RICE UNIVERSITY

Expression, Folding, and Assembly Mechanisms of Monomeric Myoglobins and Oligomeric Hemoglobins

by

Premila Priya Samuel Mohan Dass

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

John S. Olson, Ralph and Dorothy Looney Professor of BioSciences

George N. Phillips, Jr., Ralph and Dorothy Looney Professor of BioSciences

George Bennett, E. Dell Butcher Professor of BioSciences

Kathleen Matthews, Stewart Memorial Professor of BioSciences

Yizhi Tao, Associate Professor of BioSciences

Laura Segatori, Associate Professor of Chemical and Biomolecular Engineering

Houston, Texas
October 2017
ABSTRACT

Expression, Folding, and Assembly Mechanisms of Monomeric Myoglobins and Oligomeric Hemoglobins

by

Premila Priya Samuel Mohan Dass

Despite differing in quaternary structure and protein sequence, mammalian myoglobins and hemoglobins share similar overall globin folds and nearly identical active site structures. The folding mechanism for monomeric apomyoglobin is a well-characterized 2-step pathway involving a molten globule intermediate containing an unfolded heme pocket. Holomyoglobin assembly involves reversible hemin binding to both the molten globule and fully folded apomyoglobin. A wheat germ based cell-free expression assay was developed to show that production levels of folded holomyoglobins correlate quantitatively with their overall apomyoglobin stability constants. Higher cell-free expression levels were observed for myoglobin mutants with heme cavity filling mutations that significantly increase apomyoglobin stability at the expense of hemin binding affinity. The new in vitro results are consistent with previous observations of myoglobin expression in animal muscle cells and E. coli, all of which demonstrate that apomyoglobin stability is the key determinant of holoprotein expression.

In contrast to myoglobin, the individual α and β apoglobin subunits of adult human hemoglobin A (HbA) are extremely unstable, despite being structurally similar to apomyoglobin. GdnHCl induced unfolding curves were measured for human apo- and
holo- HbA, fetal hemoglobin, and recombinant hemoglobins with either heme cavity filling apolar mutations or a genetically crosslinked di-α subunit. A mathematical model for hemoglobin tetramer assembly was developed, starting with the mechanism for apohemoglobin folding and adding heme binding steps for each of the different apoprotein states. The unfolding pathway for the heterodimeric apohemoglobin is a 4-step, 5-state mechanism. The first step involves unfolding of the heme pockets to form a heterodimeric molten globule intermediate. This intermediate dissociates into mostly unfolded monomers that then either interact transiently or undergo complete unfolding. Reversible hemin binding to the folded αβ apoHb dimer facilitates formation of the tetrameric α₁β₂ interfaces, promoting the final assembly of the HbA tetramer. Both the experimental studies and mathematical modeling of hemoglobin assembly provide the framework for understanding human hemoglobinopathies arising from globin misfolding and for enhancing the production yields of heme proteins in bacterial and eukaryotic expression systems.
ABBREVIATIONS

Apo-  without heme bound
CD    circular dichroism
Dw    dwarf sperm whale
DTT   dithiothreitol
ESI-TOF MS  electrospray ionization time-of-flight mass spectrometry
GdnHCl guanidine hydrochloride
Hb    hemoglobin
HbA   adult human hemoglobin A
HbF   fetal human hemoglobin
Holo- with heme bound
HPLC  high performance liquid chromatography
Mb    myoglobin
metHb oxidized hemoglobin with heme iron in the ferric form
rHb   recombinant hemoglobin
rHb0.0 recombinant hemoglobin with V1M mutations in subunits
rHb0.1 recombinant hemoglobin with V1M mutations in subunits
     and a single glycine linker joining the C-terminus of the
     first α chain and the N-terminus of the second α chain
Sw    sperm whale
Acknowledgments

I dedicate this thesis work to my late mother, Vanathi Chitra Devi Sundaram, who always wanted the best for me, never let me give up on getting a higher education, and continuously motivated me to seek excellence in my studies. I could never have reached this point in my life if not for the support, sacrifices and selflessness of both my late parents, and the culmination of my thesis work therefore should also rightly be a celebration of their memories.

I want to first thank both my thesis co-advisors, Dr. John S. Olson and Dr. George N. Phillips, for their mentorship and for showing me there is always something to discover and look forward to in science. I feel extremely grateful to Dr. Olson for agreeing to take me on as his last graduate student before his retirement, even though he had stopped taking new students for several years when I joined the program. Simultaneously, I want to thank Dr. Phillips for his enthusiasm for the tradition of collaboration between Olson and Phillips labs to continue by agreeing to be my co-advisor.

Sometimes, you just need someone to believe in you, and I am truly lucky to have met that person in Dr. Olson when I was at one of the lowest points of my life. When I joined the graduate program in August 2012, I was experiencing a lot of personal issues due to my mother’s recent death, and thus honestly I would not have been able to survive the intensity and high stress environment of graduate school if not for the constant guidance and support from Dr. Olson. Because he believed in my potential and valued my opinions in science, he gave me the gifts of courage and confidence to pursue my
goals. I have always enjoyed the scientific debates we had, and through his mentorship, I was able to evolve as a researcher over the years from someone who always second-guessed her own reasoning to someone who is able to voice and pursue her ideas confidently. I especially want to thank Dr. Olson for helping me to discover the wonderful world of mathematical modeling in biochemistry, and for showing me, sometimes, that basic algebra can offer the most straightforward answers to complex biological problems. Finally, I also have always appreciated the countless hours he has devoted throughout my PhD training to analyzing my data, and editing my manuscripts, presentations, and documents for my committee.

Similarly, I am thankful to Dr. Phillips for all the scientific and professional advice he has given me, for pushing me beyond my comfort zone, and for encouraging me to think out of the box in terms of my experiments. He also showed me that X-crystallography is not an esoteric field. He encouraged my interests in computational biology, and even sent me to a coding bootcamp, despite its non-relevance to my PhD dissertation work. He goes the extra mile for his students, and I will always appreciate his active role in my post-doctoral job search. He even took the time to drive me once to and from Galveston so that I can set up my experiments at the Sealy Center, UTMB.

I want to thank my PhD thesis committee members – Dr. George Bennett, Dr. Kathleen Matthews, Dr. Yizhi Tao, and Dr. Laura Segatori – for their helpful advice and feedback on my dissertation work. I want to thank Dr. Kathleen Matthews for writing numerous recommendations letters for my job applications, for editing my curriculum vitae, and for giving me helpful advice on the postdoctoral job search process. I also
appreciate the great conversations I have had with my committee chair, Dr. George Bennett, at the end of each meeting with my committee.

I want to convey my gratitude to of my two favorite women in science from Olson’s lab, Dr. Jayashree Soman and Eileen Singleton, for all their support and encouragement, and for having made Olson’s lab a great environment to work in. Jayashree has been an unofficial mentor to me, and I am grateful to her for her patience in training me in spectroscopy and experimental techniques in biophysics from my early days as a naïve researcher. Similarly, I am grateful to Eileen for teaching me how to express and purify globins, and also for her immense help in making a good portion of the proteins I used in my unfolding experiments.

I want to thank Dr. Mitch Miller (Phillips’ lab colleague) for teaching me a lot about X-ray crystallography, and patiently answering all my questions, while being understanding about my lack of background in the field. I also want to thank David Xu (Phillips’ lab colleague) both for being a fun person to work with and especially for growing the dwarf sperm whale Mb crystals, thus giving me my first taste in crystallography.

I want to thank Dr. Lu Han (Phillips’ lab colleague) for being a great friend and lab mate, and for being there for those not so good days of graduate school. I am also grateful to Dr. Andres Benitez Cardenas (Olson’s lab colleague) for fixing broken instruments that I needed for my experiments without complaining. I also want to thank my former undergraduate mentee, William Ou, for his collaboration and contribution to my research.
I want to thank Dr. Neil Varnado, a former member of Olson’s lab, and the following current and former members of Phillips lab – Dr. Fengbin Wang, Sarah Alvarado, Joey Olmos, Dr. Ragothaman Yennamalli, Jonathan Clinger, and Dr. Hongnan Cao, - for being supportive colleagues.

Finally, I am grateful to Berea College for having given me the opportunity pursue a tuition-free undergraduate education, and I am extremely thankful to the following professors from my undergraduate alma matter for their mentorship: Dr. Matthew Saderholm, Dr. Ron Rosen, Dr. Jay Baltisberger, Dr. Megan Hoffman, and Dr. Martin Viellette.

Finally, I want to extend my gratitude to friends, staff, and faculty from Biosciences at Rice for all their support during the trials and tribulations of graduate school as well as for being there for the good times as well. It has been a wonderful experience being a part of this academic community.
Table of Contents

List of Figures .................................................................................................................. xii
List of Tables .................................................................................................................... xvi

Chapter 1 .......................................................................................................................... 1
Introduction and background ......................................................................................... 1
  1.1 Significance of globin stability studies ................................................................. 1
  1.2 Specific aims ............................................................................................................ 3
  1.3 Structures of apo- and holo- sperm whale myoglobin ......................................... 4
  1.4 Structures of human adult hemoglobin A and fetal hemoglobin ....................... 6
  1.5 Folding, assembly, and expression of mammalian myoglobin in vivo and in vitro as a model system for the globin subunits ......................................................... 8
  1.6 Previous studies on the folding and assembly of human hemoglobin ............... 11
  1.7 Clinical relevance of hemoglobin folding ........................................................... 13
  1.8 The molten globule state of the globin folding intermediate .............................. 15

Chapter 2 .......................................................................................................................... 19
Materials and Methods .................................................................................................... 19
  2.1 Recombinant Mb and Hb plasmids for expression in E. coli ............................... 19
  2.2 In vivo expression and purification of recombinant myoglobins ....................... 22
  2.3 In vivo expression and purification of recombinant hemoglobins and isolation of native HbA .......................................................... 26
  2.5 Analytical gel filtration analysis .......................................................................... 30
  2.6 Preparation of oxy-, met-, and deoxy Mb and Hb samples .................................. 31
  2.7 Preparation of apomyoglobins (apoMb) .............................................................. 32
  2.8 Preparation of hemichrome standards ............................................................... 33
  2.9 Unfolding measurements of apo- and holo- myoglobins .................................... 33
  2.10 Measurement parameters for fluorescence and CD spectrometers ................ 37
  2.11 Hemin loss assays .............................................................................................. 38
  2.12 Small angle X-ray Scattering (SAXS) ............................................................... 39
  2.13 X-ray crystallography of dwarf sperm whale metMb ....................................... 40

Chapter 3 .......................................................................................................................... 43
Development of a cell-free protein expression assay to investigate factors governing expression of animal myoglobins ................................................................. 43
3.1 Introduction .......................................................................................................................... 43
3.2 In vivo E. coli expression studies .......................................................................................... 44
3.3 Cell-free expression studies ............................................................................................... 45
3.4 Construction of a library of Mb cell-free expression plasmids .......................................... 49
3.5 In vitro transcription and cell-free protein expression .......................................................... 50
3.6 Development of heme addition protocol for cell-expression of metMb .......................... 51
3.7 Quantitative RT-PCR measurements of mRNA levels transcribed in vitro ................. 56
3.8 Quantification of total Mb levels expressed in cell-free system with indirect ELISA .... 56
3.9 Small scale isolation and purification of soluble cell-free expressed Mb ....................... 57
3.10 Quantification of cell-free expressed, soluble holoMb .................................................... 60
3.11 Cell-free expression of soluble holoMbs is the highest for apolar distal heme pocket mutants and deep-diving variants ......................................................... 61
3.12 Lack of correlation between protein expression and transcript level ......................... 65
3.13 Total protein synthesis, expression of folded protein, and protein aggregation ... 67
3.14 Conclusions ..................................................................................................................... 68

Chapter 4 ........................................................................................................................... 70

Correlation between expression of soluble holomyoglobin and apomyoglobin stability .................................................................................................................. 70
4.1 Introduction .......................................................................................................................... 70
4.2 Structure of dwarf sperm whale (Dw) metMb .................................................................... 72
4.3 Measurement of apoglobin stabilities .............................................................................. 76
4.4 Measurements of hemin affinity and hemin dissociation rate constants ....................... 80
4.5 Effects of L137M mutation in recombinant dwarf sperm whale Mb .............................. 82
4.6 Direct correlations between holoMb expression and apoMb stability (logK_{UN}) but not with heme affinity ......................................................................................... 85
4.7 Mechanistic interpretation of cell-free expression ............................................................ 89
4.8 Conclusions ..................................................................................................................... 98

Chapter 5 ........................................................................................................................... 101

Human apoHbA unfolding models ....................................................................................... 101
5.1 Basic models for dimeric apoHbA unfolding .................................................................... 102
5.2 Initial optimization of reversible unfolding experiments with apoHb .................. 104
5.3 Non-native disulfide formation in apoHb samples ................................. 107
5.4 Final optimized conditions for reversible apoHb unfolding .................... 110
5.5 Statistical analysis methods comparing apoHbA unfolding models ............ 110
5.6 Fitting to a simple 2-step model ....................................................... 112
5.7 Statistical analyses of alternative 3-step models .................................. 116
5.9 Multi-model analysis of the residual distributions for the fits .................. 126
5.10 Conclusion ......................................................................................... 130

Chapter 6 .............................................................................................. 131

Verification of human hemoglobin unfolding model .................................... 131

6.1 Introduction ......................................................................................... 131
6.2 Heme pocket unfolding ........................................................................ 132
6.3 Unfolding of apo-rHb0.0 .................................................................... 137
6.4 Unfolding of apo-rHb0.1, a cross-linked Hb with a glycine linker between the α subunits .......................................................... 138
6.5 Unfolding of fetal apohemoglobin ....................................................... 147
6.6 Conclusions ......................................................................................... 150

Chapter 7 .............................................................................................. 154

Human met-holohemoglobin unfolding ....................................................... 154

7.1 Introduction ......................................................................................... 154
7.2 Tetramer-dimer (α₁β₂) interface interaction as a factor governing hemin binding 156
7.3 SAXS analysis of rHb0.1 .................................................................... 157
7.2 HoloHbA assembly model including hemin binding ................................ 159
7.3 HoloHb unfolding measurements ....................................................... 163
7.4 Spectral deconvolution ...................................................................... 165
7.5 Analysis of holoHbA unfolding curves ................................................. 165
7.6 Conclusions and future directions ....................................................... 176

Chapter 8 .............................................................................................. 181

Future Directions .................................................................................... 181

8.1 Early folding events during human Hb assembly .................................... 183
8.2 Expression and assembly studies of human embryonic Hb variants .......... 184
8.3 Understanding human blood disorders arising from inefficient globin expression or globin misfolding ........................................ 185
8.4 The unfolding pathways of the members of globin superfamily ................. 186
List of Figures

Figure 1.1. Holo- sperm whale myoglobin (PDB ID 1JP6) ....................................................... 5
Figure 1.2. Human Hemoglobin A (PDB ID: 2DN1) ................................................................. 6
Figure 1.3. Fetal Hemoglobin (PDB ID: 4MQJ) ........................................................................ 7
Figure 1.4. Six-state mechanism for holoMb unfolding modified from Culbertson and
Olson⁹ ........................................................................................................................................ 9
Figure 2.1 ESI-TOF MS spectra .............................................................................................. 23
Figure 2.2. pH gradient elution of HbF off the S Sepharose cation ....................................... 28
exchange column .................................................................................................................. 28
Figure 2.3. Reverse phase HPLC chromatograms of native HbA, rHb0.1, and distal heme
pocket rHb mutants .............................................................................................................. 30
Figure 2.4. Spectral properties of hemichromes generated from HbA ................................. 32
Figure 2.5. Protein crystal of dwarf sperm whale Mb .......................................................... 41
Figure 3.1. Cell-free screening assay for holoMb incorporating in vitro transcription and
cell-free expression (ENDEXT® technology) and small-scale protein purification .......... 47
Figure 3.2. Cell-free expression of goosebeak whale Mb ...................................................... 51
Figure 3.3. Analysis of cell-free expression of sperm whale holoMb with cyanohemin
addition .................................................................................................................................... 53
Figure 3.4. Soret absorbance (Abs) of semi-purified dwarf sperm whale Mb .................... 55
Figure 3.5. Binding of holoMb to Zn²⁺ resin during small-scale purification of cell-free
expressed Mbs ...................................................................................................................... 58
Figure 3.6. SDS-PAGE analysis of semi-purified cell-free expressed Mbs ......................... 59
Figure 3.7. Absorbance spectra of semi-purified holoMb isolated from cell-free translation reactions

Figure 3.8. Correlation between holoMb cell-free expression yield and amount of mRNA transcribed in vitro

Figure 4.1 A. Structure of dwarf sperm whale oxyMb (unpublished) determined in this thesis work. B Structural alignments of dwarf sperm whale Mb (mixture of O₂ and CO bound forms) (red helix), sperm whale MbO₂ (PDB #1A6M, magenta helices), and pig metMb (PDB #IPMB, cyan helices)

Figure 4.2. Electron density maps of the active site of Dwarf whale Mb

Figure 4.3. GdnHCl induced equilibrium unfolding of selected apoMbs

Figure 4.4. GdnHCl induced equilibrium unfolding of native, recombinant wild-type, and recombinant 137M Dw apo-Mbs followed by A, circular dichroism (CD)

Figure 4.5. GdnHCl induced equilibrium unfolding of selected holoMbs

Figure 4.6. Correlation between holoMb cell-free expression yield and A, overall apoMb stability (-log(K₇₁); B, the stability of the apoMb native (N) state relative to the intermediate (I) state ((-log(K₇₁)); C, the stability of the apoMb intermediate (I) state relative to the unfolded (U) state (-log(K₇₃)))

Figure 4.7. Correlation between holoMb cell-free expression yield and -log (K₇₁). K₇₁ was determined from the analyses of the holoMb unfolding experiments done with 10 μM protein in 10 mM potassium phosphate, pH 7 at 20°C.

Figure 4.8. Heterologous expression of holoMb and correlations with overall apoMb stability
Figure 4.9. A, HoloMb expression models for both unimolecular precipitation and bimolecular aggregation. B, Comparison of observed and fitted results for relative expression versus log(K_{UN}).

Figure 5.1. Basic models of equilibrium unfolding curves for ApoHbA heterodimer (D).

Figure 5.2. Unfolding of apoHbA observed with significant variability.

Figure 5.3. Analytical gel filtration column analysis of HbA and rHb0.1 without (A,B,C) and with (D,E,F) reducing agents.

Figure 5.4. Fits of equilibrium unfolding curves for apoHbA to the simple 3-state, 2-step model 1.

Figure 5.5. Fits of equilibrium unfolding curves for apoHbA to the 3-step model with U_M \rightarrow U_c.

Figure 5.6. Fits of equilibrium unfolding curves for apoHbA to the 3-step model with 2U_M \rightarrow U_2.

Figure 5.7. Complete 4-step, 5-state mechanism for human apoHb unfolding.

Figure 5.8. Fits of equilibrium unfolding curves for apoHbA to the complete 5-state, 4-step model.

Figure 5.9. Raw residual distributions for fits of apoHbA unfolding curves.

Figure 5.10. Standardized residual distributions for fits of apoHbA unfolding data.

Figure 6.1. GdnHCl induced equilibrium unfolding of apoHbA and distal heme pocket apo-rHb mutants followed by CD changes.

Figure 6.2. Recombinant Hb 0.1 (rHb0.1) (PDB ID #1O1L).
Figure 6.4. GdnHCl induced equilibrium unfolding of apo-rHb0.1 as a factor protein concentration 143

Figure 6.5. Defining human apoHb unfolding transitions through mutagenesis. 145

Figure 6.6. GdnHCl induced equilibrium unfolding of apoHbF as function of total protein concentration 148

Figure 6.7. Comparison between GdnHCl induced equilibrium unfolding curves for apoHbA and apoHbF 149

Figure 7.1 SAXS analysis of apo- and holo-rHb0.1 159

Figure 7.2 HoloHbA assembly model with cooperative heme binding 160

Figure 7.3 Deconvolution of visible absorbance spectral measurements of holoHbA unfolding 166

Figure 7.4 Normalized CD unfolding curves for metHbA 167

Figure 7.5. Comparison of fluorescence unfolding curves for metHbA with total predicted population fractions of partially and completely unfolded states 170

Figure 7.6 Predicted population fractions of heme variants during holoHbA unfolding 173

Figure 7.7. Population fractions of holo- tetrameric and dimeric metHbA predicted from fittings of metHbA unfolding measurements 175

Figure 7.8 Comparison between apo- and holo HbA and HbF 177

Figure 7.9 HoloHb unfolding of HbA, α(H58L/V62F)β(wt) rHb and rHb0.1. 179

Figure 8.1 Assembly and misfolding of human hemoglobin A 182

Figure 8.2 Human neuroglobin (PDB #ID 1OJ6) 188
List of Tables

Table 3.1. Summary of Mb cell-free expression yields. ______________________________________ 62
Table 3.2. ELISA measurements of total Mb synthesized in the cell-free expression system.  ____________________________________________________________ 68
Table 4.1 Structure determination parameters of the dwarf sperm whale Mb __________ 73
Table 4.2. Summary of Apomyoglobin stabilities, hemin dissociation equilibrium constants, and hemin dissociation rates constants _______________________________________________ 79
Table 4.3. Effects of the L137M mutation on apomyoglobin stabilities, and heme affinities  ____________________________________________________________ 84
Table 5.1. Fitting statistics for apoHbA unfolding models ________________________________ 116
Table 5.2. Fitted parameters for apoHbA unfolding models ________________________________ 119
Table 6.1. Fitted equilibrium unfolding parameters for apoHbA, apoHbF, and other rHb variants, using the 5-state, 4-step model for ApoHb ________________________________ 135
Chapter 1

Introduction and background

Several sections of this chapter are adapted and reproduced from my following two publications:


In order to avoid redundancy, I have minimized further references to these publications within the text body.

1.1 Significance of globin stability studies

Animal hemoglobins and myoglobins belong to the globin family, and the classical three-on-three alpha helical fold characterizes most of the globin subfamilies\(^1,2\). The function of mammalian myoglobins is to provide O\(_2\) storage protein for release during times of muscle contraction when limited blood flow occurs\(^3\). The major function of animal hemoglobins encapsulated in erythrocytes circulating in the cardiovascular system is oxygen transport from lungs to respiring tissues\(^4\). Both *in vitro* and theoretical investigations of the globin folding,
assembly, and expression pathway are fundamental to achieving a quantitative understanding of the factors that govern hemoglobin and myoglobin expression levels and biological half-lives. In addition, many human hemoglobinopathies arise from detrimental mutations in the genes for hemoglobin subunits, which cause tetramer dissociation, heme loss, globin unfolding, and globin precipitation as well as changes in $O_2$ affinity. The occurrence of myoglobinopathies have not been found in literature, but many of the problems associated with rhabdomyolysis resulting from the breakdown of muscle fibers arise from the oxidative degradation of myoglobin.

A quantitative investigation of globin stability is also required for rationally designing more robust acellular recombinant hemoglobins that could potentially be used as oxygen carriers in transfusion medicine and in tissue engineering applications. The structure and $O_2$ binding characteristics of mammalian myoglobins (Mb) are similar to those of the individual hemoglobin (Hb) subunits and have been more fully characterized biophysically. The first part of my thesis research explored the relevance of apoprotein folding and hemin binding to expression of the holoprotein in vivo and in cell free translation systems using myoglobin as simple model system. The mechanisms for apohemoglobin folding and holohemoglobin tetramer assembly were then determined, analyzed, and compared to those for the simpler myoglobin monomer. Despite the ~150 year history of biochemical experiments, no one has published a quantitative description of either apo- or holohemoglobin folding and assembly, starting from the isolated $\alpha$ and $\beta$ subunits. Thus, any successful quantitative theory for describing these processes represents a significant advance in the hemoglobin field, even if some of the conclusions seem obvious in retrospect.
1.2 Specific aims

My first goals were to: (1) develop and implement a cell-free expression assay in order to investigate myoglobin assembly during expression from mRNA; and then (2) derive hemoglobin assembly and folding mechanisms, with the eventual goal of understanding the complex assembly of human hemoglobin in vivo during erythropoiesis.

The wheat-germ based cell-free translation system allows the decoupling of cellular homeostasis with protein expression, thereby reducing variability in protein expression measurements\(^8\). For this study, a cell-free holomyoglobin expression assay was successfully developed (Chapter 3) and then used to demonstrate a strong, positive, and quantitative correlation between in vitro expression levels of fully folded holoMb variants and the corresponding apoMb unfolding parameters. These parameters were measured independently with a small library of purified protein variants (Chapter 5). The experimental results were further supported by new mechanistic analyses, which show that very high apoglobin stability is required for a high yield and efficiency of holoMb expression in order to overcome the rapid aggregation and precipitation of unfolded polypeptides coming off the ribosomes.

Apo- and holomyoglobin unfolding curves for a new library of variants were measured using GdnHCl as a denaturant and circular dichroism to evaluate secondary structure. Tryptophan fluorescence emission was used to determine when heme was lost, and visible absorbance spectra were measured to look for folding intermediates containing bound heme. Hemin affinities for the various folding states during Mb assembly were determined by simultaneously measuring and analyzing apo- and holoprotein unfolding curves\(^9\).

The second half of my thesis involves the application of these approaches to determining the human hemoglobin assembly mechanism, which is much more complex due to
hetero-dimer and tetramer formation from the individual α and β monomers. I derived a mathematical model for the mechanisms of apoHb assembly, which can describe quantitatively unfolding curves as function of total protein concentrations and the dependence of curves on subunit crosslinking and heme pocket mutations. The apoHb assembly model and the independently determined apoHb unfolding parameters were then incorporated into an overall holoHb assembly mechanism, using both empirical and analytical analyses. The relationship between heme binding and tetramer formation for Hb was also investigated by gel filtration and small angle X-ray scattering (SAXS).

1.3 Structures of apo- and holo- sperm whale myoglobin

The first three-dimensional protein structure solved by X-ray crystallography was that for sperm whale holomyoglobin as reported by Kendrew et al. in 1959. Sperm whale Mb is a single domain globin, consists of 8 alpha helical segments, which are labeled A to H (Fig. 1.1). Mb contains iron protoporphyrin IX which can adopt 3 different oxidation states: ferrous (heme or Fe(II)-protoporphyrin IX), ferric (hemin or Fe(III)-protoporphyrin IX), and ferryl (Fe(IV)=O protoporphyrin IX). When the heme is bound, the iron atom is covalently coordinated to the proximal histidine, which is the 93rd residue of Mb at the F8 helical position. This F8 histidine is a conserved residue across the globin family. This amino acid is a part of a functional cluster of 6 conserved residues (others include CD1, E11, FG5, F4, and G5) which directly affect globin heme affinity and axial ligand binding properties'. On the distal side of heme ring, the iron coordinates a variety of ligands including O₂, CO, NO, and alkyl isocyanides. When the iron is oxidized to the ferric state, water, NO, imidazole, and various anions including cyanide and azide can be bound. These ligands are stabilized through
favorable electrostatic interactions (including hydrogen bonding) with the distal histidine (His64 at the E7 helical position).\textsuperscript{12-14}

Figure 1.1. **Holo- sperm whale myoglobin (PDB ID 1JP6)**

Based on circular dichroism measurements, mammalian Mbs show a \textasciitilde{}20-30\% loss of helical content when heme is removed to form the apoprotein\textsuperscript{9}. When Eliezer and Wright\textsuperscript{15} characterized sperm whale apoMb through high resolution NMR methods, they observed increased disorder localized around the proximal side of the heme pocket at the EF loop, F helix, FG loop, and the beginning of G helix\textsuperscript{16}. Unfortunately, no crystals of apoMb have been
grown, which is probably due to intrinsic disorder in the F-helical region of the apoglobin. As a result, there has been no structure determined by X-ray crystallography.

