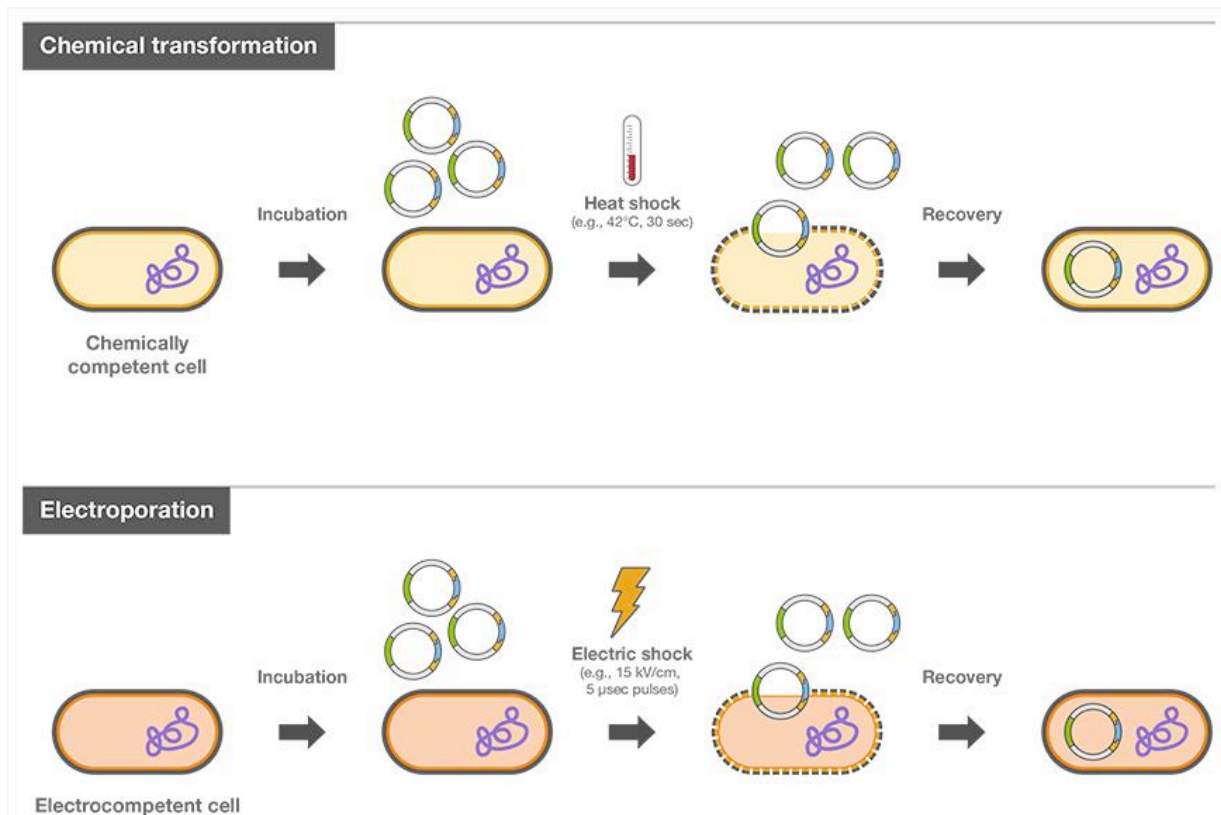


Figure 1.21. Protein transformation.^{15, 64-65} Plasmids with a desired gene can be introduced into bacteria such as *E. coli* by first ‘shocking’ the bacterium with either heat or an electric shock, then allowing the cell to rest so it will accept the plasmid into its body and recover.

After the plasmid is introduced into cells and colonies of plasmid-containing bacteria grow, a single colony is picked and grown in large scale. This larger culture will be used to synthesize and isolate the protein. The plasmid containing the protein gene also contains what is known as the lac operon (Figure 1.22).^{15, 64-65} Normally, an active repressor binds to the DNA that prevents RNA polymerase binding and protein synthesis. However, in the presence of lactose,

the bound repressor detaches from DovA and deactivates, exposing lac operon. RNA polymerase is able to bind to this operon and begin protein synthesis. Analogs of lactose, such as IPTG are used to 'induce' bacteria to express protein while they are growing. After several hours of growth, the cells are collected by centrifugation and are used for protein purification.

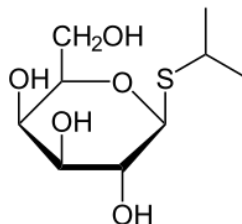
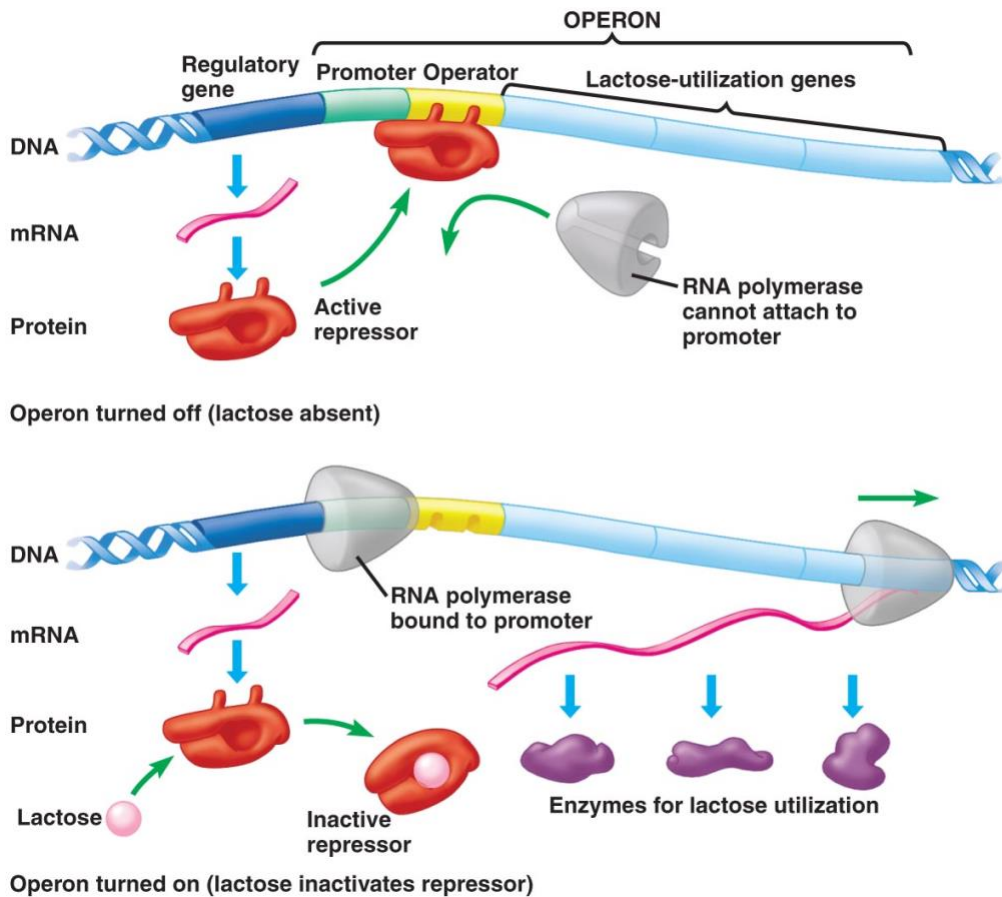


Figure 1.22. Diagram of protein induction.³ In normal conditions without lactose, a repressor binds to the operon (the sequence that signals protein synthesis) and RNA polymerase is unable to attach. In the presence of an inducing agent such as IPTG (bottom), it will bind to the repressor and allow protein synthesis to take place.

There are various methods of purifying proteins from cell paste but typically they consist of steps to first get the protein into solution by splitting open the cells and treating them, then purifying the protein from the supernatant.⁶⁶ For the purposes of purification, many proteins

have an extra sequence known as the His-tag attached to them. This His-tag consists of six histidines and a domain linking the histidines and the protein (usually with a sequence that can be cut by proteases such as tobacco etch virus or TEV).⁶⁷ The histidines are able to bind to a special resin known as Ni-NTA, which consists of nickel complexed with nitrilotriacetic acid that is attached to a solid surface such as a bead (Figure 1.23).⁶⁷ The histidine tags are able to bind to the Ni^{2+} while other impurities are not. To remove the bound protein, imidazole is used to 'elute' the protein off by competing with the His-tag for binding (Figure 1.24). Any undesired salts or buffers can then be removed using affinity chromatography.

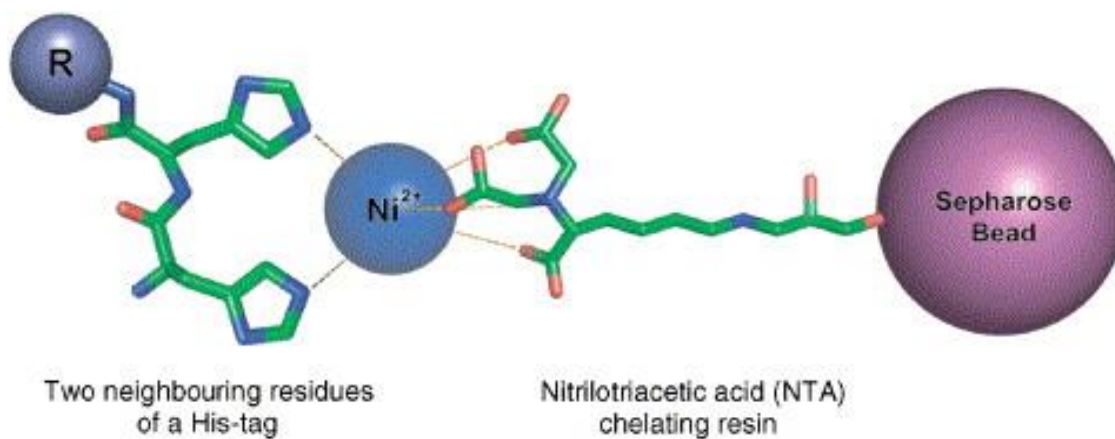


Figure 1.23. Nickel-NTA resin.⁶⁸ NTA chelates the nickel ion to the resin bead. When histidine is present, the nickel ions bind the residues while allowing other impurities to pass by. To detach the histidine, a competing compound such as imidazole is added to replace the histidine on the nickel.

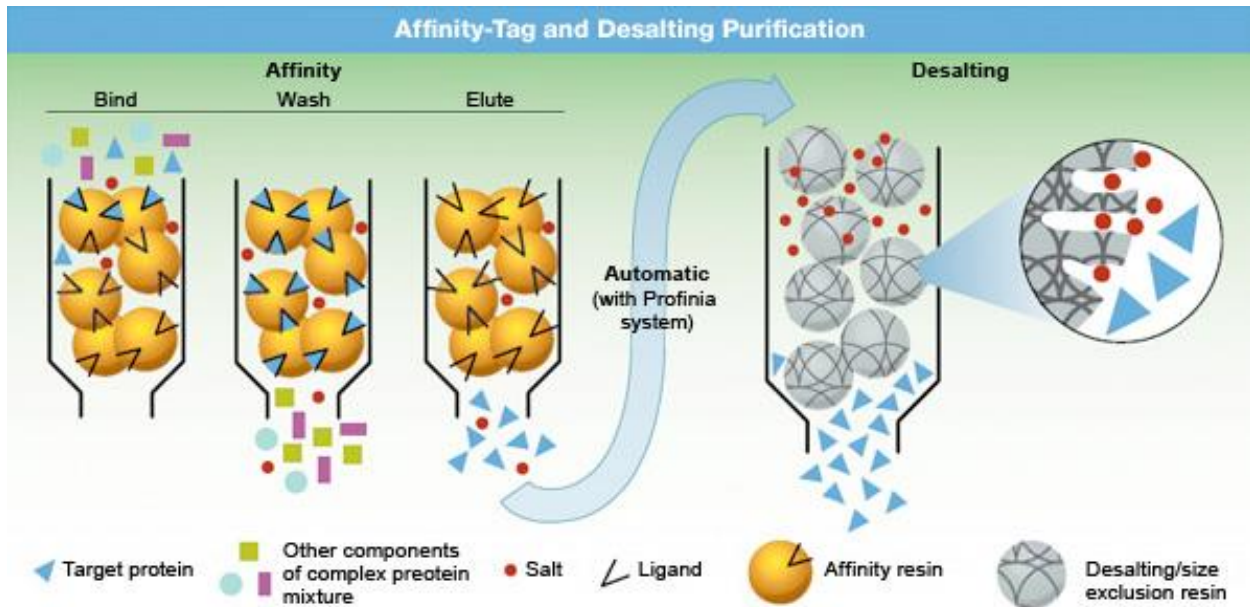


Figure 1.24. Affinity chromatography.^{1,3} The protein binds to the affinity resin while all other impurities are washed away, then the protein is eluted off using a competing molecule such as imidazole. To remove the eluting agent, the protein can then be put on a desalting column which traps smaller ions on top while the protein flows through, allowing the buffer to be 'exchanged.'

To determine purity of the protein, the protein is usually visualized using gel electrophoresis (Figure 1.25).^{1,3} The protein is loaded onto an acrylamide gel. Molecules that are larger are not as able to travel through the porous gel as smaller molecules. A detergent is employed, usually sodium dodecyl sulfate (SDS), that can break apart proteins in proportion to their molecular weight (typically one molecule of SDS binds to every two amino acid residues) and partially unfolds the proteins.¹ SDS gives the protein a large net negative charge that helps it travel through the polyacrylamide gel when a current is applied. Using this method, you can see bands from all proteins (both desired and undesired) present in sample and determine

expression of the protein of interest by comparing to a protein ladder. To visualize the bands, a dye such as Coomassie blue is added, which binds the proteins but not the gel polymer.

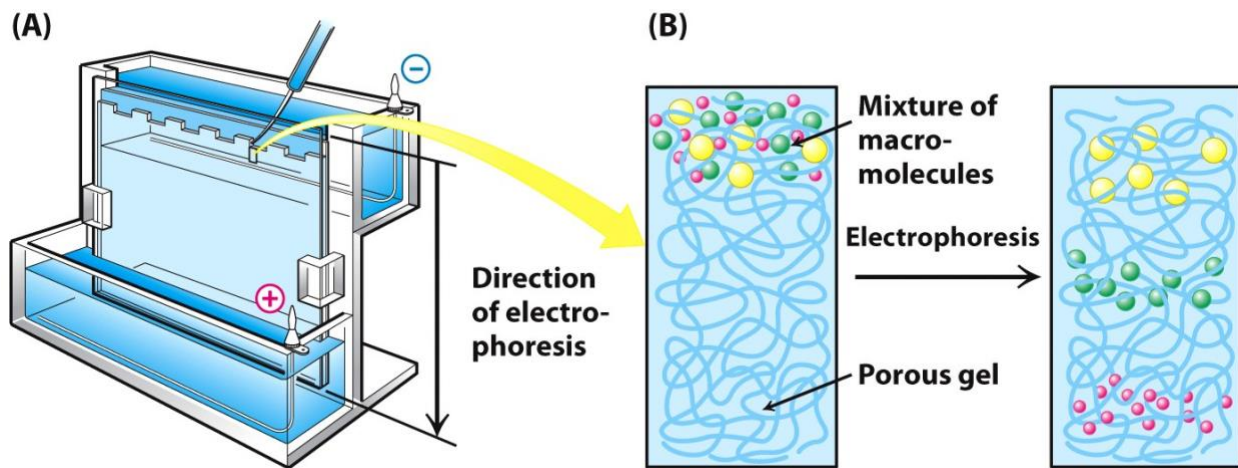


Figure 1.25. Gel electrophoresis.⁶⁹⁻⁷⁰ Protein samples are dyed and loaded onto a gel made of acrylamide polymer. An electrical current is passed through the gel, and smaller macromolecules travel further down the gel while larger ones stay near the top.

Concentration can be determined using UV-Vis spectroscopy (Figure 1.26). Specific amino acids have specific molar absorptivity that can be used to calculate the extinction coefficient of a protein at a specific wavelength (Figure 1.27).⁶⁹⁻⁷⁰ Protein absorbance is usually measured at 280 nm, which can be calculated from the absorbance of the tryptophan and tyrosine residues in the protein. At 280 nm, tryptophan and tyrosine have extinction coefficients of $5690 \text{ M}^{-1}\text{cm}^{-1}$ and $1280 \text{ M}^{-1}\text{cm}^{-1}$, respectively.

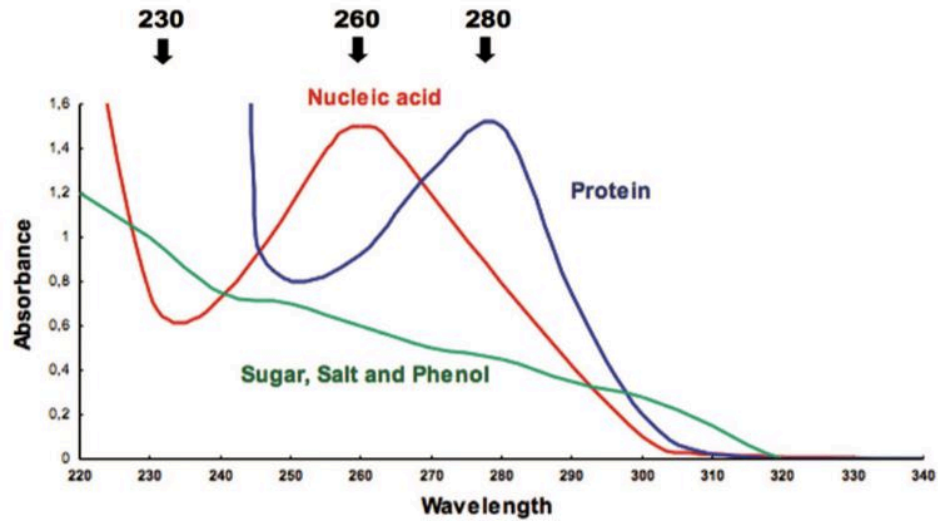


Figure 1.26. UV-Vis spectroscopy of proteins and nucleic acids.⁷⁰ Nucleic acids have an absorbance maxima at 260 nm. Proteins have absorbance maxima at 280 nm due to tyrosine and tryptophan residues.

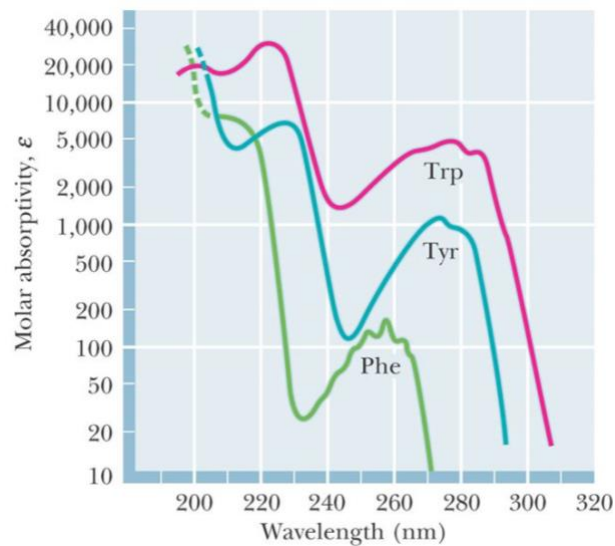


Figure 1.27. UV-Vis spectra of aromatic amino acids.⁷¹ Due to their aromatic groups, tryptophan and tyrosine both peak at 280 nm. Phenylalanine peaks at 260 nm, the same wavelength as nucleic acids, making it an unreliable measure of protein concentration.

To determine the exact mass of the isolated protein, mass spectrometry can be used.

MALDI-TOF is a common method of mass spectrometry where the protein is mixed with a matrix

(Figure 1.28).⁷¹ A laser releases the sample from the matrix (which is responsible for ionizing the sample) and the particles travel through an electric field generator towards a detector. Smaller, lighter compounds are able to reach the detector before larger fragments. The detector is then capable of determining the mass of the fragments based on how long a fragment takes to reach the detector.

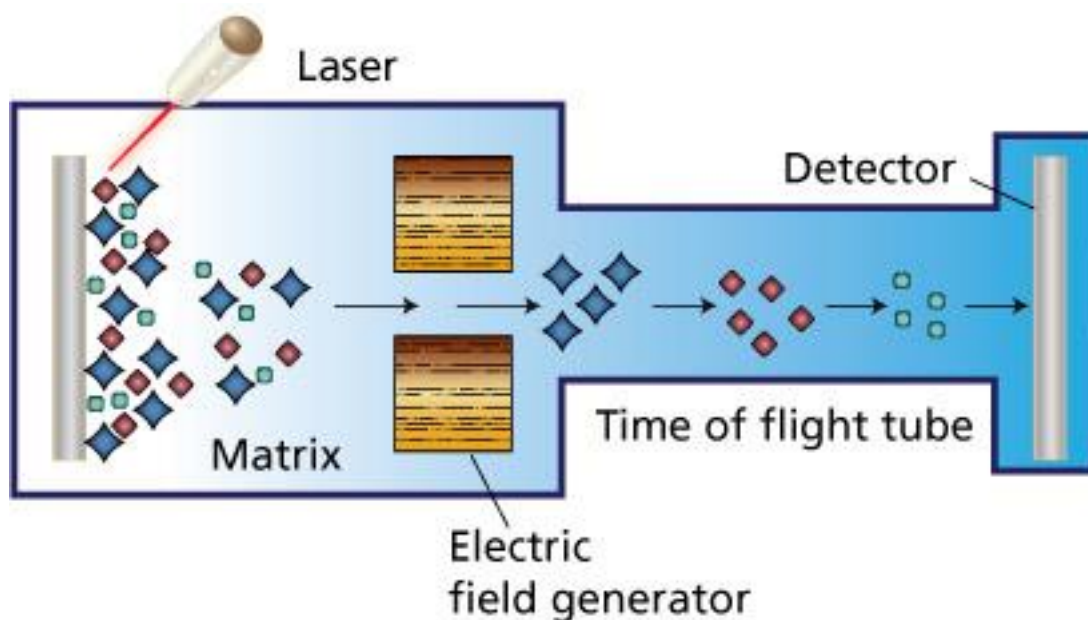


Figure 1.28. Diagram of MALDI-TOF.³¹ A laser releases sample bound to the matrix, which travels through an electric field to a detector. The detector determines the size of the sample by calculating how long it took for the sample to reach the detector.

1.4. Concluding Remarks

Enzymes are catalysts found in the natural world. They have a variety of functions due to their unique structures.¹ Because of their utility, scientists are constantly seeking new ways to design, optimize and express proteins for novel functions. This chapter gives a brief overview of how proteins function, the issues with designing proteins, and the most common ways to purify and identify proteins; techniques essential to the research projects described in this thesis. This dissertation will focus on how proteins can be made more viable for clinical and industrial uses. Chapter 2 will focus on designing a new, metal-specific protein from an existing scaffold. Chapter 3 will then apply the newly made protein from Chapter 2 for clinical use as a method of imaging and/or treating cancer. Chapter 4 focuses on how changing a protein's expression system can aid in making the protein more useable for industrial applications. Chapter 5 will focus on developing a new method of directed evolution that can be applied for optimizing other proteins more efficiently. Finally, Chapter 6 will highlight how peptide fragments can be used in conjunction with organometallic catalysts to create new potential catalysts that are cheaper to make and reuse.

1.5. References

1. Berg, J. M. T., John L; Stryer, Lubert, *Biochemistry*. 7 ed.; W. H. Freeman: 2010.
2. Voet, D. V., J.G., *Biochemistry*. 4 ed.; Wiley and Sons: 2010.
3. Nelson, D. L. C., M.M., *Lehninger Principles of Biochemistry*. 6 ed.; W.H. Freeman and Company: 2012.

4. Moran, L. A. H., R.A.; Scrimgeour, G; Merry, M., *Principles of Biochemistry*. 5 ed.; Pearson Edu: 2012.
5. Loew, O., On the chemical nature of enzymes. *Science* **1899**, *10* (261), 955-61.
6. Bruice, T. C. L., F.C, Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. *Acc Chem Res* **1999**, *32*, 127-36.
7. Parker, N. S., M; Thi Tu, A; Forster, B.M.; Lister, P; Allen, S; Auman, A; Brelles-Mariño, G; Feldman, M.A.; Flowers, P; Paterson, A; Pinchuk, G; Rowley, B; Sutherland, M; Franklund, C., *Microbiology*. OpenStax; American Society for Microbiology Press: 2016.
8. Koshland, D. E., The Key-Lock Theory and the Induced Fit Theory. *Angew Chem Int Ed Engl* **1994**, *33*, 2375-8.
9. Eschenmoser, A., One hundred years lock-and-key principle. *Angew Chem Int Ed Engl* **1994**, *33* (23), 2363.
10. Pandya, C.; Farelli, J. D.; Dunaway-Mariano, D.; Allen, K. N., Enzyme promiscuity: engine of evolutionary innovation. *J Biol Chem* **2014**, *289* (44), 30229-36.
11. Bernhard, S. A. M., R.A., Half-site reactivity and the Induced-fit hypothesis. *J Mol Biol* **1973**, *74*, 73-8.
12. Richter, M., Functional diversity of organic molecule enzyme cofactors. *Nat Prod Rep* **2013**, *30* (10), 1324-45.
13. Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009**, *460* (7257), 823-30.

14. Bronner, F., Extracellular and intracellular regulation of calcium homeostasis. *Scientific World Journal* **2001**, *1*, 919-25.
15. Lewis, M., Allosteric and the lac Operon. *J Mol Biol* **2013**, *425* (13), 2309-16.
16. Price, N. C., What is meant by 'competitive inhibition'? *TIBS* **1979**, 272-3.
17. Cornish-Bowden, A., Why is uncompetitive inhibition so rare? *FEBS J* **1986**, *203* (1), 3-6.
18. Bachan Upadhyay, L. S.; Verma, N., Enzyme Inhibition Based Biosensors: A Review. *Analytical Letters* **2013**, *46* (2), 225-41.
19. Johnson, D. S. W., E.; Cravatt, B.F., Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* **2010**, *2* (6), 949-64.
20. Michaelis, L.; Menten, M. L.; Johnson, K. A.; Goody, R. S., The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* **2011**, *50* (39), 8264-9.
21. Chothia, C., The classification and origins of protein folding patterns. *Annu Rev Biochem* **1990**, *59*, 1007-39.
22. Zorko, M., Structural organization of proteins. In *Introduction to peptides and proteins*, In Langel, U. C., B.F; Gräslund, A; von Heijne, G; Land, T; Niessen, S; Zorko, M, Ed. CRC Press: Boca Raton, 2010; pp 36-57.
23. Ahern, K. R., I.; Tan, Taralyn, Biochemistry Free for All. Oregon State University: 2017.
24. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990**, *29* (31), 7133-55.
25. Jiang, Q.; Jin, X.; Lee, S. J.; Yao, S., Protein secondary structure prediction: A survey of the state of the art. *J Mol Graph Model* **2017**, *76*, 379-402.

26. Chakravarty, S. V., R, Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* **1999**, 7 (7), 723-32.
27. Zheng, F.; Grigoryan, G., Sequence statistics of tertiary structural motifs reflect protein stability. *PLoS One* **2017**, 12 (5), e0178272.
28. Means, A. R., Calmodulin-mediated Signaling. In *Handbook of Cell Signaling*, 2 ed.; Academic Press: 2010; pp 979-81.
29. Takacs, M.; Makhlynets, O. V.; Tolbert, P. L.; Korendovych, I. V., Secretion of functional formate dehydrogenase in *Pichia pastoris*. *Protein Eng Des Sel* **2017**, 30 (3), 381-6.
30. Charnock, S. J. M., B.V., Enzymes: industrial and analytical applications. International, M., Ed. Wicklow, Ireland.
31. Choi, J. M.; Han, S. S.; Kim, H. S., Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnol Adv* **2015**, 33 (7), 1443-54.
32. Hult, K.; Berglund, P., Engineered enzymes for improved organic synthesis. *Current Opinion in Biotechnology* **2003**, 14 (4), 395-400.
33. Lodish, H. B., A; Zipursky, S.L; Matsudaira, P; Baltimore, D; Darnell, J, Functional Design of Proteins. In *Molecular Cell Biology*, 4 ed.; W.H. Freeman: 2000.
34. Mocny, C. S.; Pecoraro, V. L., De novo protein design as a methodology for synthetic bioinorganic chemistry. *Acc Chem Res* **2015**, 48 (8), 2388-96.
35. Wolfenden, R. S., M.J, The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* **2001**, 34, 938-45.

36. Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D., Minimalist active-site redesign: teaching old enzymes new tricks. *Angew Chem Int Ed Engl* **2007**, *46* (18), 3212-36.
37. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* **2012**, *109* (10), 3790-5.
38. Verma, R. B., E; George, A.J.T, Antibody engineering: Comparison of bacterial, yeast, insect, and mammalian expression systems. *J Immun Met* **1998**, *216*, 165-81.
39. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014**, *5*, 172.
40. Ahmad, M.; Hirz, M.; Pichler, H.; Schwab, H., Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* **2014**, *98* (12), 5301-17.
41. Pourmir, A.; Johannes, T. W., Directed evolution: selection of the host organism. *Comput Struct Biotechnol J* **2012**, *2*, e201209012.
42. Makhlynets, O. V.; Korendovych, I. V., Enzyme design: Functional Frankensteins. *Nat Chem* **2016**, *8* (9), 823-4.
43. Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stohr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., Short peptides self-assemble to produce catalytic amyloids. *Nat Chem* **2014**, *6* (4), 303-9.

44. Stepankova, V.; Bidmanova, S.; Koudelakova, T.; Prokop, Z.; Chaloupkova, R.; Damborsky, J., Strategies for Stabilization of Enzymes in Organic Solvents. *ACS Catalysis* **2013**, *3* (12), 2823-36.
45. Shendure, J.; Ji, H., Next-generation DNA sequencing. *Nat Biotechnol* **2008**, *26* (10), 1135-45.
46. Reeves, H. Recombinant DNA Technology site directed mutagenesis genetics vs. reverse genetics gene expression in bacteria and viruses gene expression in yeast genetic. <http://slideplayer.com/slide/6199958/>.
47. Bartel, M. A.; Weinstein, J. R.; Schaffer, D. V., Directed evolution of novel adeno-associated viruses for therapeutic gene delivery. *Gene Ther* **2012**, *19* (6), 694-700.
48. Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L., Directed evolution of enzymes: Library screening strategies. *IUBMB Life* **2009**, *61* (3), 222-8.
49. Porter, J. L.; Boon, P. L.; Murray, T. P.; Huber, T.; Collyer, C. A.; Ollis, D. L., Directed evolution of new and improved enzyme functions using an evolutionary intermediate and multidirectional search. *ACS Chem Biol* **2015**, *10* (2), 611-21.
50. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat Rev Genet* **2015**, *16* (7), 379-94.
51. Shafee, T. Evolvability of a viral protease: experimental evolution of catalysis, robustness, and specificity. University of Cambridge, 2014.
52. Sittampalam, G. S. C., N.P; Brimacombe, K; Grossman, A; Arkin, M; Auld, D; Austin, C; Baell, J; Bejcek, B; Chung, T.D.Y; Dahlin, J.L; Devanaryan, V; Foley, T.L; Glicksman, M; Hall, M.D;

Hass, J.V; Inglese, J; Iversen, P.W.; Kahl, S.D; Kales, S.C; Lal-Nag, M; Li, Z; McGee, J; McManus, O; Riss, T; Trask, O.J; Weidner, J.R; Xia, M; Xu, X, Assay Guidance Manual. Eli Lilly; National Center for Advancing Translational Sciences: Bethesda, MD, 2017.

53. Xiao, H.; Bao, Z.; Zhao, H., High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind Eng Chem Res* **2015**, *54* (16), 4011-20.

54. *Computational protein design*. Springer Nature: 2017.

55. Feldmeier, K.; Höcker, B., Computational protein design of ligand binding and catalysis. *Current Opinion in Chemical Biology* **2013**, *17* (6), 929-33.

56. Ebert, M. C.; Pelletier, J. N., Computational tools for enzyme improvement: why everyone can - and should - use them. *Curr Opin Chem Biol* **2017**, *37*, 89-96.

57. Durrenberger, M.; Ward, T. R., Recent achievements in the design and engineering of artificial metalloenzymes. *Curr Opin Chem Biol* **2014**, *19*, 99-106.

58. Davids, T.; Schmidt, M.; Bottcher, D.; Bornscheuer, U. T., Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr Opin Chem Biol* **2013**, *17* (2), 215-20.

59. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering enzymes. *Curr Opin Chem Biol* **2014**, *19*, 8-16.

60. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc Natl Acad Sci U S A* **2011**, *108* (17), 6823-7.

61. Tantillo, D. J. C., J; Houk, K.N., Theozymes and compuzymes: theoretical models for biological catalysis. *Curr Opin Chem Biol* **1998**, *2*, 743-50.

62. Brustad, E. M.; Arnold, F. H., Optimizing non-natural protein function with directed evolution. *Curr Opin Chem Biol* **2011**, *15* (2), 201-10.
63. Scientific, T. F. Bacterial Transformation and Competent Cells-A Brief Introduction. <https://www.thermofisher.com/us/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/molecular-cloning/transformation/competent-cell-basics.html>.
64. Lewis, M., A tale of two repressors. *J Mol Biol* **2011**, *409* (1), 14-27.
65. Jacob, F. M., J, Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **1961**, *3*, 318-56.
66. Blommel, P. G.; Fox, B. G., A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr Purif* **2007**, *55* (1), 53-68.
67. Bolanos-Garcia, V. M.; Davies, O. R., Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta* **2006**, *1760* (9), 1304-13.
68. Bio-Rad Mechanisms of Affinity Binding. <http://www.bio-rad.com/en-us/applications-technologies/affinity-chromatography?ID=LUSMJIDN>.
69. Armbrecht, M., Detection of contamination in DNA and protein samples by photometric measurements. AG, E., Ed. Germany, 2013.
70. Wetlaufer, D. B., Ultraviolet spectra of proteins and amino acids. *Advan Prot Chem* **1962**, *17*, 303-90.

71. Agrawal, M. V. V., K; Biswas, S., Vacuum MALDI-TOF Technology for Advanced Applications in Pharmaceuticals and Medicine. *Vacuum Technology and Coating* **2015**.

Chapter 2 Design of a Rare Earth Selective Calmodulin Derivative

2.0. Abstract

Even though metalloproteins account for at least half of all proteins in the body, strategies for evolving specific metal ion selectivity are still relatively unknown. Calmodulin (Cam) is a protein that undergoes a conformational change in presence of Ca^{2+} that allows for substrate binding. In recent years, Korendovych and co-workers modified a non-catalytic wild type Cam to enable the current generation, Ac7, to perform a Kemp elimination reaction upon addition of calcium ions. The product of Kemp elimination produces a clearly visible yellow color that can be read using a standard UV-Vis spectrometer. In addition, the Korendovych group successfully evolved the “EF hands” of the catalytic Cam mutant to bind lanthanide ions with a moderate preference over Ca^{2+} . Using these same principles, an Ac7 mutant has been developed to strongly prefer rare earth metals over Ca^{2+} and yet still catalyze the Kemp elimination.

2.1. Introduction

Approximately 40% of all proteins can be classified as metalloproteins, that is, proteins that bind to one or more metal ions.¹⁻⁶ The metal ion can be responsible for structure, electron transfer, dioxygen binding, and/or catalysis by the protein.⁵ Metalloenzymes, the enzymatic equivalent of metalloproteins, are responsible for some of the most important reactions in nature, including photosynthesis, respiration, and nitrogen fixation.³

Most metalloproteins bind to either alkaline metals (Group 1), alkaline earth metals (Group 2), or transition metals (Groups 3-12).^{2, 7} This is believed to be due to the fact that these elements are more prevalent in the environment and therefore proteins evolved to utilize the already abundant resource (Figure 2.1). This theory is supported, for example, by methanol dehydrogenase, a protein produced by *Methylobacterium extorquens*.⁸ This organism is capable of living in volcanic mud pots and is able to convert methanol to formaldehyde. Mud pots contain high amounts of soluble lanthanides, which are used in the protein as a cofactor alongside pyrroloquinoline quinone (Figure 2.2).⁹ Without this lanthanum, the protein is incapable of withstanding the extreme temperatures of its native environment.^{8, 10-11}

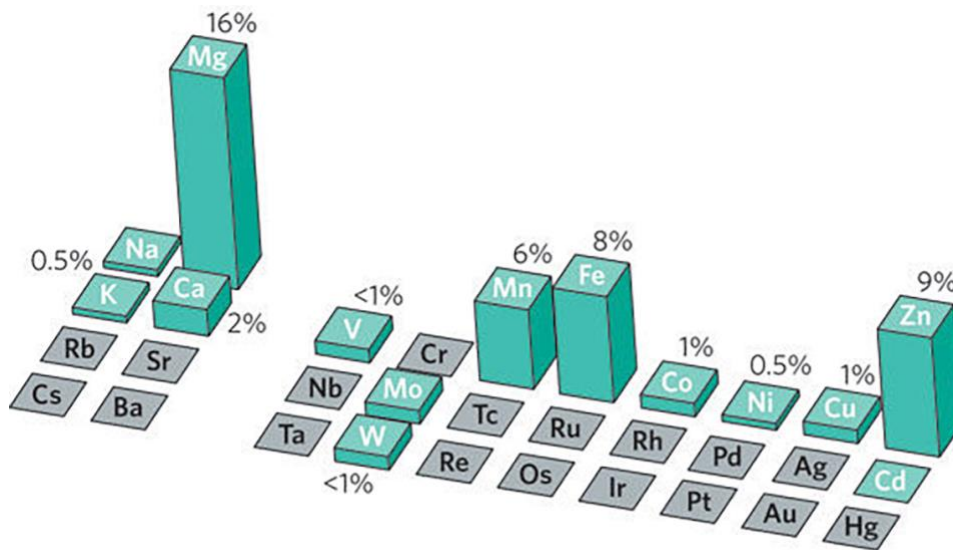
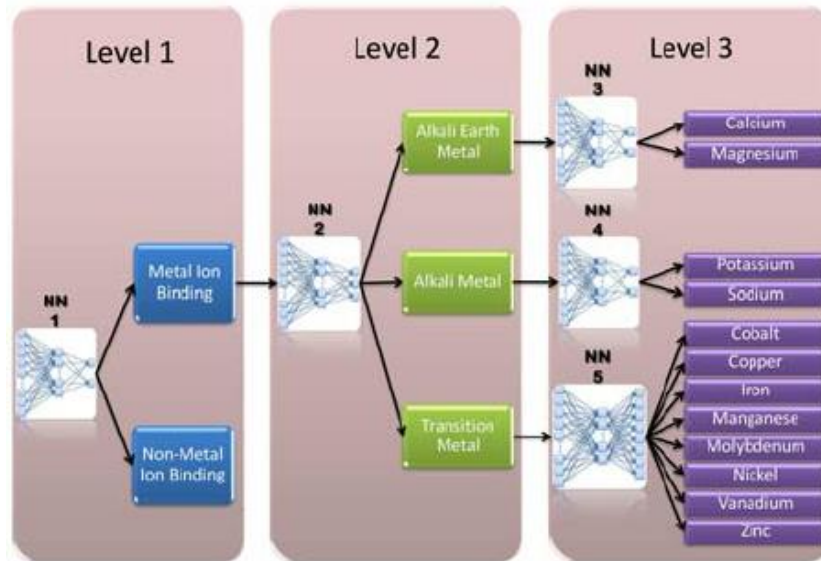


Figure 2.1. Summary of the most common metals used in metalloproteins.^{2,7}

In addition, rare metals are becoming more common in the environment due to human influence. For example, due to toxic waste, there has been an increase in bacteria and other organisms capable of living in these conditions. *Cupriavidus metallidurans* strain CH34 is a Gram-negative bacteria that has been found to grow in high levels of heavy metals such as mercury,

cadmium, and lead (Figure 2.2).¹²⁻¹⁵ The bacterium relies on proteins that can selectively bind the heavy metal ions over essential ions such as calcium or zinc.¹⁶⁻¹⁷ One protein in particular, PbrR691 of the MerR family, is especially effective in binding to lead metal thanks to its three conserved cysteine residues and its ability to allow Pb^{2+} to adopt a preferred hemi-directed geometry while bound.^{13, 17}

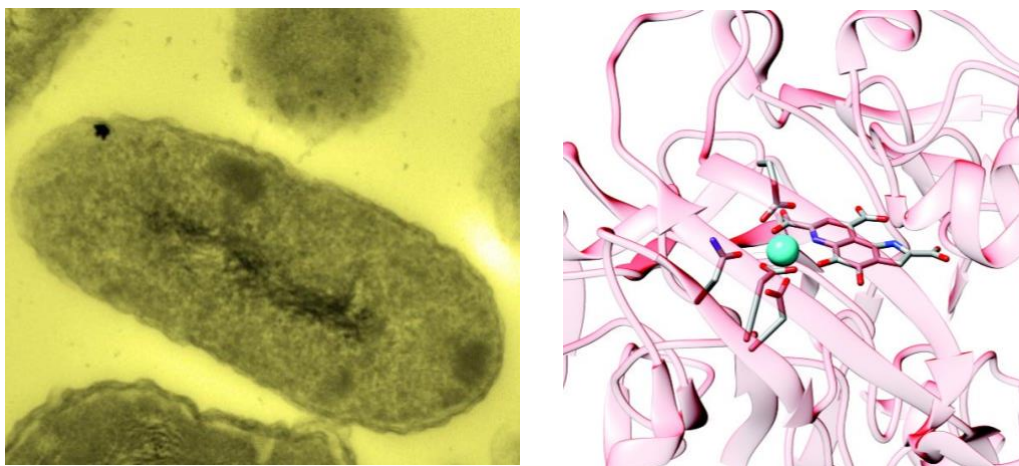


Figure 2.2. Examples of evolved organisms and proteins that bind non-typical metals specifically. Left: *Cupriavidas metallidurans*, a bacterium that is able to live in toxic waste due to its ability to selectively bind Pb^{2+} and ship it out of the cell before it does damage.¹¹ Right: The active site of Methanol dehydrogenase, isolated from *Methylobacterium extorquens*, which coordinates with a La^{3+} ion to allow it to live in extreme temperatures.¹⁰

It is the ambition of many scientists to be able to design metalloproteins from scratch.³ Being able to evolve and design metalloproteins is an important step to understanding evolution itself. One way to gain insight into this question is to transform a protein that binds a more common metal, such as calcium, into a new derivative that can bind less common metals, such as the rare earth elements. For this, we used calmodulin as a starting scaffold, and eventually mutated it into HOLIEE, a selective lanthanide-binding protein that catalyzes an unnatural reaction.

2.1.1. Calmodulin

Calcium is the most prevalent metal ion within the body and as such calcium-binding proteins are common and are widely found in eukaryotes.¹⁸⁻¹⁹ Of these calcium-binding proteins, Calmodulin (Cam) is responsible for a myriad of functions in the body including inflammation, muscle contraction, memory, nerve growth, and the immune response (Figure 2.3).²⁰⁻²¹ Cam is a highly conserved 148-residue (16.7 kDa) protein that has been extensively studied and has a well characterized structure.

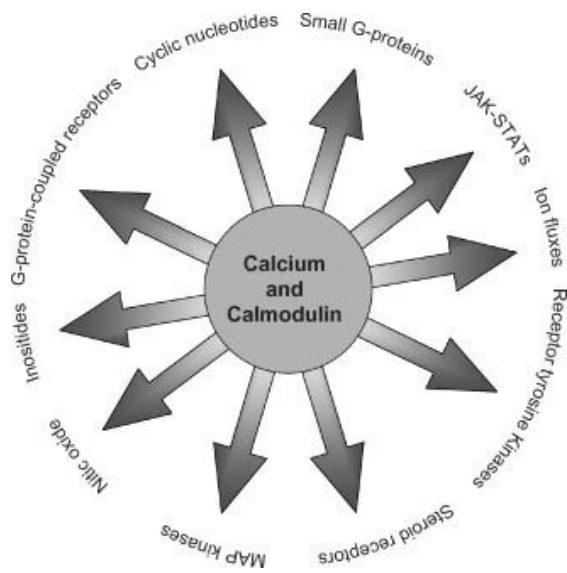


Figure 2.3. Examples of systems and proteins that calmodulin and calcium regulate.²¹

Most calcium-sensing proteins, including Cam, use EF-hand motifs to coordinate calcium ions.²² In the EF-hands of calcium-binding proteins, the amino acids are numbered according to their relative position within the loop, positions 1-12 (Figure 2.4). Positions 1, 3, 5, 7, 9, and 12 coordinate with calcium to form a pentagonal bipyramidal complex.^{20, 23-24} Loop positions 1, 3, 5, and 12 consist of amino acids with carboxyl or hydroxyl groups in the side chains.²⁴ Meanwhile,

amino acid at position 7 coordinates from the main chain carbonyl, position 9 is a bridged water molecule, and position 12 also acts as a bidentate ligand in the coordination complex. Residue 1 is almost exclusively Asp between species and is responsible for the maintaining the structure of Cam.²⁴⁻²⁵ Residues 3 and 5 are predominantly Asp residues as well, though there is more variety between eukaryotes.²⁴ Position 6 is always a Gly residue in order for the folds in the α -helix to occur properly. Finally, position 12 is usually an aspartate or glutamate residue; options for bidentate amino acids are limited.

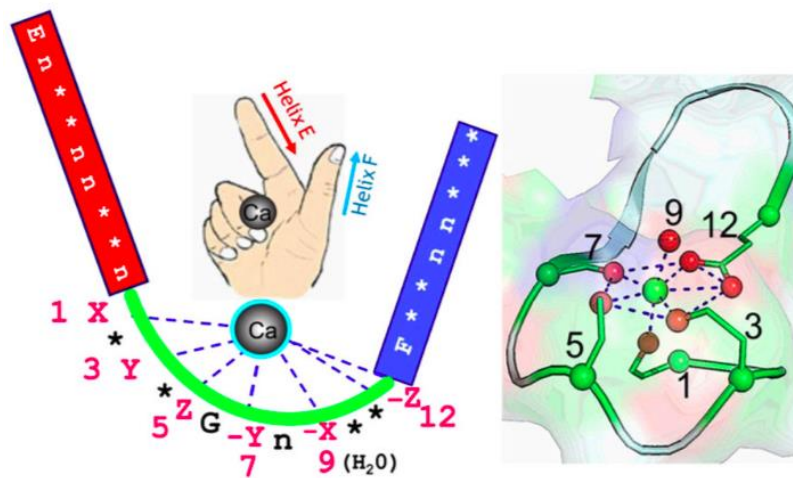


Figure 2.4. Diagram of EF-hands in calcium-sensing proteins.²¹ Residues 1, 3, 5, 7, and 12 coordinate direction with the Ca²⁺ ion to form a pentagonal bipyramidal complex. Nine represents a water molecule that is coordinated to the Ca²⁺ ion with residue 9 in the EF-hand.

Cam contains two lobes, which contain two EF-hand sites each, that are linked by an α -helix.^{20, 26} Each lobe consists of a large hydrophobic section that is important for substrate binding.²⁶ When calcium binds to the EF-hands, it induces a change in the structure of Cam and exposes the hydrophobic sections so that substrates may bind (Figure 2.5).

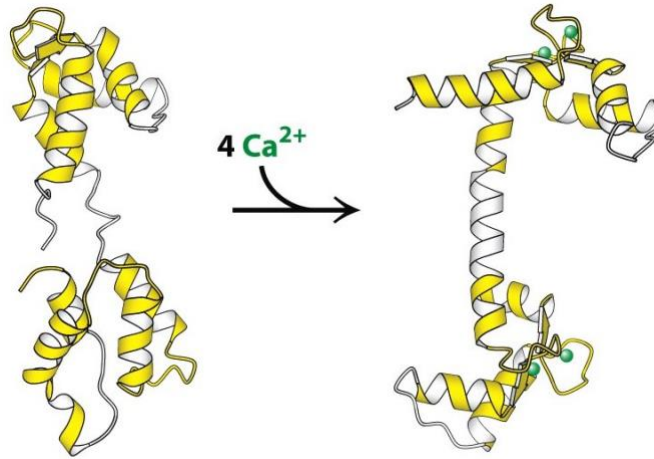


Figure 2.5. Conformational change of calmodulin in the presence of calcium.²⁴

Calmodulin is a popular scaffold for protein engineering.²⁷ Its thermostability, small size, lack of intrinsic catalytic activity, the presence of a cavity sufficient for substrate binding, and the spatial separation between the allosteric ligand-binding sites (in this case the EF-hands) and the substrate binding pocket make it an ideal protein since it is easy to work with and examine with NMR characterization.²⁸

2.1.2. AlleyCat

In 2010, Ivan Korendovych of the DeGrado group at University of Pennsylvania was able to engineer a Kemp eliminase by using calmodulin as the scaffold.²⁸ Studies suggested that a majority of the initial allosteric metal binding occurs in the C-terminal domain of calmodulin.^{4, 20} By removing the N-terminal domain of Cam, Korendovych was able to create a new protein (cCam) with a similar structure to the parent scaffold, but with only 74 residues and two EF-hands (meaning it only needs to bind two equivalents of calcium instead of four).²⁸ The new protein was then used to create a new Kemp eliminase known as AlleyCat (ALLostERically Controlled

CATalyst). This protein has been the foundation for numerous studies in the Korendovych lab as an evolvable enzyme and sensor.²⁸⁻³⁰ In addition, cCam has been used as a platform for other catalytic enzymes, such as a retro-aldolase and an esterase.³¹⁻³²

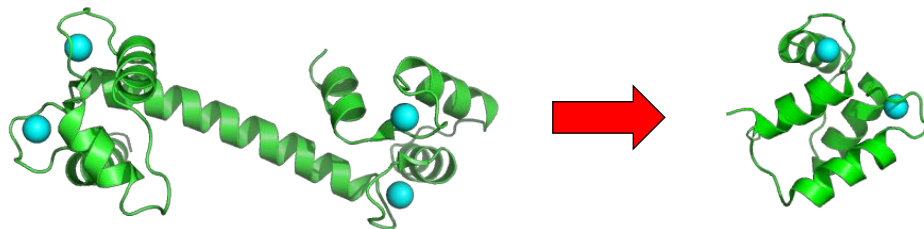
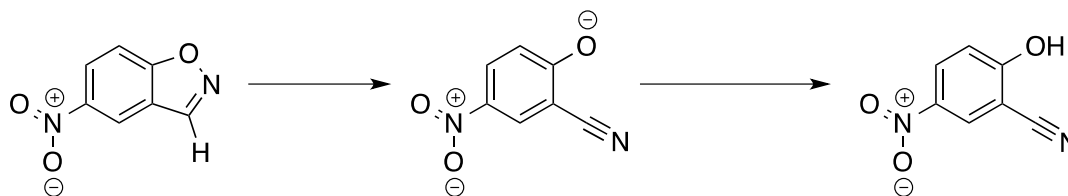


Figure 2.6. Schematic of calmodulin conversion to C-terminal calmodulin. Mutating the phenylalanine in the 92 position of C-term Cam to glutamate yields AlleyCat protein.¹⁴

Kemp elimination (Scheme 1) is a non-natural reaction for bacteria, meaning that any observed reaction of the substrate must be due to the introduced mutations.³³ For this reason, the Kemp elimination is a useful benchmark for protein engineering.^{28, 30, 33-38} The reaction requires a basic catalytic residue to break one of the C-H bonds, allowing for the formation of a nitrile whose concentration is easily measured using visible spectroscopy at 380 nm.



Scheme 1. Kemp elimination reaction.

AlleyCat's catalytic activity is from the mutation of the phenylalanine at position 92 to glutamate.^{28, 30} This residue is deep within the hydrophobic pocket, which raises the pK_a of the normally acidic Glu residue from 2.2 to 5.8 when substrate is not present.^{28, 39} In the presence of a bound substrate, the pK_a of the Glu becomes even more basic, approximately 6.9, allowing for Kemp elimination to proceed.²⁸

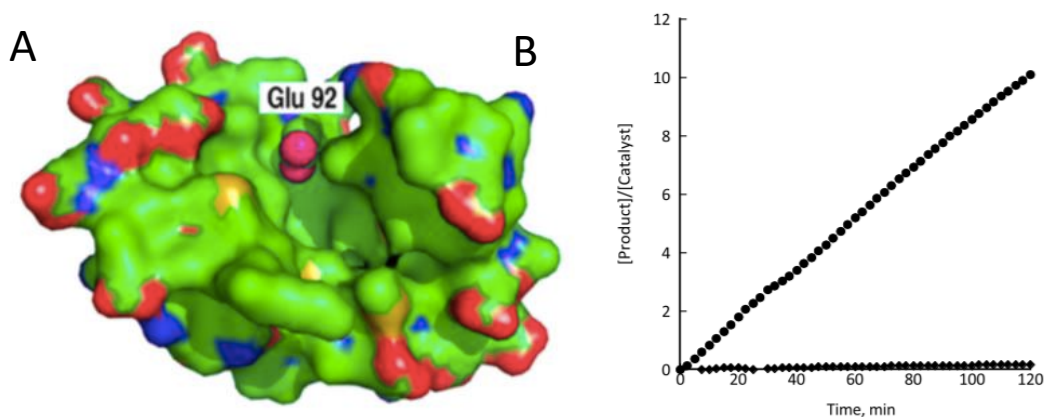


Figure 2.7. A) 3D representation of cell surface of AlleyCat. B) Comparison of catalytic activity between F92E (black circles) and negative control mutant F92Q (black diamonds).²⁵

AlleyCat has undergone many rounds of evolution in the Korendovych group in order to optimize catalytic activity.³⁰ The round of mutation is denoted by the terminal number of the name AlleyCat (Ac); AlleyCat5 (Ac5) refers to the fifth round of mutation, AlleyCat3 (Ac3) refers to the third round of mutation, and AlleyCat0 (Ac0) refers to the original protein. A comprehensive list of all mutations and sequences of AlleyCat is available in Table S.4. of the appendix. Further detail into the methods for further evolution of the AlleyCat series will be covered in Chapter 5.

One of the major features of the AlleyCat series is that Kemp elimination only occurs in the presence of calcium.³⁰ This is due to the conformational change that occurs in Cam when

calcium binds to the EF-hands.^{30,40} The substrate binding pocket of AlleyCat is only exposed when calcium is in place, and removing calcium from the environment will immediately stop catalysis and vice versa (Figure 2.8).

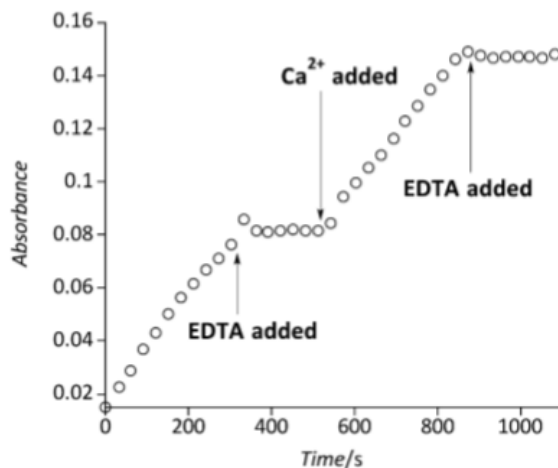


Figure 2.8. Kinetic assay of Kemp elimination showing the allosteric regulation of AlleyCat7.²⁷ Removing calcium ions from the reaction with chelator reduces kinetic activity and vice versa.

With this allosteric property in mind, AlleyCat has the potential to become a customizable metal sensor.^{29,41} Since Kemp elimination only occurs if the protein is bound to metal, modifying what metals AlleyCat is able to bind to is a promising way to create this sensor (Figure 2.9). We are able to utilize this property while characterizing the specificity of our evolved proteins, since Kemp elimination will only occur if the protein is binding a metal ion.

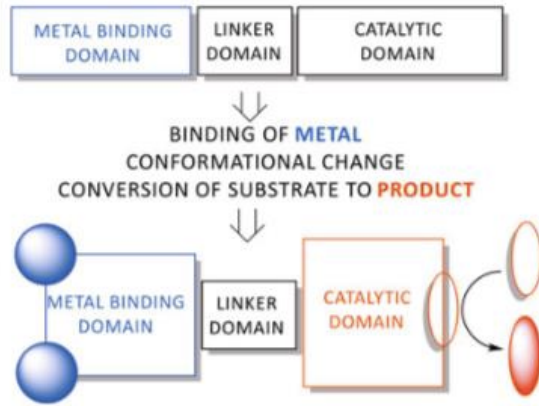


Figure 2.9. General mechanism of an engineered protein metal sensor.²⁶

2.1.3. Coordinating alternative metals in AlleyCat

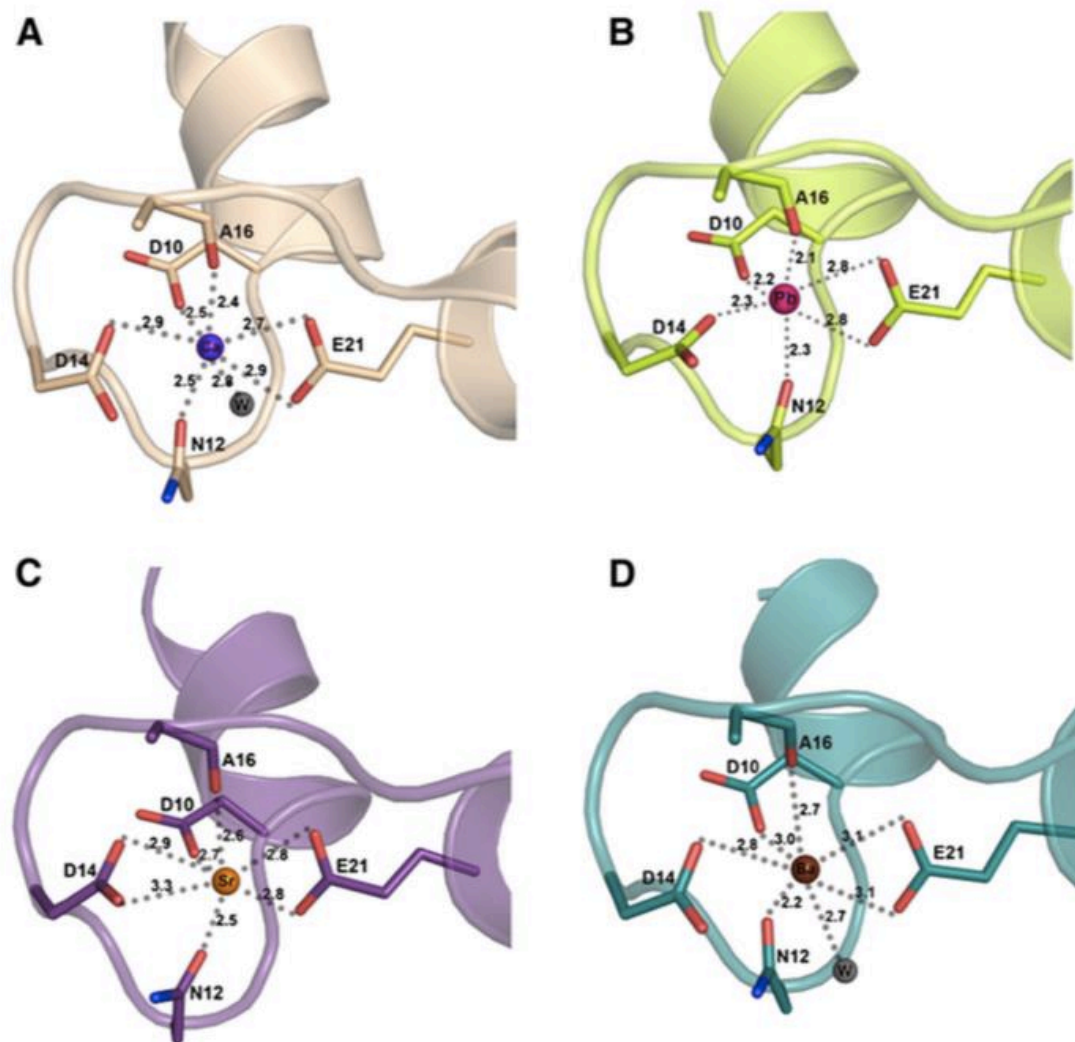


Figure 2.10. Crystal structure of the EF-hand metal binding site of calmodulin when bound to (A) Ca²⁺, (B) Pb²⁺, (C) Sr²⁺, and (D) Ba²⁺.⁴⁵

In addition to heavy metals, the binding of lanthanide ions is also well studied in regard to their interactions with EF-hands.^{18, 42} Despite being trivalent ions, lanthanides have similar ionic radii to calcium and seem to bind Cam similarly (Figure 2.10). There are relatively few convenient methods to determine the concentration of lanthanide ions in solution. Most

methods involve either the use of expensive and hard to obtain equipment or complex molecules that will give out measurable readings when bound to the metals.⁴³⁻⁴⁴ The limited availability of such methods is increasingly problematic. Rare earth metals are increasingly important in modern society, used in products such as hybrid car batteries, smart phones, magnets, and even medicines.^{8, 45-46}

Analyzing the results of previous studies on lanthanide and heavy metal binding to EF-hands, we hope to create new proteins that can be highly selective for binding and sensing by catalyzing the Kemp elimination to provide a convenient colorimetric assay. A catalytic metal sensor is advantageous over current methods such as using fluorophores since a catalyst would require much less metal. The measurable output of a catalytic sensor is limited only by the turnover rate of the catalyst, and will continually secrete a reading until it can be measured, which means that even small amounts of metal will eventually be 'sensed' since they are not being measured directly. Fluorophores and instrumental analysis measure concentrations of metal ions directly, so if the metal concentration is below the threshold of an instrument's range or not concentrated enough to get a reliable reading from the fluorophore, the metal concentration cannot be measured.

2.1.4. CuSeCat and previous work

In 2013, Korendovych lab published about a new variant of AlleyCat called CuSeCat (CUstomizable Sensor aided by CATalysis) that was able to bind lanthanides in the EF-hands²⁹. In order to accomplish this, residues S101 and N137 of the EF-hands were mutated into negatively-charged residues; this should increase the charge density of the metal binding site (Figure 2.11).

Residue 9 in the EF-hands, which normally consists of neutral residues bridged by a water molecule, is an ideal position to change since the residue itself is only loosely associated with metal coordination. In principle, charge density at that site could be increased with theoretically little change to the structure.^{19, 22, 29, 47} To accomplish the goal of increasing the negative charge of the metal binding loop, S101 and N137 were mutated into Asp and Glu residues, CuSeCatDD and CuSeCatEE, respectively.²⁹ The results are promising.

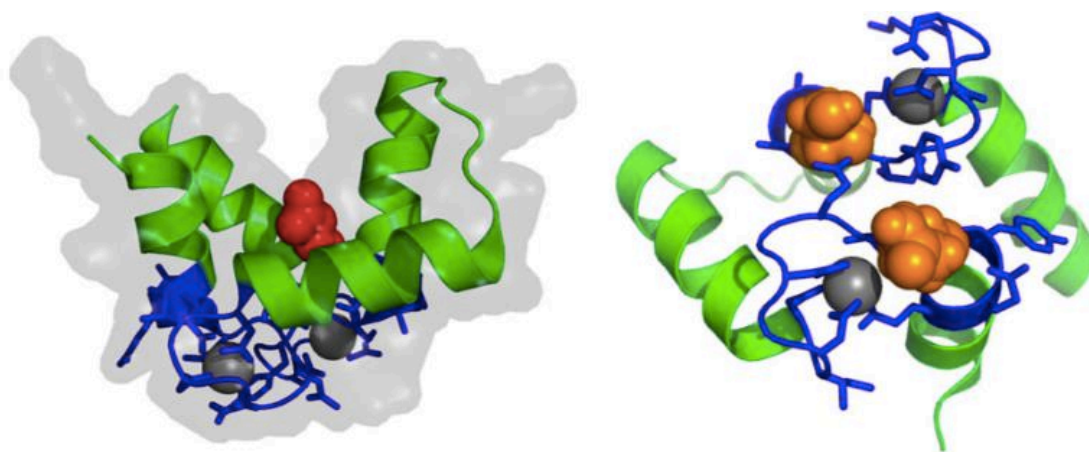


Figure 2.11. Modeling of S101 and N137 positions of AlleyCat.²⁸ The green ribbon represents the structure of CuSeCat, the blue sticks represent the active residues in the metal binding site, the red spheres represent the F92E mutation responsible for Kemp elimination, the orange spheres represent the S101E and N137E mutations that were introduced to CuSeCat, and the gray spheres represent a trivalent metal ion.

In this particular case, the goal was not to create a more efficient enzyme, but rather an enzyme that would (a) be active in the presence of the desired metal and (b) *not* be active in the presence of high amounts of calcium. Metal dependence assays, which measure the maximum $k_{\text{cat}}/K_{\text{M}}$ at different concentrations of metal, are a practical way to visualize selectivity. When the Kemp elimination with CuSeCatEE was run in the presence of CaCl_2 and YbCl_3 , it was clear that CuSeCatEE was more active with the lanthanide (Figure 2.12) reflecting stronger metal binding to Yb^{3+} .

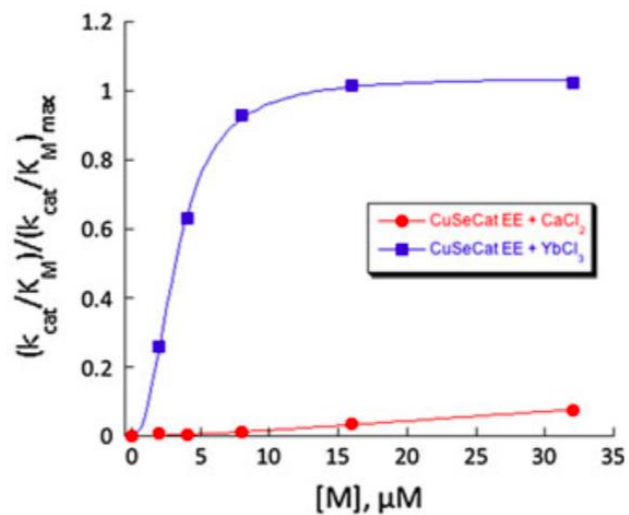


Figure 2.12. Metal dependence assay of CuSeCatEE.²⁸ In low concentrations of YbCl₃ and CaCl₂, CuSeCatEE is more active in the presences of lanthanide.

