Engineering Oxygen Reactivity in Heme-Protein Maquettes

Molly Marie Sheehan

University of Pennsylvania, sheehan.molly@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations

Recommended Citation
http://repository.upenn.edu/edissertations/1438

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/1438
For more information, please contact libraryrepository@pobox.upenn.edu.
Abstract
Fundamental questions of protein cofactor oxygen reactivity are left unanswered even after years of research due to the limitations imposed by the complexity of natural systems. Natural electron transport proteins are often large with many spectrally overlapping cofactors and fragile to mutation due to multiple roles of each amino acid. The study of the products oxygen reduction is further made difficult by inadequate available methods for kinetic resolution with quantitative yields of different reactive oxygen species (ROS). To first examine how proteins can avoid oxygen reduction for oxygen transport, as in natural globins, an oxygen transporter is designed from first principles of protein folding with each intermediate characterized. This is the first example of a transparently-crafted protein design functioning at natural rates. The ease of this design process, only three steps, and the minimalism of the design, only 8 different amino acids, calls into question many assumptions of what is necessary for a functional protein. This work suggests that functional success is easier reached by modification of gross properties of proteins rather than attempting structural exactness. This work may also aid future designs without necessitating a return to first principles because it not only produces a functional design but a toolbox from which to engineer.

Oxygen reactivity was further studied through the development of a system for differentiating the rates and yields of superoxide, hydrogen peroxide and hydroxyl radicals. A series of man-made proteins were tested for oxygen binding, each aforementioned ROS production, and ligand exchange rates to uncover differences in ROS species production and mechanisms. It was found that it takes a single repetitive mutation to design from stable oxygen binder to a rapid outer-sphere superoxide generator. Further, another small set of changes confer preferential hydrogen peroxide generation without a superoxide intermediate through an inner-sphere mechanism. The ability to elucidate mechanistic details of highly homologous proteins and resolve each species of ROS allows us to learn about protein oxygen chemistry and apply this knowledge to the interpretation of data from natural systems, as well as use this knowledge to engineer forward towards industrially- and medically-applicable designer proteins.
ENGINEERING OXYGEN REACTIVITY IN HEME-PROTEIN MAQUETTES

Molly M. Sheehan

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

P. Leslie Dutton Ph.D., FRS
Eldridge Reeves Johnson Professor of Biochemistry and Biophysics
Director of the Johnson Foundation for Molecular Biophysics

Graduate Group Chairperson

Kathryn M. Ferguson Ph.D.
Associate Professor of Physiology

Dissertation Committee:

Sergei Vinogradov Ph.D.  Associate Professor of Biochemistry and Biophysics
Ian Blair Ph.D.           A.N. Richards Professor of Pharmacology
Feng Gai, Ph.D.          Professor of Chemistry
Joseph Baur, Ph.D.       Assistant Professor of Physiology
Cecilia Tommos, Ph.D.    Research Asst Professor of Biochemistry and Biophysics
Jose Cerda, Ph.D.        Associate Professor of Chemistry and Chemical Biology at St. Joseph's University
ENGINEERING OXYGEN REACTIVITY IN HEME-PROTEIN MAQUETTES

© 2014

Molly Marie Sheehan
DEDICATION

To my father, Thomas Sheehan, who always supported my curiosity of how the world worked.
ACKNOWLEDGEMENTS

Many people have contributed to my reaching not only this point, but making it an enjoyable process that I will always look back upon fondly.

First and foremost, I would like to thank my P.I., Les Dutton, for everything he has done for me over the years. Les has supported my independent growth from day one, somehow building my confidence while challenging my ideas simultaneously. He has recruited a phenomenal lab full of diverse, creative people that I look forward to working with every day. Les has been a great role model in balancing a rich personal life in addition to fulfilling scientific career. His trust in us and our ideas allows us to grow up into creative scientists unafraid of a heretical scientific thought.

Christopher Moser has also been an invaluable asset in my life in the Dutton lab. Chris is an experimental powerhouse and I cannot recall ever asking him how to do an experiment without him finding a clever thorough solution to perform it shortly. Chris has also helped me learn about each equipment setup he has kindly built. Each scientific ‘problem’ I have come across over the years that I could not overcome myself, Chris has not only helped me through but also done so while bestowing onto me as much knowledge as possible in the most light-hearted way possible. While I think it is likely only possible that I absorbed a mere percentage of what Chris knows over these years, I will treasure it in my future scientific endeavors.

My committee has been a rich source of guidance throughout these years. While I may have bribed them with baked goods, they were always focused and thoughtful in my meetings, often offering me new experimental ideas and challenging my thought process. In particular, I would like to thank my chair, Sergei Vinogradov, for being an added source of guidance. He has been invaluable in navigating the field of ROS detection and always pushed me to be more rigorous. I’d also like to thank Feng Gai for
offering his laboratory and students to my projects, particularly the help of Paul Straus in collecting the FTIR data in this thesis.

I would also like to thank my undergraduate advisor, Robert Fairman, of Haverford College for getting me started. He initially peaked my interest in protein design and made biochemical research accessible to a mere biology major.

My labmates have also been a source of entertainment and scientific debate. In particular, aside from their friendship which I also value, Lee Solomon, Tammer Farid, Sarah Chobot Hoakanson, Bryan Fry, Bohdana Discher and Goutham Kodali have always been available throughout my stay in the Dutton lab to discuss data, debate theories and generally also make day-to-day life more pleasant. Sarah was also my mentor when I first joined and instilled upon me much wisdom at how to survive in a lab full of fairly messy men. Her insight has no doubt saved me countless hours chasing down cuvettes or data resulting from contaminants. Bryan Fry has also been incredibly helpful in troubleshooting equipment issues. Even when we cannot fix the Olis, as we often cannot, he helps me not explode in frustration as my samples oxidize.

And of course, I would like to thank all of my friends and family for being there for me after long evenings in the lab, endless frustration from stopped-flows not working from one reason or another, and realizing I’ve procrastinated something important like a big talk the day after my wedding. I cannot thank you all, but you know who you are. And last, but certainly not least, I am eternally grateful for the support of my husband, Ian, who likely sees me at my worst as still forgives it and helps me celebrate when things are going their best. I’m so glad I came to Penn and met you and you’ve been there for me over the past years. I am also appreciative of your good genes and our in utero daughter for having no complications and giving me a fairly easy pregnancy so I could finish this thesis.
ABSTRACT

ENGINEERING OXYGEN REACTIVITY IN HEME-PROTEIN MAQUETTES

Molly M. Sheehan
P. Leslie Dutton

Fundamental questions of protein cofactor oxygen reactivity are left unanswered even after years of research due to the limitations imposed by the complexity of natural systems. Natural electron transport proteins are often large with many spectrally overlapping cofactors and fragile to mutation due to multiple roles of each amino acid. The study of the products oxygen reduction is further made difficult by inadequate available methods for kinetic resolution with quantitative yields of different reactive oxygen species (ROS). To first examine how proteins can avoid oxygen reduction for oxygen transport, as in natural globins, an oxygen transporter is designed from first principles of protein folding with each intermediate characterized. This is the first example of a transparently-crafted protein design functioning at natural rates. The ease of this design process, only three steps, and the minimalism of the design, only 8 different amino acids, calls into question many assumptions of what is necessary for a functional protein. This work suggests that functional success is easier reached by modification of gross properties of proteins rather than attempting structural exactness. This work may also aid future designs without necessitating a return to first principles because it not only produces a functional design but a toolbox from which to engineer.

Oxygen reactivity was further studied through the development of a system for differentiating the rates and yields of superoxide, hydrogen peroxide and hydroxyl radicals. A series of man-made proteins were tested for oxygen binding, each
aforementioned ROS production, and ligand exchange rates to uncover differences in ROS species production and mechanisms. It was found that it takes a single repetitive mutation to design from stable oxygen binder to a rapid outer-sphere superoxide generator. Further, another small set of changes confer preferential hydrogen peroxide generation without a superoxide intermediate through an inner-sphere mechanism. The ability to elucidate mechanistic details of highly homologous proteins and resolve each species of ROS allows us to learn about protein oxygen chemistry and apply this knowledge to the interpretation of data from natural systems, as well as use this knowledge to engineer forward towards industrially- and medically-applicable designer proteins.
TABLE OF CONTENTS

Dedication...........................................................................................................................................iii
Acknowledgements.............................................................................................................................iv
Abstract..................................................................................................................................................vi
List of Topics.........................................................................................................................................viii
List of Tables..........................................................................................................................................xii
List of Figures........................................................................................................................................xiii

Chapter 1: Introduction.........................................................................................................................1
1.1 Oxygen in metabolism.......................................................................................................................1
  1.1.1 The oxidoreductase family...........................................................................................................1
  1.1.2 Heme proteins as oxidoreductases..............................................................................................3
  1.1.3 Globins.........................................................................................................................................4
1.2 Reactive oxygen species....................................................................................................................8
  1.2.1 What is ROS? ...............................................................................................................................8
  1.2.2 Purposeful ROS production proteins..........................................................................................9
  1.2.3 Mitochondrial ROS production and disease..............................................................................10
  1.2.4 Theories of mitochondrial ROS production..............................................................................12
1.3 Challenges in ROS detection and control......................................................................................16
  1.3.1 Probe delivery and quantitation..................................................................................................16
  1.3.2 Controlling oxygen reactivity for protein material engineering..............................................17
1.4 Goals of This Thesis.........................................................................................................................19
1.5 References.......................................................................................................................................21

Chapter 2: The Maquette Approach....................................................................................................26
2.1 The diverse field of protein design.................................................................................................26
2.2 Maquette design.............................................................................................................................29
  2.2.1 The approach.............................................................................................................................29
  2.2.1 The accomplishments................................................................................................................31
2.3 Applications of maquettes..............................................................................................................35
  2.3.1 Maquettes in protein engineering............................................................................................35
  2.3.2 Maquettes as a model system....................................................................................................39
2.4 What is next for maquette design...................................................................................................40
2.5 References.......................................................................................................................................40

Chapter 3: Designing Minimal Oxidoreductases..............................................................................43
3.1 Introduction ........................................................................................................ 43
  3.1.1 Virtues of minimal design .............................................................................. 43
  3.1.2 Former oxygen transporter designs .............................................................. 46
  3.1.3 Limited amino acid libraries .......................................................................... 47

3.2 Designing functional requirements independently ........................................... 47
  3.2.1 Design of a stable water-soluble $\alpha$-helical protein ................................. 50
  3.2.2 Design of hexacoordinated bis-histidine heme ligation that promotes rapid
      histidine to $O_2$ exchange .................................................................................. 52
  3.2.3 Design of high affinity heme binding in the interior of the $\alpha$-helices ........ 54
  3.2.4 Design of water-restricted heme site and development of an equilibrium
      affinity favoring $O_2$-binding at prevailing aqueous $O_2$ concentrations ........ 56

3.2 Conclusions about natural systems: Lower limits of amino acid diversity for a
functional protein ................................................................................................... 58

3.3 Conclusions about engineering: A new paradigm for protein design ............... 59
  3.3.1 Orderly reconstruction from first principles by testing intermediate designs 59
  3.3.2 Engineering through modification of gross properties ................................. 61

3.4 References ......................................................................................................... 62

Chapter 4: Designing Hyperstable Oxidoreductases ............................................ 66

4.1 Introduction ........................................................................................................ 66
  4.1.1 Stability and its role in protein design and engineering ................................. 66
  4.1.2 Thermophiles ................................................................................................ 67
  4.1.3 The search for blood substitutes ................................................................. 68

4.2 Rebuilding a functional protein without assumptions confers stability .......... 68
  4.2.1 Design ........................................................................................................... 69
  4.2.2 Stability measures ......................................................................................... 71
  4.2.3 Gas-binding measures ................................................................................. 72
  4.2.4 Ligand Exchange Dynamics ........................................................................ 75
  4.2.5 Structure by NMR ....................................................................................... 78

4.3 Conclusions ....................................................................................................... 80
  4.3.1 Simplicity confers stability, not diminishes it .............................................. 80
  4.3.2 Sliding rocks ................................................................................................ 82
  4.3.3 Suggestions for computation and directed evolution .................................. 83
  4.3.4 Potential as a platform for blood substitutes and other applications ......... 84

4.4 References ....................................................................................................... 86

Chapter 5: Development of a ROS Laboratory ..................................................... 90

5.1 Introduction ....................................................................................................... 90
  5.1.1 Maquettes as model systems ....................................................................... 90
  5.1.2 ROS kinetic resolution ............................................................................... 93

5.2 Oxygen reactivity in a series of maquettes ..................................................... 96
Appendix 2: Supplemental Data

A2.1 Gas binding
A2.2 Reactive oxygen species detection
A2.3 CO flash photolysis
A2.4 Protein characterization
LIST OF TABLES

Chapter 3
Table 3.1: Heme-binding protein characterization.............................................53

Chapter 4
Table 4.1: Ligand exchange dynamics and energies of activation.....................76

Chapter 5
Table 5.1: Sequences used in the development of a ROS laboratory.....................97
Table 5.2: Measured and reported rates for maquettes, natural globins and Nox......110
LIST OF FIGURES

Chapter 1
Figure 1.1: ROS potentials and reactions.................................................................2
Figure 1.2: Most common naturally occurring hemes..............................................4
Figure 1.3: Pentacoordinate and hexacoordinate heme binding.............................6
Figure 1.4: Heme coordination shown for pentacoordinate binders.......................7
Figure 1.5: NADPH oxidase (Nox) cartoon structure...........................................9
Figure 1.6: The first three complexes of the mitochondrial ET chain....................14

Chapter 2
Figure 2.1: Maquette design..................................................................................29
Figure 2.2: Cofactors and potentials......................................................................32
Figure 2.3: Heme b ligations to a single maquette by UV-visible spectroscopy........33
Figure 2.4: Thermal melts of a series of single chain heme b maquettes................34
Figure 2.5: Illustration of Muller’s Ratchet..............................................................36
Figure 2.6: Maquette approach to functional designs............................................37
Figure 2.7: Oxygen binding shown functional domain independence....................38

Chapter 3
Figure 3.1: Complexity of natural and man-made proteins...................................45
Figure 3.2: Heme and oxygen binding to maquettes.............................................47
Figure 3.3: Design of a minimal oxygen-binding protein from first principles........49
Figure 3.4: Heme titrations of 5, 4, 3, 2, and 5A9EA79E......................................51
Figure 3.5: Redox titrations of 3, 4, and 5...............................................................53
Figure 3.6: UV-visible spectra at -15°C.................................................................55
Figure 3.7: UV-visible kinetics of 4 and 5 by stopped-flow mixing......................56

Chapter 4
Figure 4.1: Depiction of synthetic protein residues..............................................70
Figure 4.2: Thermal stability..................................................................................71
Figure 4.3: Stability over time................................................................................72
Figure 4.4: UV-visible spectra at -15°C.................................................................73
Figure 4.5: FTIR of the thermostable protein in comparison to previous maquettes..74
Figure 4.6: Oxygen-binding kinetics upon rapid mixing......................................75
Figure 4.7: CO flash photolysis at 20%, 50% and 100% CO..................................77
Figure 4.8: Changes in structure upon heme binding..........................................79

Chapter 5
Figure 5.1: Topologies of maquettes for ROS study............................................92
Figure 5.2: SO production scheme from one-heme maquettes.............................98
Figure 5.3: ROS production possibilities from two-heme maquettes.....................99
Figure 5.4: Oxygen-binding properties of maquettes..........................................102
Figure 5.5: SO production as detected by MCLA................................................104
Figure 5.6: Hydrogen peroxide production as detected by Amplex Red and HRP....105
Figure 5.7: Hydroxyl radical production as detected by AO7.............................108
Figure 5.8: Future experiments possible with the ROS laboratory......................117
Appendix 2: Supplemental Data

Figure A2.1: CO binding data of C.1 and 5.................................................................140
Figure A2.2: Oxygen binding to C...............................................................141
Figure A2.3: Oxygen binding to C.1.................................................................142
Figure A2.4: Spectra for stopped-flow peroxide detection................................143
Figure A2.5: Spectra for stopped-flow hydroxyl radical detection.................144
Figure A2.6: CO photolysis of A.................................................................145
Figure A2.7: CO photolysis of B.................................................................146
Figure A2.8: CO photolysis of B.1.................................................................147
Figure A2.9: CO photolysis of C.................................................................148
Figure A2.10: CO photolysis of C.1.................................................................149
Figure A2.11: Circular dichroism spectra of 4 and 5........................................150
Chapter 1: Introduction

1.1 Oxygen in metabolism

1.1.1 The oxidoreductase family

Since free dioxygen became available in the Earth’s atmosphere 2.3 billion years ago, oxygen metabolism is central to the survival of aerobic life (Falkowski 2006). Prior to the first photosynthetic events producing oxygen from water, the atmosphere was highly reducing owing to the relatively poor oxidizing ability of the prior oxidant, nitrogen. Today, oxidoreductases make up over 25% of all known proteins (Dutton & Moser 2011).

Dioxygen’s powerful oxidation ability leads to reactions incorporating it into organic molecules as well as reduction and oxidation of both itself and other molecules. These reactions’ thermodynamic favorability is governed by its reactivity and ubiquitous existence throughout cells. Despite being a simple molecule, its chemistry is complex and diverse. If not properly controlled and contained, dioxygen’s oxidizing ability will lead to unfavorable side reactions oxidizing proteins, lipids and other biological molecules through the production of reactive oxygen species (ROS) (Figure 1.1B). In order to both harness and control dioxygen chemistry, a wide range of proteins has evolved to catalyze reactions from superoxide dismutation and peroxidation to respiratory and photosynthetic energy production. Because oxygen is so vital to respiratory life’s existence, additional sets of proteins, globins, have evolved to bind, transport, store and deliver dioxygen in larger organisms (Wajcman et al. 2009).

The simplest chemistry dioxygen can partake in is one electron transfer. Oxygen can evolve through four electron transfer events to water (Figure 1.1A) with highly reactive oxygen species being produced as intermediates. Dioxygen also participates in
more complex natural reactions as an oxygen donor for the addition of oxygen groups on organic molecules. Even though many of these reactions are thermodynamically favorable, most will not occur without a catalyst present in biological or room temperature conditions. These reactions do not occur because oxygen exists in its ground state in the triplet state and most products formed with organic molecules exist in the singlet state. Because a quantum-mechanically-forbidden spin flip must occur for the reaction to take place, these reactions will not occur with appreciable yields during collisional encounters. However, enzymes are able to use alternate paths to these reactions by either producing organic radicals that can freely react with triplet oxygen (Cantarella et al. 2003) or binding oxygen to paramagnetic metal ions to promote multi-electron processes, as in the case of monooxygenases like cytochrome P450 (Shaik et al. 2007). Unless these reactions are carefully controlled through fine enzymatic tuning, ROS will be released, leading to the damage discussed in section 1.2.

<table>
<thead>
<tr>
<th>A</th>
<th>Reaction</th>
<th>$E^*$ (V) vs NHE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O₂ + e⁻ → O₂⁻⁻</td>
<td>-0.16 V</td>
</tr>
<tr>
<td></td>
<td>O₂⁻⁻ + e⁻ → O₂⁻⁻O₂⁻⁻</td>
<td>+0.89 V</td>
</tr>
<tr>
<td></td>
<td>O₂⁻⁻ + e⁻ → H₂O⁻⁻</td>
<td>+0.38 V</td>
</tr>
<tr>
<td></td>
<td>H₂O⁻⁻ + e⁻ → OH⁻⁻</td>
<td>+2.31 V</td>
</tr>
</tbody>
</table>

Figure 1.1: ROS potentials and reactions. (A) The four consecutive step reduction of oxygen to water through superoxide, hydrogen peroxide and hydroxyl radical intermediates is shown with reduction potential of each step vs. normal hydrogen electrode (NHE) at pH 7, 25°C with standard state of unit activity (Bertini 2007). Oxygen atoms depicted as red circles and hydrogen atoms as grey circles with proton additions not shown. (B) The reactions dioxygen can undergo without catalysis are shown towards the formation of ROS through either one or two electron processes, as well as dismutation of two oxygens to hydrogen peroxide and additional reduction of hydroxyl radical.
1.1.2 Hemes in oxidoreductases

Oxidoreductases naturally contain a large variety of cofactors, or species that can partake in oxidation and reduction reactions, for their catalysis and electron transfer events. The most prevalent of these cofactors are flavins, quinones, iron-sulfur clusters and hemes, which are all essential for the mitochondrial electron transport chain. Flavins and quinones are capable of one- or two-electron transfer events. Iron-sulfur clusters can transfer as many electrons as they have iron atoms but typically are involved in single electron transfers. Hemes, iron-centered porphyrin macrocycles, however are obligate one-electron transfer cofactors that can access a large range of reduction potentials through a variety of macrocycle substitutions. Hemes can be either covalently, as in c-type cytochromes, or non-covalently attached to proteins, as in a- and b-type cytochromes (Figure 1.2). They are present in a variety of oxidoreductases, as they can be efficient oxygen-binders and reducers, as well as promote high rates of electron transfer due to low reorganization energies allowed by their conjugated macrocycle ligand delocalizing electrons and ability to span a large range of reduction potentials (Bertini 2007).

Hemes can be either pentacoordinate or hexacoordinate, altering the reactivity of the cofactor by exposing an axial binding site in the pentacoordinate form (Figure 1.3). Hexacoordinate hemes have all octahedral coordinate sites filled and are hence stably bound and tend to be involved in rapid electron transfer events, such as cytochromes c and b562, as well as in larger electron transport chains such as respiratory Complex III. Pentacoordinate hemes have one axial orbital unbound, free to bind a ligand, such as dioxygen. Proteins containing pentacoordinate hemes make good gaseous ligand binders for transport, such as myoglobin, and for oxygen-mediated catalysis, such as cytochrome P450.
1.1.3 Globins

Containing hexacoordinate and pentacoordinate hemes, the globin family of proteins today consists of a large variety of structurally similar proteins that perform a multitude of tasks. Globins are classically considered reversible oxygen-binders for the purpose of oxygen transport, but have more recently been found to perform electron transfer, NO scavenging and cytoprotection against ROS damage (Wajcman et al. 2009). Globins exist in Archea (Freitas et al. 2004; Moens et al. 1996), plants (Hoy et al. 2007) and vertebrates. Throughout these kingdoms they are active as monomers, dimers and tetramers, though each monomer consists of the globin fold. Discovered over 50 years ago, the globin fold consists of 8 α-helices surrounding a heme $b$ group in a 3/3 canonical formation (Kendrew et al. 1958).

Typically oxygen-binding globins are thought to be pentacoordinate heme binders, with a free axial orbital open for gaseous ligand binding as in the cases of myoglobin and hemoglobin. However, it has been found that these oxygen binders...
evolved from hexacoordinate heme binders, which may have evolved the ability to bind oxygen prior to stable pentacoordinate heme binding (Hoy et al. 2007). Likely owing to their predecessors, pentacoordinate heme binders still have a second histidine at a distal site, which acts to stabilize bound dioxygen in the oxyferrous state (Figure 1.3). The placement of this histidine in such close proximity to a pentacoordinate heme without it binding requires extreme stabilization and rigidity to the heme-binding site. This required structural tuning presents an evolutionary challenge, likely taking considerable iterations. These iterations would involve evolving through hexacoordinate oxygen-binding intermediates for which one histidine ligation was weakened but not fully removed from the iron coordination sphere.

For all oxygen transport globins, there are energetic requirements that must be met in order for both oxygen binding and release to be possible at physiological conditions. Whether a globin contains pentacoordinate or hexacoordinate heme site, for oxygen binding to occur one axial binding site must be open for oxygen to bind. In the case of hexacoordinate heme binders, such as neuroglobin, this requires the removal of one histidine. In order to lower the thermodynamic barrier for histidine release, it has been suggested that natural hexacoordinate oxygen-binders place structural strain upon one of the ligating histidines so the iron can become pentacoordinate (Pesce et al. 2004). This barrier must be sufficiently lowered that the histidine-off rate is high enough such that the iron spends enough time in the pentacoordinate form for oxygen to bind. If this condition is not met, the iron will oxidize to its ferric form, in which it cannot bind oxygen. These oxygen binders are involved in local oxygen reactivity mediation in cases of oxidative stress, so only need to bind for short periods of time (Pesce et al. 2003). However, pentacoordinate-heme-ligating oxygen binders have evolved for long-range oxygen transport and therefore must stabilize the oxyferrous state more.
Figure 1.3: Pentacoordinate and hexacoordinate heme binding. The left shows a pentacoordinate heme $b$ (red) ligation to a single histidine (green) with distal, unbound histidine as present in myoglobin. The right shows a hexacoordinate bis-histidine heme $b$ ligation between two histidines as present in neuroglobin.