1.4 Structures of human adult hemoglobin A and fetal hemoglobin

Figure 1.2. Human Hemoglobin A (PDB ID: 2DN1)

In 1968 Max Perutz published the first high resolution, three-dimensional structure of human hemoglobin (Hb) by X-ray crystallography [7,18]. Roughly 95% of the proteins in human erythrocytes are adult hemoglobin (HbA) (Fig. 1.2). HbA is composed of α and β globin chains, with each chain having 7-8 alpha helical segments with a myoglobin-like fold and
heme-binding pocket \(^{4,16,19}\). These subunits assemble first into heterodimers by forming an \(\alpha_i\beta_i\) interface with extensive inter-subunit contacts involving 34 different amino acid side chains. These dimers then associate into a compact globular tetrameric structure by forming two new \(\alpha_i\beta_2\) and \(\alpha_2\beta_i\) interfaces\(^{16,20,21}\). Tetrameric human fetal hemoglobin, HbF, is composed of \(\alpha_i\gamma_1\) dimers (Fig. 1.3)\(^{16,22}\). The \(\alpha\) subunits in both HbA and HbF are identical, but the \(\gamma\) subunits differ from the beta subunits at 39 different amino acid positions\(^{23,24}\). However, HbF is structurally homologous to HbA, with only slight structural differences at the N-terminals between the \(\gamma\) and \(\beta\) chains (PDB ID: 4MQJ)\(^{25}\).

Figure 1.3. **Fetal Hemoglobin (PDB ID: 4MQJ)**
1.5 Folding, assembly, and expression of mammalian myoglobin in vivo and in vitro as a model system for the globin subunits

Culbertson and Olson⁹ developed a six-state model (Fig. 1.4) to describe quantitatively the unfolding of holo-metMb and dissociation of hemin in vitro based their experimental measurements of GdnHCl-induced unfolding curves for a series of holo- and apoMb variants.¹³ The apoMb unfolding curves were analyzed according to the mechanism initially developed by
Figure 1.4. **Six-state mechanism for holoMb unfolding modified from Culbertson and Olson**. The three states of apoMb are native (N) with most of the helices folded, molten intermediate (I) with the heme pocket mostly unfolded, and then the completely unfolded (U) polypeptide chain. HoloMb states containing bound heme (H) are native (NH), intermediate (IH) with the heme pocket melted and a hemichrome structure, and unfolded (UH) with heme bound non-specifically.

Barrick, Baldwin, Wright, and their colleagues. ApoMb unfolds in vitro via the generation of a molten globule intermediate (I), which retains ~40% of the helical content of the native apo-structure. As described in their 6-state model, Culbertson and Olson showed that hemin can bind reversibly to the different apoglobin folding states. The binding of hemin to the I state leads to the generation of a reversible hemichrome (IH). However, the affinity of the I state for hemin is ~100 fold lower than that of the native (N) state. The hemichrome is formed when hemin iron atom axially coordinates with an endogenous amino acid side chain at its 6th coordination site creating a hexacoordinate metal complex. When the heme iron is in the reduced Fe$^{2+}$ state, this hexacoordinate species is called a hemochrome. In the absence of GdnHCl, globin unfolding can lead to irreversible aggregation of the unfolded globin states and hemin (UH, U). Free hemin released from the unfolded globin is highly toxic in vivo and can partition into membranes and promote lipid oxidation and generation of reactive oxygen species.

Solution X-ray scattering studies at small and high angles performed on various folding states of horse myoglobin showed that the native folded holoMb state has the smallest radius of gyration (17.5 Å) and is the most compact relative to the various states of the corresponding apoprotein. Native apoMb has a slightly larger radius of gyration (19.7 Å) than holoMb but still retains as a globular shape. However, the molten globule I state has a more expanded
structure with a larger radius of gyration (~23 Å°), whereas the unfolded apoglobin exhibits an
even larger radius of gyration of ~30-35 Å°.38

Scott et al.39 showed that apomyoglobins from deep diving whales are 10 to 500 times
more resistant to unfolding induced by GdnHCl than apomyoglobins originating from
terrestrial animals. They suggested that the globins most likely evolved to have increased
resistance to denaturation in order to adapt to the acidic conditions that occur in whale muscle
during prolonged dives when lactic acid accumulates due to lack of oxygen. Scott et al.39 also
noted anecdotally that the enhanced apomyoglobin stability of deep diving whales would help
explain the high expression yields of sperm whale holoMb in E. coli relative to that of pig and
human holoMbs. In contrast to the apoproteins, all mammalian Mbs appear to have similar
thermal stabilities in the holo-metMb state40,41. Hargrove et al.41 had also suggested in an earlier
study that heterologous expression of mammalian Mb in E. coli is governed more by globin
stability than by hemin affinity. However neither Hargrove et al.41 nor Scott et al.39 examined
the differences in expression of the mammalian Mbs in a quantitative and systematic manner13.

In a recent bioinformatics study, Mirceta et al.42 reported that the deep diving mammal
Mbs appear to have evolved to have higher net surface charge, which they suggested allowed
much higher expression of the protein in the muscles of these animals. For example, the level
of Mb in the skeletal muscles of sperm whale is roughly 70 mg/g of wet tissue whereas the
amount in pig muscle is only 2 to 4 mg/g of wet tissue, a greater than 10-fold difference42.
They noted that MbO₂ is the primary source of oxygen for swimming during deep dives when
blood circulation is diverted from the skeletal muscles to keep the heart and brain of the diving
animal well oxygenated. Mirceta et al.42 also reported that Mbs in aquatic mammals had, on
average, a surface charge, Zₘb, equal to +2.5 to +4.8 at neutral pH, whereas Zₘb for most Mbs
from terrestrial mammals was much smaller and closer to zero. They suggested that the net higher positive charge prevented aggregation and crystallization of myoglobin at the high protein concentrations required for O\textsubscript{2} storage in the muscles of diving animals. However, based on the previous measurements by Scott et al.\textsuperscript{39}, an alternative explanation is that this larger protein surface charge may promote higher apoglobin stability\textsuperscript{13}.

**1.6 Previous studies on the folding and assembly of human hemoglobin**

Despite the high structural homology with monomeric mammalian myoglobins, hemoglobin subunits are much less stable in both their apo- and holo globin forms as monomers. There have been reports that suggest heme binds to the nascent \(\alpha\) and \(\beta\) chains as they are being synthesized on ribosomes in order to promote and stabilize globin folding\textsuperscript{16,43,44}. However, there is no direct experimental evidence for initial heme binding to unfolded monomers before the subunits associate. Previous studies have shown that partially folded apoHbA monomers are stabilized markedly by oligomerization to form \(\alpha\beta\) heterodimers\textsuperscript{16,44,45}. Similar stabilization occurs when partially folded apo-\(\alpha\) chains bind to the alpha hemoglobin stabilizing protein (AHSP), and when apo-\(\beta\) chains self-assemble into homo-dimers and tetramers\textsuperscript{16,46-48}. The extraction of heme from HbA results in complete dissociation into apo-\(\alpha\),\(\beta\) dimers with little or no trace of either monomeric or tetrameric species at neutral pH (see Chapter 5)\textsuperscript{16,44,49,50}. Because free apoglobin heterodimers were found in erythrocytes, Winterhalter et al.\textsuperscript{51} suggested that apo-heterodimers are likely to be the initial post-ribosomal species during the Hb folding and assembly\textsuperscript{16,50}. 
When isolated $\alpha$ and $\beta$ apo-subunits are reconstituted with hemin, the monomeric subunits form unstable hemichrome states in which it appears that both the HisE7 and HisF8 side chains are coordinated to the iron atom. These forms are unstable and prone to precipitation. Hemichrome formation is also observed when native metHbA is dissociated into monomer by the addition of excess p-hydroxymercuribenzoate to react with all the cysteine side chains. The addition of dithionite to these hemichromes causes rapid reduction of the iron atoms in the individual subunits to form hemochromes, but following 1-hour dialysis under N$_2$, the heme groups become pentacoordinate and the spectra of native deoxygenated subunits are restored$^{52,53}$. These observations suggest that the ferric forms of the isolated subunits are disordered with partially unfolded heme pockets, i.e., a molten globule state containing a hexacoordinate hemin group$^{52,53}$.

By analogy with the apoMb structure, which was determined through high resolution NMR spectroscopy$^{15}$, we assume that the enhanced disorder in human apoHb dimers also occurs mainly around the F helix region and the FG loop, both regions that are critical for $\alpha_1\beta_2$ interface interactions and tetramer formation. However, no one has examined quantitatively either apo- or holohemoglobin unfolding under reversible conditions.

Past $in$ $vitro$ studies of human holoHb unfolding and denaturation did not attempt to determine in a quantitative manner the importance of heme affinity, apoglobin stability, and equilibrium subunit association constants, nor were any molten globule intermediates identified$^{16,54-63}$. For example, thermal denaturation curves for holoHb were usually fitted to 2-state models (folded and unfolded, and sometimes an irreversible aggregated state), and partially unfolded transition states were not clearly defined$^{16,54,57,59}$. These Hb unfolding studies were not done under either equilibrium or reducing conditions and seemed to have led to non-
native high order oligomers as a result of disulfide bond formation and irreversible hemichrome formation. These artifacts trap Hbs in non-native folding states, which eventually lead to globin denaturation, and their degree of occurrence is difficult to predict in a defined manner and seems to vary with sample age and oxidative conditions.\textsuperscript{16,60,61,64,65}.

Previous studies on cooperative ligand binding to adult human hemoglobin have established the critical role of the proximal histidine and F helix in the quaternary changes that occur at the $\alpha_1\beta_2$ interface.\textsuperscript{66,67} Thus, in both Mb and Hb, the binding of heme and its coordination with the proximal histidine (F8) "rescues" these disordered regions, and, in the case of hemoglobin, leads to formation of the $\alpha_1\beta_2$ interface and tetramers by interactions near the FG loops and F and C helices.\textsuperscript{16}

1.7 Clinical relevance of hemoglobin folding

Disruption of normal HbA folding and assembly \textit{in vivo} can lead to the accumulation of Heinz bodies containing insoluble patches of denatured hemoglobin and free hemin, which are deposited on the erythrocyte membranes leading to lysis or removal of the damaged red cells by the spleen. These symptoms are described as hemolytic anemias, which can arise from mutations in the hemoglobin genes that lead to unstable proteins, unequal expression of the $\alpha$ and $\beta$ globins, or loss of the ability of red cells to rapidly re-reduce metHbA after auto- or chemically-induced oxidation. These disruptions in red blood cells in terms of oxidation states of the Hbs can arise from the loss of normal reduction mechanisms (e.g., loss of glucose-6-phosphate dehydrogenase activity, cytochrome b$_5$, or the flavoprotein reductases involved in reducing metHbA) in either new or aged red blood cells. These conditions, called collectively methemoglobinemia, lead to the accumulation of oxidized hemoglobin, which can initially
unfold into reversible hemichromes, lose hemin, which is itself toxic, and generate highly unstable apoglobins. The initial hemichrome-containing globins can also become irreversibly modified and have propensity to cross-link through disulfide bonds.\(^{60,61,64,65,68}\)

Globin mutations disrupting subunit assembly, globin folding, and heme binding can also manifest clinically as Heinz body anemias. Some of these hemoglobinopathies are due to mutations that may help to define more specific apohemoglobin folding pathways. \(\beta\mathrm{G112R}\) (Hb Indianapolis) and \(\alpha\mathrm{G104C}\) (Hb Sallanches) mutations occur along the \(\alpha_1\beta_1\) interface, disrupting the apohb dimer formation, and are clinically manifested as microcytic anemia, which is due to inefficient Hb expression.\(^{69,70}\) Although the rate of \(\beta\mathrm{Indianapolis}\) synthesis is similar to \(\beta\) wild-type, radiolabeled mutant peptides have an abnormally short half-time leading to a phenotype of severe \(\beta\)-thalassemia. All these results suggest strongly that \(\alpha_1\beta_1\) apoheterodimer formation is required for not only apohb stability, but also for the successful expression of the final holoprotein.\(^{70}\) In my study, I have looked quantitatively at the role of the \(\alpha_1\beta_1\) interface and genetic crosslinking in regulating hemoglobin folding.

Previously, it has been shown that the valine to alanine mutation at the E11 position decreases the stability of the distal heme in apoMb\(^{41,71}\) due to the increase in empty space in the heme pocket. The corresponding mutation in the \(\beta\) subunits (Hb Sydney) of human hemoglobin leads to a hemolytic anemia phenotype with the Heinz bodies populated by irreversible hemichromes containing equal amounts of precipitated \(\alpha\) and \(\beta\) subunits.\(^{60,61,72}\) In this case, the apohb dimer intermediates are forming but folding of the heme pocket is inhibited, leading to denaturation and accumulation of globin precipitates. In my thesis work, I looked more carefully at how increasing the stability of the heme pocket alters the folding and
assembly of HbA, and then compared the results to the well-established mechanism for Mb folding and assembly.

1.8 The molten globule state of the globin folding intermediate

As described above, the intermediate (I) folding state for sperm whale apoMb has been characterized by NMR and consists of an unfolded heme pocket and a folded hydrophobic core of A (N-termini), G, and H (C-termini) helices\textsuperscript{13,16,73,74}. A bioinformatics study by Ptitsyn and Ting determined the existence of 6 conserved residues in the AGH nucleus across various globins of diverging protein sequence including 81 vertebrate Mbs, 259 $\alpha$ Hb subunits, 237 $\beta$ Hb subunits, and 24 $\gamma$ Hb subunits. Their conclusions lend credence to the role of AGH helices as the globin folding nucleus\textsuperscript{1}. In early 2000s, Scott et al.\textsuperscript{39} had examined the reversible unfolding curves of 13 different mammalian apomyoglobins from both terrestrial and aquatic origins. Their experimental results show that a 2-step folding pathway involving a molten globule intermediate\textsuperscript{26-29} is conserved across all the apomyoglobins examined despite their diverging protein sequences. Thus, the folding and assembly studies in this thesis almost certainly apply to all mammalian (and perhaps all vertebrate) myoglobins and hemoglobins.

Culbertson et al.\textsuperscript{9,30,31} had later demonstrated that hemin binds reversibly to the molten globule intermediate of sperm whale apomyoglobin through bis-histidyl hexacoordination, forming a hemichrome. The other key question is whether hemichrome formation by reversible hemin binding to the molten globule is preserved across the globins, and this question is addressed in both Chapters 4 and 7 in the contexts of myoglobins from different species as well as human hemoglobin.
Many studies have observed reversible hemichrome formation during partial denaturation or sometimes in frozen samples of hemoglobins, suggesting that heme pocket dynamics are a contributing factor. However, these hemichromes are not a significant sub-population of native human hemoglobin at room temperature\textsuperscript{72,75-77}. When freeze-quenching experiments were performed on metHbA, a sub-population of bis-histidine hemichromes present at room temperature was identified by rapid-freezing EPR methods in the immediately frozen samples. Following incubation reheating from 210 to 260 K, a subtle reconfiguration of the heme pocket appeared to occur with the increase of thermal energy, leading to conversion of various populations present in metHbA to mostly another form of bis-histidine hemichrome that is associated with complete absence of water in the heme pocket\textsuperscript{75}.

Pressure-induced unfolding is driven by the presence of cavities in the folded state of a protein\textsuperscript{78}. In addition to the heme pocket cavity, various apolar cavities have been identified through xenon binding sites in myoglobin and hemoglobin\textsuperscript{79-81}. Proteins are unfolded by pressure when the molar volume of their unfolded peptides is smaller relative to that of their native folded states\textsuperscript{78}. Both met- and deoxyhemoglobin show formation of hemichrome and hemichrome, respectively at high pressure as a result of the heme pocket collapse. However, pressure-induced structural changes show low reversibility in hemoglobin. The precipitates formed following high-pressure-induced unfolding of oxy- and deoxyHb could not be reconstituted to the folded state\textsuperscript{82,83}.

Dehydration has also been shown to cause unfolding of hemoglobin, again leading to the formation of hemichromes and hemochromes\textsuperscript{32,84-86}. The complete loss of water content in met-HbA has been shown to lead to about ~30\% loss of helical structure, as well as hemichrome formation. Interestingly, dehydration seems to be a biphasic process, with an
initial reversible hemichrome forming at a 30% loss of water content with little change in secondary structure. Peaks at 412 nm, 536 nm, and 565 nm characterize the visible absorbance spectrum of this intermediate. The second phase of dehydration involves a ~30% loss of helical structure coupled with red shift of the hemichrome spectral peaks (from 412 nm to 416 nm and from 536 nm to 540 nm). Thus during dehydration, hemoglobin unfolding and hemichrome formation are partially uncoupled in contrast to what is seen during GdnHCl titrations. Initial water extraction from the heme pocket promotes hemichrome formation with little heme pocket collapse. The reconstitution of commercially available lyophilized hemoglobins into functional protein in solution reflects the reversibility of this first dehydration process.

These previous studies and the work in this dissertation show that hemichromes form due to either initial unfolding of the heme pocket or heme pocket reconfiguration. Relative to human hemoglobin, sperm whale myoglobin has a lower occurrence of the intermediate hemichrome state, a feature that has been ascribed to the greater rigidity of the myoglobin heme pocket.

As in myoglobin, the Nε atom of the distal histidine in hemoglobin can coordinate with the iron to form hemichrome (Fe(III)) or hemochrome (Fe(II)) complexes. However, the Nε atom is located more than 4 Å from the iron in the folded heme pocket of native hemoglobin subunits, preventing coordination without unfolding. Hemichromes can also occur by axial coordination of the hemin iron with the sulfur atom in methionine and cysteine side chains or the phenoxy atom of the side chain of tyrosine. The closest methionine to the iron atom in HbA is at α32, the conserved 13th residue on the B helix. In the native met-α structure, this methionine sulfur is more than 8Å away from the iron. The only cysteines that are free to coordinate with the iron in HbA are either located at the α1β2 interface or on the proximal side...
of the heme pocket at the $\beta 93$ position$^{32}$. All of these structural observations indicate that hemichromes can only occur in partially unfolded or unfolded states of hemoglobin, but none of the work in literature has quantitatively correlated reversible molten globule formation with the presence of hemichrome or hemochrome spectral species. My thesis work is the first attempt in literature to show that molten globule formation also occurs reversibly during Hb assembly and that it is associated with the appearance reversible hemichrome intermediates.
Chapter 2

Materials and Methods

*Materials and methods described in this section are for previously established protocols, in the lab or in literature, that were used in this study. Protocols or experimental techniques that were newly established or modified through my thesis work are described later in the relevant chapters in this thesis.*

Several sections of this chapter are also adapted and reproduced from my following two publications as a first author:


In order to avoid redundancy, I have minimized or omitted further references to these publications within the text body.

2.1 Recombinant Mb and Hb plasmids for expression in *E. coli*

The original wild-type (synthetic gene) recombinant sperm whale (*Physeter catodon*) (Sw) Mb contains N122 instead of the native D122 residue, based on the DNA originally
This original synthetic Sw Mb gene has served as the genetic background for all of Sw Mb mutants used in my cell-free assays, Smith’s *E. coli* expression assays, as well as most other *in vitro* unfolding studies, kinetic measurements, and structure determinations done in both the Olson and Phillips groups. Previous studies showed that the differences between the D122 and N122 variants, when they occur, are very small. However, as a control for my expression-based studies discussed in Chapter 3, I did construct the D122 Sw Mb variant (see Fig. 3.1, mRNA gel) and saw insignificant differences between the *in vitro* expression yields of the D122 Sw Mb and that of wild-type Sw Mb containing N122. Therefore the N122 variant was kept as the control Sw Mb gene for my thesis work in order to be consistent with previous studies in our laboratories.

During the large-scale expression of all the whale and seal Mbs in *E. coli*, no removal of the N-Methionine occurs, whereas the initiator methionine is removed in native muscle Mbs (Fig. 2.1A). For example, in Fig. 2.1A, ESI-TOF mass spectroscopy (ESI-TOF MS) done on purified recombinant dwarf sperm whale Mb (MW=17,368.2 Da with N-Met) expressed in *E. coli*. The results show that the N-Methione is not cleaved during expression in *E. coli*. If the N-methionine had been cleaved the apo- or holo-Mb peaks (A1 and A2 in Fig. 2.1A) should have been smaller by 131.2 Da. Because the N-Met is removed in native mammalian Mbs, the numbering of amino acids starts with N-Val or N-Gly as residue 1. To be consistent with this number system, the N-Met in recombinant Mbs is listed as 0. The original pig Mb gene was a gift from Dr. Anthony Wilkinson (University of York, York, UK). Both the human Mb gene and some of the protein variants were kind gifts from Dr. Masao Ikeda-Saito.

For my work, a number of Mb genes were synthetically constructed through Integrated DNA Technologies (IDT) based on the protein sequences obtained from the UniProtKB/Swiss-
Prot database: dwarf sperm whale (Kogia simus - Q0KIY5.3) (Dw) Mb, emperor penguin
(Aptenodytes forsteri-P02199) Mb, goosebeak whale (Ziphus cavirostris - P02182.2) Mb and
grey seal (Halichoerus grypus - P68081) Mb. The codons in these genes were optimized by
IDT for expression for E. coli expression to improve globin production yields.

The dwarf sperm whale, grey seal, goosebeak whale, pig, emperor penguin, and sperm
whale Mb genes were cloned into the pVP80K vector through the Polymerase Incomplete
Primer Extension (PIPE) cloning method. Sequence-specific ribosome binding sites were
designed in the Mb-pVP80K expression plasmids utilizing the Salis Lab Calculator
(https://salislab.net/software/) to optimize further the globin expression yields in E. coli. Myoglobin mutants were constructed in these vectors using PIPE site-directed mutagenesis. The sperm whale Mb H64F/V68F mutant gene was already in the pEMBL19+ vector and had been constructed previously through cassette mutagenesis. The original pVP80K-HSBC vector was a kind gift from the Center of Eukaryotic Structural Genomics (University of Wisconsin-Madison, WI).

Recombinant fetal human hemoglobin (HbF) was expressed from the pHE9 plasmid according to protocols developed by Shen et al. α H58L/V62F and β H63L/V67F apolar heme pocket mutations were introduced by site-directed mutagenesis of human α and β genes in the pHE2 vector. The original pHE2 and pHE9 plasmids were gifts from Dr. Chien Ho’s laboratory (Carnegie Mellon University). The globins expressed from the pHE2 and pHE9 plasmids have an "extra" Met residue initially at their N-terminus. This initiator Met is later cleaved from the globin chains by E. coli Met aminopeptidase, which is co-expressed at the same level as the α and β globins from the same plasmid.
Recombinant crosslinked human hemoglobin, rHb0.1, was expressed from the pSGE1.1-E4 expression plasmid, whereas the uncrosslinked control, rHb0.0, was expressed from the pDL111-13e expression plasmid. Both of these plasmids were gifts from Somatogen Inc. (later Baxter Hemoglobin Therapeutics) to the Olson laboratory in the late 1990s. Somatogen also provided the SGE1661 E. coli cells for rHb0.0 and rHb0.1 expression.96-98 rHb0.0 and rHb0.1 were engineered with V1M mutations to include an initiator Met, which is not processed during expression in E. coli. The rHb0.1 variant was genetically crosslinked with a glycine linker between Arg-141 on the C-terminus of the first α chain (with the V1M mutation) and the Val-1 N-terminus of the second alpha chain.16,96-98

2.2 In vivo expression and purification of recombinant myoglobins

For my apo- and holoMb unfolding experiments, almost of the protein samples were obtained using the new pVP80K-Mb vectors. The variant plasmid was transformed into E. coli BL21-DE3 cells (Agilent), which were grown in 1 L Terrific Broth medium with 50 μg/ml Kanamycin for expression in 2 L flasks. Following induction with 0.8 mM isopropyl-β-D-1-thiogalactopyranoside, the cells were grown at 25 °C-28 °C post lag phase. For the sperm whale Mb H64F/V68F mutant gene in the older pEMBL19+ vector, the holoMb was expressed in TB1 cells.13,71 External heme was not added during expression of any of the Mb variants.

All the new and older Mb variants were isolated and purified according to the method of Springer and Sligar as modified by Carver et al. 99 The harvested cells were mechanically and enzymatically lysed in an AVESTIN cell breaker, and the clarified supernatant was subjected to ammonium sulfate protein precipitation. The precipitated protein was then re-suspended and dialyzed in 20 mM Tris pH 8, 1 mM EDTA pH 8. The protein was then
sequentially purified by initial elution through a DEAE Sepharose Fast Flow (GE Healthcare) column equilibrated with the dialysis buffer and then elution with a pH gradient from a cation exchange CM52 cellulose (Whatman) column equilibrated initially with 20 mM sodium phosphate, pH 6. Depending on protein purity, a further salt gradient elution was performed using a S Sepharose cation exchange column (GE Healthcare). Protein purity was determined from the ratio of the heme Soret absorbance of the MbCO sample at 420 nm to the absorbance of the protein at 280 nm and by SDS-PAGE electrophoresis. In some cases, ESI-TOF mass spectroscopy (Fig. 2.1A) and reverse phase HPLC analyses were also performed. The purified Mbs were stored in the carbon monoxide (CO) bound, ferrous form. All buffers during protein isolation and purification were kept equilibrated with 1 atm of pure CO gas. The concentration of the CO bound form was determined spectroscopically using $\varepsilon_{424\text{nm}} = 187\text{mM}^{-1}\text{cm}^{-1}$. $^{100}$

Figure 2.1 ESI-TOF MS spectra for A) recombinant dwarf sperm whale Mb; and B) recombinant $\alpha$H58L/V62F $\beta$H63L/V67F Hb. Panels A and B are shown on the next two pages. The spectra shows that N-Methionine is present in recombinant dwarf sperm whale Mb expressed in E.coli, but for the $\alpha$H58L/V62F $\beta$H63L/V67F rHb variant co-expressed with E. coli Met aminopeptidase (MAP) through the pHE2 E.coli expression plasmid, the N-Methionine residues are cleaved post-translationally by the MAP protein.
A1. Apo Dw Mb (MW= 17368.2 Da)

A2. Dw Mb with 1 hemin (MW= 616.5 Da) bound

A3. Dw Mb with 1 CN-hemin (MW= 642.5 Da) bound

A4. Dw Mb with 2 hemin bound

A. Recombinant dwarf sperm whale (Dw) Mb with initiator Met
B. $\alpha$(H58L/V62F) $\beta$ (H63L/V67F) rHb

B1. $\alpha$(H58L/V62F) subunit (15150.4 Da)

B2. $\beta$ (H63L/V67F) subunit (15891.3 Da)
2.3 In vivo expression and purification of recombinant hemoglobins and isolation of native HbA

Native holoHbA was isolated from lysed erythrocytes obtained from expired blood units from the Gulf Coast Regional Blood Bank, Houston, TX. The HbA in the lysate was then purified using purification protocols developed to purify recombinant hemoglobins (rHbs) by Looker et al. and Shen et al. 95,98.