While these results are promising, increasing the concentration of CaCl₂ results in the significant increase in Kemp elimination activity (Figure 2.13). While intracellular calcium levels can be as low as 50-100 nM, outside the cell there can be anywhere from 1-3 mM free calcium ions.⁴⁸ This means that CuSeCat could not be used because there would be too much interference from background calcium. Furthermore, due to the nature of ytterbium and other lanthanides, higher concentrations (typically exceeding 0.1mM) result in precipitation of protein.

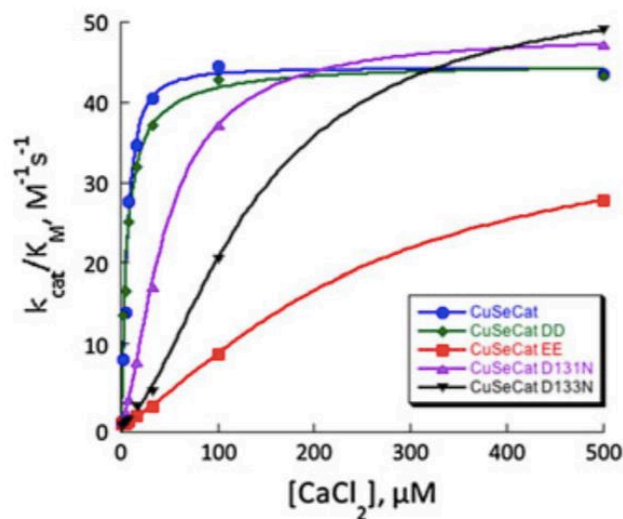


Figure 2.13. Metal dependence assay of CuSeCat series with 10,000-fold excess of CaCl₂ to protein.²⁸

2.2. Designing a Selective AlleyCat

2.2.1. Position 9 - S101 and N137

We are particularly interested in designing Kemp elimination assays that respond to heavy metals or lanthanides due to the potential applications as to environmental sensors or chelating linkers for radioimmunotherapy. AlleyCat7, the most active of the AlleyCat series to date, was used as a DNA template for all mutations. To mimic CuSeCat, aspartate and glutamate mutations were introduced into the S101 and N137 positions to create AcDD and AcEE respectively. Due to the poor solubility of lead(II) nitrate in water and the tendency for lanthanides such as yttrium(III) nitrate to precipitate proteins at higher concentrations, concentrations of metal had to be kept below 100 μM.⁴⁹⁻⁵⁰ Unlike CuSeCat, AcDD and AcEE were not selective for lead or yttrium over

calcium; in all cases, all mutants exhibited higher activity for calcium at equal concentration (Table 2.1).

Cysteine is residue that binds strongly to heavy metal ions.^{13, 20} However, cysteine mutations can be difficult to work with due to their ability to make disulfide bonds, which can cause proteins to dimerize or aggregate. Cysteine mutations to the S101 and N137 positions were expressed and purified. Unfortunately, there was evidence of dimer formation by SDS-PAGE gel (Figure S.0.2). Upon adding an equivalent of the reducing agent β -ME, the dimers broke down into the monomeric form of the protein. Nonetheless, neither the dimers nor the presence of β -ME had much effect on the activity of the proteins. S101C and N137C showed much greater activity in the presence of calcium than lead or yttrium (Table 2.1). We therefore moved on to other positions in the EF-hands.

Table 2.1. Summary of activity of mutations to position 9 after the addition of 0.1 mM $\text{Ca}(\text{NO}_3)_2$, 0.1 mM $\text{Pb}(\text{NO}_3)_2$, or 0.1 mM $\text{Y}(\text{NO}_3)_3$. All assays were run in 20mM Hepes at pH 7.0 with 5.0 μM protein and can be found in Table S.7-S.12.

$k_{cat}/K_M (M^{-1}s^{-1})$	Ca^{2+}	Pb^{2+}	Y^{3+}
<i>Ac7</i>	462±24 $M^{-1}s^{-1}$	403±62 $M^{-1}s^{-1}$	32±48 $M^{-1}s^{-1}$
<i>AcDD</i>	555±25 $M^{-1}s^{-1}$	254±56 $M^{-1}s^{-1}$	18 $M^{-1}s^{-1}$
<i>AcEE</i>	309±29 $M^{-1}s^{-1}$	---	---
<i>AcS101C</i>	327±40 $M^{-1}s^{-1}$	---	5 $M^{-1}s^{-1}$
<i>AcN137C</i>	410±49 $M^{-1}s^{-1}$	---	18 $M^{-1}s^{-1}$

2.2.2. Position 7- Y99 and Q135

Position 7 is also responsible for calcium-binding in Cam and is the next logical target for a set of mutations.¹⁶ Even though position 7 binds metals through the main chain, we reasoned that changes to the side chain may alter the tertiary structure enough to have an impact on metal binding. The Y99 and Q135 residues were targeted for mutation into either Glu or Cys. However, the Q135C site-directed mutagenesis failed in multiple attempts and the Q135E mutation, while successful, failed to express in cells. Kemp assays were performed but showed that Y99E and Y99C exhibited decreased activity in the presence of calcium (Table 2.2). Y99E and Y99C mutants were not active with lead(II) or yttrium(III).

Table 2.2. Summary of activity of position 7 mutants after addition of 0.1 mM Ca(NO₃)₂, 0.1 mM Pb(NO₃)₂, or 0.1 mM Y(NO₃)₃. All assays were run in 20mM Hepes at pH 7.0 with 5.0μM protein.

$k_{cat}/K_M (M^{-1}s^{-1})$	Ca^{2+}	Pb^{2+}	Y^{3+}
<i>Ac7</i>	462±24 M ⁻¹ s ⁻¹	403±62 M ⁻¹ s ⁻¹	32±48 M ⁻¹ s ⁻¹
<i>AcY99E</i>	195±18 M ⁻¹ s ⁻¹	---	3 M ⁻¹ s ⁻¹
<i>AcY99C</i>	369±93 M ⁻¹ s ⁻¹	---	5 M ⁻¹ s ⁻¹

2.2.3. Position 3 – D95 and D131

We know that positions 1, 3, and 5 are directly involved in binding calcium ions. Previous work in our lab has showed that position 5 mutations result in proteins that are completely inactive and position 1 is essential for protein structure.²⁵ However, little information was

available on the consequences of mutating position 3. When consulting the crystal structures of calmodulin bound to calcium vs lead (Figure 2.14), position 3 showed a difference of 0.6 angstroms between the two ions.⁵¹ This lead us to believe that this position might give us the selectivity we desired.

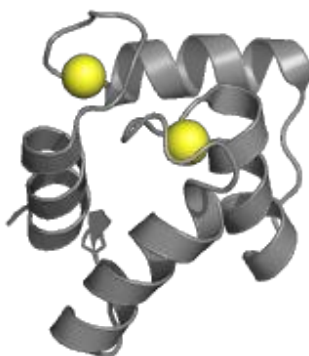


Figure 2.14. Three-dimensional structure of HolIEE constructed in PyMol using Cam structure (pdb code 1c1l) as a template. The yellow spheres represent yttrium ions.¹

Preliminary data showed that mutating D95 and D131 into either Glu or Cys almost completely eliminated activity of the protein in the presence of calcium (Table 2.3). However, while activity was decreased in comparison to Ac7, the EE and CC mutations of Ac7, called HolIEE and ReCCes respectively, showed significant activity in yttrium when compared to calcium. While ReCCes also showed potential selectivity for Pb^{2+} , the difficulty in expressing and working with this protein caused us to set it aside in favor of HolIEE.

Table 2.3. Activity of HolIEE and ReCCes in presence of calcium, lead, and yttrium. All assays were run in 20 mM Hepes at pH 7.0 with 5.0 μ M protein.

$k_{cat}/K_M (M^{-1}s^{-1})$	<i>AlleyCat 7</i>	<i>HolIEE</i>	<i>ReCCes</i>
0.1 mM Ca^{2+}	526	0.3	0
0.25 mM Ca^{2+}	580	1.7	0.5
0.1 mM Pb^{2+}	--	1.0	3.6
0.25 mM Pb^{2+}	286	0	4.0
0.1 mM Y^{3+}	190	47	32
0.25 mM Y^{3+}	102	14	0

An important property to note from Table 2.3 is that the activity of AlleyCat7, HolIEE, and ReCCes all significantly decrease as the concentration of yttrium increases. This is due to solubility issues. If concentrations exceeding 0.1 mM of lanthanide, the protein begins to precipitate out of solution (observed as a white powder) which decreases the measured activity.

2.2.4. MALDI-TOF

A MALDI-TOF analysis was also performed on the protein (Figure 2.15) to confirm its identity. The m/z of the protein on the MALDI spectrum matched the predicted molecular weight, indicating that the isolated protein was most likely HolIEE and not misidentified.

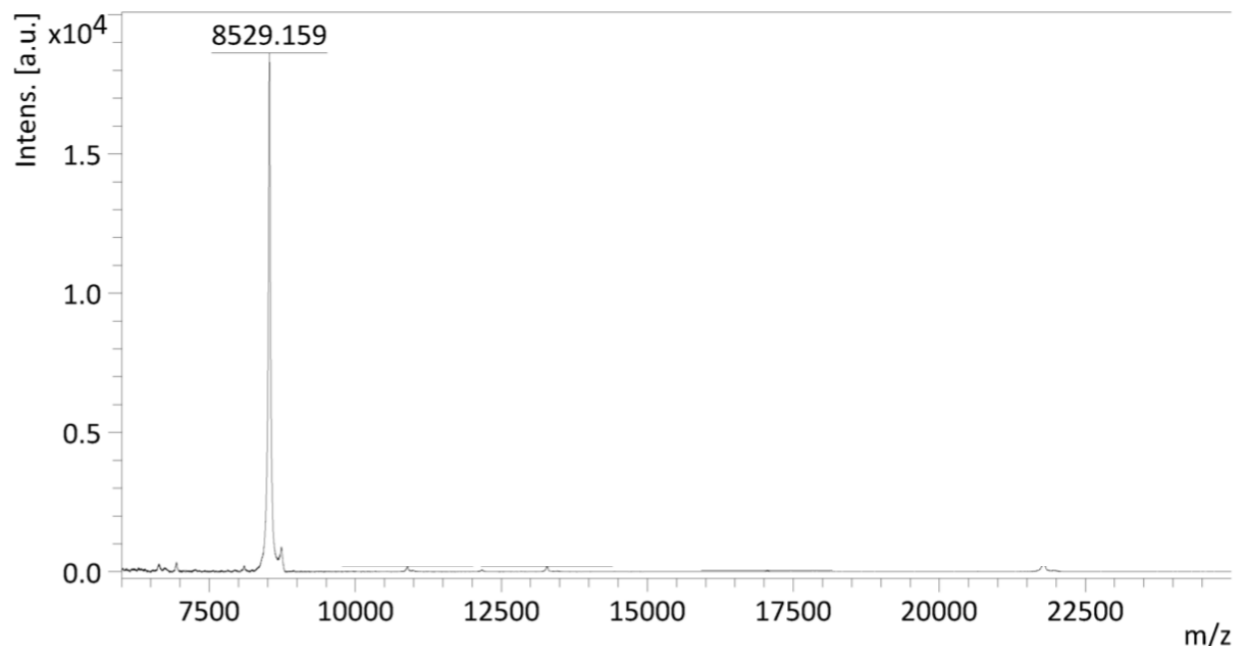
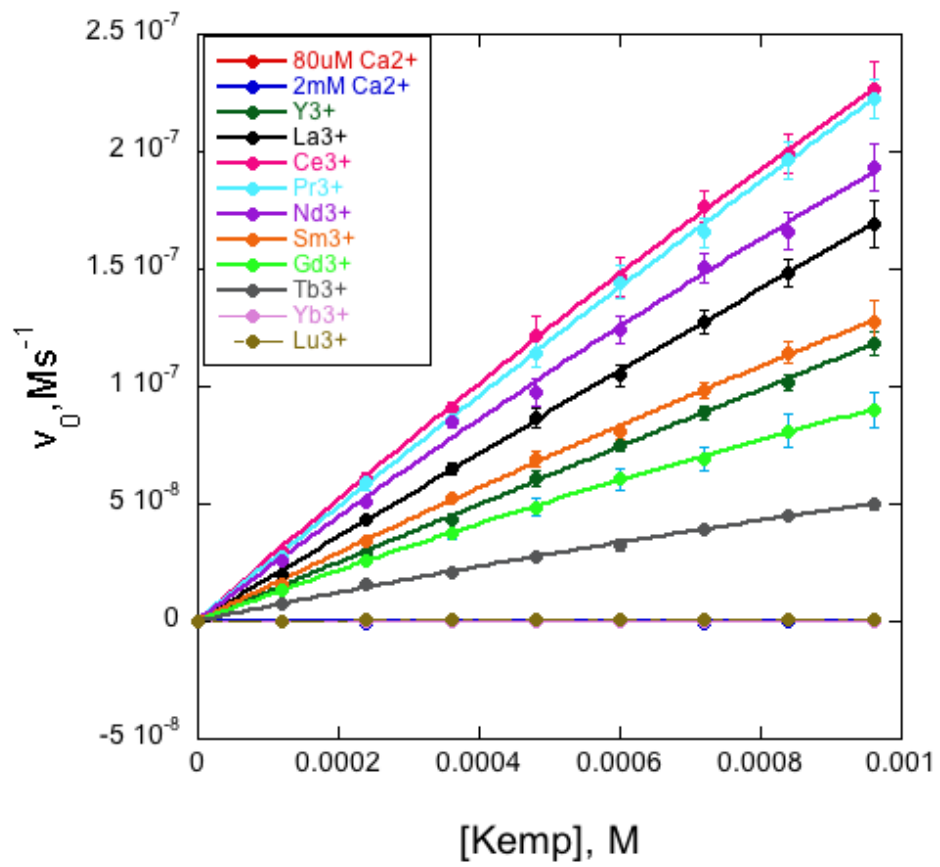


Figure 2.15. MALDI-TOF of cut HolIEE. The calculated mass of HolIEE is 8530 Da.

2.3. Characterizing HolIEE with different lanthanides

2.3.1. Kinetic assays

Kemp elimination assays of HolIEE were run with a total of ten different lanthanides: yttrium, lanthanum, cerium, neodymium, praseodymium, samarium, gadolinium, terbium, ytterbium, and lutetium. By examining the influence of these lanthanides on Michaelis-Menten kinetics of Kemp elimination, along with calcium, it can be concluded that HolIEE is highly selective for lanthanides even in concentrations of calcium typically found in the body (Figure 2.16 and Table 2.4).



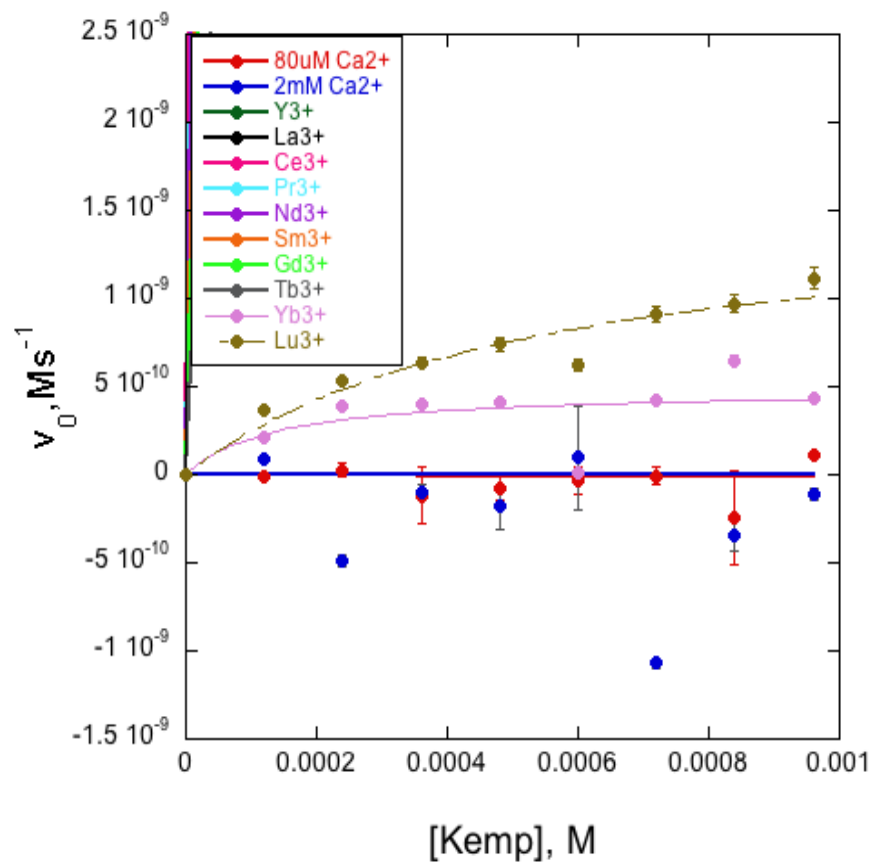


Figure 2.16. Catalytic efficiency of Kemp elimination with HolEE in the presence of various lanthanides. The left graph pictures all the metals tested while the right graph only features lutetium, ytterbium, and calcium. All samples were run in 20 mM Mops, 100 mM NaCl, pH 7.0 with 0.08 mM metal unless otherwise noted. The concentration of HolEE was 1.0 μM . All k_{cat}/K_M are reported in Table 2.4. (MT7081)

Table 2.4. Summary of HolIEE activities in 20 mM Mops, 100 mM NaCl at pH 7.0. The concentration of all metals was 0.08 mM unless otherwise noted. (MT7081)

	2mM	Ca ²⁺	Y ³⁺	La ³⁺	Ce ³⁺	Pr ³⁺	Nd ³⁺	Sm ³⁺	Gd ³⁺	Tb ³⁺	Yb ³⁺	Lu ³⁺
Ca ²⁺												
k_{cat}/K_M	0.03	0.03	123	176	236	231	199	133	94	52	3.7	2.9
(M ⁻¹ s ⁻¹)												

When the catalytic activity is correlated to the ionic radius of the +3 charged lanthanide, there is a clear trend of increased activity with increased ionic radius (Figure 2.17).⁵² Because different lanthanides have different optimal concentrations with protein, it is difficult to find one metal concentration to measure all k_{cat}/K_M at, which can explain why the yttrium is higher than the trend estimates (Y-HolIEE complex is most active at 80 μ M metal) and lanthanum is lower than expected (La-HolIEE complex is most active at higher metal concentrations).

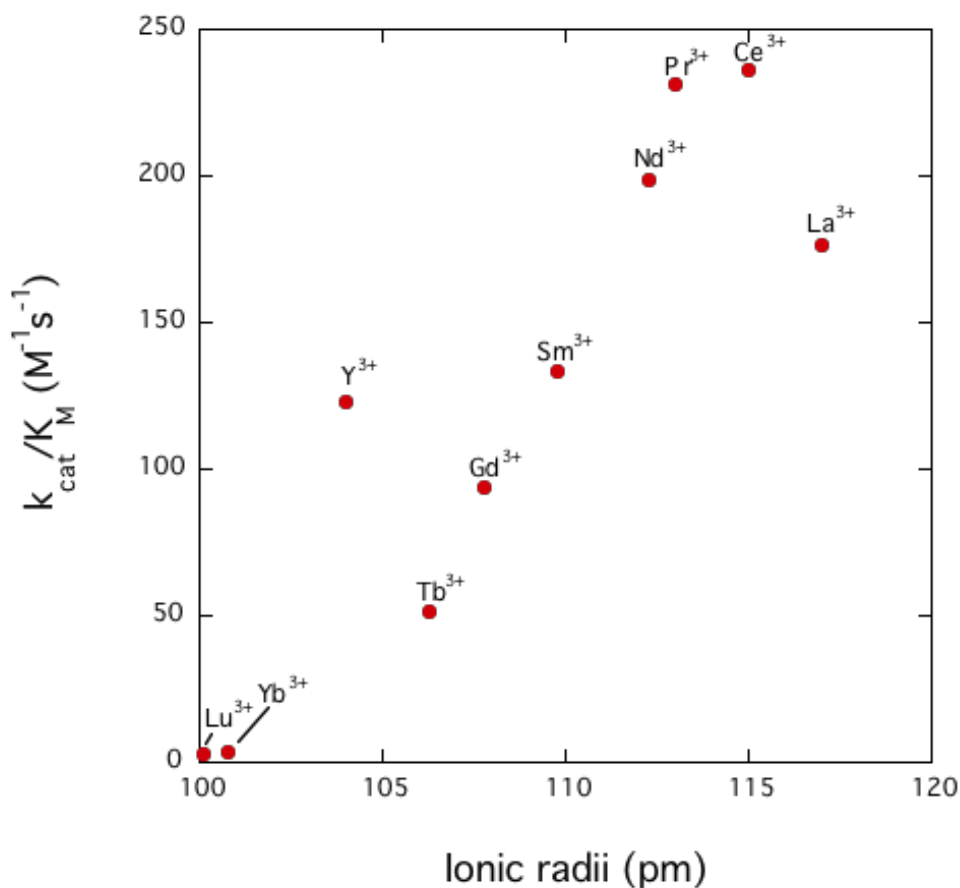


Figure 2.17. Correlation between ionic radius (pm) and catalytic activity with 0.08 mM metal concentration. All ionic radii for the lanthanides are calculated for six-coordinate octahedral geometry.⁵² (MT7081)

It is possible that the larger ions allow the protein to conform to a more optimal shape for catalysis than smaller ions (Figure 2.18). Since catalytic activity of HolIEE is dependent upon the exposure of the substrate binding pocket, which in turn is dependent upon the position of the EF-hands, a smaller ion might change shape so that Kemp elimination is harder to occur. However, this hypothesis does not necessarily mean that the binding of HolIEE to smaller ions is weaker than HolIEE to larger ions, merely that catalysis is not ideal.

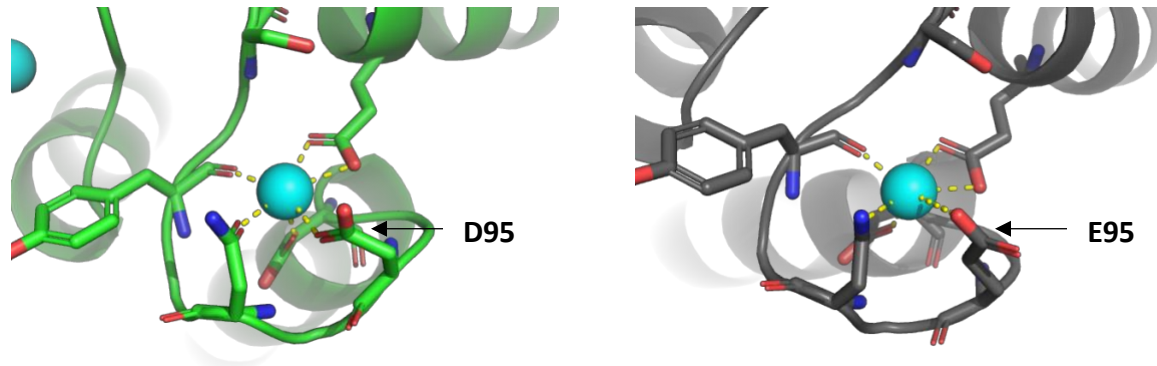


Figure 2.18. Active site of AlleyCat (left) and HolIEE (right) in the presence of calcium (cyan sphere). Both figures were generated in PyMol using 1cII as a template.

The data show that larger ions tend to retain more activity at higher concentrations of metal than smaller ions (Figure 2.19). For most of the smaller ions, higher ion concentrations are needed to achieve maximum catalytic activity, but higher concentrations of the metal tend to cause protein precipitation. Meanwhile, larger lanthanides are less likely to precipitate out protein at higher concentrations. These issues with precipitation do limit the scope of the protein to smaller concentrations of metal ion, which is desirable from a sensor standpoint, but make

comparison between metals difficult since each metal has a different optimum metal concentration.

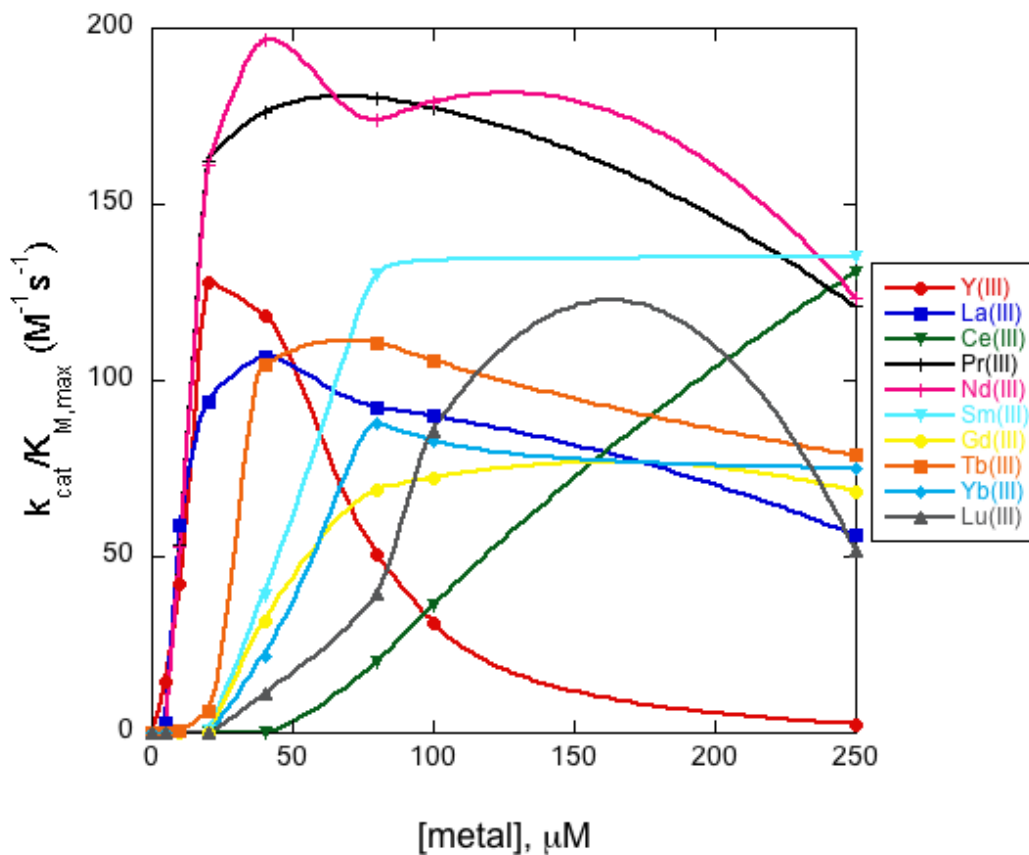


Figure 2.19. Catalytic efficiency of Kemp elimination using 2.5 μM HolEE in 20 mM Mops at pH 7.0. Kinetic activity was measured in the presence of different metals and varying concentrations. (MT7047)

2.3.2. pH profiling

Attempts were made to determine the pK_a of HolIEE using pH profiles and determine the effective pH range of HolIEE. Unfortunately, due to precipitation issues with even low concentrations of lanthanides, none of the data could be fit to a curve (described in chapter 2.5.6). The pH profile graphs of HolIEE with different metals can be found in the Appendix I, section S.2.3.

2.3.3. Isothermal Titration Calorimetry

A byproduct of chemical and physical change in reactions is the release or absorbance of heat.⁵³⁻⁵⁴ Isothermal titration calorimetry (ITC) is a favored method for determining protein-ligand interactions since it requires no labelling, is highly sensitive, and can be used with substrates that are undetected by spectroscopic methods. ITC works by slowly titrating ligand (or in this case, a metal ion) into a protein and measuring the heat released or absorbed by the how much energy is required for the instrument to maintain a temperature that is consistent with that of the solution in a reference cell (Figure 2.20).⁵⁴⁻⁵⁵ As the protein's binding sites become saturated, the heat absorbed/released approaches zero. From this isotherm, a number of properties can be determined such as the enthalpy of binding (ΔH), the entropy of binding (ΔS), the dissociation constant (K_d), binding stoichiometry, the free energy of binding (ΔG), and cooperativity (site-site interactions).⁵⁵

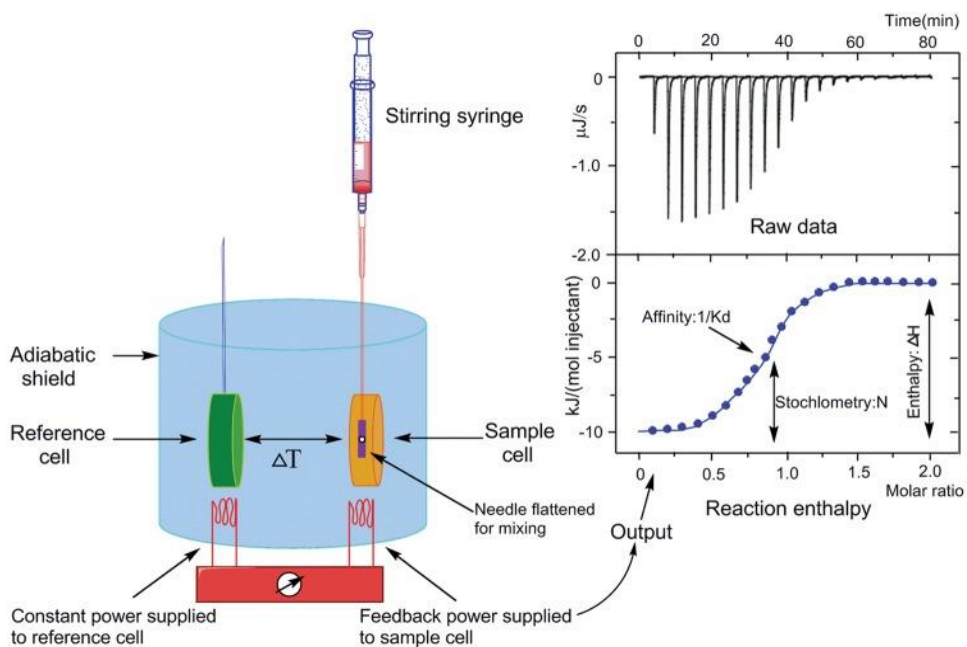


Figure 2.20. Schematic of basic experimental setup for ITC (left) and the typical data acquired from ITC (right).⁵⁵

Our expectation was that when aliquots of $\text{Ca}(\text{NO}_3)_3$ were added to HolIEE, the heat of reaction would either remain at 0 kJ/mol or would quickly level off to zero (Figure 2.20). Meanwhile, in the presence of increasing amounts of $\text{Y}(\text{NO}_3)_3$, we expected an isotherm much like that seen in AlleyCat7 when titrated with calcium (Figure S.25). The AlleyCat7 isotherm curve is indicative of either a macromolecule with two identical and independent binding sites for substrate, a macromolecule with two nonidentical binding sites, or a macromolecule with two identical sites that display negative cooperativity (the binding in the first site decreases the macromolecule's affinity for binding in the second site).⁵⁶ Since different EF-hands in Cam are known to have different binding affinities for calcium, we expect the second case to be true.²⁰

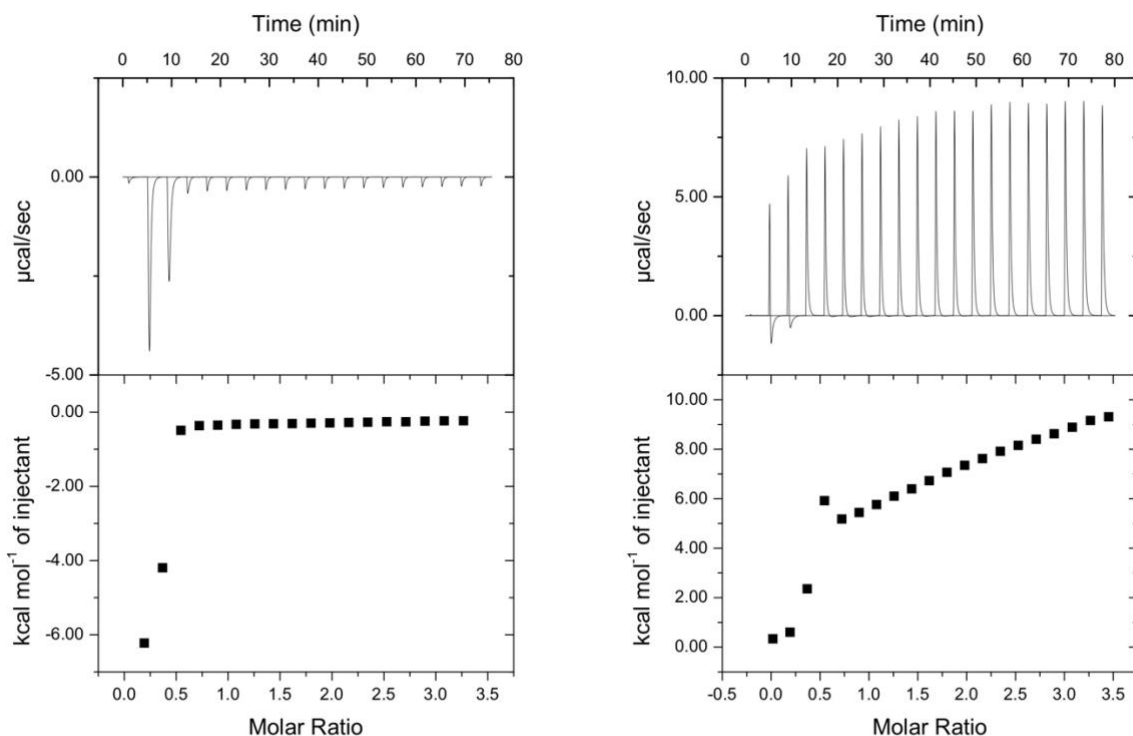


Figure 2.21. ITC data of HolIEE titrated with Ca(NO₃)₂ (left) and Y(NO₃)₃ (right). The sample cell contained 0.1 mM HolIEE in 20 mM Mops pH 7.0 at 25°C and was slowly titrated with either 5 mM Ca(NO₃)₂ or Y(NO₃)₃ in 5 µL injections with an equilibration time of 240 seconds. (MT6155)

HolIEE releases very minimal amounts of heat when titrated with calcium as expected and shows no binding affinity for calcium. The results obtained from titrating HolIEE with yttrium were not consistent with Ac7 (Figure S.25) or a protein with two nonidentical binding sites. The results are inconclusive. We know that binding must occur since HolIEE is able to catalyze the Kemp elimination reaction in the presence of yttrium but not in calcium. However, the ITC data for yttrium titration never reaches 0 kJ/mol, thus signaling the completion of reaction, even in the presence of saturating concentrations of yttrium. It was expected that HolIEE would have a 2:1 binding ratio of metal:protein, but this is not shown in Figure 2.21. Furthermore, calcium

binding in Ac7 is an exothermic process, but the positive values of yttrium titration suggest an endothermic process.

We ultimately decided that using ITC to determine dissociation coefficient, K_d , of metal ions to HolIEE is not a viable route. We were able to conclude that calcium does not bind to HolIEE, but little else. We believe that our unusual results are likely due to HolIEE's tendency to precipitate at metal concentrations exceeding 100 μM , which could impact our readings greatly since the concentration of protein is already 100 μM in our experiments.⁵⁷ Unfortunately, experimental costs limited our ability to test this hypothesis. This prompted us to determine a new method to calculate K_d using other measurable parameters.⁵⁸⁻⁵⁹

2.3.4. Circular Dichroism analysis of binding

Another way to determine the binding efficacy of HolIEE with metals is to observe the secondary structure of HolIEE for different metal ions. This can be measured using circular dichroism (CD), a technique that allows us to visualize secondary structure. As stated above, the EF-hands responsible for metal binding consist of α -helices. Secondary structures such as α -helix, β -sheet, disordered or random coils, and even denatured proteins have unique signatures in the CD spectrum (Figure 2.22).⁶⁰⁻⁶¹

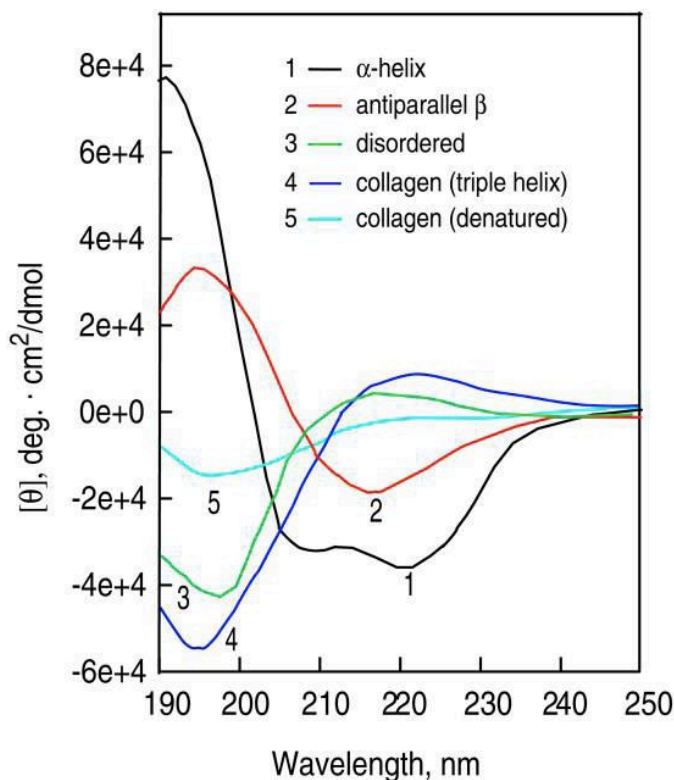


Figure 2.22. Secondary structures visualized by circular dichroism.⁶¹ Each secondary structure has a unique readout that can be used to analyze protein and peptide samples.

When CD spectra of HolIEE are measured using 100 μM concentration of different metals, there is a distinct difference between the structure of HolIEE without metal added and with metal (Figure 2.23). Without metal, HolIEE is more disordered and thus, not expected to be catalytically active for Kemp elimination based on the properties of Cam. However, HolIEE exhibits significant α -helical structure in the presence of rare earth metals. There also appears to be a correlation between Kemp elimination activity and the degree of α -helicity; metals that give greater activity tend to exhibit lower mean residue ellipticity (MRE) compared to lower activity metals. HolIEE in the presence of EDTA (a metal chelator) and calcium exhibits a similar curve to that of HolIEE without any metal added at all.

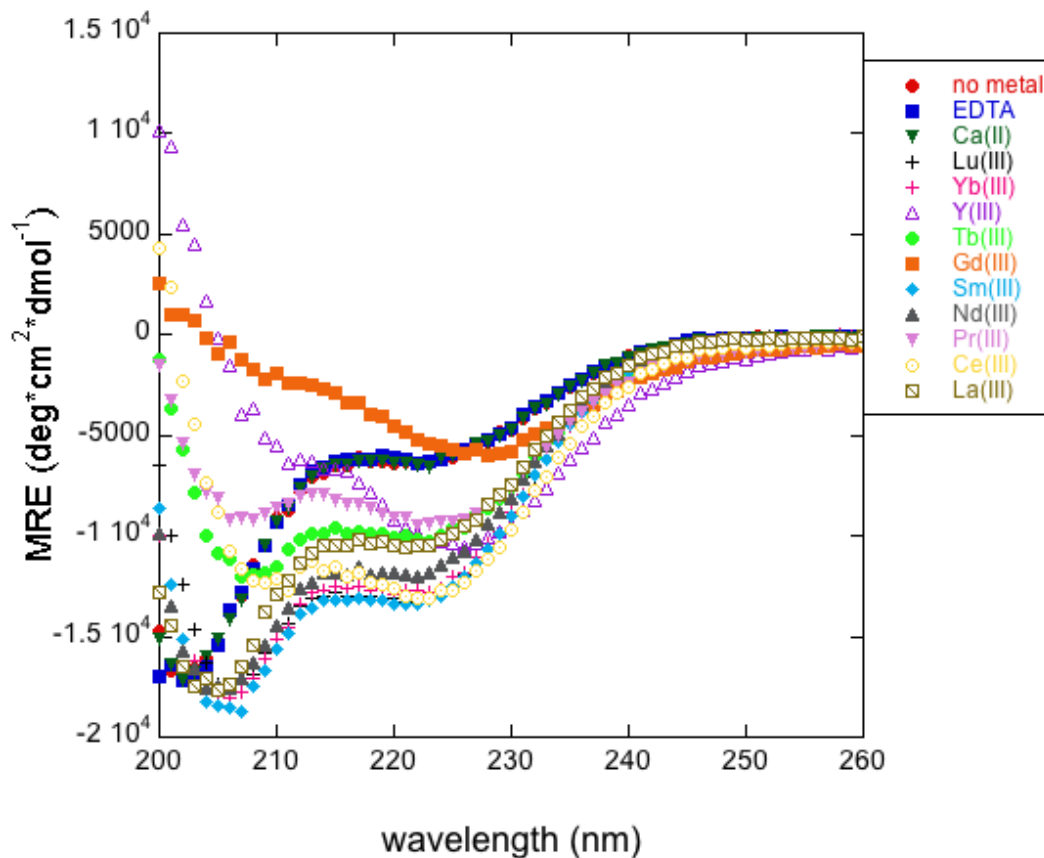


Figure 2.23. Circular dichroism of HolIEE (25 μM , 4 mM Mops, pH 7.0, 22°C) with 100 μM of different metals. (MT7011)

When HolIEE is titrated with a metal ion, lutetium for example, there is a clear trend in MRE as a function of metal ion concentration (Figures 2.24 and 2.25). Because HolIEE has two EF-hands, it requires at least two metal ions to bind to it to be fully functional. The data in Figure 2.24 shows that once the concentration of lutetium reaches 50 μM , any concentration higher than that has roughly similar MRE values (until precipitation starts to occur). This is further shown in Figure 2.25.

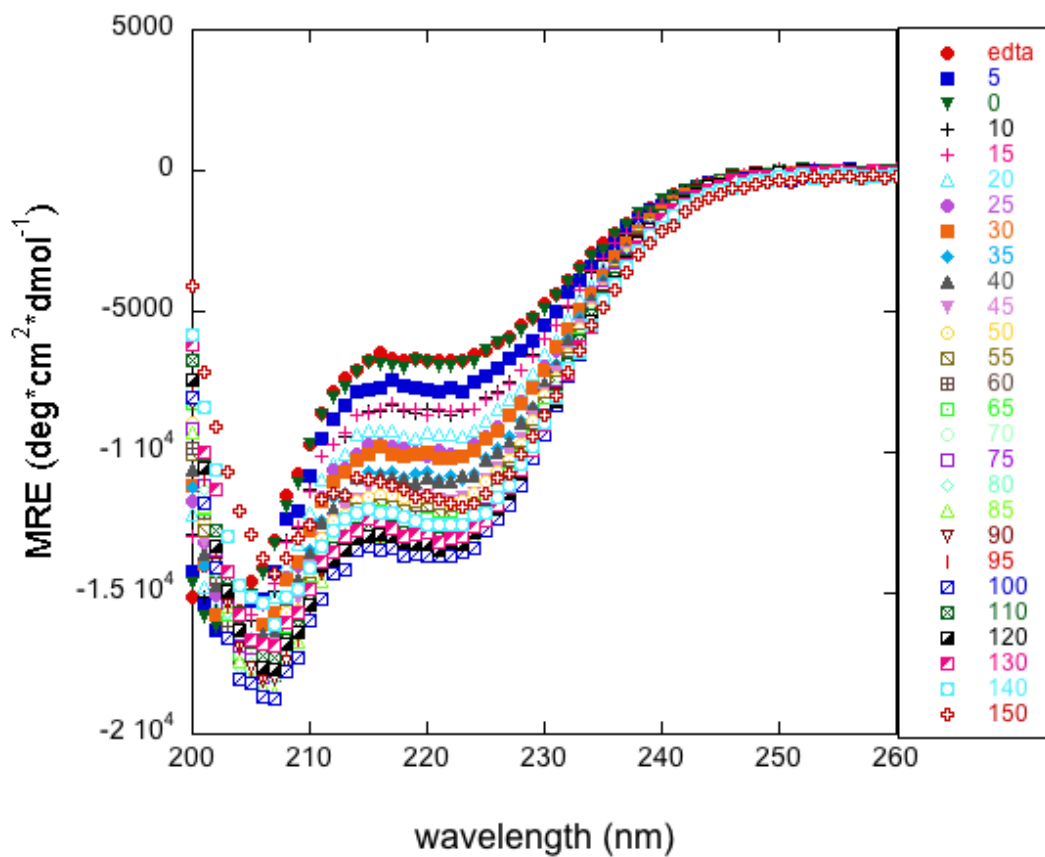


Figure 2.24. Circular dichroism of HolIIEE (25 μM) in 4 mM Mops pH 7.0 with varying concentrations of LuCl₃, which are listed in the legend (except in the case of EDTA which added a concentration of 0.1 mM EDTA instead). An increase in the concentration of metal ions results in increasingly negative MRE at 207 nm and 220 nm. (MT7026)

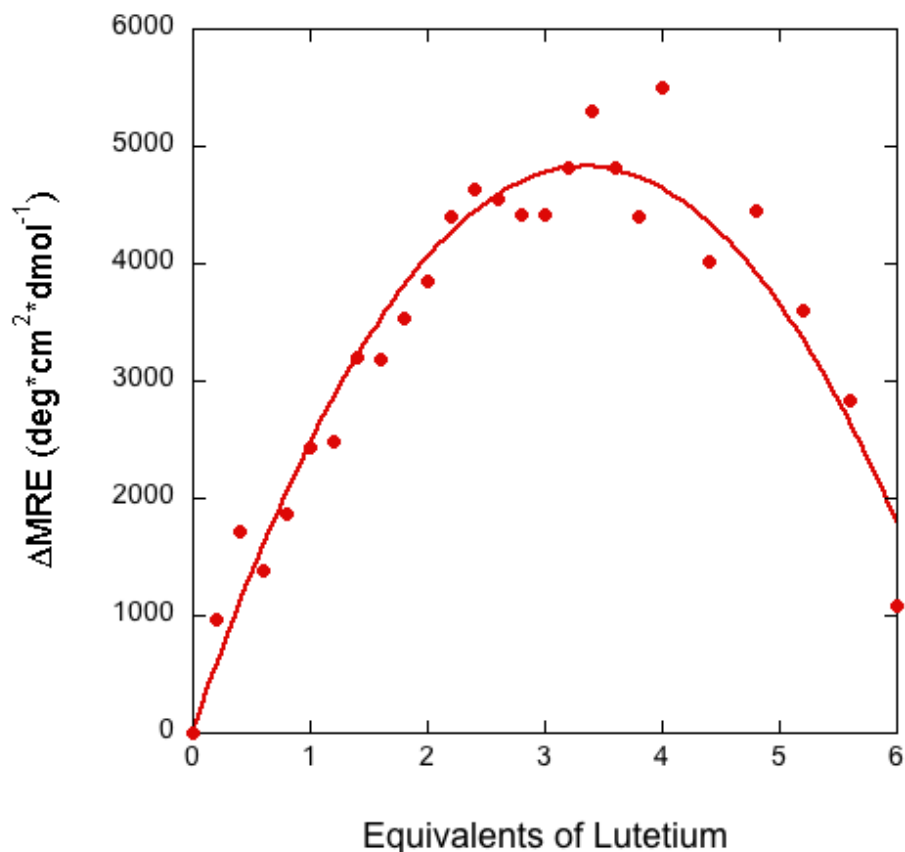


Figure 2.25. Analysis of metal titration of Figure 2.24 at 220 nm. The ΔMRE is calculated by subtracting mean residue ellipticity at 220 nm from the MRE at 220 nm of HolIEE in 0.1 mM EDTA. (MT7026)

Since there is a measurable in mean residue ellipticity with change in metal concentration, it is possible that these measurements can be used to calculate K_d .⁵⁸⁻⁵⁹ We can fit Figure 2.25 to the following curve: $\Delta MRE = (Y_1[E]K_1x + Y_2[E]K_1^2x^2) / (1 + K_1x + K_1^2x^2)$, where $[E]$ is the concentration of HolIEE, x is the concentration of metal ions added, Y_1 represents the MRE value when one metal ion is bound to HolIEE, K_1 represents the association coefficient of the first metal ion to HolIEE, Y_2 represents the MRE value when the second metal ion binds to HolIEE, and K_2 represents

the association coefficient when the second metal ion binds.⁵⁹ The association coefficient, K_a , can be found by multiplying K_1 and K_2 and by assuming that K_1 is equal to K_2 , this equation can be further simplified. Then to find the dissociation coefficient K_d , the inverse of K_a is calculated (Figure 2.26).

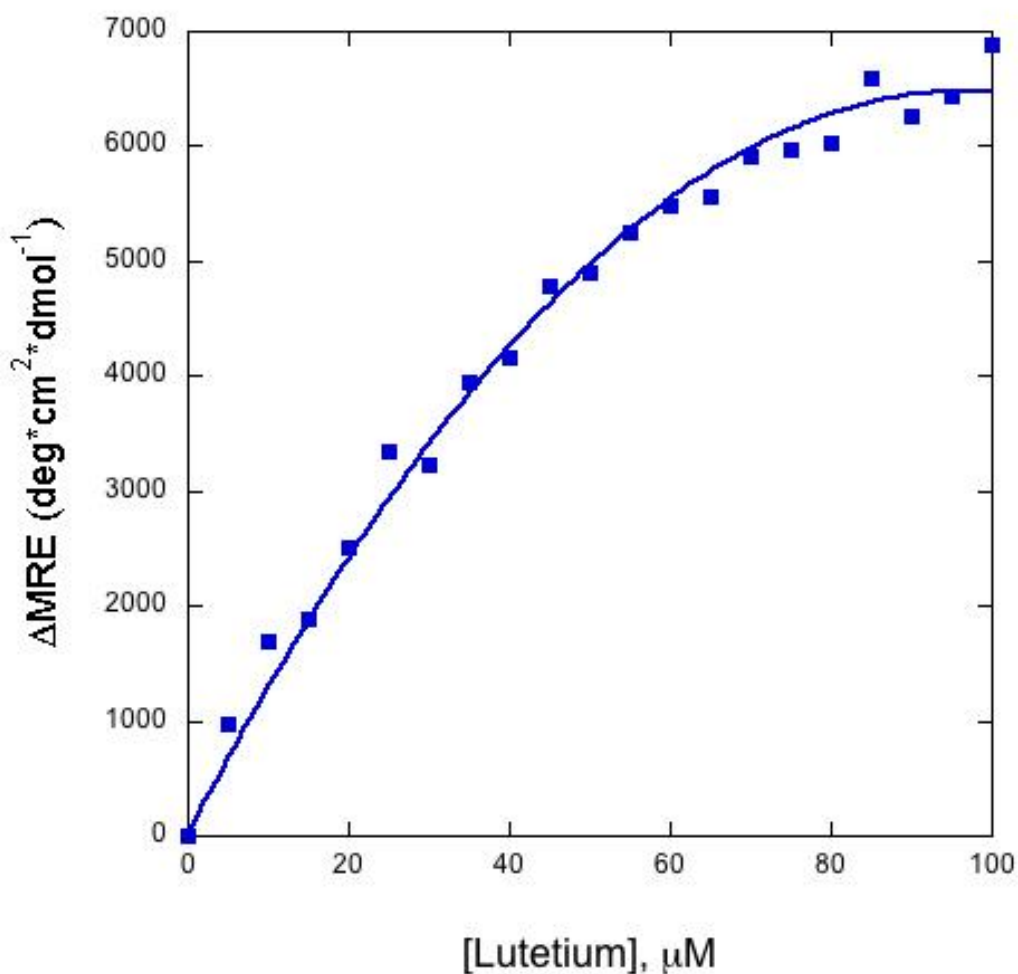


Figure 2.26. K_d analysis of HolIEE (25 μM , 4mM Mops, pH 7.0, 22 $^\circ\text{C}$) with LuCl_3 titration using mean residue ellipticity at 220nm. (MT7026)

Unfortunately, once again due to HolIEE's tendency to precipitate, it is difficult to saturate HolIEE enough to create a curve that can obtain a reasonable value. While CD metal titration is a promising method for determining K_d , this method still needs to be optimized.

2.4. Conclusions

Designing proteins that are capable of binding non-typical metals is challenging.⁴¹ The design of metalloproteins is particularly difficult but being able to control metal binding gives important clues to how metalloproteins can and may have developed metal ion specificity. By taking an existing scaffold such as Cam, we were able to alter the sequence at the metal binding site to accept exclusively rare earth metals, something not previously done with such great success. This also paves the way to creating proteins that could bind other metals, such as heavy metals, another important goal. A protein that can bind rare earth metals exclusively could have many applications, for example, as a metal sensor, in reclamation of materials, and even in clinical treatments; the latter topic is discussed in the next chapter.

2.5. Experimental

2.5.1. Site-directed Mutagenesis

AlleyCat7 gene was previously cloned into Champion pET-SUMO/CAT vector (Invitrogen). Plasmid contains a N-terminal His₆-tag separated from the gene of interest by a SUMO protease recognition fragment (GG).

The appropriate mutations were made using a standard mutagenesis protocol for *iProof* High-Fidelity DNA Polymerase, which was provided by the manufacturer. After PCR, the template

was digested with *DpnI* restriction enzyme and transformed into *E. coli* XL-10 cells. DNA was extracted from colonies using EZ-10 Spin Column Plasmid DNA kit and mutated sequences were confirmed by Genescript using Sanger sequencing technique.

2.5.2. Expression of AlleyCat7 mutants

The plasmid with a gene coding for AlleCat7 or its mutants were transformed into *E. coli* BL21(DE3) pLysS cells and expressed using LB broth. After cells reached an OD₆₀₀ of ~0.5, IPTG was added to a final concentration of 0.5 mM and the temperature was lowered to 30°C. After four hours, the cells were collected by centrifugation and stored at -80°C until needed.

Cell pellet was resuspended in lysis buffer containing 25 mM Tris (Tris was used instead of Tris-HCl due to worries of precipitation of metal when combined with chlorine ions), 20 mM imidazole, 0.5 mM PMSF, and 300 mM NaCl (pH 8). In addition, 5 µL of Benzonase (25 U/µL, Novagen) is added for every 1 g of cell pellet. Cells were lysed by sonication, and then allowed to rock gently at room temperature for 30 min. The crude cell lysate was centrifuged at 20,000 \times g for 30 minutes. The supernatant was applied to a Ni-NTA (His60 Ni Superflow resin, TaKaRa) column and washed with lysis buffer (without PMSF). The protein was then eluted with buffer containing 25 mM Tris, 250 mM imidazole, and 300 mM NaCl (pH 8). The sample was exchanged into buffer containing 50 mM Tris, 75 mM NaCl (pH 8) by applying it onto a Bio-Rad 10 DG desalting column. Protein concentration was determined using UV-Vis spectroscopy where $\epsilon_{280} = 5445 \text{ cm}^{-1}\text{M}^{-1}$.

2.5.3. Removal of His₆-tag

SUMO protease was transformed into *E. coli* BL21 RP cells and expressed using LB broth at 30°C. After cells reached OD₆₀₀ ~0.5, culture was induced with 0.5 mM IPTG and the cells harvested by centrifugation (4000xg, 20 min) after 4 hours. Cells were resuspended in 25 mM Tris, 5% glycerol, 10 mM β-ME, pH 7.6 and lysed using sonication (20 sec intervals). Debris was collected by centrifugation (4°C, 20,000xg, 30 min). Streptomycin sulfate was then added dropwise to 0.1% and stirred on ice for 30 minutes. Debris was once again collected by centrifugation (4°C, 20,000xg, 30 min). Supernatant was loaded onto Ni-NTA column and eluted off using buffer containing 25 mM Tris, 5% glycerol, 250 mM imidazole, pH 7.6. Fractions with protein were loaded onto a Q-Sepharose column and washed with original lysis buffer, then eluted with buffer containing 25 mM Tris, 10 mM β-ME, 5% glycerol, 200 mM NaCl, pH 7.6. Samples were stored at -80°C until needed.

To cleave His-tag, uncut AlleyCat sample was incubated with 1.0 mM DTT, 0.5 mM EDTA, and SUMO protease at a 200:1 ratio (absorbance ratio at 280 nm) overnight at 30°C. To ensure integrity, cleavage solution was filter sterilized using 0.22 μm PES filter before incubation.

After incubation, DTT and EDTA were removed by applying sample onto a Bio-Rad 10 DG desalting column with 50 mM Tris, 75 mM NaCl (pH 8) buffer. Desalted samples were then applied to Ni-NTA column equilibrated with same buffer and allowed to sit on column for minimum one hour. Afterwards, protein was collected in flow through and wash and the presence of protein was determined using UV-Vis spectroscopy.

To thoroughly remove metal, approximately 5 equivalents of EDTA was added to the collected fractions. The sample was then exchanged into 20 mM Mops (pH 7.0) twice using a Bio-Rad 10 DG desalting column to ensure removal of excess calcium ions and EDTA.

The purity of the protein was established by SDS-PAGE using a 15% acrylamide gel with Unstained Precision Protein Plus ladder. Protein concentration was determined using UV-Vis spectroscopy where $\epsilon_{280} = 2967 \text{ cm}^{-1}\text{M}^{-1}$.

2.5.4. Kinetic assays

Kemp substrate was synthesized using protocol outlined in Casey *et al.*⁶²

Kinetic measurements were done with either a Thermo Lab-systems Multiskan Spectrum Platerreader or a BioTek Eon Micro Platerreader monitoring the absorbance at 380 nm at 22°C using at least three independent measurements. The final protein concentration (determined by UV-Vis spectroscopy using $\epsilon_{280} = 2967 \text{ M}^{-1} \text{ cm}^{-1}$) for kinetic measurements was 1.0 μM . Final substrate concentrations ranged from 0.12 to 0.96 mM. Kinetic parameters ($k_{\text{cat}}/K_{\text{M}}$) were obtained by fitting the linear portion of data to the Michaelis-Menten equation: $[v_0 = k_{\text{cat}}[E]_0[S]_0/(K_{\text{M}} + [S]_0)]$ using Kaleidagraph 4.5 (Synergy Software) or by fitting the data to a linear curve using Kaleidagraph 4.5.

2.5.5. Metal Dependence Assays

Kinetic measurements were done with either a Thermo Lab-systems Multiskan Spectrum Platerreader or a BioTek Eon Micro Platerreader monitoring the absorbance at 380 nm at 22°C using at least three independent measurements. The final protein concentration (determined by

UV-Vis spectroscopy using $\epsilon_{280} = 5445 \text{ M}^{-1} \text{ cm}^{-1}$ for uncut protein and $\epsilon_{280} = 2967 \text{ M}^{-1} \text{ cm}^{-1}$ for cut protein) for kinetic measurements was $2.5 \mu\text{M}$. Final metal concentrations ranged from 0.001 to 0.25 mM . In addition, a concentration of $15 \mu\text{M}$ EDTA was used to establish activity without metal ions present.

2.5.6. pH profiles

Kinetic measurements were done with either a Thermo Lab-systems Multiskan Spectrum Platerreader or a BioTek Eon Micro Platerreader monitoring the absorbance at 380 nm at 22°C using at least three independent measurements. The final substrate concentration was 0.6 mM . The extinction coefficient of $15,800 \text{ cm}^{-1}\text{M}^{-1}$ was used to determine product concentration at pH 5.5 and above. At pH 4.5 and 5.0 , the product coefficients of $12,480 \text{ cm}^{-1}\text{M}^{-1}$ and $14,220 \text{ cm}^{-1}\text{M}^{-1}$, respectively, were used. Samples were run in acetate buffer (pH 4.5 - 5.5), Mes (pH 6.0 - 6.5), Mops (pH 7.0 - 7.5), and Tris (pH 8.0 - 8.5) with concentrations of 50 mM . The final protein concentration (determined by UV-Vis spectroscopy using $\epsilon_{280} = 5445 \text{ M}^{-1} \text{ cm}^{-1}$ for uncut protein and $\epsilon_{280} = 2967 \text{ M}^{-1} \text{ cm}^{-1}$ for cut protein) for kinetic measurements was $1.0 \mu\text{M}$. The curves were fit to $[(k_{\text{cat}}/K_{\text{M}})_{\text{max}} = (k_{\text{cat}}/K_{\text{M}})_{\text{protonated}} + (k_{\text{cat}}/K_{\text{M}})_{\text{deprotonated}} \times 10^{-\text{pK}_a} / (10^{-\text{pH}} + 10^{-\text{pK}_a})]$, where pK_a is the apparent pK_a value of the active residue, using Kaleidagraph 4.5 (Synergy software).

2.5.7. Circular Dichroism

The CD spectra were collected on the Jasco J-715 CD spectrometer, using a step scan mode (4 sec averaging time) averaging over three runs, using a quartz cuvette with a 0.1 cm path length. Each sample contained $25 \mu\text{M}$ protein in 4 mM Mops pH 7.0 and the final volume of each

sample was 300 μL . Samples contained varying amounts of metal chloride salts and were measured within five minutes of addition of metal ions.

Care was taken so the sample absorbance never exceeded 1.8 at all wavelengths to produce reliable ellipticity values.

2.5.8. MALDI-TOF Mass Spectrometry

MALDI-TOF spectrometry was performed on a Bruker Autoflex III instrument. The instrument was calibrated with standard samples provided by the manufacturer and sinapinic acid was used as a matrix. Protein ($\sim 10 \mu\text{M}$) was mixed with matrix at a 1:2 volume ratio.

2.5.9. Isothermal Titration Calorimetry

Apo Ac7 and HolIEE were freshly purified and buffer exchanged into 20 mM Mops pH 7.0 using dialysis (Slide-A-Lyzer Dialysis Casettee G2, 10k MWCO, ThermoScientific). Samples were frozen with liquid N_2 until the night before, then thawed at 4°C overnight. Each sample used 100 μM of protein at 25°C and were titrated with 5 mM metal nitrate stock in 5 μL increments. Each measurement was allowed to equilibrate for 240 s.

2.6. References

1. Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quijcho, F. A., Calmodulin structure refined at 1.7 \AA resolution. *J Mol Biol* **1992**, *228* (4), 1177-92.
2. Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009**, *460* (7257), 823-30.

3. Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M., Design of functional metalloproteins. *Nature* **2009**, *460* (7257), 855-62.
4. Kirberger, M.; Yang, J. J., Structural differences between Pb²⁺- and Ca²⁺-binding sites in proteins: implications with respect to toxicity. *J Inorg Biochem* **2008**, *102* (10), 1901-9.
5. Holm, R. H. K., P.; Solomon, E.I., Structural and Functional Aspects of Metal Sites in Biology. *Chem Rev* **1996**, *96*, 2239-314.
6. Tainer, J. A. R., V.A.; Getzoff, E.D., Metal-binding sites in proteins. *Curr Opin Biotechnol* **1991**, *2*, 582-91.
7. Naik, P. K.; Ranjan, P.; Kesari, P.; Jain, S., MetalloPred: A tool for hierarchical prediction of metal ion binding proteins using cluster of neural networks and sequence derived features. *Journal of Biophysical Chemistry* **2011**, *02* (02), 112-23.
8. Vu, H. N.; Subuyuj, G. A.; Vijayakumar, S.; Good, N. M.; Martinez-Gomez, N. C.; Skovran, E., Lanthanide-Dependent Regulation of Methanol Oxidation Systems in *Methylobacterium extorquens* AM1 and Their Contribution to Methanol Growth. *J Bacteriol* **2016**, *198* (8), 1250-9.
9. Pol, A.; Barends, T. R.; Dietl, A.; Khadem, A. F.; Eygensteyn, J.; Jetten, M. S.; Op den Camp, H. J., Rare earth metals are essential for methanotrophic life in volcanic mudpots. *Environ Microbiol* **2014**, *16* (1), 255-64.
10. Skovran, E. M.-G., Norma Cecelia, Just add lanthanides: Some methanol-using bacteria may depend on lanthanide elements for carbon capture and energy generation. *Science* **2015**, *348* (6237), 862-3.

11. McSkimming, A.; Cheisson, T.; Carroll, P. J.; Schelter, E. J., Functional Synthetic Model for the Lanthanide-Dependent Quinoid Alcohol Dehydrogenase Active Site. *J Am Chem Soc* **2018**, *140* (4), 1223-6.
12. Hynninen, A.; Touze, T.; Pitkanen, L.; Mengin-Lecreulx, D.; Virta, M., An efflux transporter PbrA and a phosphatase PbrB cooperate in a lead-resistance mechanism in bacteria. *Mol Microbiol* **2009**, *74* (2), 384-94.
13. Hobman, J. L. J., D.J.; Brown, N.L., Cysteine coordination of Pb(II) is involved in the PbrR-dependent activation of the lead-resistance promoter, PpbrA, from *Cupriavidus metallidurans* CH34. *BMC Microbiology* **2012**, *12*, 109-21.
14. Cerminati, S.; Soncini, F. C.; Checa, S. K., A sensitive whole-cell biosensor for the simultaneous detection of a broad-spectrum of toxic heavy metal ions. *Chem Commun (Camb)* **2015**, *51* (27), 5917-20.
15. Borremans, B.; Hobman, J. L.; Provoost, A.; Brown, N. L.; van Der Lelie, D., Cloning and functional analysis of the pbr lead resistance determinant of *Ralstonia metallidurans* CH34. *J Bacteriol* **2001**, *183* (19), 5651-8.
16. Chen, P.; Greenberg, B.; Taghavi, S.; Romano, C.; van der Lelie, D.; He, C., An exceptionally selective lead(II)-regulatory protein from *Ralstonia metallidurans*: development of a fluorescent lead(II) probe. *Angew Chem Int Ed Engl* **2005**, *44* (18), 2715-9.
17. Chen, P. R. W., E.C.; Zhao, J.; van der Lelie, D.; Chen, L.X.; He, C., Spectroscopic insights into lead(II) coordination by the selective lead(II)-binding protein PbrR691. *J Am Chem Soc* **2007**, *129*, 12350-1.