The ideal $K_D$ for oxygen binding is organism-specific depending on the oxygen concentration its environment, such that globins become oxygen-saturated in highly oxygenated tissues and release their oxygen in those tissues requiring oxygenation. No matter the environment, however, the natural $K_D$ for oxygen to a pentacoordinate heme is not low enough to allow for efficient oxygen binding. In order to improve the stability of the oxyferrous state, pentacoordinate heme binders have conserved the distal histidine at such a distance from the open axial heme site that it can hydrogen bond to dioxygen.
when bound (Perutz 1970). Additionally, when oxygen binds to the ferrous iron, an energetically favorable event of the iron fitting into the center of the porphyrin ring occurs due to the iron shrinking in size as it goes from high spin to low spin (Collman 1977). Once the oxygen settles into the porphyrin ring, the distal histidine is positioned such that it can accomplish the dual task of stabilizing the binding of dioxygen and sterically inhibiting the binding of more-linear carbon monoxide that would otherwise bind 200-fold better, posing a problem to oxygen transport even under the low normal physiological levels of CO (Perutz 1990) (Figure 1.4).

![Diagram of heme coordination](image)

**Figure 1.4:** Heme coordination shown for pentacoordinate binders. From left to right: high-spin iron out of plane of the porphyrin ring in the ferrous state, low-spin oxyferrous state with ferrous heme bound to dioxygen and also NH group of distal histidine hydrogen bonded to oxygen, low-spin ferrous heme bound to carbon monoxide with sterically-impeding distal histidine.

Not all globins are designed to bind oxygen; there are other globins that participate in ROS mediation, electron transfer and NO scavenging, as well as a whole family of globin-coupled receptors that bind dioxygen and other gaseous ligands (Hou et al. 2001). Recently, some tissue hexacoordinate-heme-binding globins have been identified including neuroglobin and cytoglobin with neural and nuclear locations, respectively. These globins appear to have evolved from the same early globins as
oxygen transport proteins and are highly conserved (Burmester et al. 2004). Both neuroglobin and cytoglobin are capable of binding oxygen, though for much shorter periods of time than typical oxygen-transport globins (Dewilde et al. 2001). While it has been postulated that neuroglobin, in particular, may be responsible to tissue-level oxygen delivery to mitochondria, its up-regulation in hypoxic conditions and biochemical traits point to its more prominent role being an electron transport protein used to scavenge ROS (Fago et al. 2004; Kiger et al. 2011). Cytoglobin, which is also up-regulated in hypoxic conditions, may play a role in providing oxygen to dioxygen-mediated enzymes such as hydroxylases in addition to ROS mediation (Fago et al. 2004; Sugimoto et al. 2004). Even though the specific biological role of these globins has yet to be identified and more of their kind are still being identified (Kiger et al. 2011), their strong conservation across species suggests a primary role in mediating oxidative damage caused by ROS in tissues that are difficult to repair and/or are prone to hypoxic injury or increased ROS production.

1.2 Reactive oxygen species

1.2.1 What is ROS?

Reactive oxygen species (ROS) have been implicated in aging (Afanas'ev 2005; Afanas'ev 2009; Lambert & M. D. Brand 2009) and a variety of diseases (Drose & Brandt 2008; Fink 2001) as well as used as a bactericidal agent by neutrophils (Cross et al. 1985). In particular, mitochondrial ROS production has been shown to be related to intracellular calcium levels, leading to opening of the mitochondrial permeability transition (MPT) pore (Fink 2001; Zorov et al. 2006; Halestrap 2006) in ischemia/reperfusion injury during cardiac arrest, heart transplantation and stroke
Superoxide (SO) is the primary ROS which leads to most damaging oxidative effects due to its reaction cascade into hydroxyl and nitrogen radicals, whereas hydrogen peroxide is a well-known signaling molecule and will not, on its own, lead to the production of damaging radicals (Bell et al. 2007; Murphy 2009; Guzy & Schumacker 2006; Adler et al. 1999) (Figure 1.1). Despite the medical relevance of mitochondrial ROS production, the source of SO production in mitochondrial electron transport (ET) is still highly debated.

1.2.2 Purposeful ROS production proteins

Not all ROS production is bad though. Some proteins intentionally produce ROS. The most efficient natural ROS generators are NADPH oxidases (Nox). Nox enzymes are present throughout eukaryotes and are critical in processes including hormone synthesis, signal transduction and propagation, and killing pathogens (Sumimoto 2008; Brown & Griendling 2009). Despite their varied purposes, many of which have yet to be defined, they share a similar structure of an extra-membrane FAD for the oxidation of NADPH and

![Figure 1.5: NADPH oxidase (Nox) cartoon structure across a lipid bilayer (blue and orange) with two hemes (red) and one flavin adenosine diphosphate (FAD, yellow), which transfer electrons from NADPH to dioxygen (red), forming superoxide.](image-url)
7 transmembrane α-helices with bis-histidine binding of two heme bs 14Å apart to transfer electrons across the membrane for the production of SO, with one exception (Figure 1.5). Nox proteins are capable of producing large volumes of SO across membranes and are capable of generating enough ROS to kill microbes in the phagosomes of neutrophils (Babior et al. 1973; Segal 2004).

The only Nox not to produce superoxide is Nox4, a constitutive hydrogen peroxide producer. Despite high structural homology to other Nox enzymes (Martyn et al. 2006), the few structural differences found have lead to speculation as to how this enzyme preferentially produces peroxide, doubly reduced oxygen, from a one-electron transfer cofactor, heme b. The top theories implicate differences in extra-membrane loop regions, particularly the E loop, which may keep one superoxide near the heme site in close enough proximity to favor dismutation upon a second superoxide’s generation (Takac et al. 2011). More specifically, it is possible that changes in this loop region may allow for stable oxygen-binding to oxygen-reducing heme, with kinetics that would favor that heme receiving a second electron while in the oxyferrous state, allowing for two electrons to transfer to oxygen concertedly (Dikalov et al. 2008).

1.2.3 Mitochondrial ROS production and disease

Much of the ROS produced in biology is not so purposeful. Every time an electron transfer reaction occurs, there is the possibility of leakage to oxygen, forming ROS. This leakage can especially lead to significant ROS production for metabolic pathways in which electrons are transported in high volumes such as the mitochondrial electron transport (ET) chain. Since the discovery of mitochondrial ROS production by Britton Chance in 1973 (Boveris & Chance 1973) the ubiquitous radical SO has garnered
much attention. However, little is known about the sources of SO and its biological importance, especially within the mitochondrial ET chain.

There is evidence that SO produced from mitochondria is involved in both cell signaling and oxidative stress, though there is debate of its primary purpose. SO may act as a cell-signaling molecule when dismuted to hydrogen peroxide (Bell et al. 2007; Murphy 2009; Guzy & Schumacker 2006; Adler et al. 1999). However, SO can also have deleterious effects. Although SO is not a potent oxidant, it can rapidly react to form highly reactive radicals under normal biological conditions such as hydroxyl radicals and peroxynitrite (Beckman & Koppenol 1996) (Figure 1.1). These radicals are implicated in oxidative stress during ischemia/reperfusion injury (M. Brand et al. 2004; Halestrap 2006). Large amounts of oxidative stress occur during cardiac transplants, arrest and stroke (Adlam et al. 2005; Ferrari et al. 2004). Much of this stress has been traced back to mitochondrial ROS production and the opening of the MPT pore (Halestrap 2006). Treatments such as quinones (Adlam et al. 2005) have been used to mitigate the oxidative damage, but choice of quinones has been through trial-and-error methods because the sources of ROS are ill defined. Recent work by Judy Hirst has substantiated that these quinones may outcompete oxygen for reduction by the flavin of complex I (CI) (King et al. 2009).

While in extreme cases such as transplant it is clear that ROS cause oxidative damage, there is debate over whether ROS production is a mode of signaling employed to avoid more damaging effects. A recent finding of mitochondrial ROS production in ischemia/reperfusion injury is characteristic of this debate. Due to an increase in succinate concentration under hypoxic conditions (low oxygen concentrations), it has been assumed that in hypoxia the mitochondrial ET chain and electrons flow in reverse through CI and forward through complex III (CIII), leading to an increase in deleterious
ROS production primarily from the flavin site of CI (Lambert & M. D. Brand 2009) (Figure 1.6). However, Budinger and colleagues recently discovered that an increase in CIII ROS production under hypoxic conditions is necessary for stabilizing a protein important for adaptation to low oxygen concentrations, hypoxia-inducible factor 1α (HIF-1α) (Bell et al. 2007). HIF-1α is stabilized through dehydroxylation mediated by hydrogen peroxide produced by a quinone site, Qo, of CIII. These findings are contradictory to two general assumptions about ROS production: ROS production is primarily deleterious and CI is the major ROS producer under reverse electron transport.

Central to this debate is the fate of ROS in vivo: how much SO and hydrogen peroxide are produced? While SO may lead to oxidative damage, hydrogen peroxide is fairly stable (Figure 1.1) and a well-known cell signaling molecule. If the primary product is hydrogen peroxide or if the majority of SO is rapidly converted to hydrogen peroxide, then cell signaling is more plausibly the primary purpose of ROS production.

1.2.4 Theories of mitochondrial ROS production

So far, understanding mitochondrial SO production and its effects have been difficult. Both CI and CIII have been implicated in SO production, as have specific cofactors within each of those enzymes. Mainly suspected are quinone (Bell et al. 2007; St-Pierre et al. 2002) and heme (Drose & Brandt 2008) of CIII and the iron-sulfur clusters or quinones (Genova et al. 2001; Ohnishi 1998; Magnitsky et al. 2002) and flavin (Grivennikova & Vinogradov 2006; Esterházy et al. 2008; Kussmaul & Hirst 2006) of CI. ROS production from these complexes is hard to study as are their complicated ET mechanisms, which have yet to be completely defined.

Most studies of respiratory ROS have used inhibitors or isolated proteins to ease study and have focused on CI and CIII, so they do not well address potential production from other sites. However, work by James Imlay suggests the flavin of complex II (CII)
directly reduces oxygen in some disease states and hyperoxic conditions (high oxygen concentrations) (Messner & Imlay 2002) and little work has been done to test if CII produces ROS under normoxic conditions.

In contrast, there has been considerable work and debate on whether CI or CIII produce the majority of ROS in forward (glucose metabolism) and reverse (succinate metabolism) ET under normoxic and hypoxic conditions. Conclusions differ on which cofactors directly reduce oxygen, when ROS production is enhanced, and which ROS is produced. The major trends for forward and reverse ET are presented in figure 1.6. However, there is little consistency in techniques between studies, so it is difficult to compare results. For CI, a few researchers suggest the flavin mononucleotide (FMN) produces ROS in reverse ET and the iron-sulfur clusters or quinones produce ROS in forward ET (Grivennikova & Vinogradov 2006, Genova et al. 2001, Lambert & M. Brand 2004a), although there is considerable variance in the amounts proposed to be produced and the importance of quinones among these studies. Despite some consensus in these studies, the findings have been disputed by work from Judy Hirst (Kussmaul & Hirst 2006; Esterházy et al. 2008) that suggests the FMN is responsible for the majority of oxygen reduction even in forward ET. Hirst found quinone-binding site inhibitors to have no effect on ROS production and that iron sulfur clusters do not form radicals. Furthermore, Hirst surmised that the FMN of CI has a role in ROS fate. She found 95% of ROS in bovine CI is SO, leading her to conclude that the reduced FMN is responsible for oxygen reduction (Kussmaul & Hirst 2006). Surprisingly, Hirst also found that *E. coli* produce 20% SO and 80% hydrogen peroxide. However, her experiments may not be quantitatively comparable due to the use of different SO probes among studies. She attributes the bovine SO production to the reduction of the FMN.
semiquinone (SQ) lifetime by the adjacent low potential iron-sulfur cluster and a possible faster SO dissociation rate.

Figure 1.6: The first three complexes of the mitochondrial electron transport chain during forward and reverse ET. Electron transfer is shown in blue arrows and ROS production is depicted as red arrows with size corresponding to qualitative consensus of the literature.
For CIII, sites are more easily deduced using SOD-sensitivity due to the directionality associated with each site. Most studies agree that heme b₅ or Qₒ are the sites of ROS production and that Martin Brand (St-Pierre et al. 2002) was correct in concluding that CIII is the major producer of ROS in forward ET after CI. Several studies have demonstrated that CIII can produce SO to both sides of the IM and is the only producer to the IMS (Han et al. 2001, Muller et al. 2004). However, CIII’s role in reverse ET ROS production is more highly debated. Rather than CI, Ulrich Brandt proposed that CIII Is the primary producer under these conditions, also implicating the Qₒ site (Drose & Brandt 2008). This work contrasts with much of the CI studies, but is supported by Bell’s finding of CIII ROS production’s involvement with hypoxic signaling (Bell et al. 2007).

These studies have left essential questions regarding ROS functions and its role in disease treatment unanswered: Is ROS production from CI deleterious and CIII beneficial? Does CII really not contribute to ROS production? What are the levels of SO and hydrogen peroxide formation from CI, CII and CIII under normoxic, hypoxic and hyperoxic conditions? Which cofactors are directly reducing oxygen or which mediators are allowing oxygen reduction? While the inconsistencies between studies create too much complexity for coherent meta analysis, interpretation is further hindered by the complexity of the complexes themselves: CI, CII, and CIII have 9, 5, and 8 constitutive cofactors excluding diffusible electron carriers. In order to answer these questions with cofactor-level resolution, a system must be devised that simplifies study and allows more basic questions regarding the abilities of each cofactor or subset of cofactors to reduce oxygen to be addressed.
1.3 Challenges in ROS detection and control

1.3.1 Probe delivery and quantitation

Determining the fate of oxygen reduction under various oxygen pressures *in vivo* is not trivial. Further, it is not clear which cofactor sites in the mitochondrial respiratory complexes are producing the majority of ROS under these conditions (Lambert & M. D. Brand 2009). Addressing this question is complicated by inconsistent use of probes and techniques in the literature that detect SO, hydrogen peroxide or total ROS via conversion of SO to hydrogen peroxide using SOD.

Although HRP/Amplex Red reagents are well regarded to detect hydrogen peroxide (Drose & Brandt 2008; Esterházy et al. 2008; Bell et al. 2007; Murphy 2009), there is no consensus of the best method for detecting mitochondrial SO production. Judy Hirst determined acetylated cytochrome c reduction is the best SO detection method (Kussmaul & Hirst 2006), but that dihydroethidium (DHE) is the best option when acetylated cytochrome c can be directly reduced (Esterházy et al. 2008). For years methyl-cypridina luciferin analog (MCLA) has been considered the most reliable detection method (Nakano 1998; Tampo et al. 1998; Muller et al. 2004), as it is more efficient at biological pHs than other probes (Oosthuizen et al. 1997). Others have avoided the selection of SO-specific probes completely by converting SO to hydrogen peroxide with excess SOD and detecting hydrogen peroxide with Amplex Red (Drose & Brandt 2008; Grivennikova et al. 2008; Lambert & M. Brand 2004b; St-Pierre et al. 2002). This last method avoids issues with SO-specificity, but only is able to measure total ROS production. Prominent researchers have not been able to agree on SO probe selection, but it is key to determining the fate of mitochondrial ROS production.
One major issue with the common SO detection methods is a lack of specificity for SO over other ROS (Afanasev 2008). Without specificity, it is difficult to quantify signals. Thus, signals reported are mainly qualitative and not readily comparable between studies; the majority of comparisons made between complexes, cofactors within them, and homologs are highly speculative. Nevertheless, many comparisons have been made and some trends of mitochondrial ROS production during both forward and reverse ET are generally agreed upon in the literature (Figure 1.6).

Probe delivery poses another challenge plaguing quantitation of ROS from natural enzymes. With isolated mitochondria only external ROS are readily measurable; detection of ROS in the intermembrane space (IMS) and matrix requires probe delivery to these locations (Murphy 2009). Instead, studies of mitochondrial ROS production are conducted on submitochondrial particles (SMPs), reverse SMPs (Drose & Brandt 2008; Grivennikova et al. 2008) or isolated enzymes (Kussmaul & Hirst 2006). SMPs are made by breaking and reforming mitochondria with either side of the inner membrane (IM) exposed to solvent. However, determining ROS directionality is still difficult, because studies typically use one type of SMP and hydrogen peroxide will diffuse across membranes but SO will not. Despite these difficulties, many claim to have found the culprits of ROS production in forward and reverse ET with varying conclusions discussed in section 1.2.

1.3.2 Controlling oxygen reactivity for protein material engineering

There are two main reasons it is desired to control oxygen reactivity in protein materials. The first is to avoid the production of damaging ROS, such as SO and hydroxyl radicals, which may cause oxidative damage. The second is to control oxygen binding towards the production of useful oxygen catalysts.
Despite oxidoreductases making up over a quarter of natural enzymes, very few useful industrial oxidoreductase catalysts have been produced (Xu 2005). The majority of commercial catalysts are hydrolases, with very few oxidoreductases available. The challenge of controlling electron transfer reactions and oxygen-mediated catalysis is formidable. Nature has become astute at controlling and utilizing these reactions, but the fields of protein design and redesign have had limited success. While we can observe the natural enzymes, reproducing or modifying their functions has proven difficult as small disruptions to their structures often plummet catalytic rates. We clearly do not understand the chemistry of oxygen-binding and oxidoreductase catalysis enough to engineer efficient enzymes.

No matter which reaction is performed by a designed protein, if electrons are transferred there is the possibility of leakage to dioxygen. If SO is produced, the subsequent reactions can lead to oxidative damage via the formation of hydroxyl or nitrogen radicals. These species can oxidize the proteins producing them as well as other organic molecules nearby. As the protein material becomes oxidized, it becomes degraded and less efficient. This is only one of the major barriers in creating synthetic electron transport systems out of proteins, as materials will be short-lived if they cannot self-repair or control their electron flow. If oxygen were a bound substrate, stable oxygen binding would increase efficiency and thus decrease leakage to ROS. For other ET reactions, even if leakage cannot be avoided it may be possible to steer ROS production towards peroxide production. Peroxide is not a great threat for protein degradation as it is a fairly weak oxidizing agent and without high levels of SO present more damaging radicals would not be produced (Figure 1.1).
1.4 Goals of This Thesis

The goals of this thesis are two-fold. First, I aim to outline the basic requirements for oxygen-binding chemistry from first principles through the systematic reengineering of a globin with full transparency. This work is presented in Chapter 3 and is further developed in Chapter 4. Second, I aim to develop a molecular ROS laboratory that can be used to determine the requirements for specific ROS species production, particularly hydrogen peroxide over superoxide. This work is presented in Chapter 5. This thesis begins with what had been previously accomplished in the Dutton laboratory related to these projects by former lab members and in collaboration.

The limited success of being able to replicate oxidoreductase function in designed proteins points to a lack of understanding of the natural oxidoreductase mechanism. Much of the difficulty in studying these enzymes comes from the complexity of the system. These complexes have been relatively unyielding to direct study due to the indistinct and overlapping spectra of multiple cofactors and incomplete structural information. However, such cofactor-level resolution is critical to resolve questions of cofactor contribution. With existing techniques, this resolution requires a simplified system.

There exists no simplified system transparent enough to uncover each engineering element’s contribution to function. Past work at designing proteins has proven that some functions can be replicated in non-natural systems, but there was still sufficient complexity of the system to obscure full understanding (Koder et al. 2009; Farid et al. 2013). This is evident by the slow progress at engineering more oxidoreductase functions using past designs. This thesis aims to demonstrate that a natural oxidoreductase function central to oxygen catalysis, oxygen binding, can be reengineered from first principles with full transparency of design. Oxygen binding was a
first choice as it has been demonstrated to be possible in simple four-helix bundle scaffolds, or maquettes, and has clear metrics of success against both past designs and a family of natural proteins. As the first step in a number of catalyses, it also has functional appeal as more enzymes can be designed using engineering guidelines uncovered by its transparent rebuilding.

In addition to outlining guidelines for stable oxygen binding, this work aims to reduce complexity and improve quantification of ROS production through the development of the first systematically variable protein scaffold, or maquette, designed to study mitochondrial ROS production in a protein environment. The proteins in this work will be inspired by the natural mitochondrial enzymes, but bear no sequential or structural resemblances to their natural counterparts. This system allows for fundamental questions of ROS production chemistry in a protein environment to be asked without obscuring natural complexity. Maquettes can be modified in a stepwise manner to introduce functional groups and cofactors relevant to sites found in mitochondrial CI and CIII that are implicated in ROS production. The simplicity and plasticity of an engineered system allows for similar scaffolds to be directly compared to one another. In addition, maquette systems can be studied at high enough concentrations to quantify SO production. These results of SO production from many cofactors will give insight into the mechanisms of mitochondrial ROS production.

This work will specifically focus on the development of a system for ROS study. The goals are to develop a methodology to determine whether SO, hydrogen peroxide or hydroxyl radicals are produced and the kinetics of their production. To accomplish this goal, heme b-containing maquettes were designed and systematically modified to alter their oxygen reactivity towards the design of obligate SO and peroxide producers. These
maquettes were also biochemically characterized to uncover the possible mechanisms for favoring peroxide over SO production.

1.5 References


Lambert, A. & Brand, M., 2004a. Inhibitors of the quinone-binding site allow rapid superoxide production from …. *Journal of Biological Chemistry*.


Magnitsky, S. et al., 2002. EPR characterization of ubisemiquinones and iron-sulfur

Martyn, K. et al., 2006. Functional analysis of Nox4 reveals unique characteristics compared to other NADPH oxidases. *Cellular Signalling*, 18(1), pp.69–82.


Muller, F., Liu, Y. & Van Remmen, H., 2004. Complex III releases superoxide to both sides of the inner mitochondrial …. *Journal of Biological Chemistry*.


St-Pierre, J., Buckingham, J. & Roebuck, S., 2002. Topology of superoxide production from different sites in the mitochondrial electron .... Journal of Biological ....


2.1 The diverse field of protein design

Protein design can have a lot of different meanings throughout the literature, as it is used for a variety of purposes. Early protein design work focused on structural engineering of protein folds from first-principles while most of the recent literature is functional redesign of natural proteins to new functions. The early structural work, pioneered by DeGrado and Regan in the late 1980s, was successful at determining requirements for helical folding without importing natural sequences in very simple designs (DeGrado et al. 1987; Regan & DeGrado 1988). They developed *de novo* four-helix bundles, which were stable and self-assembling and later they were able to design heme-binding sites into these scaffolds (Choma et al. 1994). Since then, Stroud has been able to show nonfunctional, but structurally sound proteins can be designed in small sets of amino acids as low at 7 residues (Schafmeister et al. 1997). More recent work by the David Baker laboratory has constructed some new folds not existent in biology (Kuhlman et al. 2003), but the field of structural design from first-principles has otherwise remained stagnant.

In contrast to the structural design field, the functional design literature is plentiful. However, the majority of work in this area is on the redesign of natural proteins, not design from first principles. There has been some success in redesigning proteins to perform non-natural functions, though mostly not at natural-level rates. Computational design and directed evolution have dominated these fields. The majority of work begins by computational design and then proceeds to improve activities and rates by directed evolution (Bolon et al. 2002). Some work has been done primarily with computational
design with functional success but limited rates of activity (Lassila et al. 2006; Jiang et al. 2008; Grzyb et al. 2009; Nanda & R. L. Koder 2009). Directed evolution has also also led to successes even on its own, by a method Frances Arnold coins “irrational design” (Arnold 1998). However, it’s arguable whether directed evolution can be called ‘design’ at all, if the researcher does not play a cognizant role in the engineering. Combinatorial methods between the two have had the most plentiful success however, mostly in the Baker laboratory. Baker has had success in designing Kemp eliminases (Khersonsky et al. 2010; Korendovych et al. 2011), Diels-Alderases (Siegel et al. 2010; Eiben et al. 2012) and retro-aldolases (Jiang et al. 2008; Giger et al. 2013) through the redesign of natural proteins' active sites.