Recombinant HbF was expressed from the pHE9 E. coli expression plasmid 94, and the \( \alpha \) H58L/V62F and \( \beta \) H63L/V67F apolar heme pocket mutations were introduced through site-directed mutagenesis into the recombinant human adult hemoglobin gene and expressed from the pHE2 expression plasmid 94,95. The pHE2 and pHE9 vectors were gifts from Dr. Chien Ho (Carnegie Mellon University). Expression of rHbs in E. coli requires the presence of an initiator Metionine at the N-termini of the globin chains, whereas the final N-terminal residue in native human Hb \( \alpha \) and \( \beta \) chains is a valine and in \( \gamma \) chains is a Glycine. In the pHE systems developed by Chien Ho and coworkers, the \( \alpha \) and \( \beta \) globin genes are expressed with an "extra" Met residue at their N-termini. This initiator Met is cleaved post-translationally from the \( \alpha \) and \( \beta \) polypeptides by E. coli Met aminopeptidase (MAP), which is co-expressed with the globin genes 94,95. For example, the ESI-TOF MS spectra shown in Fig. 2.1B for the purified \( \alpha \) H58L/V62F \( \beta \) H63L/V67F rHb variant, expressed with the pHE2 expression plasmid, demonstrates the N-Met residues have been cleaved by MAP because both the B1 peak for apo \( \alpha \) H58L/V62F subunit and the B2 peak for apo \( \beta \) H63L/V67F subunit in Fig. 2.1B would have had 131.2 Da higher molecular weight if N-Methionine were present.
As described previously, rHb0.1 was expressed from the pSGE1.1-E4 expression plasmid whereas rHb0.0 was expressed from the pDL111-13e expression plasmid\textsuperscript{96-98}. rHb0.0 and rHb0.1 were engineered with V1M mutations, and the Met1 amino acid is retained in the final proteins. rHb0.1 was genetically crosslinked with a glycine linker between Arg-141 on the C-terminus of the first $\alpha$ chain (with the V1M mutation) and the Val-1 N-terminus of the second $\alpha$ chain. Thus, this protein is the di-$\alpha_2\beta_2$ version of rHb0.0\textsuperscript{96-98}. Crosslinking of the $\alpha$ chains produces a highly stable tetramer, which does not dissociate into monomers except in the presence of high concentrations of denaturant (see Chapter 5). In contrast, the CO and ferric forms rHb0.0 and HbA dissociate into $\alpha_1\beta_1$ dimers in the micromolar concentration range.

pHE2 and pE9 expression plasmids were transformed into \textit{E. coli} JM109 cells (Promega). The expression of rHbF was performed using the protocols developed by Shen et al.\textsuperscript{94,95} but modified for \textit{E. coli} cell growth in 500 ml flasks. The pSGE1.1-E4 and pDL111-13e expression plasmids were transformed into SGE1661 E.coli cells, a modified variant of JM109 cells, provided by Somatogen, Inc (Baxter Hemoglobin Therapeutics). Cells transformed for expression of rHb0.0, rHb0.1, and rHb heme pocket mutants were grown in a Biostat C 20 bioreactor (B Braun Biotech International, Melsungen, Germany), using cell growth and protein expression protocols described initially by Looker et al.\textsuperscript{98} and modified by Maillett\textsuperscript{101}. Purification the various rHbs was done following the protocols of Maillett\textsuperscript{101}, Looker et al.\textsuperscript{98}, and Shen et al.\textsuperscript{94,95}. The key steps involved first loading clarified cell lysates onto a Zn$^{2+}$-charged Chelating Sepharose Fast Flow resin column (GE Healthcare) and then onto a Q Sepharose anion exchange column (GE Healthcare). rHbs expressed from the pHE2 and pHE9 plasmids were subjected to a final purification step on S Sepharose cation exchange column
(GE Healthcare) using a pH gradient (6.6 to 8.3) in 5-10 mM sodium phosphate, 0.5 mM EDTA to isolate N-Met-free rHb fractions (Fig. 2.2). ESI-TOF mass spectroscopy (Fig. 2.1B) and reverse phase HPLC (Fig. 2.3) analyses confirmed both the identity and purity of all the Hb variants and native HbA\textsuperscript{94,95}. Isoelectric focusing and SDS-PAGE gel electrophoresis were also done to ensure the quality of the samples. Hbs were kept under CO conditions during protein isolation and purification, and purified protein concentrations were determined using the Soret absorbance peak for HbCO at 419 nm, using $\varepsilon_{419} = 191$ mM$^{-1}$cm$^{-1}\textsuperscript{109}$.

Figure 2.2. pH gradient elution of HbF off the S Sepharose cation exchange column (GE Healthcare). Fractions were collected from the tallest peak only.
2.4 Reverse phase HPLC analysis

Reverse phase HPLC was done with an Agilent 1220 Infinity LC system. An initial protocol was first described by Hoffman et al.\textsuperscript{96} and Shelton et al.\textsuperscript{102}, and then optimized in our lab by Neil Varnado. A final 350 \( \mu \text{M} \) protein solution was solubilized in HPLC grade water containing 0.5 \% TFA. The protein sample was then spun down at >10,000 \( g \) for 5 minutes to separate out any precipitated heme. Then, a 20 \( \mu \text{l} \) protein sample was injected onto a reverse phase (Grace Altech Vyduc 214TP C4) column, consisting of N-capped butyl aliphatic groups polymerized onto a silica column with 300 \( \text{Å} \) pore diameter. The Vyduc 214TP C4 column was recommended for separation of hemoglobin polypeptides and variants. The reverse phase was equilibrated with mobile phase of 0.1\% TFA in 35\% acetonitrile (HPLC grade). Protein samples were eluted at room temperature with an increasing gradient of acetonitrile from 35\% to 47\% at 1 ml/min flow rate, while a high pump pressure was maintained below 100 bar. The silica column can withstand a pressure up to 355 bar, and no collapse was ever observed. After each sample run, the reverse phase column was washed with 100\% acetonitrile.

Examples of the quality of the heme pocket rHb mutant and rHb0.1 samples are shown in the reverse-phase chromatographs in Fig. 2.3. The increase in hydrophobicity of the heme pocket \( \alpha \) and \( \beta \) mutants and the increase in the molecular weight of the di-\( \alpha \) subunit of rHb0.1 are apparent from the shifts to longer elution times for the individual subunit peaks in the \( \alpha \)(H58L/V62F), \( \beta \)(H63L/V67F), and di-\( \alpha \) samples in the reverse phase HPLC chromatograms shown.
Figure 2.3. **Reverse phase HPLC chromatograms of native HbA, rHb0.1, and distal heme pocket rHb mutants.** Protein samples at a concentration of 310 μM were each loaded onto HPLC column equilibrated with 0.1% TFA in 35% acetonitrile. The polypeptide chains were eluted with an increasing gradient of acetonitrile from 35% to 47% at 1ml/min flow rate.

### 2.5 Analytical gel filtration analysis

For analysis of their oligomeric state, globin samples were loaded onto a Superose-12 HR 10/30 GL column, with a bed volume of 24 ml, which was attached to an AKTA FPLC. The column was equilibrated with 200 mM potassium phosphate, pH 7 at 4°C, and the protein samples were eluted from the column at a flow rate of 0.5 ml/min. The final concentration of
eluted protein was calculated using a dilution factor computed as the ratio of the average elution peak width at the half-height to the volume of the sample that was loaded on the column. Protein sample sizes of either 100 µl or 500 µl with a total amount ≤ 10 mg were loaded onto the column. Elution profiles are shown in Chapter 5, Fig. 5.3 for various apoHb samples.

2.6 Preparation of oxy-, met-, and deoxy Mb and Hb samples

Carbon monoxide (CO) bound to the hologlobin was replaced with dioxygen (O₂) by continuous flushing of protein solutions for ~ 30 minutes with 1 atm of pure O₂ under an intense visible light to enhance photodissociation of CO. During this process, the sample was continuously rotated in a round bottom glass vial, which was immersed in ice water to prevent protein degradation due to the heat generated by the light.

For preparation of ferric samples, HbO₂ or MbO₂ samples were rapidly oxidized with potassium ferricyanide (K₃[Fe(CN)₆])³⁻⁴. The metHb or metMb sample was centrifuged to remove any degraded material and precipitated free heme. Then, the sample was gel-filtered on a Sephadex G-25 (GE Healthcare) column to separate the ferric hologlobin from excess oxidizing agent. Protein samples were eluted from the column in 10 mM potassium phosphate, pH 7 at 4 °C.

DeoxyHbs were prepared by placing the protein sample in a glass vial fitted with a stopcock at one end. The vial was then alternatively degassed under vacuum and filled with nitrogen gas until the protein sample completely turned purple with visible absorbance peaks at 430 nm and 555 nm indicating deoxyHb formation.¹⁰⁰
2.7 Preparation of apomyoglobin (apoMb)

ApoMb was obtained by hydrolysis of the heme Fe(III)-HisF8 bond at low pH and then extraction into 2-butanone (methylethylketone). The pH of the metMb solutions was lowered to ~ 2.2-2.5 with 0.1% HCl and then selective heme extraction into the organic phase was carried out in chilled separatory flask. ApoMb remained in the lower aqueous phase. After removal of the 2-butanone, the aqueous phase was dialyzed overnight in cold 10 mM potassium phosphate at pH 7, and then centrifuged to remove precipitates. ApoMb concentration was determined spectroscopically using $\varepsilon_{280} = 15.2 \text{ mM}^{-1} \text{ cm}^{-1}$.

Figure 2.4. Spectral properties of hemichromes generated from HbA
2.8 Preparation of hemichrome standards

Six-coordinate hemin complexes (hemichromes) are generated during the reversible unfolding of most hologlobins. I needed to obtain standard spectra for these species in order to determine the fractional contribution of these species in observed sets visible spectra recorded during unfolding experiments. One approach was to generate reversible hemichromes by adding excess imidazole to metMb or metHb samples. For example, when 12 µM metHb was prepared in a solution of 200 µM imidazole, a reference hemichrome absorbance spectra was obtained with peaks in the visible region at 413 nm, 535 nm, and 565 nm (Fig. 2.4, blue spectrum). Similar hemichrome spectra are obtained by added excess imidazole to metMbs. Alternatively, when 12 µM metHb is prepared in 600 µM SDS to partially denature the protein, a hemichrome spectra is observed with absorbance peaks that overlay with those obtained after imidazole addition (Fig. 2.4 red spectrum).

2.9 Unfolding measurements of apo- and holo- myoglobins

Measurements of GdnHCl induced unfolding of apo- and holoMbs were performed in 10 mM potassium phosphate pH 7 at 20°C for comparison with previous work by Culberston and Olson. The individual samples were incubated in a water bath at different concentrations of GdnHCl (Sigma-Aldrich) ranging from 0 M to 4 M at 20 °C for 2 hours to achieve equilibrium. The fractional decrease in helical content was measured as the loss of negative ellipticity at 222 nm during unfolding using a Jasco J-810 CD spectropolarimeter. Changes in Mb fluorescence emission originating from Trp residues were measured at 341 nm for apoMb unfolding and at 355 nm for holoMb unfolding using a Varian Cary Eclipse.
spectrofluorometer. In both cases, fluorescence excitation was at 285 nm. The changes in metMb Soret peak absorbance due to hemin dissociation during holoprotein unfolding were measured using a Cary 100 Bio UV-Visible spectrophotometer.

The CD and fluorescence equilibrium unfolding data for apoMb were fit simultaneously to the 3-state, 2-step mechanism shown in the right hand column in Fig. 1.4 as described by Culbertson and Olson and Samuel et al. These fits were done in order to determine the apoMb folding constants, which are described in Fig. 1.4. CD, visible absorbance, and fluorescence equilibrium unfolding data for holoMb were then fit simultaneously together with the previously determined apoMb folding constants to the complete 6-state mechanism shown in Fig. 1.4. Fits to the holoMb unfolding data enabled the determination of the heme binding equilibrium constants to the different apoMb folding states as described in Fig. 1.4. The different apo- and holo- folding states are as described in Fig. 1.4 and section 1.5. Each folding state is assigned an intrinsic spectroscopic signal (S), dependent on the spectral measurements done (CD, fluorescence or Soret absorbance). These intrinsic signals for the different folding states are defined by their accompanying subscripts. For example, \( S_N \) for apoMb is either the CD or the fluorescence signal for the native (N) state. The CD signal for the N state and the fluorescence signal for the U state were initially normalized to 1. The intrinsic spectroscopic signals for the unfolded and intermediate apoMb folding states, \( S_I \) and \( S_U \) respectively, are assigned relative to the \( S_N \) value for the CD measurements and relative to the \( S_U \) value for the fluorescence measurements. The dependence of the overall CD or fluorescence or the Soret absorbance signal (S) change on [GdnHCl] is defined in eq 2.1 and eq 2.2 for apo- and holo- Mb unfolding measurements respectively.
Fitting of apoMb unfolding data to eq 2.1 was done with the Solver program in MS Excel 2011 to obtain the unfolding equilibrium constants, $K_{NI}$ and $K_{IU}$, extrapolated at $[\text{GdnHCl}] = 0$. The I state signal, $S_I$, the CD signal for U, and fluorescence signal for N were allowed to vary during fitting. The differential GdnHCl binding parameters $m_{NI}$ and $m_{IU}$ were set as 9.85 kJ mol$^{-1}$ M$^{-1}$ and 5.68 kJ mol$^{-1}$ M$^{-1}$ based on the work of Culbertson and Olson$^9$ and define the linear dependence of free energies on GdnHCl concentration for the N to I and I to U transitions respectively. $R$ and $T$ in eq 2.1 are gas constant (8.314 J mol$^{-1}$ K$^{-1}$) and temperature in Kelvins respectively.

$$S = \frac{S_N + S_I K_{NI} \exp\left(m_{NI} \frac{[\text{GdnHCl}]}{RT}\right) + S_U K_{NI} K_{IU} \exp\left((m_{NI} + m_{IU}) \frac{[\text{GdnHCl}]}{RT}\right)}{1 + K_{NI} \exp\left(m_{NI} \frac{[\text{GdnHCl}]}{RT}\right) + K_{NI} K_{IU} \exp\left((m_{NI} + m_{IU}) \frac{[\text{GdnHCl}]}{RT}\right)}$$  

(eq 2.1)

$$S = Y_{NH} S_{NH} + Y_N S_N + Y_{IH} S_{IH} + Y_I S_I + Y_U S_U + Y_{UH} S_{UH} + Y_{IH} S_{IH}$$  

(eq 2.2)

The hemin disassociation constants at $[\text{GdnHCl}] = 0$, $K_{NH}$ and $K_{IH}$, and the independently determined apoMb unfolding constants at $[\text{GdnHCl}] = 0$, $K_{NI}$ and $K_{IU}$, were used to derive the population fractions ($Y$) of the different unfolding states in eq 2.2 as defined by Culbertson and Olson$^9$ and described in eqs 2.3-2.4. The population fraction of a specific folding state is defined by its accompanying subscript. For example, $Y_N$ represents the population fraction of the native N state of apoMb. Data fitting of holoMb unfolding measurements to eq 2.2 was done in the Gnuplot program$^{110}$ to obtain the the hemin disassociation constants at $[\text{GdnHCl}] = 0$, $K_{NH}$ and $K_{IH}$. In eq 2.2, $S_{NH}$ was initially normalized to 1, while $S_N$, $S_I$, and $S_U$ for the CD signal measurements were renormalized relative to $S_{NH}$, consistent with apoMb helical content relative to that of holoMb. $S_H$ represents Soret absorbance for free hemin. The $K_{UH}$ value at $[\text{GdnHCl}] = 0$ was fixed at $1 \times 10^{-6}$ M based on
previous studies which showed that the non-specific binding of hemin to unfolded globins is roughly independent of amino acid sequence\textsuperscript{109,111}. The dependence of the free energies on [GdnHCl] for hemin disassociation (K\textsubscript{d}) from the hemin bound native (NH), intermediate (IH), and unfolded (UH) states are defined as m\textsubscript{NH}, m\textsubscript{IH}, and m\textsubscript{UH}. m\textsubscript{UH} was fixed to 10 kJ mol\textsuperscript{-1}M\textsuperscript{-1}, and for wild type Mbs, the m\textsubscript{NH} and m\textsubscript{IH} values were initially set at 18.42 kJ mol\textsuperscript{-1}M\textsuperscript{-1} and 16.75 kJ mol\textsuperscript{-1}M\textsuperscript{-1} respectively based on Culbertson and Olson's analyses\textsuperscript{9}.

\[
\frac{[N][H]}{[NH]} = K_{NH} \exp\left(\frac{m_{NH}[\text{GdnHCl}]}{RT}\right) \quad \text{eq 2.3}
\]

\[
\frac{[I][H]}{[IH]} = K_{IH} \exp\left(\frac{m_{IH}[\text{GdnHCl}]}{RT}\right)
\]

\[
\frac{[U][H]}{[UH]} = K_{UH} \exp\left(\frac{m_{UH}[\text{GdnHCl}]}{RT}\right)
\]

\[
\frac{[I]}{[N]} = K_{NI} \exp\left(\frac{m_{NI}[\text{GdnHCl}]}{RT}\right)
\]

\[
\frac{[U]}{[I]} = K_{IU} \exp\left(\frac{m_{IU}[\text{GdnHCl}]}{RT}\right)
\]

\[
Y_{NH} = \frac{[NH]}{[NH]+[IH]+[UH]+[N]+[I]+[U]} \quad \text{eq 2.4}
\]

\[
Y_{IH} = \frac{[IH]}{[NH]+[IH]+[UH]+[N]+[I]+[U]}
\]

\[
Y_{UH} = \frac{[UH]}{[NH]+[IH]+[UH]+[N]+[I]+[U]}
\]

\[
Y_{N} = \frac{[N]}{[NH]+[IH]+[UH]+[N]+[I]+[U]}
\]

\[
Y_{I} = \frac{[I]}{[NH]+[IH]+[UH]+[N]+[I]+[U]}
\]

\[
Y_{U} = \frac{[U]}{[NH]+[IH]+[UH]+[N]+[I]+[U]}
\]

\[
Y_{H} = \frac{[N]+[I]+[U]}{[NH]+[IH]+[UH]+[N]+[I]+[U]}
\]
For all the reversible unfolding experiments induced by denaturants, the equilibrium constants can be described in terms of the dependence on denaturant concentration and temperature, as shown in eq 2.3. However, these constants are not derived in terms of the dependence on the experimental buffer’s final salt concentration.

For the heme pocket mutants, the $m_{NH}$ and $m_{IH}$ values were initially set at 14.65 kJ mol$^{-1}$M$^{-1}$ and 13.82 kJ mol$^{-1}$M$^{-1}$, respectively, again based on previous work 9. When allowed to vary, the $m$ constants did not change much. The $K_{NI}$ and $K_{IU}$ values were fixed to the values obtained for unfolding of the corresponding apoMb species. The CD signal for the native state of the holoMb ($CD_{NH}$) was renormalized and fixed to 1 and the CD signals for the apoMb N, I, and U states were fixed at values from the apoMb unfolding studies with respect to native state of holoMb. The Soret absorbance of the native state of holo-metMb at 409 nm was also normalized and fixed to 1. The other signals at the different folding states were allowed to vary.

2.10 Measurement parameters for fluorescence and CD spectrometers

Fluorescence spectra were measured using a Varian Cary Eclipse spectrofluorometer. The excitation and emission slits were set at 10 nm and 2.5 nm, respectively, and data were collected for 310 nm-380 nm at a scan rate of 120 nm/min with 1 nm data interval.

CD spectra were collected using a Jasco J-810 CD spectropolarimeter. The data integration time was set to 1 sec, and data pitch, which signifies the data collection intervals, was set to 0.1 nm. Scanning speed for data collection was set to 100 nm/min, and a continuous mode was used for recording three consecutive spectra, which were then averaged for each sample. The spectral bandwidth, which assumes Gaussian distribution of light intensity
centered at the nominal wavelength\(^{112}\), was set to 1 nm. The sensitivity settings of the photomultiplier (PMT) can be adjusted depending on the concentration of protein sample, with low sensitivity being set for high globin sample concentrations with negative ellipticity beyond 200 mdeg at 222nm to avoid spectral saturation. The high tension (HT) voltage setting for the photomultiplier is automatically set to inversely correlate with the amount of transmitted light hitting the detector. Measurements at a HT voltage around \(\sim 700-800\) V are considered to be suspect because too little light is being transmitted to obtain accurate CD signals\(^{113,114}\). Therefore, special 0.1 to 0.01 mm path length flow-through cells were required for the higher protein concentrations maintain reasonable transmittance and high voltage settings.

### 2.11 Hemin loss assays

Rates of hemin dissociation from metMb and metHb variants were measured by reacting the ferric holoprotein with excess H64Y/V68F apoMb, which serves as a hemin scavenging agent and assumes a green coloration when it takes up hemin\(^{104,115}\). The decrease of Soret absorbance peak for the ferric hologlobin variants was monitored over time at 37°C. For the metMb reactions, 60 \(\mu\)M of sperm whale H64Y/V68F apoMb reagent was added to 6 \(\mu\)M of the variant being tested. For the metHb reactions, the concentration of the hemin scavenging agent was lowered to roughly 20 \(\mu\)M because of the much higher affinity of apoMb for hemin than apoHb. Reaction conditions for both the metMb and metHb reactions were 0.45 M sucrose in 0.15 M buffer that was either sodium acetate at pH 5 or potassium phosphate at pH 7\(^{104,109,115}\). Sucrose was added as an osmolyte to inhibit apoglobin precipitation after hemin dissociation, as described in Hargrove et al.\(^{104}\).
2.12 Small angle X-ray Scattering (SAXS)

Small angle X-ray scattering experiments were carried out with both apo- and holo-rHb0.1 to see if the apoprotein opened up and became a more linear molecule of linked α₁β₁ dimers rather than a compact spherical tetramer. Samples were initially determined to be monodisperse and free of higher order non-native oligomers by analytical gel filtration. Apo-rHb0.1 and holo-rHb0.1 samples were buffer exchanged into 10 mM potassium phosphate in the presence of 5 mM dithiothreitol (DTT). The buffers were degassed under a slight vacuum in order to reduce bubble formation, which can cause errors during SAXS measurements. Sample/buffer matching is a critical step because SAXS analysis is determined by the difference of the electron density between the protein and bulk solvent. DTT was added to reduce the radiation damage of the protein by acting as a scavenger of the free hydroxyl radicals that form due the x-ray absorption by the bulk solvent. Samples were then kept under nitrogen gas to reduce oxidation of DTT until measurements were taken.

Data collection was performed in collaboration with Professor Mark White with the Rigaku Bio-SAXS 1000 instrument, equipped with a 96-well automatic sample changer and a Rigaku BioSAXS-1000 2D-Kratky camera at the Sealy Center for Structural Biology and Molecular Biophysics. Samples were prepared at varying concentrations below 10 mg/ml in 70 μl volumes. Samples were initially screened at shorter exposure times of x-ray radiation to determine globin concentrations where transient inter-particle interactions or protein aggregation may occur. These particulates can disrupt the scattering intensities at lower q values. Data at longer exposures (16 hour) were then collected for rHb0.1 samples at concentrations between 1 mg/ml to 4 mg/ml, which were determined from the previous screening to provide good quality scattering curves. Scattering measurements were collected.
on blank buffers and protein samples alternatively to enable buffer subtraction for each protein scattering curve\textsuperscript{116,118}. SAXSLab software was used for data collection and processing of the images, and the ATSAS software suite package was used for data analysis as described following.

The scattering intensity, $I$, is defined in term of the scattering vector, $q = 4\sin\theta/\lambda$, where $\theta$ is the scattering angle and $\lambda$ is the source wavelength. PRIMUS was used for the initial processing of the scattering curves. The radius of gyration ($R_g$) was derived from Guinier plots defined by the equation, $\ln(I(q)) = \ln(I(0)) - q^2R_g^2/3$. The pair distance distribution function $P(r)$ between electrons in the sample was calculated in GNOM by applying indirect Fourier transform on the scattering curve as determined by the following equation,

$$P(r) = r(2\pi^2)^{-1} - \int_0^\infty I(q)q \sin(qr) dq .$$

CRYSOL was used to obtain theoretical scattering curves from deposited coordinates for the rHb0.1 crystal structure (PDB ID#1O1L), and then comparisons of the experimental data to these curves were performed. Rigid body modeling of rHbs based on the experimental scattering curves was done in CORAL with linkers placed between the alpha subunits for rHb 0.1\textsuperscript{117,119}.

### 2.13 X-ray crystallography of dwarf sperm whale metMb

High-throughput crystallization screening was done with the liquid handler mosquito\textsuperscript{®} Crystal (TTP Labtech). Crystals of wild-type dwarf sperm whale metMb (Fig. 2.5)
Figure 2.5. **Protein crystal of dwarf sperm whale Mb**

were obtained by David Xu in the screen using the sitting drop vapor diffusion technique. The well solution was 0.1 M sodium HEPES pH 7.5, 1 M sodium acetate trihydrate, 0.05 M cadmium sulfate 8/3-hydrate (Molecular Dimensions). A volume of 300 nl of dwarf sperm Mb in 5 mM potassium phosphate, pH 6.5 was mixed with 300 nl of the well solution in the sitting drop. The crystal was placed in a small amount of 35% sucrose as cryoprotectant condition, taken up with a loop, and X-ray diffraction experiments were done at 77 K using a GM/CA beamline at the Advanced Photon Source (Argonne National Laboratory). Data collection was done with the assistance of Sarah Alvarado using the JBlulce-EPICS software. The source X-ray wavelength was 1.0332 Å and the reflection spots were recorded using a PILATUS detector. 130 frames were collected with a φ angle increase of 0.7 between successive frames. Initial processing of the diffraction images (indexing, integration, scaling) was done by the beamline scientists at the APS laboratory. X-ray Detector Software (XDS)\(^{120}\) was used for initial indexing of the reflections in order to determine the unit cell parameters of the protein crystal, followed by measurements of the reflection intensities through integration and scaling. POINTLESS (CCP4i package) was then used to convert the xds file format of the reflections, which is the output file from XDS, to mtz format and also for further accurate confirmation of
the protein crystal space group. AIMLESS (CCP4i package) was then used for further data scaling\textsuperscript{123}.

The PHENIX\textsuperscript{122} software suite was used to solve and refine the structure of dwarf sperm whale metMb. Initially, phenix.reflection\_file\_editor was used to generate R-free flags in the reflections file by setting aside 5\% of the reflections, which were not included in the protein structure model building and refinement. This data set was later used to assess statistically and in an unbiased manner the accuracy of the protein structure prediction based on the R-free value, which is used as a statistical measure for model quality\textsuperscript{123}. The PHENIX PHASER\textsuperscript{124} package was then used to solve the structure of the dwarf sperm Mb through molecular replacement based the reference coordinates in the PDB file 2MBW for the recombinant sperm whale metMb. Dwarf sperm whale Mb shares 96\% protein sequence identity with the sperm whale Mb. AnoDe\textsuperscript{125} package, which is part of the SHELX suite, was used to determine anomalous electron density peaks corresponding to Cd, which was present in the crystallization buffer. The dwarf sperm whale Mb structure was further refined with the phenix.refine\textsuperscript{122,126} package. eLBOW\textsuperscript{122} (PHENIX package) was used to impose restraints on the heme molecules. COOT was used for manual fitting of the protein model to the electron density difference maps during refinement\textsuperscript{127}. Final stages of protein model refinement were done in REFMAC5 in order to make bulk solvent corrections\textsuperscript{128}. The latter analysis was done to address the observed higher electron density of the solvent due to the presence of cadmium (Cd) in the crystallization buffer.
Chapter 3

**Development of a cell-free protein expression assay to investigate factors governing expression of animal myoglobins**

This chapter is adapted and reproduced from my following publication:


In order to avoid redundancy, I have minimized and/or omitted further references to this publication within the text body.

### 3.1 Introduction

Previous studies have shown that the expression levels of myoglobin in the muscle cells of deep diving mammals and birds are significantly higher than in muscle cells of terrestrial mammals and birds\(^{42,129-132}\). However, very few studies have attempted to address systematically the biophysical properties that govern the levels of myoglobins in muscle cells across the animal kingdom as well as its heterologous expression in bacteria. This chapter summarizes my work on building a wheat germ-based cell-free globin expression assay to
probe the factors that govern holomyoglobin yields. The cell-free assay was built to allow for a more quantitative analysis while controlling inconsistencies and overlapping expression variables often encountered in myocytes and bacterial cells. The recombinant Mbs investigated in this study include variants reconstructed from Mb protein sequences of terrestrial and deep-diving aquatic mammals as well as distal heme pocket mutants with apolar substitutions designed to enhance apoglobin stability but decrease hemin affinity.

3.2 *In vivo* E. coli expression studies

This cell-free study was developed based on initial observations by Lucian P. Smith in his PhD thesis work. Smith surveyed the expression levels of recombinant holoMb in *E. coli* of a large library of ~250 site-directed and randomly generated heme pocket mutants of sperm whale Mb. All the variant genes were cloned into the same or closely related expression vectors (pUC19 or pEMBL19), and the goal was to examine compromises between apoglobin expression, stability, and functional O₂ binding. Mb expression in the bacterial suspensions was quantified based on amplitude of the Soret absorbance peak of holoMbCO complexes, and globin stability was correlated with the logarithm of folding equilibrium constants measured with purified apoproteins. Interestingly, Smith observed that mutants with large, apolar residues in the distal portion of the heme pocket were the highest expressers. Smith also showed that there appears to be a roughly linear correlation between the overall folding constant (1/Kₕₜ or Kₕₛ, where U and N are unfolded and native folded states, respectively) and the relative expression yield in the *E. coli*. However, the scatter in the protein expression data was substantial, and there were clear outliers.
The noise in the *in vivo* expression data could have come from a variety of factors, including variability in the bacterial growth phase, differential susceptibility of the variants to proteolysis, and alterations in light scattering by *E. coli* suspensions. In Smith's expression studies, holoMbCO was not purified from lysates, and absolute spectra were not measured. HoloMb expression in Smith’s studies was dependent on the synthesis endogenous heme in *E. coli*. In many cases, workers in the heme protein field add exogenous hemin or δ-aminolevulonic acid at the same time as induction of globin genes in *E. coli* to enhance yields of the holoprotein and overcome any limitation due to cofactor availability.