18. Bertini, I. G., Ioannis; Katsaros, Nikolas; Luchinat, Claudio; Provenzani, Alessandro, Tuning the Affinity for Lanthanides of Calcium Binding Proteins. *Biochemistry* **2003**, *42*, 8011-21.
19. Drake, S. K. F., Joseph J., Kinetic Tuning of the EF-Hand Calcium Binding Motif: The Gateway Residue Independently Adjusts (i) Barrier Height and (ii) Equilibrium. *Biochemistry* **1996**, *35*, 1753-60.
20. Kirberger, M.; Wong, H. C.; Jiang, J.; Yang, J. J., Metal toxicity and opportunistic binding of Pb(2+) in proteins. *J Inorg Biochem* **2013**, *125*, 40-9.
21. Means, A. R., Calmodulin-mediated Signaling. In *Handbook of Cell Signaling*, 2 ed.; Academic Press: 2010; pp 979-81.
22. Drake, S. K. Z., Michael A.; Miller, Cory L.; Falke, Joseph J., Optimizing the Metal Binding Parameters of an EF-Hand-Like Calcium Chelation Loop: Coordinating Side Chains Play a More Important Tuning Role than Chelation Loop Flexibility. *Biochemistry* **1997**, *36*, 9917-26.
23. Le Clainche, L.; Plancque, G.; Amekraz, B.; Moulin, C.; Pradines-Lecomte, C.; Peltier, G.; Vita, C., Engineering new metal specificity in EF-hand peptides. *J Biol Inorg Chem* **2003**, *8* (3), 334-40.
24. Zhou, Y.; Frey, T. K.; Yang, J. J., Viral calciomics: interplays between Ca²⁺ and virus. *Cell Calcium* **2009**, *46* (1), 1-17.
25. Piazza, M.; Taiakina, V.; Dieckmann, T.; Guillemette, J. G., Structural Consequences of Calmodulin EF Hand Mutations. *Biochemistry* **2017**, *56* (7), 944-56.
26. Ouyang, H. V., Hans J., Metal ion binding to calmodulin: NMR and fluorescence studies. *BioMetals* **1998**, *11*, 213-22.

27. Bhowmick, A.; Sharma, S. C.; Head-Gordon, T., The Importance of the Scaffold for de Novo Enzymes: A Case Study with Kemp Eliminase. *J Am Chem Soc* **2017**, *139* (16), 5793-800.
28. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc Natl Acad Sci U S A* **2011**, *108* (17), 6823-7.
29. Mack, K. L.; Moroz, O. V.; Moroz, Y. S.; Olsen, A. B.; McLaughlin, J. M.; Korendovych, I. V., Reprogramming EF-hands for design of catalytically amplified lanthanide sensors. *J Biol Inorg Chem* **2013**, *18* (4), 411-8.
30. Moroz, O. V.; Moroz, Y. S.; Wu, Y.; Olsen, A. B.; Cheng, H.; Mack, K. L.; McLaughlin, J. M.; Raymond, E. A.; Zhezherya, K.; Roder, H.; Korendovych, I. V., A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew Chem Int Ed Engl* **2013**, *52* (24), 6246-9.
31. Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; van Nuland, N. A.; Korendovych, I. V., New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J Am Chem Soc* **2015**, *137* (47), 14905-11.
32. Raymond, E. A.; Mack, K. L.; Yoon, J. H.; Moroz, O. V.; Moroz, Y. S.; Korendovych, I. V., Design of an allosterically regulated retroaldolase. *Protein Sci* **2015**, *24* (4), 561-70.
33. Miao, Y. M., Richard; Asano, Yasuhisa, Kemp Elimination Catalyzed by Naturally Occurring Aldoxime Dehydratases. *ChemBiochem* **2017**, *18*, 451-4.

34. Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013**, *503* (7476), 418-21.
35. Khersonsky, O.; Kiss, G.; Rothlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci U S A* **2012**, *109* (26), 10358-63.
36. Merski, M.; Shoichet, B. K., Engineering a model protein cavity to catalyze the Kemp elimination. *Proc Natl Acad Sci U S A* **2012**, *109* (40), 16179-83.
37. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* **2012**, *109* (10), 3790-5.
38. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453* (7192), 190-5.
39. Nelson, D. L. C., M.M., *Lehninger Principles of Biochemistry*. 6 ed.; W.H. Freeman and Company: 2012.
40. Berg, J. M. T., John L; Stryer, Lubert, *Biochemistry*. 7 ed.; W. H. Freeman: 2010.
41. Hennig, H. F.-K. O., Metal-binding proteins as metal pollution indicators. *Environ Health Per* **1986**, *65*, 175-87.

42. Martinez-Gomez, N. C.; Vu, H. N.; Skovran, E., Lanthanide Chemistry: From Coordination in Chemical Complexes Shaping Our Technology to Coordination in Enzymes Shaping Bacterial Metabolism. *Inorg Chem* **2016**, *55* (20), 10083-9.
43. Bünzli, J. C. G., Lanthanide luminescence for biomedical analyses and imaging. *Chem Rev* **2010**, *110*, 2729-55.
44. Lemon, C. M.; Brothers, P. J.; Boitrel, B., Porphyrin complexes of the period 6 main group and late transition metals. *Dalton Trans* **2011**, *40* (25), 6591-609.
45. Teo, R. D.; Termini, J.; Gray, H. B., Lanthanides: Applications in Cancer Diagnosis and Therapy. *J Med Chem* **2016**, *59* (13), 6012-24.
46. Rim, K.-T., Effects of rare earth elements on the environment and human health: A literature review. *Toxicology and Environmental Health Sciences* **2016**, *8* (3), 189-200.
47. Drake, S. K. Z., Michael A.; Kundrot, Craig; Falke, Joseph J., Molecular Tuning of an EF-Hand-like Calcium Binding Loop. *J Gen Physiol* **1997**, *110*, 173-84.
48. Bronner, F., Extracellular and intracellular regulation of calcium homeostasis. *ScientificWorldJournal* **2001**, *1*, 919-25.
49. Leong, M. L.; Newell, E. W., Multiplexed Peptide-MHC Tetramer Staining with Mass Cytometry. *Methods Mol Biol* **2015**, *1346*, 115-31.
50. Fairhall, L. T., The solubility of various lead compounds in blood serum. *J Biol Chem* **1924**, *60*, 481-4.

51. Kursula, P.; Majava, V., A structural insight into lead neurotoxicity and calmodulin activation by heavy metals. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2007**, *63* (Pt 8), 653-6.
52. Miessler, G. L. T., D.A, *Inorganic Chemistry*. 3 ed.; Pearson: New Jersey, 2004.
53. Freyer, M. W.; Lewis, E. A., Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. In *Biophysical Tools for Biologists, Volume One: In Vitro Techniques*, 2008; pp 79-113.
54. Ghai, R.; Falconer, R. J.; Collins, B. M., Applications of isothermal titration calorimetry in pure and applied research--survey of the literature from 2010. *J Mol Recognit* **2012**, *25* (1), 32-52.
55. Song, C.; Zhang, S.; Huang, H., Choosing a suitable method for the identification of replication origins in microbial genomes. *Front Microbiol* **2015**, *6*, 1049.
56. Freire, E.; Schön, A.; Velazquez-Campoy, A., Isothermal Titration Calorimetry: General Formalism using Binding Polynomials. In *Biothermodynamics, Part A*, 2009; pp 127-55.
57. Quinn, C. F.; Carpenter, M. C.; Croteau, M. L.; Wilcox, D. E., Isothermal Titration Calorimetry Measurements of Metal Ions Binding to Proteins. *Methods Enzymol* **2016**, *567*, 3-21.
58. Hargrove, A. E.; Zhong, Z.; Sessler, J. L.; Anslyn, E. V., Algorithms for the determination of binding constants and enantiomeric excess in complex host : guest equilibria using optical measurements. *New J Chem* **2010**, *34* (2), 348-54.
59. Thordarson, P., Determining association constants from titration experiments in supramolecular chemistry. *Chem Soc Rev* **2010**, *40*, 1305-23.

60. Bulheller, B. M.; Rodger, A.; Hirst, J. D., Circular and linear dichroism of proteins. *Phys Chem Chem Phys* **2007**, *9* (17), 2020-35.
61. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006**, *1* (6), 2876-90.
62. Casey, M. L. K., D. S.; Paul, Kenneth G.; Cox, Daniel D., The Physical Organic Chemistry of Benzisoxazoles. I. The Mechanism of the Base-Catalyzed Decomposition of Benzisoxazoles. *J Org Chem* **1973**, *38* (13), 2294-301.

The following chapter was co-authored with Z. Lengyel and A. Kulesha of the Korendovych lab.

Z. Lengyel and A. Kulesha designed the 2E10-HolLEE construct.

Chapter 3 Applications of HolLEE: Designing a Selective HCC-targeting Treatment Using Monoclonal Antibodies

3.0. Abstract

Hepatocellular carcinoma (HCC) is currently the sixth most diagnosed cancer in the world and has a very low survival rate. HCC is difficult to diagnose early and is even more difficult to treat when it finally is caught. One of the most effective treatments is using ^{90}Y radioisotopes to destroy the cancer cells. Currently, the only method of delivery is to attach the ^{90}Y to microspheres and deliver them directly to the blood supply of the tumor using a transarterial catheter. However, this process is dangerous, invasive, and can only be used if the tumor is already large and inoperable. Using our previously mentioned HolLEE protein, which is selective for lanthanides including yttrium, and attaching it to an antibody fragment that is specific for HCC tumor cells, we propose a new method of ^{90}Y delivery. This would allow the ^{90}Y to be delivered to the cancer cells much less invasively and also potentially reach metastasized cells. In this chapter we report the success of the antibody-HolLEE fusion (2E10-HolLEE) in achieving selectivity with ^{90}Y and highlight some of the issues that have currently come up regarding purification.

3.1. Introduction

3.1.1. Hepatocellular Carcinoma

Hepatocellular Carcinoma (HCC) is currently the sixth most diagnosed cancer in the world and the third most responsible for cancer-related deaths (Figure 3.1).¹ It is most commonly caused by cirrhosis of the liver and/or infection by the hepatitis virus.²⁻⁵ HCC is most prevalent in eastern Asia and Sub-Saharan Africa where Hepatitis B infection rates are the highest.¹ However, HCC diagnosis rates in the United States are increasing in correlation with the rising rates of Hepatitis C. Aside from the hepatitis virus, HCC risk factors include obesity, diabetes, alcohol related liver disease, and fatty liver disease.^{1,3}



Figure 3.1. Documented incidences of hepatocellular carcinoma for all ages and genders in 2013 worldwide.⁶

It is estimated that 94% of patients diagnosed with HCC will die within the next five years.^{1-2, 6} One of the reasons this cancer is so deadly is due to its low diagnosis rates in the early stages of the disease. A patient with earlier diagnosis is ten times more likely to receive and respond to potentially curative options of treatment.⁵ Once the disease progresses to later stages, the survival rate is reduced significantly. The best curative treatment for HCC is either resection of the liver, where the infected portion of the liver is removed, or liver transplantation. However, once the tumor(s) becomes too big or spreads, these options become impossible and less reliable means must be employed (Figure 3.2).

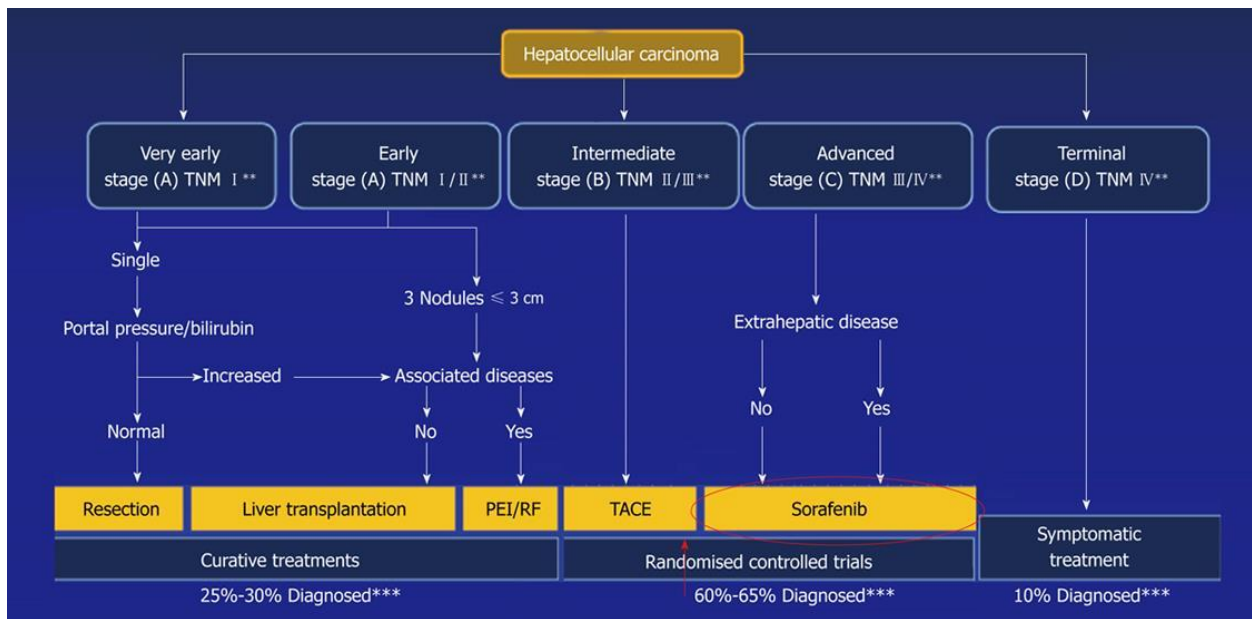


Figure 3.2. Algorithm for determining treatment options for HCC.⁷

Low diagnosis rates are due to HCC's lack of symptoms.^{1, 8} HCC is usually diagnosed from acute or chronic pain in the abdomen, which only occurs in the later stages of the disease. Furthermore, diagnosing HCC is very difficult (Figure 3.3). Liver biopsy can be intrusive and

inconclusive, and if the correct section of the liver is not biopsied, then HCC can remain undetected.⁸ α -Fetoproteins from tumor cells can act as an indicator, but screening is unreliable and non-specific to HCC.¹ CT/MRI findings can be hard to distinguish and diagnose, and MRI is not an option for people with metallic implants. PET imaging faces a similar challenge by being unable to differentiate HCC cells from regular hepatic cells in the early stages of the cancer.⁹ In addition, these methods all have the disadvantage of not being able to diagnose metathesized cancer cells.

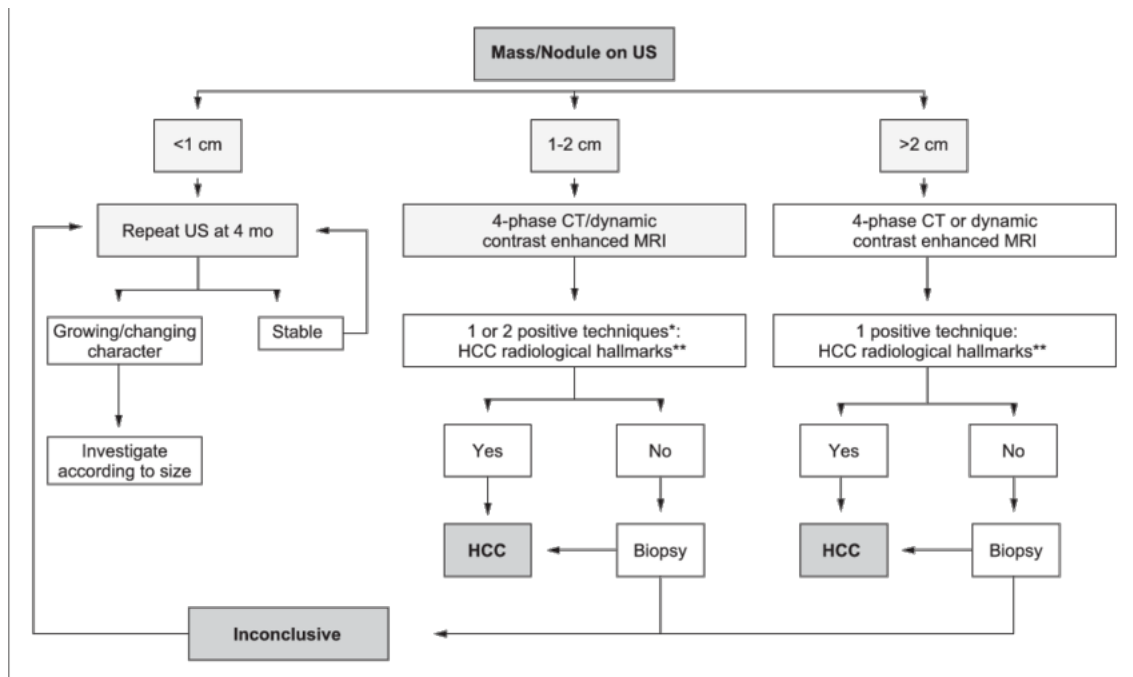


Figure 3.3. Algorithm for diagnosis of HCC.¹⁰

Once HCC is diagnosed, there are few available treatment options (Figure 3.2). If the cancer is caught early enough, resection or transplantation of the liver may be enough to cure the patient completely. However, once the disease has progressed, this no longer becomes an

option.¹ Thermal or chemical ablation where chemicals or energy such as radio frequencies are applied to the surrounding tumor tissue to induce necrosis can be utilized if the tumor is small enough. Otherwise, transarterial chemoembolization (TACE), where a catheter is fed through the aorta and hepatic artery to deliver medication directly to the blood supply of the tumor, can be used if other options are not available. This technique can be very effective in reducing the size of tumors but cannot be used if the tumor is not of sufficient size or has spread to another part of the body. One of the most common drugs used with this technique is ⁹⁰Y microspheres that release radiation to kill the tumor cell directly (Figure 3.4).^{1, 11}

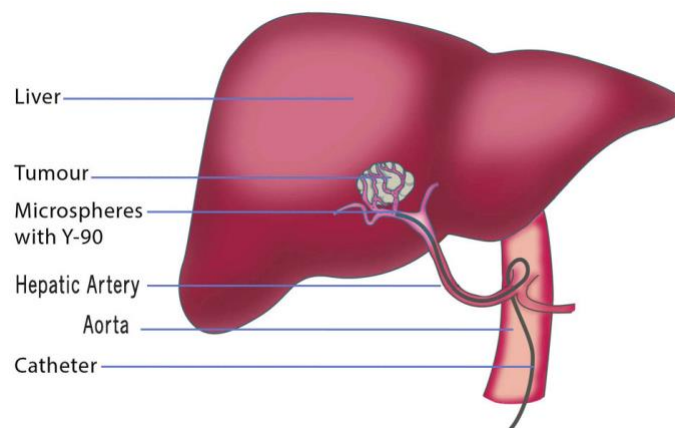


Figure 3.4. Transarterial chemoembolization of HCC using ⁹⁰Y.¹²

The key to reducing the mortality rate of HCC is two-fold: increase the ability to diagnose the disease and increase the ability to treat the disease, especially in later stages. Antibodies can be utilized as a targeted therapy for HCC tumors, and when coupled with HOLLER and ⁹⁰Y, could prove to be a successful new method of diagnosis and/or treatment.

3.1.2. Antibodies

The immune system is capable of responding to a wide range of threats in the body, including bacteria, viruses, parasitic helminths, and even tumors.¹³⁻¹⁴ The immune system consists of a variety of natural barriers, such as mucosal membranes and stomach acid, and specific defenses created over the lifetime of the individual known as the immune response.¹³ The immune response consists of α -lymphocytes, which can further be divided into B-cells, T-cells, Helper T-cells, and cytotoxic T-cells, antigen-presenting cells, phagocytic and cytotoxic cells, and natural killer cells.

Antigens and antibodies are responsible for inducing the immune response within an individual.¹³⁻¹⁴ Antigens are compounds (including proteins, peptides, and polysaccharides) that are recognized by B-cell and T-cell receptors to elicit the immune response. Antibodies are globulin proteins that act in response to these antigens.¹⁴ Antibodies protect the individual from pathogens by preventing attachment to mucosal surfaces, neutralizing toxins, facilitating phagocytosis to quarantine the foreign invader, and binding to the antigen to activate further immune responses (Figure 3.5).¹³⁻¹⁵

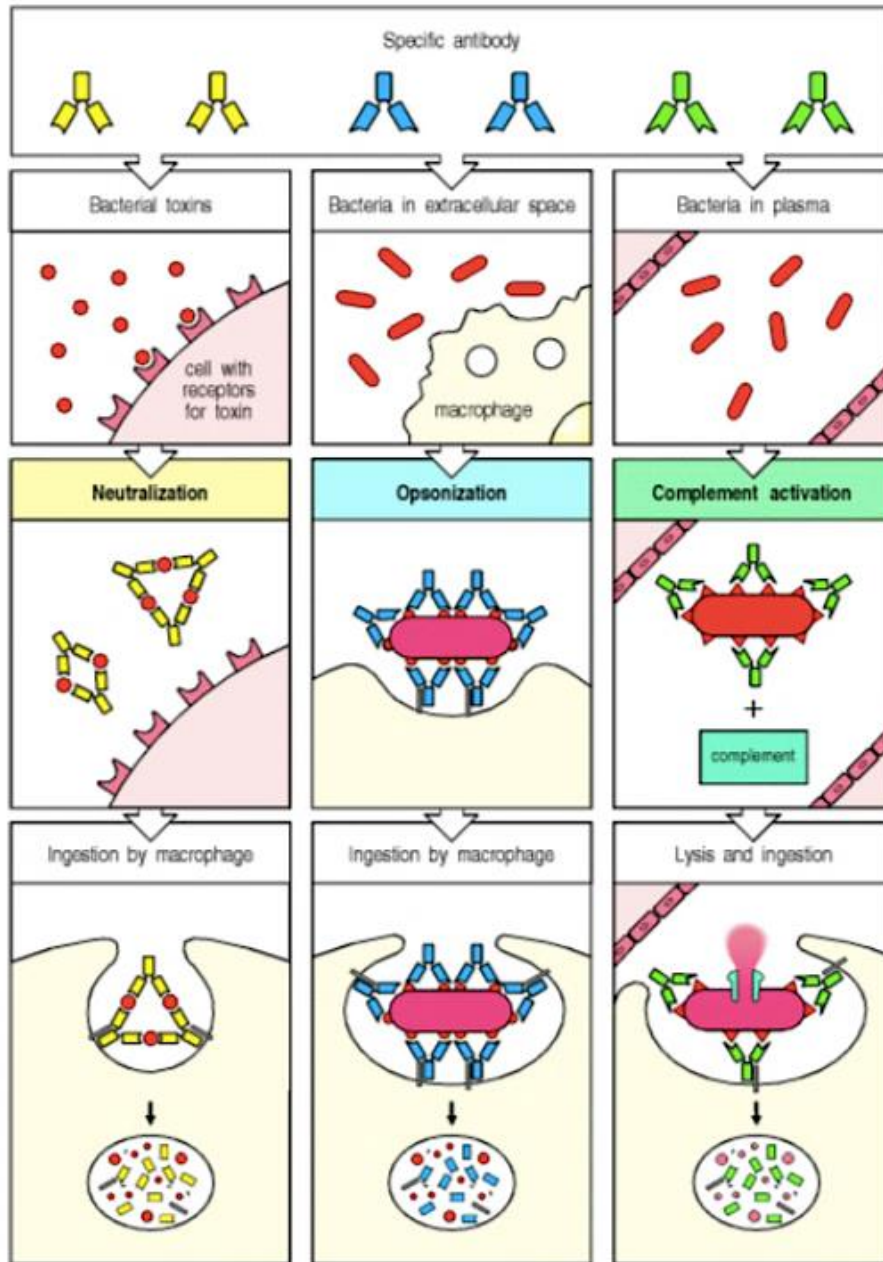


Figure 3.5. Three mechanisms of antibody defense.¹⁵ The first mechanism is when the antibodies bind to bacterial toxins so they do not damage cells and then trigger macrophages to ingest the antibody-toxins. The second method involves the antibody attacking the bacteria in the extracellular space and triggering ingestion by macrophages. The final method is when antibodies bind to bacterium in the plasma and trigger macrophages to break apart and digest the bacteria.

Antibodies consist of two identical ~55 kDa glycoproteins and two identical ~25 kDa glycoproteins that combine together to form a distinctive Y-shape (Figure 3.6).¹⁴ The chains are

known as “heavy” and “light” chains in accordance to their relative weights. This complex is held together by a collection of disulfide bonds between both heavy chains, and between the heavy and light chains. In addition, each polypeptide chain consists of a variable region on the N-terminus and a constant region on the C-terminus consisting of two amino acid sequences. The constant region of the heavy chain plays an important role in enabling the antibody to fulfill its various functions. Meanwhile, the variable regions are responsible for the specific binding of antibodies to antigens.

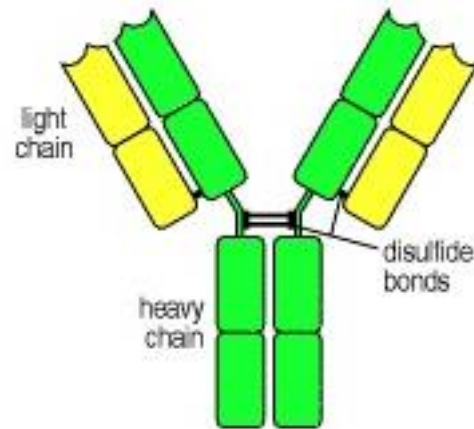


Figure 3.6. General structure of antibody. Heavy chains are featured in green and light chains are featured in yellow.¹⁵ Disulfide bonds between the chains are used to create the distinctive structure of the antibody.

Antibodies (Abs) have many diagnostic and clinical uses due to their specificity.¹⁴ They are used in immunofluorescence assays, enzyme immunoassays, radioimmunoassays, chemoluminescence assays, and Western blotting. More recently, antibodies have been explored and used for therapeutic purposes.¹⁶ There are currently over 20 FDA-approved Abs used in the treatment of various maladies such as rheumatoid arthritis, Crohn’s disease, and certain types of cancer.¹⁶⁻¹⁷ Most therapeutic antibodies are known as monoclonal antibodies (mAb) because

they made by identical immune cells from a single parent cell.^{15, 18} They are harvested from mice that have formed an immune response to the antigen and the cells that produce the antibodies are harvested and isolated.¹⁷

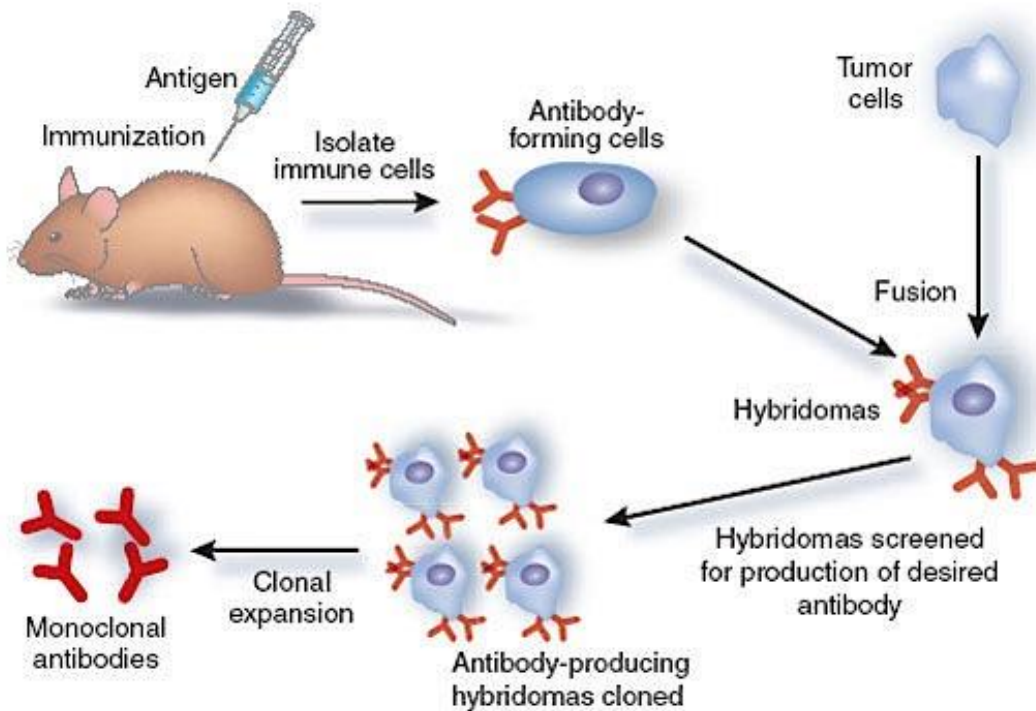


Figure 3.7. How monoclonal antibodies are made.¹⁸ Monoclonal antibodies are used because they all bind to a single epitope and are highly specific for the tumor cells. If polyclonal antibodies were used, then there would be much less specificity and even the possibility that the Ab could bind to a healthy cell instead.

Due to the high specificity of the antibody-antigen interaction and the antibodies long half-life within the host, therapeutic antibodies are an appealing alternative to chemotherapy for the treatment of tumors and cancer.^{17, 19-20} Several strategies have been employed to use antibodies to treat cancer; these are summarized and explained in Figure 3.8.¹⁷ Tumor cells produce unique antigens from regular cells, so complementary antibody-based drugs should be

extremely selective at either targeting the cancer cells primary survival functions, introducing treatments directly to the cancer cells, or redirecting the body's own immune cells to target the cancer cells.^{17, 20-21}

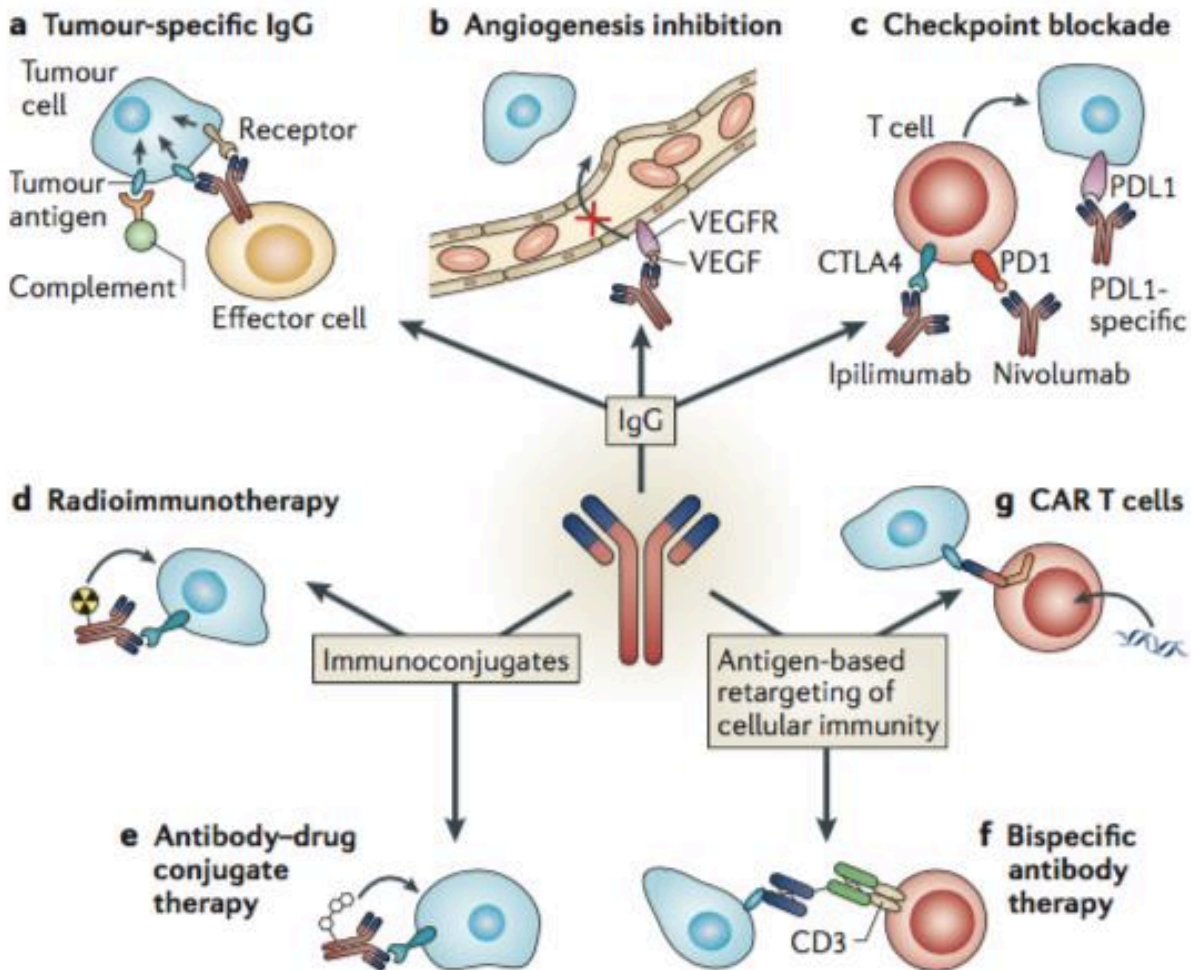


Figure 3.8. Strategies of cancer therapeutic monoclonal antibodies.¹⁷ a) The mAb binds to target and triggers antibody-dependent cytotoxicity, complement-mediated cytotoxicity, or directly signals cancer cell death. b) The mAb inhibits the development of new blood vessels that feed the tumor cells. c) The mAb elicits a response from antitumor T cells by blocking signals that would normally protect the tumor from the body's own immune system. d) The mAb is used to deliver radioisotopes directly to cancer cells for selective radiation treatment. e) The mAb is used to deliver drugs directly to the cancer cells. f) A small portion of the mAb is used to both recognize cancer cells on one end and activate antigens to trigger immune effector cells on the other end. g) The mAb variable region is attached to a signaling peptide that is transferred to T cells and makes them chimeric antigen receptor (CAR) T cells that are specific for the tumor the peptide is specific for.

Further innovations into antibody engineering have provided exciting new platform for drug design. The large size of the antibody makes it difficult to diffuse within a tumor, hindering the therapeutic antibody's ability to treat it.²² However, by removing fragments and/or chains from the original antibody structure, new "bodies" can be made that exhibit similar activity but present much higher diffusion rates within tumors and be easier to manipulate (Figure 3.9).²³⁻²⁵ Since the variable region is solely responsible for antigen selectivity, these fragmented bodies can still retain similar sensitivity while being much easier to deliver to tumors.²⁶⁻²⁷

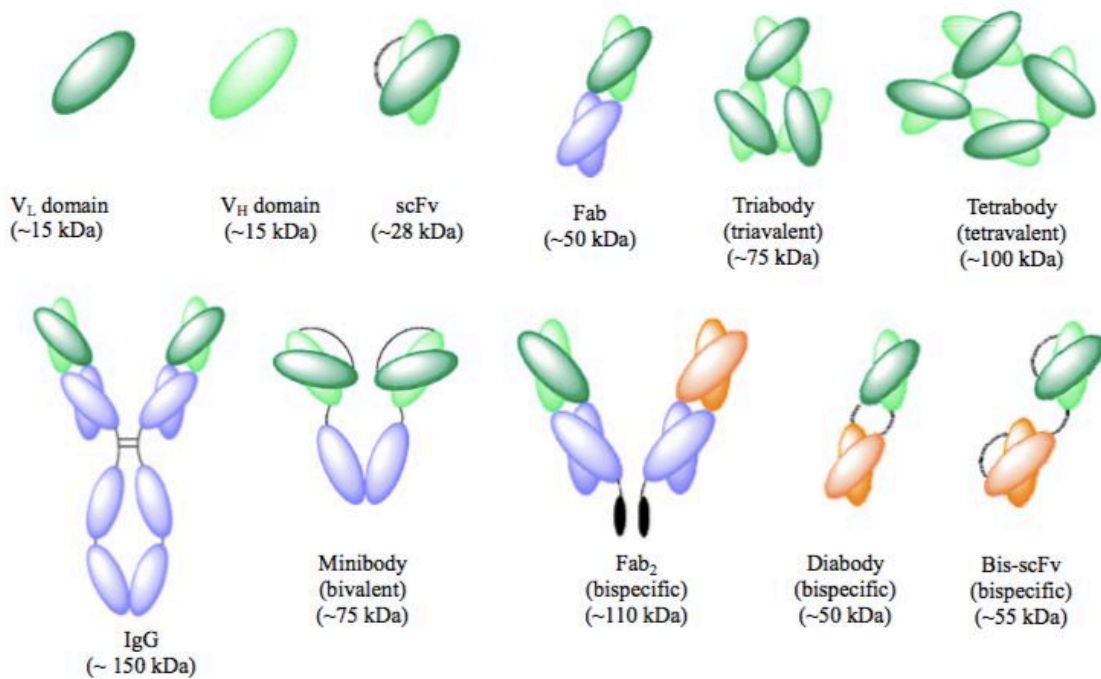


Figure 3.9. Structure of various antibody derivatives.²⁷ As long as the structure contains the variable region that is responsible for binding to antigens, the fragment will successfully bind. Smaller bodies have the advantage of being more soluble and cell permeable.

3.1.3. Radioimmunotherapy

Radioactive ions are becoming an increasingly frequent method of cancer treatment.²⁸⁻³⁰ They have been found to be effective diagnostic tools and/or therapeutic treatment for a variety of cancers such as thyroid, non-Hodgkin's lymphoma, prostate, and breast.

There are several types of radioactive decay.³¹⁻³² Alpha radioactive elements, upon excitation, releases an α -particle consisting of two protons and two neutrons. β^- -emitters release a high energy electron, while β^+ -emitters, also known as positron emission, release a high energy positively charged particle known as a positron. Gamma emission releases gamma rays, which have very high energy and frequency and great penetrative power. Finally, while not as common or widely used, the Auger effect occurs when an electron from an inner-valence shell is emitted from the cell.³²⁻³³ Most clinical radiotherapy uses α -, β -, γ -emitters, although there has been considerations about using Auger elements, which are normally too cytotoxic, as a way to treat residual cancer cells to prevent resurgence. The isotope yttrium-90 is a common radionuclide in cancer treatment.^{11, 34-35} When it decays, ^{90}Y emits a β^- -particle and the daughter ion ^{90}Zr . Yttrium-90 is especially known for its use in effectively treating HCC via radioembolization (Figure 3.10).³⁶⁻³⁸

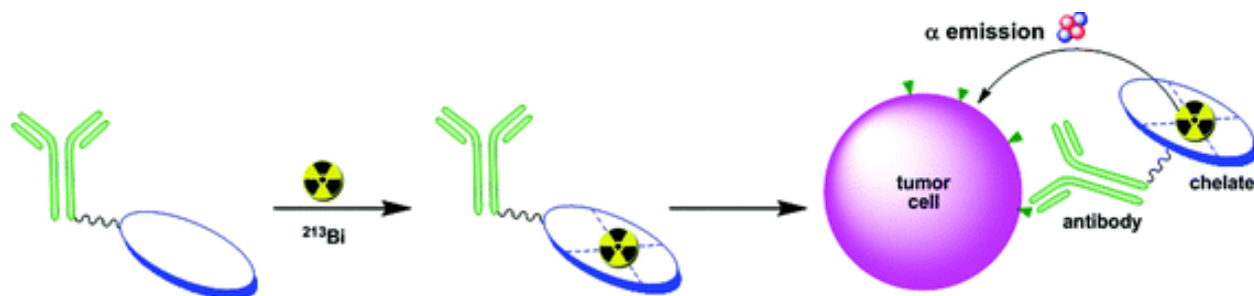


Figure 3.10. Diagram of radioimmunotherapy.³⁶

One challenge in synthesizing radioimmunotherapeutics is attaching radionuclides to the antibody with linkers.^{32, 37} Streptavidin, polymers, and short chemical linkers are all popular solutions to this problem with varying rates of success and difficulty.^{29, 38-40} Given the success of the HolIEE protein in binding lanthanides selectively, we propose that the protein could be used as a way to bind ^{90}Y to a selective antibody fragment, 2E10, to deliver the radionuclide for PET imaging and/or therapeutic purposes.⁴¹ 2E10 is a glypican-3 specific variable fragment. Glypican-3 is expressed in the cell membranes of HCC tumors and inhibiting glypican-3 in these cells can reduce proliferation of this cell line. Using a scFv of a glypican-3 specific antibody also allows the advantage of being easier to mutate since it is smaller and being less likely to trigger rejection by the immune system before it reaches its target.

Using a protein chelator also allows the advantage of production via bacteria or yeast and allows us to evolve the protein for selectivity or stability using directed evolution rather than making each individual chelator synthetically. If this is successful, other antibody fragments could also be combined with HolIEE to provide treatment for other types of cancer since many lanthanides have anticancer properties (Figure 3.11).

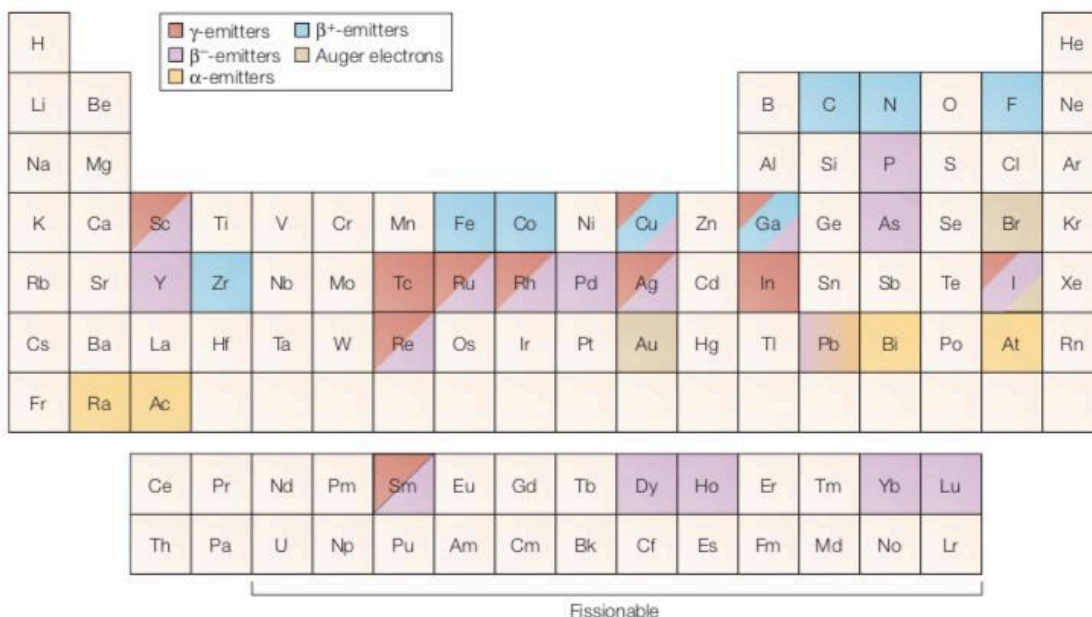


Figure 3.11 Periodic Table highlighting radionuclides of interest for cancer treatment.³²

3.2. Isolation of 2E10-HolIEE Construct

The 2E10-HolIEE fusion protein was prepared, but unfortunately, it could only be isolated from inclusion bodies. Inclusion bodies occur when a protein is overexpressed and insoluble, leading to aggregates of the protein.⁴² These can be difficult to manage and purify. In order to isolate protein from fusion bodies, the protein must be first denatured in high concentrations of guanidine chloride and then refolded with urea. Nonetheless, small amounts of 2E10-HolIEE was isolated (~0.3 μ g) and metal binding was tested using Kemp elimination. Much like the results described in Chapter 2, Kemp elimination activity only occurred in the presence of yttrium. Indeed, when activity was compared between calcium and yttrium, there was a clear distinction.

2E10-HolIEE had no activity in the presence of calcium (the negative slope is due to noise) (Figure 3.12).

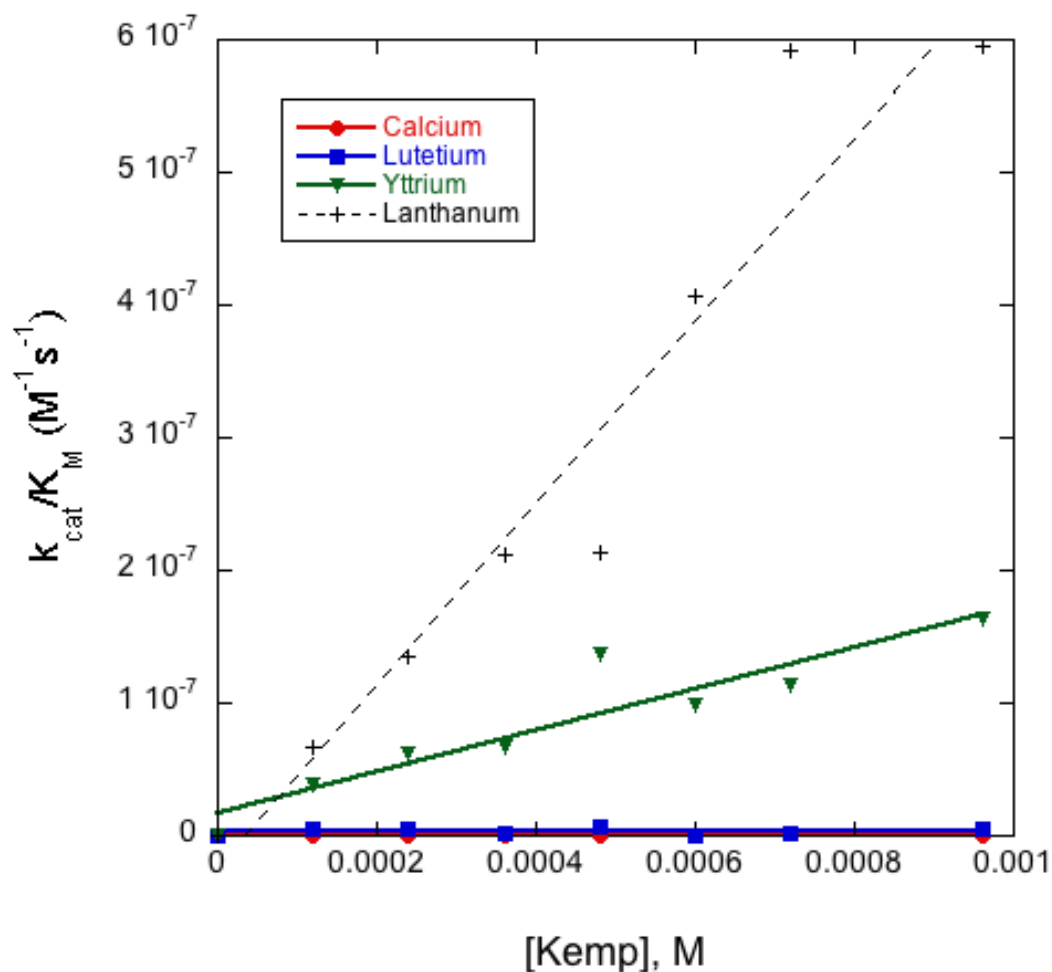


Figure 3.12. Catalytic efficiency of Kemp elimination using HolIEE (0.5 μ M) in the presence of yttrium (0.08 mM) and calcium (0.08 mM) in 20 mM MOPS, 100 mM NaCl, pH 7.0. The k_{cat}/K_M of 2E10-HolIEE was $8.0 M^{-1}s^{-1}$ in the presence of yttrium. (MT7021)

However, it is important to note that comparing 2E10-HolIEE and HolIEE, there is a significant decrease in activity (Figure 3.13). While both proteins have absolutely no activity in

the presence of calcium, 2E10-HolIEE has a 10-fold decrease in activity compared to HolIEE in the presence of yttrium. This is possibly due to the antibody portion either blocking the active site or interfering with the metal binding portion of the protein. Even with decreased activity, the fact that 2E10-HolIEE is capable of binding yttrium is promising, although further improvement will be needed.

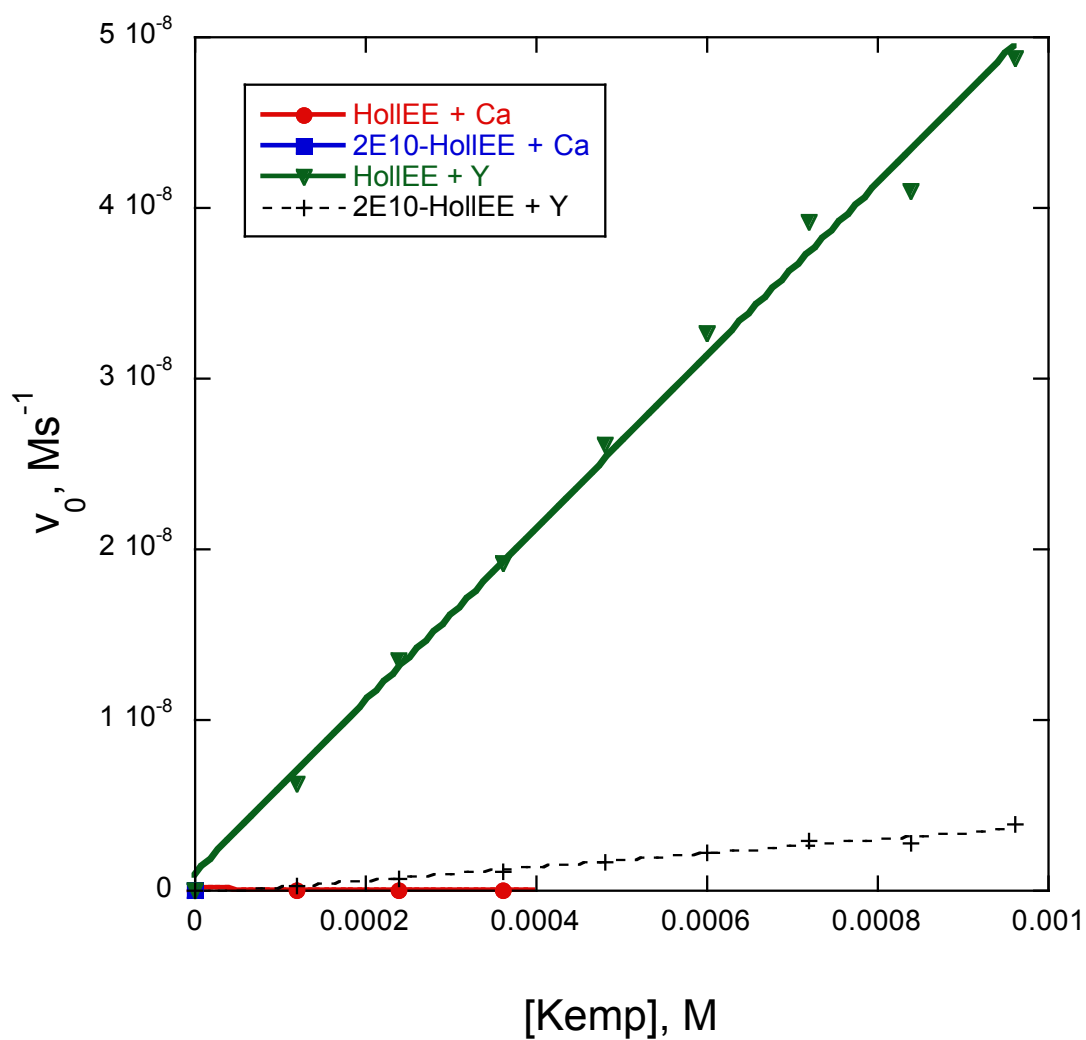


Figure 3.13. Catalytic efficiency of Kemp elimination using 2E10-HolIEE (0.5 μM) in the presence of yttrium (0.08 mM) and calcium (0.08 mM) in 20 mM MOPS, 100 mM NaCl, pH 7.0. The k_{cat}/K_M of 2E10-HolIEE was $8.0 \text{ M}^{-1}\text{s}^{-1}$ in the presence of yttrium and the k_{cat}/K_M of HolIEE was $101 \text{ M}^{-1}\text{s}^{-1}$ in the presence of yttrium. (MT7021)

The next question that we asked was if 2E10 still bound to HCC cells selectively. However, when attempts were made to purify the protein for testing, issues came up after cleavage of the His₆-tag with TEV protease (Figure 3.14). It appears that despite the low concentration of protein used (~2 mg), precipitation occurred after TEV cleavage since the protein band was now absent.

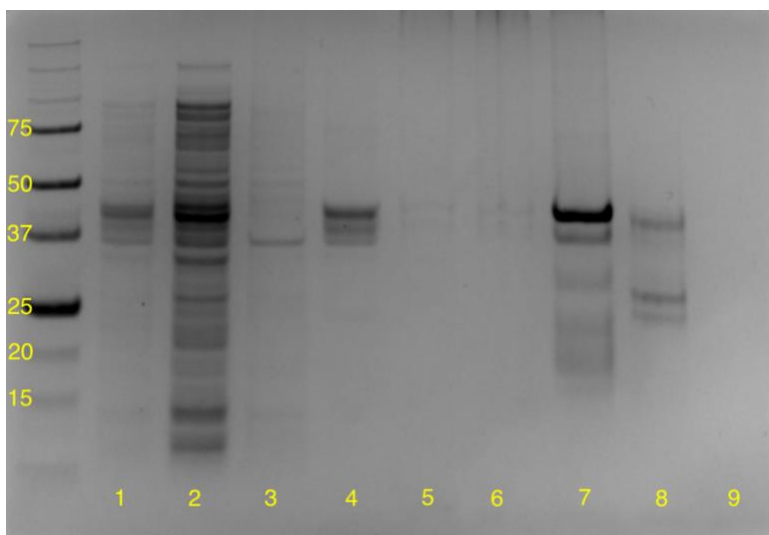


Figure 3.14. 10% acrylamide SDS-PAGE gel of 2E10-HolIEE purification: Lane 1) cells before induction, 2) cell pellet before purification, 3) supernatant after sonication, 4) supernatant before loading onto Ni-NTA column, 5) Ni-NTA column flow through, 6) Ni-NTA column wash, 7) uncut 2E10-HolIEE before TEV cleavage, 8) uncut 2E10-HolIEE after incubation overnight with TEV, 9) final sample after removing TEV. The calculated mass of 2E10-HolIEE with His₆-tag is 39,725 Da and without His₆-tag is 37,260 Da.

3.3. Conclusions

Radioimmunotherapeutics provide a powerful method of delivering radioisotopes to cancer cells selectively. By attaching our lanthanide-selective protein HolIEE to a HCC-specific antibody fragment, 2E10, we hope to create a new treatment for HCC that is selective, effective, and evolvable. We have shown that 2E10-HolIEE is selective for yttrium. Currently, we are

working towards isolating enough protein to test 2E10-HolLEE's selectivity towards HCC cells. The initial results are promising and further work is needed to develop and test the hypothesis. Work is also being done to determine the binding affinity of 2E10-HolLEE to yttrium as well as using other expression vectors such as yeast or insect cells to express protein.

3.4. Experimental

3.4.1. Expression and Purification of 2E10-HolLEE

Plasmid was transformed into Rosetta pLysS cells and grown overnight at 37°C in LB media. A 100 mL seed culture was grown at 37°C for approximately 8 hours before being used to inoculate 1 L of Zym-5052 autoinduction media. The culture was grown for 16-18 hours at 30°C and the cell pellet was collected (approximately 2 g).

Cell pellet was resuspended in lysis buffer containing 25 mM Tris-HCl (used since chloride ions were determined to be a non-issue for HolLEE), 20 mM imidazole, 0.5 mM PMSF, and 500 mM NaCl (pH 8). Cells were lysed by sonication, and then the crude cell lysate was centrifuged at 20,000xg for 30 minutes. The supernatant was decanted and the cell pellet was resuspended in Wash 2UT buffer (25 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl, 2 M urea, pH 8.0). Urea resuspension was then sonicated until solids were broken up and then centrifuged at 20,000xg for 30 minutes. The urea wash was repeated 2-3 times. The inclusion bodies were then washed with lysis buffer and centrifuged at 20,000xg for 30 minutes. The supernatant was decanted and

the inclusion body pellet was resuspended in 6G buffer (25 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl, 6 M guanidinium chloride) and allowed to rock gently overnight at room temperature.

The supernatant was centrifuged at 20,000xg for 30 minutes the following morning and applied to a Ni-NTA column. The Ni-NTA was allowed to stand for at least one hour to ensure binding occurred. The column was then washed with 6G buffer. To refold the protein, the column was washed with urea refolding buffer (25 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl, 6 M urea) and was then washed with decreasing concentrations of urea (6-0 M) in lysis buffer. The protein was then eluted with buffer containing 25 mM Tris-HCl, 250 mM imidazole, and 500 mM NaCl (pH 8). The sample was exchanged into buffer containing 20 mM Mops and 100 mM NaCl (pH 7) by applying it onto a Bio-Rad 10 DG desalting column.

3.4.2. Removal of His₆-tag

Sample was incubated with 1.0 mM DTT, 0.5 mM EDTA, and TEV protease at a 10:1 absorbance (280 nm) ratio overnight at 34°C. The cleavage solution was purified using a 0.22 µM PES sterile filter before incubation.

After incubation, DTT and EDTA were removed by applying sample onto a Bio-Rad 10 DG desalting column with 50 mM Tris, 75 mM NaCl (pH 8) buffer. Desalted samples were then applied to Ni-NTA column equilibrated with same buffer and allowed to sit on column for minimum one hour. Afterwards, protein was collected in flow through and wash and the presence of protein was determined using UV-Vis spectroscopy.

Sample was briefly incubated with 10-fold molar excess of EDTA and then exchanged into 100 mM bicarbonate buffer (pH 8.5) twice using a Bio-Rad 10 DG desalting column to ensure

removal of excess calcium ions and EDTA. Bicarbonate buffer was used because a basic environment is necessary to properly use the fluorescent labels that would be needed for the next step. The purity of the protein was established by SDS-PAGE using a 10% acrylamide gel with Unstained Precision Protein Plus ladder (BioRad). Protein concentration was determined using UV-Vis spectroscopy where the ϵ_{280} of cut protein is $62,800 \text{ cm}^{-1}\text{M}^{-1}$.

3.4.3. Kinetic assays

Kinetic measurements were done with either a Thermo Lab-systems Multiskan Spectrum Platerreader or a BioTek Eon Micro Platerreader monitoring the absorbance at 380 nm at 22°C using at least three independent measurements. The final protein concentration (determined by UV-Vis spectroscopy using $\epsilon_{280} = 64,290 \text{ M}^{-1} \text{ cm}^{-1}$ for uncut protein) for kinetic measurements was $0.5 \mu\text{M}$. Final substrate concentrations ranged from 0.12 to 0.96 mM. Kinetic parameters (k_{cat}/K_M) were obtained by fitting the data to a linear fit using Kaleidagraph 4.5 (Synergy Software).

3.6. References

1. Waghray, A.; Murali, A. R.; Menon, K. N., Hepatocellular carcinoma: From diagnosis to treatment. *World J Hepatol* **2015**, *7* (8), 1020-9.
2. Tang, Z. Y., Hepatocellular Carcinoma- Cause, Treatment, and Metastasis. *World J Gastro* **2001**, *7* (4), 445-54.
3. Blonski, W., Non-viral causes of hepatocellular carcinoma. *World Journal of Gastroenterology* **2010**, *16* (29).

4. Naik, P. K.; Ranjan, P.; Kesari, P.; Jain, S., MetalloPred: A tool for hierarchical prediction of metal ion binding proteins using cluster of neural networks and sequence derived features. *Journal of Biophysical Chemistry* **2011**, *02* (02), 112-23.
5. Sanyal, A. J.; Yoon, S. K.; Lencioni, R., The etiology of hepatocellular carcinoma and consequences for treatment. *Oncologist* **2010**, *15 Suppl 4*, 14-22.
6. Gomes, M. A., Priolli, D. G., Tralhão, J. G., Botelho, M. F., Hepatocellular carcinoma: epidemiology, biology, diagnosis, and therapies. *Rev Assoc Med Bras* **2013**, *59* (5), 514-24.
7. Rossi, L.; Zoratto, F.; Papa, A.; Iodice, F.; Minozzi, M.; Frati, L.; Tomao, S., Current approach in the treatment of hepatocellular carcinoma. *World J Gastrointest Oncol* **2010**, *2* (9), 348-59.
8. Bialecki, E. S.; Di Bisceglie, A. M., Diagnosis of hepatocellular carcinoma. *HPB (Oxford)* **2005**, *7* (1), 26-34.
9. Kolthammer, J. A.; Corn, D. J.; Tenley, N.; Wu, C.; Tian, H.; Wang, Y.; Lee, Z., PET imaging of hepatocellular carcinoma with ¹⁸F-fluoroethylcholine and ¹¹C-choline. *Eur J Nucl Med Mol Imaging* **2011**, *38* (7), 1248-56.
10. Carrilho, F. J.; Mattos, A. A.; Vianey, A. F.; Vezozzo, D. C.; Marinho, F.; Souto, F. J.; Cotrim, H. P.; Coelho, H. S.; Silva, I.; Garcia, J. H.; Kikuchi, L.; Lofego, P.; Andraus, W.; Strauss, E.; Silva, G.; Altikes, I.; Medeiros, J. E.; Bittencourt, P. L.; Parise, E. R., Brazilian society of hepatology recommendations for the diagnosis and treatment of hepatocellular carcinoma. *Arq Gastroenterol* **2015**, *52 Suppl 1*, 2-14.

11. Kuei, A. S., Sammy; Cho, Sung-Ki; Kee, Stephen T; Lee, Edward Wolfgang, Effects of Yttrium-90 selective internal radiation therapy on non-conventional liver tumors. *World J Gastro* **2015**, 21 (27), 8271-83.
12. Y90-Liver Cancer Treatment Procedure.
https://www.radconlittlerock.com/services/interventional/procedures/cancer_liver_treatment_y90/.
13. Virella, G., Introduction to Medical Immunology. 4 ed.; Virella, G., Ed. Marcell Dekker: Charleston, SC, 1998; p. 711.
14. Parija, S. C., Textbook of Microbiology & Immunology. 2 ed.; 2012.
15. Janeway, C. A. T., Paul; Walport, Mark; Shlomchik; Mark J, Immunobiology: The Immune System in Health and Disease. 5 ed.; Garland Science: New York, 2001.
16. Leavy, O., Therapeutic antibodies: past, present and future. *Nat Rev Immunol* **2010**, 10 (5), 297.
17. Weiner, G. J., Building better monoclonal antibody-based therapeutics. *Nat Rev Cancer* **2015**, 15 (6), 361-70.
18. Michnick, S. W. S., S.S., Submitting antibodies to binding arbitration. *Nat Chem Biol* **2008**, 4 (6), 326-9.
19. Hansel, T. T.; Kropshofer, H.; Singer, T.; Mitchell, J. A.; George, A. J., The safety and side effects of monoclonal antibodies. *Nat Rev Drug Discov* **2010**, 9 (4), 325-38.
20. Adams, G. P.; Weiner, L. M., Monoclonal antibody therapy of cancer. *Nat Biotechnol* **2005**, 23 (9), 1147-57.

21. Restifo, N. P.; Dudley, M. E.; Rosenberg, S. A., Adoptive immunotherapy for cancer: harnessing the T cell response. *Nat Rev Immunol* **2012**, *12* (4), 269-81.
22. Chala, G., A Review on: Antibody Engineering for Development of Therapeutic Antibodies. *International Journal of Immunology* **2015**, *3* (3).
23. Holliger, P. P., T; Winter, G, "Diabodies": Small bivalent and bispecific antibody fragments. *PNAS* **1993**, *90*, 6444-8.
24. Todorovska, A. R., Rob C; Dolezal, Olan; Kortt, Alexander A; Hoogenboom, Hennie R; Hudson, Peter J, Design and applications of diabodies, triabodies and tetrabodies for cancer targeting. *J Immun Met* **2001**, *248*, 47-66.
25. Chang, C.-H. S., Robert M; Rossi, Edmund A; Karacay, Habibe; McBride, William; Hansen, Hans J; Chatal, Jean-François; Barbet, Jacques; Goldenberg, David M, Molecular Advances in Pretargeting Radioimmunotherapy with Bispecific Antibodies. *Mol Cancer Ther* **2002**, *1*, 553-63.
26. Behr, T. M. G., David M; Becker, Wolfgang S, Radioimmunotherapy of Solid Tumors: A Review "Of Mice and Men". *Hybridoma* **1997**, *16* (1), 101-7.
27. Herrington-Symes, A. P.; Farys, M.; Khalili, H.; Brocchini, S., Antibody fragments: Prolonging circulation half-life special issue-antibody research. *Advances in Bioscience and Biotechnology* **2013**, *04* (05), 689-98.
28. Thomas, S. R., Options for Radionuclide Therapy: From Fixed Activity to Patient-Specific Treatment Planning. *Cancer Biotherapy & Radiopharmaceuticals* **2002**, *17* (1), 71-82.
29. Membreno, R.; Cook, B. E.; Fung, K.; Lewis, J. S.; Zeglis, B. M., Click-Mediated Pretargeted Radioimmunotherapy of Colorectal Carcinoma. *Mol Pharm* **2018**, *15* (4), 1729-34.