The most transparent, and also the most minimal of the functional designs, have used less natural sequence importation. These proteins also consist of the least number of amino acids of all the redesigns, perhaps not out of necessity but because their designers were cognizant of the collateral importation of interdependences with sequence complexity. Functional designs were achieved in 14 amino acids in the redesign of an SH3 domain by Baker (Riddle et al. 1997), 13 amino acids in that of an orotate phosphoribosyltransferase by Yokoyama (Akanuma et al. 2002), and 13 amino acids in the design of an oxygen transport protein by our own lab (R. Koder et al. 2009).

Computational work can lead to transparent designs if all design parameters are fully understood. However, once directed evolution methods are employed, rational and traceable design parameters are inherently obscured by non-rational mutations introduced. The only major functional design successes have avoided directed evolution and hence maintained more transparency to their designs. These design processes also began with a fundamental understanding of the chemistry they wished to engineer in the protein scaffold. The only proteins designed by these methods have been the naturally-
functioning oxygen transporter by the Dutton laboratory, by the ‘maquette’ method of rational, iterative hand-design (R. Koder et al. 2009) and the design of the non-natural chemistry of oxaloacetate decarboxylation via an imine intermediate by Steve Benner (Johnsson et al. 1993). However, as is evident by the vacuum in which these designs exist, all parameters were still not clearly outlined, preventing the easy engineering of more reactions from their designs.

The lack of designs that can be modified to perform alternate catalyses is due to the lack of understanding of the final design. All the functional design successes discussed so far have been engineered with the focus solely on the end product. These engineering processes have produced enzymatic functionality, but have not clearly outlined how it was achieved. So while they may succeed as producing functional catalysts on a one-by-one basis and may have taken the fastest route to a functional design, no engineering principles that can be applied to the design of future proteins or general protein chemistry knowledge is achieved. One protein is developed, in perhaps what is the most efficient means to that one product at the time, but the field is left in the same place. This issue is likely responsible for the slow pace of production of functional designs despite improvement of computational methods over the years.

The step which prevents the transparency of the engineering which would allow for a general understanding of protein material and for general principles to be applied to future design efforts is what Steve Benner has termed the ‘tinkering’ (Luisi & Chiarabelli 2011). Benner describes ‘tinkering’ as, “a process that alters the structure of molecules without much predictability.” He notes that each success exists in a vacuum because finishing steps obscure the full understanding of the achievement. Even when proteins are designed by somewhat rational and transparent methods, such as computational or hand-design, the researchers have always finished with some level of ‘tinkering.’ Most
frequently, this is through the employment of directed evolution to improve the catalytic rate, but it can also be in the form of mutagenesis or non-rational trial-and-error. Benner postulates that this ‘tinkering’ would not be necessary if chemical theory were adequate, as that would allow our rational design processes to succeed on their own.

Figure 2.1: Maquette design. Scheme of maquette design and synthesis (bottom) from a 4-helix bundle (bottom left) to functionalized disulfide-linked 4-helix bundles with various orientations (bottom middle). Maquettes self assemble with hydrophobic regions to their interior (purple) and positive (pink) and negative (blue) residues patterned to the outside and possible polar (yellow) interfacial residues. Apo maquettes show predicted structure by x-ray crystallography and NMR (bottom right). These and subsequent cartoons presented in this work are color-coded by amino acid type (top).

2.2 Maquette design

2.2.1 The approach

Maquette design describes the use of a non-natural scaffold to design natural or non-natural functions from first-principles. The approach relies of iterations of design in which each new intermediate design is biochemically characterized until a desired function is reached. The idealized process would involve small modifications for each
iteration in which only one variable is changed at a time, such that its contribution to the structure, function and other biochemical properties of the protein can be fully assessed. It is practically difficult to achieve such incremental iterations, however, as sometimes mutations have more long-range effects than desired or more than one variable is changed at a time due to imperfect application of the process. In the idealized situation, where no ‘tinkering’ or trial-and-error mutations be employed at the end to improve functionality, the maquette method could be used to develop fully transparent designs.

While the approach aims at transparency, past work has not avoided the seemingly inevitable ‘tinkering’ step before successful natural-level function was achieved nor the simultaneous incorporation of multiple variables between characterized designs. However, this is not an issue with the approach itself that, if followed closely, would produce fully described sequence- and structure-function relationships. Rather, it is a limitation of the approach in its idealism. In order to make small iterations and avoid ‘tinkering’, while still developing a functional product, the chemistry of the function must be well understood. The thesis work presented in Chapters 3 through 5 is made possible by the years of work previously done in the Dutton laboratory with maquettes, as well as former studies of natural proteins performing such functions. In Chapter 3 particularly, this work attempts the idealized maquette approach on an oxygen transporter as there has been much work in this field. For more advanced enzymatic designs, the maquette approach would likely not be the most efficient means to an end product at this point, as their chemistries may not yet well-understood enough to design a maquette with transparency. However, if the approach were applied to these catalyses, we would likely learn much more about the protein chemistry than by other, less laborious design processes such as computational design or directed evolution, informing future maquette design work. While the maquette approach eventually may
prove to be applicable and produce an ease of function for other catalyses, such as synthetic organic chemistry events, maquette performance of oxidoreductase catalysis is a more reasonable start point as a proof-of-principle of the efficacy of the approach. The vast work previously done by other maquette designers over the past twenty years informs the achievements highlighted in later chapters.

The maquette approach has been mainly employed to produce four-α-helix bundles that self assemble as linked tetramers or dimers and more recently single-chains. These bundles self assemble and are fairly stable due to their binary patterning. Sequences of each helix are designed through the patterning of amino acids, which have been divided into generalized groups, of hydrophobic amino acids to the center and alternating positive and negative residues up the outside of each helix for both solubility and hydrogen-bonding purposes (R. Koder et al. 2006; Lichtenstein et al. 2012) (Figure 2.1). These maquettes have also been crystallized to show linear helices and tight packing (Huang et al. 2003). From well characterized, self-assembling, non-functional bundles, functions can be designed. As the Dutton laboratory is primarily interested in bioenergetic proteins and particularly oxidoreductases, electronic cofactors have been incorporated at desired locations to promote electron transfer and gaseous-ligand binding described below.

2.2.1 The accomplishments

Maquettes have successfully incorporated a variety of cofactors through a variety of ligations. Through covalent cysteine linkages, flavins have been attached (Sharp et al. 1998); through synthetic amino acid incorporation, quinones and nicotinamide have been placed as residues in the sequence (Lichtenstein et al. 2009); through amino-acid coordination, iron-sulfur clusters have been bound (Grzyb et al. 2009); and through both
c-type covalent ligation (Anderson et al. 2013) and coordinate covalent $b$-type ligation, hemes and other macrocycles have been incorporated (Farid et al. 2013). This variety of cofactors allow for a large range of accessible redox potentials (Figure 2.2). Even for the same cofactor, it has been shown that small sequence modifications can play a large and predictable role in a single cofactor’s redox potential (Shifman et al. 2000). In addition to simply incorporating different cofactors, it has been shown recently by Goutham Kodali (unpublished work) that iron and zinc macrocycles can be ligated in a variety of fashions, even achieving single cysteine P450-like UV-visible spectra (Figure 2.3). The stability in such simple scaffolds allowing for so many amino acid coordinations in unprecedented and opens the door to the design of a variety of oxidoreductase functions.

Figure 2.2: Cofactors and potentials. Left shows range of potentials of cofactors, as divided into families, when incorporated into maquette proteins. (Low potential flavins seen in organic solvents, but all other potentials measured in aqueous solutions). Right shows cofactors that have been incorporated into maquettes (edited from (Lichtenstein et al. 2012)).
In addition to these structural successes, there have been some functional successes using maquette proteins. The highest level function achieved has been stable oxygen-binding. This was first achieved in a disulfide-linked dimer (R. Koder et al. 2009) and later in a single-chain construct (Farid et al. 2013). Additionally, photoactivation has been achieved using both flavins (Sharp et al. 1994) and zinc tetrapyrroles (Farid et al. 2013). These successes have allowed for some general guidelines to be described for the future design of other functions. Intercofactor distances optimal for electron transfer, water inaccessibility to promote stable oxyferrous state formation, and helical strain necessary for histidine de-ligation for gas binding have all been described through this work. Additionally, thermal stability control has begun to be described through the modification of hydrophobic core packing where larger aromatic residues lead to higher melting temperatures as does bis-histidine heme ligation (Figure 2.4).

While some guidelines have been described, the sequences produced were not fully understood due to the addition of more than one modification between biochemical characterizations. However, these functional successes in

![Heme b ligations to a single maquette by UV-visible spectroscopy.](image)
completely non-natural sequences have proved the principle that the functionality is not unique to the fold of its natural counterpart. For example, this work on the oxyferrous state stability has demonstrated that the globin fold is not necessary for oxygen binding and natural rates can be achieved in sequences and structures baring no resemblance.

Figure 2.4: Thermal melts as measured of loss of $\alpha$-helical content of a series of single chain heme-$b$-binding maquettes with sequence modifications (A) and cofactor incorporation (B). The $T_m$ increased from 37°C for histidine core residues to 55°C when replaced by alanine and 95°C when replaced by phenylalanines (A). Heme-binding also induces an increase of $T_m$ for each domain. The melting temperature is increased for half the protein with one heme binding such that a split of 38 and 72°C is seen. Similar results are seen for obligate one-heme binders with a split of 39 and 75°C. When both heme sites are occupied, both domains of the protein melt at the higher temperature, 72°C. Rendering reprinted from thesis of Lee Solomon (Solomon 2013) with data published thereafter (Farid et al. 2013).
2.3 Applications of maquettes

2.3.1 Maquettes in protein engineering

Maquettes have many properties to make them ideal for industrial protein engineering. Their stability, small size, expressibility and domain independence all contribute towards their ease of production and modification. As demonstrated in figure 2.4, hydrophobic core packing can be used to control thermal stability of maquettes. All these maquettes also bind cofactors, such as heme $b$, in the nanomolar regime (Farid et al. 2013) and self-assemble in seconds in solution (Solomon 2013), making them easy to work with over long periods of time. There is also the possibility of covalent attachment of flavins as well as hemes (Anderson et al. 2013), allowing for the production of constructs that will not lose their cofactors over time. Additionally, these proteins are highly expressible in E. coli with gram quantities per 2-liter culture.

High production yields and stability are important for industrial applications, but not special to the maquette method. Expression could be optimized for other protein designs or redesigns developed through alternative methods such as computation or directed evolution, as could their thermal stability and cofactor affinities be improved. What makes the maquette approach ideal is that it allows for modularity of design unlike others that ‘tinker’ or use obscuring methods to the transparency of the design.

Directed evolution and computational approaches which start with natural protein sequences inherently import natural complexity. As described in Muller’s ratchet, evolution incorporates deleterious mutations over iterations, as it is a blind process to the end product. These blind mutations also lead to interdependences as redundancies lead to the removal of previously necessary components, such that new, more complicated pathways are essential for function (Figure 2.5). Even if natural proteins
Figure 2.5: Illustration of Muller’s Ratchet. A functional assembly of parts in an efficient form as a bridge, is added to a redundant part, which can allow formerly necessary parts to be removed. Once these parts are removed, the formerly non-essential part is now necessary, creating irreversible complexity of the system as it is now more complicated and fragile than previously engineered.
could be efficient for their purposes, one natural protein may have a multitude of functions that it has evolved to perform, often with competing structural requirements. As a result, neither function is optimally executed, only sufficiently so, and neither are often robust to modification due to interdependencies. Beginning with natural proteins for industrial catalyst design inherently starts with a non-optimal protein fragile to modification.

![Figure 2.6: Maquette approach to functional designs. Many cofactors (left), such as hemes (brown), chlorines (blue), FeS clusters (red) and flavins (yellow) can be incorporated into a single scaffold (grey) by engineering in binding sites at many locations (green, non-exhaustive). These cofactors can be ligated in many combinations to engineer a variety of oxidoreductase functions (right).](image)

It is only when proteins are designed from first principles with no import of natural sequences that non-necessary interdependences can be avoided. In avoiding these interdependences, the resulting protein becomes more robust to modification and a single sequence can be used to successfully incorporate a multitude of cofactors at varying sites to perform a wide range of functions (Figure 2.6). Because these minimal sequences have little structural independences in addition to avoiding functional
redundancies, cofactors can be inserted in many locations without affecting cofactors in other locations. The result is a modular design in which each domain acts independently and designs can be constructed akin to molecular ‘Legos’. While past designs were not fully transparent, they did achieve a high level of domain independence. The thermal melts show in Figure 2.4 demonstrate structural independence, as addition of a heme to one domain does not affect the melting of the other domain. Functional independence has also been shown in oxygen-binding maquettes through the ability to bind oxygen.

Figure 2.7: Oxygen binding shows functional domain independence. In the same maquette with one-heme (top left) and two-hemes (top right) bound, the same spectral bands are seen for the oxidized (green), reduced (blue), oxyferrous (red) and carbon monoxide bound (black) with exactly two times the absorbance for the two-heme version. The oxyferrous state formation (lower left) and decay (lower right) are also within error of each other for the one-heme (green) and two-heme (black) maquettes. Data re-rendered from that previously published (Farid et al. 2013).
identically, and independently to each heme in a two-heme binding maquette. Figure 2.7 shows that in a single-heme containing oxygen transport maquette, the oxyferrous state is formed in exactly half the quantity of a two-heme binding maquette with identical UV-visible bands and oxyferrous formation rates and lifetimes.

Modularity of designs when combined with transparency of the engineering of functional elements opens the door for the ease of construction of many functions. The limit to the functions that can be engineered is the number of functional elements described with full transparency. In this work, oxygen binding will serve as the first fully-transparent functional element to be described in a protein system.

2.3.2 Maquettes as a model system

Maquettes can also allow us to learn more about natural systems. By providing a simple transparent framework in which we can modify a single variable at a time, maquettes allow us to ask fundamental questions about protein chemistry. Natural oxidoreductases are notoriously difficult to study due to their interdependences and multiple functions inherent to all natural systems as well as their incredible complexity discussed in Chapter 1. Each bioenergetic protein may contain a variety of cofactors leading to spectral degeneracy in experiments. Due to their interdependences, they also are fragile to mutagenesis, with single point mutations often having long-ranging, unpredictable effects to both the structure and function.

The use of maquettes to ask many of the same questions we would like to know in relation to natural systems can help us learn about basic protein chemistry and apply it back to understanding the natural systems. Maquettes have already been used for this some in the past, with our engineering of a oxygen-binding protein illuminating some of the functional requirements for oxygen-binding in hexacoordinate globins (R. Koder et al.
2009). Amphiphilic versions of maquettes have also been incorporated into lipid vesicles to study protein-mediated transmembrane electron transfer (Discher et al. 2005). However, to fully realize maquettes applications to understanding natural systems, all unnecessary complexities must be avoided. To achieve this, fully transparent maquettes must be designed.

2.4 What is next for maquette design

Towards both using maquettes for protein engineering as well as understanding natural systems, it has become clear that a more rigorous approach to our design is necessary. By designing in such a manner where single variables are changed between iterations and each iteration is biochemically characterized, we will develop designs for which the sequence-function relationship is transparent. If we fully understand our protein chemistry and effects of each modification, no ‘tinkering’ will be necessary in order to achieve efficient catalytic rates or yields. In these designs, we will understand each functional element employed to engineer the target function, allowing us to use these guidelines to engineer future designs without returning to ‘scratch’ as well as apply them towards understanding how natural systems may work.

2.5 References


Discher, B. et al., 2005. Design of Amphiphilic Protein Maquettes: Controlling Assembly, Membrane Insertion, and Cofactor Interactions†. *Biochemistry*.


Grzyb, J. et al., 2009. De novo design of a non-natural fold for an iron-sulfur protein: Alpha-helical coiled- ..... ... et Biophysica Acta (BBA ....


Chapter 3: Designing Minimal Oxidoreductases

3.1 Introduction

3.1.1 Virtues of minimal design

Through natural selection, nature has created proteins with the many functions required to sustain life. Yet natural selection is a blind process that typically results in complex and often relatively fragile proteins in which both the engineering behind the functions and the multiple, overlapping roles played by each amino acid are obscure (Darwin 1859; Muller 1964) (Figure 3.1). Through our work with structurally simple man-made proteins, we have found that function does not require such complexity, and that the maquette method of modular construction using intentional simplicity and iterative redesign can identify the functionally important roles played by many individual amino acids (Anderson et al. 2008; Koder et al. 2009; Farid et al. 2013).

However, we have yet to design towards the level of transparency that has been achieved for structural protein designs. Early protein design work outlining the first-principles of protein folding (DeGrado et al. 1987; Regan & DeGrado 1988; Regan et al. 1994) proved transparent and useful in turning energetic and structural folding propensities of amino acids into stable secondary peptide motifs and tertiary folded proteins. Robustness of these principles has since become evident in developments leading to the creation of proteins displaying folds absent in nature. While these achievements are notable, successful extensions with the intent to reproduce and match the functional characteristics of natural enzymes in man-made proteins have not been accomplished. Even though we have matched natural rates in designed proteins (Koder et al. 2009; Farid et al. 2013), we have not done so from first principles and therefore
have not achieved complete transparency in our designs (Figure 3.1). Achieving understanding of each step of the design process is far from trivial, as it can be incredibly laborious, particularly if the protein material or chemistry is not well understood for reasons noted in Chapter 2. Clearly we had yet to know our material.

The few cases in which natural functions and rates have been achieved in man-made enzymes have been informed by a fundamental understanding of how the natural enzyme mechanistically works prior to the engineering (Koder et al. 2009; Farid et al. 2013; Johnsson et al. 1993). This knowledge has allowed for outlining of a clear list of components to be engineered into a given scaffold. While such examples are still not as transparent as the preceding structural design work, these well founded cases hold the potential to begin to identify functional component parts and document analytical protocols that ultimately may prove general in transforming our knowledge of protein folding into man-made protein structures displaying prescribed catalytic functions.

In this chapter, the work aims to replicate a natural function in a manner as transparent as the former structural work by designing from first-principles. The chosen function for this proof-of-principle study was stable oxygen binding, akin to natural oxygen transport proteins such as myoglobin and neuroglobin (Hasinoff 1974; Dewilde et al. 2001), because it is relatively sophisticated, but previously shown to be designable in a four-$\alpha$-helix bundle (Koder et al. 2009). In order to maintain transparency through the design process, minimal changes are made to achieve each component part of the target function of oxygen binding. By doing the minimal amount necessary and using a stepwise addition of component parts, we can uncover the requirements for each without importing or designing unnecessary interdependencies between parts or sub-functions. In this way we identify the minimal design requirements for such basic physical chemical elements as cofactor ligation, residue mobility and exclusion of water from the bundle
interior, that underlie the oxygen binding engineering in proteins.

Avoiding interdependencies common to natural protein systems, we gain the ability to modify the protein more freely and engineer without the need to return to first principles. We achieve simple, modular designs which can be combined in multiple ways to achieve a variety of functions rather than just a functional end design from which one can only design forward through iterative trial and error, like past work. This feature will allow for the design of increasingly refined functions involving multiple sites with the protein and multiple cofactors without the necessity to start from scratch each time.

Figure 3.1: Complexity of natural and man-made proteins. For natural proteins, Darwin's principle of multiple utility (Darwin 1859) states that each amino acid or property (purple) contributes to multiple functions (grey) in a protein. Mullerian interdependency (Muller 1964) states each function is the result of cooperative effects of many amino acids or properties. Former designs were greatly simplified for each amino acid, though not all interdependencies were defined. For the minimal proteins designed here, gross properties were engineered rather than focusing on the amino acid level to achieve transparency of how each function was accomplished and the contributions of each property. Modified with permission (Anderson et al. 2008).
3.1.2 Former oxygen transporter designs

In this chapter, a reconstruction of a stable oxygen binding protein is presented with full transparency to its engineering components. Oxygen transport was chosen as a function as it is the most advanced of the successful man-made reproductions (Koder et al. 2009; Farid et al. 2013) and constructed in a four-α-helical bundle protein and hence structurally quite removed from that of the functionally inspiring iconic hemoglobin fold (Perutz 1970). Thus it is not surprising that despite the two quite different folds sharing the same palette of twenty amino acids and requiring analogous functional components to engineer oxygen transport, the amino acid sequences and motifs used to construct them are distinct. These distinctions, sharpened by the absence of any mimicry of the natural globins in the man-made reproduction, provide a basis for the work presented here in which an α-helical bundle protein is analyzed to disentangle and abstract individual engineering and structural elements that contribute to its oxygen transport function.

Former work has clearly outlined some of the necessary components for oxygen binding, despite not clearly identifying the sequential or structural factors contributing to their engineering. Oxygen binding was first achieved in a symmetric dimer (Koder et al. 2009) and later applied to a single chain construct with identical helical sequences (Farid et al. 2013). Work on both the dimer and single chain have outlined the importance of water restriction from the core heme-oxygen binding site for oxyferrous state stability as well as outlined how to achieve dryness (Anderson et al. 2008). Additionally, the necessity of the heme-adjacent buried glutamates in the formation of the oxyferrous state has been clearly defined through reverse engineering (Koder et al. 2009; Zhang et al. 2011; Zhang et al. 2012). These glutamates become buried upon heme-binding helical rotation, presumably inducing helical strain. In order for oxygen to bind to the
heme in these hexacoordinate bis-histidine ligating globins, one histidine must be released from the axial site of the heme. In order for this removal to happen quickly and often enough for efficient oxygen binding, there must be a low energetic barrier for its removal. It has been suggested that the burying of the glutamates lower the barrier for histidine removal to allow of effective oxygen-binding at the axial site (Figure 3.2).

![Figure 3.2: Heme and oxygen binding to maquettes. Looking down the top of a symmetric four-helix bundle (grey) with a hydrophobic interior (purple) and designed glutamate residues (red triangles), we see glutamate rotation into the hydrophobic core upon each heme-binding (red squares) event to the bis-histidine (green) site (top). Upon oxygen (red circles) binding, this rotation is reversed (bottom), releasing the glutamates back into the aqueous solvent from the protein core.](image)

3.1.3 Limited amino acid libraries

Evolution has built an extraordinary amount of complexity into the proteome (Edelman & Gally 2001), including over 1000 different folds and 20 amino acids (Marsden et al. 2006; Thornton et al. 1999; Chothia 2003). Questions as to the minimal requirements for protein evolution and variety of amino acids have related not only to the chemistry residues can perform (McDonald & Storrie-Lombardi 2010), but also whether
they constitute a ‘foldable set’ (Longo & Blaber 2012; Fan & Wang 2003; Murphy et al. 2000). Protein design has proven to be a useful tool for the determination of functional requirements (Longo & Blaber 2012). As discussed above, design has demonstrated oxidoreductase function does not require complex folds, as oxygen binding (Koder et al. 2009), interprotein ET, ROS production and light-activated charge separation (Farid et al. 2013) have been achieved at natural rates in four-α-helix bundles.

In addition to using these much-simplified folds, designed proteins have also shown that reduced sets of amino acids can produce functional enzymes. Protein database analysis by Romero and colleagues suggested that the lower bounds of amino acid variety for globular proteins to be 10 (Romero et al. 1999) while Stroud has demonstrated stable nonfunctional proteins can be formed in 7 (Schafmeister et al. 1997). Experimental designs have achieved lower limits of 14 amino acids in the redesign of an SH3 domain by Baker (Riddle et al. 1997), 13 amino acids in that of an orotate phosphoribosyltransferase by Yokoyama (Akanuma et al. 2002), and 13 amino acids in the design of an oxygen transport protein by Dutton (Koder et al. 2009).