Other variables that must be taken into account when quantifying globin expression in *E. coli* include translation initiation rate related to the ribosome binding site sequence, plasmid copy number, codon usage bias in the mRNA sequence, the rate of endogenous heme synthesis, and metabolic carbon and nitrogen sources. Another complicating factor in *E. coli* is that the expressed holoMb is kept reduced in the bacterial cytoplasm and, in some cases, binds endogenously produced CO, which could greatly stabilize the protein, particularly those variants in which the distal histidine is replaced with Phe or Leu. A complex analysis would be needed to correlate all these variables with levels of holoemb expression and in most cases it is difficult to either control or quantify some of these variables. Thus, although *E. coli* expression is cost-effective for producing large amount of recombinant globins, this *in vivo* bacterial system is not ideal for correlating key biochemical properties of the myoglobins and hemoglobins with expressions levels of the holoprotein, particularly in mammalian cells.

### 3.3 Cell-free expression studies

Eukaryotic and prokaryotic cell extracts maintain the functionality of their protein
synthesis machinery despite losing their cell integrity\textsuperscript{135}. Olson’s lab had previously collaborated with Weiss’ group to use a cell free translation system to examine the role of alpha-hemoglobin-stabilizing protein (AHSP) in stabilizing $\alpha$ subunits of HbA and in facilitating folding of the $\alpha$ and $\beta$ subunits into intact HbA tetramers\textsuperscript{136}. In my work, the wheat germ based \textit{CellFree Sciences (ENDEXT\textsuperscript{®} technology) in vitro} protein synthesis system\textsuperscript{135,137} was adapted (Fig. 3.1) to test and examine quantitatively the various features of holoMb expression that were suggested from previous, more qualitative or semi-quantitative observations of Smith\textsuperscript{71}, Scott et al.\textsuperscript{39}, and Mirceta et al.\textsuperscript{42}.

The decoupled \textit{in vitro} transcription and wheat germ-based translation system allows control over the amounts of DNA, mRNA, amino acids, ATP, and hemin present\textsuperscript{135}. Most proteases have also been eliminated\textsuperscript{137}. In addition, the soluble holoMb product can be quickly separated from the translation mixture and be partially purified, and its spectral properties in the ferric state can be measured quantitatively.
Figure 3.1. **Cell-free screening assay for holoMb incorporating in vitro transcription and cell-free expression (ENDEXT® technology) and small-scale protein purification.** DNase-RNase free pEU-Mb DNA template was used to transcribe mRNA *in vitro*. The gel electrophoresis image at the upper right corner shows Dw, pig, Sw N122, and Sw D122 Mb mRNAs transcribed without any degradation, where Dw and Sw refer to dwarf sperm whale and sperm whale Mbs, respectively. The Mb mRNA transcribed was then incubated with the wheat germ extract in a bilayer cell-free translation reaction for 20 hours at 15°C. The cell-free expressed holoMb was purified on a Zn$^{2+}$ chelating resin mini-column using a well-filter plate, and 6 µl of the purified sample was loaded on a SDS-PAGE gel. The SDS-PAGE gel image at the bottom left corner shows the successful cell-free expression of Dw, pig, and Sw holoMbs, confirmed by the sharp bands at 17kDa, with the less stable pig Mb showing a less intense band. The 4th and 5th lanes in the gel represent wheat germ extracts (We) in which no mRNA was added for the translation incubation. Hemin was added to the wheat germ extract (We) loaded in the 4th lane whereas none was added for the sample in the 5th lane. The supernatants for the control extracts were also run through the Zn$^{2+}$ chelating resin mini-column filter plates following translation incubation. The SDS-PAGE gels verify that no significant amount of protein with a MW of 17 kDa was expressed and purified when Mb mRNA was omitted.
The bilayer reaction method (Fig. 3.1) for protein expression is marketed by the Cell Free Science Co. Ltd and can easily be set up over 96 well U bottom plates, facilitating high throughput screening. The reaction involves overlaying the feeding buffer (amino acids, ATPs, GTPs, creatine phosphate, and other cofactors) on top of the wheat germ extract containing the mRNA being translated and the required factors and ribosomes. Diffusion of the translation factors, polypeptides being translated, and unwanted inhibitors occur at the phase separation between the two layers. This method generates only micrograms amount of protein due to decreasing diffusion rates over time and increasing protein concentration. However, we have shown that the amount of protein generated is sufficient for carrying out a cell free expression assay.

The pEU vector incorporating SP6 phage-specific promoter, and TMV (tobacco mosaic virus) gene sequences was designed specifically for in vitro transcription to be coupled with this wheat germ based cell-free expression. An in vivo transcribed eukaryotic mRNA will have a 5’ end capped region followed by an untranslated region (UTR) for enhancing and regulating translation, and a poly(A) tail at the 3’ end. An in vitro generated mRNA would normally lack these factors, and therefore the TMV sequences were included in the pEU vector to flank the open reading frame of the gene to be expressed in order to compensate for these absences. The in vitro transcribed mRNA from the pEU plasmid therefore will have the TMV omega sequence at the 5’ end and TMV-L sequence at 3’ end of the mRNA to facilitate translation in the plant system.
3.4 Construction of a library of Mb cell-free expression plasmids

Pig, human, sperm whale, grey seal, goosebeak whale, emperor penguin, and dwarf sperm whale Mb were selected to be examined by the cell-free expression system assay because these proteins span both a 600-fold range of measured apoglobin stabilities reported originally by Scott et al. and 10 to 20-fold differences in levels of myoglobin found in skeletal muscle as reported by Mircesta et al. for terrestrial and deep-diving mammals. My study also represents the first attempt in literature to express recombinant dwarf sperm whale myoglobin and goosebeak whale myoglobins. Scott et al. had obtained the native myoglobins for these species directly from the animal muscle tissue. Grey seal and emperor penguin Mbs were chosen at the suggestion of Dr. Michael Berenbrink and his colleagues based on their sequence comparisons, diving capacity, and preliminary unfolding studies.

For the cell-free expression system, all Mb genes were cloned into pEU vectors (Fig. 3.1) that were modified by the Center of Eukaryotic Structural Genomics (University of Wisconsin-Madison, WI). The original pEU-HSBC vector was a kind gift from the this center. The sequence sources of these Mb genes were explained in Section 2.1. The dwarf sperm whale, goosebeak whale, grey seal, and emperor penguin myoglobin genes were ordered from Integrated DNA Technologies (IDT) for insertion into pEU vectors independently. These genes were not codon-optimized for either bacterial or eukaryotic expression because the wheat germ translation system is thought to be independent of codon bias. For example, similar amounts of protein were expressed from both codon optimized and native Psf25 (a malaria protein) genes, demonstrating codon optimization was not needed for wheat germ protein expression system. Then, three distal pocket mutants, H64F/V68F sperm whale Mb and H64L and H64F/V68F dwarf sperm whale Mb, were constructed using PIPE site-directed
mutagenesis. The goal was to enhance globin stability and, at the same time, reduce hemin affinity. The specific mutations were selected based on the results described by Smith.

### 3.5 In vitro transcription and cell-free protein expression

ENDEXT® Technology Protein Research (H) Kits (CellFree Sciences (CFS)) were used to set up the decoupled small-scale mRNA transcription and cell-free translation reactions. The wheat germ extracts from these specific kits had been pretreated to remove endogenous proteins that bind to nickel (Ni\(^{2+}\)) columns. Since in this work, the cell-free expressed proteins were purified over Zn\(^{2+}\) resin, which is often used interchangeably with the Ni\(^{2+}\) resin in protein purification, less variability was observed in the expression yields for the same Mb variants compared to when untreated wheat germ extracts were initially used (data not shown).

Following Cell-Free Systems (CFS) protocols, concentrated pEU-Mb DNA template (2000 ng) was prepared to be DNase-RNase free and then used as the starting material for in vitro transcription at 37°C. Immediately following transcription, the resultant mRNA was incubated with the wheat germ extract for 20 hours at 15°C\(^{137-140,145}\). A layer of translation cofactors, including amino acids, ATP, and creatine phosphate, was placed on top of the wheat germ extract to set up a bilayer translation reaction volume of 226 µl \(^{141,145}\) (Fig. 3.1). MetMb expression requires the presence of free hemin, which was added to the upper bilayer following a protocol developed for this study (see following section). During the 20-hour incubation, hemin, along with other translation cofactors, diffuses into the wheat germ extract where the Mb mRNA is being translated into polypeptides by the ribosomes. At the end of the 20-hour incubation, the formation of brown metMb can often be seen with the naked eye (Fig. 3.2), especially for the Mb genes producing higher expressions yields. This color change can be
monitored over the 20-hour incubation period beginning with a pale green color (free hemin) at the start of the reaction to a darker brown color for intact holo-metMb. Preparation of required materials and the cell-free expression system was done in a DNase-RNase free room. All glassware was baked at 180 °C for 3 hours before use, and solutions were prepared with DNase-RNase free distilled water (Invitrogen)\(^8\).

Figure 3.2. **Cell-free expression of goosebeak whale Mb.** Results are shown for a 20-hour incubation of the bilayer translation reaction. Wheat germ extract containing the ribosomes and globin mRNA are in the bottom layer, where the concentrated brown metMb is observed.

### 3.6 Development of heme addition protocol for cell-expression of metMb

Whereas the *in vivo* protoheme synthesis pathway starting from 5-aminolevulinate (ALA) regulates holoMb expression in *E. coli*, the wheat germ extract lacks this pathway\(^{146}\). To drive holoprotein expression in the cell free system, hemin had to be added externally during translation. Hemin has poor solubility in water at neutral pH but is soluble in polar organic solvents and in basic aqueous solutions. The insolubility, dimerization, and aggregation of...
hemin in aqueous solution will increase with increasing hemin concentration and decreasing pH. In aqueous solution at high concentrations, hemin (H₂O/OH-Fe(III)PPIX) reversibly dimerizes at the unliganded axial faces giving rise to distinct spectral changes. These complexes can further aggregate and precipitate and, on longer time scales (days), form covalent complexes called μ-oxo-bridge dimers that cannot interact with apoproteins or be rapidly reduced. Thus, all hemin solutions were prepared right before use in the assays.

In order to reduce variability in heme protein expression from one cell free reaction set to another, free hemin dimerization, aggregation, and surface adsorption factors have to be minimized. The presence of strong or weak ligands in solution can reduce hemin dimerization through ligand binding at the axial sites of hemin. Initially, in order to decrease the occurrence of hemin dimerization, the cell free expression of sperm whale holoMb was performed with the addition of 23 μM dicyanohemin [CN-Fe(III)PPIX], a liganded hemin, solution. The Soret absorbance of free dicyanohemin is at 422 nm, and Soret absorbance of cyanometMb is at 423 nm. However, when the purified cell free reaction was analyzed, a Soret peak intermediate between of metMb and free dicyanohemin was obtained (Fig. 3.3A). SDS-PAGE analysis showed a protein band corresponding to sperm whale Mb molecular weight (~17 kDa) with dicyanohemin addition. The band’s intensity was somewhat higher compared to that of cell-free expressed sperm whale apoMb without hemin addition (Fig. 3.3B). The addition of dicyanohemin was facilitating expression of a mixture of metMb and cyanometMb species. DTT is present in the translation mixture to maintain creatine kinase in its active, reduced form to enable the enzyme to catalyze the generation of ATP from creatine phosphate and ADP. This presence of DTT was probably driving the reduction of either free cyanohemin or cyanometMb to produce deoxyheme and deoxyMb, and because the protein expression
proceeds in an aerated translation mixture, these reduced species reacted with O$_2$ to transiently produced oxygenated complexes that then auto-oxidized during the 20 hour incubation.

Figure 3.3. Analysis of cell-free expression of sperm whale holoMb with cyanohemin addition. A. Visible absorbance analysis of cell-free expression of holoMb (semi-purified)
following cyanohemin addition. B, SDS-PAGE analysis of cell-free expression of holoMb (semi-purified) following dicyanohemin addition. ApoMb was in loaded in lane 3, and cyanometMb was loaded in lane 5 and 7. Protein bands at ~17 kDa corresponding to Mbs are boxed in red. Lane 10 and 11 were loaded with protein purification washes obtained during the small-scale purification of the cell-free expressed Mbs.

Spectral analysis the translation products using dicyanohemin was difficult because dicyanohemin and cyanometMb have very similar peaks in the Soret region and significant amounts of metMb are also made due to oxidation reaction reactions and loss of cyanide. I also examined additions of caffeine-hemin complexes because caffeine is thought to keep hemin monomeric in solution, but there was no improvement in apparent yield; further the caffeine interfered with measures of protein concentration at 280 nm. In the end, I developed the following optimized protocol of hemin addition for cell-free expression of metMb.

As discussed above, free hemin can dimerize, aggregate, and adsorb to glass and plastic surfaces at high hemin concentrations and low pH. Thus, glassware and pH probes used for hemin preparation were rinsed with 1 M NaOH to reduce hemin adsorption. A stock solution of fresh hemin in 1 M NaOH was prepared 1-2 hours before initiating the cell-free translation reaction because irreversible hemin aggregation increases slowly with time. The pH of the stock hemin solution was then lowered to pH 8.4 by gradual addition of 50% acetic acid. Acetic acid is also a weak ligand for hemin, and therefore can reduce hemin aggregation. Next, the hemin stock solution was spun down to separate out any insoluble hemin before the solution concentration was measured spectroscopically, using $\varepsilon_{398} = 122$ mM$^{-1}$cm$^{-1}$. Finally, a total of 5.65 nmol hemin was added to the upper bilayer of the cell-free translation reaction in three aliquots, with each aliquot added hourly beginning an hour after initiation of translation. The pH of the hemin added had been optimized to ensure that both hemin solubility
and the conditions for the cell-free protein expression system were not compromised. The pH of the cell-free translation mixture from ENDEX® Technology Protein Research (H) Kit (CellFree Sciences (CFS)) was ~7.8. Fig. 3.4 compares spectra of free hemin prepared through this method with purified cell-free expressed dwarf sperm whale Mb obtained following hemin addition using this optimized protocol. These spectra show clearly that hemin was taken up by newly translated apoMb to form holo-metMb with a distinct narrow band centered at 409 nm, which is ≥ 10 nm red-shift from the free hemin bands.

Figure 3.4. Soret absorbance (Abs) of semi-purified dwarf sperm whale Mb obtained from cell-free expression free hemin addition at pH 8.4, prepared with 1M NaOH and 50% acetic acid as outlined in section 3.6.
3.7 Quantitative RT-PCR measurements of mRNA levels transcribed in vitro

After transcription, the mRNA was treated with DNase I (RNase-free) to digest the template DNA. Then, the mRNA transcript was purified with a MEGAclear™ Kit (Life Technologies) to minimize the presence of free nucleotides and digested DNA template. To measure the amount of mRNA present, a cDNA reverse transcription reaction was performed using a sequence specific primer for each Mb mRNA species. The amount of cDNA generated was measured through quantitative PCR, utilizing the SYBR® Green JumpStart™ Taq ReadyMix™ (Sigma-Aldrich) for High Throughput Quantitative PCR.

3.8 Quantification of total Mb levels expressed in cell-free system with indirect ELISA

After translation of dwarf sperm whale and sperm whale Mbs, unpurified cell-free reaction mixtures were spun down to separate the pellet containing precipitated protein from the lysate containing soluble protein. Wells in a 96-well clear flat bottom plate (Corning® Costar®) were then individually coated with 50 μl lysate and pellet dilutions in 50 mM carbonate buffer pH 9.6 (Sigma). The plates were incubated overnight at 4°C. Each well was then blocked at room temperature for 1 hour with a 1% bovine serum albumin (BSA), 0.02% azide solution in phosphate buffered saline (PBS). Next, each well was incubated at room temperature with 50 μl Santa Cruz FL-154 rabbit polyclonal IgG Mb antibody at 1:50 dilution in the blocking buffer for 2 hours and then incubated with 100 μl of Santa Cruz goat anti-rabbit
IgG-HRP at 1:2000 dilution in 1% BSA in PBS for 1 hour. Wells were washed with 100 μl PBS with 0.05% tween-20 between different steps. Finally, each well was incubated with 50 μl Thermo Scientific Pierce 1-Step™ Ultra TMB-ELISA Substrate for 15 minutes before the peroxidase reaction was stopped with 50 μl 2M sulfuric acid. Absorbance was read at 450 nm using a TECAN Infinite M1000 Pro microplate reader. The Santa Cruz FL-154 rabbit polyclonal IgG Mb primary antibody used in the assay was determined through Western blot to recognize the whale Mbs. ELISA standard curves for the different species were obtained by optimizing the primary antibody incubation time and dilution and making serial dilutions of the different Mb protein standards of known concentrations.

**3.9 Small scale isolation and purification of soluble cell-free expressed Mb**

Initially, it was difficult to isolate fully folded holoMb protein from His(6)-tag Mb constructs with the cell-free expression system due to nonspecific heme binding to the poly His tag. In the end, tagless constructs of Mb were made because my main focus was to investigate globin expression in terms of globin folding of the native proteins. Any additional solubility and purification tags in the primary amino acid sequence might have influenced both globin folding and heme insertion during expression.

Attempts were also made to pursue small-scale purification incorporating DEAE Sepharose Fast Flow (GE Healthcare) resin as was used for large-scale purification of Mbs expressed in *E. coli*, but these methods did not efficiently isolate cell-free expressed Mbs. Hemoglobins and Mbs are rich with natural Zn$^{2+}$ binding sites on their surfaces. A small-scale holoMb purification protocol utilizing Zn$^{2+}$ resin was then developed by adapting a previously published, small-scale purification protocol for His(6)-tagged proteins (S. Makino,
University of Wisconsin, Center for Eukaryotic Structural Genomics). The pH, ethylenediaminetetraacetic acid (EDTA), and salt concentration of the purification wash and elution buffers were optimized as described below in the protocol developed to reduce non-specific binding of eukaryotic proteins on the Zn$^{2+}$ resin and maximize the isolation of Mbs.

Figure 3.5. **Binding of holoMb to Zn$^{2+}$ resin during small-scale purification of cell-free expressed Mbs.** The bound metMb is brown in color.

Zn$^{2+}$ Chelating Resin (GBiosciences) was added into the wells of a MultiScreen$_{HTS}$ HV 0.45 μm filter plate (Millipore) to purify $10^{10}$ to $10^{11}$ moles of soluble holoMb generated from each cell-free translation reaction (total volume ≈ 230 μl). A 96-well U bottom plate (Grenier Bio One) was affixed below the 96-well MultiScreen$_{HTS}$ HV filter plate containing the Zn$^{2+}$ resin for collection of wash buffers and then the eluted holoMb, following centrifugation of the plates at 3,640 rpm for 1 minute. Forty μl of Zn$^{2+}$ resin was loaded in each well of the filter plate and washed with 200 μl distilled, deionized water. Next, 100 μl of each resuspended translation reaction mixture was pipetted on top of the wells containing resin, along with 140 μl 0.1 M sodium phosphate equilibration buffer, pH 7, and incubated on a plate shaker for 20 minutes before the plate was centrifuged. Then, the protein bound on the resin (Fig. 3.5) was
subjected to decreasing salt washes of 150 μl volumes: first wash buffer (0.5 M sodium chloride, 0.1 M sodium phosphate, pH 7), second wash buffer (0.3 M sodium chloride, 0.1 M sodium phosphate, pH 7), and third wash buffer (0.1 M sodium chloride, 0.1 M sodium phosphate, pH 7). Finally the resin was incubated with 60 μl elution buffer (0.1 M sodium phosphate, 15 mM EDTA, pH 7.4) for 10 minutes on a plate shaker. The plates were centrifuged again to elute the bound holoMb protein.

Figure 3.6. SDS-PAGE analysis of semi-purified cell-free expressed Mbs at ~17 kDa. The protein samples and the blank were semi-purified on the small-scale Zn²⁺ column in the well-plate filter. The blank was wheat germ extract with added hemin but without Mb mRNA present during translation reaction.
3.10 Quantification of cell-free expressed, soluble holoMb

The cell-free expression of soluble holoMb was confirmed after partial purification of the soluble protein by measurement of Soret absorbance (Abs) peaks (see Fig. 3.7) and protein bands at ~17 kDa (molecular weight of Mbs) in SDS-PAGE gels (see Figs. 3.1 (bottom left) and 3.6). Note that protein precipitates do not pass through the Zn²⁺ resin mini-columns, which also act as filters to remove particulates.

The cell-free expression level of the semi-purified, soluble holoMb was quantified as the concentration of Mb determined from the Soret absorbance (AbsSoret) normalized to an estimate of the total protein concentration present, which was quantified by the observed absorbance at 280 nm (Abs280) (eq 3.1). This ratio takes into account differences in amounts of total protein present in the partially purified sample (reflected in variability of the Abs280) due to soluble enzymes from the wheat-germ ribosomal system, which would affect translational efficiency, and any residual folded apomyoglobin. The latter should be minimal in the soluble fraction because excess hemin was present. For the native ferric holoMb, the Soret abs peak is at 409.5 nm and has an extinction coefficient (εSoret) of ε409.5 = 157 mM⁻¹cm⁻¹, whereas the 280 nm peak has an ε280 = 31.2 mM⁻¹cm⁻¹, which was used in eq 3.1 as an average ε280 for holoMb100. For the ferric holoMb heme pocket mutants H64L and H64F/V68F, the Soret absorbance peak is at 395 nm with ε395 = 103 mM⁻¹cm⁻¹162. The variance in triplicate assays was reduced significantly when eq 3.1 was used to normalize the holoMb yield by the amount of total protein present.
Figure 3.7. Absorbance spectra of semi-purified holoMb isolated from cell-free translation reactions. The protein samples and the blank were semi-purified on the small-scale Zn\(^{2+}\) column in the well-plate filter. The blank refers to the wheat germ extract with added hemin but without Mb mRNA present during translation reaction.

3.11 Cell-free expression of soluble holoMbs is the highest for apolar distal heme pocket mutants and deep-diving variants

A summary of the cell-free expression data for the small library of recombinant myoglobin variants is shown in Table 3.1. In all cases, the holoMb samples were first semi-purified on a Zn\(^{2+}\) chelating resin filter plate, and then expression was measured as the ratio of the heme protein concentration calculated at the Soret absorbance peak to an estimate of total
protein present measured by absorbance at 280 nm (eq 3.1 and Fig. 3.7). For highly expressing Mbs, the Soret absorbance peaks were sharply defined (Fig. 3.7).

The cell-free expression yields for the H64F/V68F and H64L dwarf sperm whale, and H64F/V68F sperm whale holoMb mutants were significantly higher than any of the other naturally occurring variants examined. The cell-free expression yields of human and pig holoMbs were dramatically lower than any of the holoMbs from deep diving mammals (Table 3.1). The Soret absorbance peaks for the semi-purified lysates containing pig and human Mbs and the blank wheat germ extract were not sharply defined (Fig. 3.7).

Table 3.1. *Summary of Mb cell-free expression yields.*
The cell-free reactions were done using ENDEXT® Technology Protein Research (H) Kits (*CellFree Sciences* (CFS)).

<table>
<thead>
<tr>
<th>Mb variant</th>
<th>Expression yield</th>
<th>Relative expression yielda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dw H64F/V68F</td>
<td>0.215 ± 0.036</td>
<td>2.27 ± 0.48</td>
</tr>
<tr>
<td>Sw H64F/V68F</td>
<td>0.208 ± 0.007</td>
<td>2.17 ± 0.29</td>
</tr>
<tr>
<td>Dw H64L</td>
<td>0.181 ± 0.019</td>
<td>1.77 ± 0.30</td>
</tr>
<tr>
<td>Dwarf Sw (Dw)</td>
<td>0.171 ± 0.011</td>
<td>1.62 ± 0.24</td>
</tr>
<tr>
<td>Grey seal</td>
<td>0.154 ± 0.016</td>
<td>1.38 ± 0.23</td>
</tr>
<tr>
<td>Goosebeak whale</td>
<td>0.151 ± 0.005</td>
<td>1.34 ± 0.181</td>
</tr>
<tr>
<td>Sperm whale (Sw)</td>
<td>0.128 ± 0.017</td>
<td>1.00 ± 0.18</td>
</tr>
<tr>
<td><em>Emperor penguin</em></td>
<td></td>
<td>0.90b</td>
</tr>
<tr>
<td>Human</td>
<td>0.079 ± 0.009</td>
<td>0.28 ± 0.05</td>
</tr>
<tr>
<td>Pig</td>
<td>0.071 ± 0.008</td>
<td>0.16 ± 0.03</td>
</tr>
</tbody>
</table>
a Relative expression was computed as described in Eq. 3.2 in the text.

b This cell-free reaction was done initially using ENDEXT® Technology Protein Research (S) Kits (CellFree Sciences (CFS)), which has a different blank value, and therefore only its relative expression yield to Sw Mb is presented in the table. The wheat germ extracts from the ENDEXT® Technology Protein Research (S) Kits (CellFree Sciences (CFS)) were not treated to remove endogenous proteins that bind to nickel resin.

When the semi-purified Mbs were analyzed on SDS gels, however, a small amount of the terrestrial Mbs were seen to be expressed in the system, albeit with a much less intense protein band compared to those of the other Mbs (Figs. 3.1 (bottom left panel, second column) and 3.6).

As shown in Fig. 3.7, there is variability in the A_{280} values for the translation products, and thus normalization for total protein (both expressed product and soluble enzymes from the translation reaction mixture) in the partially purified sample is needed. Similarly, it is clear from Fig. 3.7 that the H64F/V68F dwarf sperm whale metMb has a different Soret maximum (395 nm) from that of the native whale metMbs (409.5 nm), and thus the extinction coefficient difference needs to be taken into account as prescribed in eq 3.1. The broader and less intense Soret band for the H64L and H64F/V68F mutants of Mb is due to the loss of coordinated water and reflects the spectral properties of five-coordinate hemin.

The (A_{Soret}/ε_{Soret})/(A_{280}/ε_{280}) (eq 3.1) ratio for a cell-free translation reaction mixture that contains hemin but no holoMb was assigned a value of 0.06. This value is only slightly below the measured expression yields observed for pig and human Mbs and equal to or just slightly less than values computed for a variety of "blank" reactions (no mRNA) using eq 3.1. The spectrum of the “blank” in Fig. 3.6 clearly shows that there is some hemin present after purification with the Zn^{2+} resin as judged by the broad Soret absorbance. This hemin is
probably bound non-specifically to proteins that are a part of the translation system and elute from the Zn$^{2+}$ resin with holoMb in the presence of EDTA. This interpretation is supported by presence of a 280 nm absorbance peak for the blank control when no mRNA was added and no holoMb was made. The cell-free expression yield normalized using eq 3.1 is presented with no subtraction of this “blank” value in the second column of Table 3.1 to provide a clearer indication of the original sample spectra (Fig. 3.6).

To compare the cell-free results to those reported previously by Smith$^{71}$, the blank value was subtracted from all the observed yields (determined through eq 3.1) and then computed as the expression relative to a value of 1.0 for wild-type sperm whale myoglobin. This relative yield is defined as:

$$\text{Relative holoMb yield} = \frac{\text{holoMb expression yield} - 0.06}{\text{sperm whale holoMb expression yield} - 0.06}$$

$$\text{eq 3.2}$$

This relative yield of holo-protein is not dependent on the $\varepsilon_{280}$ value used for the initial calculation of the measured expression yield of soluble holoMb from eq 3.1 and allows more direct comparisons of the expression yields of the Mb variants to other values in the literature (Table 3.1, third column$^{41,42,71,134}$).

The results in Table 3.1 show quantitatively that Mbs from deep diving whales are expressed in cell-free systems to levels that are 5 to 8 time higher than Mbs from terrestrial mammals. These differences in in vitro yields correlate directly with the higher levels of Mb found in the skeletal muscles of the deep diving mammals, as described in Mirceta et al.$^{42}$. 
3.12 Lack of correlation between protein expression and transcript level

Figure 3.8. Correlation between holoMb cell-free expression yield and amount of mRNA transcribed in vitro. The amount of mRNA was determined by quantitative PCR as described in section 3.5. Expression yields were computed from eq 3-1. The dashed line represents a linear fit with an $R^2 = 0.12$ and the dashed-dotted line represents no dependence (average expression yield).