30. Carlsson, J.; Forssell Aronsson, E.; Hietala, S.-O.; Stigbrand, T.; Tennvall, J., Tumour therapy with radionuclides: assessment of progress and problems. *Radiotherapy and Oncology* **2003**, *66* (2), 107-17.
31. Tro, N. J., *Introductory Chemistry*. 5 ed.; Pearson: 2014; p 840.
32. Milenic, D. E.; Brady, E. D.; Brechbiel, M. W., Antibody-targeted radiation cancer therapy. *Nat Rev Drug Discov* **2004**, *3* (6), 488-99.
33. Yokoya, A.; Ito, T., Photon-induced Auger effect in biological systems: a review. *Int J Radiat Biol* **2017**, *93* (8), 743-56.
34. Attarwala, A. A.; Molina-Duran, F.; Busing, K. A.; Schonberg, S. O.; Bailey, D. L.; Willowson, K.; Glatting, G., Quantitative and qualitative assessment of Yttrium-90 PET/CT imaging. *PLoS One* **2014**, *9* (11), e110401.
35. Fricker, S. P., The therapeutic application of lanthanides. *Chem Soc Rev* **2006**, *35* (6), 524-33.
36. Lemon, C. M.; Brothers, P. J.; Boitrel, B., Porphyrin complexes of the period 6 main group and late transition metals. *Dalton Trans* **2011**, *40* (25), 6591-609.
37. Boros, E.; Holland, J. P., Chemical aspects of metal ion chelation in the synthesis and application antibody-based radiotracers. *J Labelled Comp Radiopharm* **2017**.
38. Chong, H.-S. M., Xiang; Le, Thien; Kwamena, Baidoo; Milenic, Diane E; Brady, Erik D; Song, Hyun A; Brechbiel, Martin W, Rational Design and Generation of a Bimodal Bifunctional Ligand for Antibody-Targeted Radiation Cancer Therapy. *J Med Chem* **2008**, *51*, 118-25.

39. Guleria, M.; Das, T.; Kumar, C.; Sharma, R.; Amirdhanayagam, J.; Sarma, H. D.; Dash, A., Effect of Number of Bifunctional Chelating Agents on the Pharmacokinetics and Immunoreactivity of ¹⁷⁷Lu-labeled Rituximab: A Systemic Study. *Anticancer Agents Med Chem* **2018**, *18* (1), 146-53.
40. Qaim, S. M., Nuclear data for production and medical application of radionuclides: Present status and future needs. *Nucl Med Biol* **2017**, *44*, 31-49.
41. Li, Y. S., D.L.; Scholler, N.; Kaplan, D.E., Validation of glypican-3-specific scFv isolated from paired display/secretory yeast display library. *BMC Biotech* **2012**, *12* (23).
42. Singh, S. M.; Panda, A. K., Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* **2005**, *99* (4), 303-10.

The work from this chapter has led to the following publication:

M. Takacs, O.V. Makhlynets, P.L. Tolbert, I.V. Korendovych. Secretion of Functional Formate Dehydrogenase in *Pichia Pastoris*. *Prot. Eng. Des. Sel.* **2017**, 30, 279-284.

Chapter 4 Functional Formate Dehydrogenase Secretion in *Pichia pastoris*

4.0. Abstract

Biofuels are an important way to reduce carbon dioxide and other greenhouse emissions. Unfortunately, current methods of biofuel production require large amounts of land and energy that can create significant amounts of its own pollution and waste. Self-regenerating systems using select enzymes have been proposed in the past to convert CO₂ to the biofuel methanol, which are less costly in terms of space and energy, but are limited by their expression in bacteria and the following purification. We have successfully expressed and secreted formate dehydrogenase (FDH), a key enzyme that converts CO₂ to formate, which can subsequently be converted to methanol, in *Pichia pastoris*. This provides an important first step in creating a self-sustaining system of organisms capable of producing biofuels with minimal resources.

4.1. Introduction

4.1.1. Common Methods of Biofuel Production

Biofuels are a diverse class of fuels that are produced from biomasses that contain photosynthesis-fixed carbons.¹ They are an important strategy to keeping CO₂ greenhouse gas emissions to a minimum, reducing atmospheric greenhouse gas concentrations, and stabilizing fuel prices. Already both the United States and European Union require fuel for cars to incorporate a percentage of renewable biofuels to reduce carbon emissions. In the last decade, many countries have focused on researching new ways to produce and/or refine biofuels.²⁻³

Current methods of producing biofuels can be divided into four generations (Figure 4.1).¹⁻
² First generation biofuels use sugar and oil feedstocks to produce alcohols and alkyl esters. One prominent example of this is the corn industry in the Midwest. Much of the corn grown by farmers is currently converted to ethanol and mixed with current fuel reserves. However, this method of biofuel production comes at a serious cost.⁴ In order to grow the amount of corn needed to make a sufficient quantity of ethanol, large amounts of land are required that compete directly with corn that would normally be used for food.^{2, 4-5} This poses problems as converting grasslands into corn fields can incur a carbon debt of up to 93 years just by clearing the land for use. Using already existing land designated for corn fields, which already has a large portion of its nutrients depleted, competes directly with food production and would still incur a carbon debt of almost 50 years.⁴ This problem is even worse in places like Brazil where healthy rainforests would have to be cut down in order to produce biofuels.⁵

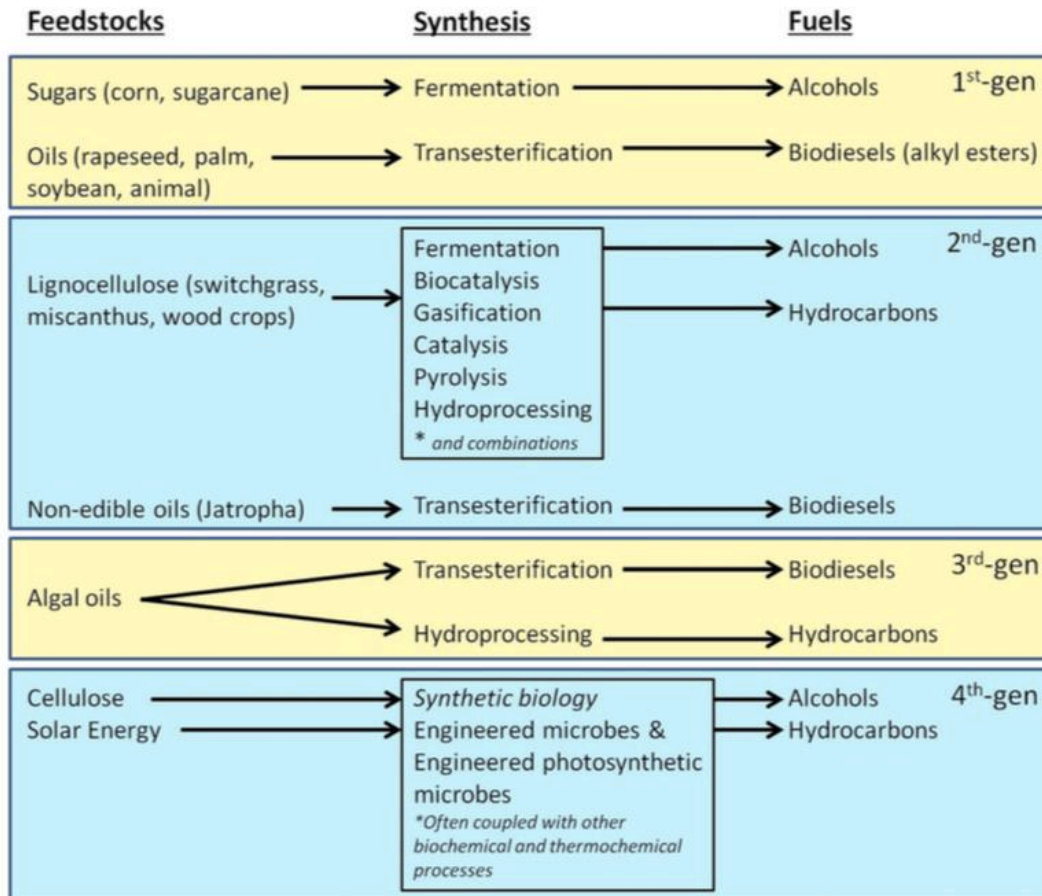


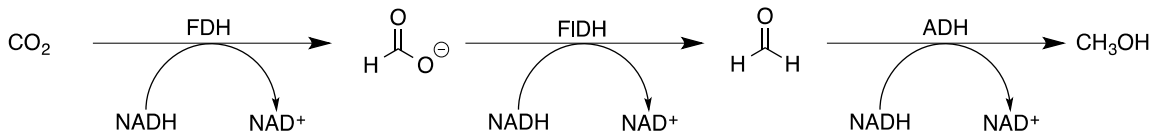
Figure 4.1. Summary of the production methods for the four generations of biofuels.¹

The second generation of biofuels is produced from lignocellulose waste such as wood and switchgrass.¹ Lignocellulose is found in all plants and consists of cellulose, hemicellulose, and lignin.⁵ It is available throughout the world and there is little cost for raw materials since waste products can be used and no land needs to be dedicated to growth.⁶ The biomass can be converted to alcohols, hydrocarbons, and alkyl esters through a variety of means including fermentation, pyrolysis, and acid or alkaline hydrolysis.^{1, 6} The downside to lignocellulose as a biofuel is that the feedstocks require pretreatment so that they are more readily digestible and can be more easily converted into sugars for enzymatic hydrolysis, minimizes inhibitors to slow

or stop enzymatic hydrolysis, the ability to recover lignin, and to reduce overall costs.⁵ Each method of processing has distinct disadvantages: pyrolysis releases carbon emissions into the environment, dissolving the lignin in ammonia can be harmful to the environment if not recycled properly, chemical treatments results in inhibitors that can lower yields, and using fungi and bacteria to degrade feedstocks can result in slow rates and low yields.⁶

The third generation of biofuels, algae, utilizes single-celled, photosynthetic organisms that produce lipids, hydrocarbons, and oils that can be converted into biofuels.⁷ Algae requires large amounts of water to live, but it can produce products even in waste water, making it economically viable.⁸ It might also be possible to use algae to reduce CO₂ in the environment by dissolving CO₂ in the water.⁹ It is efficient because algae contains RuBisCO (ribulose-biphosphate-carboxylase-oxygenase) which fixates carbon dioxide during photosynthesis. Even a 5-10% increase of RuBisCO's fixation activity would greatly reduce the current atmospheric levels of CO₂. However, after two decades, there has still been no progress in making this enzyme more active. Algae biofuel production is still relatively new, but the main challenge it faces is the massive amounts of land and water it would take to produce enough algae to sustain even the United States, as well as the processing required to make the lipids and other products into useable biofuels.^{1, 7-9}

There is a fourth, yet unrealized generation of biofuels that shows much promise: using solar energy to convert CO₂ to alcohols using enzymes.^{1, 9} It has been proposed that a sequence of three enzymes: formate dehydrogenase (FDH), formaldehyde dehydrogenase (FIDH), and alcohol dehydrogenase (ADH) can be used for this conversion (Scheme 2).⁹⁻¹⁰



Scheme 2. Overview of enzymatic conversion of carbon dioxide to methanol.

While partial hydrogenation of carbon dioxide has been accomplished through multiple pathways including heterogeneous catalysis, electrocatalysis, and photocatalysis, the proposed enzymatic process is advantageous due to its ability to operate at ambient temperatures and atmospheric pressure¹¹⁻¹². However, this multi-step reduction of carbon dioxide faces its own challenges. The synthesis requires low pH and elevated temperature for maximum efficiency, which can cause the enzymes to degrade in turn.¹³

We propose that by using *Pichia pastoris*, a methylotrophic yeast capable of expressing and secreting non-native proteins, to produce FDH, FIDH, and ADH. If such a system can be created, it has the potential to be both economic and renewable¹⁴. The yeast will continuously secrete the protein into media, replacing enzymes as they degrade and extending the lifetime of the biofuel-producing system. This approach eliminates the time-consuming step of growing, expressing, and purifying the desired proteins from bacteria.

4.1.2. Secretory pathway of *Pichia pastoris*

In the early 1970's, Phillips Petroleum utilized *Pichia pastoris* (also sometimes known as *Komagataella pastoris*) as a method of producing single cell protein (SCP) as an additive for animal feed.¹⁵ By exploiting the ability of *P. pastoris* to use of methanol as a carbon source Phillips was able to make SCP in large quantity. Years later, Cregg *et al* successfully transformed *P.*

P. pastoris into a host for DNA transformations.¹⁶ Since that development, this yeast has been utilized for the industrial production of over twenty different proteins to date, including several biopharmaceuticals.¹⁷⁻¹⁸

P. pastoris has several important characteristics that make it an ideal platform for protein production.¹⁹ One characteristic is the alcohol oxidase I (AOX1) gene that controls the expression of foreign proteins is activated upon the addition of methanol (Figure 4.2). Another factor is that *P. pastoris* prefers respiratory growth that allows it to grow at high cell densities much like other popular fermenting yeasts such as *Saccharomyces cerevisiae*. The final major advantage *P. pastoris* has over other expression systems is the presence of the α -mating factor (Figure 4.2).¹⁵ This gene signals for the organism to secrete folded (and sometimes even post-translationally modified) protein outside the yeast cell into the media where it can be more easily collected.

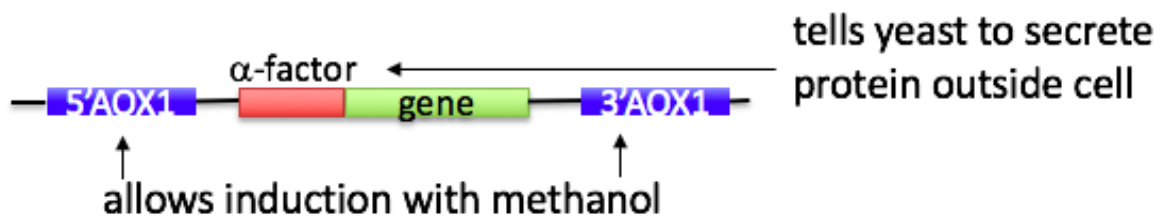


Figure 4.2. A summary of the mechanisms *Pichia pastoris* uses in order to express and secrete protein.

There are many ways to insert the desired gene into our eukaryotic factory.²⁰ One of the most common ways is to use a vector known as pPic9 (Figure 4.4). This 8 kilobase vector contains AOX1 which is necessary for methanol induction and the α -factor necessary to secrete the expressed protein outside into media. It also contains the His4 gene, which codes for histinol upon recombination into the yeast genome.¹⁹ First, one clones the desired gene using any of the

five restriction enzymes: XhoI, SnaBI, EcoRI, AvrII, and NotI.²⁰ After the gene has been successfully cloned into the vector, the entire vector is cut with either Sall or Stul, which essentially cuts the His4 gene in half. The cut vector is then transformed into a strain of yeast known as GS115 (Figure 4.3). This strain is incapable of producing histidine on its own, so when the cut vector recombines with the yeast genome, it completes the His4 gene and allows the yeast to be self-sufficient in histidine production. This allows for screening of the gene in transformed yeast cells.

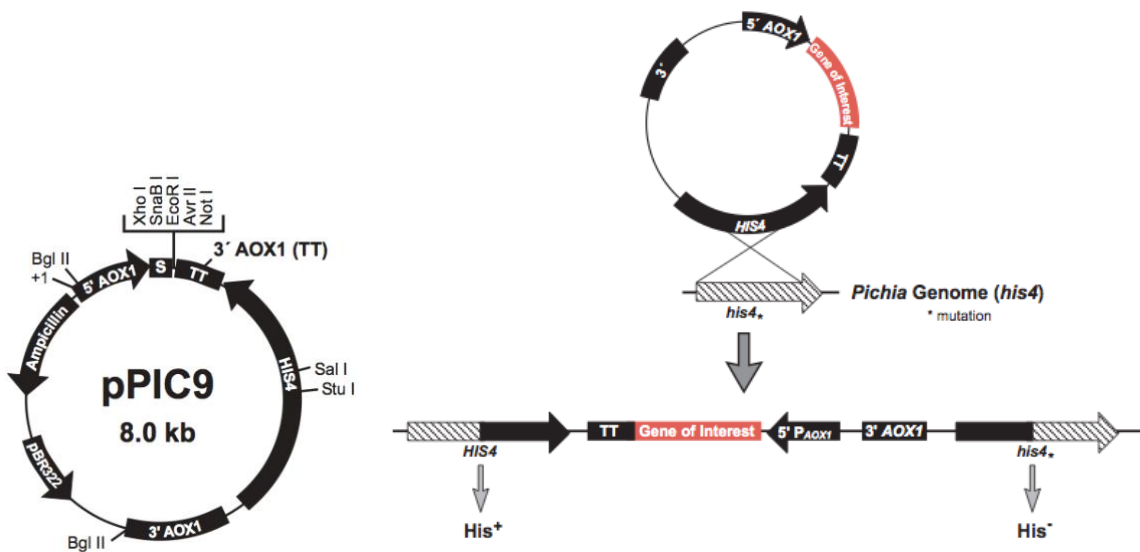


Figure 4.3. a) The pPic9 vector used in yeast transformation; b) Summary of homologous recombination of pPic9 into GS115 genome: the gene of interest is cloned into the pPic9 vector. Upon transformation, the pPic9 vector will integrate itself into the yeast genome, which contains a mutation that does not allow for the production of histidine in yeast. Addition of the pPic9 vector into the genome allows for yeast to produce histidine, which allows for screening of yeast colonies for the desired gene.²⁰

Once the gene is inserted into the yeast genome, the protein can then be expressed and secreted by *P. pastoris* by growing it in a mixture of yeast extract, peptone or tryptone, dextrose, and 0.5% methanol.²⁰ Protein synthesis occurs in the ribosome of the cell and is then moved to the endoplasmic reticulum (ER) where glycosylation and other post-translational modifications

can occur.²¹ The protein is then transported to the Golgi in vesicles (Figure 4.4). If the protein is not secreted, it will be released into the cytosol of the cell.

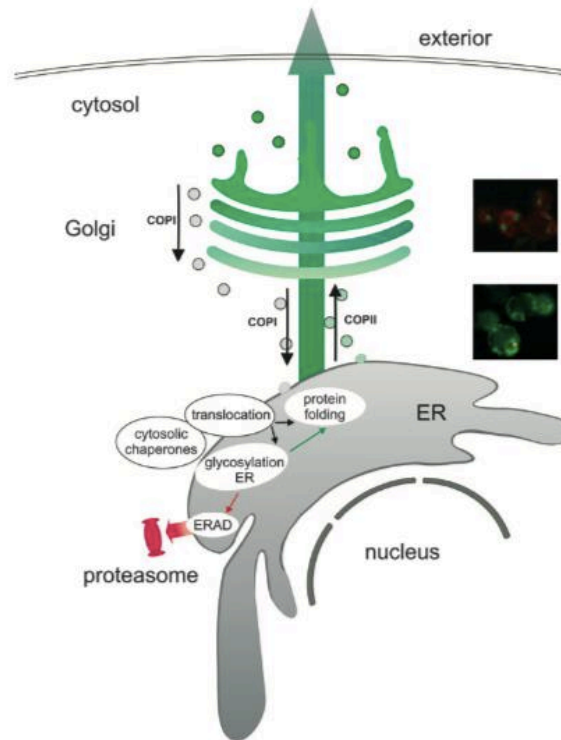


Figure 4.4. Protein synthesis in yeast from the nucleus to the Golgi apparatus (where glycosylation may occur) to outside the cell.²¹

Since yeast are eukaryotic cells that are surrounded by cell walls, the Golgi must make further modifications to the vesicles carrying the protein so that they may leave (Figure 4.5).²² From there, the vesicles are transported to the cell wall, where exocytosis occurs and the protein is released.

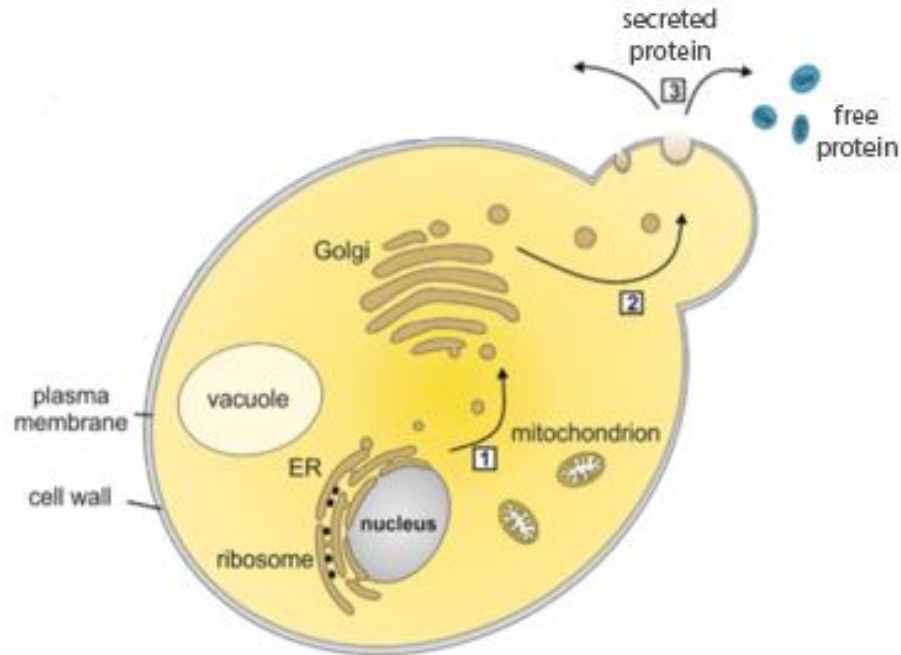


Figure 4.5. Secretory pathway of yeast.²²

4.1.3. Constructing a biofuel producing apparatus

Normally, FDH, FIDH, and ADH require the coupled oxidation of cofactor NADH to NAD⁺ to catalyze the reaction of CO₂ to CH₃OH.^{12, 23-24} Our collaborators at Lehigh University are creating a method to replace the cofactors with electrons generated photochemically by quantum dots (QDs) (Figure 4.6).¹⁴ The advantage to this method is that QDs can be continuously regenerated by bacteria and will absorb energy from visible light to create electron-hole pairs that act as cofactor equivalents.²⁰ The QDs will act as a photocatalyst for splitting water and the resulting electrons generated by splitting the water will be captured by a reduced graphene oxide

(rGO) surface.^{14, 25} The enzymes will be modified to have affinity tags that can bind to the rGO surface.¹⁴

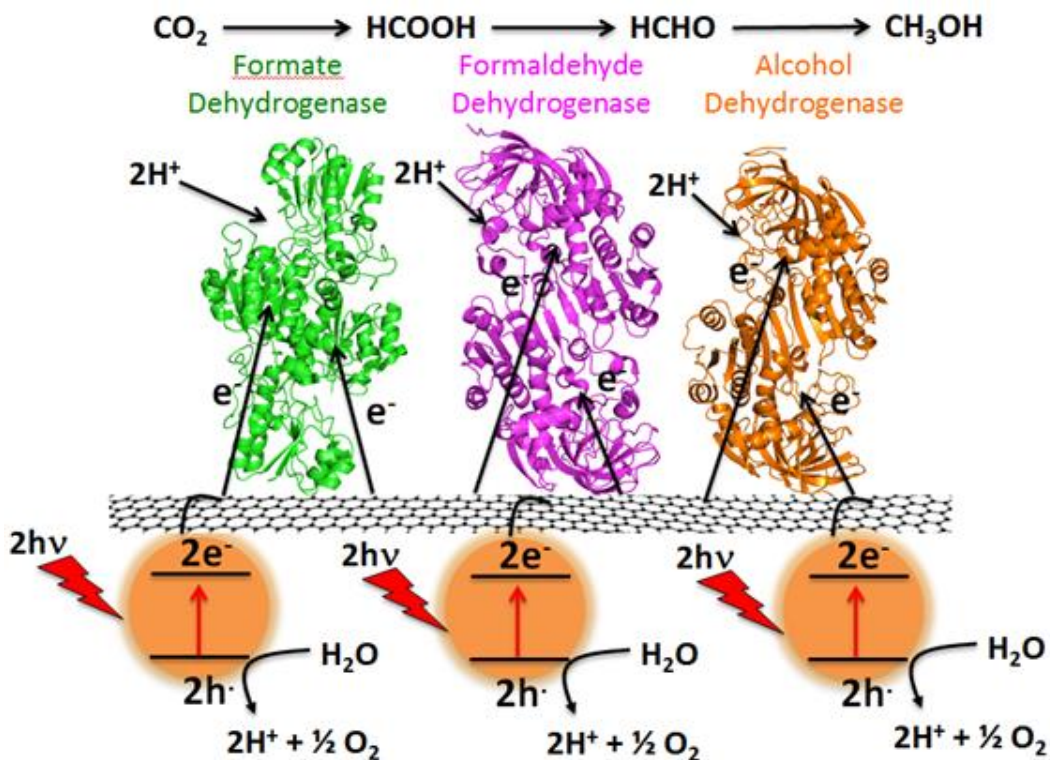


Figure 4.6. Model of QD-protein system.

The overall result of this process yields a self-sustaining system that can use light as an energy source and requires minimal space (Figure 4.7). The methanol produced by the system will go in part to signal the secretion of more protein so that it is continually being renewed without outside interference. Meanwhile, the QD surface will act as a place for protein to attach as well as providing an electron source to replace NADH.

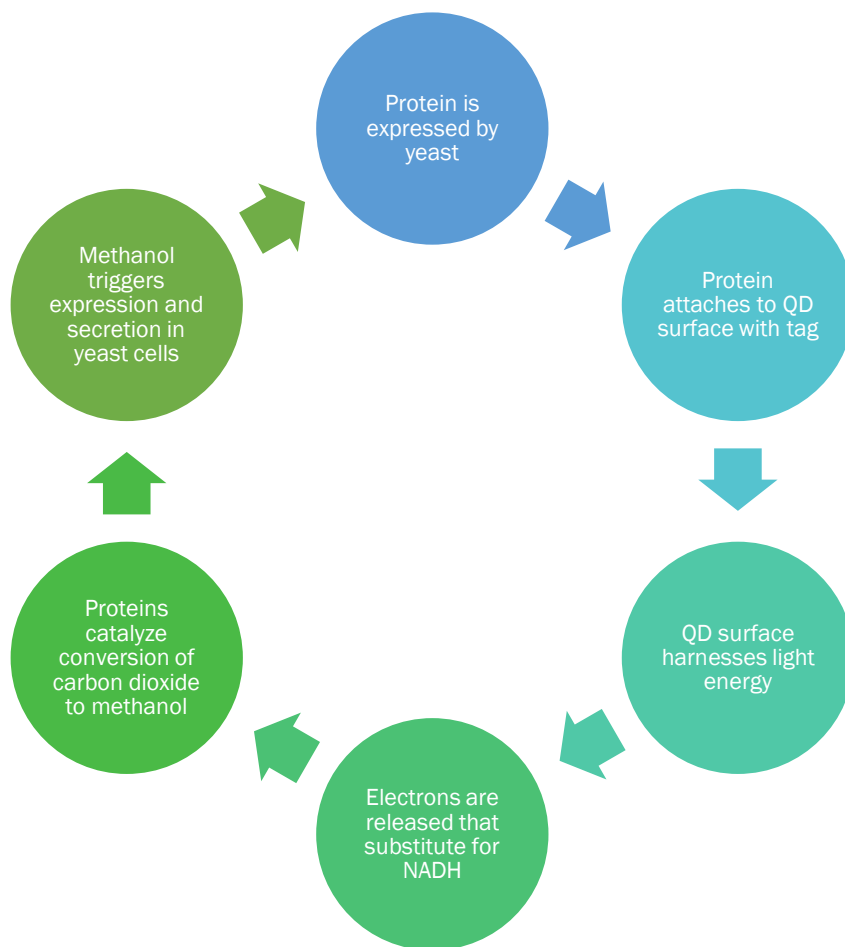


Figure 4.7. Summary of self-sustaining QD-protein system.

4.2. Design and Screening pPic9 Vector of FDH Secretion

Formate dehydrogenase is a 42 kDa protein that is normally expressed in *Candida boidinii* (Figure 4.8).²⁶ It catalyzes the conversion of formate into carbon dioxide with the cofactor NAD⁺. It is commonly used as a means of regenerating the NADH cofactor in various systems.²⁷

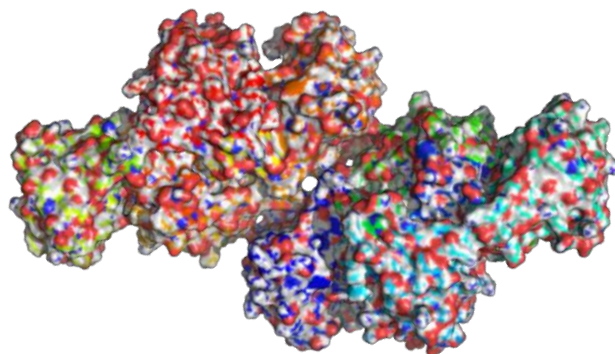


Figure 4.8. Crystallography structure of formate dehydrogenase (pdb 5DNA).²⁸

The construct consisting of the FDH sequence was successfully integrated into the chosen plasmids. FDH-N was tagged at the 5' end with a six histidine coding sequence in the pMCSG49 vector and, likewise, FDH-C was tagged at the 3' end with a His₆ coding sequence in the pET28a vector. Both expressed upon induction with 0.4 mM IPTG, which was visualized in a well-defined protein band at ~45 kDa. The protein was successfully isolated after Ni-NTA and Q-sepharose.

The specific activity was dependent upon the placement of the His₆-tag, where FDH-N was more active than FDH-C in the oxidation of NAD⁺ to NADH as assayed by the subsequent conversion of formate to carbon dioxide (Figure 4.9).

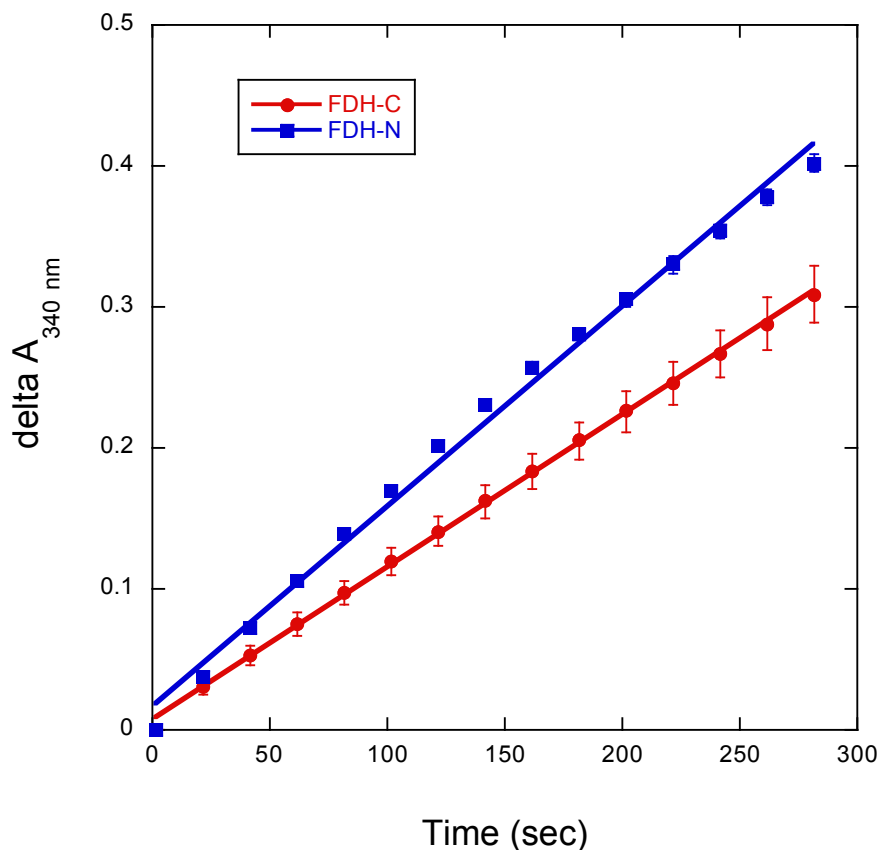


Figure 4.9. Catalytic efficiency of FDH-C and FDH-N proteins (0.5 μM) measuring the conversion of NAD^+ (1 mM) to NADH during the oxidation of sodium formate (160 mM) to carbon dioxide in 100 mM sodium phosphate buffer, pH 7. The specific activity was calculated to be 0.526 U/mg for FDH-C and 2.77 U/mg for FDH-N.²⁹

Preliminary experiments were also done to test the conversion of CO_2 to formate, the desired direction of our reaction. Sodium bicarbonate was used as a substitute for carbon dioxide since it is easier to control the concentration of bicarbonate versus aqueous CO_2 and bicarbonate anions are dominant at neutral pH where we want our reaction to take place.³⁰ As expected, the conversion of bicarbonate to formate was much lower than its preferred direction, however, some conversion did take place (Figure 4.10).

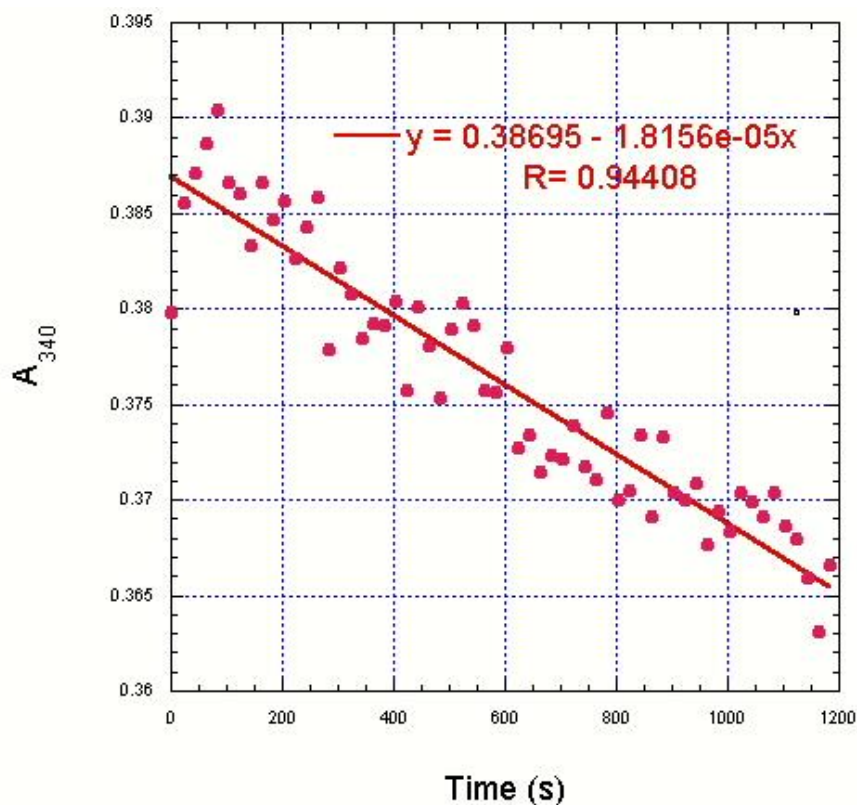


Figure 4.10. Catalytic efficiency of conversion of NADH (1.2 mM) to NAD⁺ by FDH-N (0.75 mM) using sodium bicarbonate (200 mM) in 100 mM sodium phosphate buffer, pH 7.0. The specific activity of FDH-N for NaHCO₃ to formate conversion is 0.005 U/mg. (MT2051)

4.3. Characterization of Yeast-Secreted Formate Dehydrogenase

4.3.1. Time Elapsed Activity of Secreted Yeast

We cloned both a version of FDH that used an N-terminal His₆ (FDH-N) affinity tag instead of a C-terminal tag (FDH-C) since FDH-N had greater activity than FDH-C when expressed in *E. coli*. The proteins were secreted by yeast into media over the course of five days and visualized by SDS-PAGE (Figure 4.10). After 48 hours, the approximate concentration of protein in media was 8-10 mg/L. Upon purification of FDH from the BMMY media, there was no evidence of cleavage or degradation of protein from natural yeast proteases on SDS-PAGE gel (Figure 4.11).¹¹

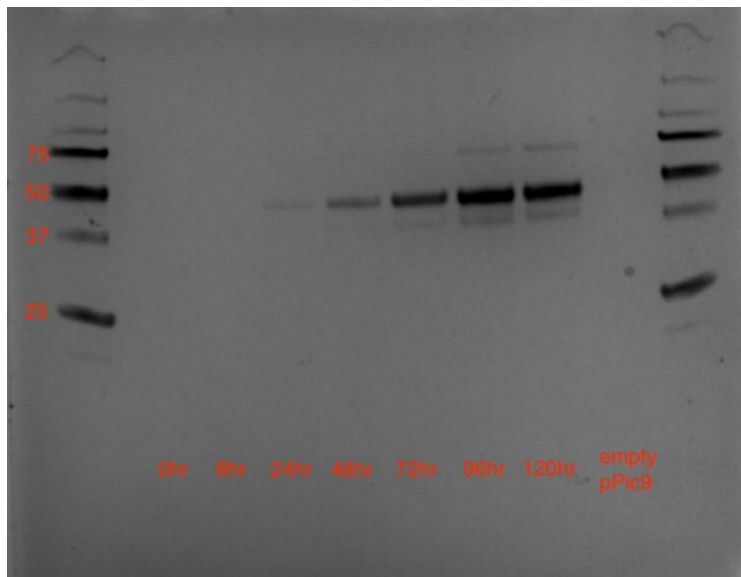


Figure 4.11. Continuous secretion of FDH from yeast over 120 hours. Empty pPic9 vector was also cloned into yeast as a control.²⁹

Proteins were purified from yeast media by Ni-NTA column and their specific activities were measured in the conversion of formate to carbon dioxide using UV-Vis spectroscopy. To measure kinetics, the absorbance of NAD⁺/NADH (340 nm) was measured over a period of time and the slope is then used to calculate specific activity. Specific activity is used to measure the amount of catalytically active protein present. One unit of activity (U) is defined as the production or consumption of 1 μ mol of NADH per minute. The specific activity of the proteins was measured at 24, 48, and 120 hours and found to be consistent over time (Figure 4.12).

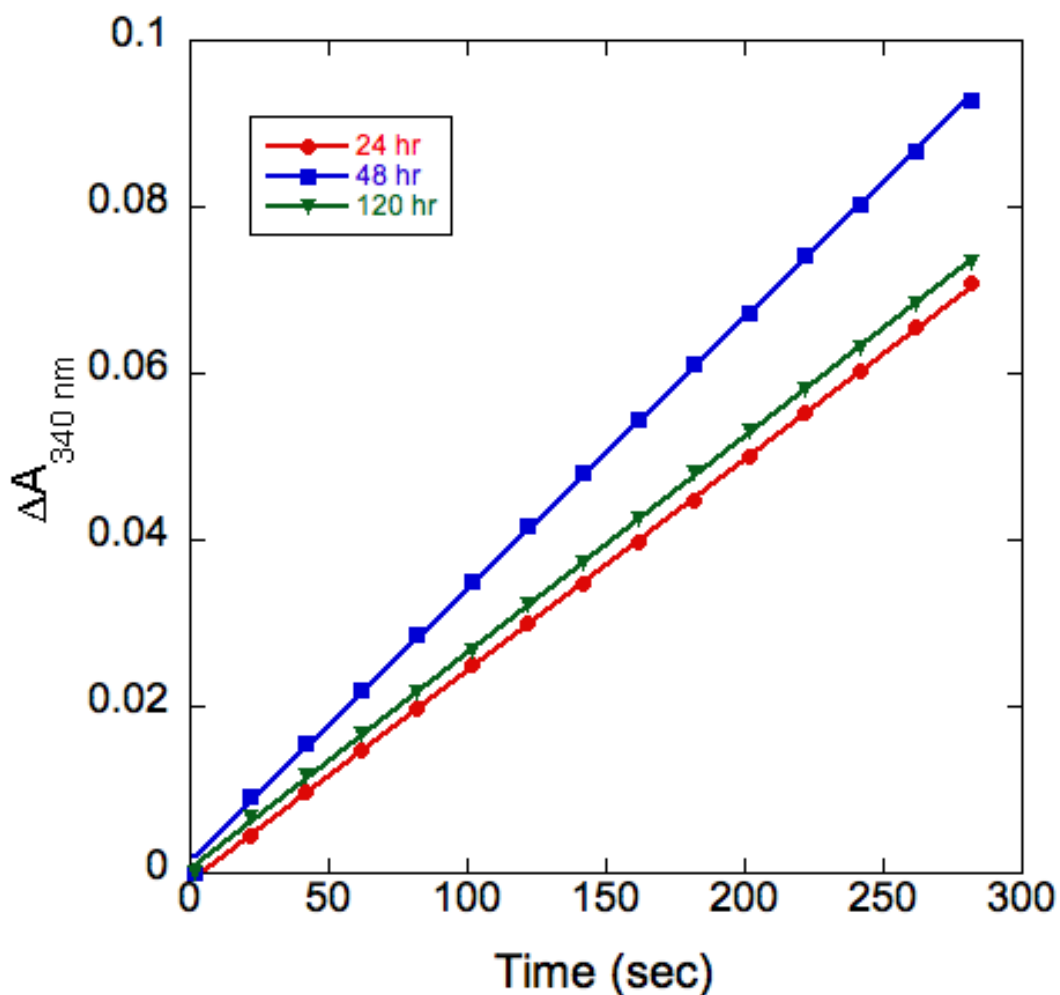


Figure 4.12. Catalytic efficiency of FDH-C- γ (0.5 μM) measuring the conversion of NAD^+ (1.0 mM) to NADH during the oxidation of sodium formate to carbon dioxide in 100 mM sodium phosphate, pH 7. The specific activity of FDH at 24 hours is 0.11 U/mg, at 48 hours is 0.13 U/mg, and at 120 hours is 0.10 U/mg.²⁹

Table 4.1. Summary of specific activities of enzymes for FDH.

Enzyme	Formate \rightarrow CO_2
FDH-N (<i>E. coli</i>)	2.710 U/mg
FDH-N (<i>P. pastoris</i>)	393 U/ μg
FDH-C (<i>E. coli</i>)	590 U/ μg
FDH-C (<i>P. pastoris</i>)	19 U/ μg

4.3.4. Investigating the Decreased Activity of Secreted Formate Dehydrogenase

When the activities of FDH expressed in bacteria and FDH secreted by media were compared, there was found to be an almost ten-fold difference in activity as shown in Figure 4.13. This is not an unusual problem in proteins secreted by yeast and can usually be attributed to either unwanted glycosylation of the protein,²⁰ denaturing proteases in the media,²⁵ or incomplete cleavage of the α -mating factor.³¹

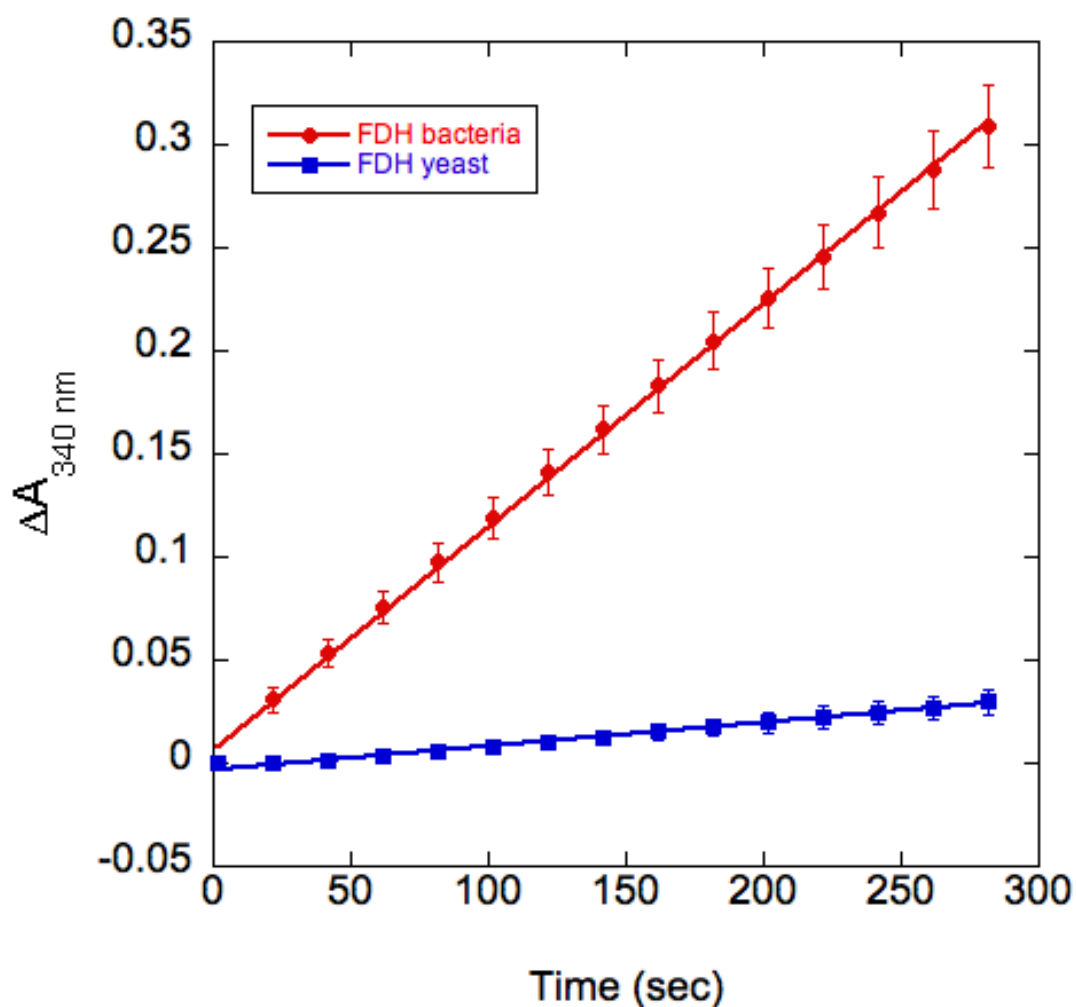


Figure 4.13. Catalytic efficiency of FDH-N (0.5 μ M) from *E. coli* and *P. pastoris* constructs measuring the conversion of NAD⁺ (1 mM) to NADH during the oxidation of sodium formate (160 mM) to carbon dioxide in 100 mM sodium phosphate buffer at pH 7. The specific activity of FDH from bacteria is 2.71 U/mg and the specific activity of FDH from yeast is 0.4 U/mg.²⁹

4.3.4.1. Glycosylation

Mass spectrometry of formate dehydrogenase protein as it is secreted by yeast reveals that glycosylation, a form of post-translational modification that can be quite common with secreted proteins, was not occurring with our protein. As Figure 4.14 shows MALDI-TOF mass spectrometry shows the molecular weight matches that of an unmodified FDH, ca 43 kDa.

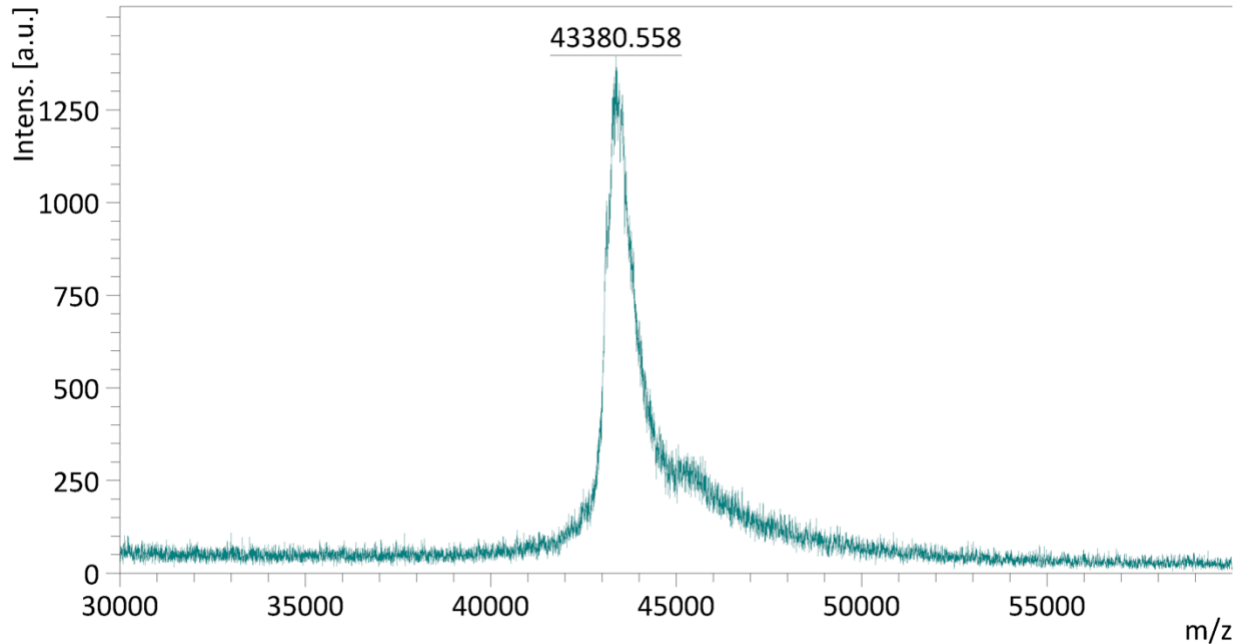


Figure 4.14. MALDI-TOF of secreted FDH C-term. The matrix used was sinapic acid. The molecular weight matches the weight of the FDH if it were unmodified (43.3 kDa).²⁹

4.3.4.2. Proteases

Another common issue in protein secretion from *P. pastoris* is undesired and uncontrolled proteolysis of the product protein.^{15, 25} This can occur either during transport of the recombinant protein outside of the cell in vesicles, by proteases in the extracellular space of the cell wall, or by proteases in the culture medium.¹⁵ This can be a very serious issue since not only does it lower protein yield, but the degradation products can be difficult to separate from the desired product.

There are several methods of reducing proteolytic activity during yeast secretion. The first way is to employ PMSF.²⁵ PMSF is commonly used when purifying proteins from bacterial cell pellets by inhibiting proteases that may be released from the cells during lysis. EDTA is a metal chelator that can inhibit metalloproteases. Casamino acids, which is a mixture of amino acids and

small peptides obtained by hydrolyzing casein, can reduce proteolytic activity by competing with the secreted protein for proteolysis.

When cultures were grown with either no additives (control), PMSF, EDTA, casamino acids, or all three, there was no change in appearance on the protein gel at 24 and 48 hours (Figure 4.15). In addition, when the catalytic activity of these proteins was tested, with the exception of the sample with all three protease inhibitors, there was little difference in activity between the four (Table 4.2). The combination of the SDS-page and kinetic assay led us to conclude that proteases played a minor role in the decreased activity of the secreted FDH, if at all.

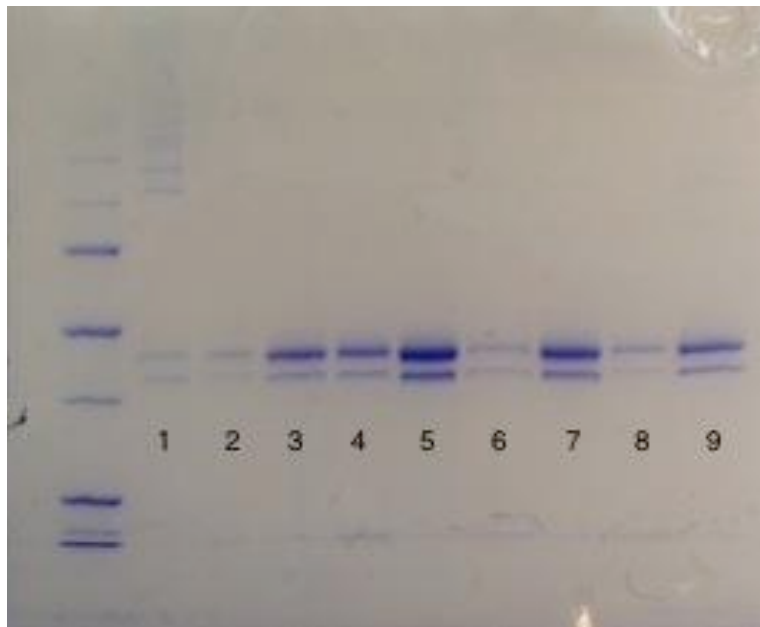


Figure 4.15. 10% acrylamide SDS-PAGE gel. Lanes: 1) control 48 hr, 2) PMSF 24hr, 3) PMSF 48 hr, 4) EDTA 24 hr, 5) EDTA 48 hr, 6) casamino acids 24 hr, 7) casamino acids 48 hr, 8) all three 24 hr, 9) all three 48 hr.

Table 4.2. Comparison of specific activity of secreted FDH in the presence of protease inhibitors.

	Control	PMSF	EDTA	Casamino acids	All three
Specific activity (U/ μ g)	19	17	22	25	8

4.3.4.3. Incomplete cleavage

With several possibilities for low specific activity of the secreted FDH discounted, we looked to other possible implications to why our protein might have reduced activity in comparison to protein we previously expressed in bacteria. Since previous experiments had shown that time did not seem to be a factor towards the protein's efficacy, we concluded that the problem must lie in the secretion method of the protein.

Other non-natural proteins that have secreted by yeast have been reported to incompletely cleave the α -secretion factor.³¹ To test whether this was true for our protein, we used tryptic digestion of our purified FDH to determine if any fragments would correspond to the α -secretion factor (Figure 4.16).³² The advantage to this approach was that α -secretion factor would have a very distinct Maldi spectra at 7624 Da compared to the rest of the fragmented protein, which would be very small in comparison (the list of all fragments is included in the appendices). Indeed, our hypothesis was proved correct and the Maldi spectrum of the tryptic digest did prominently show that key peak at 7624 m/z .

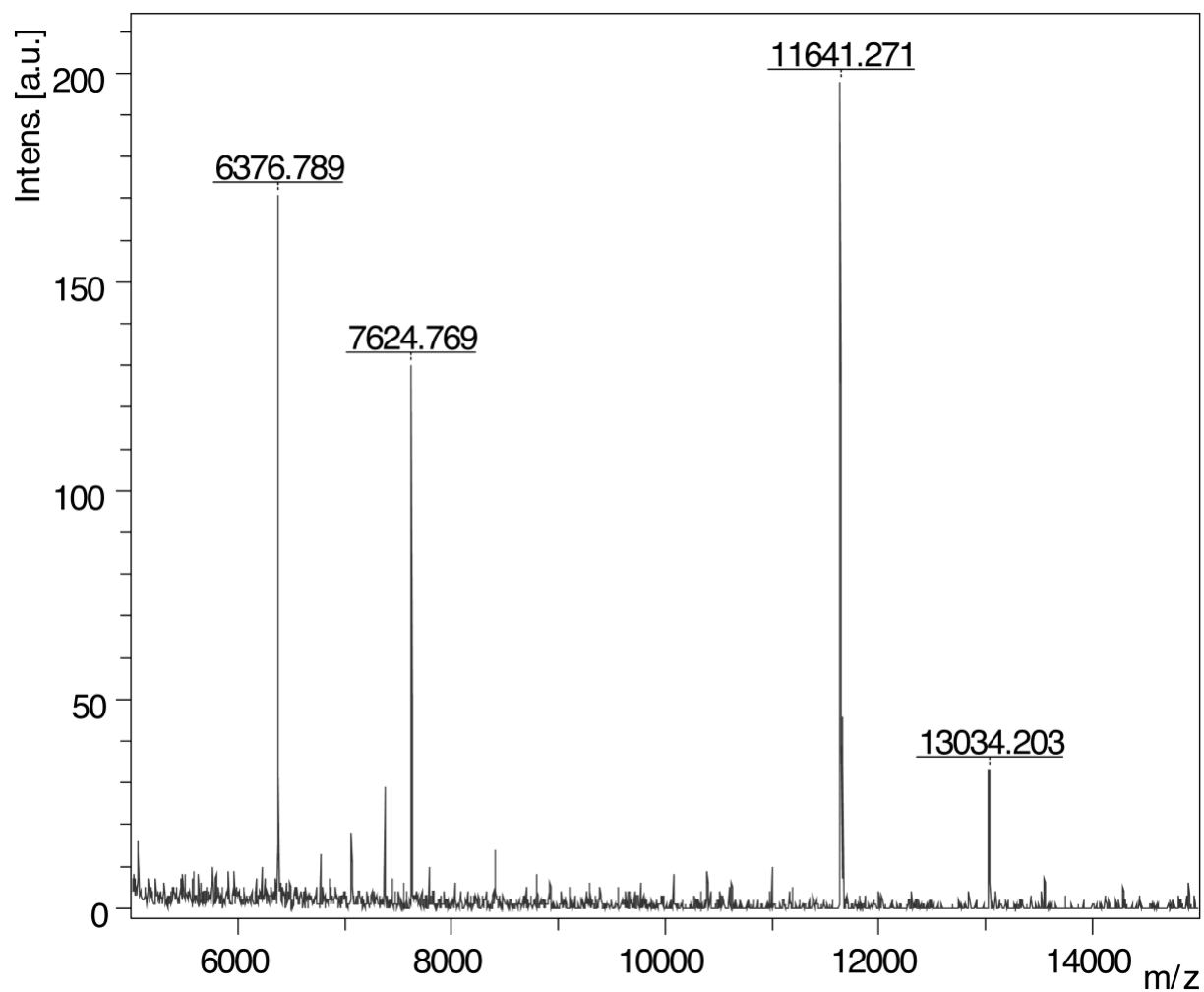


Figure 4.16. MALDI-TOF of FDH C-term digested by trypsin. The peak at 7624 Da corresponds to a digested α -factor.²⁹

4.3.5. Formaldehyde dehydrogenase and Alcohol dehydrogenase

FIDH (derived from both *P. aeruginosa* and *P. putida*) and ADH (from *S. cerevisiae*) were also cloned into *E. coli* and their specific activities were determined. The summary of their specific activities can be found in Table 4.3 below and in the appendix (S.34-S.38).

Table 4.3. Summary of specific activities of FIDH proteins and ADH in the conversion of NADH. Kinetic assays can be found in S.34 to S.38.

Protein	Reaction	Specific activity (U/mg)
<i>FIDH-aeru</i>	formaldehyde → formate	2.6
<i>FIDH-aeru</i>	formate → formaldehyde	0.026
<i>FIDH-puti</i>	Formaldehyde → formate	3.1
<i>FIDH-puti</i>	formate → formaldehyde	0.0003
<i>ADH</i>	formaldehyde → methanol	84

The same protocol for yeast transformation for FDH was used for FIDH and ADH. FIDH (EC 1.2.1.1) is responsible for the conversion of formate into formaldehyde in *Pseudomonas putida*.²³ This particular protein strain is of special interest because unlike other FIDH subtypes it does not require the cofactor glutathione (GSH). Finally, ADH (EC 1.1.1.1), also called methanol dehydrogenase, is found in *Saccharomyces cerevisiae* and oxidizes methanol into formaldehyde.²⁴ The colonies on minimal media were screened. Despite growing without histidine, only two colonies had the actual FIDH gene and no ADH colonies contained the gene based on Sanger sequencing. The positive FIDH-yeast strains were grown for 120 hours and then purified using Ni-NTA. No protein was visible on UV-Vis or gel before or after purification (Figure 4.17).

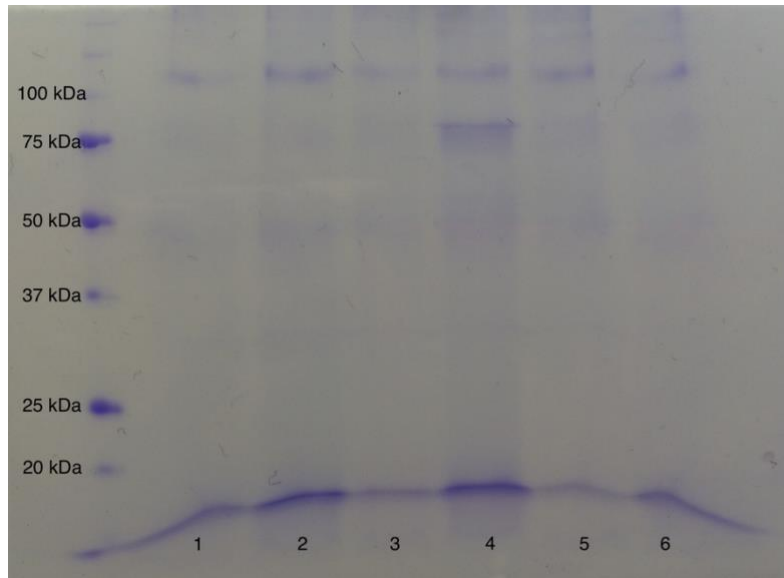


Figure 4.17. SDS-PAGE (10% acrylamide) of the secretion of FIDH-puti protein cloned into yeast. A band is expected at ~35 kDa. Lanes 1-6 represent the media secretion of different “positive” hits of FIDH-puti cloned into yeast. Even though the gene was present in the plasmids, there was no expression.

The cloning of ADH and FIDH into *P. pastoris* where secretion occurs is an ongoing effort. It is possible that other strains of yeast would be capable of this process, but they have not been investigated at this time.^{17, 33}

4.4. Conclusions

In a proof-of-principle study, we cloned the FDH gene from *Candida biodinii* into the vector pPic9, which incorporates into the *P. pastoris* genome via homologous recombination. With the addition of 0.5% methanol every 24 hours, the yeast cells continuously secrete FDH into media. The FDH can be separated from the media using a His₆-tag on FDH. *Pichia pastoris* is capable of expressing and secreting proteins into media. We found the activity of FDH remained consistent, but low due to incomplete cleavage of the α -secretion factor, over an extended

period of time. Nonetheless, this study demonstrates partial proof-of-concept and sets the stage for future studies for this multi-enzymatic pathway of biofuel production.

4.5. Experimental

4.5.1. Ligation-independent Cloning of FDH-N into pMCSG49

The pMCSG49 expression vector was digested using *SspI* restriction enzyme in Cutsmart buffer (New England Biolabs (NEB)). The mixture was incubated at 37°C for 1.5 h. Digested plasmid was purified using Nucleospin Extract II (Machery-Nagel) low yield plasmid protocol. Gene of interest and digested plasmid blunt ends were made separately using T4 DNA Polymerase in 10X T4 DNA Polymerase buffer (NEB) supplemented with 5 mM dithiothreitol (DTT), 0.05 ng/μL Bovine Serum Albumin (BSA), and 10 mM dGTP (for pMCSG49) or 10 mM dCTP (for genes of interest). The reactions were incubated at 22°C for 30 min, followed by incubation at 75°C for 20 min. Digested vector was added to gene insert in a 1:8 ratio (based upon mass) and incubated at room temperature for 5 min. The reaction was stopped with addition of 1.2 mM EDTA and incubating the mixture for 5 min. Plasmid was transformed into *E. coli* XL-10 cells and purified using EZ-10 Spin Column Plasmid DNA kit (Bio Basic) low plasmid yield protocol. The sequence was confirmed using Sanger sequencing provided by Genescript.

4.5.2. Cloning of FDH-C into pET-28a

The pET28a vector and gene of interest were both digested using *SalI* and *NcoI* restriction enzymes in Buffer 3.1 (NEB). Digests were purified using Nucleospin Extract II low yield protocol.

Ligation of pET28a and gene insert was done using T4 DNA ligase manufacturer protocol (NEB) with a 1:8 vector to insert ratio (based on mass) and incubating for 4 h at room temperature. Plasmid was transformed into *E. coli* XL-10 cells and purified using EZ-10 Spin Column Plasmid DNA kit low plasmid yield protocol. The sequence was confirmed using Sanger sequencing provided by Genescript.

4.5.3. Protein Expression and Purification

4.5.3.1. FDH in *E. coli*

The appropriate vectors containing FDH were transformed into *E. coli* BL21(DE3) cells and expressed in LB. After cells reached an OD₆₀₀ of ~0.6, IPTG was added to a final concentration of 0.4 mM and the temperature was lowered to 30°C. After four hours, the cells were collected by centrifugation and resuspended in lysis buffer containing 50 mM Hepes and 5% glycerol (pH 7.6). PMSF was added to a final concentration of 0.5 mM. Cells were lysed by sonication, and then the crude cell lysate was centrifuged at 20,000xg for 30 min. A 5 mL solution of streptomycin sulfate diluted in lysis buffer was added to supernatant (on ice) at a final concentration of 1% w/v and allowed to stir for 15 min. Cells were pelleted by centrifugation at 20,000xg for 30 min. The supernatant was applied onto a Ni-NTA column and washed with lysis buffer. The protein was eluted with increasing concentrations of imidazole in 50 mM Hepes, 5% glycerol (pH 7.6) buffer with concentration ranging from 30-300 mM. Protein fractions were diluted two-fold and applied onto a Q-sepharose Fast Flow column, then washed with lysis buffer. Protein was eluted with 50

mM Hepes, 5% glycerol, and 300 mM NaCl (pH 7.6). The presence and purity of the protein was confirmed using SDS-PAGE.

4.5.3.2. FIDH-aeru, FIDH-puti, and ADH

The same procedure as 4.5.3.1. was used to clone and purify FIDH-aeru, FIDH-puti, and ADH.

4.5.4. Cloning of FDH Genes into pPic9

The *E. coli* vectors containing the genes of interest and the empty pPic9 vector were digested using *EcoRI* and *NotI* restriction enzymes in Buffer 3.1 (NEB) at 37°C overnight. Digested vector was purified using Nucleospin Extract II kit per the manufacturer's protocol.