Reduced amino acid sets may produce protein with greater potential catalytic activity in addition to simplicity. Recent work in directed evolution libraries has shown that a restricted amino acid set not only accelerates the evolutionary process, but results in products with greater rates and specificity (Reetz et al. 2008). Here we attempt to determine how few amino acids and which design elements are necessary for function. In exploring these questions, we have improved from our own former designed oxygen transport protein, further suggesting sequence complexity is encumbering to function.

3.2 Designing functional requirements independently

Engineering of a hexacoordinate oxygen transporter, typified by neuroglobin,
Figure 3.3: Design of a minimal oxygen-binding protein from first principles. Cartoons are color coded by hemes (brown) and amino acids that are charged (grey), hydrophobic (purple), histidines (green) and alanines (black).
began with a rational evaluation of supporting physical characteristics or functional requirements that must be met to achieve the desired terminal function. Informed by natural oxygen transport proteins (Kiger et al. 2011) and our prior oxygen-binding designed maquette proteins (Farid et al. 2013), we determined five functional requirements for the engineering of a oxyferrous heme state on the order of tens of seconds: first, we start with a stable water-soluble α-helical protein; second, hexacoordinate bis-histidine heme ligation must be achieved in such a way as to introduce strain to the histidine-heme ligation to promote rapid histidine to O$_2$ exchange; third, we establish high affinity heme binding in the interior of the α-helices; fourth, water must be restricted from the heme site of the protein to stabilize the oxyferrous heme state; fifth, we develop an equilibrium affinity favoring O$_2$-binding and formation of the oxyferrous heme site at prevailing aqueous O$_2$ concentrations. Each of these functional elements was individually engineered in a sequential manner, with the exception of the fourth and fifth which were achieved through the same engineering element, and evaluated for contributing factors to each achievement. This chapter will focus on the step-by-step design of protein-induced heme properties with protein structural and biochemical characterization presented in the following chapter due to the surprising nature of the results, such that they can be discussed more fully.

3.2.1 Design of a stable water-soluble α-helical protein

Design began with a sequentially and structurally minimal protein, which can self-assemble, is highly soluble in water, can be modified without perturbing the tertiary structure and is capable of binding multiple cofactors. Four-α-helix proteins have previously been shown to meet these characteristics, so were chosen as a starting point (Farid et al. 2013; Huang et al. 2004; Shifman et al. 2000; Gibney et al. 2001). The
starting helical protein, 1, was developed from binary-patterned α-helical sequence capable of self-assembly into a bundle of four α-helices based on work by DeGrado and Regan (Regan & DeGrado 1988). They had previously shown that a bundle containing only the amino acids leucine, lysine and glutamate would self-assemble and be water-soluble. Their sequence was elongated to the length appropriate for two cofactor binding as demonstrated through prior work from Dutton and colleagues (Huang et al. 2004) (Figure 3.3). These helices were then linked into an expressible, single-chain four-helix protein, 2, with 9 residue loops containing glycine and serine, as has been previously shown to be optimal for self-assembly (Farid et al. 2013), and a single tryptophan was added to the second loop for spectroscopic identification purposes. The protein 2 does not show any heme binding because of the lack of any heme-ligating residues and thus further characterization of gas binding properties was not possible (Figure 3.4).

![Graph](image)

Figure 3.4: Heme Titrations of (A) 5, (B) 4, 3, (C) 2 and 5A9EA79E. Protein concentrations were 10 nM for 5 and 1 µM for 1, 2, 3, 4, and 5A9EA79E in 20 mM TRIS, 20 mM NaCl pH 8. Data was fit using Equation 3.1.

\[
\text{Abs} = [H] \times \varepsilon_{\text{free}} + (\varepsilon_{\text{bound}} + \varepsilon_{\text{free}}) \times \frac{(K_D + [P] + [H]) - \sqrt{(K_D + [P] + [H])^2 - 4[P][H]}}{2}
\]

Equation 3.1: Fitting equation for single heme binding affinity constant $K_D$, where [P] is the protein concentration ad [H] is the heme concentration in moles, for plots of the Soret band maximum absorbance for bound heme (Abs$_{412}$) plotted against concentration of heme titrated.
3.2.2 Design of hexacoordinated bis-histidine heme ligation that promotes rapid histidine to O₂ exchange

Hexacoordinate histidine-heme ligation was introduced in two symmetrical sites through the mutation of the leucine in the 6ᵗʰ position of each helix to histidine, forming maquette 3 (Figure 3.3). The 6ᵗʰ position was chosen based on prior work discussed in section 3.1.2 that demonstrated the hexacoordinate heme ligation drove an unfavorable >50° rotation of the adjacent glutamate-7 toward the hydrophobic interior to induce a helical strain (Figure 3.2) (Farid et al. 2013; Koder et al. 2009; Zhang et al. 2012). The burial of these glutamates has been verified in previous maquette proteins by observance of their pKₐ modulation upon heme binding (Gibney et al. 2001). This in turn promotes a lowering of the energetic barrier for histidine to come off and be displaced for gas binding; upon binding the dissociated histidine is thought to become a ‘distal’ histidine for bound dioxygen.

Maquette 3, whose only pro-heme-binding engineering element is histidine addition, forms a bis-histidine ligation with a Kᵅ of 1.12 µM (Table 3.1, Figure 3.4). This relatively low affinity does not meet the functional standard set by former oxygen-binding protein designs or natural globins which have Kᵅ values in the nanomolar range (Dewilde et al. 2001), creating some difficulty in characterization of the heme site. The oxidation-reduction (redox) midpoint potential (Eₘₚₜₚ₉₈) of 2 heme b-binding maquette 3 was determined by redox potentiometry in both the reducing and oxidizing directions (Figure 3.5). Because of poor heme-binding affinity, 3 showed a hysteresis between the course of reduction (Eₘ of -154 mV ) and oxidation, (-175 mV). The mean potential of -164 mV is consistent with other earlier designed proteins that did not contain large hydrophobic core residues for the exclusion of water (Gibney et al. 2001; Shifman et al. 2000).
UV-visible spectra at -15°C in 30% ethylene glycol were able to be observed for both O₂ and CO binding, although a mixture of free and bound heme were observed because the heme-protein affinity was too low for full retention of heme throughout the experiment in 30% ethylene glycol solutions (Figure 3.6). Maquette 3 demonstrates a minimal ability to bind oxygen. This was not entirely surprising since maquette 3 contains the glutamates in placement previously demonstrated to be necessary for oxygen binding because the histidines were introduced adjacent to glutamates, as described above. In an environment of 50% CO and 50% O₂ at -15°C maquette 3 binds CO and O₂ with a 7:1 ratio, despite oxyferrous state formation in oxygen-saturating conditions, likely due to water access to the core destabilizing the oxyferrous state. Although 3 formed a more prominent oxyferrous state by low temperature UV-visible spectroscopy, the heme loss at 15°C was too rapid to determine oxyferrous lifetime by stopped-flow analysis.

![Graph](image)

Figure 3.5: Redox titrations of 3 (black), 4 (blue) and 5 (red) in the reducing (square) and oxidizing (circle) directions as observed by reduced heme b Q-band absorbance at 558 nm.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Kₜ Heme b (µM)</th>
<th>Eₜ (mV)</th>
<th>O₂/CO Preference</th>
<th>kₜ e⁻ (s⁻¹)</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.12</td>
<td>-164</td>
<td>12/88</td>
<td>*</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>0.13</td>
<td>-213</td>
<td>0/100</td>
<td>0.313</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>0.00017</td>
<td>-202</td>
<td>55/45</td>
<td>0.045</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 3.1: Heme-binding protein characterization
3.2.3 Design of high affinity heme binding in the interior of the α-helices

In order to improve heme-binding affinity in maquette 4, a heme-binding pocket was introduced to the core and the heme-binding rotational barrier (Figure 3.2) was lowered by changing hydrophobic residues that undergo solvent exposure upon heme binding (Figure 3.3). The heme pocket was introduced by removing hydrophobic bulk of residues near the heme site through the introduction of alanine at the 9 position of each helix. Heme-binding rotation was eased through the removal of solvent-exposed leucines at the 24 positions of helices 1 and 2, where they were replaced with highly soluble glutamates. Glutamates at the 7, 14 and 21 positions of each helix, which experience differing degrees of desolvation upon heme binding, were maintained because they have been shown to be necessary for oxygen binding in prior work, presumably through lowering the energetic barrier for breaking the histidine ligation during oxygen-binding (Koder et al. 2009; Zhang et al. 2012).

Removing these barriers to heme binding for 4 increases the heme-binding affinity, lowering the $K_D$ value to 130 nM (Table 3.1, Figure 3.4). This greater affinity allowed for greater ease in characterization of both the midpoint potential and gas-binding. A redox titration showed no hysteresis and a potential of -213 mV for 4 (Figure 3.5). Contributions to the depression of the redox potential in 4 relative to 3 is may result from the lowered charge state from +4 to +2 and increased burial of the heme leading to increased hydrophobicity of its environment.

Low-temperature UV-visible spectra of 4, like 3, also showed a mixture of free and bound heme as the heme-binding affinity is lowered in the 30% ethylene glycol environment (Figure 3.6). 4 displays a strong tendency to bind CO not only in CO-saturating conditions, but also when exposed to a 50% O$_2$ 50% CO gas mixture. In contrast, 4 does not show any detectable oxyferrous state formation in oxygen-
saturating nor 50% O$_2$ 50% CO conditions. This suggests the increased access of water to the protein interior with the addition of the heme-binding pocket, which removed some of the hydrophobic core bulk near the gas-binding site. However, even though oxyferrous state formation was not observed at -15°C, a transient oxyferrous state was observed from its characteristic Q-band at 574 nm in the seconds time domain by stopped-flow UV-Visible spectroscopy (τ = 3.2 s) (Figure 3.7).

![Figure 3.6: UV-Visible spectra at -15°C at 5 µM protein with 2 hemes bound in 25 mM sodium phosphate pH8, 30% ethylene glycol. Spectra in the oxidized (green), reduced (blue), carbonmonoxyferrous (black) and oxyferrous (red) states of (A) 3, (B) 4, and (C) 5. (D) Spectra when exposed to 50% CO, 50% O$_2$ gas mixture of 5 (blue), 4 (black), and 3 (red).](image-url)
Figure 3.7: UV-visible kinetics of 4 (A) and 5 (B) as observed after stopped flow mixing of reduced 10 µM protein with oxygen-saturated 20 mM Tris, 20 mM NaCl pH 8 at 15°C of the reduced (blue), oxyferrous (red) and oxidized (green) heme state. Data is presented as fraction of total as determined by SVD analysis and confirmed by monitoring of Q bands of reduced heme at 558nm and oxyferrous state at 574nm.

3.2.4 Design of water-restricted heme site and development of an equilibrium affinity favoring O₂-binding at prevailing aqueous O₂ concentrations

In order to lengthen the lifetime of the oxyferrous state, the O₂-binding affinity at natural aqueous concentrations was improved through restricting water access to the core in protein 5. Introduction of a heme-binding pocket in 4 may have also allowed for increased water access to the core in comparison to 3 as seen through the destabilization of the oxyferrous state. For 5, we desired to add back some core bulk without removing the heme-binding pocket (Figure 3.3). The H+7 and H+14 positions of
each helix, which are internal to the protein directly above the histidine residues, were all mutated to phenylalanine to add hydrophobic core bulk, as has been suggested in previous work (Koder et al. 2009). These positions for phenylalanine were also chosen due to their ability to interact through \( \pi \)-stacking with the porphyrin ring of the heme bound to the opposing helix at the H+14 position and improve the heme-binding affinity. These mutations removed the alanines introduced on opposing helices in 4, but maintained the alanines introduced at the 9 positions.

Further improvement of heme-binding affinity was seen with the addition of phenylalanines in 5, which has a \( K_d \) of 170 pM (Figure 3.4). This affinity is greater than any designed oxygen-binding protein and allowed for much cleaner determination of its character. Because the design of 4 from 3 including the removal of both solvent exposed hydrophobic residues and addition of a heme-binding pocket and such a strong improvement of affinity was observed with the addition of core phenylalanines in 5, the necessity of the heme-binding pocket was tested through the removal of half of the pocket by mutating interfacial alanines on the first and third helix to glutamate. 5A9EA79E did not bind heme (Figure 3.4), indicating all of the alanines at the 9 positions of each helix are necessary for heme binding.

5 showed a similar redox potential to 4 at -202 mV as it has a net charge of +2 and both have well-buried hemes (Figure 3.5). There was also little difference in their thermal melting temperatures, despite previous designs showing a drastic increase in melting temperature with phenylalanine content. In this case, both 4 and 5 have high melting temperatures, showing only uncooperative fraying with increased temperatures (Figure A2.11). Strong oxygen-binding and heme affinity of 5 is evident in the UV-visible spectra taken at -15°C. Clearly defined oxidized, reduced, oxyferrous and carbonmonoxyferrous states are seen (Figure 3.6) which correspond to those observed
for natural hexacoordinate globins (Dewilde et al. 2001). The addition of phenylalanines in 5 prevents water access to the heme while allowing for tight heme binding, preventing heme loss and resulting in singular state spectra. UV-Visible spectra taken at -15°C of 5 when exposed to an environment of 50% CO and 50% O₂ showed little preference between the two gasses, with the curve fitting best to a 55% oxyferrous state and a 45% carbonmonoxyferrous state model. By stopped-flow kinetic measurement of the Q-band at 574nm, an oxyferrous lifetime of 22.4 seconds was observed for 5 (Figure 3.7). This matches those observed for natural hexacoordinate globins (Uzan et al. 2004; Kiger et al. 2011; Dewilde et al. 2001) and is longer than that observed for former, more sequentially complex, designs of 14 seconds (Farid et al. 2013; Koder et al. 2009). We observed a near complete oxyferrous state formation.

3.2 Conclusions about natural systems: Lower limits of amino acid diversity for a functional protein

Here it has been shown that the lower limit of amino acid variety to produce a stable and functional protein is less than 10. This finding suggests that a function can be achieved in a simple four-α-helix bundle protein of 6 to 8 amino acids (excluding tryptophan which was only added for spectroscopic detection of protein concentration). We show that primitive globin activity is achieved only with heme binding to an initial sequence of 6 amino acids. Furthermore, we show improvements to both heme affinity and oxyferrous-state lifetime only require an increase to 8 amino acid types. By modifying gross properties of the sequence to achieve tight heme binding and adding hydrophobic core bulk, we can achieve natural functions.

This work demonstrates that a sequentially and structurally simple protein can achieve high-level natural function. Oxygen-binding globin function was achieved in a
14.5 kDa four α-helix protein containing only 8 amino acids. We also show that oxygen binding to a hexacoordinate heme-ligated protein can be achieved with only 6 amino acid types. In order to modify a functionless sequence into one that can bind oxygen, only histidines to bind heme were required. 3 bound heme with a K_D of over 1.12 µM and did not fully exclude water from its core, and so both heme and oxygen bound states were unstable. Improvement of heme binding in 4 to a K_D of 130 nM only required the introduction of four alanines and two glutamates at interfacial regions, to remove steric constraints to heme binding by creating a pocket. The removal of hydrophobic leucines destabilized the oxyferrous state in 4, likely due to water access. In order to maintain a watertight core while improving the heme affinity, phenylalaines were added at the 13 and 20 positions of each helix to add core hydrophobic bulk. 5 bound heme even tighter, possibly due to phenylalanine stacking with porphyrin ring constituents, with a K_D of 170 pM. The tight ligation of heme did not preclude histidine coming off for oxygen to bind, as complete oxyferrous state formation is observed. The phenylalanines also prevented water access to the core more than any previously designed protein (Farid et al. 2013; Koder et al. 2009), as the oxyferrous lifetime of 22.4 s is the longest observed for any designed protein and is comparable to that of the best oxygen-binding natural hexacoordinate globins (Kiger et al. 2011).

3.3 Conclusions about engineering: A new paradigm for protein design

3.3.1 Orderly reconstruction from first principles by testing intermediate designs

Here we present the orderly reconstruction of an oxygen transporter from first principles. This approach differs from past non-natural reproductions of oxygen transport function discussed earlier, which had success in replicating the oxyferrous state
formation with natural rates without resorting to mimicry or natural sequence or structure incorporation, but was not engineered in such a manner as to allow a separation of functional component parts. Designs towards other functions have also used and end-focused approach (Riddle et al. 1997; Akanuma et al. 2002; Siegel et al. 2010), with the exception of metal-free oxaloacetate debarboxylase design by Benner that described necessary design components for the reaction before the target function (Johnsson et al. 1993). By designing each functional requirement for the target parent function in a sequential manner with characterization of each intermediate design, the sequence is transparent in its construction. This transparency allows for the removal of individual functional elements as well as the incorporation of such elements into other proteins for other design purposes.

The reconstruction method utilized here develops a practical relationship between abstracted functional component-parts with the earlier transparently designed structural folding elements absent of intended function (DeGrado et al. 1987). The analysis began from a DeGrado design which had been defined from first principles to be the simplest heptad repeat to spontaneously fold into a four-α-helix bundle. Engineering elements were introduced incrementally, by the smallest separable individual functional component parts, into the helical structure. At each stage the minimal set of sequence changes are selected that meet engineering requirements as gauged against the standards expressed by natural globins and man-made α-helix protein reproductions. The analysis achieves a new level of transparency in describing the individual roles of amino acids or structural features.

As a result of this work, for future designs or alternative functions we need not start from scratch. Because this design process outlines specific requirements for intermediate functional elements in addition to the finished product with the main target
function, we can use these resulting building blocks to engineer future designs that may or may not use the same starting sequence. This method leads to an understanding of the material, not just the end function, towards the production of a molecular toolbox generalizable for any protein design.

3.3.2 Engineering through modification of gross properties

We were able to be successful in a simple orderly reconstruction because we did not focus our alterations at the single amino acid level. Prior rational protein design work has constructed through point mutations and using angstrom-resolution models of active sites (Giger et al. 2013; Privett et al. 2012; Zastrow et al. 2011; Siegel et al. 2010; Faiella et al. 2009; Jiang et al. 2008). While there has been modest functional success, these achievements require ‘tinkering’ after the rational design process, usually through random mutagenesis or directed evolution (Bommarius et al. 2011), which obscure the transparency of the design. In contrast, here we approached a single design change as the modification of a single gross property of the protein. These changes, particularly when made to a repetitive sequence, can take the form of a group of mutations, each at analogous sites within each domain to adjust a property of the protein in general. While this requires more mutations per iteration, it is arguably a simpler change than a single point mutation without fully defined effects if it is observed to not effect other properties in unexpected ways.

Here we chose a set of requirements to be met and then engineer them through modifications of the protein’s overall properties, not residue-by-residue mutation. It was this designing by gross properties which allowed for the ease of the design process. Rather than seeing proteins as a special material due to the natural evolution of sequence and folds, we can treat them as any other polymer and design, as polymer
chemists have been doing for years (Maranas 1996; Bergbreiter et al. 1998; Wulff 2002; Toorisaka et al. 2003; Madhavan et al. 2008), by modification of each repetitive unit in tandem to effect large-scale property change. This approach side steps many of the problems of using high-resolution structures to engineering angstrom-level change of an active site, such as crystal structures not being dynamically accurate and snapshots biased by the precipitation conditions. Even nature designs in a manner more akin to this process than by point mutation; diseases are often caused by point mutations having long-ranging unexpected effects to a protein’s function (DePristo et al. 2005) and homologs of proteins with identical functions of organisms which live in different conditions have proteins which differ in their gross properties throughout the sequence, such as thermophile proteins lack of polar residues (Fukuchi & Nishikawa 2001; Dominy et al. 2004; Sadeghi et al. 2006).

This work sets the stage for the following chapter, which demonstrates despite significant simplification of this four-α-helix oxygen transporter and engineering at a gross level without high-resolution structural information, these designs can display hyperthermal stability and maintain their heme ligation and structure for months.

3.4 References


Huang, S.S. et al., 2004. The HP-1 maquette: from an apoprotein structure to a


Reetz, M.T., Kahakeaw, D. & Lohmer, R., 2008. Addressing the Numbers Problem in


4.1 Introduction

4.1.1 Stability and its role in protein design and engineering

Stability is an essential design feature of any practical engineered protein. For \textit{in vitro} engineering of cofactor-binding proteins, such as ones for light-harvesting in solar cells (R. Koder et al. 2009; Sharp et al. 1998), high thermal stabilities and slow degradation are needed for robust materials able to be exposed to harsh temperatures for prolonged periods of time. For \textit{in vivo} designs, such as artificial bloods (R. Koder et al. 2009), functional rescue of cellular functions and light harvesting, stable cofactor binding proteins would guard against cytotoxic levels of free cofactor as seen, for example, with heme (Natanson et al. 2008; Gladwin et al. 2012; Vandegriff et al. 2006).

Despite the practical need for stability, protein-based materials have been traditionally viewed as unacceptably fragile (Grzyb 2012; Wulff 2002). This view was countered by early sequence minimal designs of non-natural proteins that displayed conspicuously high thermal stabilities (Huang et al. 2003; DeGrado et al. 1987; Regan & DeGrado 1988; Nagi & Regan 1997; Skalicky et al. 1999; Schafmeister et al. 1997; Nautiyal & Alber 1999). However, work to engineer function into non-natural protein frameworks suggested that dynamic structures, and therefore lowered melting temperatures were essential to support a variety of catalytic functions (Nanda & R. L. Koder 2009; Gibney et al. 1997; Choma et al. 1994; Farid et al. 2013). For example, Baker's redesign of a thermophilic protein to confer novel Kemp eliminase catalytic activity ended up significantly decreasing melting temperatures (Khersonsky et al. 2010). Here we suggest that this effect is not necessary. By applying insight from recent
discoveries of hyperthermal stability in natural proteins, such as high melting temperatures of holo hexacoordinate globins (Hamdane et al. 2005), redesigned natural proteins (Malakauskas & Mayo 1998; Gribenko et al. 2009; Nakajima et al. 2008) and other naturally thermophilic enzymes (Razvi & Scholtz 2006; Sadeghi et al. 2006; Dominy et al. 2004; Fukuchi & Nishikawa 2001), we can maintain rigid secondary structures while performing functions requiring rapid core movement.

4.1.2 Thermophiles

In the effort to design hyperstable protein systems for industrial and medical applications, we can ask how nature achieves hyperthermal stability in functional enzymes. By looking at the differences between the proteins of organisms that live in typical conditions and those that live in extreme temperatures, or thermophiles, we can begin to understand how it may be possible to engineer such stability. There is a clear pattern that arises when looking at thermophilic proteins’ in which polar residues are absent and charged residues dominate the surface (Mamonova et al. 2013). The thermophiles have scant use of polar residues throughout their structures, with sequences being mostly defined by hydrophobic interiors and charged surfaces. While this hydrophobic residue usage is similar to typical proteins, the exteriors are decorated with binary patterned charges to maximize hydrogen bonds and solubility (Fukuchi & Nishikawa 2001; Sadeghi et al. 2006; Dominy et al. 2004). These hydrogen-bonding networks have been shown to induce helical rigidity (Mamonova et al. 2013), stabilizing the tertiary structures thermally through rigidification of the secondary structure. It is the goal of this chapter to show how in simplification of the patterning of a maquette protein, as described in Chapter 3, these same principles of thermostability can be applied to a functional and dynamic protein.
4.1.3 The search for blood substitutes

Perhaps the most clinically relevant target of our former protein design work is to create a blood substitute. Globally, 85 million units of blood are given each year (Carson et al. 2012) and there is a demand for far more. Hemorrhage is the worldwide leading cause of maternal (Khan et al. 2006) and preventable battlefield death (Alam et al. 2005). Many of these deaths would be preventable if blood transfusions were readily available. Due to viral blood infections, short shelf life, the need for refrigeration of human blood, transfusions are often not available in the developing world or at war (Chen et al. 2009). A synthetic substitute stable for an extended period of time without refrigeration is necessary to meet high demand for transfusions.