To ensure that higher levels of holoMb were not due to higher levels of mRNA generated during in vitro transcription, the total amount of mRNA generated in the first step of the decoupled transcription-translation system (Figure 3.1) was analyzed quantitatively. The results of this analysis are shown in Figure 3.8. Although the variability in the mRNA present was greater than expected, there was no correlation between the extent of cell-free expression
of soluble holoMb and the amount of Mb mRNA present. The amount of mRNA transcribed \textit{in vitro} varied across the different species. In fact, the low expressing holoMb proteins originating from pig and human were translated from amounts of mRNA comparable to those of the high expressing H64L dwarf sperm whale Mb and sperm whale Mb. The higher expressing dwarf sperm whale Mb and H64F/V68F dwarf sperm whale Mb were transcribed from lower amounts of mRNA compared to the other Mbs expressed (Fig. 3.8).

This lack of correlation between the total mRNA measured and expression of holoMb (Fig. 3.8) might be attributed to several factors that are most likely unique to the eukaryotic wheat germ based cell-free expression system. In comparison to prokaryotic mRNAs, eukaryotic mRNAs generally are much more stable, with globin mRNAs in reticulocytes having half-lives of 50 hours\textsuperscript{163-167}. The stability, translation efficiency, and regulation of degradation of eukaryotic mRNAs are often attributed to their 5’ and 3’ untranslated regions (UTRs). The \textit{in vitro} mRNA transcribed was engineered to include UTRs at the 5’ (omega (Ω) sequence) and 3’ ends derived from tobacco mosaic virus (TMV) positive sense RNA, but without the 5’ end cap and poly(A) tail for \textit{in vitro} translation efficiency and to bypass the issues of decapping and shortening of the poly(A) tail of the mRNAs during cell-free expression (Fig. 3.1)\textsuperscript{135,139,140}. Therefore, a dependence of \textit{in vitro} expression yield on mRNA transcript concentration is less likely, above a certain threshold, due to high mRNA half-life and consistent 5’ and 3’ TMV UTR sequences, which confer similar stability across the different transcripts. In effect, the ribosomes in our cell free system were likely "saturated" with stable mRNAs at the levels present in our assays.

In addition, previous studies using electron microscopy have also shown efficient recycling of multiple ribosomes translating mRNA simultaneously through the formation of
circular polysomes linking the mRNA’s 3’ and 5’ ends in the cell-free wheat germ extract\textsuperscript{168}. Poly(A) binding protein I (PABI), present in all eukaryotic cells, is postulated to place the disengaged ribosome at the 3’ back to the 5’ front of the mRNA, and Madin et al.\textsuperscript{135,168-171} have shown \textit{in vitro} that this phenomenon is not dependent on the presence of 5’ cap and 3’ poly (A) tail for the mRNA. Therefore, continuous ribosomal recycling on a cell-free pool of mRNA most likely decouples the translation yield dependency on the mRNA transcript concentration, above a certain threshold level.

### 3.13 Total protein synthesis, expression of folded protein, and protein aggregation

ELISA measurements of the total amount of soluble, folded holoMb in the wheat germ extract lysate correlated with and thereby confirmed the holoMb cell-free expression yields defined by eqs 3.1 and 3.2 for the wild type and heme pocket mutants of sperm whale and dwarf sperm whale Mb (Table 3.2). The heme pocket mutants had a higher amount of polypeptide in the soluble form relative to wild type Mbs. Minimization of proteases in the pre-treated wheat germ extract\textsuperscript{137} also allowed measurement of unfolded proteins and enabled a better estimate of the total polypeptide (both folded and unfolded/aggregated protein) synthesized using the indirect ELISA measurements. The total amount of polypeptide synthesized for both sperm whale Mb and dwarf sperm whale Mb are within a similar range. However, the H64F/V68F mutation in both sperm whale Mb and dwarf sperm whale Mb and the H64L mutation in dwarf sperm whale Mb appeared to double the total amount of polypeptide synthesized. The cause of this effect is unclear; however, the ELISA assay is
difficult to quantitate precisely because of the differences in specificity of the Mb antibody. Further, quantification of the unfolded polypeptides might be affected by the degree of protein aggregation that will reduce the binding sites of the antibody to the polypeptide.

Table 3.2. **ELISA measurements of total Mb synthesized in the cell-free expression system.**

Unpurified cell-free translation mixtures were spun down after expression incubation to obtain pellets containing precipitated protein and a lysate containing soluble protein. Santa Cruz FL-154 rabbit polyclonal IgG Mb antibody was incubated with the polypeptides contained in the lysates and pellets at 1:50 dilution in 1% bovine serum albumin (BSA), 0.02% azide solution in PBS. For the detection of Mb polypeptides, incubation was next done against Santa Cruz goat anti-rabbit IgG-HRP at 1:2000 dilution in 1% BSA in PBS. Quantification of Mb polypeptides was next done through absorbance measurements following incubation with Thermo Scientific Pierce 1-Step™ Ultra TMB-ELISA Substrate.

<table>
<thead>
<tr>
<th>Species Mb</th>
<th>Total moles in lysate (nm)</th>
<th>Total moles in pellet (nm)</th>
<th>Total protein synthesized (nm)</th>
<th>Ratio of Mb in lysate to pellet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sw</td>
<td>0.057 ± 0.015</td>
<td>0.055 ± 0.012</td>
<td>0.11 ± 0.03</td>
<td>1.04 ± 0.35</td>
</tr>
<tr>
<td>Sw H64F/V68F</td>
<td>0.15 ± 0.02</td>
<td>0.09 ± 0.009</td>
<td>0.24 ± 0.024</td>
<td>1.67 ± 0.28</td>
</tr>
<tr>
<td>Dw</td>
<td>0.096 ± 0.027</td>
<td>0.041 ± 0.0098</td>
<td>0.14 ± 0.03</td>
<td>2.34 ± 0.86</td>
</tr>
<tr>
<td>Dw H64F/V68F</td>
<td>0.18 ± 0.07</td>
<td>0.091 ± 0.052</td>
<td>0.27 ± 0.12</td>
<td>1.98 ± 1.37</td>
</tr>
<tr>
<td>Dw H64L</td>
<td>0.20 ± 0.065</td>
<td>0.094 ± 0.012</td>
<td>0.29 ± 0.08</td>
<td>2.13 ± 0.74</td>
</tr>
</tbody>
</table>

3.14 Conclusions

The results in this chapter demonstrate that holoMb expression can be examined quantitatively using a cell-free wheat germ translation system. As previously observed in mammalian muscle\(^{42,129-132}\) and *E. coli* cells\(^{172}\), the deep diving species have Mb genes that
express to much higher levels than the genes from terrestrial animals. Roughly 6 to 10-fold higher levels of Mb expression were observed for genes from deep-diving animals in the cell-free expression assay. Similar differences were observed for the expression of the native Mbs in animal muscle cells. My results in the cell-free assay for the apolar distal pocket mutants (Table 3.1 and Table 3.2) also correlated with Smith's previous observation that His64 and Val68 to Leu or Phe mutations caused ~2.0-2.5 fold increases in expression in E. coli. Interestingly, there is no correlation between mRNA transcript levels and cell-free expression levels of holoMb when all other levels of transcription variables are standardized, and these results likely apply to animal muscle cells as well because the wheat germ extract is also an eukaryotic translation system.

My results in Tables 3.1, 3.2, and the previous work of Scott et al., Smith, and Mirceta et al. strongly suggest that amino acid sequence and stability of the resultant globin fold directly influence holoprotein expression. These studies highlighted the need to examine in a more systematic manner the mechanism of globin expression in terms of the biophysical properties of the protein. The cell-free expression assay was therefore developed to try to determine the key biochemical properties the govern production of hologlobins. In the next chapter, unfolding and hemin binding parameters were measured systematically for the Mb variants Table 3.1, and correlations are made with the relative cell-free expression yields.
Chapter 4

Correlation between expression of soluble holomyoglobin and apomyoglobin stability

This chapter is adapted and reproduced from my following publication:


In order to avoid redundancy, I have minimized and/or omitted further references to this publication within the text body.

4.1 Introduction

Multiple factors can contribute to in vivo expression yields of holoMb, including apoprotein stability, hemin affinity, auto-oxidation, ligand binding, as well as transcription and translation factors. Most studies on gene expression, especially in systems biology, focus on the regulation of mRNA synthesis by promoter sequences and transcription factors, mRNA structure and stability, and rates of translation on ribosomes. In contrast, my work focuses on the final steps of expression that involve the folding and assembly of fully functional hologlobins, which depend on the amino acid sequence of the polypeptide chain that specifies the secondary, tertiary, and quaternary structure of the protein.

In this chapter, I discuss my observation of a strong linear correlation between cell-free expression levels of soluble metMb variants and their corresponding apoglobin stabilities.
determined from the analysis of *in vitro* unfolding curves. In contrast, the expression levels show little or no dependence on either the Mb transcript levels (Fig. 3.8) or the affinities of apoMb for heme, which were also measured from unfolding curves. Culbertson and Olson\(^{9,30,31}\) suggested that holoMb is primarily generated by insertion of the heme into the folded heme pocket of apoMb because the molten globule has a much lower affinity for the prosthetic group. The higher affinity of the native state is due to the rigidity of the heme pocket, which also inhibits the occurrence of hemichromes\(^{9,30,32,87}\). Based on this idea, a simple holoMb expression model was developed to analyze quantitatively the cell-free expression data.

The *in vitro* model for native apoMb folding\(^{9,13,27,39,41}\) assumes that the unfolded globin polypeptide (likely with transient helicity) first comes off the ribosomes and then initially folds into a molten globule formed by the AGH helices, which includes the N- and C-termini of the globin. Then, the heme pocket forms in order to generate the fully folded native apoMb (N) state as shown in Fig. 1.4. This N state can rapidly bind heme to produce the ferric holoMb (metMb), which in turn can rapidly be reduced and bind \(\text{O}_2\) or \(\text{CO}\) to be stabilized the final holoprotein product. As shown in Fig.1.4, the net synthesis of holo-metMb and its reduction competes with aggregation and precipitation of both unfolded apoprotein and free hemin. In order to have high levels of expression of the holoprotein, Mbs from deep diving mammalians have evolved to have very stable apoglobin structures to reduce the fraction of unfolded polypeptides and inhibit irreversible precipitation of denatured states.
Figure 4.1 A. Structure of dwarf sperm whale oxyMb (unpublished) determined in this thesis work. The side chains shown are different from those at similar positions in Sw Mb, which are Val21(B2), Ser35(B16), Thr51(D1), Ala121(GH3), and Asn132(H8), and these differences lead to marked difference in folding and expression parameters. B Structural alignments of dwarf sperm whale Mb (mixture of O$_2$ and CO bound forms) (red helix), sperm whale MbO$_2$ (PDB #1A6M, magenta helices), and pig metMb (PDB #1PMB, cyan helices).

4.2 Structure of dwarf sperm whale (Dw) metMb

In early 2000s’, Scott et al.\textsuperscript{39} had shown that dwarf sperm whale Mb has the highest stability ($K_{UN}$ folding constant), whereas pig Mb had the lowest $K_{UN}$ value among the 13 mammalian myoglobin variants investigated in vitro. In an independent study, Mirceta et al.\textsuperscript{42} showed that the Mbs from deep diving mammals have evolved sequences that result in markedly increased in net positive surface charge ($Z_{Mb}$) at pH 6.5. For example, Mirceta et al.\textsuperscript{42} had computed the surface net charge at pH 6.5 for dwarf sperm whale Mb, native sperm whale Mb (the D122 variant), and pig Mb to be +4.24, +4.15, and -0.02, respectively, based on their sequences and assumed surface side chain pKa values. There are exactly 5 amino acid differences between native sperm whale Mb and dwarf sperm whale Mb (highlighted by
yellow residues in Fig. 4.1A) and 22 amino acid differences between pig Mb and native sperm whale Mb. These differences in sequence clearly cause marked differences in expression yields and apoglobin stability. The structures of pig and Sw holoMbs have already been determined and to explore the stability issue further, I determined the structure of recombinant dwarf sperm whale oxyMb in the ferrous state (Fig. 4.1, Table 4.1).

Table 4.1 Structure determination parameters of the dwarf sperm whale Mb. Parameters in parentheses are for the outer, lowest resolution shell.

Crystal data:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>27.75 - 1.88 (1.947 - 1.88)</td>
</tr>
<tr>
<td>Space group</td>
<td>P 41 21 2</td>
</tr>
<tr>
<td>Unit cell parameters (Å, º)</td>
<td>a=85.995, b=85.995, c=109.01</td>
</tr>
<tr>
<td>Reflections (measured/unique)</td>
<td>33947/218789</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.85 (99.91)</td>
</tr>
<tr>
<td>Mean &lt;I&gt;/σ(I)&gt;a</td>
<td>15.22 (1.53)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.0 (2.0)</td>
</tr>
<tr>
<td>$R_{merge}$ (%)</td>
<td>8.7 (126.6)</td>
</tr>
</tbody>
</table>

Refinement:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>29.27 - 1.88</td>
</tr>
<tr>
<td>R-factor (%)</td>
<td>19.03 (27.99)</td>
</tr>
<tr>
<td>$R_{free}$ (%)</td>
<td>23.62 (32.65)</td>
</tr>
</tbody>
</table>

**RMS deviations from ideal values:**

| Bond length (Å)                  | 0.007                           |
| Bond angles (º)                  | 0.9                             |

Ramachandran plot:

| Residues in most favorable region (%) | 96.38 |
| Residues in additional allowed region (%) | 3.62  |

Hemin $RSCC^d$ (%) (Chain 1/Chain 2) 98/97

---

$^a$Signal to noise ratio of the intensities

$^b$Rmerge is an indicator of the data quality in terms of merging symmetry related reflections, and using the best reflection intensity

$^c$Root mean square (RMS)
Real space correlation coefficient (RSCC), representing the electron density map correlation between that of calculated and experimental.

The crystal structure of recombinant holo dwarf sperm whale Mb was solved at 1.88Å through molecular replacement using the sperm whale Mb PDB ID# 2MBW as a reference. The protein crystalized in the P 41 21 2 space group, and there are two chains of the dwarf sperm whale Mb in the asymmetric unit of the unit cell. In contrast, liganded native sperm whale Mb crystalizes in the P 21 space group, and pig Mb crystalizes in the I 21 space group, also with two chains in the asymmetric unit of the unit cell. Interestingly, the R state of human hemoglobin also crystalizes in the P 41 21 2 space group. Fig. 4.2 shows the quality of the electron density map of the new Dw metMb structure. The parameters for the data collection and refinement are given in Table 4.1. The protein structure was solved with a R-free and R-work values of 0.236 and 0.19, with R (residual index) representing the difference between observed and calculated data. These R values are a measure of the fit quality of the solved protein structure to the data observed. Approximately, 5% of the reflections were flagged and not included in the refinement of the protein model and were then used to calculate the unbiased R_free value, whereas the data set used for refinement was used to calculated R_work.

The heme-bound ligand density appears to represent partial occupancy of O2, even though crystal trays had been set with CO bound dwarf whale Mb. The bound oxygen is a result of exposure to air during handling and data collection. It should be noted however that the tertiary structures of Sw MbCO, MbO2, and aquometMb are virtually identical (e.g., compare PDB IDs: #1A6G, #1A6K, #1A6M,) with only small variability in the orientation of
the distal histidine relative to the heme plane\textsuperscript{176}. Electron density was also observed for the N-Methionine residue in the recombinant dwarf sperm whale Mb structure corresponding to the ESI-TOF MS analysis (Fig. 2.1A), further demonstrating that N-Methionine residues are not cleaved for Mbs during \textit{E. coli} expression.

Figure 4.2. \textbf{Electron density maps of the active site of Dwarf whale Mb.}

A ribbon drawing for Dw metMb is shown in Fig. 4.1A, with the side chains that are different from Sw Mb shown as yellow sticks. The substitutions are mostly conservative in terms of size and polarity (see legend to Fig. 4.1A and Scott et al.\textsuperscript{39}). The backbone structures of pig, sperm whale (Sw), and dwarf (Dw) sperm metMbs are shown in Fig. 4.1B and are
virtually identical, despite the marked differences in apoglobin stability and expression levels of these naturally occurring globin homologues (see Tables 3.1 and 4.2). Clearly the amino acid changes must influence the folding process as the apoglobins come off the ribosome or as the polypeptide unfolds in GdnHCl titration experiments. Solvent content of the different Mbs might also influence protein stability parameters, with cavities being filled with apolar residues, either by nature or through molecular biology, increasing protein stability. However, all these effects are not directly obvious when examining the three crystal structures. Thus, direct measurements of stability are required to quantify the differences and are presented in this Chapter.

4.3 Measurement of apoglobin stabilities

In order to look for correlations between holoMb expression yield and globin stability, I measured the folding parameters for the Mb variants listed in Tables 3.1. Intrinsic protein stability was determined by analyzing GdnHCl induced unfolding curves for these apoMb variants using the mechanism in Fig. 1.4 and the equations and approaches described in Chapter 2, sections 2.9 and 2.10. Much of this work was published in Samuel et al.13.
Figure 4.3. GdnHCl induced equilibrium unfolding of selected apoMbs was followed by A, CD and B, Fluorescence changes. Full circles are the observed data while the solid lines are fitted curves using eq 2.1 and the parameters listed in Table 4.2. Unfolding measurements were performed with 10 μM protein in 10 mM potassium phosphate, pH 7 at 20°C to be comparable to the work of Culbertson and Olson. The CD signal at the initial native state (N) and maximum fluorescence at the intermediate state (I) is re-normalized to 1 in these curves.

The changes in the CD and fluorescence (Fig. 4.3) signals suggest two major unfolding processes. Intrinsic fluorescence emission at 341 nm for all Mb species (except grey seal) increases initially with increasing GdnHCl concentration and then decreases as the protein completely unfolds. The result is a defined bell-shaped curve (Fig. 4.3B), despite the high scatter in fluorescence data. The peak in the fluorescence curve helps to define the emission of the molten globule intermediate (I)\textsuperscript{9,27,41}, complementing the CD profiles, in which the I state is defined by an inflection point in the decrease in negative CD\textsubscript{222nm} signal amplitude. The intrinsic fluorescence of Mb likely originates from Trp7 and Trp14, both of which are conserved across all the native Mb variants examined. Previous studies have proposed that the increase in fluorescence for the I state is due to either the movement of a quenching amino acid side chain away from Trp7 or an increase in the flexibility of the indole side chain, allowing it
to move into a more apolar environment. In the completely unfolded U state, the fluorescence of the exposed Trp residues is quenched by surrounding solvent. A bell-shaped fluorescence curve was not observed for the unfolding of grey seal apoMb. Instead, higher fluorescence was observed for the N state which then decreased monotonically with increasing [GdnHCl], suggesting the absence of Trp7 quenching in the N apoMb state. The protein sequence of grey seal Mb and harbor seal Mb are identical, and holoMb protein crystal structure of harbor seal (Phoca vitulina) has already been solved (PDB ID #1MBS). Since the globin fold is conserved across all the mammalian Mbs, there is no obvious structural explanation for the absence of Trp 7 quenching in the seal Mb compared to what is observed for the whale Mbs, and mutagenesis studies are required to understand these differences.

The overall apoMb unfolding constant, $K_{NU}$, is defined as $K_{NI}K_{IU}$, and overall stability is empirically represented as $-\log(K_{NU})$ or $\log(K_{UN})$, which is proportional to the free energy released during folding to the native conformation, where $K_{UN}$ is the equilibrium constant for the $U \rightarrow N$ reaction. The results for $K_{UN}$, shown in Table 4.2, reiterate that wild type Mbs originating from deep diving mammals have significantly higher apoglobin stabilities, with dwarf sperm whale Mb being one of the most stable native Mbs investigated. Similar results were observed under high salt unfolding conditions (200 mM potassium phosphate, pH 7) by Scott, Paster, and Olson.

Ten out of the thirteen apomyoglobins examined by Scott et al. were obtained from native animals, and these myoglobins did not have an initiator Met (N-met) compared to the majority of recombinant myoglobins that I examined in Fig. 4.3 which do have an extra N-met for expression in E. coli (i.e., Dw H64F/V68F, Dw H64L, Sw H64F/V58F, Sw, and grey seal apoMbs). However, Scott et al. showed that presence or absence of the N-Met had little or no
effect on the measured folding constant ($K_{UN}$) for the transition from unfolded (U) to native, folded (N) for Sw apoMb, and I later obtained similar results with Dw Mb (Fig. 4.4, Table 4.3).

ApoMb variants constructed with large apolar substitutions at the E7 and E11 helical positions also showed significantly higher apoglobin stabilities than the original wild-type proteins (Fig. 4.3, Table 4.2).

Table 4.2. Summary of Apomyoglobin stabilities, hemin dissociation equilibrium constants, and hemin dissociation rates constants.

<table>
<thead>
<tr>
<th>Mb</th>
<th>$\log(K_{IN}^a)$</th>
<th>$\log(K_{UI}^a)$</th>
<th>$\log(K_{UN}^a)$</th>
<th>$K_{NH}^b$ pM</th>
<th>$K_{HH}^b$ pM</th>
<th>$k_{H^+}$, pH 7$^c$ h$^{-1}$</th>
<th>$k_{H^+}$, pH 5$^c$ h$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dw H64F/V68F</td>
<td>3.16</td>
<td>2.26</td>
<td>5.42</td>
<td>15</td>
<td>57</td>
<td>(1.2)</td>
<td>40</td>
</tr>
<tr>
<td>Sw H64F/V68F</td>
<td>3.05</td>
<td>1.67</td>
<td>4.72</td>
<td>0.19</td>
<td>10</td>
<td>0.62 ± 0.05</td>
<td>15.4 ± 1.7</td>
</tr>
<tr>
<td>Dw H64L</td>
<td>2.46</td>
<td>2.11</td>
<td>4.57</td>
<td>0.52</td>
<td>27</td>
<td>(0.8)</td>
<td>26</td>
</tr>
<tr>
<td>Dwarf sperm whale (Dw)</td>
<td>2.29</td>
<td>1.67</td>
<td>3.97</td>
<td>0.048</td>
<td>5.5</td>
<td>0.035 ± 0.003</td>
<td>0.66 ± 0.03</td>
</tr>
<tr>
<td>Sperm whale (Sw)</td>
<td>1.68</td>
<td>1.72</td>
<td>3.40</td>
<td>0.022</td>
<td>1.6</td>
<td>0.049 ± 0.012</td>
<td>1.0$^d$</td>
</tr>
<tr>
<td>Emperor penguin$^e$</td>
<td>1.98</td>
<td>1.72</td>
<td>3.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goose-beak whale</td>
<td>1.77</td>
<td>1.85</td>
<td>3.62</td>
<td>0.045</td>
<td>4.8</td>
<td>0.044 ± 0.004</td>
<td>2.7 ± 0.3</td>
</tr>
<tr>
<td>Grey seal</td>
<td>1.92</td>
<td>1.66</td>
<td>3.58</td>
<td>0.023</td>
<td>4.4</td>
<td>0.082 ± 0.009</td>
<td>4.7 ± 1.3</td>
</tr>
<tr>
<td>Human</td>
<td>1.18</td>
<td>0.92</td>
<td>2.10</td>
<td></td>
<td></td>
<td>~0.01$^d$</td>
<td>3.0$^d$</td>
</tr>
<tr>
<td>Pig</td>
<td>0.90</td>
<td>1.11</td>
<td>2.01</td>
<td></td>
<td></td>
<td>~0.01$^d$</td>
<td>1.0$^d$</td>
</tr>
</tbody>
</table>
The unfolding experiments were done at the experimental conditions of 10 μM Mb, 10 mM potassium phosphate at 20 °C.

The final m_{NH} and m_{IH} values used in fitting the unfolding curves’ to derive the heme affinity constants are as follows: SW Mb (4.4 kcal mol^{-1}M^{-1} and 4 kcal mol^{-1}M^{-1}); Dw Mb (4.0 kcal mol^{-1}M^{-1} and 3.4 kcal mol^{-1}M^{-1}); goosebeak whale Mb (4.3 kcal mol^{-1}M^{-1} and 3.3 kcal mol^{-1}M^{-1}); grey seal Mb (5.7 kcal mol^{-1}M^{-1} and 4.7 kcal mol^{-1}M^{-1}); Dw H64F/V68F Mb (3.8 kcal mol^{-1}M^{-1} and 2.6 kcal mol^{-1}M^{-1}); Dw H64L (4.5 kcal mol^{-1}M^{-1} and 2.9 kcal mol^{-1}M^{-1}); and SW H64F/V68F Mb (4.5 kcal mol^{-1}M^{-1} and 3.2 kcal mol^{-1}M^{-1}).

Heme loss assays were done at 37°C in 0.45 M sucrose in 0.15 M buffer of either sodium phosphate (pH 7) or sodium acetate (pH 5). The values in parentheses for the Dw mutants were measured for the protein containing the L137M mutation as described in Section 4.4.

The k_{H} value at pH 5 and pH 7 was taken from Hargrove, Wilkinson, and Olson.

Since emperor penguin has cysteine at residue 108, non-native disulfide bonds were observed to occur in the denatured protein when analyzed with SDS-PAGE gel. These apoglobin unfolding results were obtained initially without the presence of thiol reducing agents and therefore the experiments need to be repeated.

### 4.4 Measurements of hemin affinity and hemin dissociation rate constants

In order to determine hemin affinities for the N and I states, holoMb unfolding was examined using the analyses developed by Culbertson and Olson and the scheme in Fig. 1.4. ApoMb unfolding parameters obtained from the independent experiments and analyses (Figure 4.3, Table 4.2, 2nd and 3rd columns) were used to analyze the holoMb unfolding curves shown in Figure 4.5 and to estimate equilibrium hemin dissociation constants for the native holo- and intermediate holo-metMb states (i.e., K_{NH} and K_{IH} in Figure 1.1 and Table 4.2). Titration curves for the unfolding of six key holo-metMb variants are shown in Fig. 4.5 where the lines represent fits to eq 2.2 using the apoMb unfolding parameters, K_{NH}, and K_{IH} values listed in Table 4.2. From the holoMb unfolding experiments, the K_{NH} values for all the naturally occurring holoMb variants were determined to be approximately ~ 10^{-14} M while K_{IH} is ~100 fold larger (~10^{-12} M). The distal heme pocket mutants had larger K_{NH} values, on the order of ~10^{-11}~10^{-12} M and K_{IH} values on the order of ~10^{-11} M.
Kinetic experiments were also performed in order to measure rates of hemin dissociation \( (k_{H}) \) at pH 5 and pH 7. At pH 7, the rates of dissociation are very slow and precipitation of the resultant apoprotein can interfere with analyses of the time courses, whereas at pH 5.0 the rate of hemin loss is much faster and easier to measure for the wild-type variants.

The distal heme pocket mutations, H64L and H64F/V68F, increased the hemin dissociation rates and equilibrium hemin dissociation constants significantly (Table 4.2). As shown previously\(^7\), filling up the heme pocket with large apolar residues increases apoglobin stability but compromises the physiological function of Mb by decreasing O\(_2\) affinity, heme affinity, and resistance to autooxidation\(^14,89,182\). The \( k_{H} \) values for the Mbs from deep diving mammals were more variable than expected. Interestingly, the naturally occurring variants with the higher apoMb stabilities sometimes did not have the lowest hemin loss rates, i.e. grey seal Mb and goosebeak whale Mb (Table 4.2).

![Figure 4.4. GdnHCl induced equilibrium unfolding of native, recombinant wild-type, and recombinant 137M Dw apo- Mbs followed by A. circular dichroism (CD), and B. fluorescence. The unfolding experiments were done at the experimental conditions of 10 μM Mb, 10 mM potassium phosphate at 20 °C as done by Culbertson and Olson.](image-url)
Figure 4.5. **GdnHCl induced equilibrium unfolding of selected holoMbs.** The solid circles are the observed data and the solid lines are fitted curves using eq 2.2 and the $K_{\text{NH}}$ and $K_{\text{H}}$ parameters listed in Table 4.2. Unfolding measurements were done with 10 $\mu$M protein in 10 mM potassium phosphate, pH 7, 20°C as in Culbertson and Olson$^9$.

### 4.5 Effects of L137M mutation in recombinant dwarf sperm whale Mb

In the study by Scott et al.$^9$, whale Mbs were obtained from the native organisms directly prior to the enactment of the endangered species act, and these samples were gifts from Dr. Frank N. Gurd and Dr. Jay Berzofsky. My study was the first attempt in the literature to express recombinant dwarf sperm whale, goosebeak whale, grey seal, and emperor penguin
Mbs in different expression systems. For expression in *E. coli* systems, the genes were constructed by IDT with codon optimization. Unfortunately, during the dwarf sperm whale Mb gene construction at IDT, the CTG codon for the apolar 137 L residue undergone a single nucleotide mutation to ATG causing the dwarf sperm whale Mb expressed in the *E. coli* system to have an apolar M137L mutation initially. The nucleotides immediately prior to this codon are rich in adenines, and, as a result, it was difficult to discriminate between A and C in the codon for the 137 residue from the sequencing results alone, and the presence of this change was originally missed.