The digested pPic9 vector and gene of interest were ligated using a 1:20 vector to gene ratio (based on mass) with T4 DNA Ligase and T4 ligase buffer. The reaction incubated at room temperature for four hours. After ligation, transformed into *E. coli* XL-10 cells. DNA was extracted from colonies using EZ-10 Spin Column Plasmid DNA kit low plasmid yield protocol and mutated sequences were confirmed by Genescript using Sanger sequencing technique.

4.5.5. Protein Expression and Purification of FDH in *P. pastoris*

The pPic9 vector containing the genes of interest was cut using *SalI* restriction enzyme in buffer 3.1 (NEB) at 37°C for 2 h. Cut vector was purified using Nucleospin Extract II kit protocol.

The transformation setup is outlined by Invitrogen.²⁰ GS115 yeast cells were grown in YPD (1% yeast extract, 2% tryptone, and 1% dextrose) at 28.5°C until OD₆₀₀ reached ~1.5. Culture was pelleted at 3,000xg. The cell pellet was resuspended in LiAc-DTT buffer (100 mM lithium acetate,

10 mM dithiothreitol, 600 mM sorbitol, 10 mM Tris pH 7.5) and allowed to incubate at room temp for 40 minutes. Cells were centrifuged at 3,000xg and pellet was washed twice with 1 M sorbitol. Cells were resuspended in 1 M sorbitol until the concentration at OD₆₀₀ was between 100-200. Approximately 3.5 µg of plasmid was mixed with GS115 cells and a 1.5 kV was applied to the cells using an Eppendorf 2510 electroporator. Cells were plated onto Regeneration Dextrose Base (RDB-) plates (1 M sorbitol, 2% dextrose, 1.34% Yeast Nitrogen Base (YNB), 4 x 10⁻⁵% biotin, 0.005% amino acids, 2% agar) according to the protocol set forth by Invitrogen.⁸ These plates lack histidine, so only successful recombinations will be able to grow on these plates. To determine if our sequence was present, colonies that grew on the RDB- plates were suspended in 20 µL of sterile water and analyzed by running PCR. The sample contained 10 µL of DNA template (in the inoculated water), 1.2 µL of forward primer (T7 promoter, 100 ng/ml), 1.2 µL of reverse primer (100 ng/mL) and 12 µL of GoTaq Polymerase master mix (Promega). PCR was run using an initial denaturation cycle of 2 min (95°C), 15 cycles 30 seconds denaturation (95°C), 30 seconds annealing (57°C), and 45 seconds extension (72°C) and a final extension cycle of ten minutes (72°C). Samples that contained the gene displayed a band at approximately 300 bp on a 10% acrylamide gel were deemed positive hits and used to inoculate YPD according to protocol from Invitrogen.²⁰

Cells were pelleted at 2,000xg and resuspended in buffered methanol-complex medium (BMMY; 1% yeast extract, 2% tryptone, 100 mM sodium phosphate pH 6.0, 1.34% YNB, 4 x 10⁻⁵% biotin, 0.5% methanol) and allowed to grow from 120 hours at 28.5°C. Every 24 hours, methanol was added to BMMY to a final concentration of 0.5%.

To purify secreted protein in BMMY, culture was centrifuged at 5,000 g for 2 min. Ni-NTA was equilibrated with lysis buffer containing 50 mM Hepes and 5% glycerol (pH 7.6). Ni-NTA was added to media and rocked on ice for 2 h. Media-resin mixture was washed with lysis buffer. Protein was eluted with buffer containing 50 mM Hepes, 5% glycerol, and 300 mM imidazole (pH 7.6). Presence and purity of protein was confirmed using UV-Vis spectrophotometry and SDS-PAGE.

4.5.6. Kinetic Assays for FDH

4.5.6.1. Conversion of formate to carbon dioxide

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 20 s over a period of 10 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 1.0 mM NAD⁺ and 160 mM formate. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 1.0 μ M FDH ($\epsilon_{280} = 54,445 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 43,845 g/mol.

4.5.6.2. Conversion of sodium bicarbonate to formate

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 20 s over a period of 20 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 0.12 mM NADH and 120 mM mM bicarbonate. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 10.0 μM FDH ($\epsilon_{280} = 54,445 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 43,845 g/mol.

4.5.7. Kinetic Assays for FIDH-aeru

4.5.7.1. Conversion of formaldehyde to formate

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 10 s over a period of 2.5 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 1.0 mM NADH and 100 mM formaldehyde. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 0.25 μM FIDH ($\epsilon_{280} = 44,544 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific

activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 34,755 g/mol.

4.5.7.2. Conversion of formate to formaldehyde

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 10 s over a period of 5 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 0.2 mM NADH and 160 mM formate. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 1.0 μM FIDH ($\epsilon_{280} = 44,544 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 34,755 g/mol.

4.5.8. Kinetic assays for FIDH-puti

4.5.8.1. Conversion of formaldehyde to formate

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 20 s over a period of 5 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 1.0 mM NADH and 100

mM formaldehyde. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 0.1 μM FIDH ($\epsilon_{280} = 34,755 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 42,328 g/mol.

4.5.8.2. Conversion of formate to formaldehyde

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 20 s over a period of 2 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 0.1 mM NADH and 100 mM formate. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 12.0 μM FIDH ($\epsilon_{280} = 34,755 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 42,328 g/mol.

4.5.9. Kinetic assay for ADH

4.5.9.1. Conversion of formaldehyde to methanol

All kinetic measurements were done on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Unless specifically stated, kinetic measurements were recorded every 10 s over a period of 1.5 min. Enzyme activity was characterized in a buffer containing 100 mM phosphate (pH 7) with 1.0 mM NADH and 100 mM formaldehyde. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 16.0 μM ADH ($\epsilon_{280} = 50,350 \text{ M}^{-1}\text{cm}^{-1}$). To calculate specific activity, all kinetics measurements were fit to a linear model and calculated with the following equation: [specific activity = $((\text{slope}/6220 \text{ M}^{-1}) * 60\text{s}/\text{min} * V_f) / m_f$] where V_f is the final volume of the enzymatic assay and m_f is the mass (in mg) of the final concentration of protein. The molecular weight used to calculate specific activity was 39,385 g/mol.

4.5.10. Tryptic digestion of FDH

Protein stock ($\sim 0.1 \mu\text{g}$ protein), buffer (100 mM Tris, 6 M urea, pH 7.8), and reducing agent (100 mM Tris, 200 mM DTT, pH 7.8) were combined was combined in a 1:1:0.1 volume ratio (for a total volume of 105 μL) and allowed to sit at room temperature for one hour. To alkylate cysteines, 20 μL of alkylating agent (100 mM Tris, 200 mM iodoacetamide, pH 7.8) was added and mixture was allowed to sit for 1 h. To remove unreacted iodoacetamide, 20 μL reducing agent was added. The sample was diluted to 775 μL using water and 100 μL of porcine trypsin (SAFC, 200 ng/ μL , 100 mM Tris, pH 7.8) was added. The mixture was allowed to digest

overnight at 37°C and was then stopped with the addition of acetic acid until the pH was below six.

4.5.11. MALDI-TOF of FDH

A Bruker Autoflex III Smartbeam MALDI-TOF mass spectrometer was used to confirm the molecular weights of proteins and protein fragments. Samples (~10 µM concentration) were mixed with a sinapic acid solution in a 1:2 volume ratio. Ovalbumin (44.3 kDa) was used as an external standard).

4.5. References

1. Oehlschlaeger, M. A.; Wang, H.; Sexton, M. N., Prospects for Biofuels: A Review. *Journal of Thermal Science and Engineering Applications* **2013**, 5 (2).
2. Naik, S. N.; Goud, V. V.; Rout, P. K.; Dalai, A. K., Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* **2010**, 14 (2), 578-97.
3. Smeets, E.; Tabeau, A.; van Berkum, S.; Moorad, J.; van Meijl, H.; Woltjer, G., The impact of the rebound effect of the use of first generation biofuels in the EU on greenhouse gas emissions: A critical review. *Renewable and Sustainable Energy Reviews* **2014**, 38, 393-403.
4. Fargione, J. H., Jason; Tilman, David; Polasky, Stephen; Hawthorne, Peter, Land Clearing and the Biofuel Carbon Debt. *Science* **2008**, 319, 2.

5. Bhatia, L. J., Sonia; Ahmad, Rumana, An economic and ecological perspective of ethanol production from renewable agro waste: a review. *AMB Express* **2012**, *2* (65).
6. Sarris, D.; Papanikolaou, S., Biotechnological production of ethanol: Biochemistry, processes and technologies. *Engineering in Life Sciences* **2016**, *16* (4), 307-29.
7. Singh, A.; Nigam, P. S.; Murphy, J. D., Renewable fuels from algae: an answer to debatable land based fuels. *Bioresour Technol* **2011**, *102* (1), 10-6.
8. Zhang, B.; Wang, L.; Riddicka, B.; Li, R.; Able, J.; Boakye-Boaten, N.; Shahbazi, A., Sustainable Production of Algal Biomass and Biofuels Using Swine Wastewater in North Carolina, US. *Sustainability* **2016**, *8* (5).
9. Aresta, M.; Dibenedetto, A.; Angelini, A., Catalysis for the valorization of exhaust carbon: from CO₂ to chemicals, materials, and fuels. technological use of CO₂. *Chem Rev* **2014**, *114* (3), 1709-42.
10. Baskaya, F. S.; Zhao, X.; Flickinger, M. C.; Wang, P., Thermodynamic feasibility of enzymatic reduction of carbon dioxide to methanol. *Appl Biochem Biotechnol* **2010**, *162* (2), 391-8.
11. Chaudhary, Y. S.; Woolerton, T. W.; Allen, C. S.; Warner, J. H.; Pierce, E.; Ragsdale, S. W.; Armstrong, F. A., Visible light-driven CO₂ reduction by enzyme coupled CdS nanocrystals. *Chem Commun (Camb)* **2012**, *48* (1), 58-60.
12. El-Zahab, B.; Donnelly, D.; Wang, P., Particle-tethered NADH for production of methanol from CO(2) catalyzed by coimmobilized enzymes. *Biotechnol Bioeng* **2008**, *99* (3), 508-14.

13. Lightcap, I. V.; Kosel, T. H.; Kamat, P. V., Anchoring semiconductor and metal nanoparticles on a two-dimensional catalyst mat. Storing and shuttling electrons with reduced graphene oxide. *Nano Lett* **2010**, *10* (2), 577-83.
14. Zhou, M.; Ghosh, I., Quantum dots and peptides: a bright future together. *Biopolymers* **2007**, *88* (3), 325-39.
15. Ahmad, M.; Hirz, M.; Pichler, H.; Schwab, H., Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* **2014**, *98* (12), 5301-17.
16. Cregg, J. M. B., Kevin J.; Hessler, Anita Y.; Madden, Knut R., *Pichia pastoris* as a Host System for Transformations. *Molecular and Cellular Biology* **1985**, *5* (12), 3376-85.
17. Buckholz, R. G. G., Martin A. G., Yeast Systems for the Commercial Production of Heterologous Proteins. *Nat Biotech* **1991**, *9*, 1067-72.
18. Fidan, O.; Zhan, J., Recent advances in engineering yeast for pharmaceutical protein production. *RSC Advances* **2015**, *5* (105), 86665-74.
19. Cregg, J. M. C., Joan Lin; Shi, Jianying; Higgins, David R., Recombinant Protein Expression in *Pichia pastoris*. *Molec. Biotech.* **2000**, *16*, 23-52.
20. Invitrogen, *Pichia* Expression Kit. **2010**, 1-102.
21. Delic, M.; Valli, M.; Graf, A. B.; Pfeffer, M.; Mattanovich, D.; Gasser, B., The secretory pathway: exploring yeast diversity. *FEMS Microbiol Rev* **2013**, *37* (6), 872-914.

22. Lambertz, C.; Garvey, M.; Klinger, J.; Heesel, D.; Klose, H.; Fischer, R.; Commandeur, U., Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol Biofuels* **2014**, *7* (1), 135.
23. Ito, K. T., M.; Yoshimoto, T.; Tsuru, D., Cloning and high-level expression of the glutathione-independent formaldehyde dehydrogenase gene from *Pseudomonas putida*. *J Bacteriol* **1994**, *176* (9), 2483-91.
24. de Smidt, O.; du Preez, J. C.; Albertyn, J., The alcohol dehydrogenases of *Saccharomyces cerevisiae*: a comprehensive review. *FEMS Yeast Res* **2008**, *8* (7), 967-78.
25. Sinha, J.; Plantz, B. A.; Inan, M.; Meagher, M. M., Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: case study with recombinant ovine interferon-tau. *Biotechnol Bioeng* **2005**, *89* (1), 102-12.
26. Schirwitz, K.; Schmidt, A.; Lamzin, V. S., High-resolution structures of formate dehydrogenase from *Candida boidinii*. *Protein Sci* **2007**, *16* (6), 1146-56.
27. Tishkov, V. I.; Popov, V. O., Protein engineering of formate dehydrogenase. *Biomol Eng* **2006**, *23* (2-3), 89-110.
28. Guo, Q.; Gakhar, L.; Wickersham, K.; Francis, K.; Vardi-Kilshtain, A.; Major, D. T.; Cheatum, C. M.; Kohen, A., Structural and Kinetic Studies of Formate Dehydrogenase from *Candida boidinii*. *Biochemistry* **2016**, *55* (19), 2760-71.
29. Takacs, M.; Makhlynets, O. V.; Tolbert, P. L.; Korendovych, I. V., Secretion of functional formate dehydrogenase in *Pichia pastoris*. *Protein Eng Des Sel* **2017**, *30* (3), 381-6.

30. Weiner, J. H.; Choe, H.; Joo, J. C.; Cho, D. H.; Kim, M. H.; Lee, S. H.; Jung, K. D.; Kim, Y. H., Efficient CO₂-Reducing Activity of NAD-Dependent Formate Dehydrogenase from *Thiobacillus* sp. KNK65MA for Formate Production from CO₂ Gas. *PLoS ONE* **2014**, *9* (7).
31. Lin-Cereghino, G. P.; Stark, C. M.; Kim, D.; Chang, J.; Shaheen, N.; Poerwanto, H.; Agari, K.; Moua, P.; Low, L. K.; Tran, N.; Huang, A. D.; Nattestad, M.; Oshiro, K. T.; Chang, J. W.; Chavan, A.; Tsai, J. W.; Lin-Cereghino, J., The effect of alpha-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene* **2013**, *519* (2), 311-7.
32. Levene, P. A., The chemical nature of enzymes. *JACS* **1901**, *23* (7), 505-8.
33. Kavcscek, M.; Strazar, M.; Curk, T.; Natter, K.; Petrovic, U., Yeast as a cell factory: current state and perspectives. *Microb Cell Fact* **2015**, *14*, 94.

Chapter 5 NMR Directed Evolution of AlleyCat Series

5.0. Abstract

Directed evolution of proteins is a powerful method for optimizing desired traits, for example, enzymatic activity. However, the process is often lengthy and requires multiple steps. Computational analysis of proteins can help to narrow the possible sites for site-selected mutation; however, even this method is not perfect. We propose to use a method, employing ^{15}N -HSQC NMR, to identify possible residues in the vicinity of weakly binding analogs of a desired substrate. In this chapter we test an already known “hot spot” of the AlleyCat series, A128, which when mutated to threonine increases activity dramatically. We also target a new residue for mutation V108, as a possible method for creating a new Kemp elimination catalyst from the calmodulin (Cam) scaffold.

5.1. Introduction

5.1.1. Limitations of directed evolution and rational design

While the computational methods that inform rational design are extremely powerful, this approach also tends to produce proteins exhibiting only low activity.¹ Often subsequent mutations, designed without the aid of sophisticated algorithms, are needed to fulfill the true potential of the engineered protein scaffold. Using, for example, theozyme- and iterative-based computational design, which calculate the most stable protein conformations based on the

transition state of the substrate, various “hot spots” can be identified for mutation, limiting the rounds of mutagenesis needed.²⁻³ This method assumes that the scaffold, or at least the area around the active site, will retain its structural integrity after evolution.⁴ However, identifying potential mutations beyond the immediate active site is beyond the scope of computational program. Therefore, even in the age of computers, being able identify possible mutation sites for exploration is still very important.

Directed evolution utilizes a three-step process: first creating genetically diverse “libraries” with random mutagenesis and then identifying possible hits based on desired traits such as enzymatic activity.⁵ First, many variations of a gene are created using mutagenic PCR such as site-directed mutagenesis, random mutagenesis, error-prone PCR, and DNA shuffling (for a more in depth explanation of these methods, please refer to Chapter 1.2.1).⁶⁻⁹ Using an assay designed to screen for a desired trait, genes that contain positive hits continue onto the next stage while unfit variants are eliminated. The positive hit gene is then isolated and subjected to additional mutagenesis for further optimization.

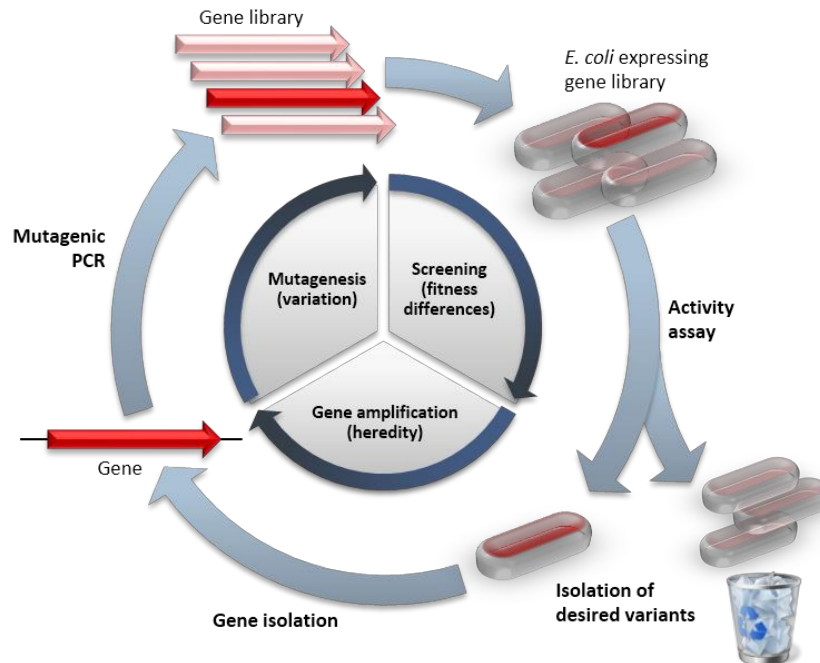


Figure 5.1. Summary of the steps of directed evolution.¹⁰

The issue with directed evolution is that the overall process is time consuming and requires vast libraries and a large number of samples.¹¹ There is also an issue with choosing the right form of mutagenesis. Error-prone PCR is good at introducing mutations into the entire gene, but due to the degeneracy of codons, this method can show bias towards some amino acids substantially reducing the number of variants. DNA shuffling requires that the different sequences have a high amount of homology between them, and DNA fragmentation is nonrandom for certain DNase enzymes.¹² Finally, site-directed mutagenesis can be a powerful tool but requires one to know the precise placement of the mutation needed.

Identifying new methods that take into account the dynamic changes in protein structures is key to optimizing directed evolution. NMR is an important structural technique in chemistry that allows for the identification of organic compounds, inorganic complexes, and proteins.¹³

Certain nuclei have a property called spin such that when placed in a magnetic field, these nuclei give rise to spin energy levels. NMR excites nuclei within the compound with a specific radio frequency and then the energy that is released when the nuclei falls back down in energy is measured. Depending on the nature of the nuclei and its local environment (i.e. neighboring atoms, hydrogen bonds, etc.) and the local magnetic field they contribute, the excited state will fall back to ground state releasing a photon of a specific frequency (i.e., energy). The frequencies are translated into chemical shifts, which are indicative of the electronic environment of the nucleus.

There are multiple types of NMR that can be used learn about a compound.¹³ The most common is ^1H NMR, which can give information such as the number of protons attached to the same atom and the number of neighboring hydrogens on adjacent carbons. Other nuclei that are able to be measured using this method include ^{13}C , ^{15}N , ^{19}F , and ^{31}P . Measuring the spin of these isotopes gives information about the type of functional groups present and their placement.

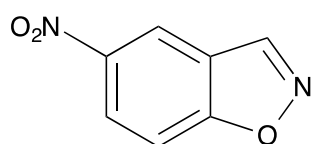
Small proteins are often measured using ^1H , ^{13}C , and ^{15}N to determine structure.¹⁴⁻¹⁵ By studying the correlations between two different types of spectra, a lot of insight is gained based on how different isotopes interact with each other. One of the most popular methods is ^1H - ^{15}N HSQC, which allows for the measurement of the peptide bonds in proteins and tells us how amide protons interact with one another in the back bone and side chains.

HSQC takes measurements of positions and energies of proteins in their average current state and if one were to take a HSQC of a protein when it is bound to the substrate (or a substrate analog), and then take another HSQC of the protein without anything bound to it, the resulting

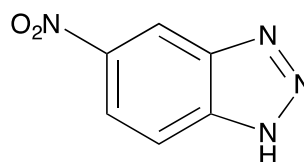
difference in chemical shifts between the two structures can provide clues to which residues are involved with substrate binding and activation. This could be potentially useful for identifying residues outside the active site that computational analysis often can't interpret. The hypothesis is that residues with larger chemical shifts between bound and unbound states could be appropriate places to begin directed evolution of the protein.

5.1.2. "Hot" and "Cold" Spots of AlleyCat

Previously, Korendovych lab has successfully carried out several projects on the re-design of Cam into an allosterically regulated catalyst for Kemp elimination.^{14, 16-17} Since the initial mutation, this new catalyst (named AlleyCat) has undergone seven rounds of mutation using various methods of directed evolution to Kemp elimination optimize activity. We characterized all rounds of AlleyCat extensively using ¹⁵N-¹H HSQC NMR with and without a substrate analog bound. Since the substrate of Kemp elimination, 5-nitro-1,2-benzisoxazole, would rapidly undergo catalyzed elimination, we used a similar structure, 6-nitro-1H-benzotriazole, as an inhibitor that would bind to the same active site but not react (Figure 5.2).



5-nitro-1,2-benzisoxazole



6-nitro-1H-benzotriazole

Figure 5.2. Structures of 5-nitro-1,2-benzisoxazole (left) and Kemp elimination antagonist 6-nitro-1H-benzotriazole (right), which is used during NMR titrations to determine the chemical shifts of residues in the AlleyCat series.

We take a HSQC of AlleyCat7 without any substrate added in 20 mM Hepes, 100 mM NaCl, pH 7.0 followed by adding two equivalents of 6-nitro-1H-benzotriazole and taking another HSQC. We observed differences in the N-H chemical shifts of each amino acid during the inhibitor titration and found that certain locations on the protein were ‘hot spots’, that is, residues that exhibit particularly drastic chemical shifts changes with and without the inhibitor, which can be calculated using Equation 5.1. The shift perturbation data are summarized in Figure 5.3. As seen in the graph, one residue (i.e., A128) was found to be an especially ‘hot spot’ in comparison to other amino acids in the sequence.

Equation 5.1. Equation used to determine the change in chemical shift between analog bound and nonbound AlleyCat7 when determining hot and cold spots.

$$\Delta\text{chemical shift of residue} = \sqrt{\frac{(dH)^2 + \left(\frac{dN}{5}\right)^2}{2}}$$

where $dH = {}^1H$ shift of analog bound residue – 1H shift of nonbound residue

and $dN = {}^{15}N$ shift of analog bound residue = ${}^{15}N$ shift of nonbound residue

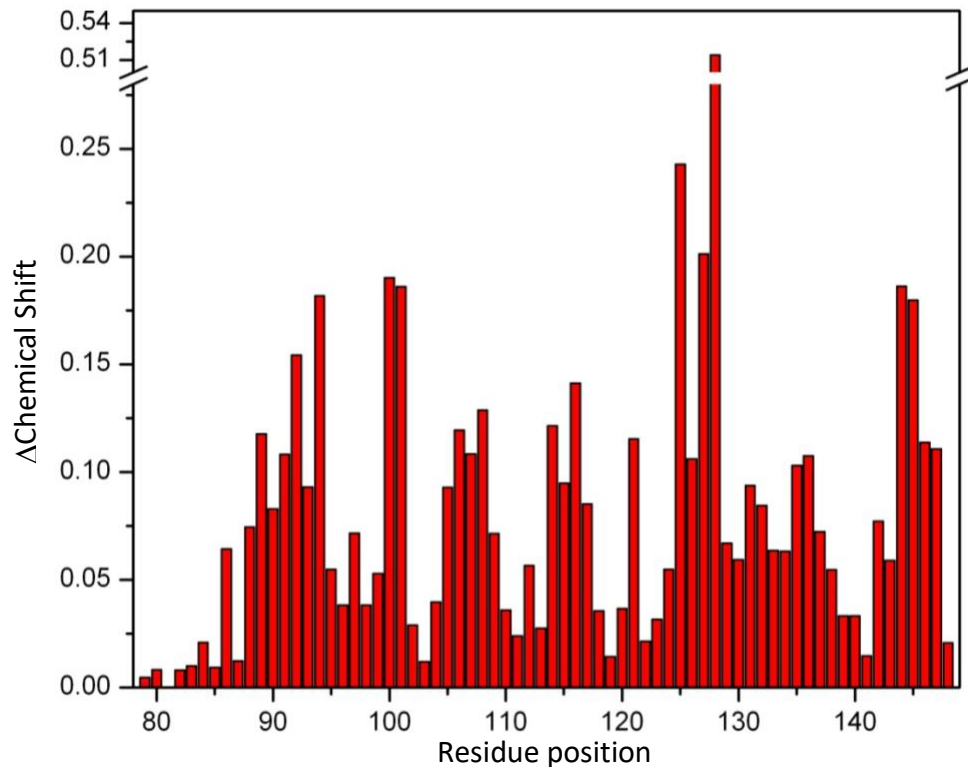


Figure 5.3. ¹⁵N-¹H HSQC shift data for all residues in AlleyCat7. NMR data was collected and analyzed by I. Korendovych and O. Makhlynets.

5.2. Analysis of the importance of A128T mutation in AlleyCat

To determine whether NMR ‘hot spots’ is a viable method for identifying residues for mutation, the Kemp elimination activities for the AlleyCat series of proteins were compared before and after the mutation of A128. It was the fifth round of mutation in the AlleyCat series that introduced the A128T mutation resulting in an almost a two-fold increase in activity over its predecessor, AlleyCat4. In principle, introducing that same mutation to previous generation in the AlleyCat series should demonstrate a similar jump in activity. Conversely, removing the mutation from a later generation, that is, AlleyCat6 and AlleyCat7, should show a significant

decrease in activity. These specific mutants were prepared, purified, and characterized in the Kemp elimination assay to determine the kinetic activity of each protein. The kinetic data are shown in Figure 5.4.

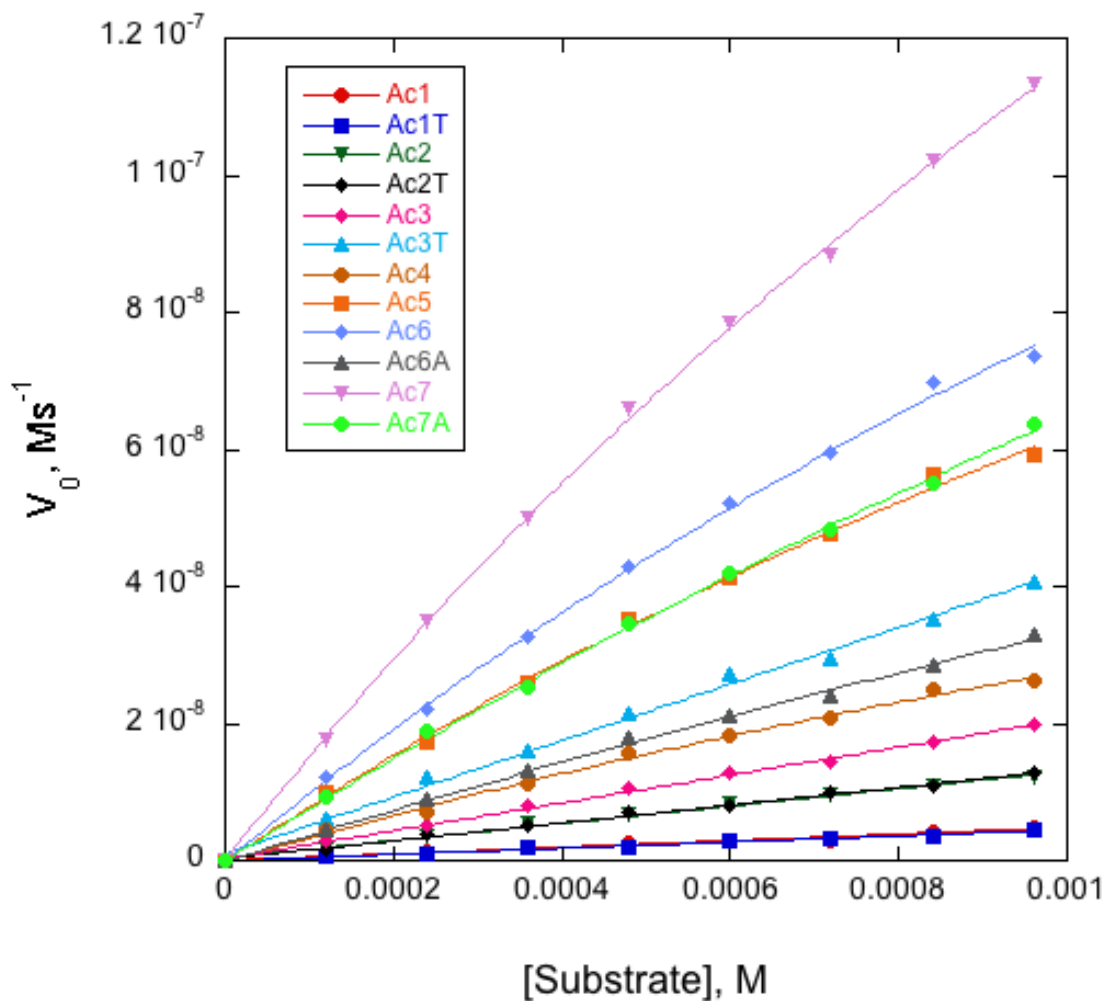


Figure 5.4. Catalytic efficiency of Kemp elimination catalyzed by the AlleyCat series with and without the A128T mutation (0.2 μ M) in 20 mM Hepes buffer, pH 7.0 with 100 mM NaCl and 10 mM of CaCl_2 . The values for k_{cat}/K_M are listed in Table 5.1. Ac5 is the equivalent of Ac4T. (MT4100)

Figure 5.5 shows the Michaelis-Menten graph of Kemp elimination catalyzed by AlleyCat0 with and without the A128T mutation. It is remarkable the even introducing this mutation to

AlleyCat0, which only has the initial F92E mutation present and exhibits only a low Kemp elimination activity, shows a significant increase after the A128T mutation.

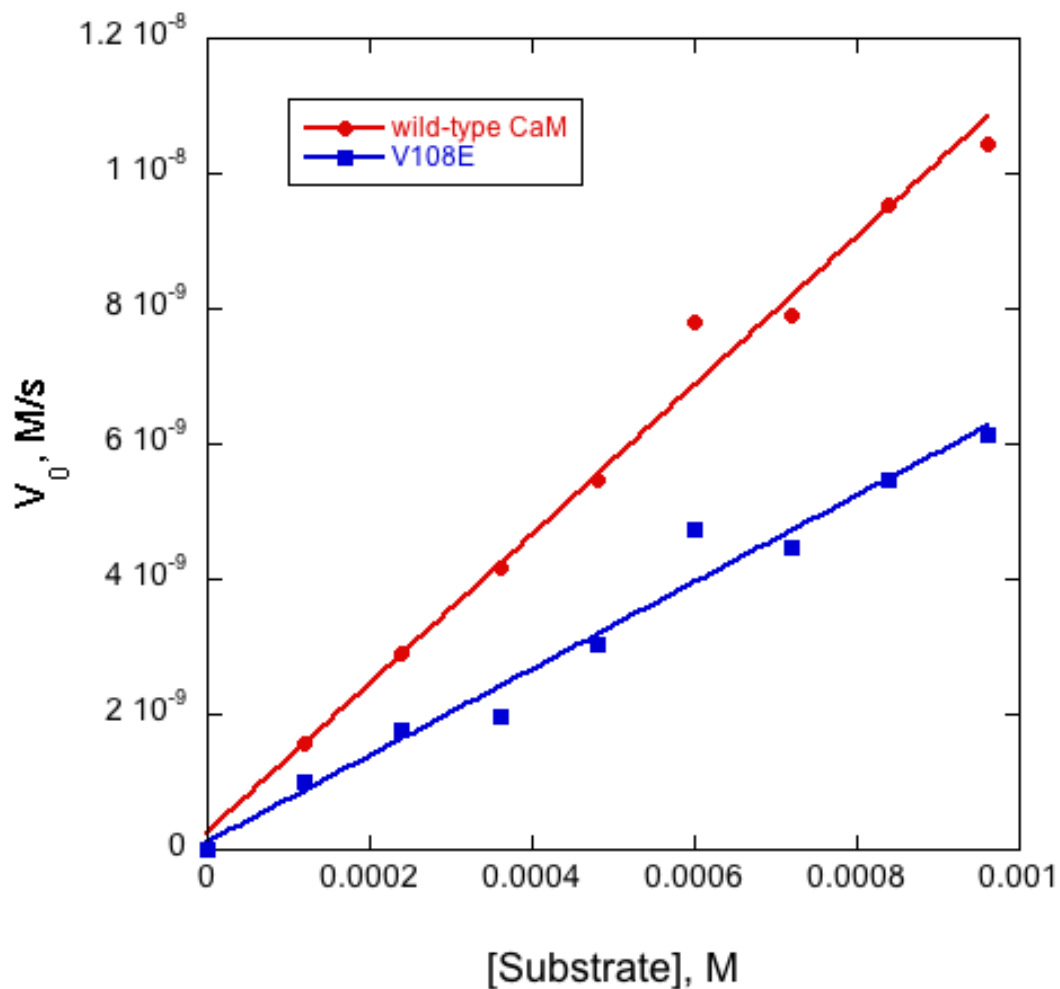


Figure 5.5. Catalytic efficiency of Kemp elimination catalyzed by AlleyCat0 with and without the A128T mutation (20 μ M) in 20 mM Hepes buffer, pH 7.0 with 100 mM NaCl and 10 mM of CaCl_2 . The values for k_{cat}/K_M are listed in Table 5.1. (MT4128)

Table 5.1 summarizes the Kemp elimination kinetic data of AlleyCat mutants with and without Thr in 128 position. The Michaelis-Menten kinetics of the Kemp elimination catalyzed by

each of these proteins fits our expectation based on the ‘hot spots’ hypothesis. Introducing the A128T mutation to Ac0 and Ac3 gives the mutants (Ac0T and Ac3T) that exhibit increased catalytic activity. Conversely, removing the mutation from Ac6 and Ac7 gives a mutant protein (Ac6A and Ac7A) that shows a significant decrease in catalytic activity.

Table 5.1. Summarized data of AlleyCat mutants with and without T in 128 position. Ac0 and Ac0T were at a concentration of 20 μ M; Ac6, Ac6A, Ac7, and Ac7A were at a concentration of 0.2 μ M; all other proteins were at a concentration of 4.0 μ M.

Protein	k_{cat}/K_M ($\text{M}^{-1}\text{s}^{-1}$)	Protein	k_{cat}/K_M ($\text{M}^{-1}\text{s}^{-1}$)
Ac0	1.42	Ac0T	6
Ac1	33 ± 11	Ac1T	30 ± 12
Ac2	84 ± 21	Ac2T	91 ± 30
Ac3	133 ± 59	Ac3T	251 ± 61
Ac4	190 ± 60	Ac5	387 ± 106
Ac6	512 ± 107	Ac6A	190 ± 103
Ac7	786 ± 97	Ac7A	390 ± 124

5.3. Analysis of V108 position as a site for Kemp elimination catalysis

The A128T mutation test, a mutation that been previously done in the AlleyCat series and was also indicated by changes in NMR shift upon inhibitor binding, gives some confidence in using the method. We therefore moved on to exploring mutations at other residues that had not been

previously mutated in AlleyCat series for which the NMR titration method data obtained for Cam and the Ac0 protein indicated a 'hot spot'.^{6,7} Figure 5.6 shows the relative 'hot spots' of calmodulin by examining the NMR chemical shifts of ¹⁵N HSQC with and without the 5-nitro-1H-benzotriazole. Two residues stand out as having particularly large chemical shift values between the bound and non-bound protein: F92 and V108. In the AlleyCat series, F92E is the mutation responsible for the subsequent Kemp elimination. It stands to reason that V108 might also be relevant for mutation since it shares similar chemical shift perturbations and is in a similar position.

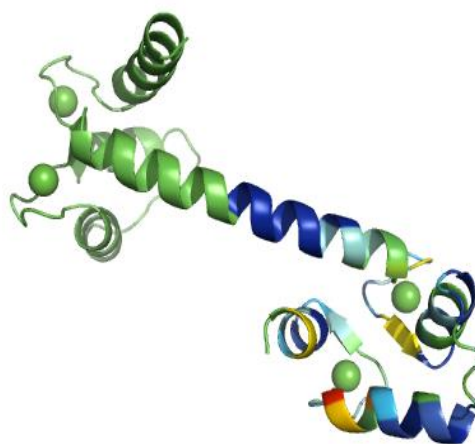


Figure 5.6. A diagram of the relative “hot” spots of calmodulin as determined by NMR titration. Red, orange, and yellow colors signify “hot spots” with high chemical shift perturbations and cyan and blue colors signify “cold spots” with small chemical shift perturbations (pdb 1c1l).¹⁸

We mutated V108 to glutamate in wild-type Cam and Ac0 to determine if the mutations would allow for catalysis much like F92E had. We reasoned that glutamate was the original mutation in AlleyCat that gave it catalytic activity and that due to the hydrophobic nature of the

active site pocket, it would work well for the Kemp elimination reaction. The V108E mutation actually *decreased* the Kemp elimination activity of the Ac0-V108E mutant relative to Ac0 (Figure 5.7, 5.8, and Table 5.2) and of the Ac1-V108E mutant relative to Cam (Figure 5.8). While this position is unsuitable for creating a new catalyst, this valine seems to play a minor role in the stability of the transition state during catalysis. While the failure of V108E mutation does not discount our hypothesis completely, it does show that care must be taken when using this NMR-directed evolution and that it would work best in conjunction with other techniques.

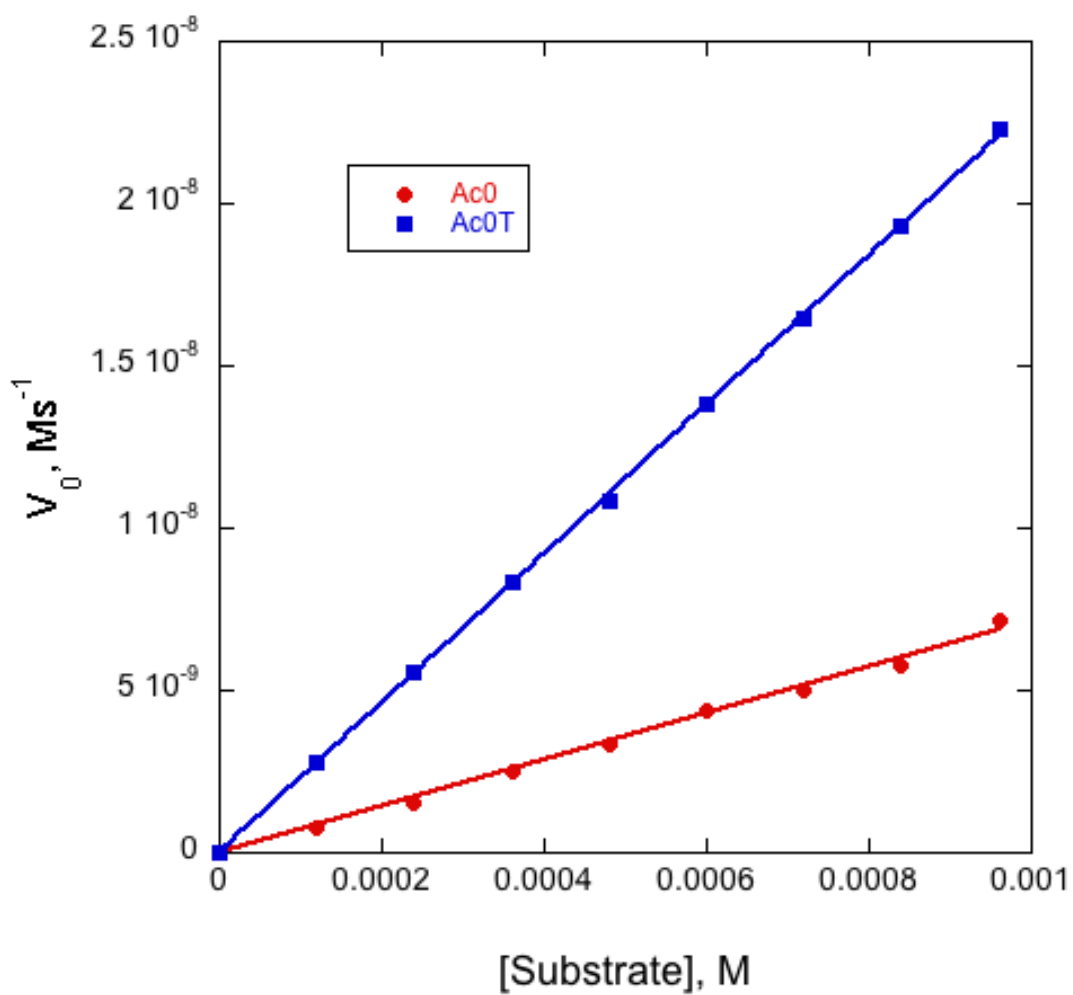


Figure 5.7. Catalytic efficiency of Kemp elimination catalyzed AlleyCat0 (20 μM) with and without the V108E mutation in 20 mM Hepes buffer pH 7.0 with 100 mM NaCl and 10 mM of CaCl_2 . The values for k_{cat}/K_M are listed in Table 5.2. (MT5051)

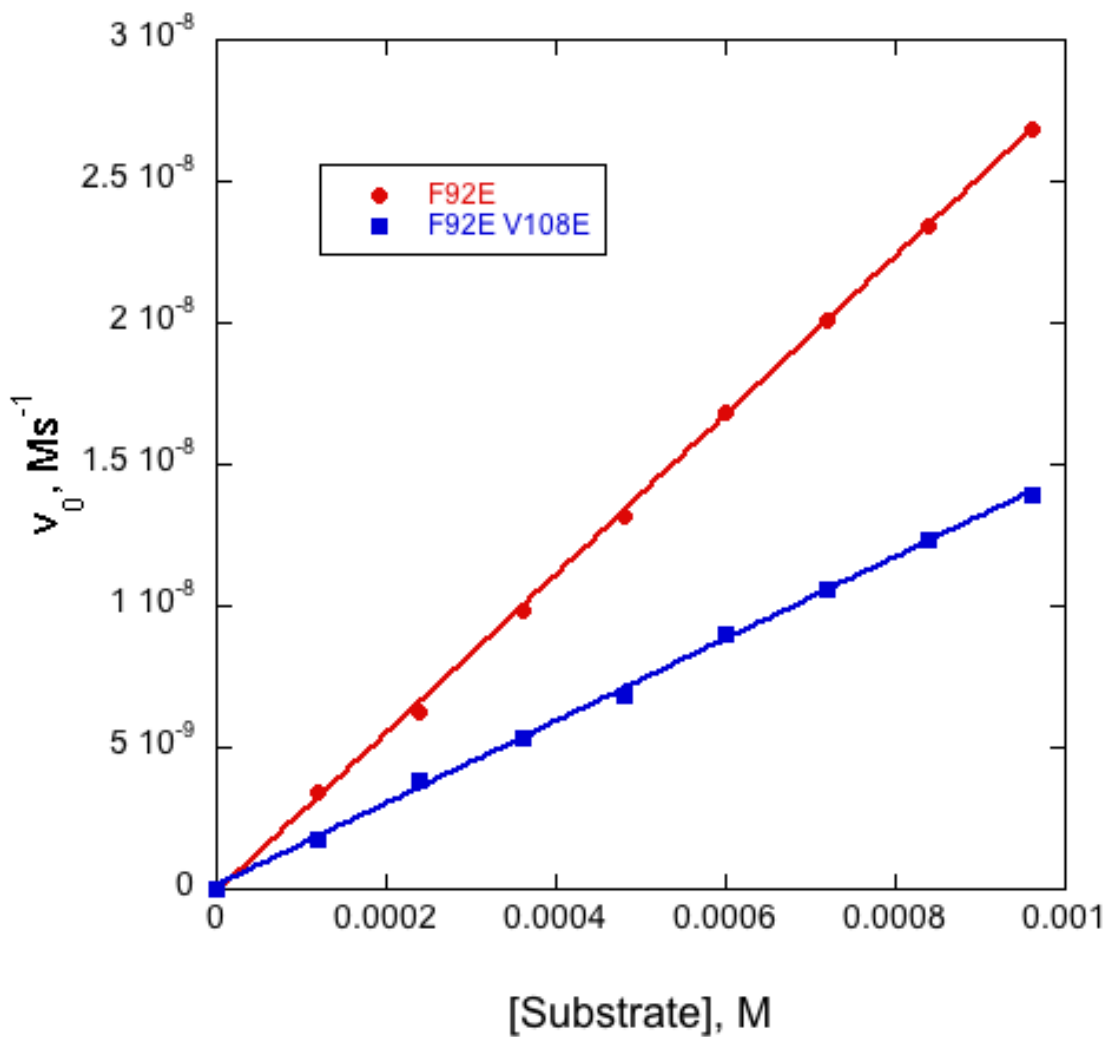


Figure 5.8. Catalytic efficiency of Kemp elimination catalyzed wild-type Cam (20 μM) with and without the V108E mutation in 20 mM HEPES buffer, pH 7.0 with 100 mM NaCl and 10 mM of CaCl_2 . The values for k_{cat}/K_M are listed in Table 5.2. (MT5142)

Table 5.2. Summarized data of V108E mutants with and without Glu in 108 position. All proteins were at 20 μ M concentration.

Protein	Mutations	k_{cat}/K_M ($M^{-1}s^{-1}$)
Wild-type Cam	none	0.22
V108E	V108E	0.13
AlleyCat0	F92E	1.42
Ac0 V108E	F92E V108E	0.76

5.4. Conclusions

AlleyCat is a small protein that has been well characterized by NMR in the presence and absence of substrate and inhibitors. The presence of inhibitor leads to perturbations in the chemical shift that could be indicative of possible spots for mutagenesis, especially for residues outside the active site. To test this theory, we focused on the residue exhibiting the largest chemical shift change upon binding inhibitor (i.e., A128) and mutated it into Thr and back to Ala in all generations of AlleyCat. The presence of Thr increased the Kemp elimination activity of the Ac0T and Ac3T mutants approximately two-fold. However, when this technique was applied to another identified position, V108E, the protein had no Kemp elimination activity. This project is currently being continued in lab and promising results have been found using another ‘hot spot’ that has been identified. This new protein is almost twice as active and leads further credence to our hypothesis. The next step is to examine another protein, the computationally designed Kemp eliminase Baker’s KE07,¹⁹ and use our method to attempt to increase its catalytic activity.

5.5. Experimental

5.5.1. Site-Directed Mutagenesis

The appropriate mutations were made using a standard mutagenesis protocol for *pfu* high-fidelity DNA Polymerase, which was provided by the manufacturer (NEB). After PCR, the template was digested with *DpnI* restriction enzyme (NEB) and transformed into *E. coli* XL-10 cells. DNA was extracted from colonies using EZ-10 Spin Column Plasmid DNA kit and mutated sequences were confirmed by Genescript using Sanger sequencing technique.

5.5.2. Purification of AlleyCat series

The pEXP5-NT vectors containing the genes of interest were transformed into *E. coli* BL21(DE3) pLysS cells and expressed using LB broth. After cells reached an OD₆₀₀ of ~0.5, IPTG was added to a final concentration of 0.5 mM and the temperature was lowered to 30°C. After four hours, the cells were collected by centrifugation and stored at -80°C until needed.

Cell pellet was resuspended in lysis buffer containing 25 mM Hepes, 20 mM imidazole, 0.5 mM PMSF, 10 mM CaCl₂, and 300 mM NaCl (pH 8). Cells were lysed by sonication, and then the crude cell lysate was centrifuged at 20,000xg for 1 h. The supernatant was applied to a Ni-NTA column and washed with lysis buffer (without PMSF). The protein was then eluted with buffer containing 25 mM Hepes, 250 mM imidazole, 10 mM CaCl₂, and 300 mM NaCl (pH 8). The sample was exchanged into buffer containing 25 mM Hepes, 10 mM CaCl₂, 100 mM NaCl (pH 7) by applying it onto a Bio-Rad 10 DG desalting column.

The purity of the protein was established by SDS-PAGE using a 15% acrylamide gel with Unstained Precision Protein Plus ladder (BioRad). Protein concentration was determined using UV-Vis spectroscopy where $\epsilon_{280} = 5445 \text{ cm}^{-1}\text{M}^{-1}$.

5.5.3. Kinetics for Kemp elimination reaction

Kinetic measurements were done with either a Thermo Lab-systems Multiskan Spectrum Platerreader or a BioTek Eon Micro Platerreader monitoring the absorbance at 380 nm at 22°C using at least three independent measurements. The final protein concentration (determined by UV-Vis spectroscopy using $\epsilon_{280} = 5445 \text{ M}^{-1} \text{ cm}^{-1}$) for kinetic measurements was 0.25 μM . Final substrate concentrations ranged from 0.12 to 0.96 mM. Kinetic parameters (k_{cat} and K_{M}) were obtained by fitting the data to the Michaelis-Menten equation: $[v_0 = k_{\text{cat}}[E]_0[S]_0/(K_{\text{M}} + [S]_0)]$ using Kaleidagraph 4.5 (Synergy Software) and dividing the calculated k_{cat} and K_{M} values, or by fitting the data to a linear slope to find the $k_{\text{cat}}/K_{\text{M}}$.

5.6. References

1. Brustad, E. M.; Arnold, F. H., Optimizing non-natural protein function with directed evolution. *Curr Opin Chem Biol* **2011**, *15* (2), 201-10.
2. Tantillo, D. J. C., J; Houk, K.N., Theozymes and compuzymes: theoretical models for biological catalysis. *Curr Opin Chem Biol* **1998**, *2*, 743-50.

3. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* **2012**, *109* (10), 3790-5.
4. Linder, M., Computational Enzyme Design: Advances, hurdles and possible ways forward. *Comput Struct Biotechnol J* **2012**, *2*, e201209009.
5. Lutz, S., Beyond directed evolution--semi-rational protein engineering and design. *Curr Opin Biotechnol* **2010**, *21* (6), 734-43.
6. Reeves, H. Recombinant DNA Technology site directed mutagenesis genetics vs. reverse genetics gene expression in bacteria and viruses gene expression in yeast genetic. <http://slideplayer.com/slide/6199958/>.
7. Porter, J. L.; Boon, P. L.; Murray, T. P.; Huber, T.; Collyer, C. A.; Ollis, D. L., Directed evolution of new and improved enzyme functions using an evolutionary intermediate and multidirectional search. *ACS Chem Biol* **2015**, *10* (2), 611-21.
8. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat Rev Genet* **2015**, *16* (7), 379-94.
9. Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L., Directed evolution of enzymes: Library screening strategies. *IUBMB Life* **2009**, *61* (3), 222-8.
10. Shafee, T. Evolvability of a viral protease: experimental evolution of catalysis, robustness, and specificity. University of Cambridge, 2014.
11. Reetz, M. T.; Kahakeaw, D.; Lohmer, R., Addressing the numbers problem in directed evolution. *ChemBiochem* **2008**, *9* (11), 1797-804.

12. Joern, J. M.; Meinhold, P.; Arnold, F. H., Analysis of shuffled gene libraries. *J Mol Biol* **2002**, *316* (3), 643-56.
13. Silverstein, R. M. W., F.X; Kiemle, D.J, *Spectrometric identification of organic compounds*. 7 ed.; Wiley and Sons: Hoboken, NJ, 2005.
14. Wüthrich, K., The way to NMR structures of proteins. *Nature* **2001**, *8* (11), 923-5.
15. Ouyang, H. V., Hans J., Metal ion binding to calmodulin: NMR and fluorescence studies. *BioMetals* **1998**, *11*, 213-22.
16. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc Natl Acad Sci U S A* **2011**, *108* (17), 6823-7.
17. Moroz, O. V.; Moroz, Y. S.; Wu, Y.; Olsen, A. B.; Cheng, H.; Mack, K. L.; McLaughlin, J. M.; Raymond, E. A.; Zhezherya, K.; Roder, H.; Korendovych, I. V., A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew Chem Int Ed Engl* **2013**, *52* (24), 6246-9.
18. Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quioco, F. A., Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol* **1992**, *228* (4), 1177-92.
19. Alexandrova, A. N. R., D; Baker, D; Jorgensen, W.L, Catalytic Mechanism and Performance of Computationally Designed Enzymes for Kemp Elimination. *J Am Chem Soc* **2008**, *130*, 15907-15.

The following chapter was co-authored with M. Dolan of Korendovych lab.

M. Dolan performed GC analysis.

Chapter 6 Asymmetric Hydrogenation of Ketones Using Catalytic Ruthenium- and Iridium-Peptide Aggregates

6.0. Abstract

Chiral catalysts are one of the most powerful tools at a synthetic organic chemist's disposal. They can impart the ability not only to speed reactions, but also impart enantioselectivity onto products while being present in less than stoichiometric quantity as the reagent. Despite these marked advantages, chiral catalysts also have several distinct disadvantages. Organometallic catalysts in particular are a problem, because while they work very effectively, they are also expensive, and their synthesis can be time consuming. However, chiral amino acids are abundant, very cheap, easy to use, and share many characteristics with other popular organic ligands. Furthermore, short peptides, for example the phenylalanine dimer, aggregate in solution. Aggregation imparts structure and with a chiral peptide, it imparts a chiral environment for catalysis. We propose a new catalyst system which uses a tripeptide as a ligand. The first amino acid, 2,3-diaminopropionic acid, will complex transition metals such as Ru^{2+} and Ir^{3+} that are often used as chiral catalysts. The two phenylalanine subunits will branch off the C-terminal of the amino acid-ligand and impart aggregating properties onto the new catalyst. In addition to creating a chiral environment for catalysis, the design will enable the

catalyst to work in water and allow for easy separation using ultracentrifugation or by binding the catalyst to other surfaces.

6.1. Introduction

6.1.1. Organometallic Catalysts

Catalysts are one of the most important tools in both nature (in the form of enzymes) and in the laboratory. Like enzyme catalysts discussed in Chapter 1, chiral chemical catalysts in the laboratory are able to lower the activation energy of reactions that would normally be difficult to overcome and impart enantioselectivity onto reactions and products. Coordinated organometallic catalysts use a large selection of metal ions, ligands, and substrates to achieve a wide variety of different reactions, making them one of the most important tools at a synthetic chemist's disposal.

Most current available catalysts have distinct disadvantages. The catalyst and their precursors can cost a significant amount of money and be difficult to use in scale up. Furthermore, the ligands, which are responsible for reactivity and selectivity, can be difficult and time consuming to make. While catalysts are in principle reusable and can be recollected from reactions after use, this step can be time consuming and impractical in industrial settings at large scale. Therefore, creating a catalyst that is both cheap to make and easily collectible is advantageous for a variety of reasons. While expensive metals are necessary, ligands can impart these properties unto a catalyst.

6.1.2. Chiral Ligands

The ability to synthesize R- or S-chiral compounds selectively is one of the biggest problems facing synthetic chemists today. To solve this problem, many chemists use metal catalysts attached to specialized ligands, known as privileged ligands (Figure 6.1).¹⁻² In general, most privileged ligands possess rigid structures and contain multiple oxygen, nitrogen, and phosphorous functional groups that allow them to coordinate with metal ions.

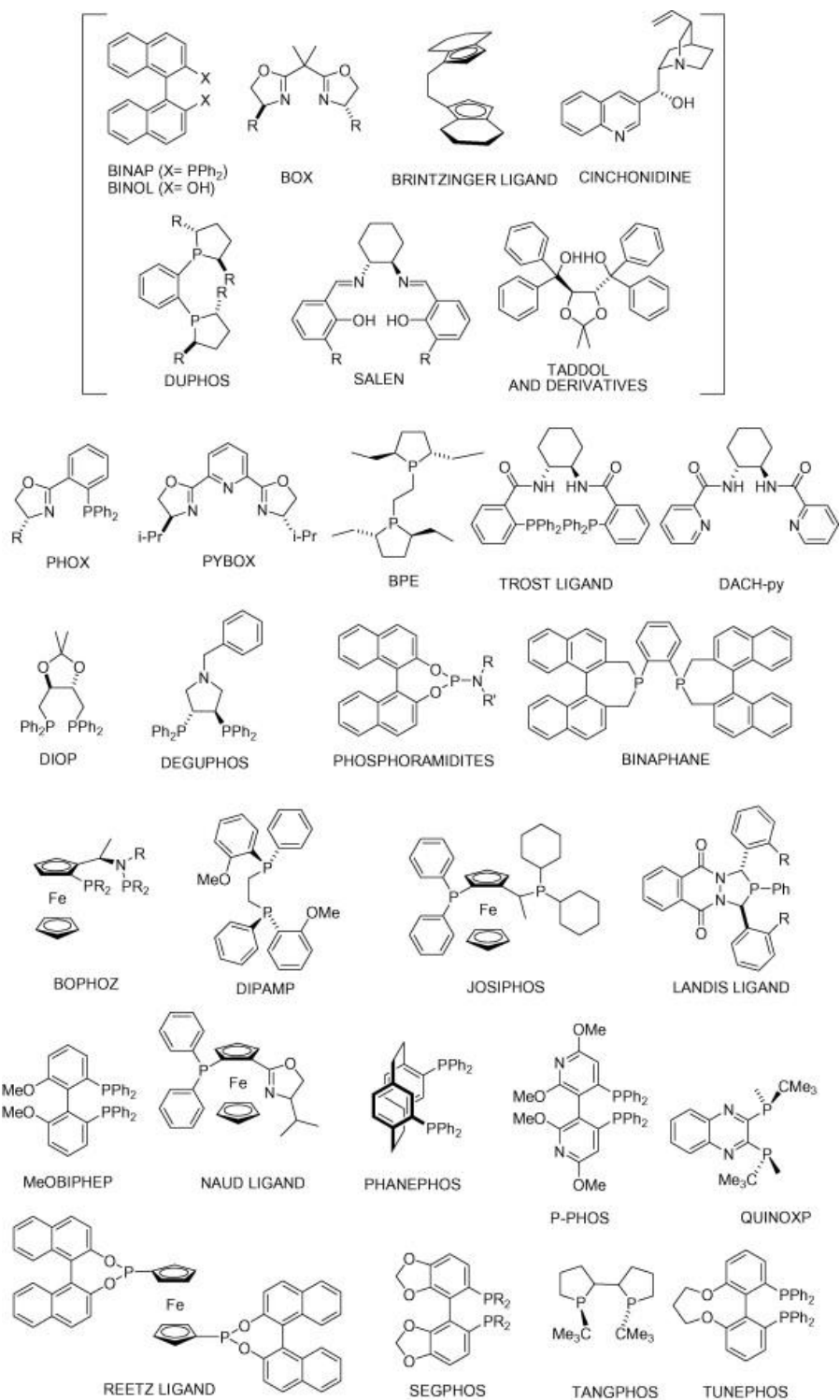


Figure 6.1. Examples of privileged ligands.³

Amino acid derived ligands are a popular alternative to complex ligands.⁴ Amino acids are well known for their ability to coordinate to metals already (i.e. metalloproteins) and contain carboxyl and amino groups like most privileged ligands do. Furthermore, amino acids are also naturally chiral, readily available, and rather cheap.

6.1.3. Phenylalanine Aggregation

Phenylalanine chains are well documented for their tendency to form amyloid structures in nature.⁵ With as little as two phenylalanine subunits, amyloids are formed between the aromatic amino acids to create complex amyloid structures (Figure 6.2).⁶⁻⁸

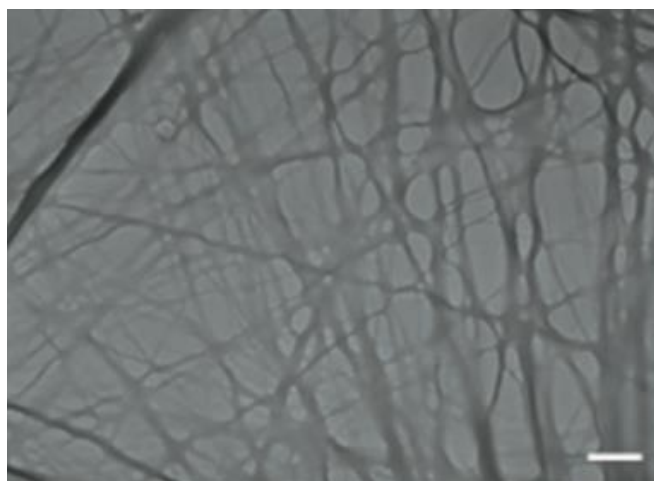


Figure 6.2. TEM image of phenylalanine fibrils.⁵

By using small peptide chains as ligands, we are proposing a system of chiral ligands that will not only be cheaper and easier to make than most popularly used ligands but will also solve the problem of catalyst recovery. Amino acids, for example threonine, lysine, and serine, share many of the functional groups that more complex ligands have. Therefore, amino acids should be able to act as suitable replacements for many of the commonly used ligands.^{4,9} While one end

of the peptide complexes with the metal ion to form the catalytic portion of the small molecule, the phenylalanines (Phe, F) will be attached at the other end of the peptides. Therefore, by attaching the diphenylalanine chain to the peptide (Figure 6.3), the catalyst will be able to aggregate and could be filtered out of solution once the reaction of interest complete.

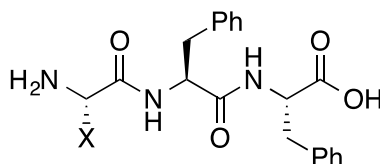


Figure 6.3. Example of new ligand design. X represents variable amino acid side chain.

6.1.4. Noyori catalysts

Ryōji Noyori won the Nobel Prize in Chemistry in 2001 for the development of an asymmetric hydrogenation catalyst that utilized Ru²⁺ and Ir³⁺ chiral centers.⁹⁻¹¹ The catalyst can carry out selective asymmetric hydrogenation and reductions of ketones, aldehydes, and imines.^{10, 12-13} The catalysts contain a Ru²⁺ or Ir³⁺ chiral center with a N-tosylated diphenylethylenediamine (pTsDPEN) at two of the metal ion's coordination sites (Figure 6.4). The pTsDPEN ligand can have either have the R,R-stereochemistry or the S,S-stereochemistry, which imparts enantioselectivity unto the final product of the reaction. One of the benefits from this catalyst, aside from the high percent enantiomeric excess of the products, is its ability to catalyze reactions in water. It can be considered a 'green catalyst.'¹⁴⁻¹⁵ The similarity of the pTsDPEN functional groups to those found in amino acids, as well as the capability for this reaction to be

run in water, makes the Noyori ruthenium catalysts an ideal starting point for testing our design hypothesis.

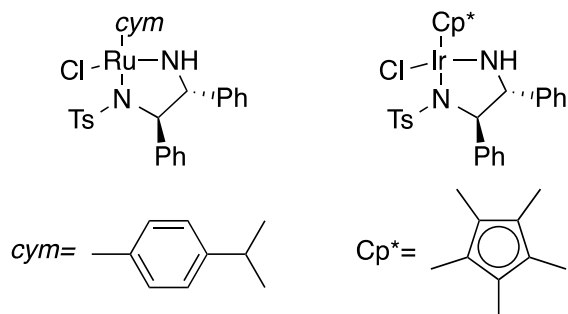


Figure 6.4. Examples of common Noyori catalysts with ruthenium and iridium metal centers.

While peptide-metal complexes have been explored as catalyst systems in the past, our method targets not only reactivity, but practicality.⁹ When applied, our catalytic constructs will solve a major problem in organic synthesis in preventing waste and the time-consuming step of recollecting and recycling catalysts. In the future, we aim for a catalyst that can simply be collected by centrifugation of a reaction mixture. In principle, the peptide catalysts can be bound to a membrane or filter and a substrate can simply be ‘pushed through’.¹⁶ Such a dream process would eliminate the need for extractions and provide a catalytic system that is easy to clean and reuse. We hope that our system can be applied effectively not only to the Noyori catalysts, but other common catalysts as well.

6.2. Design of Tripeptide Ruthenium Catalysts

6.2.1. NMR

We chose 2,3-diaminopropionic acid (Dap) as a base amino acid from which to build upon our ligand. Dap (Figure 6.5.a) is an “unnatural amino acid” that is not found in normal genetic code but instead found in secondary metabolites such as zwittermicin A and tuberactinomycin^{34,35}. It is commercially available and structurally similar to ethylenediamine, which is a backbone for TsDPEN. To create our ligand, we purchased Dap with a Boc-protecting group on the main chain, added a Tosyl group to the side chain amine, and then removed the Boc-protecting group to form L-Dap(Ts)-OH (Figure 6.5.b). The ligand was dissolved in water and coupled with 0.4 equivalence of either $[\text{RuCl}_2(\text{p-cymene})]_2$ or $[\text{Cp}^*\text{IrCl}_2]_2$ to form two new metal complexes (Figure 6.5.c).