Humans have been searching for blood substitutes since the 1600s, trying such extreme measures as beer and milk (Oberman 1969). In the 1960s, synthetic substitutes such as perfluorochemicals were used with limited success due to retention in the reticuloendothelial system and lowering of platelet counts (Chang 2000). Hemoglobin-based artificial bloods have raised considerable safety concerns and are not approved for use in the US (Natanson et al. 2008; Olson et al. 2004; Sloan et al. 1999). The utility of these substitutes is limited by blood bank supplies and natural protein characteristics such as the lag time to oxygenation and metabolism of free hemoglobin causing iron overload. There is a great need for a synthetic hemoglobin substitute that can oxygenate immediately, keeps its heme attached and does not require blood donations.

4.2 Rebuilding a functional protein without assumptions confers stability

We show here for the first time that a non-natural protein can achieve hyperthermal stability while performing oxidoreductase function at rates akin to natural proteins. We present a minimalist four-α-helix bundle protein which is thermally stable
past 95°C in both its secondary structure and retention of cofactor (weeks), yet displays rapid dynamics, including millisecond rates of heme cofactor binding and heme-ligand exchange during millisecond oxygen binding kinetics associated with significant protein conformational change. This protein’s characteristics of being thermally and temporally stable while supporting function requiring rapid movement allow it to maintain the functional diversity of previous protein designs (Farid et al. 2013) while being an improved candidate for practical application. We demonstrate achieving simultaneous high stability and dynamics is not the product of sophisticated engineering, but a consequence of basic principles of minimal design.

4.2.1 Design

This protein was designed, as described in detail in Chapter 3, based on a basic four-α-helix bundle sequence from Regan and Degrado in 1987 with LEK binary patterning, such that the hydrophobic regions are interior to the bundle and charged residues solvent exposed (DeGrado et al. 1987) (Figure 4.1). Similar sequences have been confirmed to adopt this conformation through crystallography (Huang et al. 2003) and NMR (Skalicky et al. 1999). As in former designs (Huang et al. 2004; R. Koder et al. 2009; Farid et al. 2013), histidines were added at the 6 position of each helix for bis-His hexacoordinate ligation of heme. To stabilize the holo form over the apo form, leucines that become solvent exposed upon heme binding at the 9 positions and two of the 23 positions of each helix were changed to alanine and glutamate, respectively. Phenylalanines were added at the 13 and 20 positions to add complementary hydrophobic core bulk to the heme bound form. Loops were taken from a previously published sequence (Farid et al. 2013) and the second loop modified to contain tryptophan as a spectroscopic tag for concentration determination.
In order to bind heme in the bis-His hexacoordinate manner seen spectroscopically in Figure 4.4, the helices of the apo form must rotate to bring histadines into position (Figure 4.1A, Figure 3.2). Similarly to former oxygen-binding sequences (R. Koder et al. 2009; Farid et al. 2013), the glutamates at the 7 position that are buried on rotation were maintained, as they have been shown to be necessary for oxygen binding (Figure 3.2).

Figure 4.1: Depiction of synthetic protein residues as hydrophobic (purple), hydophillic (blue), heme-binding (green), and nonreactive residues and loops (black) and hemes (brown) (A) Left: Cartoon of structure with (holo) and without (apo) heme, showing rotation of helices upon heme binding. Right: Sequence of single helix in 3.6 residue/turn register with a possible hydrogen-bonding pattern (green dashes). (B) Sequence of single chain protein reads left to right.
4.2.2 Stability measures

The most unexpected property of this protein was its extreme thermal stability. Circular dichroism temperature melts from 5°C to 95°C show helical fraying common with such simple tertiary structures, but no clear melt, a clear difference from cooperatively melting former designs (Bender et al. 2007; Farid et al. 2013) and hemoglobin (Ajloo et al. 2002) (Figure 4.2). We had previously seen a trend in our designs of lower melting temperatures conferring better heme binding, especially when melting temperatures were raised upon heme binding (Farid et al. 2013; Solomon 2013), leading us to conclude that the heme can enter a dynamic protein more easily and then rigidify the structure. However, the rapid binding of heme to this protein with such a high
affinity indicates that intrahelical rigidity can exist with interhelical dynamism. This quality is most likely due to salt bridges up the outsides of the helices that result from its consistent binary patterning of lysines and glutamates (Figure 4.1B).

This protein binds 2 hemes with a binding constant of less than 170 pM (Figure 3.4). High affinity results in its ability to bind heme stably. Over months at 37°C, the protein undergoes an exponential loss of 40% of its heme with a time constant of 1.03 weeks (Figure 4.3B), presumably aggregating or precipitating as we do not see any free heme spectra. After one month, the bound heme Soret maximum has little change, but protein aggregation is seen as a sloped baseline towards shorter wavelengths (Figure 4.3A). This heme loss profile is remarkably different than that of natural hemoglobin and derivatives, which lose most of their heme in hours (Figure 4.3B).

Figure 4.3: Stability over time. (C) UV-visible spectra showing heme loss over time at 0 (black), 2 (blue), 6 (green), 29 (purple) and 117 (red) days of 10 µM protein in 20 mM TRIS, 20 mM NaCl at 37°C. (D) Corresponding heme bound over time as determined by Abs$_{412}$ (black) compared to serum free hemoglobin (red) and PEG-hemoglobin MP4 (blue) (Vandegriff et al. 2006).

4.2.3 Gas-binding measures

UV-Visible spectra of the ferric, ferrous, oxyferrous and carbomonoxyferrous states were observed stably at -15°C (Figure 4.4) to be similar to other hexacoordinate globins (Trent et al. 2001) just as this comparison has been shown for former oxygen-
binding protein designs (R. Koder et al. 2009; Farid et al. 2013). This indicates a similarity or even more water-exclusive core, as has been previously shown by NMR (Anderson et al. 2008; R. Koder et al. 2009; Farid et al. 2013). Additionally, this protein shows greater hydrophobicity in its core through the observation of the C-O bond stretch when carbon monoxide is bound to the heme by Fourier Transform infrared spectroscopy (FTIR) in comparison to both the transiently oxygen-binding tetramer (Huang et al. 2004) and stably oxygen-binding single chain (Farid et al. 2013) maquettes (Figure 4.5). However, unlike former designs, which showed a strong preference for oxygen over carbon monoxide, this protein shows no preference (Figure 4.4B). We suggest that the shift in gas-binding preference is due to its 90mV higher midpoint potential at -200 mV (Figure 3.5). This finding is consistent with the observation of natural globins with higher potentials having CO-binding preference. We have previously shown higher potentials to be associated with more positive charges on the heme-binding helices (Farid et al. 2013; Solomon 2013). The

![Figure 4.4: UV-Visible spectra at -15°C at 5 µM protein with 2 hemes bound in 25 mM sodium phosphate pH8, 30% ethylene glycol. (A) Spectra in the oxidized (green), reduced (blue), carbomonoxyferrous (black) and oxyferrous (red) states. (B) Steady-state spectra when exposed to 50/50 mix of oxygen and carbonmonoxide (black) at 15°C and best fit using saturated spectra of 45% carbomonoxyferrous and 55% oxyferrous (red).]
charge of this sequence is +2 compared to -14 for oxygen-preferring designs.

Although theoretically a higher preference for O₂ over CO would be desirable, this trait may be necessary for many repeated oxygen transport turnovers. The higher potential of these designs, closer to that of natural globins, also confers a slower autooxidation rate. This protein binds oxygen more stably with an autooxidation time constant of 22.4 seconds at 15°C as observed by the decay of the Q-band at 574 nm after stopped-flow mixing with oxygen-saturated buffer (Figure 4.6B). Formation of the oxyferrous state was observed to be 157 +/- 2 ms at 15°C through decay of the reduced heme Q-band at 558 nm, along with growth of the oxyferrous heme peak at 574 nm (Figure 4.6A). The formation and decay of the oxyferrous state does not go through any spectrally observable intermediates (Figure 4.6A). Well after the decay fully plateaus at 90s, there is still a small peak at 574 nm, suggesting that there may be a small population that binds oxygen for a prolonged time. Although this population is too small to observe clear kinetics, this result suggests there is state in which these proteins can bind oxygen for an extended period of time closer to that of hemoglobin or myoglobin.

![Figure 4.5: FTIR of the thermostable protein presented here in comparison to previously published single chain and tetramer oxygen-binding maquettes. Maquettes were degassed with argon, reduced with sodium dithionite, bubbled with carbon monoxide for saturation and plated between KCl wells under deoxygenated conditions. FTIR spectra were averaged over 100 scans and each had the baseline of the respective maquette in the same conditions without carbon monoxide subtracted out to highlight the band of CO bound to ferrous heme.](image-url)
4.2.4 Ligand Exchange Dynamics

We see a moderate temperature dependence of oxygen binding. Over a 25°C, we observe approximately a 5-fold change in the oxygen off rate, in the range of natural
globins (Uzan et al. 2004). Arrhenius analysis of the oxygen on and off rates (Figure 4.6C) show near identical activation energies (E<sub>a</sub>) for oxygen coming on and off of 15.57 kcal/mol and 14.33 kcal/mol, respectively (Table 4.1). The E<sub>a</sub> for oxygen dissociation is similar to that of myoglobin at 20.3 kcal/mol and slightly less than the relatively temperature-insensitive neuroglobin at 24.5 kcal/mol. Meanwhile, the E<sub>a</sub> for oxygen binding is slightly higher than those of neuroglobin and myoglobin, which range from 7.5 to 8.5 kcal/mol. The similar on and off energies of activation likely account for some the difference between our oxygen binding lifetime and those of natural globins. However, the finding that the energies of activation are in the same ballpark as these natural oxygen carriers implies that the dynamicism has been maintained, allowing for ease of oxygen binding and release. It is worth noting that these values of energy of activation were achieved with a significantly depressed E<sub>m</sub> of about -200 mV compared to neuroglobin at -115 mV (Halder et al. 2006) and myoglobin at +70 mV (Isogai & Ishida 2009).

CO flash photolysis experiments (Figure 4.7) were performed at 20%, 50% and 100% CO to observe histidine on and off rates. Spectra were analyzed by singular value decomposition (SVD) and single wavelength analysis to obtain histidine on and off and CO on rates as described by Hargrove (Dewilde et al. 2001). In comparison with former

<table>
<thead>
<tr>
<th>Protein</th>
<th>k&lt;sub&gt;His on&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>k&lt;sub&gt;His off&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>k&lt;sub&gt;CO on&lt;/sub&gt; (μM&lt;sup&gt;-1&lt;/sup&gt;s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>E&lt;sub&gt;A,O&lt;sub&gt;2&lt;/sub&gt; on&lt;/sub&gt; (kcal/mol)</th>
<th>E&lt;sub&gt;A,O&lt;sub&gt;2&lt;/sub&gt; off&lt;/sub&gt; (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermostable</td>
<td>240</td>
<td>680</td>
<td>1.9</td>
<td>15.57</td>
<td>14.33</td>
</tr>
<tr>
<td>Dimer</td>
<td>310</td>
<td>17</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Chain*</td>
<td>210</td>
<td>50</td>
<td>4.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroglobin</td>
<td>2000</td>
<td>4.5</td>
<td>65</td>
<td>8-8.5</td>
<td>24.5</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>n/a</td>
<td>n/a</td>
<td>0.5</td>
<td>7.5</td>
<td>20.3</td>
</tr>
</tbody>
</table>

Table 4.1: Ligand exchange dynamics and energies of activation of the thermostable protein presented here in comparison to previously published dimer (R. Koder et al. 2009), single chain (Farid et al. 2013), neuroglobin and myoglobin (Kiger et al. 2011; Uzan et al. 2004; Dewilde et al. 2001).
Figure 4.7: CO flash photolysis on protein at 20%, 50% and 100% CO at 1 atm in 20 mM Tris, 20 mM NaCl pH 8. Delta absorbance spectra were plotted out from red (1 µs) to purple (100 ms) delays after flash. Fractions of 5-coordinate (red), 6-coordinate (green) and CO-bound heme (blue) were found via SVD analysis for 3 major components (inset). Absorbance value of peaks at 424.5 and 432 nm were used to determine rates of histidine and CO ligation to heme b.

designed oxygen-binding four helix bundles and human neuroglobin (Table 4.1) the thermostable protein presented here has similar CO and histidine on rates and a somewhat faster histidine off rate. The similarity of the histidine ligand exchange rates demonstrated a preservation of interhelical mobility despite an increased thermal
stability in both the apo and holo forms. While neuroglobin also has a high melting temperature in the holo form, it has a significantly lower apo melting temperature, indicating its thermal stability is induced through hexacoordination of heme. In contrast, the protein presented here has apo thermal stability from its protein structure alone as observed by CD thermal melting (Figure 4.2), which would be preserved in both the apo and holo states. The rigidity of the helices and exhaustive hydrogen bonding up each helix may be the reason for a relatively small difference between histidine on and off rates in comparison to other proteins, as there are likely not interhelical hydrogen bonds changing between helical rotation states. This ligand exchange data further supports that stability and helical rigidity promote rapid and low activation energy interhelical dynamics.

4.2.5 Structure by NMR

$^1$H NMR spectra taken of the apo and holo heme b forms of the protein both show structure with some gain of structure in the holo form relative to the apo form, most notably in the appearance of some peaks in the aromatic 6 to 10 ppm region and refinement of the lysine triplet around 2.6 ppm (Figure 4.8A,B). Structuring is less apparent than in previous designed four helix bundles, in which a stark refinement of peaks in all regions occurs with the addition of heme b (Figure 4.8E,F). Although it is more difficult to determine extent of structuring in this protein due to high sequence repetition, we qualitatively observe that this protein is more structured in the apo and similarly structured in the holo forms compared to previous designs. This structuring behavior more similarly resembles a natively structured apo protein that gains minimal structure upon heme addition, such as myoglobin (Figure 4.8C,D).
Figure 4.8: Changes in structure upon heme binding. $^1$H NMR at 22°C of 400 µM protein in 80% D$_2$O/20% H$_2$O, pD6.6 PBS with (B) and without (A) heme is compared to whale myoglobin (Ribeiro et al. 2003) (C, D) and previous proteins (Huang et al. 2004) (E, F) with and without heme, respectively.
4.3 Conclusions:

4.3.1 Simplicity confers stability, not diminishes it

The French engineer Antoine de Saint-Exupéry once wrote, “A designer knows he has achieved perfection not when there is nothing left to add, but when there is nothing left to take away” (Saint-Exupery 1949). It is an open question in protein design: just how much sequence detail is needed to accomplish any desired structure or function (Romero et al. 1999; Plaxco et al. 1998). A cursory examination of natural protein structures can easily give the impression that complicated assemblies of many hundreds of amino acids arranged with angstrom-level precision is essential for appropriate function; the common result that protein functions can be severely disrupted with just a few amino acid changes further supports that view. Yet the forces of natural selection that have shaped natural protein design are biased to accumulate rather than reduce complexity and have no forethought to apply simplifying engineering principles.

The design presented here reflects intentional vigilance to remove unnecessary complexity while still supporting sophisticated function that is associated with naturally evolved proteins. It shows that desirable engineering characteristics, including stability and robustness to modification, follow directly from the physical properties of a few natural amino acids arranged along a binary-patterned scaffold, and that spare changes to this elementary frame confer surprisingly sophisticated function. Patterns of residues are assembled to achieve gross characteristics rather than structural precision, only requiring single amino acid representatives for functional groups. In this case, only 8 functional amino acids were necessary to achieve both stability and oxygen-binding function as described in Chapter 3. This design is able to function below the lower limit of 10 suggested to be necessary and sufficient for folding into globular proteins having
both rigid side chain packing and biological function (Romero et al. 1999), because residues were treated as a gross matrix more akin to polymer catalyst design (Wulff 2002; Madhavan et al. 2008) rather than attempting exactness.

Our creation of the first thermally hyperstable designed protein that displays native-like oxidoreductase function is a consequence of applying the principle of minimal sequence complexity to functionalize a 4-alpha-helix framework based on two types of elementary binary patterning: a binary hydrophobic/hydrophilic patterning of a hydrophobic leucine-dominated interior with a charged exterior dominated by lysines and glutamates, and a binary exterior charge patterning such that each negative glutamate is paired with a positive lysine on the same helix. This pattern simplicity is in keeping with the maquette method, which seeks to reduce protein complexity to create a platform more robust to iterative redesign by minimizing interdependencies between domains. Similar maquettes with more traditional stabilities have already demonstrated a broad range of heme ligations and midpoint potentials (Shifman et al. 2000; Gibney et al. 2001; Gibney et al. 1998; Reedy et al. 2008; Solomon 2013).

A direct result of this choice of even binary patterning is an extensive network of hydrogen bonds up the outsides of the helices. This patterning is analogous to that found in thermophilic proteins (Mamonova et al. 2013), discussed in section 4.1.2, which maximizes hydrogen bonds and solubility (Fukuchi & Nishikawa 2001; Sadeghi et al. 2006; Dominy et al. 2004). These hydrogen bond networks seem to induce helical rigidity in this design, supported by CD results in Figure 4.2, similarly to thermophilic proteins (Mamonova et al. 2013). Compared to maquette sequences with greater sequence diversity and extensive use of polar, non-charged residues, we see increased thermal tolerance while maintaining oxygen-binding kinetics parameters (R. Koder et al. 2009; Farid et al. 2013) and improving oxygen binding lifetime by 50%.
4.3.2 Sliding rocks

The extreme thermal stability and rigidity of secondary structure combined with the finding that this protein is capable of nanosecond ligand exchange dynamics by CO flash photolysis and oxygen binding results (Figures 4.6 and 4.7) raise the question as to how such properties can be found in the same protein. Rigidity of secondary structure, even when found in functional thermophilic proteins, is generally equated to higher thermodynamic barriers for conformational changes (Razvi & Scholtz 2006; Mamonova et al. 2013), which may not pose a problem if such functions need only perform at high temperatures. However, this work suggests that even mesophilic-functioning proteins, those having optimal functioning ability at ambient temperatures, can exhibit hyperthermal stability without impeding their ability to undergo rapid conformational changes. So why did so many former protein designers see an antithetical relationship between thermal stability and dynamic ability? The answer suggested by this work is that the location of the salt bridges formed by the charged residues across the protein surface is vitally important.

Here we demonstrate that designing intrahelical charge pairing can assist both hyperthermal stability and rapid dynamics. The binary patterning which allows for exhaustive hydrogen bonding up each helix (Figure 4.1) was a natural consequence of the starting sequence outlined by DeGrado and Regan in 1988 (Regan & DeGrado 1988) despite their initial design not intending to promote helical independence nor informed by our present understanding of natural thermophilic proteins. The finding that simple sequences tend to be thermostable is not new, either. In 1998, Stroud conferred that other four-helix bundles made from reduced sets of amino acids can be extremely thermostable (Schafmeister et al. 1997). However, the use of a minimal library does not necessarily confer stability. Other designed proteins developed using supercharging
computational algorithms are not as thermostable and do not retain heme as well (French & R. L. Koder 2013). These algorithms eliminate polar residues, but do not pay attention to the location of the charges, such that not every positive charge has a negative charge with which to form a salt bridge within the same helix. As a result, many interhelical salt bridges can be formed and not all charged residues have partners, which leads to a less thermostable structure and the rigidification of interhelical movement hence preventing effective heme binding.

It is only in the case of exhaustive, not just extensive, hydrogen bonding by the alternation of negatively and positively charged residues up the outside of each helix that each potential hydrogen bond is formed within each helix. This pattern will enhance stability along but not between helices, allowing for domain segregation. In this case, no hydrogen bonds can be formed between helices, so they are less ‘sticky’ to one another than in a less hydrogen-bonded case. As a result of exhaustive intrahelical hydrogen bonding, domains, here individual helices, can move freely relative to one another. The pattern that is allowing for hyperthermal stability in this case is also promoting the clean rotation of helices necessary for heme binding and oxygen binding to hemes. These helices act as an assembly of ‘sliding rocks’, rigid and stable but with no possibility for interaction between each other, promoting both stability and rich dynamic behavior.

4.3.3 Suggestions for computation and directed evolution

Both computational design and directed evolution methods can include these principles of thermophilic design and helical segregation. Supercharging design algorithms intended to enhance solubility could discriminate to favor intra- rather than inter-helical hydrogen bonds. While directed evolution and computational work that begin with thermophilic proteins typically lose thermal stability, constraining searches to retain
exterior charge patterns may save time and improve stability. Although directed evolution techniques, such as those used for the Kemp eliminase (Khersonsky et al. 2010), may initially favor more dynamic sequences to allow nascent functions to evolve faster, these techniques could be constrained to maintain stability-enhancing exterior binary charge patterning while allowing interior and inter-helical motion.

4.3.4 Potential as a platform for blood substitutes and other applications

Here we demonstrate that a designed heme-binding protein is capable of binding oxygen for tens of seconds, thermally stable past 95°C, binds heme with a $K_D$ in the picomolar range, and weighs only 14.7 kDa. Unlike modified free hemoglobins, this protein is unlikely to cause heme toxicity due to its high heme-binding affinity and minimal heme loss over many days at 37°C. The heme retention combined with the great thermal stability not only give it the potential to circulate effectively throughout the body, but also be stored and transported without refrigeration. This protein’s thermal stability beyond inhabitable conditions would also allow it to be easily be resterilized.

This designed globin is able to oxygenate rapidly, with on rates in the millisecond regime. Hemoglobin in natural blood transfusions can take up to 24 hours to fully oxygenate due to 2,3,-phosphoglycerate depletion (Klocke 1972; McConn & Derrick 1972; Hess 2010). The lag time is an insurmountable problem when loss of blood and tissue oxygenation is extreme. Cases of severe hemorrhage are common in trauma and childbirth. Synthetic globins would be able to immediately carry oxygen to ischemic tissues. The small size of this artificial globin also may also allow it to penetrate poorly vascularized regions of the body, aiding in photodynamic cancer therapies.

Because this globin is built from a dynamic platform, it has great potential for modification and tweaking. We have already demonstrated a broad range of heme
ligations and midpoint potentials within analogous proteins (Shifman et al. 2000; Gibney et al. 2001; Gibney et al. 1998; Reedy et al. 2008; Solomon 2013; Farid et al. 2013). With increased midpoint potential from its current -200 mV, we should be able to lengthen the oxyferrous lifetime and shift its preference between CO and O₂ from its current lack of preference. Lengthening the oxyferrous lifetime would be necessary for it function as a substitute for hemoglobin although its current form may prove useful as a neuroglobin analogue for mediating oxidative damage in ischemic injury (Ferrari et al. 2004). Shifting the KＤ and binding preference between CO and O₂ may make it possible to create artificial bloods for extreme situations, such as high altitude, SCUBA, or CO-rich scenarios in which researchers and military personnel may need to live.

Perhaps the greatest potential of these proteins stems from their ability to be expressed in E. coli. Although we have not yet optimized the expression yields, we currently are able to purify grams of protein per 2L culture. Compared to hemoglobin, they have twice the oxygen-binding capacity by weight. These proteins can also be dissolved from a lyophilized powder into various buffer conditions with complete solvation in minutes to hours and heme binding in seconds. Lyophilized powders are stable for years in refrigeration and weeks to months at room temperature. These powders create a greater opportunity for storage and transport of transfusion materials than ever before. These proteins are ideal candidates from which to engineer blood substitutes for developing world and on the battlefield.

4.4 References


Gibney, B.R. et al., 2001. Hydrophobic modulation of heme properties in heme protein


Sharp, R.E. et al., 1998. Design, synthesis, and characterization of a photoactivatable flavocytochrome .... *Proceedings of the ....*


89
Chapter 5: Development of a ROS Laboratory

5.1 Introduction

5.1.1 Maquettes as model systems

As simplified, highly modifiable systems, maquettes avoid many of the difficulties inherent to the study of natural systems. They allow for single variable modifications and control of degeneracy between cofactors in their designs. Additionally, they can be concentrated to hundreds of micromolar concentrations without aggregation issues and can be expressed in high quantities. As such, they allow us to ask fundamental questions about protein cofactor chemistry that are either too difficult or too convoluted for study in the natural systems.

Of broad interest in the field of bioenergetics, yet woefully poorly understood, is reactive oxygen species (ROS) production from electron transfer proteins. As discussed in Chapter 1, it is still not known where in the mitochondrial electron transport chain ROS is produced or which ROS species is produced. The latter is fundamental to the context of the former, as not all ROS production is deleterious. Hydrogen peroxide is a poor oxidant and used as a cell-signaling molecule while superoxide (SO) can lead to the formation of highly reactive species (Figure 1.1) (Beckman & Koppenol 1996). However, differentiating between peroxide and superoxide production in vivo or in vitro from natural systems let alone distinguishing from where it was produced is difficult due to issues with probe delivery, low concentrations, and kinetic resolution due to SO dismutation to peroxide and evolution into other radicals.