When I compared the unfolding curves of this recombinant protein with that of native Dw Mb, obtained directly the muscle cells of the Dw whale, the curves were identical (see Fig. 4.4 and Table 4.3). The native Dw Mb sample was from protein stocks sent by Dr. Frank N. Gurd and Dr. Jay Berzofsky and had been stored in liquid nitrogen. The M137L mutation was discovered when Olson’s laboratory and our colleagues started routinely doing mass spectrometry analysis of our globin samples in 2016. We then re-sequenced the genes for all our constructs. Only the Dw gene variants used for expression in *E. coli* (and not any of those used for the cell-free system or any of the other Mb species) had this inadvertent mutation, which appears to be silent.

To examine this problem more systematically, I mutated the Met codon back to that for Leu at position 137 for all the recombinant dwarf whale Mb variants. When unfolding experiments were done with the corrected recombinant wild-type dwarf whale Mb, the results were identical to those for the native Dw Mb and the recombinant Dw Mb with the inadvertent L137M mutation (Fig. 4.4 and Table 4.2). Thus, the L137M is silent in the native protein as might be expected because both Met and Leu are apolar residues. The side chain of Leu137 is
located along the H helix in the AGH hydrophobic core (Fig. 4.1) that remains intact in the molten globule intermediate and far away the heme pocket. In retrospect, the lack of effect of the L137M is not unexpected. Whitaker et al.\textsuperscript{183} observed little or no effect of the M55L on the

Table 4.3. **Effects of the L137M mutation on apomyoglobin stabilities, and heme affinities.** Experimental conditions were the same as in Table 4.2. The parameters for native, wild-type (wt with extra N-met), and the L137M mutant (also with extra N-met) are identical within experimental error. Differences were observed for heme binding to the Dw H64F/V68F double mutant, which could be due in part to uncertainties in the fitting because of the similarities of the curves for the apo- and holoproteins which indicates very weak cofactor binding compared to the wild-type protein.

<table>
<thead>
<tr>
<th>Mb</th>
<th>log($K_{IN}$)</th>
<th>log($K_{UL}$)</th>
<th>log($K_{UN}$)</th>
<th>$K_{NH}$\textsuperscript{a}</th>
<th>$K_{IH}$\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dw</td>
<td>native</td>
<td>wt</td>
<td>wt</td>
<td>L137M native</td>
<td>L137M wt</td>
</tr>
<tr>
<td>Dw</td>
<td>2.29</td>
<td>1.67</td>
<td>3.97</td>
<td>0.048</td>
<td>5.5</td>
</tr>
<tr>
<td>Dw</td>
<td>2.32 L137M</td>
<td>1.69 L137M</td>
<td>4.02 L137M</td>
<td>0.048 L137M</td>
<td>5.3 L137M</td>
</tr>
<tr>
<td>Dw H64F/V68F</td>
<td>3.16 wt</td>
<td>2.26 wt</td>
<td>5.42 wt</td>
<td>15.00 wt</td>
<td>57 wt</td>
</tr>
<tr>
<td></td>
<td>4.33 L137M</td>
<td>1.37 L137M</td>
<td>5.70 L137M</td>
<td>0.83 L137M</td>
<td>80 L137M</td>
</tr>
<tr>
<td>Dw H64L</td>
<td>2.46 L137M</td>
<td>2.11 L137M</td>
<td>4.57 L137M</td>
<td>0.52 L137M</td>
<td>27 L137M</td>
</tr>
<tr>
<td></td>
<td>2.92 L137M</td>
<td>1.82 L137M</td>
<td>4.73 L137M</td>
<td>0.54 L137M</td>
<td>26 L137M</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The final $m_{NH}$ and $m_{IH}$ values used in the unfolding curves’ fittings to derive the heme affinity constants are as follows: Dw Mb (4.0 kcal mol\textsuperscript{-1}M\textsuperscript{-1} and 3.4 kcal mol\textsuperscript{-1}M\textsuperscript{-1}); Dw H64F/V68F Mb (3.8 kcal mol\textsuperscript{-1}M\textsuperscript{-1} and 2.6 kcal mol\textsuperscript{-1}M\textsuperscript{-1}); Dw H64L (4.5 kcal mol\textsuperscript{-1}M\textsuperscript{-1} and 2.9 kcal mol\textsuperscript{-1}M\textsuperscript{-1})

stability or ligand binding properties of Sw Mb, and in this case, the apolar side chain is located in the middle of the hydrophobic core of the CD corner.

To be thorough, I re-measured all folding and heme affinity parameters for the native Dw Mb, wild-type recombinant Dw Mb (Fig. 2.1A), and the H64L and H64F/V68F distal pocket variants of wild-type recombinant Dw Mb, and the results are shown in Table 4.3. The
results for native Dw and these recombinant distal pocket Dw mutants are presented in Table 4.2 and were used for comparisons with the cell-free expression data in this chapter.

There are some small differences seen in the folding and kinetic parameters for the distal pocket Dw Mb mutants, but no differences were observed between wild type, native, and L137M Dw apoMbs and there were no significant changes in the observed correlations with the expression data shown in Figs. 4.6, 4.7, or 4.9 when either the wild-type or L137M variant parameters were used. Dw apoMb is still the most stable naturally occurring variant, and the Dw distal pocket mutants show quantitatively the same dramatic increases in overall apoglobin stability ($K_{UN}$) and large decreases in heme affinity.

4.6 Direct correlations between holoMb expression and apoMb stability ($logK_{UN}$) but not with heme affinity

As shown in Fig. 4.6, there is a strong linear correlation between the expression yields of soluble holoMb measured using eq 3.1 and the logarithm of overall apoglobin stability as measured by $log(K_{NL})$ or $log(K_{UN})$. The stability of the apo-N state relative to the intermediate molten globule, as measured by $-log(K_{NI})$, has a more significant effect on expression yield than the relative stability of the apo-I state, as measured by $-log(K_{II})$ (Figs. 4.6A and 4.6B). This observation is consistent with structural interpretations of the apoMb I state, in which most of the heme pocket is highly disordered and hemin is more weakly bound, forming a non-specific hemichrome complex. Thus, the stability of the native (N) apoMb state is more critical for high expression of the holoprotein than the stability of the molten globule state.
Figure 4.6. Correlation between holoMb cell-free expression yield and A, overall apoMb stability (-log(K_{NU})); B, the stability of the apoMb native (N) state relative to the intermediate (I) state ((-log(K_{NI})); C, the stability of the apoMb intermediate (I) state relative to the unfolded (U) state (-log(K_{IU})). The apoMb equilibrium unfolding constants for all species (Table 4.2) were measured with 10 μM protein in 10 mM potassium phosphate, pH 7 at 20°C. The overall apoMb stability is empirically represented as (-log(K_{NU})) which is equivalent to log(K_{UN}) and proportional to free energy of folding. The expression yields were calculated using eq 3.1, and blank reactions with added hemin but no mRNA gave values of ~0.06 as shown by the gray dashed line in panel A. The points in panel B following the Mb variant labeling in panel A. In panel C the two points at low values of −log(IU) correspond to human and pig apoMbs (and the cluster of points at high −log(IU) correspond to the 7 whale and seal Mb variants.
Figure 4.7. Correlation between holoMb cell-free expression yield and -log (K_{NH}). K_{NH} was determined from the analyses of the holoMb unfolding experiments done with 10 μM protein in 10 mM potassium phosphate, pH 7 at 20°C.

Interestingly, there was no positive correlation between hemin affinity and cell-free expression yields (Fig. 4.7). The distal pocket, H64F/V68F and H64L mutants of dwarf sperm whale and sperm whale Mbs have the highest expression yields (Table 3.1) and equilibrium folding constants, but also the largest hemin dissociation rates (k_{H}) and lowest hemin affinities (Table 4.2). The expression yields of the more stable, naturally occurring holoMbs from deep diving mammals are significantly higher than those of the less stable holoMbs from the terrestrial mammals (Fig. 4.6). Again, these quantitative observations verify previous qualitative observations of the difficulty of expressing of pig and human Mbs in *E. coli* in fully folded holooglobin forms because these less stable proteins either undergo proteolytic degradation or aggregate very rapidly to form insoluble inclusion bodies.

These results are also consistent with Smith's and others’ work on expression of holoMb variants in *E. coli* [39,41,42,71], where both qualitative and semi-quantitative measurements suggested that apoglobin stability was a key factor in governing expression yields. In this
earlier work, however, the correlations were much poorer and the scatter in the observed data much larger (Fig. 4.8). The results in Fig. 4.8 indicate that the cell-free expression assay has decreased the variability in expression yields enough to observe stronger correlations with key biochemical parameters. However, the conclusions are the same. Apoglobin stability ($K_{UN}$) is the key determinant of holoMb expression yield.

![Graph](image)

**Figure 4.8. Heterologous expression of holoMb and correlations with overall apoMb stability.** A, *In vivo* expression of Sw Mb mutants in *E. coli* relative to expression of wild-type Sw Mb. Gray and black circles represent apoMb unfolding data measured at low (20 mM) and high (200 mM) potassium phosphate concentrations respectively, at 25°C, pH 7.0 with 5 μM apoMb. Data were taken from Smith. B, Relative expression of mammalian Mbs in the cell-free/wheat germ system. The relative expression values in panel B were calculated using eq 3.2 and are also given in Tables 3.1. Approximate concentration in mg/(g of wet tissue) of native Mbs found in the muscles of terrestrial (4-8 mg/g) and deep diving mammals (40-70 mg/g) are given in parenthesis in red and were taken from Mirceta et al. The apoMb equilibrium unfolding constants for all species were measured with 10 μM protein in 10 mM potassium phosphate, pH 7 at 20°C. The overall apoMb stability is given by log($K_{UN}$), which is proportional to the free energy released during folding from the unfolded (U) to the native (N) state.
4.7 Mechanistic interpretation of cell-free expression

Proposed models for cell-free expression of holoMb are shown in Fig. 4.9 and all the rate constants discussed below are as described in Fig. 4.9. The rate constant for translation, \( k_{\text{translation}} \), in Fig. 4.9 should be zero order until all the translation cofactors and nutrients are consumed. The newly formed unfolded apoMb (U) will will either fold into the native folded apoMb (N) state with a folded heme pocket to bind free hemin (H) and form the stable holo- metMb product (NH) or the U states will aggregate and precipitate irreversibly into a separate insoluble phase (Figure 4.9A). In these models, hemin binding to the holoMb is assumed to be effectively irreversible. At pH 7-8, the rate of hemin dissociation is extremely slow (\( t_{1/2} \approx 1 \) to 20 hours) and the \( K_d \) for hemin binding is on the order of \( 10^{-12} \) to \( 10^{-14} \) M (Table 4.2). Native holoMbs have been shown to be highly resistant to precipitation and proteolysis, whereas the apo U state, which has only residual or transient local helical structure, is prone to rapid self-aggregation and precipitation through its various extended conformations. The equilibrium-folding constant from the U to N state of apoMb is described as \( K_{UN} \) (\( K_{UN} = N/U \)) for the models shown in Fig. 4.9A, and can be used to derive the population fractions of folded (\( Y_N = \frac{[N]}{[U]+[N]} = \frac{K_{UN}}{1+K_{UN}} \)) and unfolded (\( Y_U = \frac{[U]}{[U]+[N]} = \frac{1}{1+K_{UN}} \)) states.
Figure 4.9. A, HoloMb expression models for both unimolecular precipitation and bimolecular aggregation. B, Comparison of observed and fitted results for relative expression versus log(K_{UN}). In the simple expression models, folding rates are assumed very large compared to hemin binding, precipitation, and aggregation rates. The dotted and dashed lines represent the fittings to the simple rapid folding equilibrium expression models for unimolecular precipitation (eq 4.2) and bimolecular aggregation (eq 4.3), respectively.

In the simplest model, the rate constants, as described in Fig. 4.9A, for folding to the apoMb native N state (k_N) and unfolding to the apoMb U state (k_U) are assumed to be much larger than the rate constants for hemin binding (k'_{H}[H]) and precipitation (k_{ppt}). Under these conditions, the rate of holoMb formation, d[holoMb]/dt, can be derived by using the population fraction of the N state (Y_N):

\[ \frac{d[\text{holoMb}]}{dt} \approx \frac{k_N}{k_N + k_U} \]
A steady state approximation is then assumed for the total concentration of translated apoMb ([U] + [N]) during the expression process.

If the precipitation (ppt) process is first order, then rate of U precipitation, d[ppt]/dt, is determined as $k_{ppt}[U]$ or $k_{ppt}(1/(K_{UN}+1))[apoMb]$, and [apoMb] in the steady state can be solved as follows in eq 4.2:

\[
\begin{align*}
\left( \frac{d[apoMb]}{dt} \right)_{\text{ steadystate }} &= 0 = k_{\text{translation}} - \frac{d[holoMb]}{dt} - \frac{d[ppt]}{dt} \\
\left( \frac{d[apoMb]}{dt} \right)_{\text{ steadystate }} &= 0 = k_{\text{translation}} \left( k'_{H}[H] \frac{K_{UN}}{1+K_{UN}} [apoMb] + k_{ppt} \left( \frac{1}{1+K_{UN}} \right) [apoMb] \right) \\
0 &= k_{\text{translation}} \left( k'_{H}[H] \frac{K_{UN}}{1+K_{UN}} [apoMb] + k_{ppt} \left( \frac{1}{1+K_{UN}} \right) [apoMb] \right) \\
\left( k'_{H}[H]K_{UN} + k_{ppt} \right) [apoMb] &= k_{\text{translation}} \left( 1 + K_{UN} \right) \\
\left( k'_{H}[H]K_{UN} + k_{ppt} \right) &= k_{\text{translation}} \left( 1 + K_{UN} \right) \\
\left[ apoMb \right] &= \frac{k_{\text{translation}} \left( 1 + K_{UN} \right)}{k'_{H}[H]K_{UN} + k_{ppt}}
\end{align*}
\]

(eq 4.2)

For the holoMb expression model with unimolecular precipitation in Fig 4.9, the steady state expression for the rate of holoMb formation, d[holoMb]/dt, (eq 4.3) can be derived using eqs 4.1 and 4.2:

\[
\frac{d[holoMb]}{dt} = k'_{H}[H][N] = k'_{H}[H]Y_N [apoMb] = k'_{H}[H] \left( \frac{K_{UN}}{1+K_{UN}} \right) [apoMb]
\]

(eq 4.1)
Thus, the rate of holoMb formation should depend hyperbolically on $K_{UN}$, with the exact dependence being determined by the ratio of the rate constants for precipitation and hemin binding ($k_{ppt}/k'_{H}[H]$) in denominator of eq 4.3. Assuming that this rate of holoMb formation applies throughout the assay, the relative expression yield can be computed from $d[\text{holoMb}]/dt$ for a variant with a higher or lower $K_{UN}$ value, divided by the computed $d[\text{holoMb}]/dt$ for sperm whale Mb. The dotted line in Figure 4.9B shows a fit of the observed relative expression data to eq 4.3 and indicates that, in principle, the rate of holoMb formation and the relative expression yield should not show a strictly linear dependence on $\log(K_{UN})$ but reach a limiting value equal to $k_{\text{translation}}$ at very high $K_{UN}$ values.

This non-linear dependence on $\log K_{UN}$ (Fig. 4.9B) will also occur if the aggregation process is modeled more realistically as a bimolecular process with a second order rate constant $k'_{\text{aggregation}}$ (as described in Fig. 4.9A), which is what most workers in the field assume (see 194). The rate of aggregation, $d[\text{aggregation}]/dt$, then can derived in terms of the population fraction of the unfolded state, $Y_{U}$, as $k'_{\text{aggregation}} (Y_{U} [\text{apoMb}])^2 = k'_{\text{aggregation}} \left( \frac{1}{1+K_{UN}} \right)^2 [\text{apoMb}]^2$.

In this case, the steady state equation for $[\text{apoMb}]$ is a quadratic (eq 4.4).
\[ 0 = k_{\text{translation}} \left( k'_H[H] \frac{K_{UN}}{1+K_{UN}} [\text{apoMb}] \right) + k'_{\text{aggregation}} \left( \frac{1}{1+K_{UN}} \right) [\text{apoMb}]^2 \]

\[ k'_{\text{aggregation}} \frac{1}{1+K_{UN}} [\text{apoMb}]^2 + k'_H[H] \frac{K_{UN}}{1+K_{UN}} [\text{apoMb}] - k_{\text{translation}} = 0 \]

\[ [\text{apoMb}] = \frac{-k'_H[H]K_{UN} + \sqrt{(k'_H[H]K_{UN})^2 + 4 \left( k'_{\text{aggregation}} \frac{1}{1+K_{UN}} k_{\text{translation}} \right) (1+K_{UN})}}{2 k'_{\text{aggregation}}} \]

\[ [\text{apoMb}] = \frac{\left(1+K_{UN}\right) \left[ (k'_H[H]K_{UN})^2 + 4 k'_{\text{aggregation}} k_{\text{translation}} - k'_H[H]K_{UN} \right]}{2 k'_{\text{aggregation}}} \]

(eq 4.4)

For the holoMb expression model with bimolecular precipitation in Fig. 4.9, the steady state expression for the rate of holoMb formation, \( \frac{d[\text{holoMb}]}{dt} \), (eq 4.5) is derived using eqs 4.1 and 4.4.

\[ \frac{d[\text{holoMb}]}{dt} = k'_H[H] \frac{K_{UN}}{1+K_{UN}} [\text{apoMb}] \]

\[ \frac{d[\text{holoMb}]}{dt} = k'_H[H] \frac{K_{UN}}{1+K_{UN}} \left(1+K_{UN}\right) \left[ (k'_H[H]K_{UN})^2 + 4 k'_{\text{aggregation}} k_{\text{translation}} - k'_H[H]K_{UN} \right] \]

(eq 4.5)

\[ \frac{d[\text{holoMb}]}{dt} = k'_H[H]K_{UN} \left( \sqrt{(k'_H[H]K_{UN})^2 + 4 k'_{\text{aggregation}} k_{\text{translation}} - k'_H[H]K_{UN}} \right) \]

This more complex expression also asymptotically approaches \( k_{\text{translation}} \) at very high \( K_{UN} \), fits the observed data, and is shown in Fig. 4.9B as the dashed line. This square root function shows a steeper dependence on \( K_{UN} \) and an increased differential expression between...
the more stable variants and the less stable ones. For both the unimolecular precipitation and bimolecular aggregation models, \( k_{\text{ppt}} \) and \( k'_{\text{aggregation}} \) have to be greater than the effective rate of hemin binding \( (k'_H[H]) \) to obtain the almost linear dependence on \( \log(K_{UN}) \) observed experimentally.

Similar expressions are obtained if the rates of folding and unfolding are considered in the analysis, keeping the definition \( K_{UN}=k_N/k_U \) (Fig. 4.9A). Then, \( d[N]/dt \) and \( d[U]/dt \) can be described as \( k_N[U] \) and \( k_c[N] \) respectively. In these derivations, we assume steady state expressions for \([U]\) and \([N]\), which are derived in eqs 4.6 and 4.7 for unimolecular precipitation and bimolecular aggregation of \( U \) respectively.

\[
\begin{align*}
\left( \frac{d[N]}{dt} \right)_{\text{steady state}} &= 0 = k_N[U] - (k_U[N] + k'_H[N][H]) \quad \text{eq 4.6} \\
[N] &= \frac{k_N[U]}{(k_U + k'_H[H])} \\
\left( \frac{d[U]}{dt} \right)_{\text{steady state}} &= 0 = k_{\text{translation}} + k_U[N] - (k_N + k_{\text{ppt}})[U] \\
[U] &= \frac{k_{\text{translation}} + k_U[N]}{k_N + k_{\text{ppt}}} \\
[U] &= \frac{k_{\text{translation}} k_U + k_{\text{translation}} k'_H[H] + k_U k_N[U]}{(k_N k_U + k_N k'_H[H] + k_{\text{ppt}} k_U + k_{\text{ppt}} k'_H[H])} \\
[U] &= \frac{k_{\text{translation}} (k_U + k'_H[H])}{(k_N k'_H[H] + k_{\text{ppt}} k_U + k_{\text{ppt}} k'_H[H])} \\
[N] &= \frac{k_{\text{translation}} k_N}{(k_N k'_H[H] + k_{\text{ppt}} k_U + k_{\text{ppt}} k'_H[H])}
\end{align*}
\]
\[
\left( \frac{d[N]}{dt} \right)_{\text{steady state}} = k_N[U] - k_U[N] - k'_H[H][N] = 0 \quad \text{eq 4.7}
\]

\[
[N] = \frac{k_N[U]}{k_U + k'_H[H]}
\]

\[
\left( \frac{d[U]}{dt} \right)_{\text{steady state}} = k_{\text{translation}} + k_U[N] - k_N[U] - k'_{\text{aggregation}}[U]^2 = 0
\]

\[
k'_{\text{aggregation}}[U]^2 + k_N[U] - k_{\text{translation}} - \frac{k_U k_N[U]}{k_U + k'_H[H]} = 0
\]

\[
[U] = \frac{-\left( \frac{k_N k'_H[H]}{k_U + k'_H[H]} \right) + \sqrt{\left( \frac{k_N k'_H[H]}{k_U + k'_H[H]} \right)^2 + 4k'_{\text{aggregation}}k_{\text{translation}}}}{2k'_{\text{aggregation}}}
\]

\[
[N] = \frac{k_N}{2k'_{\text{aggregation}}(k_U + k'_H[H])} \left( \sqrt{\left( \frac{k_N k'_H[H]}{k_U + k'_H[H]} \right)^2 + 4k'_{\text{aggregation}}k_{\text{translation}}} - \frac{k_N k'_H[H]}{k_U + k'_H[H]} \right)
\]

Using eq 4.6, the rate for holoMb formation for the holoMb expression model with unimolecular precipitation of U when the folding/unfolding rates are considered is described in eq 4.8.

\[
\frac{d[\text{holoMb}]}{dt} = k'_H[H] \frac{k_{\text{translation}} k_N}{(k_N k'_H[H] + k_{\text{ppt}} k_U + k_{\text{ppt}} k'_H[H])}
\]

\[
\frac{d[\text{holoMb}]}{dt} = \frac{k_{\text{translation}} K_{\text{UN}} k'_H[H]}{K_{\text{UN}} k'_H[H] + k_{\text{ppt}} \left( 1 + \frac{k'_H[H]}{k_U} \right)} \quad \text{eq 4.8}
\]

\[
\frac{d[\text{holoMb}]}{dt} = \frac{k_{\text{translation}} K_{\text{UN}}}{K_{\text{UN}} + \frac{k_{\text{ppt}}}{k'_H[H] k_U}} \left( \frac{k_{\text{ppt}}}{k_U + k'_H[H]} \right)
\]
Similarly, the rate for holoMb formation for the holoMb expression model with bimolecular precipitation of U when the folding/unfolding rates are considered is described in eq 4.9 by using eq 4.8.

\[
\frac{d[\text{holoMb}]}{dt} = \frac{k_Nk'_H[H]}{2k'_{\text{aggregation}}(k_U+k'_H[H])} \left( \frac{k_Nk'_H[H]}{k_U+k'_H[H]} \right)^2 + \frac{4k'_{\text{aggregation}}k_{\text{translation}}}{k_U+k'_H[H]} - \frac{k_Nk'_H[H]}{k_U+k'_H[H]}
\]

\[
\frac{d[\text{holoMb}]}{dt} = \frac{K_{UN}k'_H[H]}{2k'_{\text{aggregation}}(1+k'_H[H]/k_U)} \left( \frac{K_{UN}k'_H[H]}{1+k'_H[H]/k_U} \right)^2 + \frac{4k'_{\text{aggregation}}k_{\text{translation}}}{1+k'_H[H]/k_U} - \frac{K_{UN}k'_H[H]}{1+k'_H[H]/k_U}
\]

\[
\frac{d[\text{holoMb}]}{dt} = \frac{k_Uk'_H[H]}{(k_U+k'_H[H])}K_{UN} \left( \frac{k_Uk'_H[H]}{k_U+k'_H[H]} \right)^2 + \frac{4k'_{\text{aggregation}}k_{\text{translation}}}{k_U+k'_H[H]K_{UN}} - \frac{k_Uk'_H[H]}{k_U+k'_H[H]K_{UN}}
\]

\[
= \frac{2k'_{\text{aggregation}}}{k_Uk'_H[H]}K_{UN}
\]

\[
\text{eq 4.9}
\]

In both cases (eqs 4.8 and 4.9), the forms of the expression are the same as the corresponding equations for rapid interconversion of the U and N states. What changes is the effective rate of hemin binding, which is attenuated by \( k_U/(k_U+k'_H[H]) \). This expression describes the competition between the hemin binding to the N state and its unfolding back to the U state. When \( k_U \) becomes very large, the rapid equilibrium assumption for the N and U states applies, and eqs 4.8 and 4.9 are reduced to eqs 4.3 and 4.5, respectively. When \( k_U \) becomes very small, the rate of folding becomes limiting for holoMb formation and eq 4.8 reduces to \( d[\text{holoMb}]/dt = k_{\text{translation}}K_{UN}/(K_{UN}+k_{\text{pp}}/k_U) \) or \( k_{\text{translation}}K_N/(k_N+k_{\text{pp}}) \) because \( k_UK_{UN} = k_N \). Thus, even when the rates of folding are taken into account, a hyperbolic-like dependence
on $K_{\text{UN}}$ will occur and the rate of aggregation must still be very large compared to the effective hemin binding rate in order to fit the observed data.

The theoretical analyses in eqs 4.1-4.9 and in Fig. 4.9 are clearly first approximations. The I state of apoMb and the IH hemichrome state were not considered in our models to keep them simple. Wild-type apo- and holoMb unfolding experiments have shown that the I and IH states are not highly populated and their detection can be difficult and requires multiple spectral measurements. The I and IH states also rapidly appear and decay in kinetic folding experiments. Inclusion of intermediates will alter the shape of the dependence of relative expression on log($K_{\text{UN}}$) but would not change the asymptotic nature of the process nor change the conclusion that aggregation of the unfolded state is very fast.

Our models emphasize the importance of increasing both the rate of translation by adding more amino acids, ATP, and other metabolites and engineering greater apoprotein stability ($K_{\text{UN}}$) to enhance hologlobin expression yields. The availability of free hemin is important, both from a stoichiometry point of view and in terms of the rate of hemin binding versus aggregation, but in most cases, the hemin affinity is so great ($K_d \leq 10^{-11}$ M, Table 4.2) that it does not directly influence the expression yields (Fig. 4.7).

Another key result from these analyses is that the rate of precipitation must be roughly 2500 times faster than the rate of hemin binding in order to fit our observed data to eq 4.3. A similar, very large ratio is required for the bimolecular aggregation mechanism in eq 4.5 (i.e. $k'_{\text{aggregation}}/k_{\text{H}}[H] \approx 5 \times 10^9$ M). Free hemin forms dimers with a $K_d \approx 2-3 \times 10^{-7}$ M at pH 7. Free hemin forms dimers with a $K_d \approx 2-3 \times 10^{-7}$ M at pH 7. Free hemin forms dimers with a $K_d \approx 2-3 \times 10^{-7}$ M at pH 7. Free hemin forms dimers with a $K_d \approx 2-3 \times 10^{-7}$ M at pH 7. Thus, the effective first order rate of hemin binding to apoMb is given by the rate of dimer dissociation. Hargrove et al. have shown that in the μM concentration range, the rate of hemin binding to apoMb is a first order process with a rate constant $\approx 10$ s$^{-1}$. Using this first
order rate as the effective value of $k'_H[H]$, $k_{ppt}$ in eq 4.3 would be $25,000 \text{ s}^{-1}$ (assuming fast protein conformational transitions), and $k'_{\text{aggregation}}$ in eq 4.5 would be $\sim 10^{10} \text{ M}^{-1}\text{s}^{-1}$, which is a reasonable estimate for a diffusion controlled bimolecular aggregation process and similar to rates of apoMb aggregation estimated in previous studies.\footnote{196}

Alternatively, the strong dependence of expression levels on $K_{\text{UN}}$, even when the values are very high, could be due to slow rates of folding compared to heme binding. Under these conditions, the ratios $k_{\text{ppt}}/k_U$ and $k'_{\text{aggregation}}/k_U$ determine the observed dependence of rate of holoMb formation on $K_{\text{UN}}$. However, even if the value of $k_U$ was on the order of only $1 \text{ s}^{-1}$, the values of $k_{\text{ppt}}$ and $k'_{\text{aggregation}}$ would still be very large and on the order of $250,000 \text{ s}^{-1}$ and $10^{10} \text{ M}^{-1}\text{s}^{-1}$ respectively.