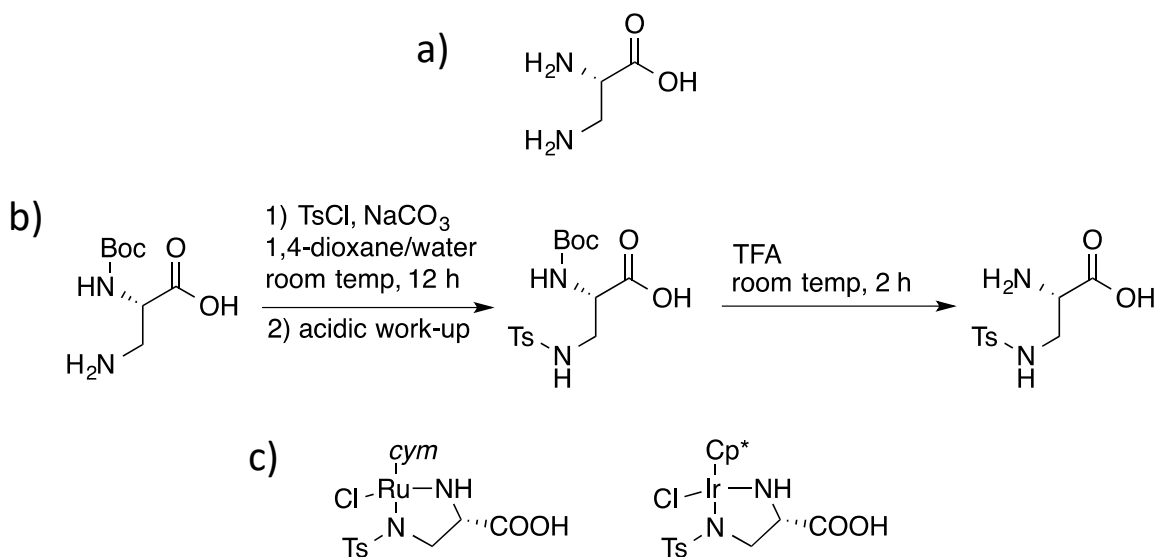


Figure 6.5. a) Structure of L-2,3-diaminopropionic acid. b) Synthesis of L-Dap(Ts)-OH from L-(Boc)Dap-OH.

c) Proposed structure of Ru^{2+} and Ir^{3+} after coupling with L-Dap(Ts)-OH

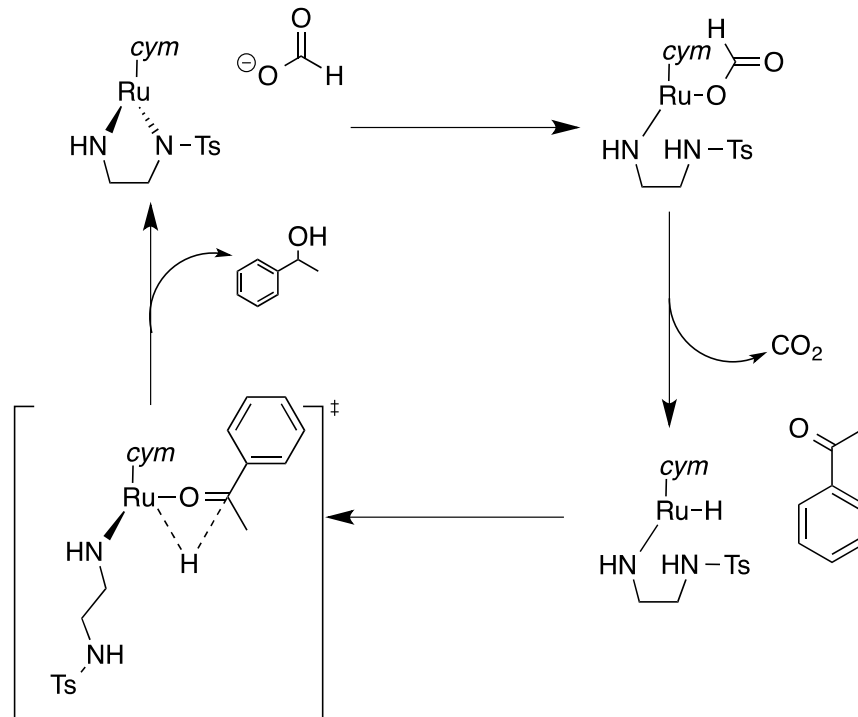


Figure 6.6. Mechanism of acetophenone hydrogenation and subsequent catalyst regeneration using sodium formate as a proton source. Iridium can be used in place of ruthenium.

Acetophenone and sodium formate were then added to the *in situ* generated catalysts and stirred for one hour at 40°C (Figure 6.6). Afterwards, the reaction was quickly worked up and the reduction of acetophenone to 1-phenylethanol was determined by H¹ NMR. In addition to multiplets in the aromatic region, acetophenone has an observable singlet at 2.58 ppm, while 1-phenylethanol has a quartet at 4.88 ppm and a doublet at 1.48 ppm (Figure 6.7).

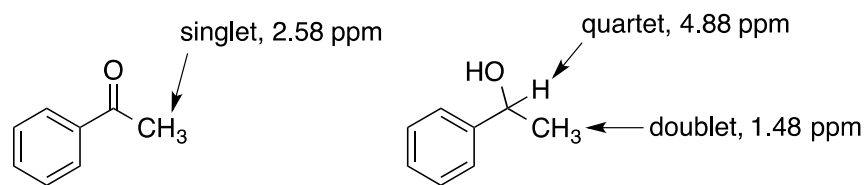


Figure 6.7. Identifying ¹H NMR chemical shifts of acetophenone (left) and 1-phenylethanol (right).

This latter resonances were used to determine the extent to which reaction occurred.³³ To get an estimate to the percent reduction that took place, we compare the integrated area of the peak due to the methyl group in the phenethyl alcohol (a doublet at 1.48 ppm) to the total integrated area for peaks due to CH₃ groups in the acetophenone (a singlet at 2.58 ppm) and phenethyl alcohol times 100. The quartet at 4.88 ppm was set to an integrated area of 1.0 hydrogen to confirm a 3:1 ratio of areas due to the methyl group doublet and methine proton quartet as required for the phenethyl alcohol product. While this does not give a true quantitative yield, the use of an internal standard is necessary for that, it gives an good estimate whether our catalyst is capable of reducing acetophenone. As a control, the metal precursors were coupled with pTsDPEN and the reaction with acetophenone carried out under the same conditions. The NMR spectral data are shown in Figures 6.8-6.11. We find the following results:

- Ru²⁺ coupled with (R,R)-pTsDPEN: 28% conversion of acetophenone to 1-phenylethanol
- Ru²⁺ coupled with L-Dap(Ts)-OH: no conversion of acetophenone to 1-phenylethanol
- Ir³⁺ coupled with (R,R)-pTsDPEN: 37% conversion of acetophenone to 1-phenylethanol
- Ir³⁺ coupled with L-Dap(Ts)-OH: 5% conversion of acetophenone to 1-phenylethanol.

While the yield is low, ¹H NMR shows that iridium coupled with L-Dap(Ts)-OH will catalyze the reduction of acetophenone (Figure 6.8-6.11).

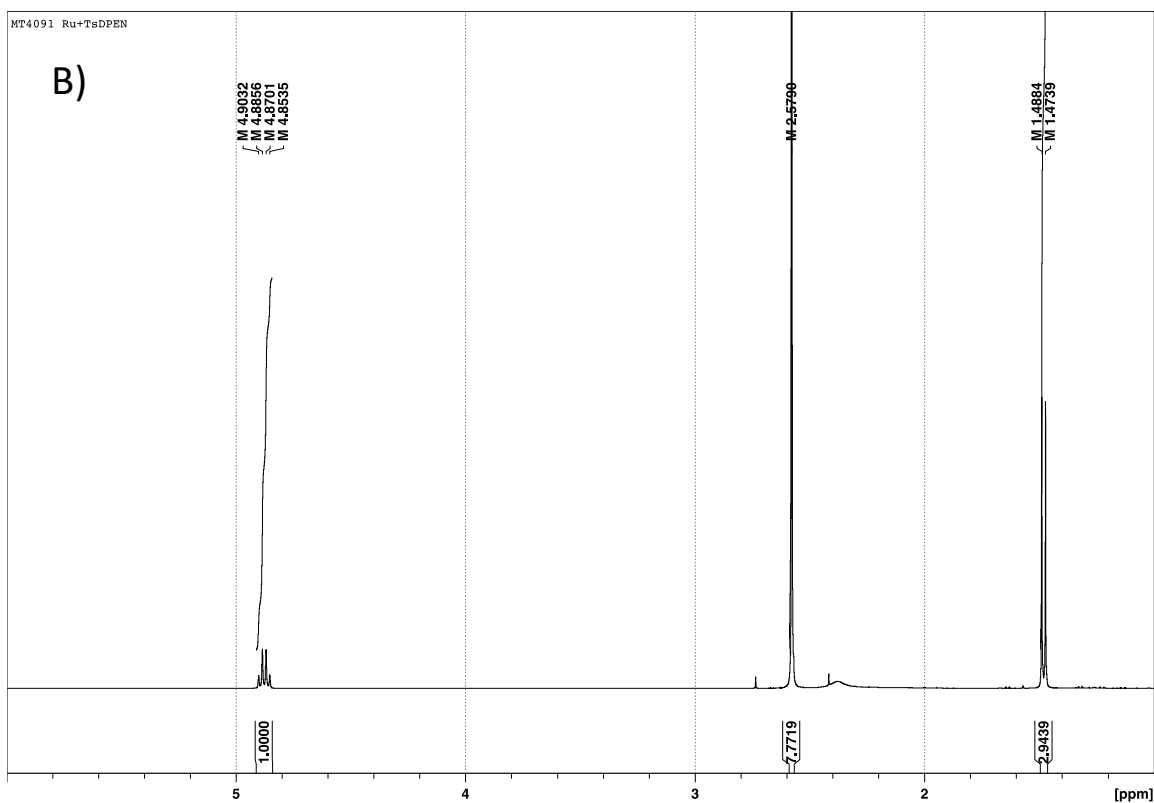
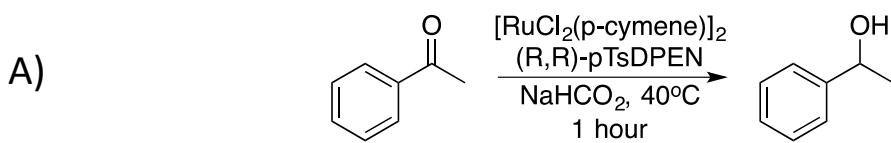


Figure 6.8. A) Synthesis of 1-phenylethanol from acetophenone using Ru^{2+} coupled with (R,R)-pTsDPEN. B) ^1H NMR spectrum of acetophenone reduction after one hour (the aromatic region is not shown). There was a 28% conversion of acetophenone to 1-phenylethanol.

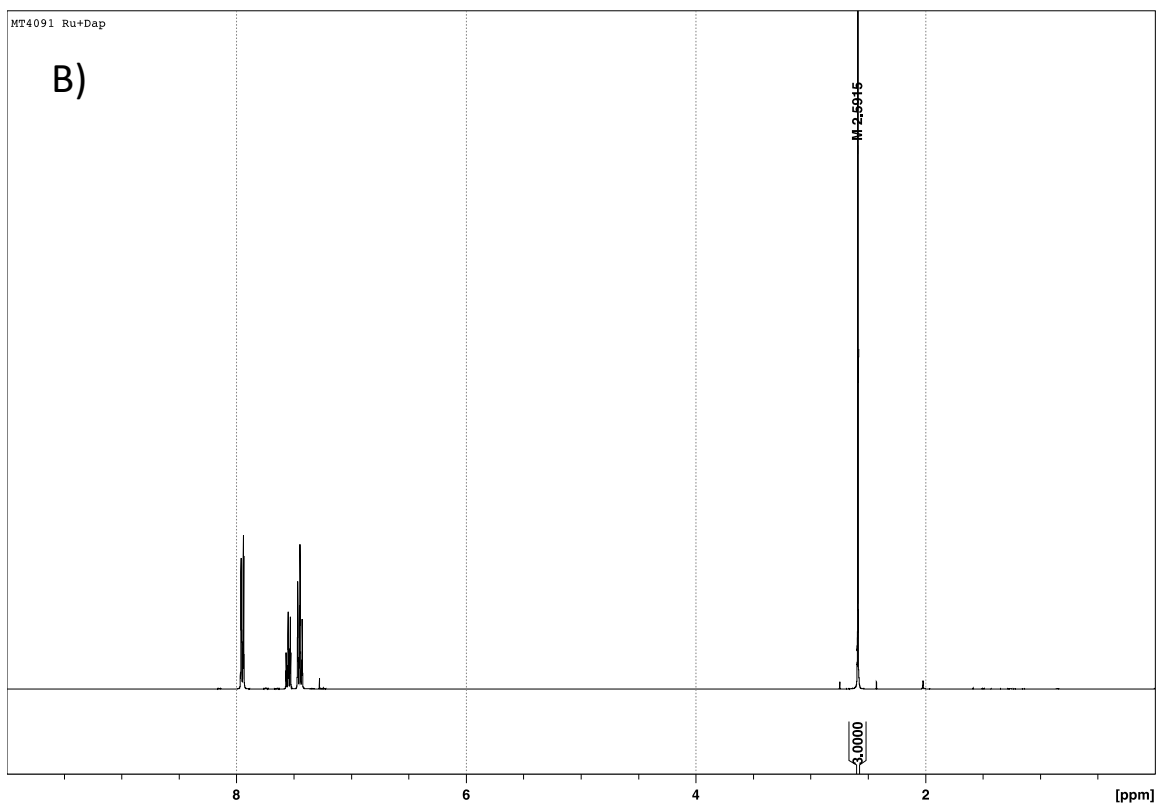
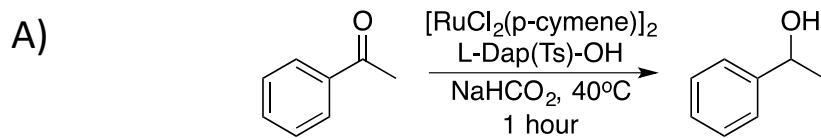


Figure 6.9. A) Synthesis of 1-phenylethanol from acetophenone using Ru^{2+} coupled with L-Dap(Ts)-OH. B) ^1H NMR spectrum of acetophenone reduction after one hour. There was no conversion of acetophenone to 1-phenylethanol.

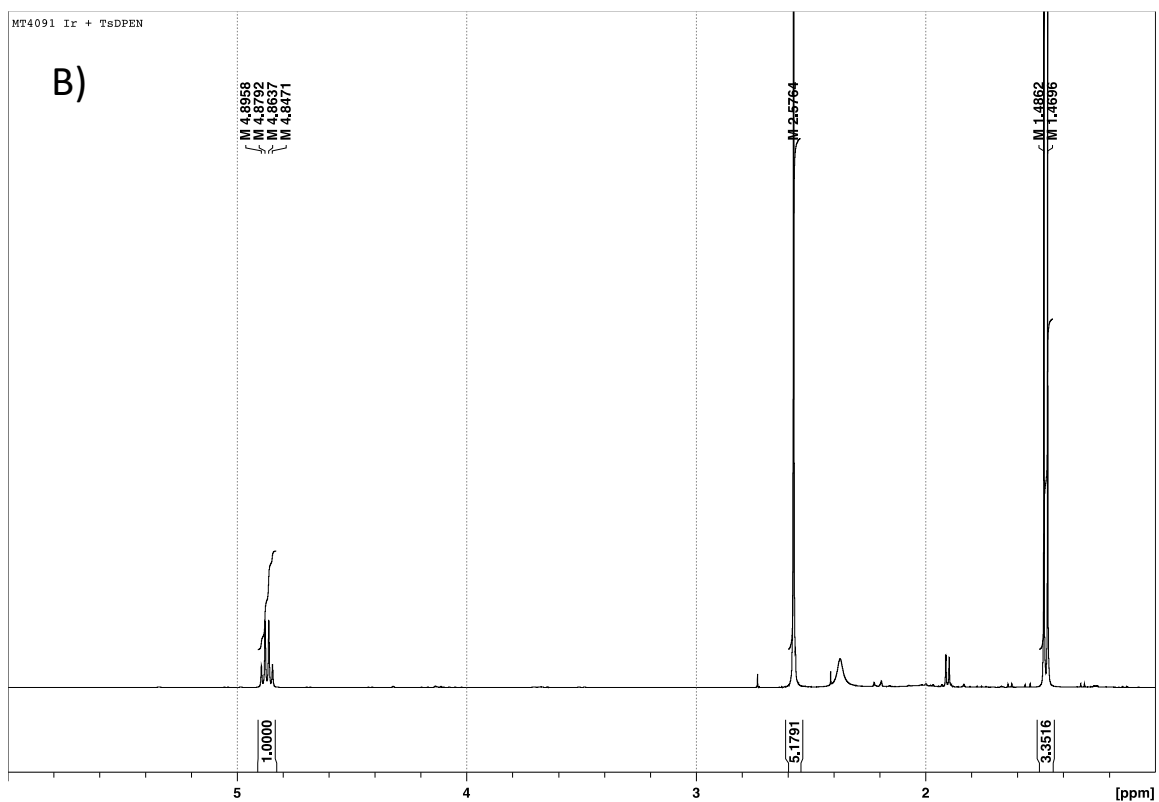
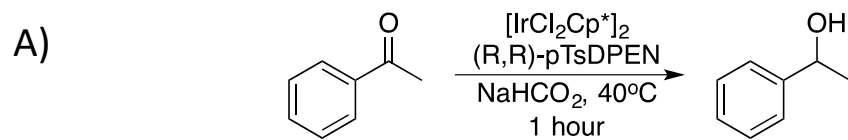


Figure 6.10. A) Synthesis of 1-phenylethanol from acetophenone using Ir³⁺ coupled with (R,R)-pTsDPEN. B) ¹H NMR spectrum of acetophenone reduction after one hour (the aromatic region is not shown). There was a 37% conversion of acetophenone to 1-phenylethanol.

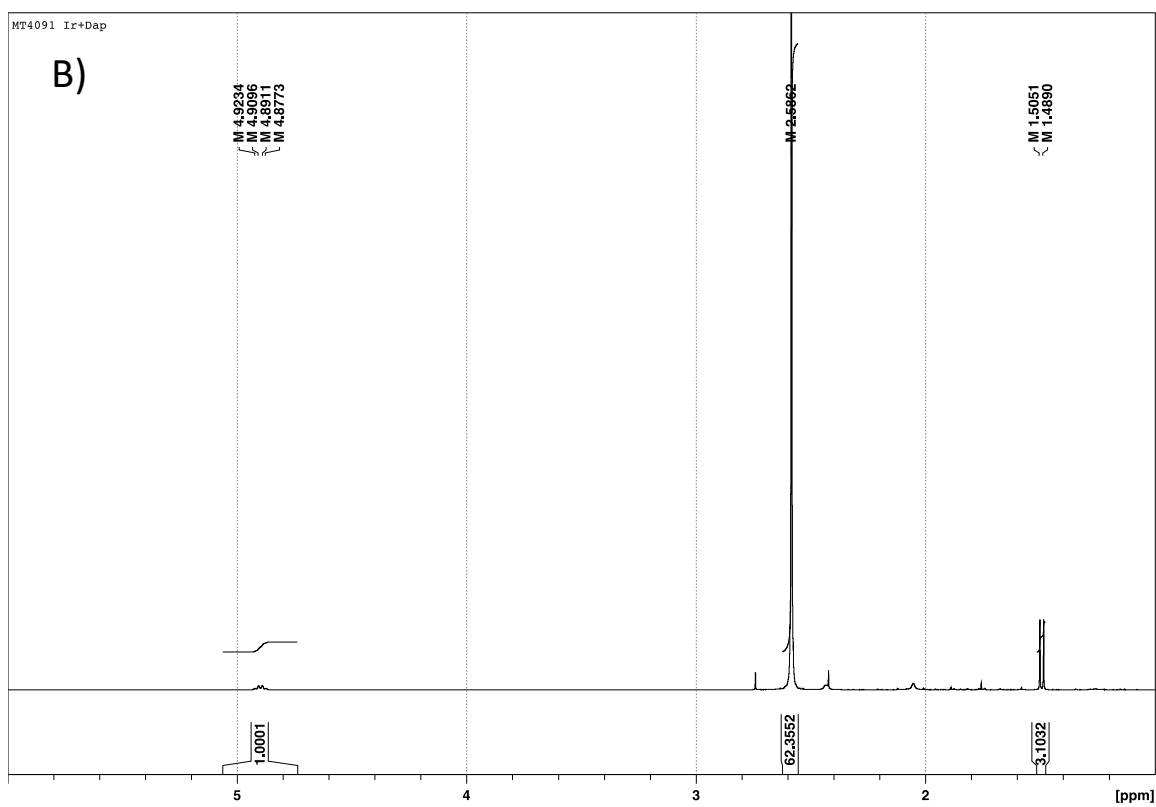
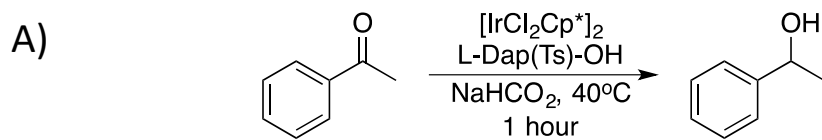


Figure 6.11. A) Synthesis of 1-phenylethanol from acetophenone using Ir³⁺ coupled with L-Dap(Ts)-OH. B) ¹H NMR spectrum of acetophenone reduction after one hour (the aromatic region is not shown). There was a 5% conversion of acetophenone to 1-phenylethanol.

6.3. Design of Tripeptide Iridium Catalysts

With the success of L-Dap(Ts)-OH in catalyzing the hydrogenation of acetophenone, the next step was to attach an amino acid that can coordinate to iridium and attach two phenylalanine residues to the C-terminal end of the peptide. We chose to use L-*allo*-threonine, which has the *S,S*-configuration and should theoretically impart enantioselectivity on any products. When analyzed by atomic force microscopy, there was evidence of fibril formation (Figure 6.11). With optimized reaction conditions, catalyst using L-*allo*-TFF as a ligand gave a yield of 85%. However, the product was only 5% ee making this an unsuitable catalyst for our needs.

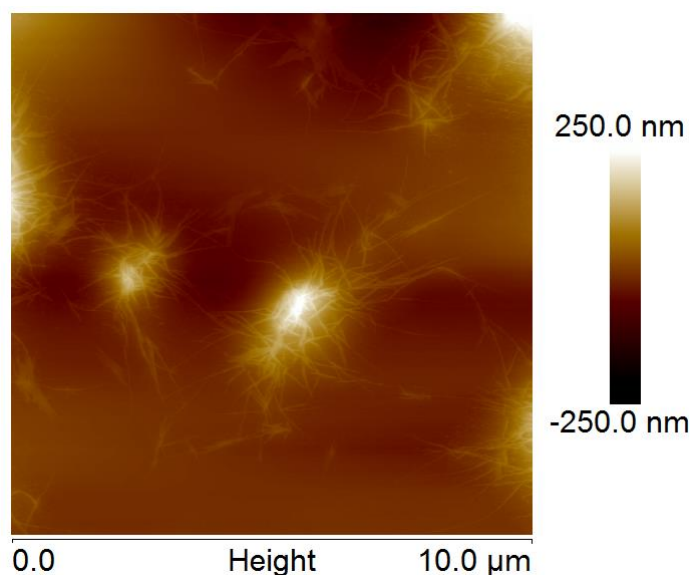


Figure 6.11. Atomic force microscopy image of L-*allo*-TFF peptide. Image was taken by Z. Lengyel.

6.4. Conclusions

We proposed a new catalyst system which uses a tripeptide as a ligand. The first amino acid, 2,3-diaminopropionic acid, will complex typical transition metals such as Ru^{2+} and Ir^{3+} that are often used in chiral catalysts. The two phenylalanine subunits branch off the C-terminal of the amino acid-ligand and impart aggregating properties onto the new catalyst. In addition to creating a chiral environment for catalysis, the design will enable easy separation using ultracentrifugation or by binding the catalyst to other surfaces. In preliminary studies, iridium coupled with L-Dap(Ts)-OH catalyzed the reduction of acetophenone in water. When analyzed by atomic force microscopy, there was evidence of fibril formation by the L-*allo*-TFF iridium complex. Under optimized reaction conditions, L-*allo*-TFF gave a yield of 85% for the reduction of acetophenone, however, the product was formed in only 5% ee. Further studies were continued by M. Dolan in the Korendovych lab.

6.5. Experimental

6.5.1. Solid-phase peptide synthesis

Peptides were synthesized manually using Fmoc (fluorenylmethyloxycarbonyl) protection and rink amide resin was used for solid support.¹⁷ The resin was first swelled in DMF (dimethylformamide) for at least 30 minutes and then was deprotected using 5% piperazine in

DMF for 5 minutes at 65°C. Amino acids were coupled to the resin using an AA:HCTU:DIEA:Resin (2:1.8:4:1) volume ratio in DMF for 4 minutes at 65°C. The resin was washed 4 times with DMF for 30 seconds between deprotection and coupling steps. After the final residue was coupled to the resin, the amino acid was deprotected. Cleavage of peptides and side chain deprotection were achieved by treating the resin with a solution of TFA (trifluoroacetic acid), water, and TIS (triisopropyl silane) in a 95:2.5:2.5 v/v ratio for two hours at room temperature. The resin was filtered off with a glass wool plug. TFA was removed from the peptide by bubbling N₂ into the solution until it was dried. Crude peptide was stored at -20°C.

6.5.2. Preparation of L-Dap(Ts)OH

L-(Boc)Dap-OH (2.45 mmol) was first dissolved in 15 mL of 1:5 1,4-dioxane:hexane solvent. Sodium carbonate (12.25 mmol) was added to solution¹⁸. The reaction was cooled to 0°C and TsCl (6.1 mmol) dissolved in 1,4-dioxane was slowly added to the reaction. The reaction mixture was allowed to warm to room temperature and stir overnight. The reaction was worked up by decreasing the pH of the reaction until it becomes cloudy (pH ~3), then extracted with ethyl acetate 3 times. The organic fractions were combined and washed with brine and water, then dried using Na₂SO₄. The product had a 92% yield and was a yellow oil.

To cleave the Boc protecting group, the compound was stirred in a mixture of TFA/TIS/water (95:2.5:2.5, vol/vol) for two hours.¹⁹ The compound was precipitated with N₂ flow until all TFA was removed and compound was yellow oil. The compound had a yield of 99% after Boc removal (91% overall yield).

6.5.3. Preparation of L-*allo*-Thr(Ts)-OH

Protocol was designed based on Roemmele et. al.²⁰ Amino acid L-*allo*-Thr-OH was dissolved in 5 mL of water. Sodium carbonate was added at 2.25 equivalents and stirred until it was dissolved. Meanwhile, 1.1 equivalents of TsCl (p-toluenesulfonyl chloride) was dissolved in 2.5 mL of 1,4-dioxane. TsCl solution was added to L-*allo*-Thr-OH solution dropwise and then allowed to stir overnight at room temperature.

To isolate the compound, 20 mL of water was added to the mixture. The mixture was extracted twice with MTBE (methyl tert-butyl ether). The aqueous layer was acidified to pH ~1 using phosphoric acid. Solid crystals were isolated by allowing the mixture to cool and recrystallize, then isolating the crystals via vacuum filtration.

6.5.4. Hydrogenation of acetophenone

Protocol was designed based on Wu et. al.²¹ A suspension of either $[\text{RuCl}_2(\text{p-cymene})]_2$ (0.005 mmol) or $[\text{IrCl}_2\text{Cp}^*]_2$ (0.005 mmol) and either (R,R)-TsDPEN (0.012 mmol) or L-Dap(Ts)-OH (0.012 mmol) in 2 mL H_2O was degassed three times using the freeze-pump-thaw method and then stirred in the shaker at 40°C. After one hour, NaHCO_2 (5.0 mmol) and acetophenone (1.0 mmol) were added and solution was degassed again. Reaction mixture was then shaken at 40°C for another hour. Organic compounds were extracted three times using diethyl ether (3 mL), concentrated, and analyzed by ^1H NMR in CDCl_3 . As these were preliminary experiments to determine if our peptides were plausible catalysts, no internal standard was used (future experiments by M. Dolan did use internal standard to obtain yields).

6.5.4. NMR

All samples were run on a 400 MHz Bruker NMR in acetone-d₆. To estimate yield, the quartet at 4.88 ppm was equibrate to an area of 1.0. To determine the estimated conversion, the area of the doublet at 1.48 ppm was divided by the total area of the singlet (2.58 ppm) and the doublet (1.48 ppm).

6.5.5. Gas Chromatography

All experiments were done using Nancy Totah's Agilent 7820A instrument with an Agilent J&W Cyclodex-B GC dolumn (30 m x 0.25 mm x 0.25 μm, 7 inch cage) and performed by M. Dolan of Korendovych lab. GC was run at isothermal conditions at 115°C with a 1.0 mL/min flow rate of He gas. The split ratio was 100:1. The retention time of the R-isomer was 29.7 minutes and the retention time of the S-isomer was 31.3 minutes.⁹⁻¹⁴ ee% was calculated by taking the area of the peak from the R isomer ([R]) and the area of the peak from the S isomer ([S]) in the following equation: $([R]-[S])/([R]+[S])*100$.

6.6. References

1. Yoon, T. P. J., Eric N, Privileged Chiral Catalysts. *Science* **2003**, *299*, 1691-3.
2. Pfaltz, A. D. I., William J, Design of chiral ligands for asymmetric catalysis: From C2-symmetric P,P- and N,N-ligands to sterically and electronically nonsymmetrical P,N-ligands. *PNAS* **2004**, *101* (16), 5723-6.

3. Benessere, V.; De Roma, A.; Ruffo, F., Carbohydrates as building blocks of privileged ligands for multiphasic asymmetric catalysis. *ChemSusChem* **2008**, *1* (5), 425-30.
4. Zhong, Z.; Yang, H.; Zhang, C.; Lewis, J. C., Synthesis and Catalytic Activity of Amino Acids and Metallopeptides with Catalytically Active Metallocyclic Side Chains. *Organometallics* **2012**, *31* (21), 7328-31.
5. Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caflisch, A.; Frenkel, D.; Gazit, E., Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nat Chem Biol* **2012**, *8* (8), 701-6.
6. Makin, O. S.; Serpell, L. C., Structures for amyloid fibrils. *FEBS J* **2005**, *272* (23), 5950-61.
7. Gazit, E., A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB* **2002**, *16*, 77-83.
8. Reches, M. G., Ehud, Casting Metal Nanowires Within Discrete Self-Assembled Peptide Nanotubes. *Science* **2003**, *300*, 625-7.
9. Hoarau, M.; Hureau, C.; Gras, E.; Faller, P., Coordination complexes and biomolecules: A wise wedding for catalysis upgrade. *Coordination Chemistry Reviews* **2016**, *308*, 445-59.
10. Haack, K.-J. H., Shohei; Fujii, Akio; Ikariya, Takao; Noyori, Ryoji, The Catalyst Precursor, Catalyst, and Intermediate in the Ru^{II}-Promoted Asymmetric Hydrogen Transfer between Alcohols and Ketones. *Angew Chem Int Ed Engl* **1997**, *36* (3), 285-8.

11. Hashiguchi, S. F., Akio; Haack, Karl-Josef; Matsumura, Kazuhiko; Ikariya, Takao; Noyori, Ryoji, Kinetic Resolution of Racemic Secondary Alcohols by Ru^{II}-Catalyzed Hydrogen Transfer. *Angew Chem Int Ed Engl* **1997**, *36* (3), 288-90.
12. Mao, J. B., David C, A Chiral Rhodium Complex for Rapid Asymmetric Transfer Hydrogenation of Imines with High Enantioselectivity. *Org Lett* **1999**, *1* (6), 841-3.
13. Matharu, D. S. M., David J; Kawamoto, Aparecida M; Clarkson, Guy J; Wills, Martin, A Stereochemically Well-Defined Rhodium(III) Catalyst for Asymmetric Transfer Hydrogenation of Ketones. *Org Lett* **2005**, *7* (24), 5489-91.
14. Zimbron, J. M.; Dauphinais, M.; Charette, A. B., Noyori-Ikariya catalyst supported on tetra-arylphosphonium salt for asymmetric transfer hydrogenation in water. *Green Chemistry* **2015**, *17* (6), 3255-9.
15. Wu, X. L., Xiaoguang; Hems, William; King, Frank; Xiao, Jianliang, Accelerated asymmetric transfer hydrogenation of aromatic ketones in water. *Org Biomol Chem* **2004**, *2*, 1818-21.
16. Lengyel, Z.; Rufo, C. M.; Moroz, Y. S.; Makhlynets, O. V.; Korendovych, I. V., Copper-Containing Catalytic Amyloids Promote Phosphoester Hydrolysis and Tandem Reactions. *ACS Catalysis* **2017**, *8* (1), 59-62.
17. Korendovych, I. V. K., Y.H; Ryan, A.H; Lear, J.D; DeGrado, W.F.; Shandler, S.J., Computational design of a self-assembling beta-peptide oligomer. *Org Lett* **2010**, *12* (22), 5142-5.
18. Loos, P.; Riedrich, M.; Arndt, H. D., Aza-Wittig access to chiral imidazol(in)es. *Chem Commun (Camb)* **2009**, (14), 1900-2.

19. Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stohr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., Short peptides self-assemble to produce catalytic amyloids. *Nat Chem* **2014**, *6* (4), 303-9.
20. Roemmele, R. C. R., H, Removal of N-arylsulfonyl groups from hydroxy alpha-amino acids. *J Org Chem* **1988**, *53* (10), 2367-4.
21. Wu, M.; Liu, W.; Yang, G.; Yu, D.; Lin, D.; Sun, H.; Chen, S., Engineering of a *Pichia pastoris* expression system for high-level secretion of HSA/GH fusion protein. *Appl Biochem Biotechnol* **2014**, *172* (5), 2400-11.

Chapter 7 Appendix

Appendix I. Copyright Permissions

All figures from other sources were used with permission from the publisher.

Appendix II. Supplemental Data

S.1. Sequencing

Table S.1.1. Sequencing for Chapter 2

CuSeCat	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQRM TAK
CuSeCat DD	MKDTDSEEEI REAFRVEDKD GNGYIDAAEL RIVMTNLGEK LTDEEVDEMI READIDGDGQ VDYEEFVQRM TAK
CuSeCat EE	MKDTDSEEEI REAFRVEDKD GNGYIEAAEL RIVMTNLGEK LTDEEVDEMI READIDGDGQ VEYEEFVQRM TAK
Ac7	MKDTDSEEEI REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK
AcDD	MKDTDSEEEI REQFRVEDKD GNGYIDAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VDYEEFVQRM TAK
AcEE	MKDTDSEEEI REQFRVEDKD GNGYIEAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VEYEEFVQRM TAK
AcCC	MKDTDSEEEI REQFRVEDKD GNGYICAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VCYEEFVQRM TAK
Y99E	MKDTDSEEEI REQFRVEDKD GNGEISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK
Y99C	MKDTDSEEEI REQFRVEDKD GNGCISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK

Q135E	MKDTDSEEEEL LTDEEVDELI	REQFRVEDKD RETDIDGDGE	GNGYISAAEL VNYEEFVQRM	RIVMTNRGEK TAK
HolIEE	MKDTDSEEEEL LTDEEVDELI	REQFRVEDKE RETDIEGDGQ	GNGYISAAEL VNYEEFVQRM	RIVMTNRGEK TAK
ReCCes	MKDTDSEEEEL LTDEEVDELI	REQFRVEDKC RETDICGDGQ	GNGYISAAEL VNYEEFVQRM	RIVMTNRGEK TAK

Table S.1.2. Sequencing for Chapter 3

2E10-HolIEE	MHHHHHSSG	VDLGTENLYF	QSNDIVMTRS	PLSLPVTLGQ
	PASISCRSSQ	SLVYSDGNTY	LNWFQQRPGQ	SPRRLIYKVS
	NRDSGVPDRF	SGSGSGTDFT	LKISRVEAED	VGVYYCMQGT
	HWPPVFGPGT	KVDIKGGSSR	SSSSGGGGSG	GGGQVQLQES
	GGPLVKPSET	LSLTCTVSGG	SISSSNYWYG	WIRQPPGKGL
	EWIGSIYYSG	STYYNPSLKS	RVTISVDTSK	NQFSLKLSSV
	TAADTAVYYC	ARVHDGHRYG	TYYYYGLDVW	GQGTAVTVSS
	GSAGSAAGSG	EFMKDTDSEE	ELREQFRVED	KEGNGYISAA
	ELRIVMTNRG	EKLTDEEVDE	LIRETDIEGD	GQVNYEEFVQ
	RMTAK			

Table S.1.3. Sequencing for Chapter 4

FDH-N	HHHHHSSGV	DLGTENLYFQ	SNAMKIVLVL	YDAGKHADE
	EKLYGCTENK	LGIANWLKDQ	GHELITTSK	EGETSELDKH
	IPDADI IITT	PFHPAYITKE	RLDKAKNLKL	VVVAGVGS DH
	IDLDYINQTG	KKISVLEVTG	SNVVSVAEHV	VMTMLVLVRN
	FVPAHEQIIN	HDWEVAAIAK	DAYDIEGKTI	ATIGAGRIGY
	RVLERLLPFN	PKELLYDYQ	ALPKEAEEKV	GARRVENIEE
	LVAQADIVTV	NAPLHAGTKG	LINKELLSKF	KKGAWLVNTA
	RGAICVAEDV	AAALESGQLR	GYGGDVWFPQ	PAPKDHPWRD
	MRNKYGAGNA	MTPHYSGTTL	DAQTRYAEGT	KNILESFFTG
	KFDYRPQDII	LLNGEYVTKA	YGKHDKK	
	FDH-C	MKIVLVLYDA	GKHAADDEEKL	YGCTENKLG I
LITTSKKEGE		TSELDKHIPD	ADIIITTPFH	PAYITKERLD
KAKNLKLVVV		AGVGS DHIDL	DYINQTGKKI	SVLEVTGSNV
VSVAEHV VMT		MLVLVRNFVP	AHEQIINH DW	EVAAIAKDAY
DIEGKTIATI		GAGRIGYRVL	ERLLPFNPKE	LLYYDYQALP
KEAEEKVGAR		RVENIEELVA	QADIVTVNAP	LHAGTKGLIN

	KELLSKFKKG GDVWFPPAP TRYAEGTKNI HDKKVDKLAA	AWLVNTARGA KDHPWRDMRN LESFFTGKFD ALEHHHHHH	ICVAEDVAAA KYGAGNAMTP YRPQDIILLN	LESGQLRGY HYSGTTLDAQ GEYVTKAYGK
FDH-N-y	EAEAYVEFHH AGKHAADEEK ETSELDKHIP VAGVGS DHID TMLVLRNFV IGAGRIGYRV RRVENIEELV GAWLVNTARG PKDHPWRDMR ILESFFTGKF	HHHSSGVDL LYGCTENKLG DADIIITTPF LDYINQ TGKK PAHEQIINH D LERLLPFNPK AQADIVTVNA AICVAEDVAA NKYGAGNAMT DYRPQDIILL	GTENLYFQSN IANWLKDQGH HPAYITKERL ISVLEVTGSN WEVAAIAKDA ELLYYDYQAL PLHAGTKGLI ALESGQLRGY PHYSGTTLDA NGEYVTKAYG	AMKIVLVLYD ELITTS DKEG DKAKNLKLVV VVSVAEHVVM YDIEGKTIAT PKEAEEKVGA NKELLSKFKK GGDVWFPPA QTRYAEGTKN KHDKK
FDH-C-y	EAEAYVEFMK WLKDQGHELI YITKERLDKA LEV TGSNVVS AAIAKDAYDI YYDYQALPKE AGTKGLINKE SGQLRGYGGD SGTTLDAQTR YVTKAYGKHD	IVLVLYDAGK TTSDKEGETS KNLKL VV VAG VAEHVVM TML EGKTIATIGA AEEKVGARRV LLSKFKKGAW VWFPPAPKD YAEGTKNILE KKG ENLYFQS	HADEEEKLYG ELDKHIPDAD VGS DHIDL DY VLVRNFVPAH GRIGYRVLER ENIEELVAQA LVNTARGAIC HPWRDMRNKY SFFT GKFDYR HHHHHH	CTENKLGIAN IIITTPFHPA INQ TGKKISV EQIINH DWEV LLPFNPKELL DIVTVNAPLH VAEDVAAALE GAGNAMTPHY PQDIILLNGE
FIDH-aeru	MSGNRGVVYL VVSTNICGSD VETMKIGDLV GGAYGYVDMG KIRDLTCLSD AAAASARLLG TPLHEQIAAL ATVLNSLMGI SIRFGLGWAK DIVGVEVITL	GPGKVEVQNI QHMRGRRTA SVPFN VACGH GWVGGQAEYV ILPTGYHGAV AAVVIVGDVN LGEPEVDCAV TRVAGKIGIP SHSFHTGQTP DDAPKGYGEF	PYPKMQDPQG PEGLVLGHEI CRTCKEQHTG LVPYADFNLL TAGVGP GSTV PTRLAHAKKQ DAVGFEARGH GLYVTE DPGA VMKYNRQLMQ DAGVPK KFVI	RQIDHGVILR TGEVVEIGRG VCLTVNPARA KLPNREAME YIAGAGPVGL GFEIADLSKD GHSQSQQEAP VDAAAKHGAL AIMWDRIKIA DPHNL FRAA
FIDH-puti	MSGNRGVVYL VVSTNICGSD VERMQIGDLV GGAYGYVDMG KIRDLTCLSD AAAASARLLG	GAGKVEVQKI QHMRGRRTA SVPFN VACGR DWTGGQAEYV ILPTGYHGAV AACVIVGDLN	DYPKMQDPRG QVGLVLGHEI CRSCKEMHTG LVPYADFNLL TAGVGP GSTV PARLAHAKSQ	KKIEHGVILK TGEIVEKGRD VCLTVNPARA KLPERDKAME YVAGAGPVGL GFEVVDLSKD

	TPLHEQIVDI	LGEPEVDCAV	DAVGFEARGH	GHEGAKHEAP
	ATVLNSLMQV	TRVAGNIGIP	GLYVTEPGA	VDAAAKIGAL
	SIRFGLGWAK	SHSFHTGQTP	TMKYNRQLMQ	AIMWDRINIA
	EVVGVQVINL	DQAPEGYGEF	DAGVPPKFVI	DPHKMWGAA
ADH	MSIPETQKGV	IFYESHGKLE	YKDIPVPPKPK	ANELLINVKY
	SGVCHTDLHA	WHGDWPLPVK	LPLVGGHEGA	GVVVGMGENV
	KGWKIGDYAG	IKWLNGSCMA	CEYCELGNES	NCPHADLSGY
	THDGSFQQYA	TADAVQAAHI	PQGTDLAQVA	PILCAGITVY
	KALKSANLMA	GHWVAISGAA	GGLGSLAVQY	AKAMGYRVLG
	IDGGEGKEEL	FRSIGGEVFI	DFTKEKDIVG	AVLKATDGGA
	HGVINVSVSE	AAIEASTRYV	RANGTTVLVG	MPAGAKCCSD
	VFNQVVKSIS	IVGSYVGNRA	DTREALDFFA	RGLVKSPIKV
	VGLSTLPEIY	EKMEKGQIVG	RYVVDTSK	

Table S.1.4. Sequencing data for Chapter 5

cCam	MKDTDSEEEI	REAFRVFDKD	GNGYISAAEL	RHVMTNLGEK
	LTDEEVDEMI	READIDGDGQ	VNYEEFVQMM	TAK
Ac0	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RHVMTNLGEK
	LTDEEVDEMI	READIDGDGQ	VNYEEFVQMM	TAK
Ac0T	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RHVMTNLGEK
	LTDEEVDEMI	RETDIDGDGQ	VNYEEFVQMM	TAK
Ac1	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RHVMTNLGEK
	LTDEEVDEMI	READIDGDGQ	VNYEEFVQRM	TAK
Ac1T	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RHVMTNLGEK
	LTDEEVDEMI	RETDIDGDGQ	VNYEEFVQRM	TAK
Ac2	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RIVMTNLGEK
	LTDEEVDEMI	READIDGDGQ	VNYEEFVQRM	TAK
Ac2T	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RIVMTNLGEK
	LTDEEVDEMI	RETDIDGDGQ	VNYEEFVQRM	TAK
Ac3	MKDTDSEEEI	REAFRVEDKD	GNGYISAAEL	RIVMTNRGEK
	LTDEEVDEMI	READIDGDGQ	VNYEEFVQRM	TAK

Ac3T	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDEMI RETDIDGDGQ VNYEEFVQRM TAK
Ac4	MKDTDSEEEEL REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDEMI READIDGDGQ VNYEEFVQRM TAK
Ac5	MKDTDSEEEEL REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDEMI RETDIDGDGQ VNYEEFVQRM TAK
Ac6	MKDTDSEEEEL REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK
Ac6A	MKDTDSEEEEL REAFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI READIDGDGQ VNYEEFVQRM TAK
Ac7	MKDTDSEEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK
Ac7A	MKDTDSEEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI READIDGDGQ VNYEEFVQRM TAK
c-V108E	MKDTDSEEEI REAFRVFDKD GNGYISAAEL RHEMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMM TAK
Ac0-V108E	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RHEMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMM TAK
Ac1-V108E	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RHEMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQRM TAK

S.2. Supplemental data for Chapter 2

S.2.1. Gel Electrophoresis

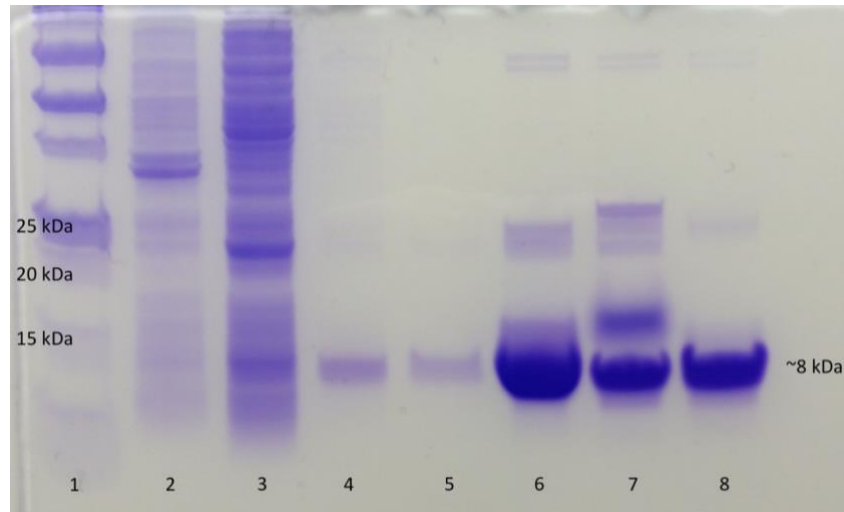


Figure S.1. 15% acrylamide SDS-PAGE gel of Ac7 purification. Samples were taken at various stages of the purification and labelled as follows: 2) Ac7 after induction, 3) Ac7 cell debris, 4) Ac7 Ni-NTA flow through, 5) Ac7 Ni-NTA wash, 6) Ac7 uncut, 7) Ac7 + TEV protease, 8) Ac7 cut pure.

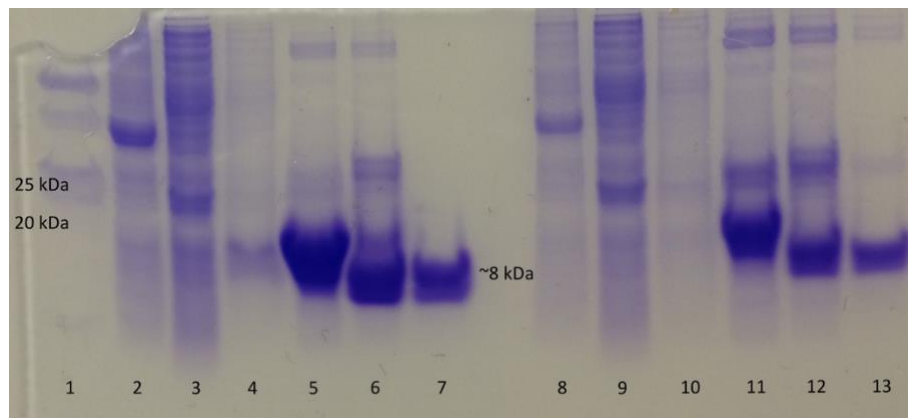


Figure S.2. 15% acrylamide SDS-PAGE gel of AcDD (lanes 2-7) and AcEE (lanes 8-13) purification. Aliquots were taken at various stages of the purification and labelled as follows: 2) AcDD after induction, 3) AcDD cell debris, 4) AcDD Ni-NTA flow through, 5) AcDD uncut, 6) AcDD + TEV protease, 7) AcDD cut pure, 8) AcEE after induction, 9) AcEE cell debris, 10) AcEE Ni-NTA flow through, 11) AcEE uncut, 12) AcEE + TEV protease, 13) AcEE cut pure.

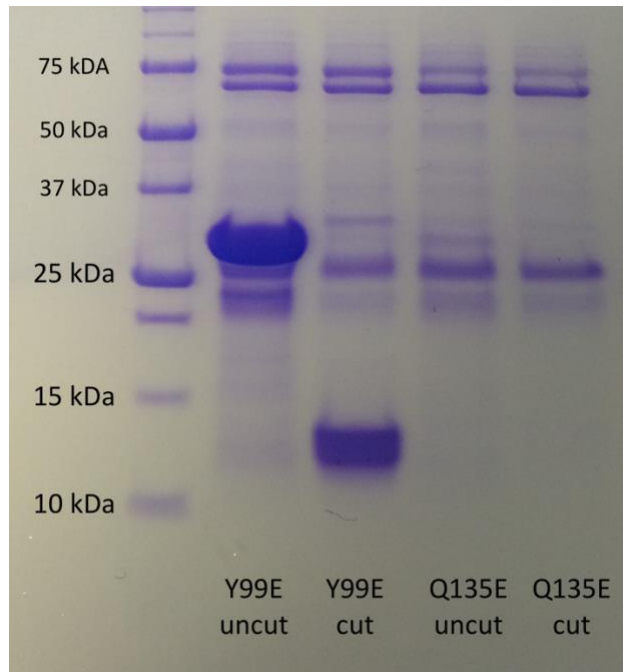


Figure S.3. A 15% acrylamide SDS-PAGE gel of purified Y99E and Q135E. “Uncut” refers to proteins with SUMO-tag still attached to the protein while “cut” proteins have SUMO-tag removed. No expression was seen for the Q135E mutant.

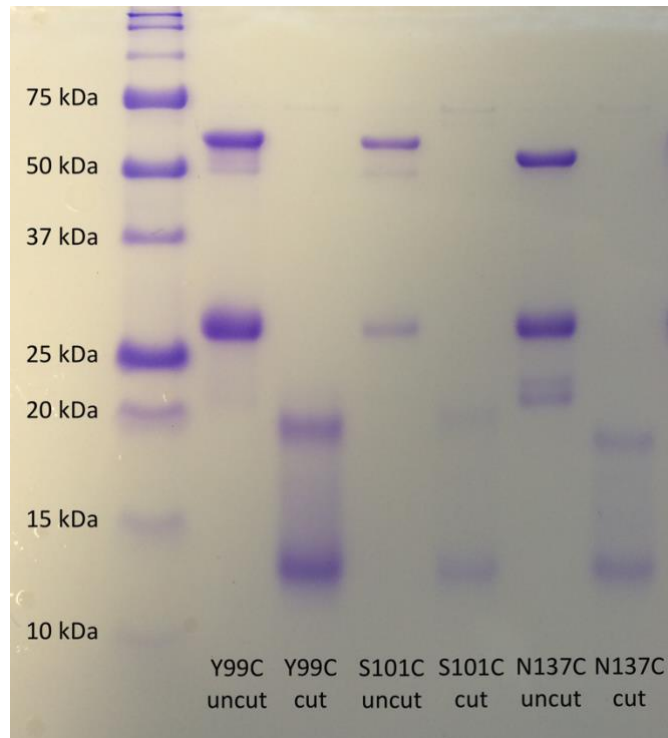


Figure S.4. A 15% acrylamide SDS-PAGE gel of purified Y99C, S101C, and N137C. While the protein is normally expected around 8-12 kDa, bands at double that molecular weight were seen. “Uncut” refers to proteins with SUMO-tag still attached to the protein while “cut” proteins have SUMO-tag removed.

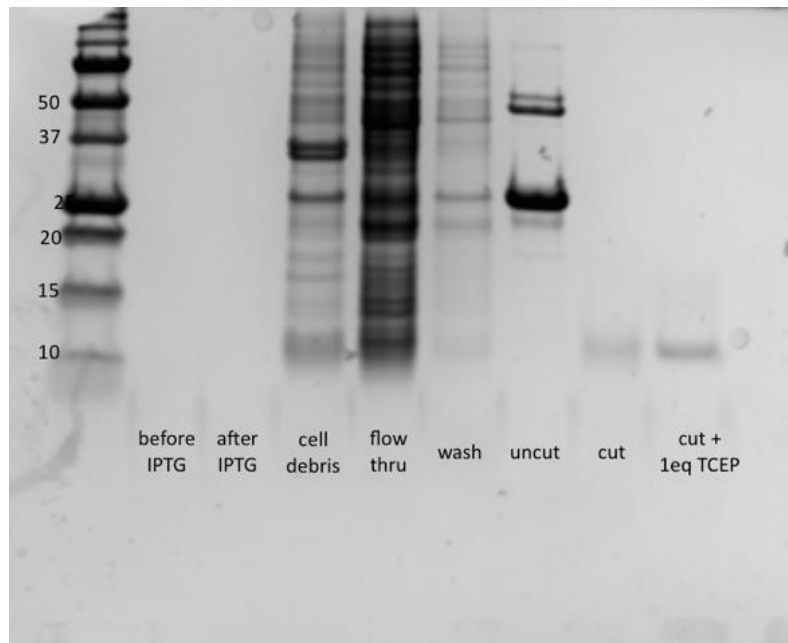


Figure S.5. A 15% acrylamide SDS-PAGE gel of ReCCes purification. The expected molecular weight of ReCCes is approximately 8 kDa after cleavage. Before cleavage of the His-tag, ReCCes forms a dimer due to disulfide bonds between cysteine residues.

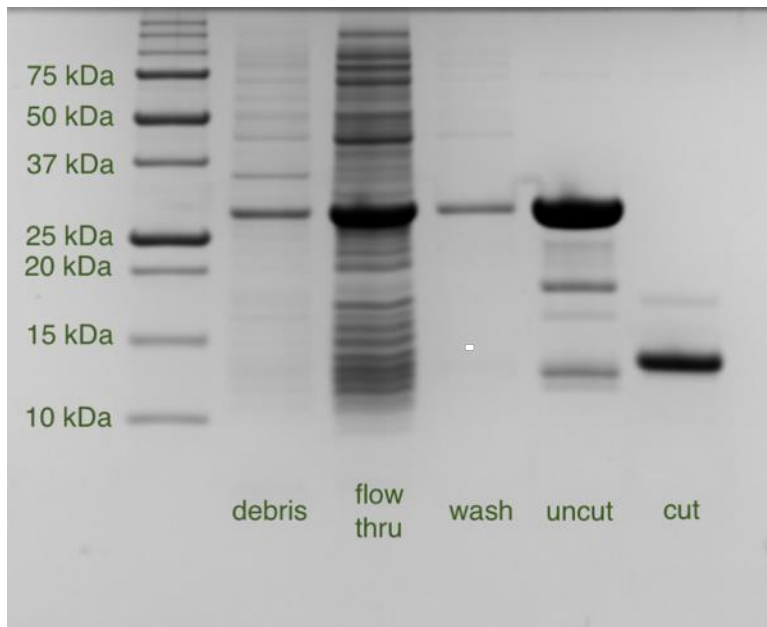


Figure S.6. A 15% acrylamide SDS-PAGE gel of HolIEE purification. The expected MW of HolIEE is 8 kDa after cleavage of His-tag. Debris refers to the cell debris after sonification and centrifugation, and flow through and wash refer to steps during Ni-NTA purification.

S.2.2. Kemp elimination kinetics

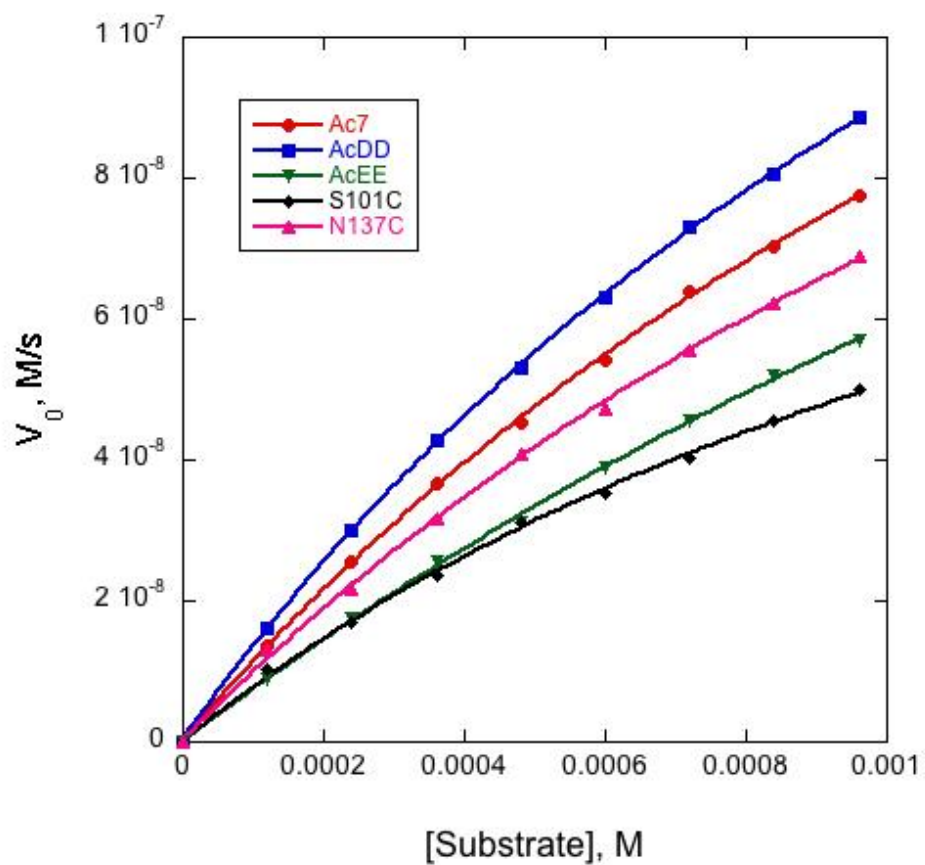


Figure S.7. Catalytic efficiency graph of Kemp elimination catalyzed by Ac7, AcDD, AcEE, S101C, and N137C (after His-tag cleavage) in 20 mM HEPES buffer pH 7.0 with addition of 0.2 mM $\text{Ca}(\text{NO}_3)_2$ and without reducing agent (β -ME). (MT3114-8)

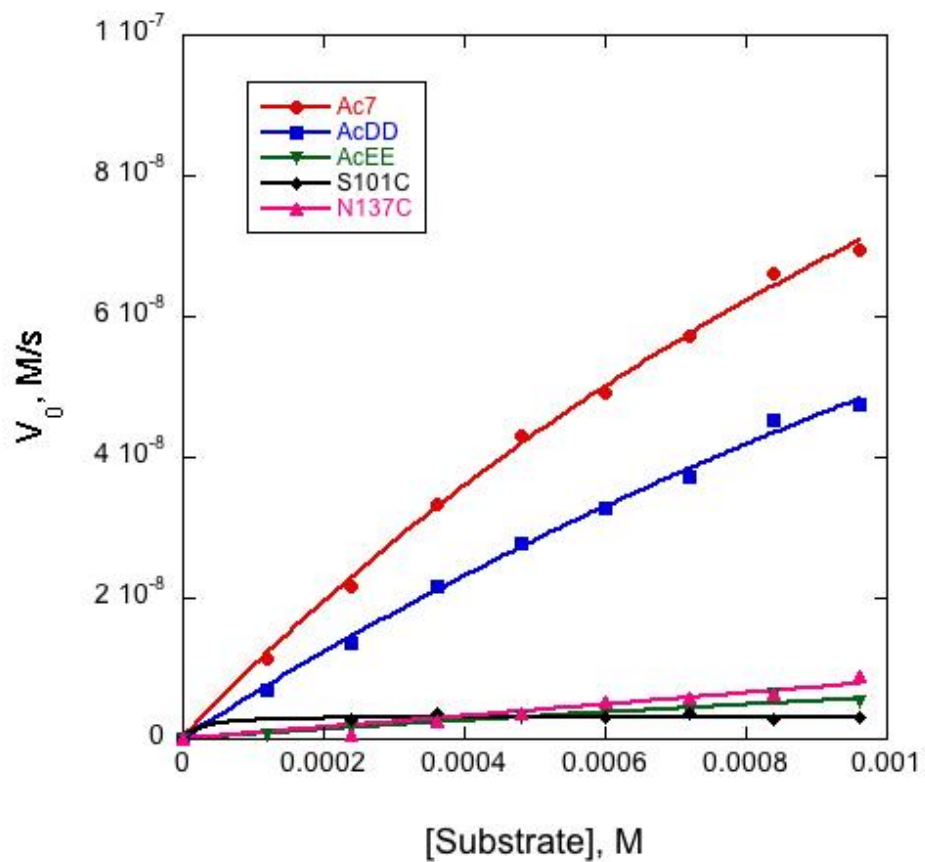


Figure S.8. Catalytic efficiency of Kemp elimination catalyzed by Ac7, AcDD, AcEE, S101C, and N137C in 20 mM HEPES buffer, pH 7.0 with addition of 0.2 mM $\text{Pb}(\text{NO}_3)_2$ and without reducing agent (β -ME). (MT3117-8)

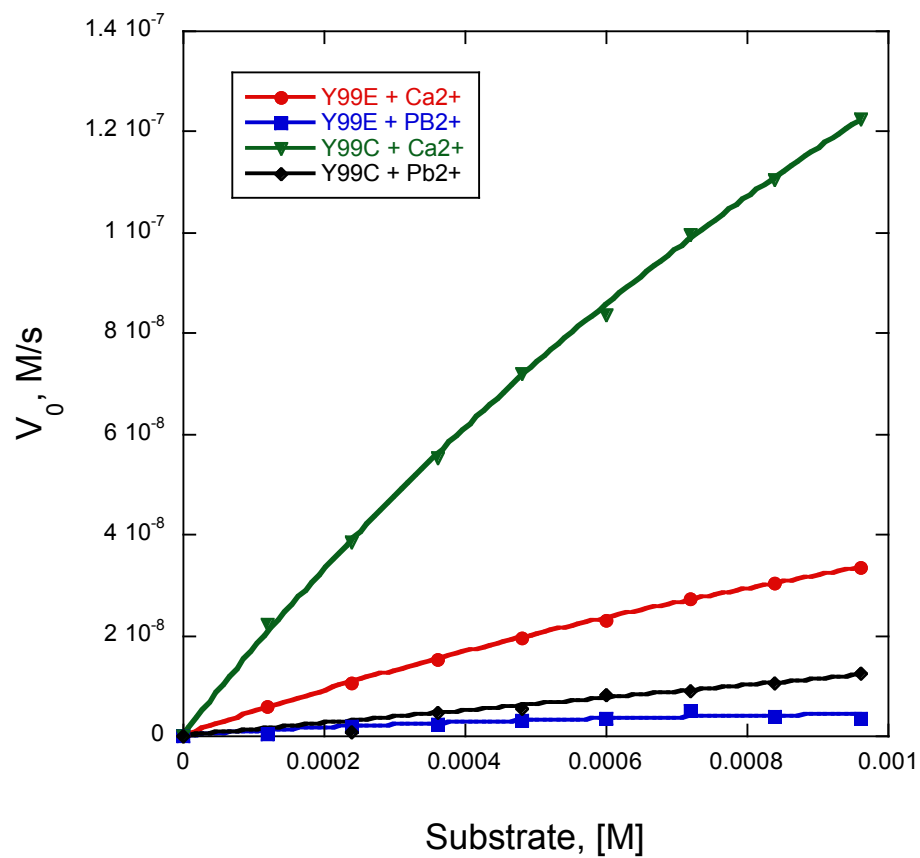


Figure S.9. Michaelis-Menten graph of Kemp elimination catalyzed by Y99E and Y99C in 20 mM HEPES buffer, pH 7.0, with addition of 0.2 mM $\text{Ca}(\text{NO}_3)_2$ or 0.2 mM $\text{Pb}(\text{NO}_3)_2$ and without reducing agent (β -ME).