When ROS species have been successfully distinguished, as by Judy Hirst in the finding that bovine complex I produces SO and E.coli complex I produces primarily
peroxide, there is still not enough fundamental understanding of basic protein cofactor-oxygen chemistry to do more than speculation as to the reason (Esterházy et al. 2008). Even for much more simple proteins with much higher ROS yields, such as NADPH oxidases (Nox), it is still not known why some produce SO and others hydrogen peroxide as discussed in Section 1.2.2. In fact, it is not known how any two homologous proteins can produce different ROS species despite ROS production being implicated in numerous diseases, particularly from mitochondria (Lin & Beal 2006; P. H. Reddy & T. P. Reddy 2011).

Maquette study is ideal for elucidating fundamental mechanisms of ROS production from protein systems. By asking a series of concise questions, we can provide a library of experimental data from which to put natural ROS production in context. We can independently modify cofactors, reduction potentials, inter-cofactor distances, protein charge, core dryness and oxyferrous state formation among other properties. In order to fully describe the engineering principles necessary for the specific production of each ROS species, maquettes would need to be engineered from first principles as done in Chapter 3 for an oxygen-binding maquette. These maquettes could then be iteratively modified and tested for ROS production properties. However, this work was begun in parallel to that described in Chapters 3 and 4, blind to the engineering properties elucidated in their designs. Even if these minimal maquettes were available, we did not have a firm enough grasp on peroxide-specific production chemistry prior to this work to systematically engineer a constitutive peroxide producer from a SO producer. The reengineering of each ROS producer would be the next step in fully understanding the functional requirements.

In this chapter, a series of maquettes related to those previously published (Huang et al. 2004; Koder et al. 2009; Farid et al. 2013) were used to develop the first
methodology for observing ROS production kinetically with species-specific resolution for SO, hydrogen peroxide and hydroxyl radicals. Here will also be presented as proof-of-principle that control over which ROS species is produced can be achieved in similar maquettes. This chapter describes for these maquettes the biochemical properties leading to their differing ROS productions and protocols for distinguishing multiple mechanisms of production.

Figure 5.1: Topologies of maquettes for ROS study. (A) Topology A is a homotetramer linked into an effective homodimer by disulfide bonds that can orient in syn (not shown) or anti (shown) configurations (Huang et al. 2004). Topology B is a homodimer linked by a disulfide into a ‘candelabra’ effective monomer structure (Koder et al. 2009). Topology C is a single-chain (Farid et al. 2013). (B) shows rotation of glutamates (red triangles) into the hydrophobic core (purple) upon each heme (brown) binding, which occurs for each topology.
The proteins chosen for this work were those with oxygen-binding capabilities previously described, structural data available with a range of topologies (Figure 5.1), allowing for differing water-accessibility to the core (Anderson et al. 2008). It should be noted that amphiphilic maquettes are also available for study and will be necessary for asking questions pertaining to directionality of ROS across membranes (Discher et al. 2005), but for the purposes of designing the initial system soluble maquettes were preferable to avoid potential issues with probe delivery and production yields. All of the proteins selected for development can also be produced in high enough quantities to avoid issues with signal-to-noise ratios, allowing for use of ROS probes and techniques not available for natural systems due to high quantities of protein needed. However, even with these improved characteristics over natural systems, detecting ROS specifically and kinetically is still challenging.

5.1.2 ROS kinetic resolution

Determining the quantities and kinetics of reactive oxygen species (ROS) production under various oxygen pressures in vitro, let alone in vivo, is not trivial. While horseradish peroxidase (HRP)-based assays such as the conversion of Amplex Red to resorufin are reliable and specific for detecting hydrogen peroxide (Murphy 2009; Drose & Brandt 2008; Esterházy et al. 2008; Bell et al. 2007) both in vitro and in vivo, there is no equivalent system for the detection of oxygen radicals. Detecting singly reduced oxygen, superoxide (SO), is of particular importance because it is the primary product that leads to more damaging radicals in vivo (Beckman & Koppenol 1996).

Further, it is desirable to have a probe for SO detection that is specific to SO and allows kinetic observation. Specificity requires the probe to not be directly reduced by cofactors and proteins. Kinetic observation requires that the reaction be rapid and
detectable in real time. Much of the work on SO probes has been conducted on the respiratory complexes because they are a prime site for ROS production related to aging and disease \textit{in vivo}. Various probes have been used to detect mitochondrial ROS production, but each has their own shortcomings. Usually, specificity and temporal resolution must be sacrificed to improve the other.

Most reduced probes will directly re-oxidize in the presence of oxygen, but cytochrome \textit{c} does not due to its conformational stabilization of the reduced state (Berghuis & Brayer 1992). Cytochrome \textit{c} is not ideal however; it will be directly reduced by complex III and other low potential protein cofactors. It can be acetylated to minimize this reaction (Grivennikova & Vinogradov 2006; Prakash & Mazumdar 2009; Azzi et al. 1975; Pepelina et al. 2009), but then gains non-native interactions with complex I allowing for direct reduction (Esterházy et al. 2008; Kussmaul & Hirst 2006). All of these species are readily reduced directly by hemes in maquettes, however, due to the proteins’ small sizes and heme partial solvent exposure (Bryan Fry, \textit{unpublished work}). Judy Hirst determined that reduction of dihydroethidium (DHE) is the best option when acetylated cytochrome \textit{c} can be directly reduced by complex I (Esterházy et al. 2008). However, DHE is significantly less reliable than cytochrome \textit{c} and there are other options available for the study of maquettes not possible at the concentrations used for natural respiratory complex study.

For years methyl-cypridina luciferin analog (MCLA) has been considered the most reliable detection method (Nakano 1998; Anon 1998; Muller et al. 2004), as it is more efficient at biological pHSs than other probes available (Oosthuizen et al. 1997). The chemiluminescent reaction time is not ideal for kinetic resolution, however, as its reaction time is much longer than most oxygen reduction events. Many have avoided the selection of SO-specific probes completely by converting SO to hydrogen peroxide with
excess SOD and detecting hydrogen peroxide with Amplex Red (Drose & Brandt 2008; Grivennikova et al. 2008; Lambert & Brand 2004; St-Pierre et al. 2002). This last method avoids issues with SO-specificity, but only is able to measure total ROS production, not SO specifically.

One major issue with the common SO detection methods is a lack of specificity for SO over other ROS (Afanasev 2008). Without specificity, it is difficult to quantify signals using standard assays. While it has been suggested that MCLA may react with singlet oxygen as well as SO, MCLA is specific for SO over other radicals such as hydroxyl radicals which may be produced in downstream reactions (Nakano 1998). Additionally, MCLA has the advantage of being chemiluminescent, so there is not spectral degeneracy caused by using large concentrations of cofactor-containing proteins. While chemiluminescence experiments are difficult to perform on natural proteins due to low signal-to-noise ratios inherent to photon-counting experiments, maquette study avoids this issue by allowing for high concentrations of protein without high cost or aggregation.

There are other chemiluminescent probes on the market, but MCLA is the best candidate for SO detection. Probes such as Lucigenin or Luminol do not work well at biologically relevant pHs (Oosthuizen et al. 1997), have much lower solubility and show high rates of autocatalysis (Esterházy et al. 2008). None of these probes are superior to MCLA in specificity either. There are some reagents that may be more specific for SO, such as commercially available tetrazolium salts like WST-1, but these reagents are easily directly reduced by complex I (Esterházy et al. 2008) and would thus be reduced by maquettes. Electron paramagnetic resonance (EPR) probes that form spin-adducts through nucleophillic reactions are also highly specific, but their reactions are extremely slow (~10 M⁻¹s⁻¹) and the products short-lived, so no kinetic information can be attained
(Afanasev 2008). So, while MCLA can only provide qualitative comparisons, its avoidance of spectral degeneracy, side reactions or direct reduction make it the best commercially-available probe for use in SO detection from maquettes.

While there is an abundance of literature attempting to detect SO radicals, there is a dearth of hydroxyl radical detection. The majority of work done has either looked at oxidation of lipids to survey damage, although this is not specific to hydroxyl radicals, or used trapping reagents to quantify hydroxyls (Tang et al. 2005; Manevich et al. 1997). Neither of these methods allows for kinetic information to be obtained and there are no quantitative hydroxyl radical studies on complex natural proteins. Although the bioenergetic ROS field has so far ignored hydroxyl radicals experimentally, hydroxyl radicals have been studied in terms of their ability to degrade pollution. In particular, these radicals are potent oxidizers of azo dyes used in the garment industry (Guivarch et al. 2003; Konstantinou & Albanis 2004). The most well-studied of this family of dyes is Acid Orange 7 (AO7), which is degraded readily and rapidly by hydroxyl radicals and highly colored before degradation (Wu et al. 2012; Stylidi 2004; Hammami et al. 2008). In this work, the observation of its decolorization through Fenton chemistry will be observed in real time to quantify hydroxyl radical production.

5.2 Oxygen reactivity in a series of maquettes

5.2.1 Design and predictions

Maquettes were designed with a range of properties, in part due to their topologies. Oxygen-binding maquettes in both the candelabra (B) and single-chain (C) topologies (Figure 5.1) were used as published with identical helical sequences (Table 5.1). In order to abolish oxygen binding, both of these sequences were then modified to
remove the glutamates that become rotated in upon heme binding at the 7, 14 and 21 positions and have been shown to be necessary for stable oxygen binding in previous work and Chapter 3. These glutamates have been suggested to induce helical strain necessary to lower the energetic barrier for histidine removal as discussed in previous chapters. All three glutamates per helix were mutated to alanines to relieve this helical strain (Figure 5.1), forming sequences B.1 and C.1 for the candelabra and single-chain topologies, respectively. The maquette used with topology A has previously been shown to allow water access into the core by hydrogen exchange NMR experiments in which no core peaks are observable after 30 minutes (Anderson et al. 2008). This exchange time contrasts to B and C, which have not exchanged most of their core hydrogens in 2.5 hours. This protein, A, has poor interfacial helical packing caused by the replacement of interfacial glutamines with charged residues as well as increased inter-helical dynamics due to its homotetramer topology and ability to flip conformations.

<table>
<thead>
<tr>
<th>Topology</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CGGG EIWKQHEEALKKFEEALKQFEELKKL</td>
</tr>
<tr>
<td>B</td>
<td>EIWKQHEDALQKFEEALNQFEDLKQL GGSGCGSGG EIWKQHEDALQKFEEALNQFEDLKQL</td>
</tr>
<tr>
<td>B.1</td>
<td>EIWKQHADALQKFEEALNQFADLKQL GGSGCGSGG EIWKQHADALQKFEEALNQFADLKQL</td>
</tr>
<tr>
<td>C</td>
<td>EIWKQHEDALQKFEEALNQFEDLKQL GGSGCGSGG EIWKQHEDALQKFEEALNQFEDLKQL GGSGCGSGG EIWKQHEDALQKFEEALNQFEDLKQL</td>
</tr>
<tr>
<td>C.1</td>
<td>EIWKQHADALQKFEEALNQFADLKQL GGSGCGSGG EIWKQHADALQKFEEALNQFADLKQL GGSGCGSGG EIWKQHADALQKFEEALNQFADLKQL</td>
</tr>
</tbody>
</table>

Table 5.1: Sequences used in the development of a ROS laboratory. Sequences are numbered by topology corresponding to Figure 5.1 and then number of sequence designed in that topology. Oxygen-binding sequences are presented as ‘parent’ sequences in all black and all other sequences are referenced to their helical sequences with mutations in red.
This set of maquettes allows for a variety of factors to be tested as they can effect ROS production. Topological influences can be seen between A, B, and C, particularly between B and C as they share the same helical sequences but have differing levels of heme burial. Core dryness can be tested of A against B and C. Glutamate removal can be tested as C.1 against C and B.1 against B. The contribution of the second heme-binding site can be tested through one-heme versions of maquettes.

Figure 5.2: SO production scheme from one-heme maquettes. Maquettes are depicted top-down helices (grey) with hydrophobic cores (purple) and single hemes (square: brown oxidized, red reduced) bound. Possible reactions involving ET between hemes and dioxygen (red circles) are shown as they may produce superoxide. Rates leading to these reactions for histidine (green) removal and ligation from hemes, oxygen binding and oxidation of heme are shown as a pseudo-thermodynamic box.
Figure 5.3: ROS production possibilities from two-heme maquettes. Top shows predictions for low-strain maquettes where glutamates (red triangles) are mutated to alanines (black triangles) to produce SO (red circles) and hydrogen peroxide (red and grey circles) through outer-sphere ET. Bottom shows predictions for water-accessibly maquettes that bind oxygen transiently to produce SO or peroxide.
All of these factors can be assessed in terms of their contribution to ROS production at each stage in which ROS can be produced. Figure 5.2 outlines, in a pseudo-thermodynamic box, pathways from fully reduced heme to oxidized heme with the possibility of oxygen binding. ROS can be produced through outer-sphere tunneling to dioxygen if histidine removal does not occur to allow oxygen binding, as predicted for B.1 and C.1, or if oxygen cannot bind in the time the histidine is off. Either the histidine off rate or the oxygen on rates can affect this production. ROS can also be produced through inner-sphere ET to dioxygen through oxidation of the oxyferrous heme state. This production can be rapid if the oxyferrous lifetime is very short, as is predicted for water-accessible oxygen-binding maquettes, such as A.

While Figure 5.2 shows the possibilities for single electron transfer from single-heme binding maquette, but in a two-heme binding maquette, the additional possibility of concerted oxidation of the two hemes may allow for hydrogen peroxide production directly through mechanisms described in Figure 5.3. In a two-heme maquette that reduces dioxygen through outer-sphere ET, as predicted for strain-removed maquettes B.1 and C.1, simultaneous oxidation of the two hemes creates the possibility of rapid dismutation between the two locally-produced SO molecules. The more rapid the ferrous heme oxidation rate, the more likely this will occur. A similar mechanism of rapid hydrogen peroxide production can occur if SO is produced through inner-sphere ET, as predicted for A due to water-catalyzed oxyferrous state oxidation. In this case, the probability of this occurring is not only limited by the ferrous heme oxidation rate, but by the apparent oxygen on rate (and thus histidine off rate) and oxyferrous heme oxidation rate. Even in the case of a rapid oxyferrous oxidation rate, if the oxygen binding to the heme is slow and outer sphere ET is suppressed, SO may be produced too slowly for
rapid dismutation rates. Dismutation can either occur outside of the protein akin to the outer-sphere ET SO producers or within the protein core before SO can diffuse away. These discussed mechanisms of peroxide production would show apparent peroxide production because dismutation would occur before any probe could react with the SO. However, true direct peroxide production may also be possible for maquettes in which two hemes are in close enough proximity to allow inter-cofactor ET (less than 14Å, satisfied for all cases) on the same order of the rates of oxygen-binding and oxyferrous oxidation. In this case, inter-cofactor ET can occur when the oxyferrous state can be described as a SO bound to a ferric heme as electrons are delocalized prior to SO leaving the heme. At this time, if an ET event occurs from the still ferrous heme to the now ferric heme, the site of oxygen binding will have a SO adjacent to a ferrous heme. In the case of a water-accessible core, protons would also be present, allowing for a second ET event to SO and protonation, producing hydrogen peroxide.

5.2.2 Oxygen binding

The first step in characterizing the oxygen reactivity of each presented maquette is to test their oxygen-binding capabilities. Both B and C, previously-published oxygen-binding maquettes, show distinct spectra for each state by stopped-flow mixing of reduced anaerobic protein with oxygen-saturated buffer at 15°C as hemes progress from the reduced to oxyferrous and then oxidized states (Figure 5.4A). The hemes are also indistinguishable from one another spectrally (Farid et al. 2013). B forms the oxyferrous state in a rate of 300 mM⁻¹s⁻¹ which then decays with a rate of 0.1 s⁻¹ (Koder et al. 2009). C binds oxygen slightly faster, with a rate of 800 mM⁻¹s⁻¹, but oxidizes at the same rate.

Strain-removed maquettes B.1 and C.1 were also spectrally identical; neither had observable oxyferrous state formation upon stopped-flow mixing (Figure 5.4B). When
reduced protein was mixed with oxygenated buffer, both B.1 and C.1 transitioned from the ferrous to ferric forms with no detectable intermediates in rates of 20 s^{-1} and 9.7 s^{-1}, respectively. In the same experiment, A showed oxyferrous state formation with a half time of 50 ms (apparent rate of 43 mM^{-1}s^{-1} compared to apparent rate of 16-17 mM^{-1}s^{-1} for B and C). This state was transient and decayed in a rate of 10 s^{-1} (Figure 5.4C). Likely due to the similarity of these rates, oxyferrous state formation is never complete for A, unlike B and C, and only goes to 88% relative completion even at -15°C. All rates are summarized in Table 2 as compared to some natural oxygen-reactive proteins.

![Figure 5.4: Oxygen-binding properties of maquettes.](image)

5.2.3 Superoxide production

SO was detected from each maquette upon stopped flow mixing at 20°C in 20mM Tris, 20mM NaCl pH 8 buffer of anaerobic reduced 10μM protein and 15μM MCLA with air-saturated buffer. Chemiluminescence was detected with a photomultiplier and signals were integrated over time. Signals were then normalized to the reduced
heme band at 558 nm, as observed by UV-visible spectroscopy prior to stopped-flow injection to control for concentration of reduced heme. Signals were averaged over at least two separate days and at 20 runs. All results are shown in Figure 5.5.

Oxygen-binding proteins B and C did not show any detectable SO formation. However, even if all heme oxidation singly reduced oxygen, SO would likely be undetectable due to the production rate leading to too low an MCLA signal to be detected over the background noise. Because the observed signal is photons produced per unit time, slower production times lead to lower signal amplitudes. If this rate is slow enough, all signals will be lost in the fairly high background inherent to photon counting experiments due to background light in the mixing cell of the stopped-flow apparatus. Not all ports can be sealed effectively in a stopped-flow mixer without sacrificing speed of loading to injection. Because MCLA is not very stable, the effective concentration is changed with added delays between sample preparation and mixing. As a result, only rapid bursts of SO are detectible by this method.

Strainless maquettes B.1 and C.1 both produce large amounts of SO rapidly. B.1 produced SO with an apparent rate of 8 s\(^{-1}\) (kinetics are convolved with MCLA chemiluminescence kinetics) and C.1 with an apparent rate of 1.34 s\(^{-1}\). While these both produce appreciable amount of SO, their detectable yields are quite different. When compared to the maximum signal observed for any maquette, C.1, the dimer B.1 only produces 43%. This is likely due to the increased dismutation rate caused by higher concentrations of SO resulting from rapid SO production. If SO dismutates more quickly, the MCLA cannot compete with that reaction as well, and therefore it is not observable.

The homotetramer A, with a less dry core, produces SO at a similar rate to C.1, at 2.4 s\(^{-1}\). This was expected as the auto-oxidation rate for A is similar to the oxidation rate for C.1. However, despite similar oxidation and SO production rates, the observed
yield for A is considerably lower, at only 20% that observed for C.1. As initial SO concentrations cannot explain this yield related to the rate, it was suspected that some hemes were going oxidized without producing SO, possibly by producing hydrogen peroxide or other radicals.

5.2.4 Hydrogen peroxide production

Hydrogen peroxide was detected from each maquette upon stopped flow mixing at 20°C in 20mM Tris, 20mM NaCl pH 8 buffer of anaerobic reduced 10µM protein with air-saturated buffer containing 10µM Amplex Red and 5 U/mL horseradish peroxidase (HRP). In the presence of peroxide, colorless Amplex Red is converted to highly colored resorufin by HRP. The conversion to resorufin was monitored by its absorbance at 574nm for each protein as well as a 20µM standard peroxide solution. Signals were normalized to amount of reduced heme as determined by the reduced heme band at 558
nm, as observed by UV-visible spectroscopy prior to stopped-flow injection. Signals were averaged over at least two separate days and 20 runs with results shown in Figure 5.6.

The peroxide control clearly shows that rapid kinetics are detectible as HRP has a rapid catalytic rate (Votyakova & Reynolds 2004). Unlike the MCLA SO detection system, which only proved qualitative kinetic comparisons, this HRP system is capable of detecting kinetics significantly more rapid than those observed for any of the proteins studied or is possible with stopped-flow apparatus dead time upon mixing. All proteins were observed to produce the same amount of peroxide per reduced heme, as is expected given the rapid dismutation rate of SO to peroxide and the high concentrations of HRP used, removing peroxide from the system rapidly after formation preventing progression to other radicals in significant quantities.

Figure 5.6: Hydrogen peroxide production as detected by Amplex Red and HRP. (Left) Each trace monitors absorbance of resorufin at 574 nm over time after stopped-flow mixing of reduced 10µM 2-heme-bound maquette with 10 µM Amplex Red and 5U/mL HRP in 20mM Tris, 20mM NaCl pH8. Signals were normalized to reduced concentration as determined by the 558nm peak at time zero and the total amount of peroxide produced as determined by UV-visible spectroscopy after 5 minutes for a 10µM peroxide control (black), A (dark blue), A with one heme (light blue), B.1 (green) and C.1 (red). (Right) The same protocol was followed for two-heme containing A at 2µM (magenta), 4µM (purple), 10µM (blue) and 20µM (light green).
Both B.1 and C.1 produce peroxide in a biphasic manner indicative of lag times before peroxide can be observed. This is consistent with their rapid SO production, as it would mean that peroxide is only being produced through the dismutation of SO external to the maquette. A does not exhibit as pronounced a lag phase and appears to mostly produce peroxide as a single exponential. The rapid kinetic with no dismutation lag is indicative of direct peroxide production internal to the maquette. The small biphasic nature of the curve is likely due to the 15-20% SO generation observed in Figure 5.5, which is being produced through an alternate reaction of reduced heme with dioxygen.

To test whether A was producing peroxide internally, as opposed to producing SO more rapidly than B.1 and C.1, peroxide production was observed from a series of concentrations from 2µM to 20µM two-heme protein. Little concentration-dependence was seen. This finding is consistent with the process of internal dismutation or direct peroxide production. To test if this phenomenon was dependent upon A containing two hemes, supporting the theory that peroxide production occurs internal to the maquette, a one-heme containing solution of 20µM A (same heme concentration) was observed. When A contains one heme, it also exhibits a lag phase and nearly identical kinetics to C.1, suggesting it only can produce peroxide directly when it has two hemes and demonstrating that the heme oxidation rate is similar to C.1.

5.2.5 Hydroxyl radical production

Hydroxyl radicals were detected from each maquette upon stopped flow mixing at 20°C in 20mM Tris, 20mM NaCl pH 8 buffer of anaerobic reduced 25µM protein with air-saturated buffer containing 20µM Acid Orange 7 (AO7). AO7 has its central azo bone cleaved hydroxyl radicals, eliminating its absorbance in the 400nm to 536nm
region of the spectrum (Wu et al. 2012; Hammami et al. 2008). Although its peak absorbance is at 486nm, its degradation was monitored at 519nm, to avoid spectral overlap with the Soret bands of hemes in maquettes. Signals were plotted as fraction of original AO7 absorbance lost compared to initial 519nm absorbance and averaged over 10 runs. All results are shown in Figure 5.7.

Both B.1 and C.1 produce large amount of hydroxyl radicals at similar rates of 6.5 mM⁻¹s⁻¹ and 9.7 mM⁻¹s⁻¹ and yields of 0.73 and 0.54, respectively. As a yield is defined as the fraction of AO7 oxidized, these yields imply efficient hydroxyl radical generation. A 25µM protein sample contains 50µM heme, which is therefore able to generate 16.6µM hydroxyl radicals if 100% of electrons from their hemes went to the formation of hydroxyl radicals because it is a 3-electron process. Therefore, oxidation yields of a 20µM AO7 of 0.73 and 0.54 correspond to efficiencies of 88% and 65%

These are remarkably high considering three SOs must react sequentially and the resulting hydroxyl radical must react with the probe prior to oxidizing the protein or another SO to form water. The greater yield of B.1 compared to C.1 is consistent with its more rapid rate of SO generation leading to more rapid dismutation to peroxide, such that its more likely to experience all three reactions due to higher temporary concentrations of SO and peroxide after mixing.