\section*{4.8 Conclusions}

The results for cell-free expression of holoMb in this study confirm unambiguously that apoglobin stability is the key factor governing heterologous expression, and similar correlations occur for Mb expression in muscle tissues of mammals.\footnote{42} In contrast, hemin affinity does not appear to be an important factor as long as there is a threshold of minimum $K_d$ maintained. These conclusions likely apply to all globins, given that a sequence alignment of 728 globins showed that 13 amino acids related to heme binding are conserved across the globins, and 6 of the residues are directly involved in heme binding.\footnote{1} As shown in Figs. 4.1 and 4.2, the heme pocket cavity structure with bound heme is conserved across the globins regardless of apoglobin stability, and this structure likely determines the $K_{\text{NH}}$ threshold. On the other hand, apoglobin structure might have some variability across the different native variants,
given my preliminary CD analysis showing variants such as native Dw apoMb having a somewhat higher degree of helical content compared to Sw apoMb (data not shown). However, as Fig. 4.1 shows, the static folded structure alone cannot predict the overall stability parameters.

In my current analyses, the ratio of the rates of precipitation or aggregation and hemin binding were assumed to the same for all the variants examined, and only the independently measured value of $K_{\text{UN}}$ was allowed to vary. However, the increased surface charge of the native Mbs from deep diving mammals could be playing a role by decreasing $k'_\text{aggregation}$ (or $k_{\text{ppt}}$), and these effects could partially explain the large differences in relative expression of pig and human Mbs ($Z \leq +1$) versus the whale and seal Mbs ($Z \geq +4$) 42. The increased $Z_{\text{Mb}}$ is probably also increasing $K_{\text{UN}}$ through favorable surface electrostatic interactions that stabilize $\alpha$-helices 197-199. The higher surface charge on the whale Mbs probably also prevents crystallization and precipitation of the folded holoMbs when they accumulate to high concentrations in myocytes.

Regardless of the exact interpretation of $Z_{\text{Mb}}$, the marked increases in relative expression of the H64L and H64F/V68F distal pocket mutations for both sperm whale and dwarf sperm whale Mb suggest that overall folding stability, driven in these cases by a more apolar heme pocket, is the dominant parameter in the heterologous cell-free and E. coli expression systems (Fig. 4.6, Tables 3.1 and 4.2). These observations support the idea that the hydrophobic effect is a major driving force for folding to the native apoprotein state 200. The results for these distal pocket mutants also show that the presence of His64 clearly compromises apoglobin stability in favor of functionality as an $O_2$ storage protein. Even though they express well, the H64F and H64L variants bind $O_2$ poorly, autooxidize quickly, and lose heme rapidly 14,89,182.
Perhaps more importantly, this study demonstrates that the cell-free hologlobin expression assay is an ideal platform for high throughput screening of large libraries of any heme protein of interest, from Hb-based oxygen carriers to NO, CO, and O$_2$ gas sensors, in order to explore apoprotein stability as measured by increased holoprotein yields. This assay can also be adapted to examine other properties of heme proteins, including ligand binding, autooxidation, and hemin dissociation. For example, a reducing agent could be added to the translation reaction mixture and then removed during purification to allow generation of ferrous samples for O$_2$, binding and autooxidation measurements using a microplate spectrophotometer. Similarly, prolonged incubation of partially purified ferric forms could be used to follow hemin dissociation, either by the loss of Soret absorbance or by the addition of a heme-scavenging reagent such as H64Y/V68F apoMb$^{104}$.
Chapter 5

Human apoHbA unfolding models

This chapter is adapted and reproduced from my following publication:


In order to avoid redundancy, I have minimized and/or omitted further references to these publications within the text body.

In order to define the factors that govern human hemoglobin expression and assembly, unfolding mechanisms for both the apoHbA heterodimer and tetrameric holoHbA have to be established and analyzed quantitatively and simultaneously. As described in Figure 1.4, a 6-state unfolding mechanism for holoMb was derived, based on the 3-state apoMb unfolding scheme, and used to estimate heme affinity constants for the native, partially unfolded, and unfolded states of apoMb. This mechanism was then used successfully to examine the structural features that regulate holoMb expression *in vivo* and *in vitro*, as was described in Chapters 3 and 4. In order to extend this analysis to HbA folding and assembly, the added complexity of hetero-dimer and tetramer formation had to be taken into account. The overall goal of the work in this chapter was to establish a general mechanism for human apoHb unfolding by using GdnHCl to induce reversible unfolding and then measuring the accompanying loss of secondary structure by CD spectroscopy. To my knowledge, no one has
previously tried to measure reversible unfolding of apoHb, nor has anyone proposed and then verified a structural mechanism similar to the one that applies to apoMb. Establishing a mechanism for apoHb is required for incorporation into a larger scheme for the folding and assembly of the holoprotein that can be used to examine the effects of hemin binding on the various intermediate folding states observed in vitro and during erythropoiesis.

5.1 Basic models for dimeric apoHbA unfolding

Previous studies\textsuperscript{44,49,50} and the results in this work have established that human apoHbA is a $\alpha_1\beta_1$ heterodimer and that formation of a tetramer requires hemin binding. A key mechanistic question for human apoHbA unfolding is whether or not a molten globule occurs and is similar to the monomeric myoglobin unfolding intermediate. If it does form, a second issue is whether this intermediate occurs before or after dissociation of human apoHbA dimers into monomers. There are three basic models for unfolding of a apoHbA heterodimer: (1) partial unfolding of the dimer (D) to generate a hetero-dimeric intermediate state (I_D) followed by a second phase involving I_D dissociation into 2 unfolded monomers, with each unfolded monomer (U_M) representing either the $\alpha$ or $\beta$ subunit (Fig. 5.1A); (2) dissociation of D into 2 partially unfolded, intermediate monomers (2I_M) followed by a second phase representing complete unfolding of each I_M to U_M (Fig. 5.1B); and (3) dissociation of the D into 2U_M in a single bimolecular dissociation process (Fig. 5.1C). The simulations of these models are shown in Fig. 5.1 and were carried out as described by Culbertson and Olson’s theories\textsuperscript{9,30} to show how protein concentration should affect GdnHCl-induced unfolding of apoHb dimers.
Figure 5.1. Basic models of equilibrium unfolding curves for ApoHbA heterodimer (D). A, Model 1 (biphasic) – D unfolds through a dimeric intermediate folding state (I_D) into 2 completely unfolded monomers (2U_M). B, Model 2 (biphasic) – D dissociates into two monomeric molten globules (2I_M) which then completely unfold into 2U_M. C, Model 3 (monophasic) - D directly dissociates into two unfolded monomers, 2U_M, without any intermediates. These mechanisms were originally proposed by Culbertson and Olson^9,30.^
The simulations in Fig. 5.1 used expressions similar to those described in section 5.6 with detailed analyses. In order to distinguish between these initial mechanisms, I examined the unfolding of apoHbA at protein concentrations ranging from 1 µM to 140 µM.

5.2 Initial optimization of reversible unfolding experiments with apoHb

ApoHb was initially prepared from ferric samples as described in Chapter 2. The apoHb concentration was determined using $\varepsilon_{280} = 12.7 \text{ cm}^{-1} \text{ mM}^{-1}$ per subunit for apoHbA and the other apo-rHb variants based on HbA protein sequence. Heme disassociation from holoHbA is coupled with an ~20% loss of $\alpha$-helicity, presumably involving part of the heme pocket and including the F-helix (see Chapter 6 and 7). In addition, apoHbA is unstable at room temperature due to the intrinsic protein disorder and readily precipitates, making measurements near room temperature impossible.

All measurements were made at 10° C, and great care was needed to ensure that samples remained at or below this temperature. Even with these precautions, the poor stability of apoHbA was the major cause of scatter in my initial unfolding data points because irreversible denaturation can occur before the denaturant is high enough to solubilize unfolded states. The final apoHbA unfolding conditions were optimized to be at 10° C in 200 mM potassium phosphate, pH 7, buffer with a 1 hour incubation time before spectral measurements were made. High buffer concentration was chosen because high ionic strength has an overall stabilizing effect on apoglobins, especially for the molten globule intermediates. GdnHCl was used as the denaturant because it prevents precipitation of dissociated hemin and allows formation of reversible of hemichromes, whereas hemichrome precipitates following
denaturation with urea⁵. As result, GdnHCl titrations of both holo- and apoHbA can be done under reversible conditions and allow combined analyses to obtain hemin binding parameters for each folding intermediate, as described in Chapter 7. Final stock solutions of 6 M GdnHCl were prepared in 200 mM potassium phosphate, pH 7, taking care to initially re-adjust the pH after addition of the denaturant. [GdnHCl] was varied from 0 to 4.8 M.

Unlike apoMb, apoHbA did not show a bell-shaped increase and decrease in Trp fluorescence during GdnHCl-induced unfolding⁹,¹³,³⁹,⁴¹. Instead, only very small changes in fluorescence were observed and difficult to analyze quantitatively. As result, I did not try to measure or analyze fluorescence spectra as function of [GdnHCl].

CD spectra were recorded for each fully equilibrated apoHb:GdnHCl mixture, using a Jasco J-810 CD spectropolarimeter. The extent of unfolding of secondary structure was monitored as a decrease in negative ellipticity at 222 nm (peak for α helical secondary structure). Despite my efforts at optimization, my initial apoHbA unfolding measurements showed significant variability, particularly for the second phase of unfolding at higher protein concentrations (Fig. 5.2). Due to this variability, it was difficult to determine the extent of protein concentration dependence observed during either the first or second phase of unfolding.
Figure 5.2. Unfolding of apoHbA observed with significant variability. Unfolding measurements were done at 10°C in 200 mM potassium phosphate, pH 7. The solid circles were the observed CD measurements while the dashed lines were fittings of apoHbA measurements attempted to Model 1 as described in Fig. 5.1A and in section 5.6 later. Global fitting for apoHb A was done to data points at protein concentrations of 1.5 µM, 7 µM, and 21.9 µM. The data points for apoHbA unfolding at 50 µM (blue circles) and 99 µM (purple circles) concentrations were not included in the global fitting due to the increased scatter and overlap with unfolding data points at lower protein concentrations. However, it was difficult to determine which unfolding model in Fig. 5.1 was correct from this set of data due to the variability, although it was clear that two phases occur.
5.3 Non-native disulfide formation in apoHb samples

One possible factor contributing to the variability observed in my initial unfolding measurements (Fig. 5.2) and seen in previous work initially done by Phillip Graves\(^6\), a former graduate student in the Olson laboratory, was the formation of non-native disulfide linkages in the apoHb samples. Such covalent crosslinking would lead to formation of irreversible, non-native oligomers. Hence, crosslinking between apoHb dimers would be promoted at higher protein concentrations. To examine this possibility, apoHb samples were analyzed using analytical gel filtration, with and without a disulfide reducing agent as described in section 2.5.

HbA and HbF have sulfhydryl groups at β Cys93 and γ Cys93, respectively, which are located at the α₁β₂ and α₁γ₂ interfaces, and at α Cys104 and β Cys112, which are located in the α₁β₁ and α₁γ₁ dimer interfaces. Because apoHbA exists predominantly as an α₁β₁ dimer\(^{21,44,49,50}\), the β and γ Cys93 thiol groups are exposed and will be able to form disulfides even in the native, folded apodimer. The other thiol side chains are exposed during dissociation of the dimer interface into monomers, which could lead to even more crossed linked states at high [GdnHCl]. Furthermore, these buried thiols groups could also be exposed during the apoglobin sample preparation itself since the heme extraction with organic solvent at low pH is a procedure that may temporarily lead to partial unfolding of the apoHb dimer samples before overnight dialysis in 10 mM potassium phosphate, pH 7 at 4°C.

To calibrate the elution profiles in my gel filtration column, I examined metHbA at high and low concentrations where tetramers and dimers, respectively, are the major oligomeric species. The value of equilibrium constant of the dissociation of tetramers into dimers, K\(_{4,2}\), has been reported to be \(~10\) µM for metHbA\(^{21}\). Concentrated metHbA (final concentration = 82 µM) eluted from the gel filtration column at 13.67 ml, which should
represent primarily a tetrameric species (MW = 64.6 kD; Fig. 5.3A). For dilute metHbA at a final elution concentration of 0.52 µM, the elution peak shifted to 14.17 ml due to the presence of > 85% dimers. Monomeric apomyoglobin (MW=17.3 kD) eluted at ~14.93 ml, irrespective of protein concentration. These positions serve as approximate standard elution volumes for tetramers, dimers, and monomers, respectively.

When my initial apoHbA samples were loaded on the gel column at different concentrations, there were always two prominent elution peaks, one at ~13.23 ml and another at ~14.44 ml (Fig. 5.3B). The 14.44 ml elution peak was always the major peak and represents a dimer species (apo-αβ, MW = 31 kD). The smaller and more variable 13.23 ml elution peak appears to represent a tetrameric species. Interestingly, these 2 peaks for apoHbA were completely distinct. If the dimers and tetramers had been in equilibrium, a single broad elution peak would have appeared somewhere between the 13.23 ml and 14.44 ml and been dependent on the initial apoHb concentration\textsuperscript{22,202}. When apo-rHb0.1 was loaded onto the gel filtration column, the major elution peak appeared at ~13.50 ml (Fig. 5.3C), demonstrating that this apoprotein was a tetramer. However, a significant amount of this recombinant protein eluted with a peak at ~12.43 ml (Fig. 5.3C), which probably represents disulfide crosslinked tetramers (\textit{i.e.} octamers in terms of Hb subunits).

The apoHbA stock sample for Fig. 5.3B was incubated with 50 mM TCEP (tris(2-carboxyethyl)phosphine), a disulfide reducing agent, at 10 °C for 30 minutes and then loaded onto the gel filtration column. In this case, the higher molecular weight peaks at ~13.23 ml for apoHbA became significantly diminished and the smaller peaks near the void volume disappeared (Fig. 5.3D). Similarly, after apoHbF was treated with reducing agent, only a single elution peak was observed and represented a dimer. These results indicate that a reducing agent
should be included during apoHb sample preparations and the unfolding titration experiments as described in next section. Dithiothreitol (DTT) was chosen because TCEP is unstable in phosphate buffers\(^{203}\), and at 1 mM concentration, DTT does not interfere with the far UV CD signals\(^{204}\). This reducing agent was not needed during almost all of the apoMb sample preparations because all of the Mb variants examined, except the emperor penguin Mb, do not have cysteine residues.

Figure 5.3. **Analytical gel filtration column analysis of HbA and rHb0.1 without (A,B,C) and with(D,E,F) reducing agents.** A, Elution profile of concentrated ferric HbA (K\(_{d2}=10\mu M\))\(^{21}\), was composed of a single peak. B, Elution profile of apoHbA sample showing two non-equilibrating peaks. C, Elution profile of apo-rHb 0.1 sample composed of two major peaks. D, Elution profile of apoHbA following incubation with 50 mM TPEC. E, Elution profile of apoHbA sample following dialysis in 1mM DTT, 10 mM potassium phosphate, pH 7 buffer. F, Elution profile of aporHb0.1 sample following dialysis in 1mM DTT, 10 mM potassium phosphate, pH 7 buffer. All samples were loaded onto a 24 ml Superose-12 HR 10/30 GL column equilibrated with 200 mM potassium phosphate, pH 7 using a 100 \(\mu l\) sample loop. Final elution concentrations of (A) 82 \(\mu M\) ferric HbA; (B) 52 \(\mu M\) apoHbA; (C) 23 \(\mu M\) aporHb0.1; (D) 54 \(\mu M\) apoHbA; (E) 38 \(\mu M\) apoHbA; and (F) 27 \(\mu M\) aporHb0.1. Samples were eluted with the equilibration buffer.
5.4 Final optimized conditions for reversible apoHb unfolding

Immediately after heme extraction into 2-butanol, all freshly isolated apoHb samples were dialyzed in 10 mM potassium phosphate buffer, 1 mM DTT, pH 7, buffer overnight at 4°C. When these DTT-treated samples were analyzed on the gel filtration column, no elution peaks characteristic of non-native, higher-order oligomeric species were observed (see Figures 5.3E, 5.3F for apoHbA and apo-rHb0.1). Then all apoHb samples were buffer exchanged into pH 7 200 mM potassium phosphate buffer containing 1 mM DTT. The final apoHb unfolding conditions for all of the variants examined were 200 mM potassium phosphate, 1 mM DTT at pH 7, 10° C. All the DTT containing solutions were freshly prepared before use to prevent oxidation and generation of radical oxygen species prior to use. The individual apoHb/GdnHCl mixtures containing DTT were incubated in a water bath at 10° C for 1 hour to ensure equilibrium. CD spectral changes upon addition of apoHb to the denaturant solution were complete in a few minutes and reversed rapidly when the apoHb samples in concentrated GdnHCl were diluted into buffer.

5.5 Statistical analysis methods comparing apoHbA unfolding models

In order to distinguish between the different unfolding models discussed in this chapter, I examined the unfolding of native apoHbA at protein concentrations ranging from 1 µM to 140 µM. Weighted fittings to all these models were performed in GnuPlot 5.04 using the program’s non-linear least squares Marquardt-Levenberg algorithm. The standard deviation for each data point, $S_{\text{measured}}$, which is the fraction CD signal change relative to the folded state at a specific protein concentration ($P_0$), was estimated from the triplicate measurements for the
GdnHCl titrations at $P_0 = 1.9 \, \mu M$, $P_0 = 12 \, \mu M$, and $P_0 = 60 \, \mu M$. For unfolding curves at $P_0 = 1.9 \, \mu M$, $P_0 = 12 \, \mu M$, and $P_0 = 60 \, \mu M$, each $S_{\text{measured}}$ data point was the average of the fractional CD change measured in triplicate with a standard deviation $s$, which was assigned a weight ($w$) equal to $1/s^2$ as described by Bevington. The measured $1/s^2$ values from the triplicate titration measurements were averaged at each [GdnHCl], divided by 4, and then used as the weights for the observed data points for the titrations at $P_0 = 108 \, \mu M$ and $P_0 = 140 \, \mu M$, for which only single measurements were made because of the large amount hemoglobin required for these concentrations. In most of the analyses, a Gaussian distribution is assumed for the final residuals, taking into account both the model and experimental errors. The raw residual was defined as the difference between the measured and predicted fractional CD change or $S_{\text{measured}} - S_{\text{model}}$, and the standardized residual was calculated as $w^{1/2}(S_{\text{measured}} - S_{\text{model}})$ as described by Bevington.

Reduced chi-square, $\chi_r^2$, was calculated as the weighted sum of squared residuals ($\Sigma w^*(-\text{raw residual})^2$) divided by the number of degrees of freedom. This $\chi_r^2$ value signifies the goodness of fit of the data to a specific model, with values around 1.0 corresponding to better fits. For the statistical analyses in Table 5.1, I conservatively took the number of degrees of freedom as the difference between the number of data points ($N$) and the total number of model variables ($K$). $N$ is 142 for the apoHbA unfolding curves.

Small sample Akaike's information criterion (AICc) values were calculated for each model fit and used as a measure of the likelihoods of the predictions by the four different models to be discussed in the following sections. These AICc values take into account the differences in numbers of variable parameters between the models. An abbreviated log likelihood estimation of each model prediction was used to calculate AICc using eq 5.1.
Burnham and Anderson\textsuperscript{208} assumed that \(\frac{\sum w\cdot (\text{raw residual})^2}{N}\) is approximately the residual variance \(\sigma^2\) of each fit and that the mean of the normalized residuals \(\mu\) is 0 when deriving this equation. They introduced this approach to calculate AICc for analyzing least-square fitting to models when the residual distributions are assumed Gaussian. The best fit model in a multi-model analysis should have the smallest AICc value among the models compared.

\[
\text{AICc} = N \log\left(\frac{\sum w\cdot (\text{raw residual})^2}{N}\right) + 2 K + \frac{(2K(K+1))}{(N-K-1)} \quad \text{eq 5.1}
\]

Because the mean of the residuals, \(\mu\), is almost never exactly 0 and \(\sigma^2\) differs more from \(\frac{\sum w\cdot (\text{raw residual})^2}{N}\) for “bad” models, AICc was also calculated using the maximum likelihood estimation (MLE) with eq 5.2\textsuperscript{207,208}. The MLE value for each model fitting was computed in MATLAB R2016a (using Statistics and Machine Learning Toolbox, The MathWorks\textsuperscript{®}, Inc., Natick, MA) by fitting the standardized residuals from the weighted fits to a Gaussian distribution\textsuperscript{207,208}.

\[
\text{AICc (using MLE)} = -2 \log(\text{MLE}) + 2 K + \frac{(2K(K+1))}{(N-K-1)} \quad \text{eq 5.2}
\]

### 5.6 Fitting to a simple 2-step model

All the unfolding measurements with apoHbA under reducing conditions showed two major phases, with only the second process showing a dependence on protein concentration (Fig. 5.4). These results are consistent with the two-step Model 1 shown in Fig. 5.1A. Thus, the apoHbA dimer (D) appears to partially unfold into a dimeric molten globule (I\textsubscript{D}), which in turn dissociates into 2 unfolded monomers (2U\textsubscript{M}). The CD signals were normalized to the total \(\alpha\) helical content of the folded apoHbA dimer and then analyzed by computing the fractions \(Y\)
of \( D, I_D, \) and \( U_M \) populations in terms of the equilibrium folding constants for the two steps described, using eqs 5.3 -5.6 given below.

Figure 5.4. Fits of equilibrium unfolding curves for apoHbA to the simple 3-state, 2-step model 1. The solid circles and triangles are the observed data, and the solid and dashed lines are the fitted curves for the 2-step apoHb heterodimer unfolding mechanism (Fig. 5.1A) obtained from global fitting of all the five curves to \( S_{[\text{GdnHCl}]} \) defined in eq 5.6. These unfolding titrations were done in 200 mM potassium phosphate, 1 mM DTT, pH 7 at 10°C and the total protein concentration (\( P_0 \)) is listed by symbol for the five different \( P_0 \) titrations (i.e., \( P_0 \) increased from 1.9 to 140 \( \mu \)M subunits). The observed data for \( P_0 = 1.9, 12, \) and 61 \( \mu \)M represent the average of triplicate titrations. Note that the simple 2-step model cannot describe the final third phase for the loss of CD signal and that the observed data show a significantly smaller dependence on total protein concentration than is predicted by the mechanism.
The equilibrium constants at a given [GdnHCl] are defined in terms of folding, starting from $U_M$, and the m values define the dependence of the folding free energy on [GdnHCl] and absolute temperature $(T)^{209}$.

$$K_{I_{I,D}} = \frac{[D]}{[I_D]} = K_{I_{I,D}}^0 \exp \left( \frac{-m_{I_{I,D}} [\text{GdnHCl}]}{RT} \right)$$

$$K_{2U_M,I_D} = \frac{[I_D]}{[U_M]^2} = K_{2U_M,I_D}^0 \exp \left( \frac{-m_{2U_M,I_D} [\text{GdnHCl}]}{RT} \right)$$

**eq 5.3**

The total monomeric protein concentration, $P_0$ is defined by a quadratic expression in $[U_M]$ (eq 5.4). At each data point, $[U_M]$ can be computed as the positive root of eq 5.4, defined by $P_0$ and the equilibrium constants at a given [GdnHCl] (eq 5.5).

$$P_0 = 2[D] + 2[I_D] + [U_M]$$

$$[I_D] = K_{2U_M,I_D} [U_M]^2; \quad [D] = K_{I_{I,D}} D K_{2U_M,I_D} [U_M]^2$$

$$P_0 = 2K_{I_{I,D}} D K_{2U_M,I_D} [U_M]^2 + 2K_{2U_M,I_D} [U_M]^2 + [U_M]$$

$$0 = 2\left(K_{I_{I,D}} D K_{2U_M,I_D} + K_{2U_M,I_D}\right)[U_M]^2 + [U_M] - P_0$$

$$[U_M] = \frac{-1 + \sqrt{1 + 8P_0 \left(K_{I_{I,D}} D K_{2U_M,I_D} + K_{2U_M,I_D}\right)}}{4\left(K_{I_{I,D}} D K_{2U_M,I_D} + K_{2U_M,I_D}\right)}$$

**eq 5.5**

The total CD signal $S_{[\text{GdnHCl}]}$ is then computed from the fractions of each folding species $(Y)$ multiplied by the corresponding intrinsic CD signal of each species (i.e. $S_D \approx 1.0$; $S_{ID} \approx 0.5$; and $S_{UM} \approx 0$) by inserting eq 5.5 into eq 5.6.
Algorithms for fitting observed data to eqs 5.3-5.6 were written in Gnuplot 5.04\textsuperscript{205}. As shown in Fig. 5.4, global fitting to the simple 2-step model, described in Fig. 5.1A and eqs 5.3-5.6, provides a preliminary description of the observed unfolding curves and an interpretation of the two major processes. However, a third unfolding phase is observed experimentally and represents loss of the last 10-15\% of the CD signal change (Figure 5.4). This third phase does not appear to depend on protein concentration and requires additional steps in the apoHbA unfolding mechanism. The observed data also show significantly less dependence on protein concentration for the second unfolding phase than predicted by the 2-step model (see inset to Fig. 5.4).

The global fit to this 2-step model also shows the largest $\chi^2$ value ($\chi^2 = 36$) in my multi-model analysis, indicating that this simple model is clearly inadequate (Table 5.1). The number of model variables, K, was assigned as 7 for this model for statistical analysis purposes (see Table 5.2).
### Table 5.1. Fitting statistics for apoHbA unfolding models

The number of data points, N, is 142. The total number of model variables, K, is 7 for the 2-step model, 10 for each of the 3-step models, and 13 for the 4-step model.

<table>
<thead>
<tr>
<th>Fitting Statistics</th>
<th>ApoHb unfolding model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-step</td>
</tr>
<tr>
<td>Reduced chi-square&lt;sup&gt;a&lt;/sup&gt;, $\chi_{r}^2$</td>
<td>36.3</td>
</tr>
<tr>
<td>AICc&lt;sup&gt;b&lt;/sup&gt;</td>
<td>518</td>
</tr>
<tr>
<td>ΔAICc&lt;sup&gt;c&lt;/sup&gt;</td>
<td>464</td>
</tr>
<tr>
<td>ΔAICc (using MLE)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>450</td>
</tr>
</tbody>
</table>

<sup>a</sup>Reduced chi-square was calculated as $(\Sigma w•(raw residual)²)/(N-K)$.

<sup>b</sup>AICc was calculated using eq 5.1.

<sup>c</sup>ΔAICc for each model was calculated as the difference between its AICc<sup>b</sup> value and the 4-step model (best fitting model) AICc<sup>b</sup> value.

<sup>d</sup>ΔAICc (using MLE) value for each model was determined from the difference of the AICc values calculated using eq 5.2.