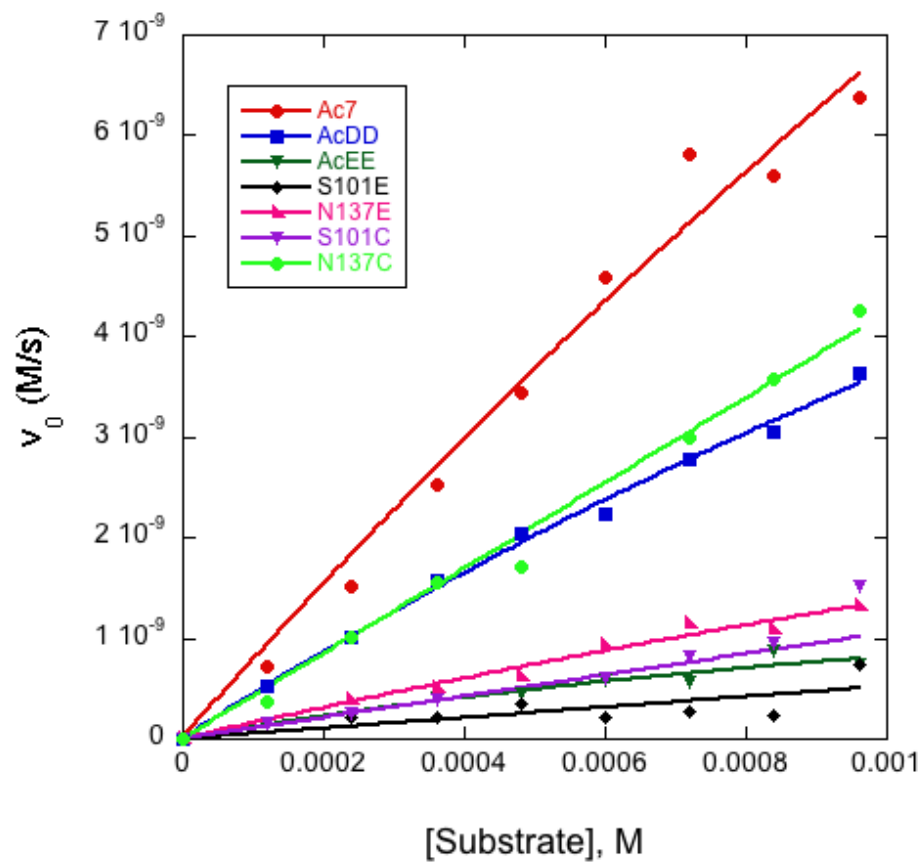


Figure S.10. Catalytic efficiency of Kemp elimination catalyzed by Ac7, AcDD, AcEE, Y99E, Y99C, S101C, and N137C (2.5 μM) in 20 mM Hepes buffer pH 7.0 with addition of 0.2 mM $\text{Y}(\text{NO}_3)_2$. The k_{cat}/K_M values are $32 \text{ M}^{-1}\text{s}^{-1}$ (Ac7), $18 \text{ M}^{-1}\text{s}^{-1}$ (AcDD), $5 \text{ M}^{-1}\text{s}^{-1}$ (AcEE), n/a (S101E), $6 \text{ M}^{-1}\text{s}^{-1}$ (N137E), n/a (S101C), and n/a (N137C). (MT3131)

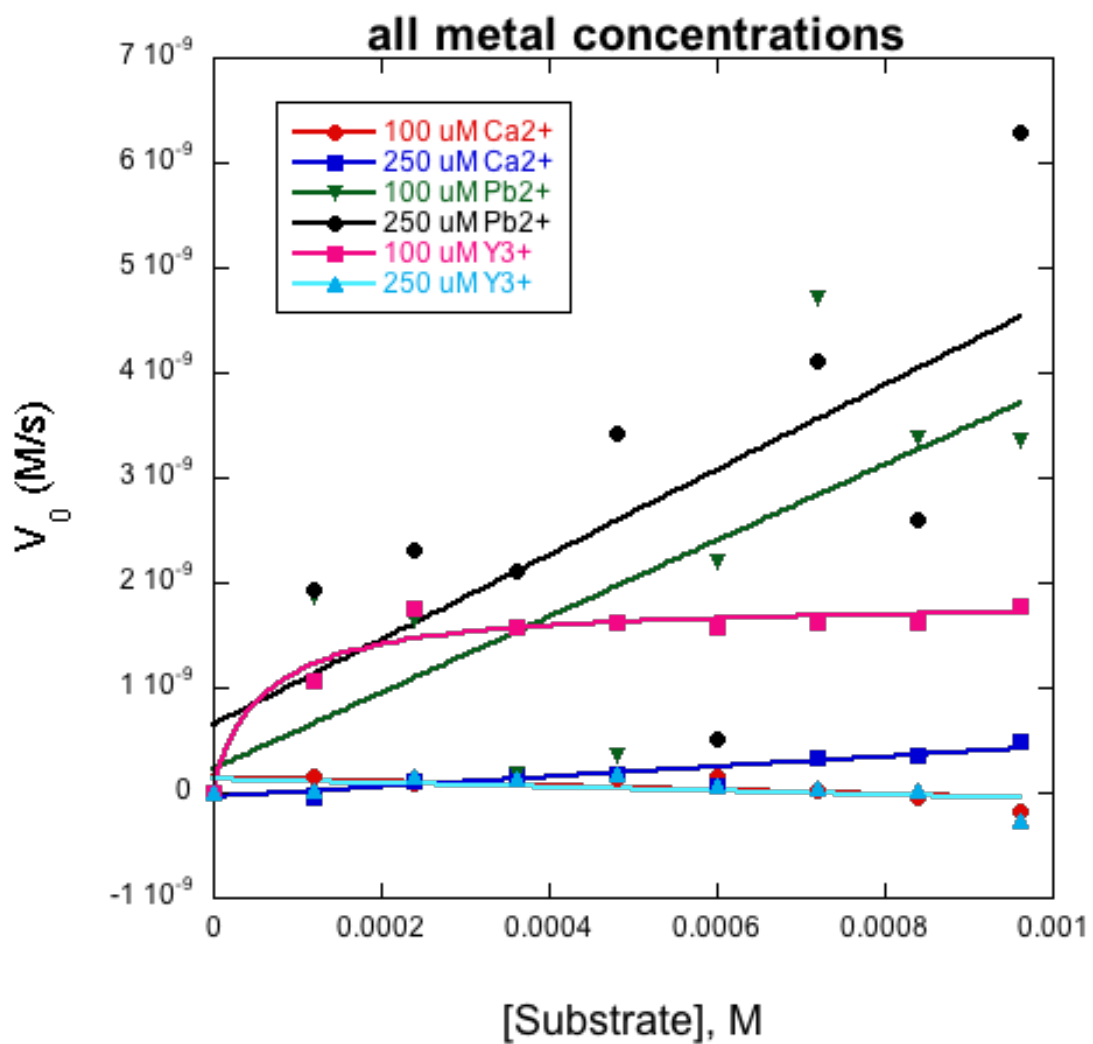


Figure S.11. Catalytic efficiency of Kemp elimination catalyzed by D95C D131C mutant (5.0 μ M) in 20 mM Hepes buffer with addition of 0.1 mM or 0.25 mM of either $\text{Ca}(\text{NO}_3)_2$, $\text{Pb}(\text{NO}_3)_2$, or $\text{Y}(\text{NO}_3)_2$.

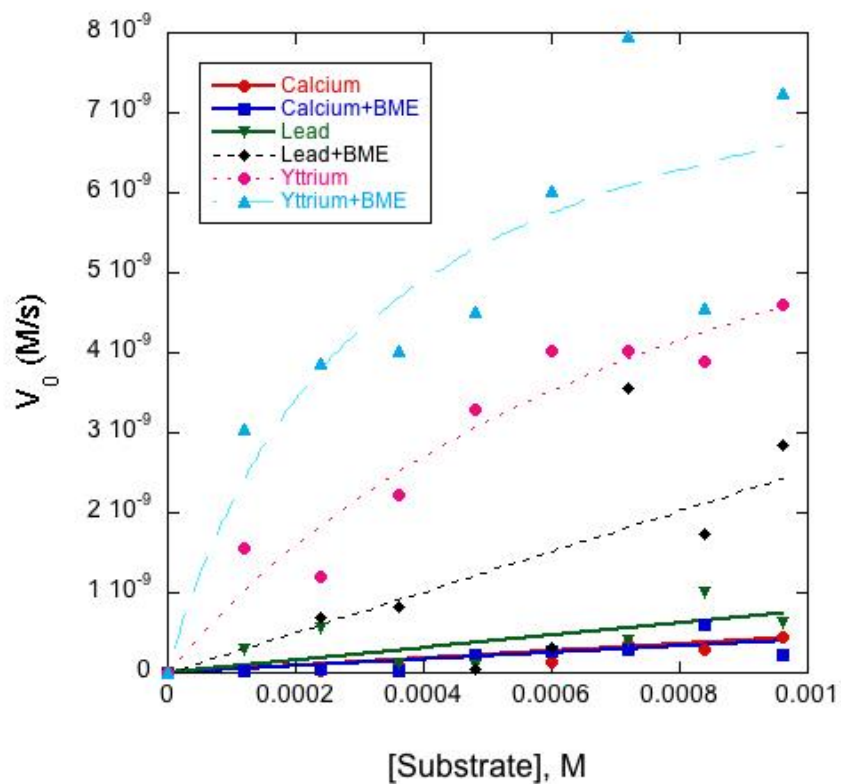


Figure S.12. Catalytic efficiency of Kemp elimination catalyzed by D95C D131C mutant (5.0 μ M) in 20 mM HEPES buffer and 0.25 mM of either $\text{Ca}(\text{NO}_3)_2$, $\text{Pb}(\text{NO}_3)_2$, or $\text{Y}(\text{NO}_3)_3$, with the addition or omission of 1.2 equivalence of β -mercaptoethanol. (MT5107)

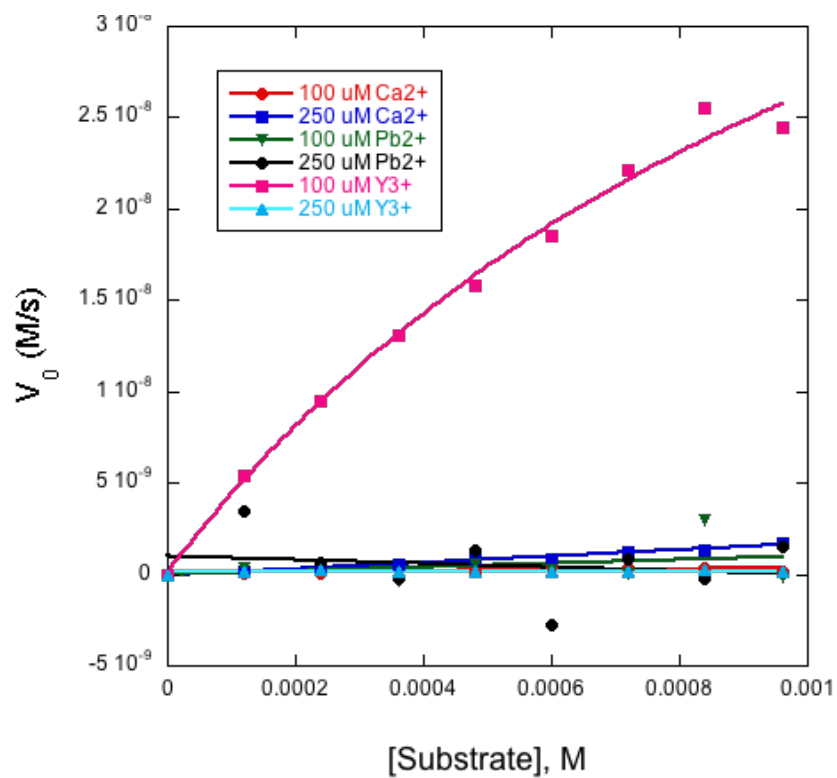


Figure S.13. Catalytic efficiency of Kemp elimination catalyzed by D95E D131E mutant (5.0 μM) in 20 mM Hepes buffer with addition of 0.1 mM or 0.25 mM of either $\text{Ca}(\text{NO}_3)_2$, $\text{Pb}(\text{NO}_3)_2$, or $\text{Y}(\text{NO}_3)_2$.

S.2.3. pH profiles of HolIEE

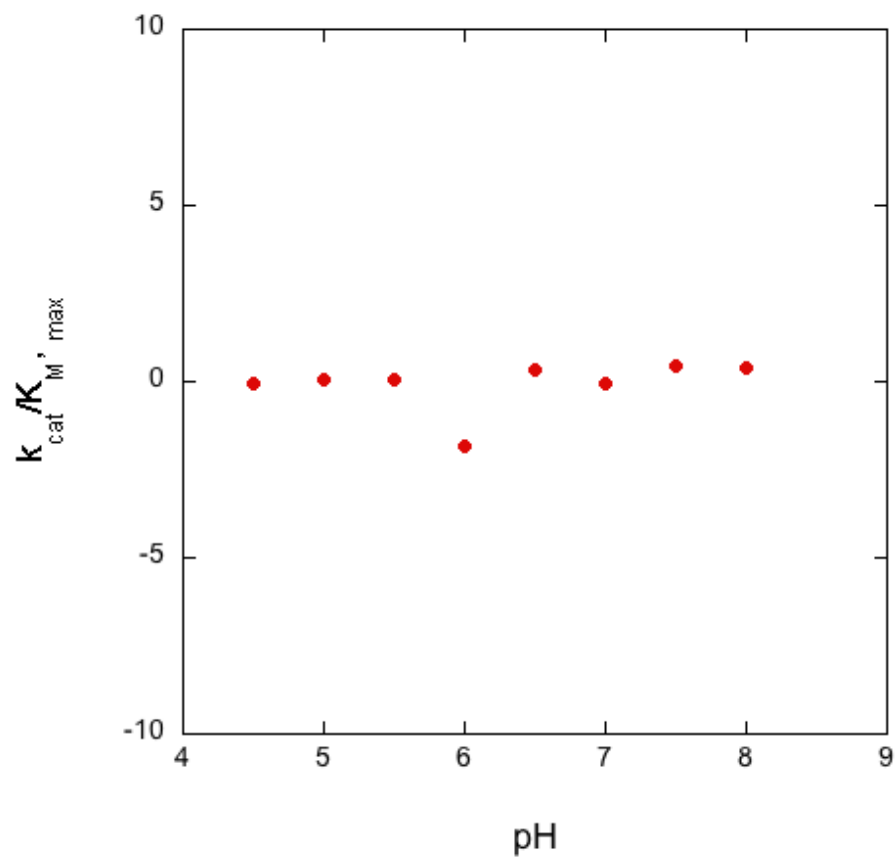


Figure S.14. pH rate profile of HolIEE (1 μ M) with 80 μ M calcium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

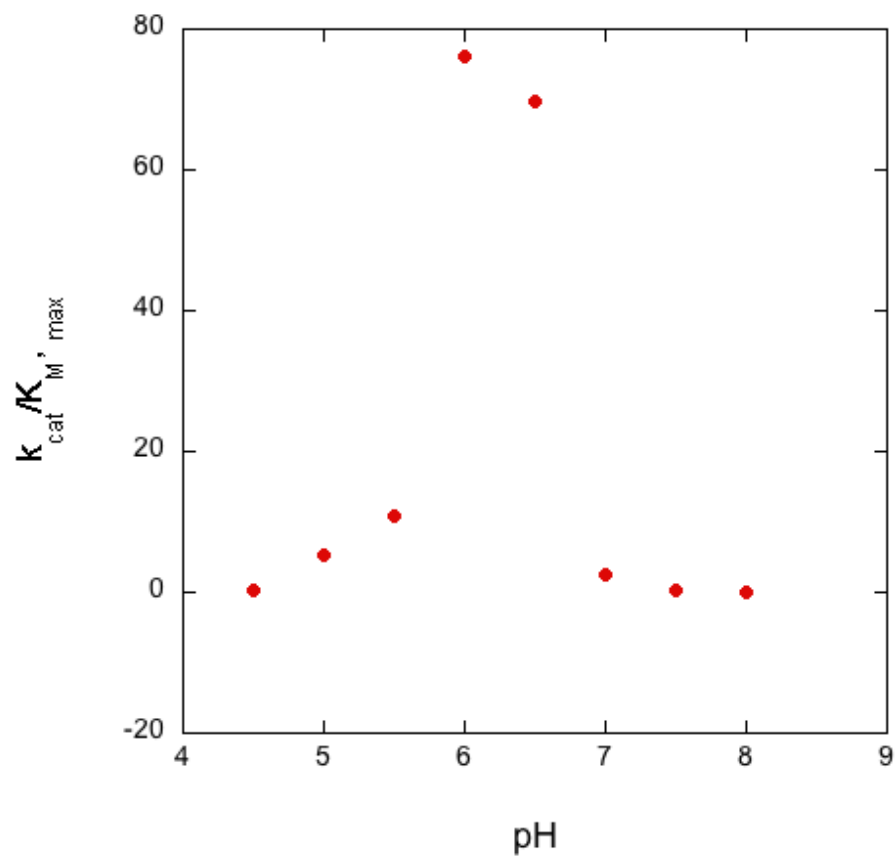


Figure S.15. pH profile of HolIEE (1 μ M) with 80 μ M lutetium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used. (AG1049)

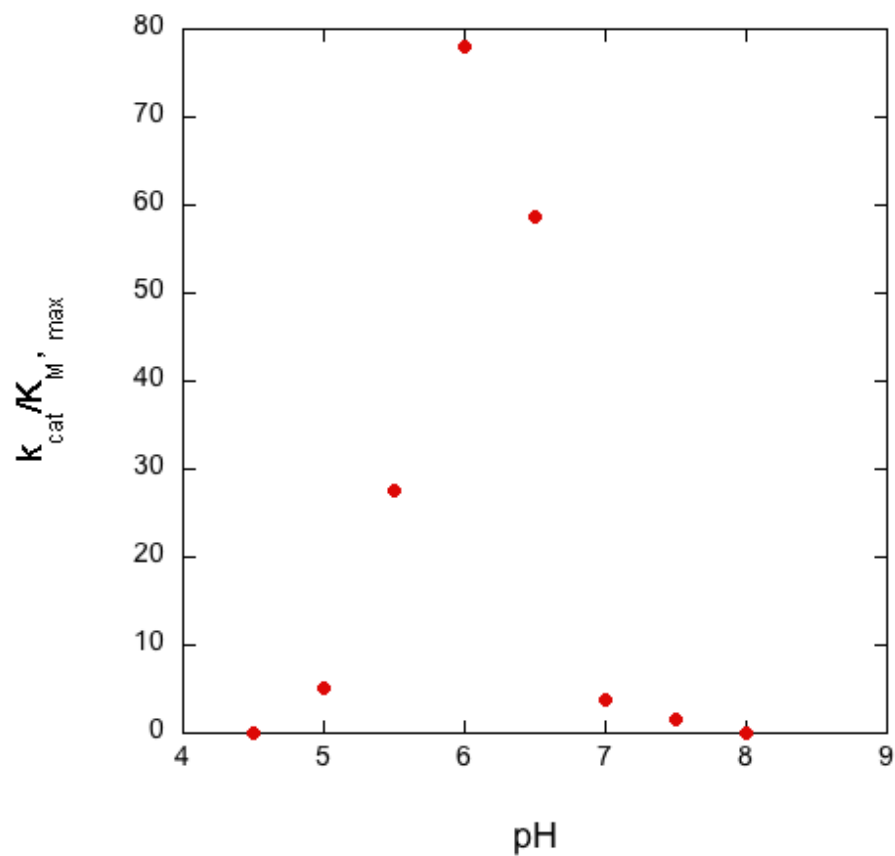


Figure S.16. pH profile of HolIEE (1 μ M) with 80 μ M ytterbium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used. (AG1049)

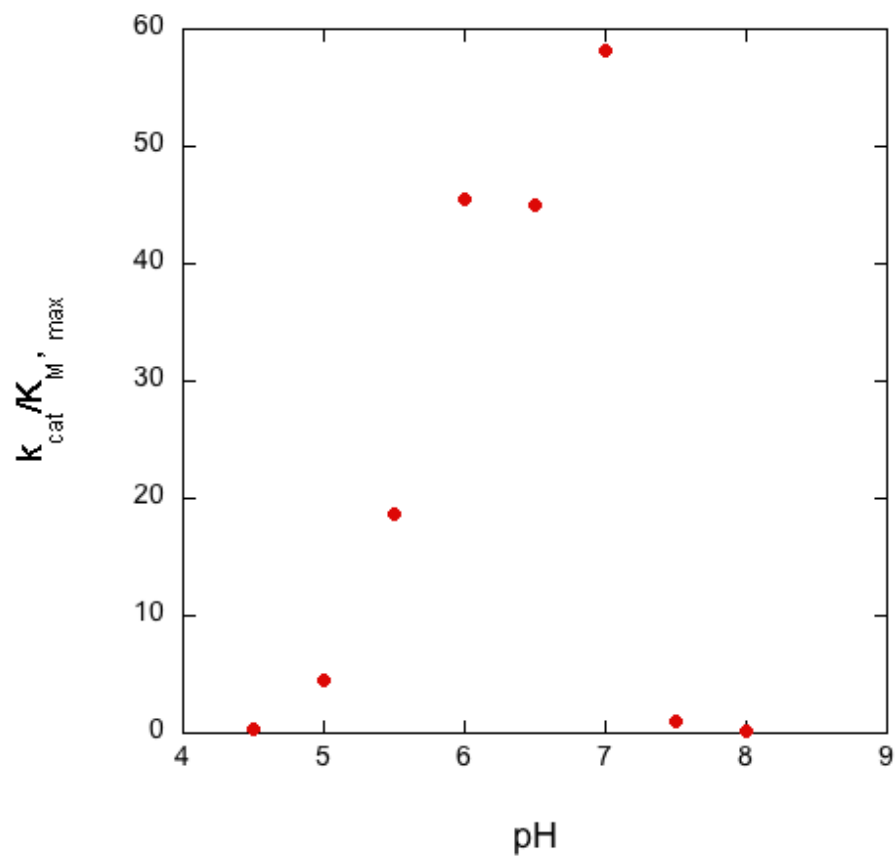


Figure S.17. pH profile of HolIEE (1 μ M) with 80 μ M terbium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used. (AG1049)

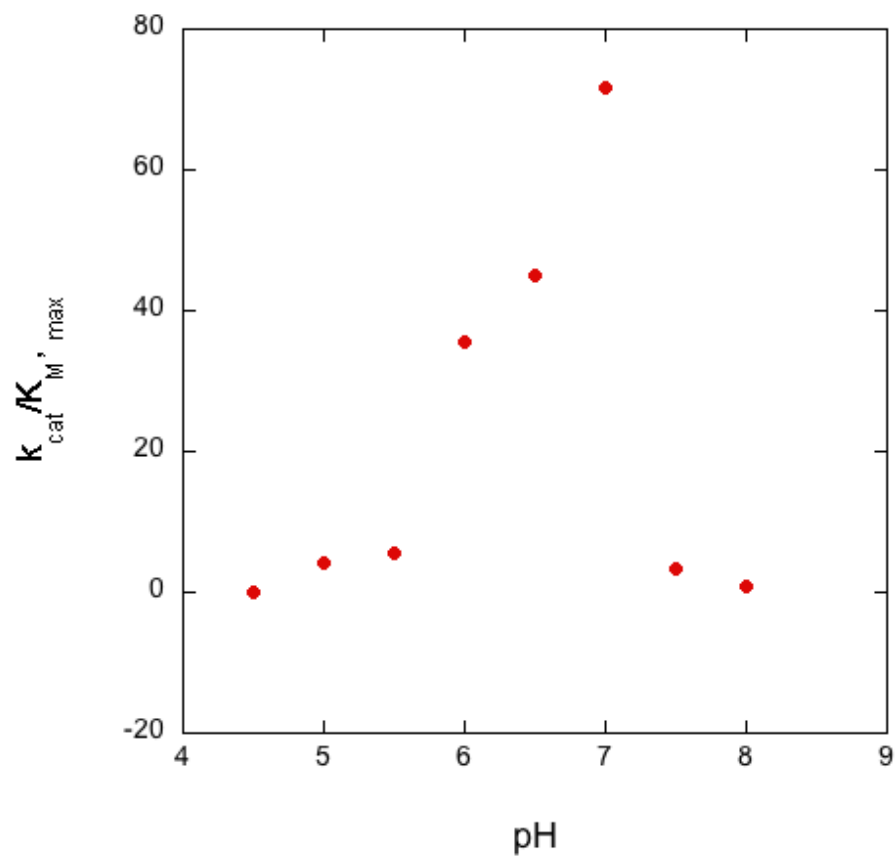


Figure S.18. pH profile of HolIEE (1 μ M) with 80 μ M gadolinium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used. (AG1049)

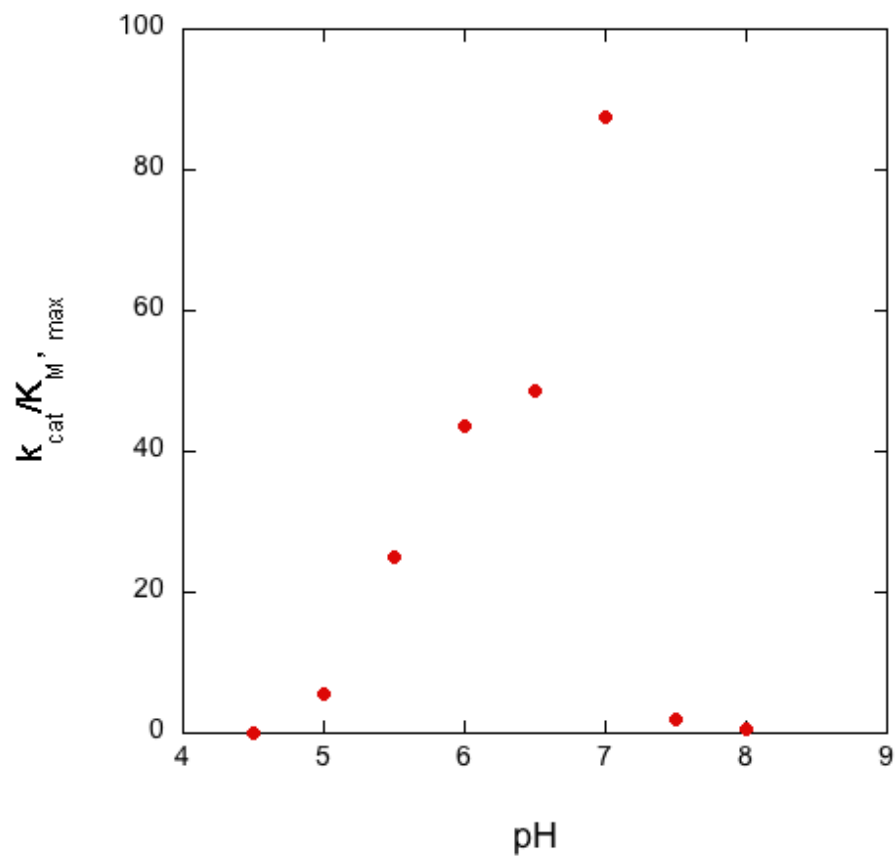


Figure S.19. pH profile of HolIEE (1 μ M) with 80 μ M samarium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

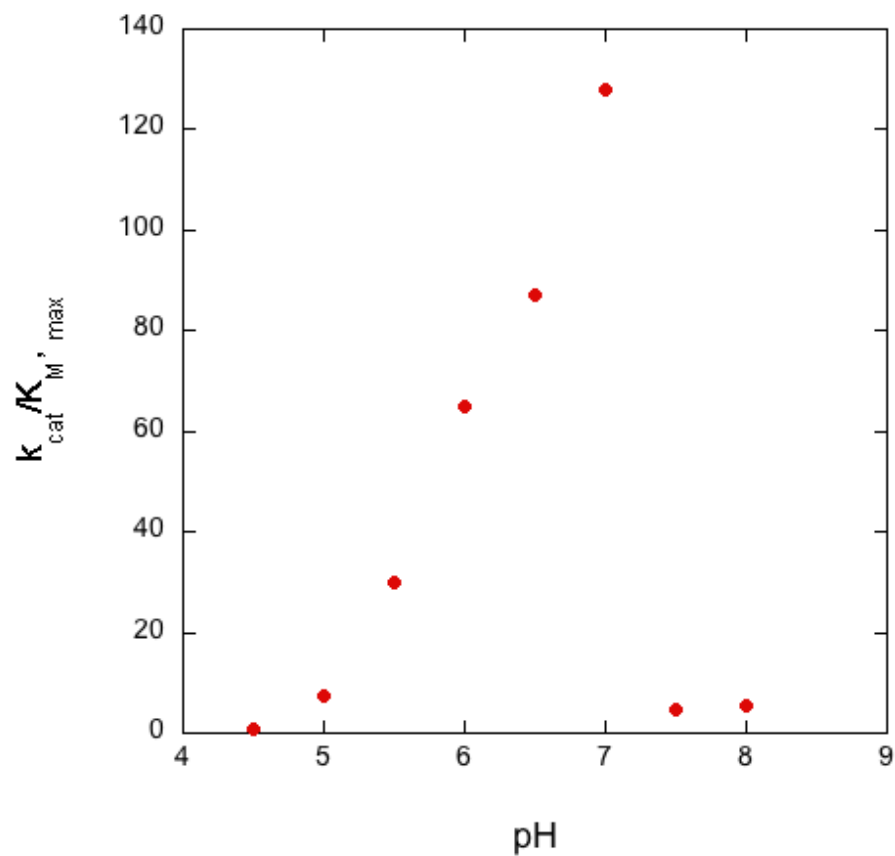


Figure S.20. pH profile of HolIEE (1 μM) with 80 μM neodymium. At different pH, the rate of 600 μM final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

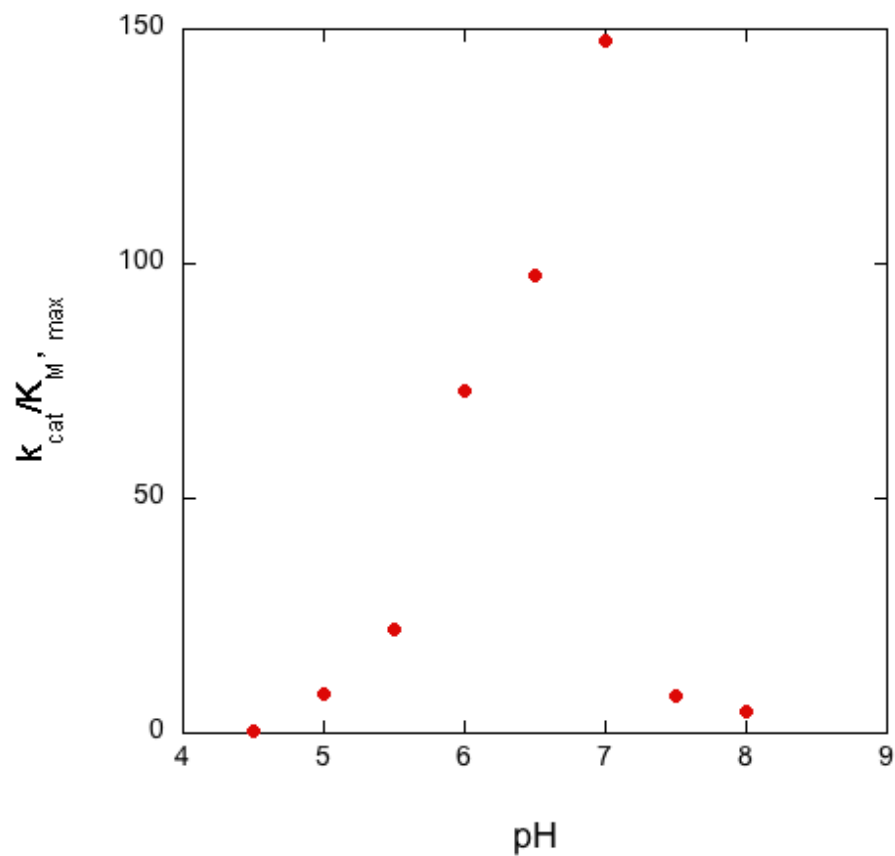


Figure S.21. pH profile of HolIEE (1 μM) with 80 μM praseodymium. At different pH, the rate of 600 μM final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

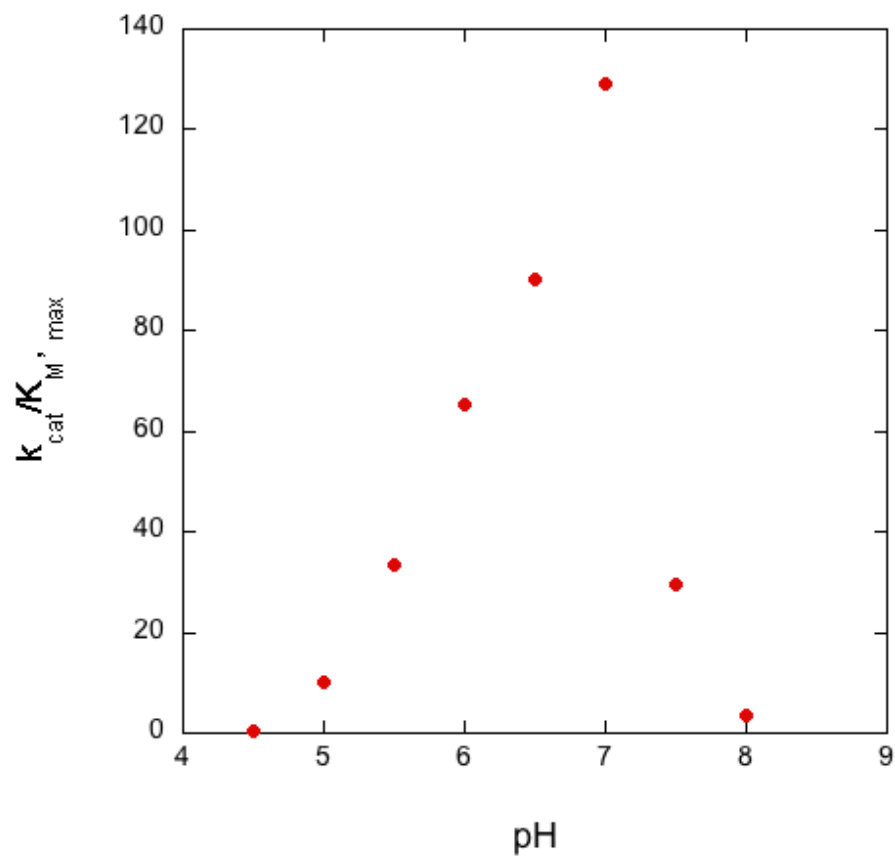


Figure S.22. pH profile of HolIEE (1 μ M) with 80 μ M cerium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

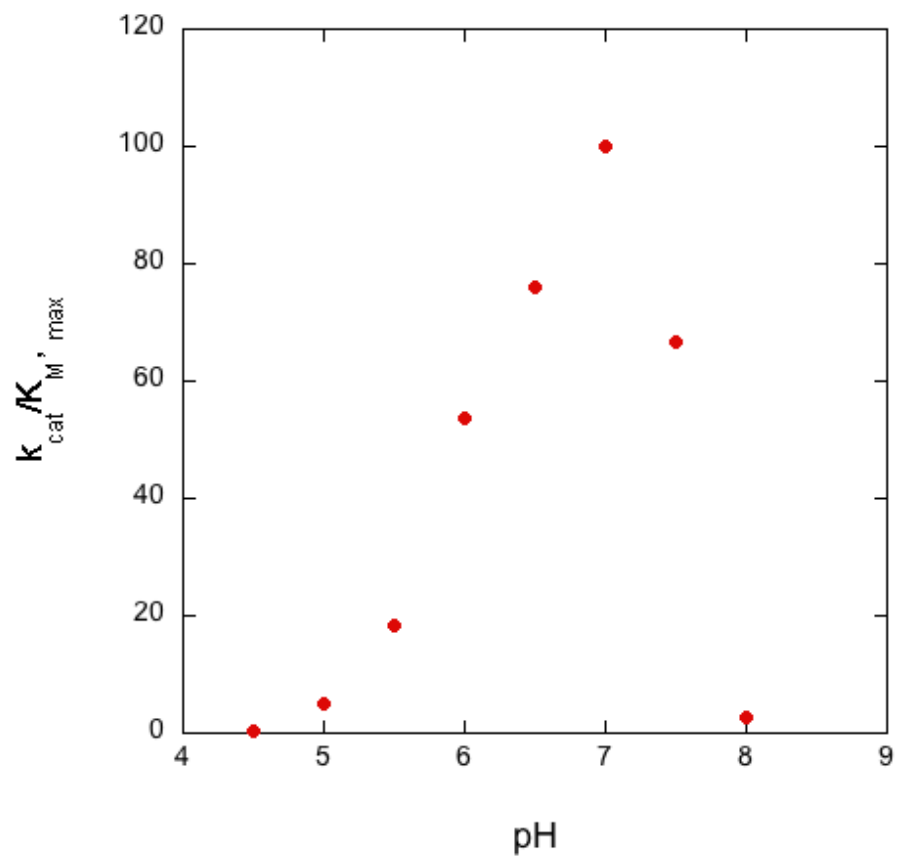


Figure S.23. pH profile of HolIEE (1 μ M) with 80 μ M lanthanum. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

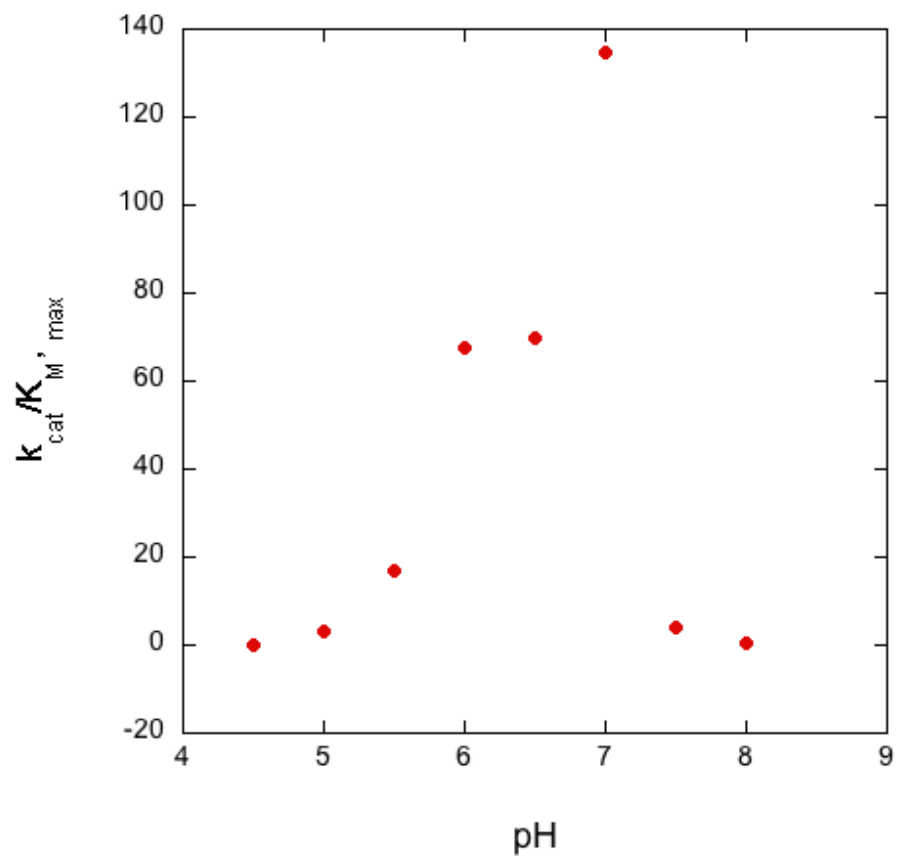


Figure S.24. pH profile of HolIEE (1 μ M) with 80 μ M yttrium. At different pH, the rate of 600 μ M final concentration of Kemp was found. At buffers 4.5-5.5, 20 mM acetate buffer was used. At pH 5.5 and 6.0 20 mM Mes buffer was used. At pH 7.0 and 7.5, 20 mM Mops buffer was used. At pH 8.0, 20 mM Tris buffer was used.(AG1049)

S.2.4. Isothermal Titration Calorimetry

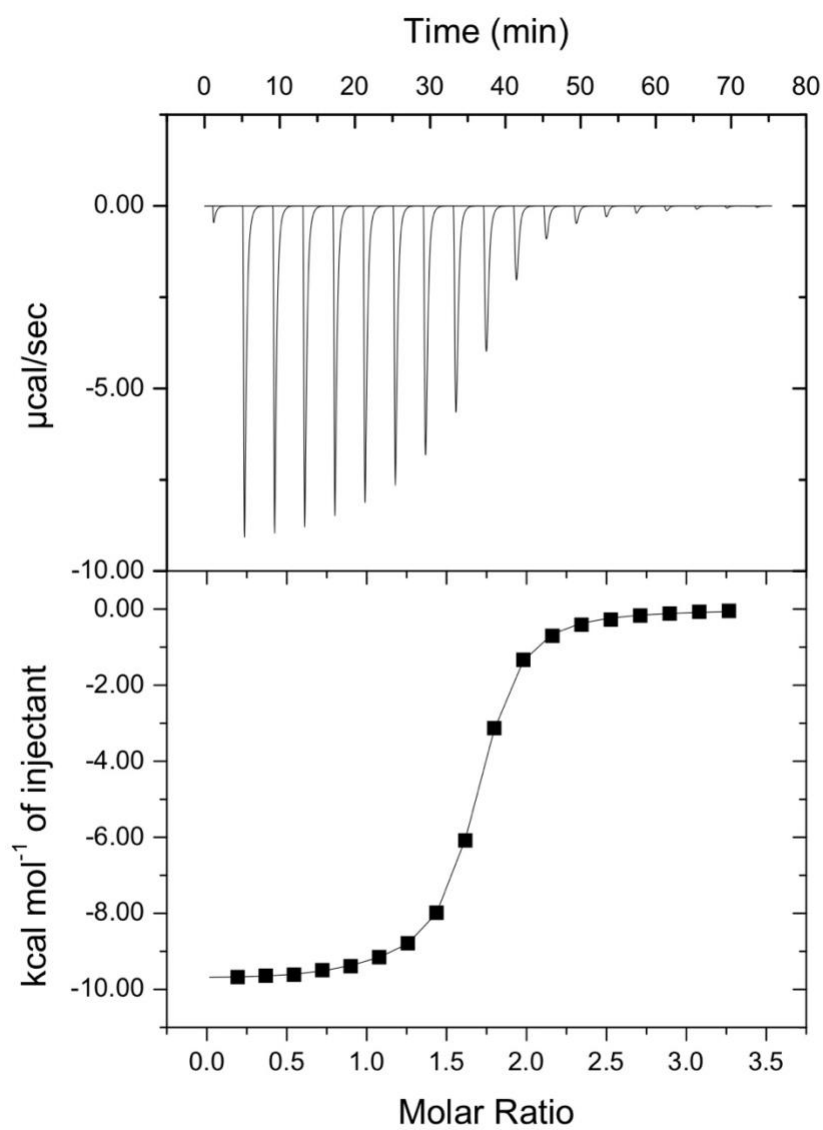


Figure S.25. ITC data of AlleyCat7 titrated with $\text{Ca}(\text{NO}_3)_2$. The sample cell contained 0.1mM Ac7 in 20mM Mops pH 7.0 at 25°C and was slowly titrated with 5mM $\text{Ca}(\text{NO}_3)_2$ in 5 μL injections with an equilibration time of 240 seconds.

S.2.4. Circular dichroism

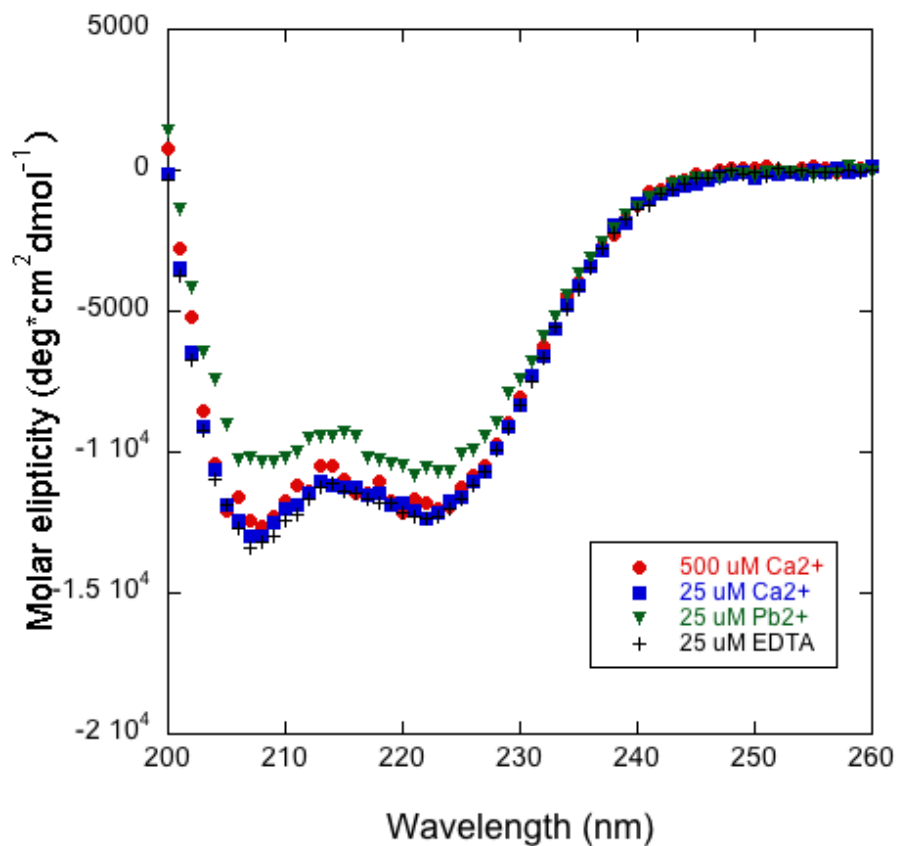


Figure S.26. Circular dichroism spectrum of Ac7 (4.75 μM) with 4mM Hepes pH 7.0 in presence of 500 μM Ca²⁺ (red), 25 μM Ca²⁺ (blue), Pb²⁺ (green), and 25 μM EDTA (black). Ac7 demonstrates an α-helical structure in the presence of all the metal ions tested. Despite no Ca²⁺ ions in the EDTA sample, Ac7 kept its structure. (MT3045)

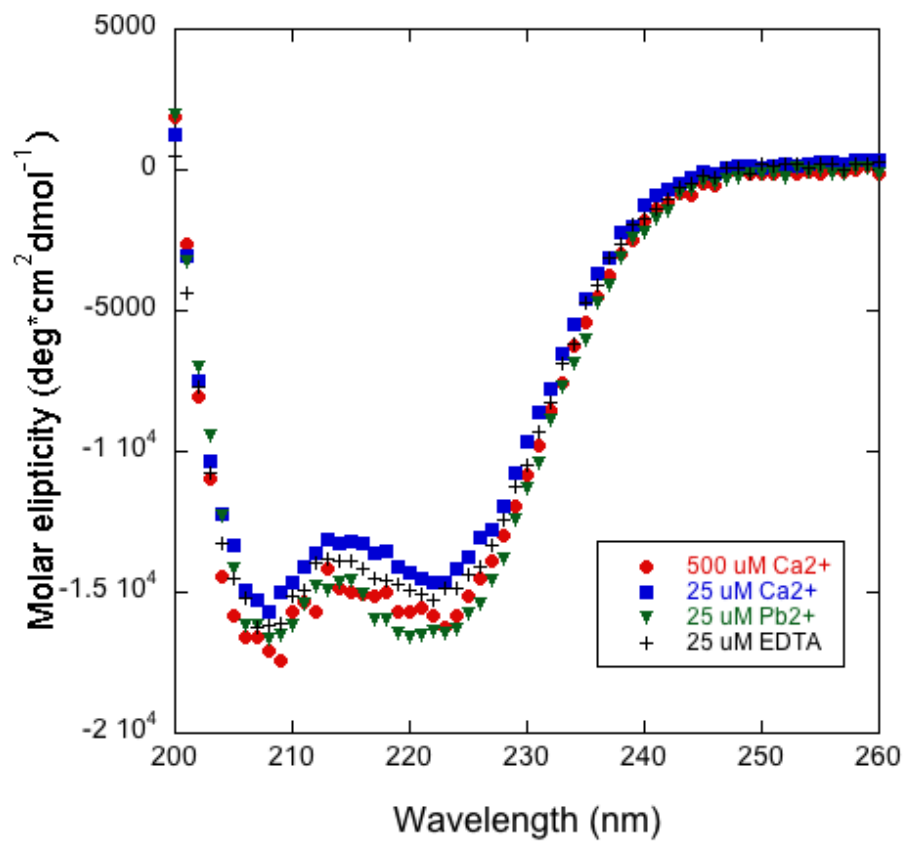


Figure S.27. Circular dichroism spectrum of AcDD (4.75 μM) in 4mM Hepes pH 7.0 in presence of 500 μM Ca^{2+} (red), 25 μM Ca^{2+} (blue), Pb^{2+} (green), and 25 μM EDTA (black). AcDD demonstrates an α -helical structure in the presence of all the metal ions tested as well as EDTA. (MT3045)

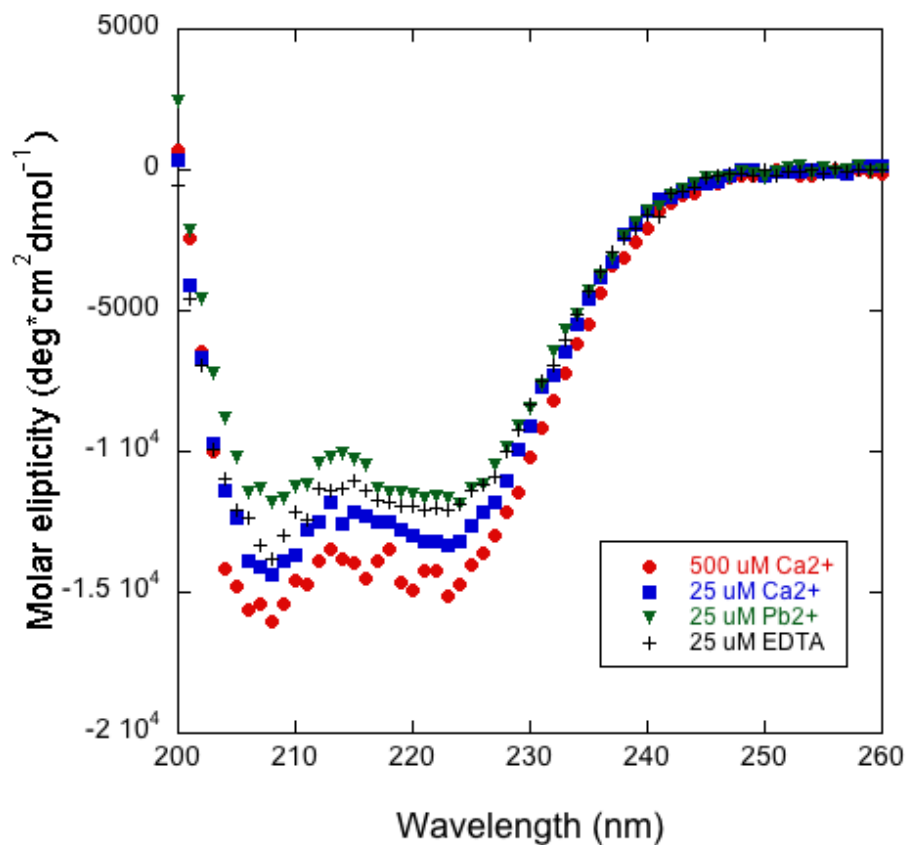


Figure S.28. Circular dichroism spectrum of AcEE (4.75 μM) with 4mM Hepes pH 7.0 in presence of 500 μM Ca²⁺ (red), 25 μM Ca²⁺ (blue), Pb²⁺ (green), and 25 μM EDTA (black). AcEE demonstrates an α -helical structure in the presence of all the metal ions tested as well as EDTA. (MT3045)

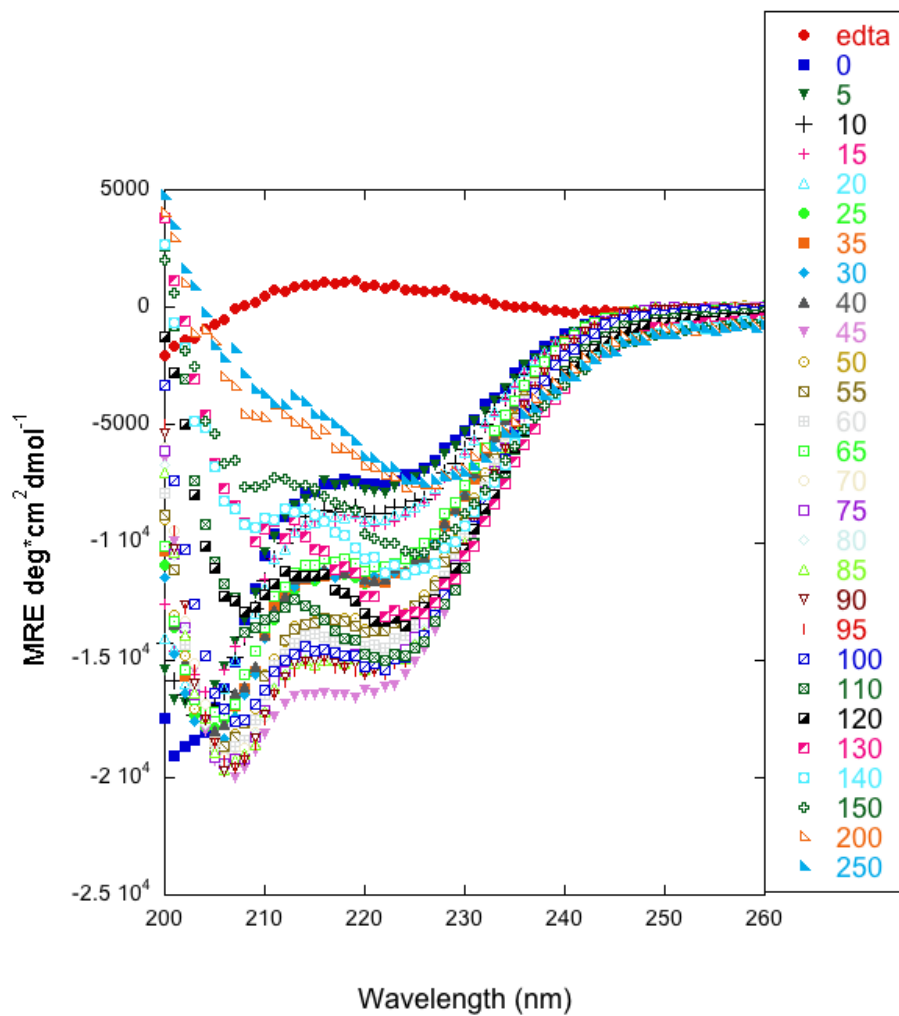


Figure S.29. Circular dichroism of HolIIEE (25 μM) in 4 mM Mops pH 7.0 with YCl_3 titration. Increasing the concentration of metal leads to changes at 207 nm and 220 nm that can be potentially evaluated for K_D . (AG1018)

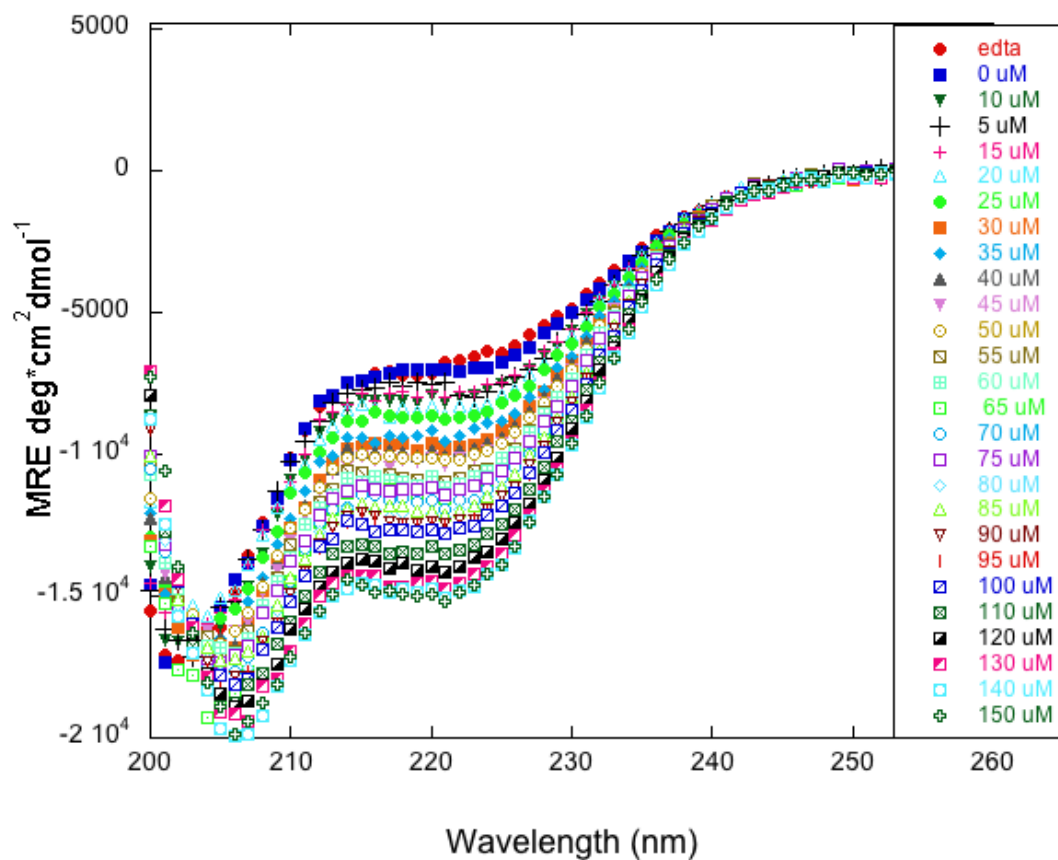


Figure S.30. Circular dichroism of HolIIEE (25 μ M) in 4 mM Mops pH 7.0 with PrCl_3 titration. Increasing the concentration of metal leads to changes at 207 nm and 220 nm that can be potentially evaluated for K_D . (AG1022)

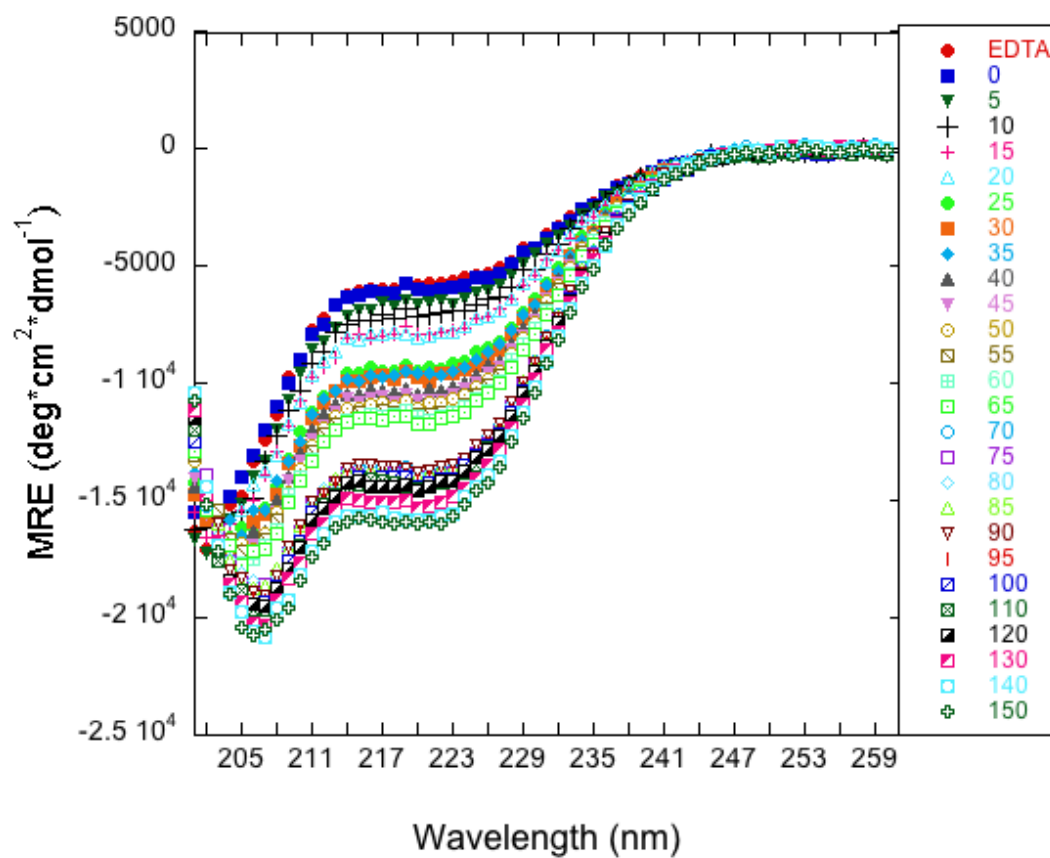


Figure S.31. Circular dichroism of HolIEE (25 μM) in 4 mM Mops pH 7.0 with LaCl₃ titration. Increasing the concentration of metal leads to changes at 207 nm and 220 nm that can be potentially evaluated for K_D. (AG1020)

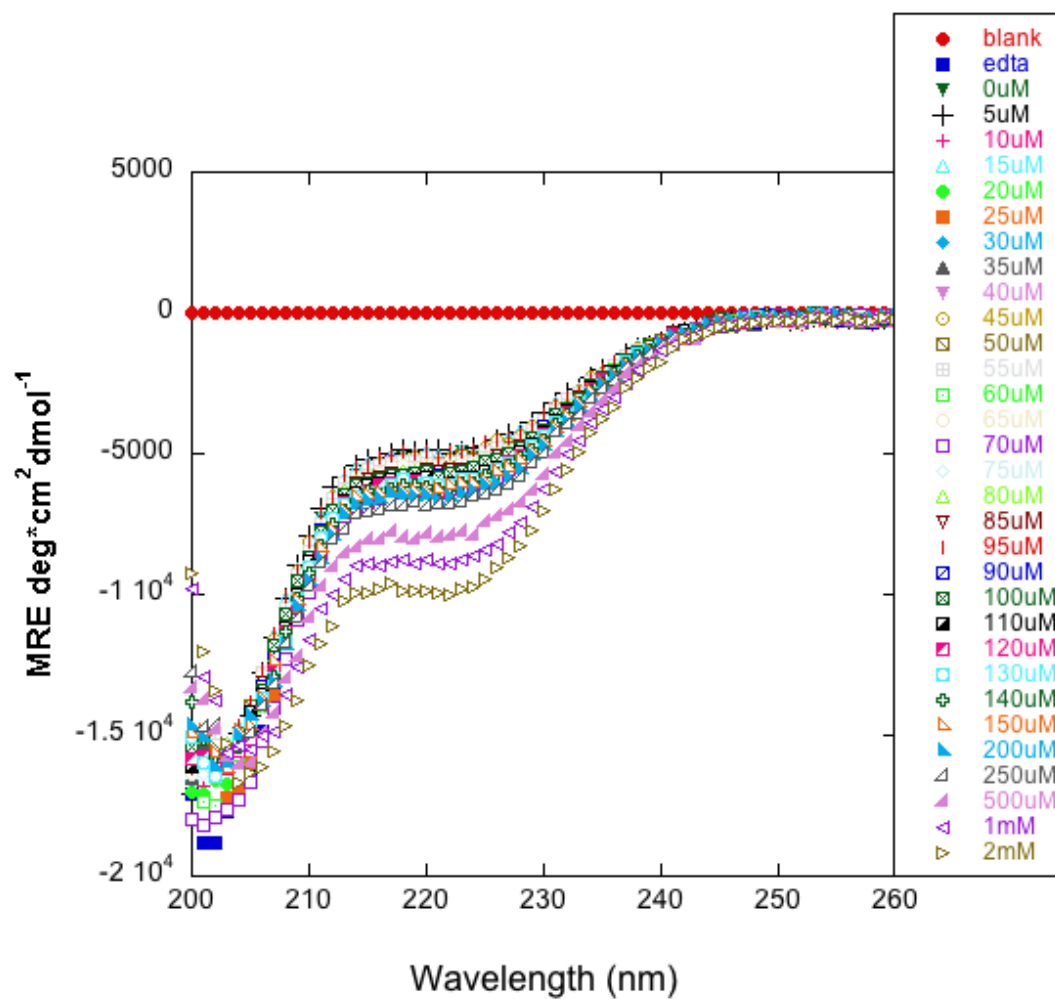


Figure S.32. Circular dichroism of HolIIEE (25 μ M) in 4 mM Mops pH 7.0 with CaCl_2 titration. (AG1025)

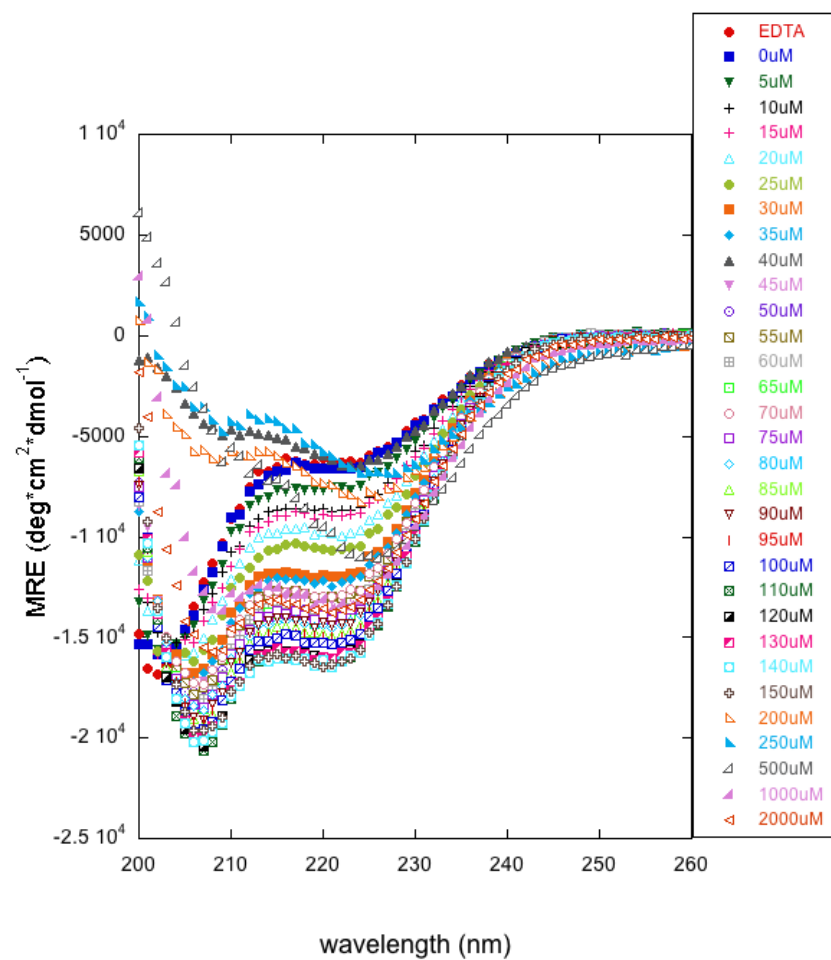


Figure S.33. Circular dichroism of HolIEE (25 μ M) in 4 mM Mops pH 7.0 with CeCl_3 titration. Increasing the concentration of metal leads to changes at 207 nm and 220 nm that can be potentially evaluated for K_D . (MT7012)

S.3. Supplemental data for Chapter 4

S.3.1. Gel Electrophoresis

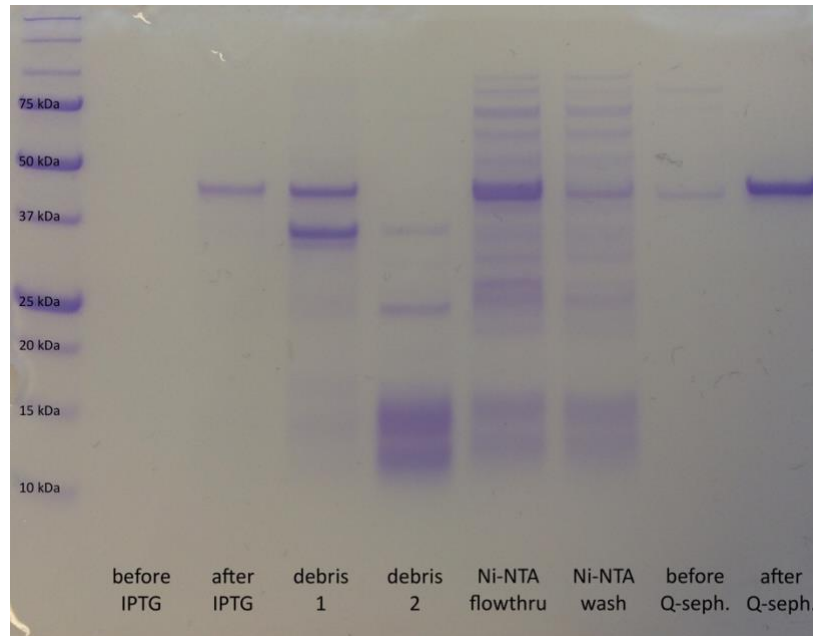


Figure S.34. A 10% acrylamide SDS-PAGE gel of purification of FDH-N expressed in BL21(DE3) *E. coli* cells. Aliquots were taken before IPTG induction, after IPTG induction, from the pellet after sonication and centrifugation (labeled debris 1 on gel), from the pellet after streptomycin sulfate addition and centrifugation (labeled debris 2), from the Ni-NTA column flow through, from the Ni-NTA column wash, and before and after Q-sepharose column. The protein band was ~45 kDa, where it was expected to appear.