Not surprisingly, A did not show appreciable yields of hydroxyl radical formation although the small about was produced rapidly. If A directly produces peroxide and only contains two electron-carrying centers, then there is no opportunity for a three-electron process; each reaction must be a multiple of two electrons. The amount leading to hydroxyl radicals must be electrons contributing through the subpopulation that produces SO in Figure 5.5 and shows biphasic behavior of peroxide production in Figure 5.6.
Figure 5.7: Hydroxyl radical production as detected by AO7. Each trace monitors the decay of AO7 visible absorbance at 519nm after stopped-flow mixing of 25µM reduced 2-heme-bound maquette against 20µM AO7 in 20mM Tris, 20mM NaCl pH8. Plots are percent of initial AO7 absorbance remaining over time. Traces are shown for A (blue), B.1 (green), C.1 (red), and a 50/50 mixture of C.1 and A.

Because hydroxyl radicals are formed through the reaction of peroxide with SO, the rate of production should be accelerated if there is no lag phase for SO dismutation to peroxide. To test this hypothesis, the AO7 oxidation ability was tested of a 25µM 50/50 mixture of peroxide-generating A and SO-generating C.1, which oxidize at the same rate. This mixture did produce the fastest hydroxyl generation rate of 200 mM⁻¹s⁻¹, considerably faster than either protein on its own. The yield was considerably lower, only oxidizing approximately 20% of the AO7. This was expected, as one protein directly produces peroxide, two-thirds of the SO generated from C.1 must react with peroxide for 100% efficiency. SO will also self-react to form peroxide and this reaction serves as competition for the SO-peroxide reaction. Additionally, the rapid rate of hydroxyl radical generation may simultaneously increase the temporary early concentrations of hydroxyl radical and SO, which can then react to form water.
5.2.6 Histidine and gas binding dynamics

To determine which rates govern the preference towards stable oxygen binding, SO production or peroxide production depicted in Figure 5.2, the microscopic binding rates were determined for each protein for histidine on, histidine off and CO on using CO flash photolysis. These experiments were performed according to protocols developed by Mark Hargrove for natural hexacoordinate globins (Hargrove 2000). CO was bound to a reduced protein sample by bubbling a known percent of CO in argon gas and then reducing with dithionite. These samples were then flashed with a laser at 532nm to remove CO from the heme center. Difference spectra were taken with a charge coupled device (CCD) camera at prescribed delay intervals from 30ns to 300ms and averaged over 20 runs for each delay. Transitions were followed at 432nm and 424nm to monitor the transition of each protein from the pentacoordinate ferrous state to the hexacoordinate ferrous state and then to the CO-bound state. These transitions were confirmed with singular value decomposition (SVD) analysis. Single exponential rates were then found for histidine ligation and removal from the heme center. To determine the second order rate constant for CO binding, the experiment was repeated for a series of CO concentrations, 20%, 50% and 100%. All rates are reported in Table 5.2.

CO binding was much slower for B.1 and C.1 compared to A, B, and C. This is consistent with the CO-binding times previously observed by stopped-flow mixing for B and B.1 (Koder et al. 2009). Design changes to remove helical strain between B/C and B.1/C.1 in order to improve histidine affinity did not have as drastic effects as expected. Both modifications lead to slightly slower histidine on and off rates, but are still more similar to their oxygen-binding precursor designs than natural globins are to one another that exhibit the same differences in oxygen reactivity. CO binding was considerably slower, however, indicating difficulty with gas binding despite similar amount of time.
spent in the pentacoordinate state for each maquette. Therefore, it is more likely that the inability of these proteins to bind oxygen stems from poor soluble gas access to the heme site or a poor oxygen affinity than competition with histidine ligation.

A shows similar histidine on rates to the other maquettes, but much more rapid histidine off rates and CO-binding rates than any other maquette or natural globin. This finding is consistent with the also observed more rapid oxygen-binding rates than other maquettes as observed by stopped-flow mixing of reduced protein with oxygen-saturated buffer, as discussed earlier. The similar histidine on and off rates result in the heme spending far more time in the pentacoordinate state than other proteins. However, this is not sufficient to support stable oxyferrous state formation, as demonstrated by its rapid oxidation time akin to those that no not bind oxygen at all, such as C.1. Therefore, oxygen binding is destabilized through an alternate means than histidine competition.

<table>
<thead>
<tr>
<th>Protein</th>
<th>( k_{\text{His on}} ) (s(^{-1}))</th>
<th>( k_{\text{His off}} ) (s(^{-1}))</th>
<th>( k_{\text{CO on}} ) (mM(^{-1})s(^{-1}))</th>
<th>( k_{\text{O2 on}} ) (mM(^{-1})s(^{-1}))</th>
<th>( k_{\text{O2 off}} ) (s(^{-1}))</th>
<th>( k_{\text{oxidation}} ) (s(^{-1}))</th>
<th>( k_{\text{SO Production}} ) (s(^{-1}))</th>
<th>SO yield by MCLA</th>
<th>k(_{\text{H2O2 Production}}) (mM(^{-1})s(^{-1}))</th>
<th>k(_{\text{HO2 Production}}) (mM(^{-1})s(^{-1}))</th>
<th>HO(_2) yield by AO7</th>
<th>Yield by MCLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>350</td>
<td>200</td>
<td>500000 (2500)</td>
<td>8600 (43)</td>
<td>*</td>
<td>10</td>
<td>2.4</td>
<td>0.2</td>
<td>660 (380)</td>
<td>150</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>*310</td>
<td>*17</td>
<td>*160000 (950)</td>
<td>300 (17)</td>
<td>*</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.1</td>
<td>200</td>
<td>8</td>
<td>2900 (360)</td>
<td>*</td>
<td>*</td>
<td>&gt;20</td>
<td>8.01</td>
<td>0.43</td>
<td>450</td>
<td>6.5</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2100</td>
<td>50</td>
<td>215000 (4300)</td>
<td>800 (16)</td>
<td>*</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.1</td>
<td>960</td>
<td>7</td>
<td>1500 (210)</td>
<td>*</td>
<td>*</td>
<td>9.7</td>
<td>1.34</td>
<td>1</td>
<td>390</td>
<td>9.7</td>
<td>0.54</td>
<td></td>
</tr>
</tbody>
</table>

°Human Ngb 1800 0.6 50000 170000 0.5 0.025 0.025 * * *
°Human Cygb 140 1.5 650 2500 0.9 0.02 0.02 * * * *
°GLB-26 20000 0.35 20000 40000 * 2 2 * * * *
°Horse Mb * * 500 14000 11 7E-05 0.00007 * * *
Nox-2 * * * * * * 1455 *13 * * * *
^Nox-4 * * * * * * 0.93 * 460 * *

Table 5.2: Measured and reported rates for maquettes, natural globins and NADPH oxidases (Nox). Symbols denote citations as follows; all other rates were determined in this work: °(Kiger et al. 2011); †(Koder et al. 2009); ‡(Koshkin et al. 1997); §(Cross et al. 1985); *rate of production (Nisimoto et al. 2010) and rate ratio of SO/H\(_2\)O\(_2\) (Dikalov et al. 2008). *rates or yields not measured or not applicable.
5.3 Resolving ROS

5.3.1 Resolvable species of ROS produced

This work establishes the first system for the kinetic resolution of SO, hydrogen peroxide and hydroxyl radicals from a protein system. It would be ideal to use a probe for SO that has faster kinetics, such that quantitative SO production rates can be resolved. However, the simplicity of the maquette system allows for qualitative comparison of both SO production rates and yields with the MCLA system. Hydrogen peroxide has long been the easiest ROS to specifically detect and the maquette system is just as amenable to peroxide detection as any other protein system with HRP and Amplex Red. The detection of hydroxyl radicals was quantitative and kinetically resolvable using highly colored dyes degradable by high-potential oxidizing agents. This was not surprising as the incredibly high reduction potential of hydroxyl radicals, above that of other possible species produced, allows for their specific detection with any probe requiring such a potent oxidizer. While there is a dearth of kinetically resolved hydroxyl radical detection from natural proteins in the literature, with main detection means relying on the quantitation of end damage such as oxidized lipids, this did not need to be the case. The azo dye system used here could also be used for the resolution of hydroxyl radicals from natural proteins. This simply seems to be the first example of literature crossover between the Fenton chemistry of azo dye destruction and the desire to measure hydroxyl radical production kinetics.

The final step of this process, the reduction of hydroxyl radicals to water, is theoretically also resolvable using this system. However, these experiments would require isotopic doping and are likely not kinetically resolvable on the milliseconds-to-seconds timescale, so were not pursued in this work. To quantify end products, it may
actually be easier to quantify damage caused by hydroxyl radicals and the amount of hydrogen peroxide remaining in a system let to completion without probe interference. The remaining electrons not accounted for by these products could be assumed to be used in the four-electron reduction of oxygen to water.

The ability to kinetically resolve SO, hydrogen peroxide and hydroxyl radicals has allowed for some conclusions to be drawn from their rates and detectible quantities. For SO, faster rates of production lead to greater a detectible yield. This is evident for the comparison of **B.1** and **C.1**, for which **B.1** produced SO at a faster rate but has less than half the detectible yield. Both likely produce SO directly in equal quantities, but the more rapid rate of generation from **B.1** leads to higher temporary concentrations of SO, which in turn increase the dismutation rate. As the SO dismutes faster, this reaction more successfully competes with the reaction of SO with the MCLA probe, hence less is detectible. Kinetic resolution of peroxide confirms this hypothesis, as both produce the same quantity of hydrogen peroxide and both exhibit a lag phase indicative of dismutation rather than direct peroxide production, with **B.1** producing peroxide with a slightly shorter lag time and faster rate. **A**, in contrast, produces SO at a rate similar to **C.1**, but only 20% of the total yield is observable. The rate therefore cannot account for the difference in yield due to rapid dismutation, so peroxide must be directly produced. This is confirmed by peroxide production kinetics in which **A** does not exhibit a lag phase and produces peroxide more rapidly than **B.1** or **C.1**. The SO produced from **A** must therefore be from a population undergoing a different oxidation process than the remaining 80%.

These theories are consistent with the kinetics and yields observed for hydroxyl radical production. The rapid generation of SO leads to higher detectible yields of hydroxyl generation, just as it did for peroxide production, with **B.1** producing the highest
observable yield at 73% which corresponds to 88% of total electrons going to the formation hydroxyl radicals. **C.1**, with slightly slower SO generation rates produces slightly less hydroxyl radicals. **A** on the other hand, hardly produces any hydroxyl radicals. This was to be expected, since the direct production of peroxide accounts for 80% of its electrons, even if every remaining SO were able to react with them, **A** would still produce only a 20% total yield of hydroxyl radicals. Rather, each time SO self-reacts to produce a peroxide, two SO molecules are removed from potentially reacting to produce hydroxyl radicals. Hence the majority of the electrons from **A** likely result in peroxide formation instead of hydroxyl radicals or water.

5.3.2 *Resolvable mechanisms of ROS production*

This system also allows for mechanistic resolution of how these proteins are producing their respective ROS products. As the first protein system in which potentially ROS-producing factors can be systematically varied, fundamental questions of protein cofactor ROS production can be tested. Because the proteins are simple and can be designed to contain a small number of cofactors, it is possible to spectrally resolve each cofactor as it undergoes electron changes in real time.

The design sequences began with those that bind oxygen stably, suppressing oxygen reduction to levels that would not physiologically cause damage. This is evident by **B** and **C**, oxygen-binding maquettes, stably binding oxygen for tens of seconds and not generating any detectable MCLA signal. While the engineering elements required for the oxygen-binding function of these proteins have been elucidated to some extent in past work and earlier chapters of this thesis, the means by which the oxyferrous state stability can be broken to generate ROS rapidly had not. Here we were able to show that ROS can be produced rapidly by two different mechanisms in similar proteins.
In B.1 and C.1, glutamates adjacent to the heme sites that had been shown to be necessary for stable oxygen binding were mutated to alanines. It was previously thought that these glutamates rotate into the hydrophobic core of the bundle upon heme binding to induce helical strain required to promote histidine de-ligation. Removing these glutamates did abolish all oxygen binding and slowed CO binding, as expected, and rapidly produce SO when exposed to dioxygen. Exposure of reduced maquette to oxygen does not result in any observable oxyferrous formation, so this SO generation must be through outer-sphere ET to oxygen. B.1, a tethered dimer, produces SO more rapidly than C.1, a single chain construct, likely due to the more dynamic helical bundle exposing heme center so that dioxygen can diffuse in closer and rapidly to the ferrous heme site.

To show that these glutamates were inducing helical strain, histidine on and off rates were measured using flash photolysis (Table 5.2). It was expected that these proteins generated SO rapidly because they did not spend enough time in the pentacoordinate state to allow for oxygen binding, akin to natural globin GLB-26 that does not bind oxygen. For GLB-26 the histidine off rate is similar to that of other globins, but the histidine on rate is orders of magnitude faster. Therefore, its histidine affinity for ferrous heme is much higher than other globins and it spends the majority of its time in the hexacoordinate state. It was possible through removing helical strain that the histidine off rate would be much lower than oxygen-binding maquettes. While they did have approximately half to one-eighths the histidine off rates, their on rates were also slowed to one-to-two-thirds the original rates. As a result, the slower histidine rates do not appreciably change the histidine affinities from their oxygen-binding counterparts and all proteins are more similar than any are to a natural globin. Even if histidine on rates were unchanged, the change in off rates is not enough to account for lack of oxygen
binding given the oxygen on rates of B and C. So, while glutamate rotation may be inducing some helical strain, this is not the reason for their necessity in order for B and C to stably binding oxygen. Rather glutamates may be stabilizing the distal histidine for oxygen binding or sterically obstructing the heme from burying into the center as far as an alanine mutant that would have a larger heme binding ‘pocket’. Either way, rather than being dictated by histidine on and off rates, the ability to bind oxygen appears to be dictated by the rate of oxygen binding, by either the oxygen affinity or gas access to the heme site.

In A, the oxyferrous state does form, though transiently as observed through stopped-flow mixing of reduced protein with oxygenated buffer. Not all hemes likely form the oxyferrous state as the oxyferrous Q-band at 574nm is weak. The remaining population likely to undergo outer-sphere ET to oxygen, like B.1 and C.1, as is observed as SO in MCLA experiments. The oxygen that does bind only stays bound for about 100ms before the heme oxidizes due to water access to the core allowed by the tethered homotetramer topology and poor interfacial packing (Anderson et al. 2008). The water-catalyzed oxidation of the oxyferrous heme implies an inner-sphere mechanism of ROS generation. The predominant ROS generated from A is peroxide, not SO, suggesting an internal mechanism for peroxide production. The direct internal mechanism is supported by experiments showing a lack of concentration dependency of peroxide production rates as well as the elimination of direct peroxide production kinetics of a one-heme A.

There are two means by which this could occur outlined in the lower two lines of Figure 5.3. First, both hemes could simultaneously release their SO into the core of the protein such that they are in such close proximity that they dismute readily and leave as peroxide. Second, one heme could be in an oxyferrous state with the electron delocalized to the oxygen such that it could be described as a ferric-SO state or the SO
could have just released as the second reduced heme undergoes ET to it. This would create a ferrous-SO state, which can then take protons from the water-accessible core to generate peroxide directly. As the oxyferrous state is not clearly formed for the majority of hemes at any one time due to the similar oxygen on and off rates, the second mechanism is more plausible. The first mechanism could still be the case if simultaneous oxidation of the hemes is not necessary due to a SO-binding pocket on or in the protein, retaining the SO until the second heme can oxidize. Given the highly dynamic nature of A and rapid water access to all residues, this also seems less likely than an internal ET event given the close proximity of the hemes to one another.

5.4 Conclusions

5.5.1 Learning about natural systems

This system allows for inferential conclusions to be drawn about natural systems by uncovering fundamental principles of protein chemistry. Here we established that relief of helical strain is not sufficient to prevent oxygen binding, as previously thought, for globins. The strain does not affect the histidine affinity to a large enough extent to explain lack of oxygen binding. Given the differences in oxygen binding and histidine affinities between natural hexacoordinate globins, this work suggests too great an emphasis has been placed on the distal histidine affinity in all but extreme cases, such as GLB-26. Rather, oxygen access to the heme site and oxyferrous state stabilization factors, both variables considered more heavily for pentacoordinate globins, play a larger role for the hexacoordinate globins as well then previously thought.

The placement of the hemes in these maquettes is analogous to those in NADPH oxidases (Nox), allowing for some insight into their ET processes. As discussed earlier,
there is little understood of the mechanism of constitutive peroxide production from Nox4 despite high sequence and structural homology to other SO-producing Nox enzymes. One of the suggested mechanisms involves ET from the internal to external heme while the external heme binds oxygen. Nox4 has not been shown to bind oxygen, which does not preclude its binding as experiments were not rapid enough to detect a transient oxygen-binding state. This work suggests that 14Å heme separation in a transient oxygen-binder, A, is sufficient to promote constitutive peroxide production instead of SO production. If oxygen on and auto-oxidation rates are very similar, τ of 50ms and 100ms in this case, then peroxide can likely be formed through an interprotein ET event. Additional experiments testing the first steps of the ET events in Nox enzymes could be done in flavin-containing maquettes using this system. Further experiments to uncover

Figure 5.8: Future experiments possible with the ROS laboratory. A non-exhaustive group of illustrated examples are depicted, as the system can be poised prior to oxygenation. Protein maquettes are grey outlined boxes; reduced and oxidized hemes are red and brown squares; zinc porphyrins are purple squares; NADPH is orange; flavins are yellow; membranes are grey solid; oxygens are red spheres; protons are grey spheres.
factors effecting Nox ROS production can be done using amphiphilic versions of these maquettes. A Nox model system could be established in which bound flavins undergo reduction by NADPH and transmembrane ET to oxygen is observed (Figure 5.8).

Observing flavin ROS production may also help to uncover ROS production mechanisms from mitochondrial complexes, particularly complex I. Oxygen reactivity of heme sites, such as in complex III, can also be explored further with variability of these heme-containing maquettes. This system also leaves open the possibility of photoactivation using light-activated cofactors, such as zinc tetrapyrrroles, to look at rapid timescales of ROS production following ET events. A non-exhaustive schematic of possible experiments is outlined in Figure 5.8.

5.5.2 Engineering more robust protein systems

Aside from leaning about natural systems, understanding ROS generation may help us design more robust protein systems for industrial purposes. Potent radicals, such as hydroxyl radicals, rapidly degrade engineered ET proteins. Outside of a natural system, there is no ability to repair proteins akin to photosystems; so avoiding these reactions is paramount. While we may not be able to avoid ‘leak’ of electrons to oxygen completely, we can likely suppress some of the reduction of oxygen as well as bias the leaks that do occur towards peroxide production. This system would allow for the detection of each product on a kinetic scale and the systematic testing of variables to decrease the ROS yields from industrial catalysts.

Alternatively, Fenton chemistry has its uses in industrial protein engineering. Laccase enzymes producing hydroxyl radicals are used for a range of industrial
processes from paper processing to pollution remediation (Prousek 2007). This system would also allow for the testing of hydroxyl-forming chemistry as well as the direct development of Fenton proteins to do participate these industrial reaction processes, possibly with greater catalytic efficiency.

5.5 References


Discher, B. et al., 2005. Design of Amphiphilic Protein Maquettes: Controlling Assembly, Membrane Insertion, and Cofactor Interactions†. *Biochemistry*.


Esterházy, D. et al., 2008. ... Reactive Oxygen Species by Complex I (NADH: Ubiquinone Oxidoreductase .... *Biochemistry*.


St-Pierre, J., Buckingham, J. & Roebuck, S., 2002. Topology of superoxide production from different sites in the mitochondrial electron .... *Journal of Biological* ....


Chapter 6: Conclusions

6.1 New rules of protein design

Using natural selection, nature designs functional proteins with a wide range of chemistries. While this process produces sufficiently effective proteins, it is blind to efficiency and robustness. Natural selection generally results in proteins that are highly complex and more fragile than necessary. The multiple, overlapping roles that each amino acid performs prevent natural proteins from withstanding many mutations and obscure our ability to uncover these roles experimentally (Muller 1964; Darwin 1859). Past maquette work has demonstrated that man-made de novo protein design offers an alternative to studying natural proteins by allowing us to construct from scratch simple molecules with minimal complexity that confer natural functionality. By building from the bottom-up rather than studying from the top-down, we can uncover the functional roles of individual amino acids or patterns of residues (Koder et al. 2009; Farid et al. 2013). However, until this work, no maquette study was complete for there was always left a level of complexity not understood.

In this work, to create a fully transparent sequence in which each design element has been validated we determine the minimal requirements needed to take a functionless, simple sequence with just 4 types of amino acids sufficient to fold into a structurally simple 4-helix bundle and, by changing just a few amino acids, confer what many biologists would consider to be a sophisticated function, namely relatively stable oxygen binding to ferrous heme, akin to natural oxygen transport proteins such as myoglobin and neuroglobin. In this way we identify the minimal design requirements for basic physical chemical elements as cofactor ligation, residue mobility and exclusion of
water from the bundle interior, that underlie the oxygen binding engineering in proteins.

By uncovering the basic protein engineering underlying oxidoreductase function, we achieve a simple, modular design. These designs avoid the complex amino acid interdependences commonly found in natural proteins by attempting to ascribe a few functions per amino acid as possible. Without interdependences, engineering can move forward in a step-by-step fashion to increasingly refined functions involving multiple sites within the protein and multiple cofactors. Here we have presented a protein representation of the minimal design requirements for oxygen transport.

The work presented here is a distilled and rigorous examination of past work in protein design and an example of a truly step-by-step, from-scratch iterative design process developing a fully justified functional protein. This work demonstrates that in removing complexity from protein systems and designing into a non-natural sequence, both natural functions can be achieved while being understood and physical robustness need not be sacrificed. This highly simple protein confers hyperthermal stability in both the apo and holo forms and can retain cofactor for months. These parameters are much greater than have been achieved for any previous protein design and did not require any functional sacrifice. This work demonstrates that in removing complexity and interdependencies, not only can domains be modularly controlled but functional and structural attributes can as well. For example, one can modify the function of a protein independently of its structural stability or rigidity, allowing for highly robust, functional designs.

The success of this work raises the question as to why it was possible in this system to engineer robust and functional proteins in only a few stages without the need to ‘tinker’, as has been necessary in past design work (Luisi & Chiarabelli 2011). There are two reasons. First, this work was done on the design of a function that had been
achieved previously in a non-natural protein. This is not to say that it could not have been done for a newly designed function, but the use of a previously designed function allowed this work to begin with engineering insight, though incomplete, from past experiments on oxygen binding. We understood the function we were attempting to design fairly well. While not all protein contributions were known yet, the chemistry of the requirements were well understood from past work and the large body of literature on heme oxygen binding in natural proteins.

The second reason for this work’s ease of success, only requiring three directed design iterations adding sub-functional elements, is that it did not attempt high level structural resolution or tuning. Each element was added independently, but not as a single amino acid mutation nor as a modification attempting structural exactness. Each element was conferred by a change in a gross structural property: heme binding was improved through removal of bulk from the histidine sites; water was restricted from the core by hydrophobic core bulk; structural thermal stability was maximized through binary patterning up each helix. None of these elements required crystal structures or high-resolution molecular modeling. Not only did they not require it, this work suggests engineering using such data would have impeded the design process.

Perhaps the issue with past work that required ‘tinkering’ was not just that we did not fully understand our material as Steve Benner has suggested, but that we have been expecting too much of it. Our strong inclination at protein biochemists is to ascribe importance and beauty to the evolution of natural proteins as beacons of efficiency and intelligent engineering. This leads us to look at each angstrom-level element as potentially important and design our man-man proteins in the same manner. However, perhaps the beauty of evolution is that it is not fooled by such precision. Evolutionary
processes are impeded by the need for exactness and proceed through randomness to their solutions, which are remarkable in their efficacy but often not in their efficiency.