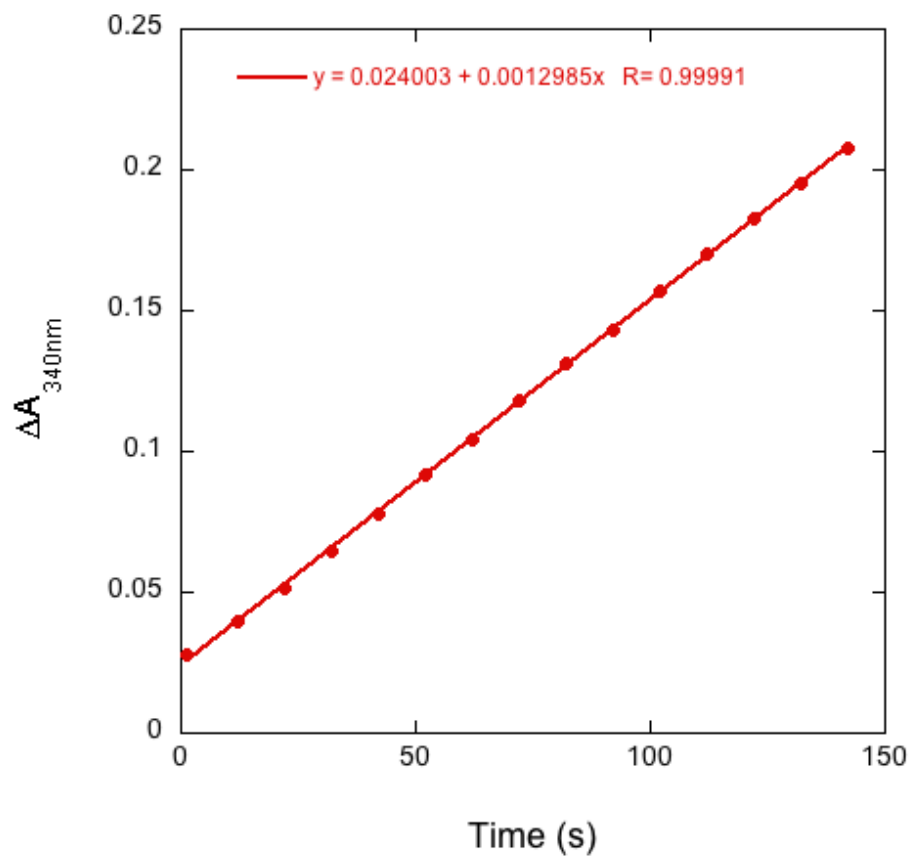


Figure S.35. Catalytic efficiency of FIDH-aeru (10 μ M) measuring the conversion of NAD⁺ (1 mM) to NADH during the oxidation of formaldehyde(100 mM) to formate in 100 mM sodium phosphate buffer at pH 7. The specific activity is 2.6 U/mg. (MT1102)

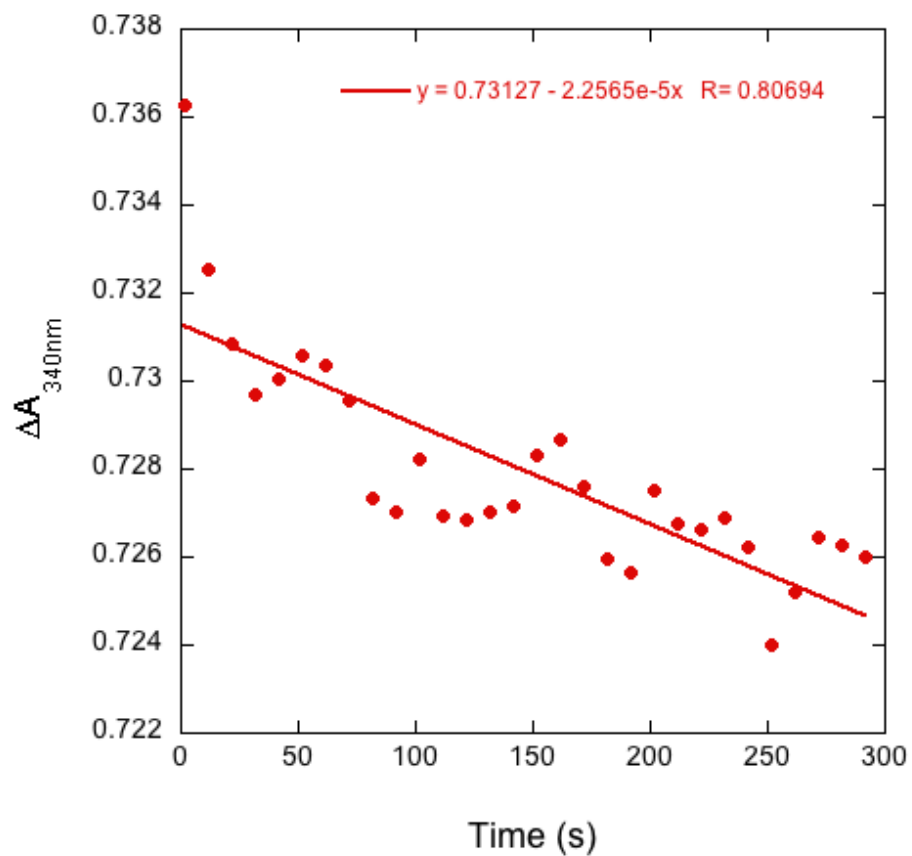


Figure S.36. Catalytic efficiency of FIDH-aeru (1 μ M) measuring the conversion of NADH (0.2 mM) to NAD⁺ during the reduction of formate (160 mM) to formaldehyde in 100 mM sodium phosphate buffer at pH 7. The specific activity is 0.26 U/mg. (MT1102)

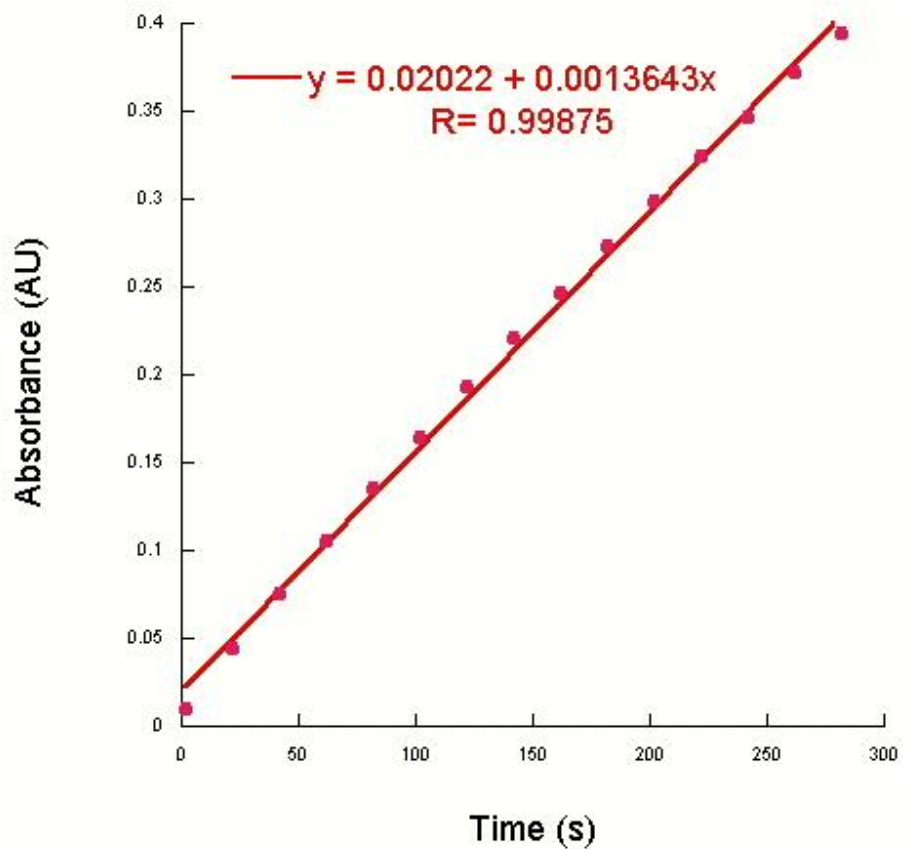


Figure S.37. Catalytic efficiency of FIDH-puti (0.1 μ M) measuring the conversion of NAD⁺ (1 mM) to NADH during the oxidation of formaldehyde (100 mM) to formate in 100 mM sodium phosphate buffer at pH 7. The specific activity is 3.13 U/mg. (MT1146)

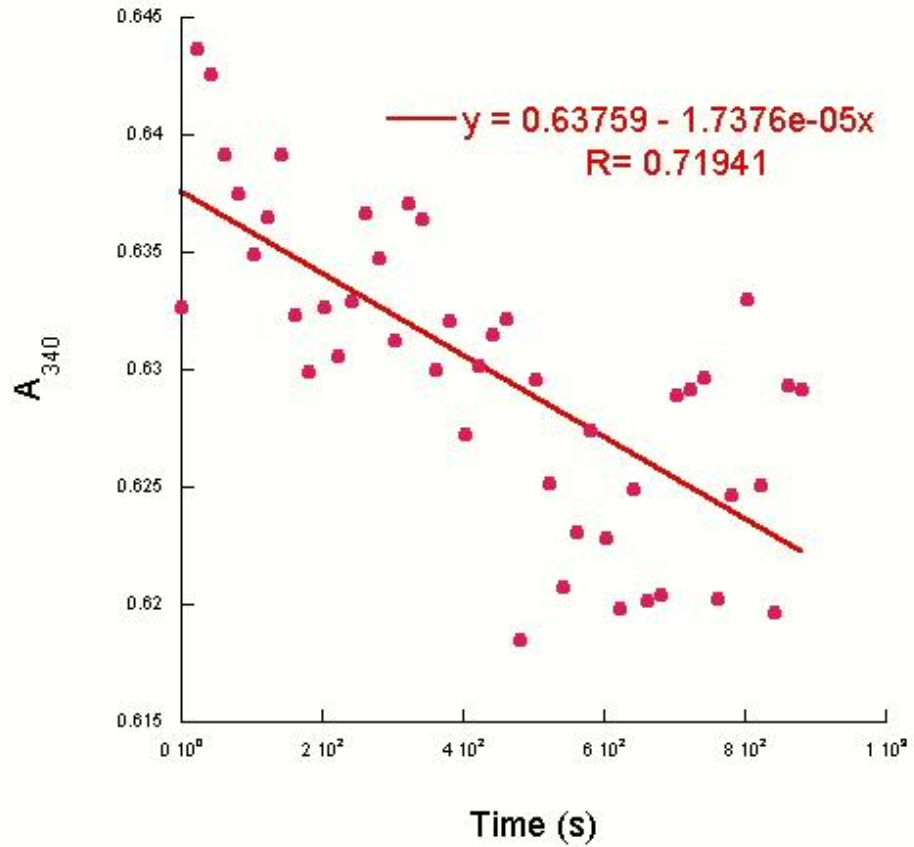


Figure S.38. Catalytic efficiency of FIDH-puti (12 μ M) measuring the conversion of NADH (0.1 mM) to NAD⁺ during the reduction of formate (100 mM) to formaldehyde in 100 mM sodium phosphate buffer at pH 7. The specific activity is 0.0003 U/mg. (MT1146)

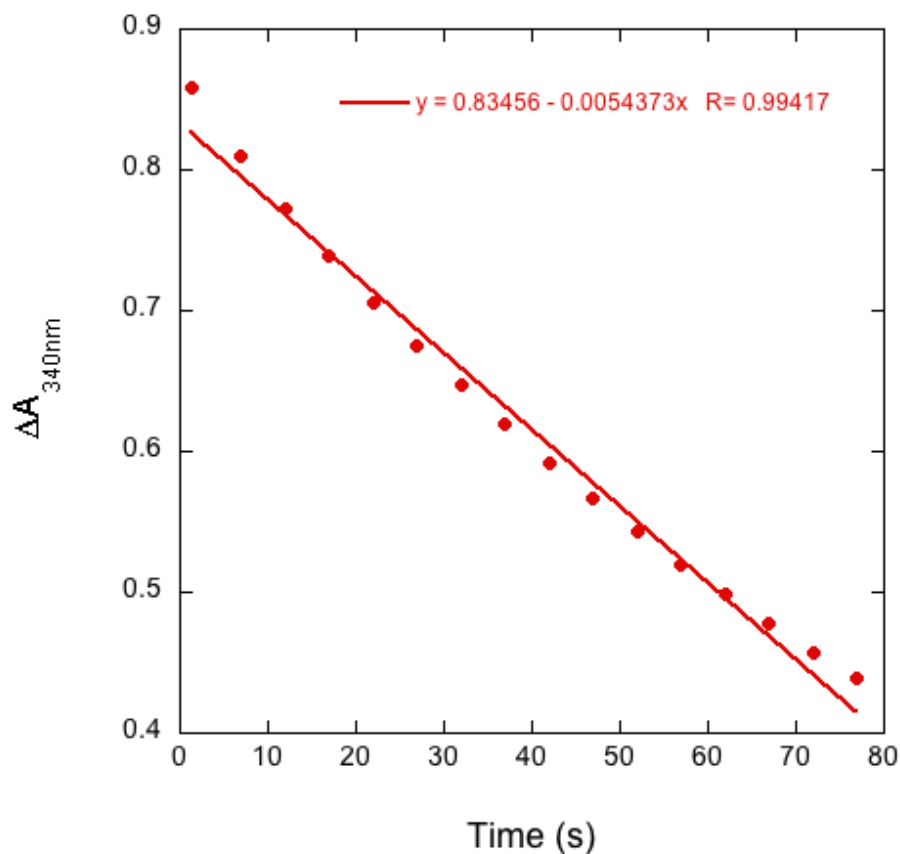


Figure S.39. Catalytic efficiency of ADH (16 nM) measuring the conversion of NADH (0.2 mM) to NAD⁺ during the reduction of formaldehyde (100 mM) to methanol in 100 mM sodium phosphate buffer at pH 7. The specific activity is 84 U/mg. (MT1064)

S.3.2. Tryptic digest

Table S.3.2.1. List of peptide fragments of FDH-C-y after trypsin digestion.

Position of cleavage site	Peptide sequence	Peptide length [aa]	Peptide mass [Da]
2	MR	2	305.395
76	FPSIFTAVLFAASSALAAPVNTTTEDETAQIP AEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFI NTTIAASIAAK	74	7623.502

84	EEGVSLEK	8	889.958
85	R	1	174.203
95	EAEAYVEFMK	10	1216.371
105	IVLVLYDAGK	10	1090.328
112	HAADEEK	7	798.808
120	LYGCTENK	8	927.04
128	LGIANWLK	8	914.115
140	DQGHELITTS DK	12	1343.413
149	EGETSEL DK	9	1007.019
169	HIPDADIIITTPFHPAYITK	20	2263.621
171	ER	2	303.318
174	LDK	3	374.437
176	AK	2	217.268
179	NLK	3	373.453
201	LVVVAGVGS DHIDL DYINQTGK	22	2313.593
202	K	1	146.189
229	ISVLEVTGSNVVSVAEHVVMTMLVLR	27	2882.47
250	NFVPAHEQIINH DWEVAAIAK	21	2402.695
258	DAYDIEGK	8	909.948
267	TIATIGAGR	9	858.993
271	IGYR	4	507.59
275	VLER	4	515.61
282	LLPFNPK	7	828.022
294	ELLYDYQALPK	12	1515.726
299	EAE EK	5	604.615
303	VGAR	4	401.466
304	R	1	174.203
329	VENIEELVAQADIVTVNAPLHAGTK	25	2631.966
334	GLINK	5	543.664
339	ELLSK	5	588.702
341	FK	2	293.366
342	K	1	146.189
351	GAWLVNTAR	9	987.126
370	GAICVAEDVAAALESQ LR	19	1873.111
384	GYGGDVWFPPAPK	14	1518.692
389	DHPWR	5	709.762
392	DMR	3	420.484
394	NK	2	260.293
415	YGAGNAMTPHYS GTTLDAQTR	21	2212.378
421	YAEGTK	6	667.717

431	NILESFFTGK	10	1155.316
435	FDYR	4	599.644
449	PQDIILLNGEYVTK	14	1602.848
453	AYGK	4	437.496
456	HDK	3	398.419
457	K	1	146.189
471	GENLYFQSHHHHHH	14	1779.854

S.4. Supplemental data for Chapter 5

S.4.1. Gel electrophoresis

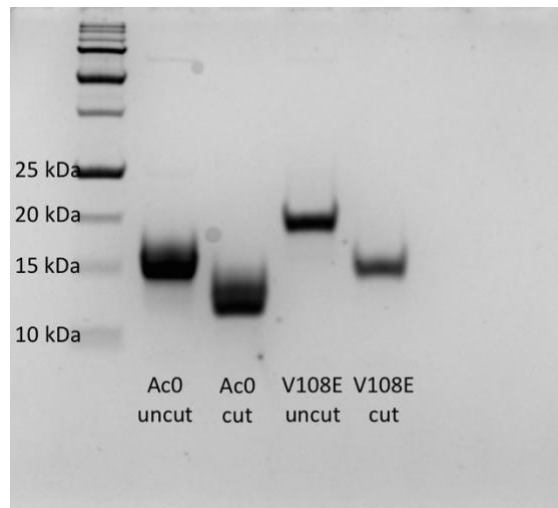


Figure S.40. 15% acrylamide SDS-PAGE gel of Ac0 and Ac0-V108E purification. The expected MW of Ac0 and Ac0-V108E uncut is ~12 kDa and the expected MW of the cut proteins is ~8 kDa.

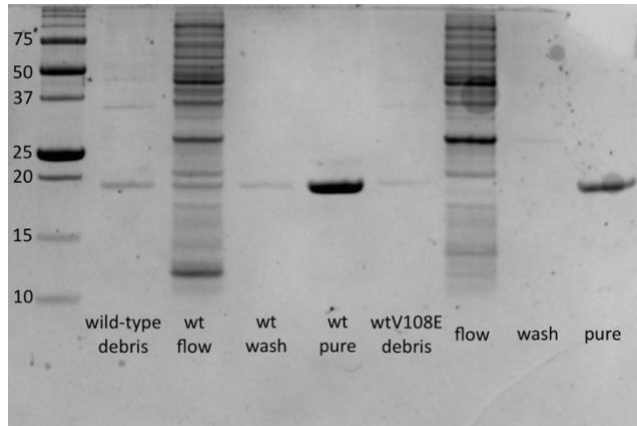


Figure S.41. 15% acrylamide SDS-PAGE gel of wild-type (wt; cCam) and wtV108E (cV108E). Debris refers to the pellet left behind after sonication and centrifugation, and flow and wash refer to steps in Ni-NTA purification. The expected masses of the uncut proteins is ~12 kDa.

Appendix III. Published Works

Appendix IV. Curriculum Vitae

Original Article

Secretion of functional formate dehydrogenase in *Pichia pastoris*

Michelle Takacs, Olga V. Makhlynets, Patricia L. Tolbert, and Ivan V. Korendovych*

Department of Chemistry, Syracuse University, 111 College Place, Syracuse, NY 13244, USA

*To whom correspondence should be addressed: E-mail: ikorendo@syr.edu

Edited by: Dagmar Ringe

Received 17 October 2016; Revised 14 January 2017; Editorial Decision 18 January 2017; Accepted 25 January 2017

Abstract

Biofuels are an important tool for the reduction of carbon dioxide and other greenhouse emissions. NAD⁺-dependent formate dehydrogenase has been previously shown to be capable of the electrochemical reduction of carbon dioxide into formate, which can be ultimately converted to methanol. We established that a functional enzyme, tagged for immobilization, could be continuously secreted by *Pichia pastoris*. The protein can be easily separated from the growth media and its activity remains constant over an extended period of time. This is an important first step in creating a self-sustaining system capable of producing biofuels with minimal resources and space required.

Key words: Formate dehydrogenase, *Pichia pastoris*, protein engineering

Introduction

Enzymes are capable of facilitating chemical reactions with exceptional versatility and efficiency. While the repertoire of enzymes is vast, many practically useful reactions still cannot be catalyzed with sufficient efficiency. Moreover, even if an enzyme is known to catalyze a particular chemical transformation well, production cost and stability issues may preclude its use in laboratory and industry.

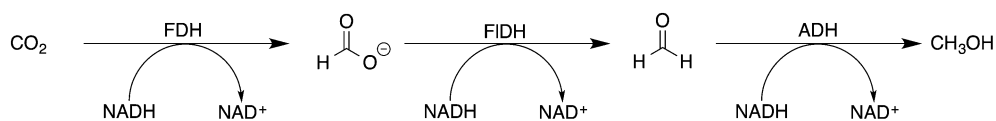
Biofuel production is an important strategy for keeping CO₂ greenhouse gas emissions to a minimum. Most biofuel in the USA is produced by fermentation of sugars obtained from corn (Bhatia *et al.*, 2012; Kondratenko *et al.*, 2013). However, making land useable for these crops requires clearing current flora by either burning or microbial decomposition, both of which release more CO₂ into the atmosphere than the produced biofuels can compensate for (Fargione *et al.*, 2008). While biofuels are important for the future of environmental protection, alternative, more sustainable methods of production are urgently needed.

Capture of CO₂ by employing enzymes provides a very appealing opportunity. Direct capture of carbon dioxide is the ultimate carbon neutral process; however, apart from Rubisco, which is the primary tool for CO₂ capture in plants, there are no other known means for enzymatic capture of carbon dioxide. Initial fixation of carbon

dioxide in solution is critically important; once the gas is trapped, it could be fed into multiple one carbon metabolic pathways.

Immobilized nicotinamide adenine dinucleotide (NAD)-dependent formate dehydrogenase (FDH, EC 1.2.1.2) has been recently used to electrochemically reduce carbon dioxide to formate, which can be in turn enzymatically reduced to methanol, a usable biofuel, by utilizing NAD⁺-dependent formaldehyde dehydrogenase (FIDH, EC 1.2.1.1) and NAD⁺-dependent alcohol dehydrogenase (ADH, EC 1.1.1.1) (Scheme 1) (Kim *et al.*, 2013; Liu *et al.*, 2013). The native function of all these enzymes is to produce NADH by oxidation of their respective substrates: formate, formaldehyde and methanol. The overall strategy is based on the fact that catalysts equally facilitate both forward and reverse reactions, thus given excess of the product and sufficient reducing power, NAD⁺-dependent dehydrogenases can reduce carbon dioxide to methanol.

While partial hydrogenation of carbon dioxide has been accomplished through multiple pathways such as heterogeneous catalysis, electrocatalysis and photocatalysis, the enzymatic process is advantageous due to its ability to operate under mild conditions (Obert and Dave, 1999; Cazelles *et al.*, 2013; Aresta *et al.*, 2014). However, this multi-step reduction of carbon dioxide faces its own challenges, such as a low pH and elevated temperature required for maximum efficiency, which can cause the enzymes to degrade (Baskaya *et al.*, 2009).



Scheme 1 Overview of the proposed bioengineered pathway for the production of methanol from CO₂ utilizing a multi-step enzyme pathway.

Our goal is to develop an economic and renewable approach to produce active dehydrogenase through continuous secretion in yeast. The yeast will continuously secrete the protein into media, which eliminates the time-consuming step of growing, expressing and extensive purification of the desired proteins from *Escherichia coli*. Yeast will also replace enzymes as they degrade and extend the lifetime of any device engineered to capture CO₂. We chose *Pichia pastoris*, a methylotrophic yeast capable of expressing and secreting non-native proteins, due to its many important benefits: (i) recombinantly produced enzymes (including large proteins with molecular weight >50 kDa) can be secreted; (ii) *P. pastoris* is tolerant to high concentrations of methanol; (iii) methanol can serve as the only energy source for the culture, eliminating the need for complex media formulations that would increase production costs; (iv) a typical *P. pastoris* strain would use methanol as an inducer for recombinant production of proteins; (v) *P. pastoris* cultures grow to high-density improving the costs and diminishing the size of fermentors and (vi) *P. pastoris* generally does not secrete its own proteins, simplifying downstream purification of the secreted dehydrogenases (Kovar et al., 2010).

Development of an efficient secretion system for functional enzymes is commonly plagued with difficulties related to posttranslational processing (most often glycosylation) of proteins in yeast that may interfere with activity. Therefore, optimization of expression is often necessary. Additionally, we set on determining the optimal position of the affinity tag necessary for attaching the protein to solid support.

We focused our efforts on FDH from yeast *Candida boidinii*, which has been previously structurally and functionally characterized (Slusarczyk et al., 2000; Schirwitz et al., 2007). In this work, we established that a tagged, functional enzyme could be continuously secreted by *P. pastoris*. The protein can be easily separated from the growth media using Ni-NTA resin and enzyme's activity remained constant even after prolonged growth at 29°C.

Materials and methods

Cloning of FDH gene into pMCSG49 and pET28a vectors

cDNA of *C. boidinii* FDH (without His₆-tag) was synthesized by GenScript. To obtain FDH with N-terminal His₆-tag (FDH-N), FDH gene was cloned into pMCSG49 vector (DNASU plasmid repository) through ligation-independent cloning site (Stols et al., 2002). To obtain FDH with C-terminal His₆-tag (FDH-C), the FDH gene was cloned into pET28a (Novagen) using *SalI* and *NcoI* restriction enzymes. The DNA sequences were confirmed using Sanger sequencing.

Expression of FDH-N and FDH-C in *E. coli* and protein purification

The appropriate vectors containing the genes of interest were transformed into *E. coli* BL21(DE3) cells and plated on Luria-Bertani (LB) agar plates containing ampicillin (FDH-N) or kanamycin (FDH-C).

For the FDH seed culture, LB (50 ml) supplemented with antibiotic (100 µg/µl Amp for FDH-N or 50 µg/µl Kan for FDH-C) was inoculated with a single colony and incubated overnight at 37°C

with shaking at 230 r.p.m. The seed culture (10 ml) was then diluted into LB (1 l) supplemented with antibiotic and the culture was grown at 37°C. When the culture reached an OD₆₀₀ of ~0.6, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM. FDH was expressed for 4 h at 30°C and the cells were then collected by centrifugation (4000 g × 20 min), flash frozen in liquid nitrogen and stored at -80°C. The typical yield of wet cell paste was 4 g per liter of culture.

Cells were resuspended in working buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HEPES, 5% glycerol, pH 7.6) at a ratio of 5 ml of buffer for every 1 g of cells. Phenylmethanesulfonyl fluoride (PMSF) was added to a final concentration of 0.5 mM. Cells were lysed by sonication on ice for 10 min (20 s lyse cycle, 20 s rest), and then the crude cell lysate was centrifuged at 20 000 g for 30 min. To precipitate DNA, a solution of streptomycin sulfate in working buffer was added to the supernatant on ice to a final concentration of 1% w/v and the mixture was allowed to stir for 15 min. Precipitate was pelleted by centrifugation at 20 000 g for 30 min. The supernatant was applied onto a Ni-NTA column (2 ml, Clontech) and washed with 2 column volumes (CV) of working buffer. Column was washed with 30 mM imidazole in working buffer to remove impurities, and then FDH was eluted with 200 mM imidazole in buffer (pH 7.6). Fractions containing protein were identified using Pierce BCA Protein Assay kit (Thermo Scientific). Combined protein fractions were diluted two-fold and applied onto a Q Sepharose Fast Flow column (6 ml, GE Healthcare), then washed with working buffer. Protein was eluted with 300 mM NaCl in working buffer. The purity (>95%) of the final protein was checked on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and the protein was concentrated to 500 µM using a 30 kDa MWCO spin filter (Vivaspin, Sartorius). Protein concentrations were determined by measuring the absorbance at 280 nm using the calculated (ExPASy) extinction coefficient 51 465 M⁻¹ cm⁻¹ (FDH-C) and 52 955 M⁻¹ cm⁻¹ (FDH-N).

Cloning of FDH into pPIC9

Sequences coding for FDH-N and FDH-C were cloned into pPIC9 vector (Invitrogen) using *EcoRI* and *NotI* restriction sites. This procedure inserts FDH gene in frame with the secretion signal open reading frame. Gene product is a fusion of α-factor with Glu-Lys-Arg*Glu-Ala-Glu-Ala at the end, followed by FDH protein. The preliminary cleavage of the signal sequence occurs between arginine and glutamine and is denoted with an asterisk. Glu-Ala repeats can be further cleaved off. FDH gene flanked by *EcoRI* and *NotI* restriction sites and empty pPIC9 vector were digested using *EcoRI* and *NotI* restriction enzymes (New England Biolabs; NEB). The digested products were purified using Nucleospin Extract II (BioBasic) kit.

The digested pPIC9 vector and FDH gene were mixed in 1:8 vector to gene ratio (based on DNA concentration in ng/µl) and then ligated using T4 DNA Ligase (NEB). The mixture was incubated at room temperature for 30 min and transformed into *E. coli* XL-10 cells. DNA was purified using EZ-10 Spin Column Plasmid DNA kit (BioBasic) and the sequence confirmed by Sanger sequencing.

Transformation of pPIC9 into *P. pastoris*

Transformation into *P. pastoris* was done using previously published protocol with minor modifications (Wu and Letchworth, 2004). The pPIC9 vector containing the genes of interest was cut using *SalI* restriction enzyme (NEB). Cut vector was purified using Nucleospin Extract II kit protocol.

Pichia pastoris GS115 yeast culture was started from a glycerol stock by inoculating 2 ml of yeast extract peptone dextrose (YPD) media and the culture was grown at 28.5°C with shaking at 230 r.p.m. overnight. Next day this culture was transferred into 100 ml of YPD and grown until OD₆₀₀ reached 1.5. The cells were pelleted at 3000 g and then resuspended in ice-cold sterile LiAc-DTT buffer (100 mM lithium acetate, 10 mM dithiothreitol; DTT, 600 mM sorbitol, 10 mM Tris pH 7.5) and allowed to incubate at room temperature for 40 min. Cells were centrifuged at 3000 g and the pellet was washed twice with 1 ml of ice-cold sterile 1 M sorbitol. Cells were resuspended in 400 µl of 1 M sorbitol to reach OD₆₀₀ between 100 and 200. Digested plasmid (3.5 µg) was mixed with GS115 cells (200 µl), the mixture was incubated on ice for 10 min and then transferred into Electroporation cuvette (2 mm electrode gap, 450 µl, Bulldog Bio). The electroporation pulse was applied to the cells at 1.5 kV using an Eppendorf 2510 electroporator. After transformation, the culture was immediately diluted 10-fold using 1 M sorbitol and variable amounts were plated on Regeneration Dextrose Base (RDB-) plates.

The same transformation procedure was repeated using empty pPIC9 digested with *SalI*. This step generated *P. pastoris* strain that does not have any proteins attached to α -factor. The strain was used to evaluate background secretion.

Screening for *P. pastoris* Mut⁺ phenotype (ability to grow on methanol as a sole carbon source)

Colonies that successfully grow on RDB- plates are assumed to have undergone successful homologous recombination upon transformation. To determine if transformed *P. pastoris* was capable of growing in the presence of methanol, several colonies from RDB- plates were streaked on Minimal Methanol (MM) plates and the plates were allowed to incubate for 3–4 days at 30°C. A colony from MM plate was used to prepare glycerol stock, 50 µl aliquots were flash frozen using acetone/dry ice bath and kept at –80°C until needed.

To confirm the presence of FDH gene, a single colony was used to inoculate 20 µl of sterile H₂O. The yeast cells were lysed by freezing the inoculated solution for 1 min in liquid N₂ followed by thawing for 1 min in a 42°C water bath. This process was repeated 15–20 times for best results. For PCR screening, 10 µl of lysate was mixed with 1.2 µl α -factor forward primer (Integrated DNA Technologies; IDT), 1.2 µl 3'AOX1 reverse primer (IDT) and 12 µl GoTaq DNA Polymerase master mix (Promega). After PCR, samples were analyzed using a 1% agarose DNA gel run in TAE buffer. A DNA band for the FDH gene was present at ~1.2 kb.

Expression and purification of FDH secreted from *P. pastoris*

A small volume of YPD (2 ml) was inoculated with glycerol stock (10 µl). After growing the culture for 16 h, the cells were pelleted at 2000 g and resuspended in 20 ml of buffered methanol-complex medium (BMMY; 1% yeast extract, 2% peptone, 1.34% yeast nitrogen base, 4 × 10⁻⁵ biotin, 100 mM sodium phosphate pH 6.0, 0.5% methanol) and allowed to grow for 120 h at 28.5°C. Every 24 h, an aliquot was taken for analysis and methanol was added to BMMY to a final concentration of 0.5%.

The cells were removed by centrifugation at 5000 g for 2 min and the media was decanted. After pH of the media was raised to approximately pH 7–8 using NaOH, it was added to Ni-NTA resin (2 ml) pre-equilibrated with working buffer (50 mM HEPES, 5% glycerol, pH 7.6). The mixture was then rocked on ice for 2 h. Media-resin mixture was washed with working buffer (~100 ml buffer for every 1 ml of resin). Protein was eluted with 300 mM imidazole in working buffer (pH 7.6). The purity (>95%) of the final protein was checked on SDS-PAGE. The protein was concentrated using a 30 kDa MWCO spin filter (Vivaspin, Sartorius) and the imidazole concentration adjusted to ~3 mM. Protein concentrations were determined by measuring the absorbance at 280 nm using the calculated (ExPASy) extinction coefficient 54 445 M⁻¹ cm⁻¹ (FDH-C and FDH-N).

Kinetic assays

All kinetic measurements were done at room temperature (22°C) on an Agilent 8453 UV-Vis Spectrophotometer monitoring absorbance at 340 nm with a background absorbance at 800 nm. Enzyme activity was characterized in a buffer containing 100 mM sodium phosphate (pH 7), 1.0 mM NAD⁺, 160 mM formate. Unless specifically stated, the final concentration of proteins in the enzymatic assay was 0.5 µM FDH and kinetic measurements were recorded every 20 s over a period of 5 min. Change in absorbance at 340 nm was fit to a linear model. One unit of activity (U) is defined as the production/consumption of 1 µmol of NADH per min. The specific activity of enzyme was calculated with the following equation:

$$\text{specific activity} = \frac{\text{slope} * V_f}{6220 * m_f},$$

where V_f is the final volume of the enzymatic assay and m_f is the mass (mg) of FDH enzyme.

Tryptic digestion

For tryptic digestion, 50 µl of protein stock (~0.1 µg protein) was mixed with 50 µl buffer (6 M urea, 100 mM Tris, pH 7.8) and 5 µl of reducing agent (200 mM DTT, 100 mM Tris, pH 7.8) and allowed to sit at room temperature for 1 h. To alkylate cysteines, 20 µl of alkylating agent (200 mM iodoacetamide, 100 mM Tris, pH 7.8) was added and protein mixture was allowed to sit at room temperature for 1 h. To consume unreacted iodoacetamide, 20 µl of reducing agent was added to the mixture. The sample was then diluted with water (775 µl) and 100 µl of porcine trypsin (SAFC, 200 ng/µl final concentration, 100 mM Tris, pH 7.8) was added. Protein mixture was allowed to digest overnight at 37°C. To stop the reaction, acetic acid was added to the mixture until pH was below six.

Mass spectrometry

The molecular weights of the purified proteins were confirmed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry using a Bruker Autoflex III Smartbeam MALDI-TOF mass spectrometer. Purified protein samples and tryptic digest samples (~10 µM concentration) were mixed with saturated solution of sinapic acid in a 1:2 volume ratio. Ovalbumin (molecular weight; MW of 44.3 kDa) was used as an external standard.

Protein sequences of the constructs used in this work

FDH-N in pPIC9 vector (MW = 43 915 Da $\epsilon_{280} = 54 445 \text{ M}^{-1}$)
EAEAYVEFH HHHHSSGV DL GTENLYFQSN AMKIVLVLYD
AGKHAAD E EK LYGCTENKLG IANWLKDQGH ELITSDKEG

ETSELDKHIP DADIIITPPF HPAYITKERL DKAKNLKLVV VAG VGSDDHD LDYINQTKGK ISVLEVTGSN VVSVAEHVVM TML VLVRNFV PAHEQIINH D WEVAIAKDA YDIEGKIAT IGAGR IGYRV LERLLPFPNK ELLYDYDQAL PKEAEKVGVA RRVENIEELV AQADIVTVNA PLHAGTKGLI NKELLSKFKK GAWLVNTARG AICVAEDVAA ALESGQLRGY GGDVWFPQPA PKDHPWRDMR NKYGAGNAMT PHYSGTTLDA QTRYAEGTKN ILESFFTGFK DYRPQDIILL NGEYVTKAYG KHDKK.

FDH-N in pMCSG49 vector (MW = 43 108 Da $\epsilon_{280} = 52\,955\text{ M}^{-1}$)
 HHHHHHSSGV DLGTENLYFQ SNAMKIVLVL YDAGKHADE EKLYGCTENK LGIANWLKDDQ GHELITTSK EGETSELDKH IP DADIIIT PPFPAYITKE RLDKAKNLKL VVAVGVSDDH IDLD YINQTG KKISVLEVTG SNVVSVAEHV VMTMLVLVRN FVPA HEQIIN HDWEVAIAK DAYDIEGKTI ATIGAGRIGY RVLERLL PFN KPELLYDYDQ ALPKEAEKV GARRVENIEE LVAQADIVTV NAPLHAGTKG LINKELLSKF KKGAWLVNTA RGAICVAEDV AAALESGQLR GYGGDVWFPQ PAKDHPWRD MRNKYGAGNA MTPHYSGTTL DAQTRYAEGT KNILESFFTGF KFDYRPQDII LL NGEYVTKA YGKHDKK.

FDH-C in pET28a vector (MW = 42 104 Da $\epsilon_{280} = 51\,465\text{ M}^{-1}$)
 MKIVLVLYDA GKHAADDEKL YGCTENKLGIANWLKDDQGH E ITTSKKEGET SELDKHIPDA DIIITPPFHP AYITKERLDK AKNLK LVVVA VGSDDHDLD YINQTKGKIS VLEVTGSNVV SVAEHVV MTM LVLVRNFVPA HEQIINH DWE VAAIAKDAYD IEGKIATIG AGRIGYRVL RLLPFPNKEL LYYDYQALPK EAEKVGARR VEN IEELVAQ ADIVTVNAPL HAGTKGLINK ELLSKFKKGA WLVN TARGAI CVAEDVAAA ESGQLRGYGG DVWFPQAPK DHPWR DMRNK YGAGNAMTPH YSGTTLDAQT RYAEGTKNIL ESFFTGF KFDY RPQDIILLNG EYVTKAYGKH DKKVDKLAAL LEHHHHHH.

FDH-C in pPIC9 (MW = 43 071 Da $\epsilon_{280} = 54\,445\text{ M}^{-1}$)
 EAEAYVEFMK IVLVLYDAGK HAADEEKLYG CTENKLGIAN WLKDDQGH E LI TTSDKEGETS ELDKHIPDAD IIITPPFHPA YIT KERLDKA KNLKLVVAVG VGSDDHDLDY INQTKGKISV LE VTGSNVVS VAHVVMTML VLVRNFVPAH EQIINH DWEV AA IAKDAYDI EGKIATIGA GRIGYRVL RLLPFPNKEL LYYDYQ ALPKE AEEKVGARRV ENIEELVAQA DIVTVNAPLH AGTKG LINKE LLSKFKKGAW LVNTARGAIC VAEDVAAA LE SGQLRG YGGD VWFPQAPKD HPWRDMRNKY GAGNAMTPHY SGTTL DAQTR YAEGTKNILE SFFTGFKFDYR PQDIILLNGE YVTKA YGKHD KKGENLYFQS HHHHHH.

Results and Discussion

FDH from *C. boidinii* is fairly small (43 kDa) and thermostable up to 50°C (Slusarczyk et al., 2000; Tishkov and Popov, 2004). It has been structurally characterized (Schirwitz et al., 2007; Guo et al., 2016), and its specific activity was reported to be 4.4 U/mg (Krahulec et al., 2008) and 6.5 U/mg at pH 7.5 at 30°C (Slusarczyk et al., 2000). The location of affinity tags in the sequence of the protein can interfere with enzymatic activity, thus we have cloned the FDH gene into the pMCSG49 vector and into the pET28a vector to be able to obtain a protein with N-terminal His₆-tag (FDH-N) and C-terminal His₆-tag (FDH-C), respectively. Both constructs were successfully expressed in *E. coli* upon induction with 0.4 mM IPTG, which was visualized in a well-defined protein band at ~45 kDa. The proteins were successfully purified using a two-step purification

(Ni-NTA affinity chromatography followed by ion exchange on Q Sepharose), as seen in Fig. 1.

Specific activity of FDH was dependent upon the placement of the His₆-tag, where FDH-N was more active than FDH-C in the reduction of NAD⁺ to NADH and consequently the conversion of formate to carbon dioxide (Fig. 2). Both FDH-C and FDH-N exhibited slightly lower activities compared to previously reported value (Krahulec et al., 2008) measured at higher temperature (30°C) and higher pH (7.5).

Next, we introduced the genes encoding FDH-C and FDH-N into the genome of *P. pastoris*. While FDH-N had a higher specific activity in bacteria, we determined the expression yield of FDH-N in yeast was much lower when compared to FDH-C, thus we focused our efforts on characterization of FDH-C. In both cases, FDH was expressed with α -factor at the N-terminus to induce extracellular secretion in *P. pastoris*. The α -factor signal derived from *Saccharomyces cerevisiae* consists of a 19 amino acid sequence (pre-signal) responsible for signaling that the expressed protein must be transferred to the endoplasmic reticulum after translation, and a 67 residue pro-signal, which is responsible for transferring the protein into the Golgi compartment where it can then be secreted outside the cell (Cregg et al., 2000; Lin-Cereghino et al., 2012; Ahmad et al., 2014).

At pH 6, the *P. pastoris* strain encoding FDH-C was able to secrete FDH upon the addition of 0.5% methanol to the media. Over time, the amount of the enzyme in the media increased, as visualized by SDS-PAGE (Fig. 3). After 48 h, the approximate concentration of protein in media was 8–10 mg/l. Upon purification of FDH-C from the BMMY media, the major product had the same size as the enzyme isolated from *E. coli*, therefore it was not degraded by natural yeast proteases (Fig. 3). Specific activity of secreted FDH was measured using the purified protein to ensure an accurate reading of concentration. On occasion, the purified protein would have a slight yellow color leftover after Ni-NTA column, however, this did not affect enzymatic activity of the protein. Overall, the secreted protein showed about a 10-fold decrease in activity compared to protein expressed in *E. coli* (Table I). Secreted FDH retains its activity in media despite the prolonged incubation at elevated temperature (29°C) and proteases. A kinetic assay of purified protein isolated from media grown for 24, 48 and 120 h shows little difference in specific activity for the reduction of NAD⁺ (Fig. 4, Table I).

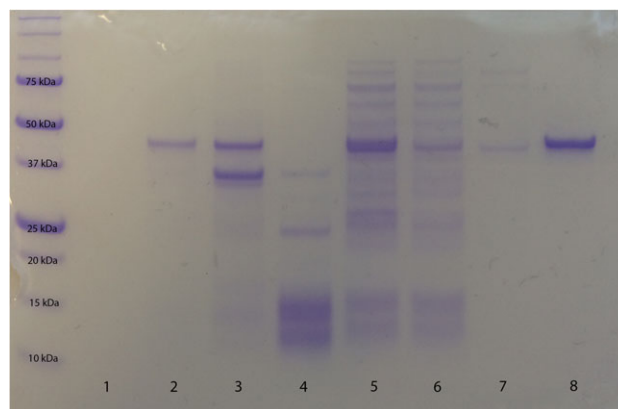


Fig. 1 SDS-PAGE analysis (10% acrylamide gel) of FDH-N at various stages of purification. FDH-N was expressed in *E. coli* BL21(DE3). Lane 1– before IPTG induction; Lane 2– after IPTG induction; Lane 3– cell debris; Lane 4– after streptomycin sulfate precipitation; Lane 5– Ni-NTA column flow through; Lane 6– Ni-NTA column wash; Lane 7– before Q Sepharose column; Lane 8 – after Q Sepharose column, final protein.

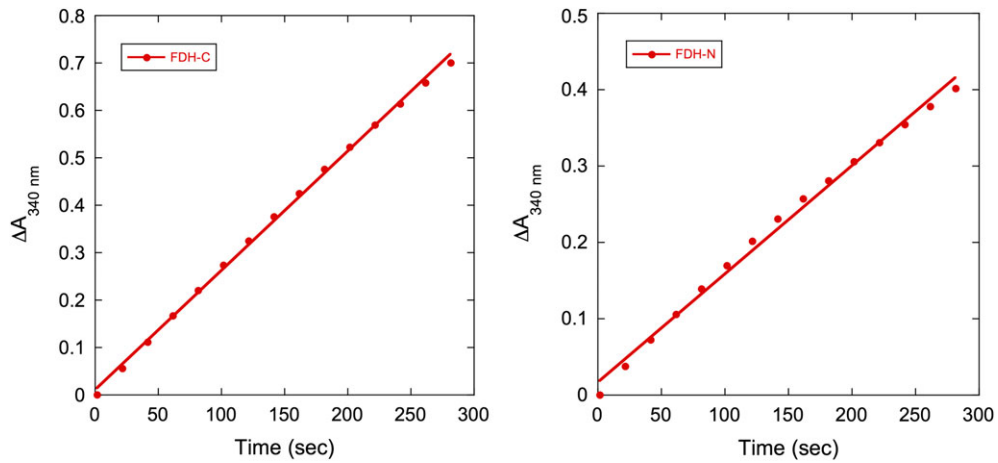


Fig. 2 The kinetic assay graph of FDH-C (0.5 μM, left) and FDH-N (0.1 μM, right) proteins measuring the conversion of NAD⁺ to NADH during the oxidation of sodium formate to carbon dioxide at pH 7. The specific activity of FDH-C is 1.1 U/mg and the specific activity of FDH-N is 2.7 U/mg.

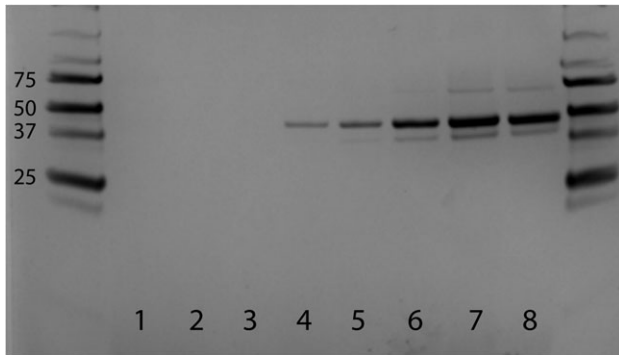


Fig. 3 SDS-PAGE (10% acrylamide) analysis of media after growing yeast with empty pPIC9 for 120 h (Lane 1); media after growing yeast with FDH-C for 0 h (Lane 2), 8 h (Lane 3), 24 h (Lane 4), 48 h (Lane 5), 72 h (Lane 6), 96 h (Lane 7) and 120 h (Lane 8).

Table I. Activity of various FDH constructs expressed in *E. coli* and *P. pastoris*

Construct	Expression host	Specific Activity, U/mg
FDH-N	<i>E. coli</i>	2.7
FDH-C	<i>E. coli</i>	1.1
FDH-C, 24 h	<i>P. pastoris</i>	0.11
FDH-C, 48 h	<i>P. pastoris</i>	0.13
FDH-C, 120 h	<i>P. pastoris</i>	0.10

Puzzled by the significant drop in enzymatic activity of the FDH secreted from *P. pastoris*, we hypothesized that the protein produced in yeast may be glycosylated. However, the MALDI-TOF analysis of the protein expressed in *P. pastoris* (Fig. 5) shows no signs of significant posttranslational modifications. Another possibility for the reduced activity is imprecise cleavage of the prepro-signal on the N-terminus. The placement of linkers or fusion proteins in relation to the protein of interest can affect secretion efficiency and yeast-secreted proteins can be degraded and/or proteolyzed on their N- or C-terminus (Park *et al.*, 1997; Moua *et al.*, 2016). To determine the possibility of the incomplete or impartial cleavage of the α-factor, we digested the secreted enzyme with trypsin. Indeed, we have observed a peak with a mass of 7624 Da in the MALDI spectrum of the tryptic digest, which

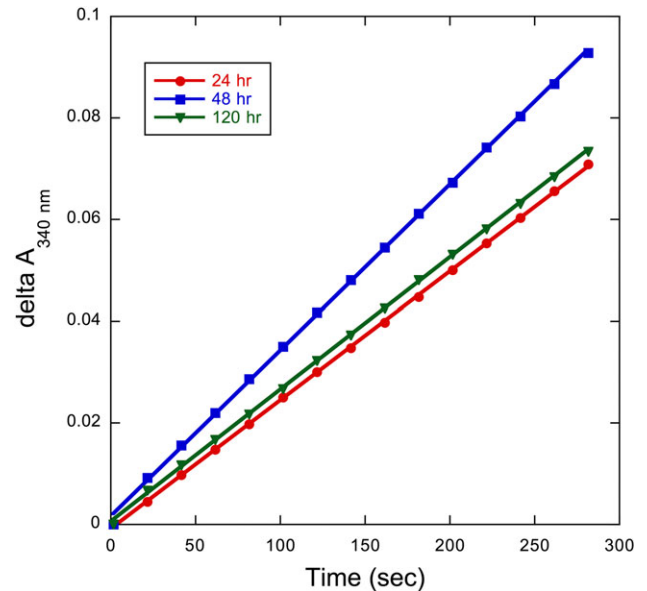


Fig. 4 The kinetic assay graphs measuring the conversion of NAD⁺ to NADH and the oxidation of sodium formate to carbon dioxide at pH 7 as catalyzed by secreted FDH-C protein (0.5 μM). The specific activity of FDH at 24 h is 0.11 U/mg, at 48 h is 0.13 U/mg and at 120 h is 0.10 U/mg.

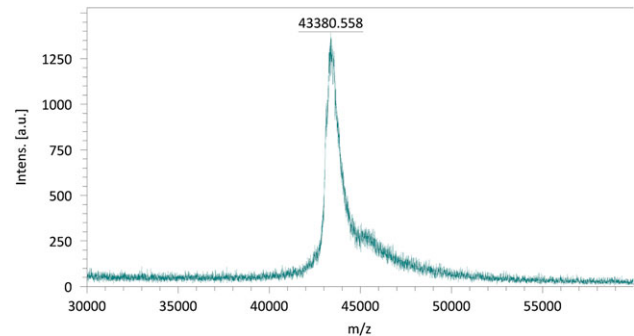


Fig. 5 MALDI-TOF spectrum of FDH-C secreted from *P. pastoris*. The calculated mass for the protein is 43.07 kDa.

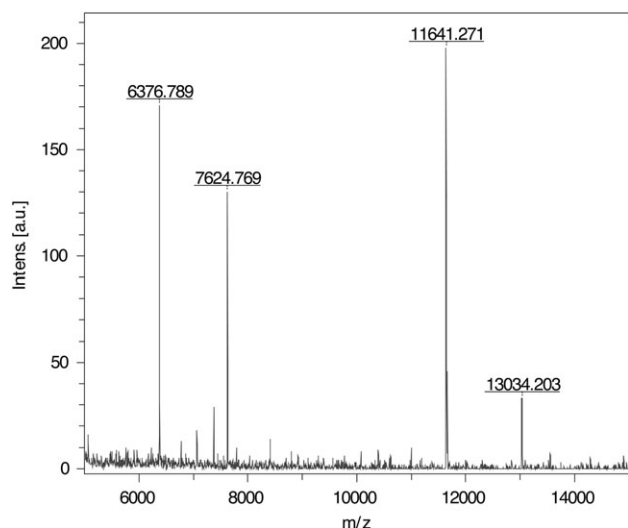


Fig. 6 MALDI-TOF spectrum of FDH-C secreted from *P. pastoris* after tryptic digestion. A peak at 7624 Da corresponds to a large fragment of α -factor.

corresponds to a large fragment of α -factor (Fig. 6). The presence of α -factor in secreted FDH-C protein suggests that this sequence was not completely removed during secretion process. This observation could explain, at least in part, reduced activity of secreted FDH-C and the presence of additional bands on SDS-PAGE gel (Fig. 3).

In summary, we have shown that functional FDH can be continuously secreted from an engineered *P. pastoris* strain. Moreover, while the activity of FDH is diminished relative to the enzyme produced in *E. coli*, the FDH secreted by *P. pastoris* maintains its activity in the culture medium for a prolonged period of time. Enzyme secretion by yeast is a promising method for creating multi-enzyme devices for biofuel production. A yeast that only needs methanol for growth will continuously secrete functional enzyme to replenish the FDH inactivated by the effect of pH, temperature and prolonged electrochemical cycling.

Funding

This work was supported by the National Science Foundation [Grant number 1332349 to I.V.K.]; the National Institute of Health [grant number

GM119634 to I.V.K.]; ORAU Ralph E. Powe Junior Faculty Enhancement award and a Humboldt Fellowship to I.V.K.

References

- Ahmad, M., Hirz, M., Pichler, H. and Schwab, H. (2014) *Appl. Microbiol. Biotechnol.*, **98**, 5301–5317.
- Aresta, M., Dibenedetto, A. and Angelini, A. (2014) *Chem. Rev.*, **114**, 1709–1742.
- Baskaya, F.S., Zhao, X., Flickinger, M.C. and Wang, P. (2009) *Appl. Biochem. Biotechnol.*, **162**, 391–398.
- Bhatia, L., Johri, S. and Ahmad, R. (2012) *AMB Exp.*, **2**, 65–84.
- Cazelles, R., Drone, J., Fajula, F., Ersen, O., Moldovan, S. and Galarneau, A. (2013) *New J. Chem.*, **37**, 3721–3730.
- Cregg, J.M., Cereghino, J.L., Shi, J. and Higgins, D.R. (2000) *Molecular Biotechnol.*, **16**, 23–52.
- Fargione, J., Hill, J., Tilman, D., Polasky, S. and Hawthorne, P. (2008) *Science*, **319**, 1235–1238.
- Guo, Qi, Gakhar, L., Wickersham, K., Francis, K., Vardi-Kilshtain, A., Major, D. T., Cheatum, C.M. and Kohen, A. (2016) *Biochemistry*, **55**, 2760–2771.
- Kim, Y.H., Campbell, E., Yu, J., Minter, S.D. and Banta, S. (2013) *Angew. Chem. Int. Ed.*, **52**, 1437–1440.
- Kondratenko, E.V., Mul, G., Baltrusaitis, J., Larrázabal, G.O. and Pérez-Ramírez, J. (2013) *Energy Environ. Sci.*, **6**, 3112–3135.
- Kovar, K., Looser, V., Hyka, P., Merseburger, T. and Meier, C. (2010) *Chimia (Aarau)*, **64**, 813–818.
- Krahulec, S., Armao, G.C., Weber, H., Klimacek, M. and Nidetzky, B. (2008) *Carbohydr. Res.*, **343**, 1414–1423.
- Lin-Cereghino, G.P., Stark, C.M., Kim, D., et al. (2012) *Gene*, **519**, 311–317.
- Liu, F., Banta, S. and Chen, W. (2013) *Chem. Commun.*, **49**, 3766–3768.
- Moua, P.S., Gonzalez, A., Oshiro, K.T., et al. (2016) *Protein. Expr. Purif.*, **124**, 1–9.
- Obert, R. and Dave, B.C. (1999) *J. Am. Chem. Soc.*, **121**, 12192–12193.
- Park, C.S., Chang, C.C., Kim, J.Y., Ogrydziak, D.M. and Ryu, D.D. (1997) *J. Biol. Chem.*, **272**, 6876–6881.
- Schirwitz, K., Schmidt, A. and Lamzin, V.S. (2007) *Prot. Sci.*, **16**, 1146–1156.
- Slusarczyk, H., Felber, S., Kula, M.R. and Pohl, M. (2000) *Eur. J. Biochem.*, **267**, 1280–1289.
- Stols, L., Gu, M., Dieckman, L., Raffin, R., Collart, F.R. and Donnelly, M.I. (2002) *Protein Expr. Purif.*, **25**, 8–15.
- Tishkov, V.I. and Popov, V.O. (2004) *Biochemistry (Moscow)*, **69**, 1252–1267.
- Wu, S. and Letchworth, G.J. (2004) *Biotechniques*, **36**, 152–154.

MICHELLE TAKACS

mtakacs9@gmail.com | 805.259.9866 | 5601 N 18th St, Lincoln, NE 68521

ACADEMIC CREDENTIALS

Syracuse University, Syracuse, NY
Doctorate of Philosophy in Biochemistry, August 2018
Masters of Philosophy in Biochemistry, May 2015

California Polytechnic State University, San Luis Obispo, CA
Bachelor of Science, Chemistry and Biochemistry double major, June 2013

PUBLICATIONS

Takacs, M., Makhylnets, O.V., Tolbert, P.L., and Korendovych, I.V. "Secretion of functional formate dehydrogenase in *Pichia Pastoris*." *Prot. Eng. Des. Sel.* 2017, 30, 279-284.

"NMR-Based Kinetic Experiments for Undergraduate Chemistry Laboratories," Kantorowski, Eric J.; Ghaffari, Bijan; Macrorie, Allee; Candee, Kellan N.; Petraitis, Jennifer; Miller, Melanie M.; Warneke, Gayle; **Takacs, Michelle**; Hancock, Vanessa; Lusth, Zoe A. in *NMR Spectroscopy in the Undergraduate Curriculum*, Chapter 14, 2013, pp 211-228, ACS Symposium Series, Volume 1128.

Takacs, M., Makhylnets, O.V., Gigon, A.R., and Korendovych, I.V. "Optimizing an Allosteric Protein for Lanthanide Binding by Rationally Designed Modification of the Metal Binding Site." *In preparation*.

PREVIOUS POSITIONS

Research Assistant, Syracuse University

JAN 2014-Present

- Cloning, expression, and purification of proteins in *E. coli* and *P. pastoris*
- Site-directed and saturated mutagenesis of enzymes for kinetic studies
- Design and execution of enzymatic kinetic assays
- Peptide synthesis (solid-phase), purification, and analysis
- Organic synthesis of compounds for kinetic analysis

Head Teaching Assistant, Syracuse University

JAN 2016-Present

- Organize other organic chemistry TAs and preparation of lectures for lab
- Teach and supervise organic chemistry lab students
- Proctor and grade organic chemistry lecture classes

Teaching Assistant, Syracuse University

SEPT 2013-DEC 2015

- Teach and supervise organic chemistry lab students
- Proctor and grade organic chemistry lecture classes

Research Assistant, California Polytechnic State University

MAR 2010-JUN 2013

- Syntheses of substrates for kinetic and mechanistic studies using Grignard reagents
- Design of GC- and NMR-based mechanistic experiments for undergraduate students studying dehydration kinetics and solvent effects
- Experience in mentoring new students in research group

Research Assistant, University of California-Santa Barbara

Summer 2011

- Study of hydroxylation reactions in water using metal additives to increase efficiency
- Synthesis of Witting reagents in water

Research Assistant, University of Nebraska-Lincoln

Summer 2009 & 2012

- Study the enantioselectivity of hydroboration reactions using chiral rhodium catalysts
- Explored new palladium-catalyzed methodology for substrate synthesis

Student Assistant, California Polytechnic State University

Winter 2013

- Hired to assist visually-impaired professor in lecture class and during office hours

Cal Poly Orientation Leader*Fall 2009-2011*

- Participated in 10-week training program each year; more than 100 hours of training in total
- Planned and carried out week long introductory events for incoming freshmen orientation

ORGANIZATIONS

Alpha Chi Sigma Profession Chemistry Fraternity*APR 2011-Present*

- Elected to Master of Ceremonies position (June 2012-June 2013)
- Planned and executed events for incoming members of Alpha Chi Sigma
- Various community outreach activities: Volunteer tutor for general and organic chemistry students; Community outreach to Boy Scouts of America for chemistry merit badge; Volunteer at Wildflower 2012 triathlon

American Chemical Society Member*APR 2010-Present*

- Attended ACS National meetings in San Francisco, Anaheim, Philadelphia, and Boston

Women in Science and Engineering*SEPT 2015-Present***MISCELLANEOUS SKILLS**

Operation and Interpretation

- Protein expression and purification, Bench synthesis, Solid-phase peptide synthesis, Circular Dichroism (CD), Nuclear Magnetic Resonance [^1H , ^{13}C , ^{31}P , ^{19}F , HMBC, HMQC, COSY, NOESY], Spectroscopy [atomic absorption, UV/Vis absorption, fluorescence], Chromatography [gas, affinity, thin layer, preparative thin layer, high-performance liquid, and column], Mass spectrometry, Glove box experience, Inert atmosphere techniques, Matrix-assisted laser desorption/ionization (MALDI)

Computer Proficiency

- Microsoft Office, Maple 14, SciFinder, ChemDraw, Spartan, PSI4 coding, Molecular dynamics calculations

PRESENTATIONS

National ACS Meeting Presentation in Boston, MA*AUG 2018*

- Optimizing an Allosteric Protein for Lanthanide Binding by Rationally Designed Modification of Metal Binding Site
- Engineering Lanthanide-binding Proteins for Hepatocellular Carcinoma Detection and Treatment

National ACS Meeting Presentation in San Francisco, CA*APR 2017*

- Secretion of Functional Formate Dehydrogenase in *Pichia pastoris*

National ACS Meeting Presentation in Boston, MA*AUG 2015*

- Re-design of Calmodulin as a Lead Sensor

Midwest ACS Regional Meeting Presentation in Omaha, NE*OCT 2012*

- Catalytic Asymmetric Hydroboration of O-Allylic Oxime Ethers

National ACS Meeting Presentation in San Diego, CA*APR 2012*

- Kinetics and Rearrangement of 1-arylbicyclo[3.1.0]hexanes

2011 Southern California Undergraduate Research**Conference in Chemistry and Biochemistry Presentation***APR 2011*

- Dehydration Kinetics of 1,1-diaryl-1-propanol and 1,1-diaryl-1-ethanol
- Electrophilic Opening of 1-arylbicyclo[4.1.0]hexanes

National ACS Meeting Presentation in Anaheim, CA*APR 2011*

- Dehydration Kinetics of 1,1-diaryl-1-propanol and 1,1-diaryl-1-ethanol
- Electrophilic Opening of 1-arylbicyclo[4.1.0]hexanes