Other fields of industrial chemical engineering do not hold themselves this standard of precision and have had more success. Non-protein polymer design has been able to design a multitude of protein-like catalysts through rational gross property design processes (Madhavan et al. 2008) and molecular imprinting (Wulff 2002). When polymers are completely non-natural, even when they attempt natural catalytic reactions, the scientific community does not ask for atomistic detail of their designs or consider modification of a monomer to be the smallest desirable unit of change. The success of polymer design suggests we may be holding ourselves back by our own perceived rigor.

Proteins should be capable of higher levels of catalysis and more intricate design patterns than polymers can be easily constructed to obtain, due to polymers needing to be highly repetitious for ease of construction. Rather, proteins have the benefit of access of natural synthetic efficient machinery capable of any sequence. However, that should not mean that we should to exhaustively test every possible sequence or design each residue separately. This work suggests reaching catalytic success will be much easier if we design in patterns to adjust properties.

Additionally, those patterns need to not necessarily be of specific residues. While the functionality or structural nature of a residue is important, it is likely that many residues contain the required characteristics. There is likely considerable freedom in amino acid choice when we design, which will lead to equal levels of functionality. This freedom is evident in the limited number of amino acids needed in this work to confer a high level function, considerably less than had been suggested by the literature for any function (Romero et al. 1999). It may also be the reason intuitive and rational hand design has succeed over computational design since our minds are capable of ascribing
gross chemical properties to sequences whereas we feel it necessary to explore a large space of variations once we involve computers to be able to say the final sequence is optimal. Just as evolution does not need to test each amino acid substitution at each site, neither should we. Rather, we should start with defining desired parameters for a site, whether that site is a specific amino acid or a whole domain, and choosing from families of residues to satisfy the requirements. Letting go of residue-specific importance in all but some cases may relieve some of the constraints to our design processes slowing our progress.

As we move forward in designing man-man proteins, this work suggest we can treat proteins as generalized polymer scaffolds for active sites instead of attempting structural exactness, so long as it forms a protein with desired general properties. We do not expect a polymer chemist to have tried every single possible polymer unit or combination to accomplish their goal and provide an atomistic model of what it will do, so we should not constrain ourselves by these standards.

6.2 New applications for protein design

After describing a transparent oxygen-binding protein, this work describes how to ‘break’ the function in multiple ways. For the first time, maquettes have been used to ask a series of questions related to oxygen reactivity. The system developed here has allowed for the first kinetic and yield descriptions of SO, hydrogen peroxide and hydroxyl radical production from a protein system. Because maquettes are simple and modifiable, these proteins differ much less than natural proteins do from one another. As a result, mechanisms of ROS production can be uncovered and protein structural elements contributing to each mechanism can be elucidated.
While this work is not exhaustive, it is sufficient to establish a methodology for ROS study. This work also has elucidated some ROS production mechanisms as they pertain to NADPH oxidases (Nox) and globins. For the first time, it has been described here how two homologous proteins with minimal changes that both produce high levels of ROS can produce predominantly different species. It was found that peroxide production can be favored over SO production, even from single-electron cofactors, such as hemes, by the formation and decay of the oxyferrous state occurring on the same timescale as each other and interprotein ET for hemes 14Å apart. This suggests that Nox4 may produce peroxide through a similar mechanism in which one heme binds oxygen as a second electron is transferred to it from the other heme.

This system has also allowed for the further description of oxygen-binding requirements. By comparing a protein without the glutamates suggested to induce helical strain that rapidly produces SO to its oxygen-binding precursor design, the role of these glutamates can be uncovered. While the glutamates are clearly necessary for oxygen binding and the histidine off rate does slow in their absence, there is not a large enough difference in histidine affinity between oxygen-binding and SO-producing maquettes for strain to account for their difference in oxygen reactivity. Since the potentials are both highly favorable for oxygen reduction from both hemes, the oxygen on rate or stabilization must be affected by the glutamates. Possibilities include glutamate occlusion of the heme from fully burying, allowing greater soluble gas access, and glutamate hydrogen-bonding to the 'distal' histidine upon oxygen binding, stabilizing its hydrogen-bonding to bound dioxygen akin to myoglobin. These mechanisms could be further elucidated by the use of a similarly sized residue that cannot hydrogen bond to histidine in glutamates place. Whichever is the case, it is clear from this work that far too great an emphasis has been placed on histidine dynamics related to oxygen-binding of
hexacoordinate globins and not enough work has been done related to oxygen access to the heme site or potential stabilizations of the oxyferrous state akin to pentacoordinate globins.

The conclusions drawn from this system already suggest that it is a viable molecular laboratory for the study of ROS generation and oxygen reactivity from protein systems. With kinetic resolution of singly, doubly and triply reduced oxygen generation, more questions pertaining to natural proteins can be asked in an incremental fashion such that we can uncover possible mechanisms of protein-oxygen chemistry.

6.3 Future directions

This work attempts to describe each step of the design process, such that neither linear design from one functional design to another nor starting from scratch for each new design is necessary. We aim to create a toolbox from which new designs can draw the desired construction elements without needing to return to first principles. We also aim to demystify the steps between designs such that each earlier iteration is not a failure on the way to success so much as a sub-success on the way to the target function, which can be returned to in order to create alternative designs not requiring later-introduced design elements.

The functional description is not yet complete for ROS-generation proteins. Now that it has been shown to be possible to design specific SO and peroxide producers, we can go back and transparently redesign, in the same fashion as for oxygen transport in Chapters 3 and 4, their functionalities to remove obscuring complexity from their designs. From the existing designs, more can also be created in which to test the oxygen reactivity of other cofactors, other arrangements of cofactors, and proteins spanning membranes. However, it is suggested to reconstruct the designs in a
transparent manner first, such that a ROS generation engineering toolbox can be created first. Further understanding the design principles of ROS generation may also help us avoid these reactions in future designs. This is particularly important for the elongation of the oxyferrous lifetime necessary for the design of useful blood substitutes.

Lessons learned from this thesis about the design process and thermal stability can also be used to design forward to other oxygen reactivities. By designing by the modification of gross properties in combination with other work ongoing in the lab to alternative ligations, progress should accelerate towards success in achieving other catalytic dioxygen chemistries, such as monooxygenase function. The principles uncovered here for drastically improving the thermal stability of proteins without sacrificing inter-domain dynamics can also be applied to other proteins, both man-made and natural. This work describes a method of ‘thermophilizing’ helical sequences that could be used to rationally hand design hyperthermally stable proteins or improve the thermal stability of natural proteins. The principle that removing polar residues improves thermal stability without sacrificing dynamic function if charged residues are placed intelligently could also be used to restrict computational design algorithms for thermally stable designs and restrict directed evolution amino acid libraries from using polar residues. Each of these approaches may allow for the design of industrial and medical proteins with higher levels of stability more quickly with higher levels of success.

6.4 References


Appendix 1: Materials and Methods

A1.1 Protein expression and purification

For all single-chain proteins, genes were designed by hand and ordered from DNA 2.0® with a 6 histidine tag and tobacco etch virus (TEV) protease cleavage site in a pET32b(+) plasmid vector with ampicillin resistance. Plasmids were transformed into BL21(DE3) E. coli cells by gently mixing 10ng DNA into 50uL competent cells, incubating on ice of 30 minutes, heat shocking cells at 42°C for 25 seconds, placing on ice for 5 minutes and then shaking at 37°C for 1 hour in 1 mL cell S.O.C. media (Invitrogen® 15544). Cells were grown on ampicillin luria broth (LB) plates overnight and single colonies then grown up in LB media with 75mg/L ampicillin to an OD$_{600}$ of 0.4 to 1. Production of protein was then induced with 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) for 4 to 6 hours. Protein was purified using a NiNTA column (GE HisTrap FF®) by fast protein liquid chromatography (FPLC), dialyzed into 50 mM tris(hydroxymethyl)aminomethane (TRIS), 0.5 mM ethylenediaminetetraacetic acid (EDTA) pH 8 and cleaved for 2 days with TEV protease in the presence of dithiothreitol. Protein was purified again by FPLC using NiNTA column and then further purified on a Grace C18® column by high performance liquid chromatography (HPLC). Peaks were mass confirmed by Matrix-assisted laser desorption/ionization (MALDI) spectroscopy and lyophilized.

A1.2 Peptide Synthesis

All non-single-chain proteins were synthesized on Liberty Peptide Synthesizer® using solid phase peptide synthesis with HBTU (O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro -phosphate) chemistry. The proteins were cleaved off of the resin
using a mix of trifluoroacetic acid (TFA), thioanisole, anisole and ethane dithiol in a ratio of 9 : 0.5 : 0.2 : 0.3. They were then purified via HPLC on a Grace C18\textsuperscript{©} column and the identity of the protein was confirmed with MALDI spectroscopy. Proteins were then lyophilized to remove acetonitrile and TFA. The proteins were resuspended in borate and dialyzed overnight in 3500 MWCO tubing to facilitate dimerization. Concentrations of dimers were determined using Agilent 8453 UV-Vis Spectrophotometer\textsuperscript{©}.

**A1.3 Heme titrations**

Protein was dissolved to desired concentration in 20 mM TRIS, 20 mM NaCl pH 8 in a 1 mL quartz cuvette. UV-visible spectra were taken from 0 to 3 equivalents heme at 0.1 to 0.5 equivalent increments after equilibrium was reached, anywhere from 10 minutes to one hour, at room temperature. Heme was added from 0.7 mM 10\% DMSO, 90\% 20 mM TRIS 20 mM NaCl stock. The absorbance at 412 nm was plotted against heme concentration and best fit for one-heme association constant using equation 1, where H is heme, P is protein and extinction coefficients are for heme in bound and free state. UV-visible spectra were taken on a Varian Cary 50 Bio Spectrometer\textsuperscript{©}.

(1)

\[
Abs = [H] \times E_{\text{free}} + (E_{\text{bound}} + E_{\text{free}}) \times \frac{(K_D + [P] + [H]) - \sqrt{(K_D + [P] + [H])^2 - 4[P][H]}}{2}
\]

**A1.4 Redox titrations**

All proteins was dissolved to 15 \mu M in 50 mM TRIS, 50 mM NaCl pH 8 with two hemes. Proteins were prepared by adding 2.2 equivalents of heme, letting site at room temperature for 20 minutes and then removing excess heme by filtering through a GE PD-10 Sephadex\textsuperscript{©} column. Samples for all subsequent experiments were prepared in
this fashion to the appropriate concentration. Redox mediators (indigo trisulfate, antraquinone sulfonate, hydroxynapthoquinone, phenazine and benzyl viologen) were added to 1 µM and the sample was degassed with argon. The potential was titrated with sodium dithionite to reduce and ferrous cyanide to oxidize and determined using a calomel and platinum electrode set. UV-visible spectra were taken on a Varian Cary 50 Bio Spectrometer© from -400 mV to +100 mV and back to 400 mV. The absorbance of the peak of the reduced state’s Q-band at 558 nm were plotted and best fit to a single midpoint curve.

A1.5 Low-temperature UV-Visible Spectroscopy

UV-visible spectra were taken on a Varian Cary 50 Bio Spectrometer© from 200 nm to 1000 nm of 5 µM holo protein, prepared as above in 25 mM sodium phosphate pH 8 and 30% ethylene glycol at -15°C. Oxidized spectra were taken after stirring under argon for 90 minutes and reduced spectra were taken after titrating sodium dithionite to minor excess. Carbon monoxide was bubbled for 5 minutes and then carbonmonoxy spectra were taken. Then, oxygen was bubbled through for 5 minutes and oxyferrous spectra were taken after carbon monoxide was flashed off using a Dolan Jenner Model 180© fiberoptic quartz halogen light source.

For gas mixture spectra, a fresh sample was prepared, reduced with minimal excess dithionite and bubbled with the contents of a 50 mL gas syringe preloaded with 50% CO and 50% O₂ gases. Spectra were taken every 5 minutes until equilibrium was reached. Spectra were fit to weighted multiplications of oxidized and reduced spectra of the respective protein and oxyferrous and carbonmonoxyferrous spectra taken at -15°C.
A1.6 Stopped-flow UV-Visible Spectroscopy

Each protein was prepared to 10 µM in 20 mM TRIS, 20 mM NaCl pH 8 as described above and stirred under argon for 90 minutes, then minimally reduced with dithionite. These samples were then rapidly mixed with oxygen-saturated 20 mM TRIS, 20 mM NaCl pH 8 using an OLIS RSM-1000 Stopped-Flow Spectrophotometer© at 15°C. Spectra were obtained from 386 nm to 611 nm each millisecond. Single variable deconvolution (SVD) analysis was performed to determine kinetics of oxyferrous formation and decay and corroborated with kinetics of the reduced and oxyferrous Q-band maximums of 558 nm and 574 nm, respectively. Decays were fit to a single exponential.

A1.7 Circular Dichroism (CD)

Apo, without heme, and holo samples, with 2 hemes, were prepared to 30 µM protein in 100 mM sodium phosphate pH8, 50 mM NaCl. Spectra were taken on an Aviv 410 Circular Dichroism Spectrometer© every 5°C from 5°C to 95°C. Complete spectra were taken at 5°C and 95°C.

A1.8 Heme Retention

Spectra were taken over 117 days of 5 µM protein with 2 hemes bound in 20 mM TRIS, 20 mM NaCl pH 8 incubated at 37°C in a 1 mL quartz cuvette in dark.

A1.9 Nuclear Magnetic Resonance (NMR) Spectroscopy

Protein was dissolved from lyophilized powder in 80% D$_2$O phosphate buffered saline (PBS) pD6.6 and 20% H$_2$O PBS pH7. Sample was diluted in same buffer mixture to 400 µM. Holo protein was assembled by adding heme from KOH hemin (Sigma
Aldrich®) stock titrated to 2:1 heme:protein ratio. Apo NMR was obtained using 250µL in a microcell tube and holo NMR was obtained using standard 500 µL tube. Holo spectra were taken with and without trimethylsilyl (TMS) capillary and aligned to determine chemical shifts without TMS peak in final spectra. Apo spectra were aligned by consistent peak at 3.7ppm. Spectra were taken using BBI® probe on Avance DMX 400 (9.4 Tesla) Bruker NMR®. 700 scans were taken of 16,384 points after 8 dummy scans using ESCULPT® water suppression sequence and 45° flip angle.

A1.10 CO Flash Photolysis

Samples of 5 µM holo protein (prepared as above) in 20 mM TRIS, 20 mM NaCl pH 8 were titrated with varying concentrations of CO to Argon gas (20, 50 100%) by bubbling the respective gas mixture while stirring in 1 cm pathlength 3 mL quartz cuvette. Samples were then reduced with dithionite with reduction conferred by UV-visible spectroscopy. Samples were illuminated for CO flash photolysis with Quanta-Ray INDI® Nd:YAG laser coupled to a BasiScan® nanosecond Optical Parametric Oscillator® to provide 1-20 ms pulses continuously tunable over the visible range to remove CO. Samples were then observed at varying delay times from nanoseconds to seconds set by a Stanford Research System DG535® digital delay generator. Spectra were taken using a timed Xe flash routed through a fiber optic bundle and flattened to a single line of fibers at the entrance slit of an Acton SP-2156® spectrograph, which was then was focused through the exit monochromater on a Princeton Instruments PiMax-3 ICCD® camera. Sample was prevented from overexposure to light with a ThorLabs® beam shutter. ICCD camera was normalized to exposure with no Xe or laser light to remove bias. Spectra were taken both with and without laser flash exposure and subtracted to remove noise for each concentration and delay interval.
Data were then followed as decays and growths between collected time points at 432nm and 424nm. SVD analysis was also performed to confirm the respective representations of these numbers of transitions from pentacoordinate to hexacoordinate heme ligation and CO recombination. The sum and product of the rates found at 432nm and 424nm were then plotted again CO concentration and fit to lines. The histidine on, histidine off and CO on rates were then found using the slope and y-intercept of these line derived from the sums and the slope found from the products as described by Mark Hargrove (Hargrove 2000) using the following equations:

\[
\begin{align*}
\gamma_1 + \gamma_2 &= k_{-H} + k_H + k'_{CO}[CO] \\
\gamma_1 \cdot \gamma_2 &= k_{-H} \cdot k'_{CO}[CO]
\end{align*}
\]

where \( \gamma_1 \) is the rate found from the 432nm transition and \( \gamma_2 \) is the rate found form the 424nm transition.

### A1.11 Fourier Transform Infrared (FTIR) Spectroscopy

Protein samples were prepared by adding 2 equivalents of heme and subsequently dialyzing three times in 95% \( \text{D}_2\text{O} \), 5% \( \text{H}_2\text{O} \) 10mM TRIS, 10mM NaCl buffer to remove any excess dimethylsulfoxide (DMSO) or heme. Maquettes were degassed with argon, reduced with sodium dithionite, and bubbled with carbon monoxide for saturation. A control solution was also prepared without carbon monoxide. Samples were plated between KCl wells under deoxygenated conditions of a continuous nitrogen gas flow. The FTIR was blanked to air after purge. FTIR spectra were averaged over 100 scans and each had the baseline of the respective maquette in the same conditions reduced but without carbon monoxide subtracted out to highlight the band of CO bound to ferrous heme.
A1.12 Superoxide Detection

2.2 equivalents heme was added and excess was eluted using a PD-10 column with 20mM TRIS, 20mM NaCl at pH 8. For reduced fractions, solutions of the maquettes containing both peptide and heme were then flushed with argon and reduced using minimal dithionite. Reduction was monitored using UV-Vis spectroscopy of the reduced Q-band for heme b at 558nm. A control solution of equivalent dithionite concentration was prepared at this time. Experimental solutions were brought to 15μM methyl-Cypridina-luciferin analogue (MCLA) from Invitrogen and mixed with air-saturated pH 8 100mM potassium phosphate buffer solutions in stop-flow experiments for effective 50% concentrations in the mixing cell.

Stop-flow experiments were performed on an OLIS RSM-1000 Stopped-Flow Spectrophotometer© at 20°C. MCLA undergoes a chemiluminescent decay when exposed to superoxide anions (Oosthuizen et al. 1997; Nakano 1998). Chemiluminscencce was observed through the side port under dark conditions using a photomultiplier directly connected to the mixing cell by a polished quartz light guide. Detection was optimized with a mirror on the opposite side of the mixing cell. Control experiments of buffers, oxidized protein and dithionite without protein were performed same-day. All traces were normalized to the amount of reduced protein as determined prior to the experiment injection by UV-visible spectroscopy. Traces were integrated for presentation to show total superoxide produced over time. Data were plotted over time and fit to single exponentials.

A1.13 Hydrogen Peroxide Detection

Samples were prepared in the same fashion as for superoxide detection experiments. Reduced protein was rapidly mixed via stopped-flow mixing just as in the
superoxide experiments with a solution of air-saturated buffer containing 5U/mL horseradish peroxidase (HRP) and 10µM Amplex Red. Absorbance of the highly colored catalytic product of Amplex Red and hydrogen peroxide in the presence of HRP, resorufin, was followed at its maximum peak at 574 nm (Votyakova & Reynolds 2004). All final concentrations of resorufin were determined by UV-visible spectroscopy without stopped-flow mixing. The same concentrations that exist in the mixing cell, half the injection concentrations, were mixed in a 1 mL quartz cuvette and the resorufin concentration was determined at 5 minutes. These data were used for normalization of the stopped-flow traces. Data were plotted over time and fit to appropriate exponentials.

A1.14 Hydroxyl Radical Detection

Hydroxyl radicals were detected in a manner identical to hydrogen peroxide. Samples were prepared in the same manner and mixed by stopped flow with a solution of 20µM Acid Orange 7 (AO7). Hydroxyl radicals cleave the central azo bone of AO7 (Hammami et al. 2008); this creates two pale products from the original highly colored orange species. The maximum absorbance of AO7 is 486nm but in order to avoid convolution of single wavelength kinetics with heme spectral transition, the decolorization of AO7 was monitored at 519nm. Data were plotted over time and fit to single exponentials.

A1.15 References


Appendix 2: Supplemental Data

A2.1 Gas binding

Figure A2.1: CO binding data of C.1 (Chapter 5) and 5 (Chapters 3 and 4). Reduced protein samples were mixed with CO-saturated buffer in 20 mM TRIS, 20 mM NaCl at 15°C via rapid stopped-flow mixing. The growth of the carbon-monoxyferrous state was monitored by a growth of the absorbance at 568 nm.
Figure A2.2: Oxygen binding to C. (Top) C shows clear reduced, oxyferrous and oxidized spectra by low-temperature spectroscopy at -15°C in 25mM sodium phosphate pH 8 at 30% ethylene glycol. (Bottom) SVD analysis results of determined fraction existing in reduced, oxidized and oxyferrous states using spectra above for stopped-flow mixing of oxygen-saturated buffer with reduced C at 15°C in 20mM TRIS, 20mM NaCl pH8.
Figure A2.3: Oxygen binding to C.1. C.1 does not bind oxygen upon stopped-flow mixing of oxygen-saturated buffer with reduced C.1 at 15°C in 20mM TRIS, 20mM NaCl pH8. (Top) Spectra in grey as C.1 progresses from reduced (blue) to oxidized (green) with no oxyferrous state intermediate. (Bottom) Progression of reduced decay by 558nm Q-band absorbance corresponds directly to growth of the oxidized species by fraction.
A2.2 Reactive oxygen species detection

Figure A2.4: Spectra for stopped-flow peroxide detection. All spectra correspond to t=0 (blue) and t=3.9s (red) for proteins in Figure 5.6 which shows plot of 574nm maxima.
Figure A2.5: Spectra for stopped-flow hydroxyl radical detection. All spectra correspond to t=0 (blue) and t=15s (red) for all proteins in Figure 5.7 which shows plot of 519nm decay.
A.2.3 CO flash photolysis

Figure A2.6: CO photolysis of A. Corresponding CO flash photolysis data for A (Chapter 5) used to calculate histidine on and off rates and CO on rates presented in Table 5.2. Data is discussed in section 5.2.6 and procedure is presented in A1.10. Delta absorbance spectra were plotted out from red (1 µs) to pink (300 ms) delays after flash. Pink 50% and 100% traces were excluded from data analysis due to noise.
Figure A2.7: CO photolysis of B. Corresponding CO flash photolysis data for B (Chapter 5) used to calculate histidine on and off rates and CO on rates presented in Table 5.2. Data is discussed in section 5.2.6 and procedure is presented in A1.10. Delta absorbance spectra were plotted out from red (1 µs) to purple (100 ms) delays after flash. 50% CO sample became partially oxidized so was excluded from data analysis.
Figure A2.8: CO photolysis of B.1. Corresponding CO flash photolysis data for B.1 (Chapter 5) used to calculate histidine on and off rates and CO on rates presented in Table 5.2. Data is discussed in section 5.2.6 and procedure is presented in A1.10. Delta absorbance spectra were plotted out from red (1 μs) to purple (100 ms) delays after flash.
Figure A2.9: CO photolysis of C. Corresponding CO flash photolysis data for C (Chapter 5) used to calculate histidine on and off rates and CO on rates presented in Table 5.2. Data is discussed in section 5.2.6 and procedure is presented in A1.10. Delta absorbance spectra were plotted out from red (1 µs) to purple (100 ms) delays after flash.
Figure A2.10: CO photolysis of C.1. Corresponding CO flash photolysis data for C.1 (Chapter 5) used to calculate histidine on and off rates and CO on rates presented in Table 5.2. Data is discussed in section 5.2.6 and procedure is presented in A1.10. Delta absorbance spectra were plotted out from red (1 µs) to purple (100 ms) delays after flash.
A.2.4 Protein characterization

Figure A2.11: Circular dichroism (CD) spectra of 4 and 5. Left shows CD spectra of 4 in holo, 2-heme bound form (red) and apo form (orange) and 5 in the same holo (black) and apo (blue) forms at 5°C (dark) and 95°C (pale) of 30 μM protein in 100 mM sodium phosphate pH8, 50 mM NaCl. Right shows % helicity of thermal melt as determined using 222nm molar ellipticity and percent helicity using Dichroweb software.