### 5.7 Statistical analyses of alternative 3-step models

A third unfolding phase is apparent in the measured curves (Fig. 5.4), and therefore, the initial model was expanded to two alternative 3-step models to try to account for this final process. The first two steps still involve partial unfolding of the dimer (D→I<sub>d</sub>) followed by dissociation into monomers (I<sub>d</sub>→2U₉₃) with ~10% α-helicity. The first 3-step model added a final step, U₉₅→U₅₉, in which monomers (U₉₃), which still retain ~10% of the original secondary structure, completely unfold into monomeric chains (U₅₉). This model is described by eqs 5.3 and 5.7-5.11. Each monomer $K_{U₅₉,U₉₃}$ equilibrium folding constant is defined in eq 5.7 and the total monomeric protein concentration is now amended to include the U₅₉ species (eq 5.8). The
other folding constants are described previously in eq 5.3. $[U_M]$ for this model is then derived from eq 5.9. By describing the folding population fractions (eq 5.10) in terms of $U_m$ and the definitions for the equilibrium folding constants, the total CD signal (eq 5.11) is predicted. Fits of the apoHbA curves to this model, as described by eq 5.11, are shown in Fig. 5.5. For this model, the predicted dependence on protein concentration ($P_0$) (eq 5.8) is greater than what is observed (Fig. 5.5B). The addition of a third monomeric unfolding step ($U_M \rightarrow U_C$) alone cannot explain the smaller than expected dependence on protein concentration for the second phase, which appears to reach a limit at ~60 µM $P_0$ in the observed data (Figs. 5.4 and 5.5).

\[
K_{U_C,U_M} = \frac{[U_M]}{[U_C]} = K_{U_C,U_M}^0 \exp \left( \frac{-m_{U_C,U_M}[\text{GdnHCl}]}{RT} \right) \quad \text{eq 5.7}
\]

\[
P_0 = 2[D] + 2[I_D] + [U_M] + [U_C] \quad \text{eq 5.8}
\]

\[
2\left(K_{I_D,D}K_{2U_M,1_d} + K_{2U_M,1_d}\right)[U_M]^2 + \left(1 + \frac{1}{K_{U_C,U_M}}\right)[U_M] - P_0 = 0 \quad \text{eq 5.9}
\]

\[
Y_D = \frac{2[D]}{2[D] + 2[I_D] + [U_M] + [U_C]}
\]

\[
Y_{I_D} = \frac{2[I_D]}{2[D] + 2[I_D] + [U_M] + [U_C]}
\]

\[
Y_{U_M} = \frac{[U_M]}{2[D] + 2[I_D] + [U_M] + [U_C]}
\]

\[
Y_{U_C} = \frac{[U_C]}{2[D] + 2[I_D] + [U_M] + [U_C]}
\]

\[
S_{[\text{GdnHCl}]} = S_D Y_D + S_{I_D} Y_{I_D} + S_{U_M} Y_{U_M} + S_{U_C} Y_{U_C} \quad \text{eq 5.11}
\]
However, the $\chi^2$ value for fits to this model with 10 model variables is 6.6 (Tables 5.1 and 5.2), which is significantly smaller than the value for the simple 2-step model.

Figure 5.5. Fits of equilibrium unfolding curves for apoHbA to the 3-step model with $U_M \rightarrow U_c$. The solid circles and triangles are the observed data, and the solid and dashed lines are the globally fitted curves for the 3-step apoHb unfolding mechanism with $U_M \rightarrow U_c$ representing the final phase of unfolding. All the unfolding titrations were done in 200 mM potassium phosphate, 1mM DTT, pH 7 at 10°C. The model predicts a larger dependence on protein concentration during the second phase of unfolding compared to the observed data (i.e. dashed lines compared to the open circles and closed triangles).
Table 5.2. **Fitted parameters for apoHbA unfolding models.**
The number of data points, \( N \), is 142. The total number of model variables, \( K \), is 7 for the 2-step model, 10 for each of the 3-step models, and 13 for the 4-step model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>2-step model</th>
<th>3-step model ((U_M \rightarrow U_C))</th>
<th>3-step model ((2U_M \rightarrow U_2))</th>
<th>4-step model ((U_2 \text{ and } U_C))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_{1D,D} )</td>
<td>231</td>
<td>197</td>
<td>102</td>
<td>171</td>
</tr>
<tr>
<td>( K_{2U_M,I_D} )</td>
<td>(7360 \times 10^8 \text{ M}^{-1})</td>
<td>(1790 \times 10^8 \text{ M}^{-1})</td>
<td>(0.956 \times 10^8 \text{ M}^{-1})</td>
<td>(6.4 \times 10^8 \text{ M}^{-1})</td>
</tr>
<tr>
<td>( K_{U_C,U_M} )</td>
<td>N/A</td>
<td>346</td>
<td>N/A</td>
<td>1190</td>
</tr>
<tr>
<td>( K_{U_2,U_M} )</td>
<td>N/A</td>
<td>N/A</td>
<td>(9.63 \times 10^{-7} \text{ M})</td>
<td>(5.55 \times 10^{-7} \text{ M})</td>
</tr>
<tr>
<td>( m_{1D,D} )</td>
<td>(17.27 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(16.92 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(14.05 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(16.22 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
</tr>
<tr>
<td>( m_{2U_M,I_D} )</td>
<td>(22.76 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(21.22 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(5.89 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(11.55 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
</tr>
<tr>
<td>( m_{U_C,U_M} )</td>
<td>N/A</td>
<td>5.00</td>
<td>N/A</td>
<td>5.39</td>
</tr>
<tr>
<td>( m_{U_2,U_M} )</td>
<td>N/A</td>
<td>N/A</td>
<td>(-0.003 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(-3.91 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
</tr>
<tr>
<td>( S \text{ for } D )</td>
<td>1.003</td>
<td>1.010</td>
<td>1.012</td>
<td>1.005</td>
</tr>
<tr>
<td>( S \text{ for } I_D )</td>
<td>0.370</td>
<td>0.372</td>
<td>0.408</td>
<td>0.431</td>
</tr>
<tr>
<td>( S \text{ for } U_M )</td>
<td>0.064</td>
<td>0.108</td>
<td>0.083</td>
<td>0.089</td>
</tr>
<tr>
<td>( S \text{ for } U_2 )</td>
<td>N/A</td>
<td>N/A</td>
<td>(0.008 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
<td>(0.019 \text{ kJ mol}^{-1} \text{ M}^{-1})</td>
</tr>
<tr>
<td>( S \text{ for } U_C )</td>
<td>N/A</td>
<td>0.008</td>
<td>N/A</td>
<td>0.004</td>
</tr>
</tbody>
</table>

The other 3-step model is described by eqs 5.3 and 5.12 - 5.16 and has an alternative third step, \(2U_M \rightarrow U_2\), in which partially unfolded monomers non-specifically associate to form a completely unfolded dimer \((U_2)\). This step was added to dampen the dependence on \(P_0\) during the second phase of unfolding and make the computed curves more consistent with the experimental observations (Figure 5.6). The equilibrium disassociation constant for the non-specific association of the monomers is defined in eq 5.12, and total protein concentration, \(P_0\),
for this model (eq 5.13) and the populations fraction of each folding species (eq 5.15) are amended to include \([U_2]\). The other folding constants are described in eq 5.3.

\[
K_{U_2,U_M} = \frac{[U_M]^2}{[U_2]} = K_{U_2,U_M}^0 \exp \left( -\frac{-m_{U_2,U_M}[\text{GdnHCl}]}{RT} \right) \quad \text{eq 5.12}
\]

\[
P_0 = 2[D] + 2[I_D] + [U_M] + 2[U_2] \quad \text{eq 5.13}
\]

\[
0 = 2 \left( K_{I_D,D} K_{2U_M,I_D} + K_{2U_M,I_D} + \frac{1}{K_{U_2,U_M}} \right) [U_M]^2 + [U_M] - P_0 \quad \text{eq 5.14}
\]

\[
Y_D = \frac{2[D]}{2[D] + 2[I_D] + [U_M] + 2[U_2]}
\]

\[
Y_{I_D} = \frac{2[I_D]}{2[D] + 2[I_D] + [U_M] + 2[U_2]} \quad \text{eq 5.15}
\]

\[
Y_{U_M} = \frac{[U_M]}{2[D] + 2[I_D] + [U_M] + 2[U_2]}
\]

\[
Y_{U_2} = \frac{2[U_2]}{2[D] + 2[I_D] + [U_M] + 2[U_2]}
\]

\[
S_{\text{GdnHCl}} = S_D Y_D + S_{I_D} Y_{I_D} + S_{U_M} Y_{U_M} + S_{U_2} Y_{U_2} \quad \text{eq 5.16}
\]

Global fitting of the apoHbA unfolding data to eq 5.16 showed that this model predicts a protein concentration dependence (eq 5.13) that is smaller than that observed experimentally (Fig. 5.6), and, at low \(P_0\), complete unfolding does not occur because \(U_M\) becomes the dominant final state and retains helicity (Fig. 5.6B). The value of \(\chi^2\) for fitting to this model is 7 which is similar to the other alternate 3-step model (Table 5.2).
Although both 3-step models have significantly smaller $\chi^2$ values than the 2-step model, these values, 6 and 7, still deviate significantly from unity, suggesting a more complex model is needed to accurately describe the data.

Figure 5.6. *Fits of equilibrium unfolding curves for apoHbA to the 3-step model with $2U_M \rightarrow U_2$. The solid circles and triangles are the observed data, and the solid and dashed lines are the global fitted curves for the 3-step apoHb unfolding mechanism, with the last step being
2U_M→U_2. All the unfolding titrations were done in 200 mM potassium phosphate, 1mM DTT, pH 7 at 10°C. The model predicts a slightly smaller dependence on protein concentration during the second phase of unfolding compared to the observed data and does not accurately predict the third phase of unfolding at low protein concentrations.

5.8 Fitting to a 4-step model, 5-state model

Figure 5.7. Complete 4-step, 5-state mechanism for human apoHb unfolding. The apoHb heterodimer (D) initially unfolds into a dimeric, molten globule state (I_d). I_d dissociates into unfolded monomers (2U_M), which still retain a small fraction of α-helicity. These unfolded monomers then either transiently form a heterodimer unfolded species (U_2) through non-specific interactions or completely unfold into unstructured peptide chains (2U_C).

When both steps, U_M→U_C and 2U_M→U_2, are included in the 4-step model described in Fig. 5.7 and eqs 5.17-5.20, the observed dependence of the second phase on protein concentration as well as the amplitude and shape of third phase for complete unfolding are accurately described (Fig. 5.8). All the folding constants considered in eqs 5.17-5.20 were described in eqs 5.3, 5.7, and 5.12. The U_M→U_2 step is needed to reduce the dependence of the
second phase on \( P_0 \), and represents transient non-specific association of unfolded monomers with residual helicity to form a mostly unfolded dimeric species (Fig. 5.7). Transient formation of the \( U_2 \) species will facilitate unfolding at higher protein concentrations, partly compensating for the inhibition of the \( I_D \rightarrow 2U_M \) transition as \( P_0 \) is increased. However, as [GdnHCl] increases, the weak non-specific interactions in the \( U_2 \) state will be disrupted and lead to dissociation back to monomers, which then completely unfold in the final \( U_M \rightarrow U_C \) step. Both the \( U_2 \) and \( U_C \) folding species are included in the quadratic expression defining \( P_0 \) in eq 5.17.

\[
P_0 = 2[D] + 2[I_D] + [U_M] + 2[U_2] + [U_C]
\]

\[
0 = 2\left(K_{I_D,D}K_{2U_M,I_D} + K_{2U_M,I_D} + \frac{1}{K_{U_2,2U_M}}\right)[U_M]^2 + \left(1 + \frac{1}{K_{U_C,U_M}}\right)[U_M] - P_0 \quad \text{eq 5.17}
\]

\([U_M]\) is then derived (eq 5.18) as the root of this quadratic expression (eq 5.17).

\[
[U_M] = \frac{-\left(1 + \frac{1}{K_{U_C,U_M}}\right) + \sqrt{\left(1 + \frac{1}{K_{U_C,U_M}}\right)^2 + 8P_0 \left(K_{I_D,D}K_{2U_M,I_D} + K_{2U_M,I_D} + \frac{1}{K_{U_2,2U_M}}\right)}}{4\left(K_{I_D,D}K_{2U_M,I_D} + K_{2U_M,I_D} + \frac{1}{K_{U_2,2U_M}}\right)}
\]

\[
\text{eq 5.18}
\]
The folding population fractions (eq 5.19) for this 4-step mechanism were then derived using the \([U_M]\) expression to obtain a final description for the CD signal (eq 5.20) in terms of the 13 model parameters for this more complex model (Fig. 5.7 and Table 5.2), including the appropriate equilibrium constants and intrinsic CD signals of each species.

\[
Y_D = \frac{2[D]}{2[D]+2[I_D]+[U_M]+2[U_2]+[U_C]} \\
Y_{I_D} = \frac{2[I_D]}{2[D]+2[I_D]+[U_M]+2[U_2]+[U_C]} \\
Y_{U_M} = \frac{[U_M]}{2[D]+2[I_D]+[U_M]+2[U_2]+[U_C]} \\
Y_{U_C} = \frac{[U_C]}{2[D]+2[I_D]+2[U_2]+[U_C]} \\
Y_{U_2} = \frac{2[U_2]}{2[D]+2[I_D]+[U_M]+2[U_2]+[U_C]} \\
\]

(eq 5.19)

The global fit to the 4-step model gave a \(\chi^2\) value near 1 (Table 5.1), indicating the validity of this mechanism for describing the observed data\textsuperscript{206}. The fit to the 4-step model also has the “best” or lowest AICc value relative to the other models when using either eq 5.1 or 5.2 (Table 5.1)\textsuperscript{207,208}. The 2-step model and both the 3-step models all have large \(\Delta\text{AICc}\) values exceeding 200, regardless of whether eqs 5.1 or 5.2 were used (Table 5.1). Burnham and
Anderson have argued that an alternative to the best fit model has no empirical support if the AICc difference between that of the alternative model and that of the best fit model, $\Delta$AICc, is larger than 10.

Figure 5.8. Fits of equilibrium unfolding curves for apoHbA to the complete 5-state, 4-step model. The solid circles and triangles represent observed data, and the solid and dashed lines are curves predicted from the 4-step apoHb heterodimer unfolding mechanism (Fig. 5.7) and were obtained by global fitting of all the five curves to eq 5-20. As shown, this 4-step model is able to describe the smaller dependence on protein concentration of the second phase of unfolding and accurately represent the final third phase for loss of CD signal as described in the text.

The final fitted parameters for this 4-step model are given in Table 5.2. All of the equilibrium constants, $K$, in this table represent the $K^0$ values defined in eqs 5.3, 5.7, and 5.10 as extrapolated to [GdnHCl]=0. The superscript 0 in these eqs is implied when these $K$ values are discussed in the text. The values for $m_{2UM,10}^0$, $m_{1D,10}^0$, $m_{U2UM}^0$, and $m_{UC2UM}^0$ for the 4-step
model were determined to be 11.55 kJ mol\(^{-1}\) M\(^{-1}\), 16.22 kJ mol\(^{-1}\) M\(^{-1}\), -3.91 kJ mol\(^{-1}\) M\(^{-1}\), and 5.39 kJ mol\(^{-1}\) M\(^{-1}\), respectively, while \(K_{Uc,Um}\) is 1190 and \(K_{U2,2Um}\) is \(5.55 \times 10^{-7}\) M (Table 5.2). The negative m value for the dissociation of the U\(_2\) dimers describes the inhibitory effect of the denaturant on the formation of unfolded dimer aggregates.

The following analysis was done to obtain more accurate experimental errors for the \(S\), \(K_{2Um,Id}\), and \(K_{Id,D}\) values obtained from fits to the 4-step model. Values of these parameters obtained from fitting the data set with the triplicate data point averages across the different \(P_0\) were used as initial values for fitting individual sets of single titration curves across the 3 different apoprotein concentrations used for the triplicate measurements. The values for \(K_{Uc,Um}\), \(K_{U2,2Um}\), \(m_{2Um,Id}\), \(m_{Id,D}\), \(m_{U2,2Um}\), and \(m_{Uc,2Um}\) were fixed from the global analysis of the data point averages. The experimental errors obtained for \(K_{Id,D}\) and \(K_{2Um,Id}\) were \(164 \pm 3\) and \(6.6 \pm 0.1 \times 10^8\) M\(^{-1}\), respectively, determined from fitting the individual sets of titration curves and then computing the variability of the fitted parameters for the three different sets of data. Using the same analysis, the following \(S\) values and errors were estimated: \(S_D = 1.003 \pm 0.002; S_{Id} = 0.430 \pm 0.002; S_{Um} = 0.089 \pm 0.0001; S_{Uc} = 0.0034 \pm 0.0002;\) and \(S_{U2} = 0.021 \pm 0.001\).

The 4-step apoHb folding model described in Fig. 5.7 and eq 5.20 does not take into account the difference in stability between the \(\alpha\) and \(\beta\) subunits because it is not possible to resolve subunit differences with CD measurements.

**5.9 Multi-model analysis of the residual distributions for the fits**

The residual distributions for fits to all four models are shown in Figs. 5.9 and 5.10. The raw residuals for the first phase of unfolding (0 to ~ 1 M GdnHCl) are large but appear
random and model independent (Fig. 5.9). The standardized residuals in Fig 5.10 incorporate corrections for the large fluctuations in the raw residuals (Fig. 5.9) due to standard deviations in the experimental measurements. These corrections show that the scatter seen in the raw data during the first phase of unfolding is due to experimental errors and not systematic fluctuations due to poor models. The standardized residuals during the first phase of unfolding appear model independent, as demonstrated by similar values predicted by all the different models for \( m_{10,D} \) and \( K_{10,D} \), which indicates that the first phase of unfolding is well established as partial unfolding of apoHb dimers (Table 5.2).

Figure 5.9. **Raw residual distributions for fits of apoHbA unfolding curves** to the: (A) 2-step mechanism; (B) 3-step mechanism with only \( U_M \rightarrow U_C \); (C) 3-step mechanism with only \( U_M \rightarrow U_2 \); and (D) 4-step mechanism with both \( U_M \rightarrow U_C \) and \( U_M \rightarrow U_2 \). The raw
residual is calculated as the difference between the measured CD signal and the CD signal calculated for the model with optimized parameters. The simple 2-step model is described by eqs 5.3-5.6 and the 3-step models are described by eqs 5.7-5.16 and Fig. 5.7 with only the $U_M \rightarrow U_C$ step or only the $U_M \rightarrow U_2$ step. The 4-step model is described structurally in Fig. 5.7 and all the expressions in eqs 5.17-5.20.

In contrast, systematic fluctuations are seen for both the 2-step and 3-step models during the second and third unfolding phases in both the normalized and raw distribution of residuals. In the case of the 2-step model, there is a clear indication of a systematic negative distribution of residuals beyond ~2.5 M [GdnHCl] showing that this model cannot accurately describe the data during the final unfolding phase (Fig. 5.9A and 5.10A). A similar but smaller systematic negative residual distribution is observed for the 3-step model with the $2U_M \rightarrow U_2$ transition at low $P_0$ (Figs. 5.9C and 5.10C). In contrast, the residual distribution is clearly random in this high [GdnHCl] region for both the 3-step and 4-step models incorporating the $U_M \rightarrow U_C$ transition (Figures 5.9B, 5.9D, 5.10B, and 5.10D), further strengthening our interpretation that complete unfolding of apoHb monomers occurs during the third and final unfolding phase.
Figure 5.10. **Standardized residual distributions for fits of apoHbA unfolding data** to the (A) 2-step mechanism; (B) 3-step mechanism with only \( U_M \rightarrow U_c \); (C) 3-step mechanism with only \( U_M \rightarrow U_2 \); and (D) 4-step mechanism with both \( U_M \rightarrow U_c \) and \( U_M \rightarrow U_2 \). The standardized residual is calculated as the raw residual multiplied by the square root of the weight used for fitting each data point (i.e. \( w = 1/s^2 \) where \( s \) is the standard deviation of the fractional CD signal change measured in triplicate titrations). Note that the y-axis scale is 75% larger in panel A, demonstrating the poor quality of the fit to the 2-step model. The simple 2-step model is described by eqs 5.3-5.6 and the 3-step models are described by eqs 5.7-5.16 and structurally in Fig. 5.7 with only the \( U_M \rightarrow U_c \) step or only the \( U_M \rightarrow U_2 \) step. The 4-step model is described structurally in Fig. 5.7 and all the expressions in eqs 5.17-5.20.

When the residuals for the 2-step and 3-step models are examined in terms of protein concentration dependence, systemic variations are detected during the second unfolding phase. For the 2-step model, the observed CD signals are systematically greater than the values computed from the model at the lowest protein concentration (positive residuals) and then are always smaller (negative residuals) at higher protein concentrations (Figs. 5.9A and 5.10A). A similar pattern is observed for the 3-step model with the \( U_M \rightarrow U_c \) transition (Figs. 5.9B and
Thus, as described previously, both these models predict larger dependences on protein concentration than observed experimentally. In contrast, the residuals in the second phase for the 3-step model with the $U_m \rightarrow U_2$ transition show the opposite distribution as a function of $P_0$ (Figures 5.9C and 5.10C). In this case, the model predicts a protein concentration dependence that is smaller than what is observed. In contrast to other models, the standardized residuals for the fit to the 4-step mechanism appear much more random in all three phases, supporting the use of this more complex mechanism as the "best" model for fitting the observed unfolding curves (Figure 5.10D).

### 5.10 Conclusion

The $\chi^2_r$ values, small sample Akaike's information criteria (AICc and $\Delta$AICc values) analysis, and the residual distributions, all indicate that the 4-step model described in Fig. 5.7 is required for an accurate representation of the observed data for apoHbA dimer unfolding. The unfolding measurements and all four models show clearly that the initial step involves partial unfolding into a molten globule dimer and that this process is followed by dissociation into monomers which have residual helical structure. The 4-step model interprets the dampened protein concentration dependence for the second unfolding phase in terms of transient formation of a unfolded dimeric species, $U_2$, with residual helicity and indicates that the third and final phase results in loss of residual secondary structure of monomers to form completely unfolded polypeptide chains. These structural interpretations were verified in the next chapter by examining the unfolding characteristics of hemoglobin variants containing stabilized heme pockets and strengthened dimer interfaces.
Chapter 6

Verification of human hemoglobin unfolding model

This chapter is adapted and reproduced from my following publication:


In order to avoid redundancy, I have minimized and/or omitted further references to these publications within the text body.

6.1 Introduction

Protein folding and assembly occur through various non-covalent interactions, including secondary and tertiary structure formation, cofactor binding, and oligomerization. These interactions are critical for biological function and inhibit protein disorder in localized regions\(^ {211-213}\). Human Hb provides an ideal model framework for studying the effects of subunit interfaces and cofactor binding on protein folding and assembly.

In the previous chapter, a general 4-step mechanism for human apohemoglobin unfolding model was proposed. As discussed, removal of heme from human hemoglobin results in formation of an apoglobin heterodimer. Titration of this apo-dimer with GdnHCl leads to biphasic unfolding curves indicating two major distinct steps. By analogy with apoMb, I proposed that the heme pocket unfolds and generates a dimeric intermediate in which ~50% of the original helicity is lost, but the $\alpha_i\beta_i$ interface ($\alpha_i\gamma_i$ in the case of HbF) is still intact. At
higher [GdnHCl], this intermediate then dissociates into mostly unfolded monomers that then completely unfold during a third minor phase.

The goal of the work in this chapter is to verify the structural interpretations of this model using a series of recombinant Hbs. The initial set of variants contained apolar mutations, which markedly stabilize the native conformation of the heme pockets in the α and β subunits. Then a recombinant Hb crosslinked with a single glycine linker between the α chains (rHb0.1) was used to inhibit dissociation into monomers.

The differences in stability between human adult (HbA) and fetal (HbF) hemoglobin were also examined. Previous studies with holoHbA and holoHbF suggested that the α₁γ₂ tetramer interface of HbF is stronger than the corresponding α₁β₂ interface of HbA, based on the observation that the equilibrium constant for tetramer to dimer dissociation (K_{4,2}) of HbF is significantly smaller than that for HbA in the liganded forms. Past studies in the Hb field have also suggested that the α₁γ₁ dimer interface of HbF is also significantly stronger than α₁β₁ interface of HbA. However, this latter idea is more controversial due to the difficulty of measuring dimer to monomer dissociation of the holoproteins and then interpreting the data. In this work, we have attempted to make a direct comparison of strengths of apodimer interfaces during unfolding of apoHbA and apoHbF.

### 6.2 Heme pocket unfolding

Introduction of large apolar amino acids in the heme pocket increases apomyoglobin stability, partially compensating for the loss of heme. NMR characterization of the intermediate folding state of H64F apoMb showed that replacement of the highly polar E7
imidazole side chain with an apolar benzyl group leads to significant stabilization of the E helix\textsuperscript{27}. In contrast to apoMb, there have been no structural characterizations published for native or mutant human apoHbs or their folding intermediates.

In order to examine whether the initial unfolding phase for apoHb dimers involves melting of the heme pocket, large apolar and aromatic amino acids were introduced at the E7 and E11 helical positions in the individual \(\alpha\) or \(\beta\) subunits and then in both subunits together. The replacements were His(E7)\(\rightarrow\)Leu and Val(E11)\(\rightarrow\)Phe. The following three variants were constructed, expressed, and purified: \(\alpha\)(H58L/V62F)\(\beta\)(wt), \(\alpha\text{(wt)}\beta\text{(H63L/V67F)}, \) and \(\alpha\text{(H58L/V62F)}\beta\text{(H63L/V67F)}\). These mutations did not alter apoHb quaternary structure based on analytical gel filtration. For example, at a final concentration of 60 \(\mu\)M, \(\alpha\text{(H58L/V62F)}\beta\text{(H63L/V67F)}\) apo-rHb, eluted at 14.43 ml, which corresponds to a dimer elution peak for HbA.

Triplicate GdnHCl unfolding titrations for \(\alpha\text{(H58L/V62F)}\beta\text{(wt)}, \alpha\text{(wt)}\beta\text{(H63L/V67F)}, \) and \(\alpha\text{(H58L/V62F)}\beta\text{(H63L/V67F)}\) apo-rHbs and apoHbA at \(P_0 = 12 \mu\)M are shown in Fig. 6.1. There is a dramatic right shift of the first phase of the unfolding curve towards much higher [GdnHCl] as the heme pocket is made more apolar. These shifts suggest strongly that the first phase of apoHb unfolding involves melting of the heme pockets leading to the formation of a molten globule dimer state, which retains \(~50\%\) of the original apoprotein helicity.
Figure 6.1. GdnHCl induced equilibrium unfolding of apoHbA and distal heme pocket apo-rHb mutants followed by CD changes. The circles are the observed data, and the solid lines are the fitted curves to the 4-step apoHb heterodimer unfolding mechanism (Fig. 5.7) using eq 5.20. Unfolding measurements of these apoHbs were performed using 12 μM protein in 200 mM potassium phosphate, 1 mM DTT, pH 7 at 10°C. The individual curves are labeled in the figure.
Table 6.1. Fitted equilibrium unfolding parameters for apoHbA, apoHbF, and other rHb variants, using the 5-state, 4-step model for ApoHb. ApoHb 4-step unfolding model used is shown in Fig. 5.7. Parameters enclosed in parentheses are for rHb0.1 using the model shown in Fig. 6.3.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>HbA</th>
<th>HbF</th>
<th>rHb 0.0</th>
<th>rHb0.1</th>
<th>rHb α(H58L/V62F)</th>
<th>rHb β(H63L/V7F)</th>
<th>rHb α(H58L/V62F)β (H63L/V7F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K₁₁,D</td>
<td>164 ± 3</td>
<td>128 ± 11</td>
<td>319 ± 68</td>
<td>(K₁₁,T)</td>
<td>5,523 ± 154</td>
<td>586 ± 114</td>
<td>14,900 ± 2,300</td>
</tr>
<tr>
<td>K₂₂UM,₁D</td>
<td>6.6 ± 0.1 x 10^8 M⁻¹</td>
<td>83.6 ± 21.5 x 10^8 M⁻¹</td>
<td>7.99 ± 0.84 x 10^8 M⁻¹</td>
<td>(K₃₃UM,₁D)</td>
<td>6.5 ± 0.3 x 10^8 M⁻¹</td>
<td>8.6 ± 0.2 x 10^8 M⁻¹</td>
<td>19 ± 5 x 10^8 M⁻¹</td>
</tr>
<tr>
<td>K₁₂UM</td>
<td>1190</td>
<td>921</td>
<td>1190</td>
<td>(266)</td>
<td>1190</td>
<td>1190</td>
<td>1190</td>
</tr>
<tr>
<td>K₂₂₂UM</td>
<td>5.55 x 10⁻⁷ M⁻¹</td>
<td>0.165 x 10⁻⁷ M⁻¹</td>
<td>5.55 x 10⁻⁷ M⁻¹</td>
<td>(K₃₃₃UM)</td>
<td>5.55</td>
<td>5.55</td>
<td>5.55</td>
</tr>
<tr>
<td>m₁₁,D</td>
<td>16.22</td>
<td>17.48</td>
<td>16.22</td>
<td>(m₁₁,₁D)</td>
<td>16.22</td>
<td>16.22</td>
<td>16.22</td>
</tr>
<tr>
<td>m₂₂₂UM</td>
<td>11.55</td>
<td>12.68</td>
<td>11.55</td>
<td>(m₃₃₃UM,₁D)</td>
<td>11.55</td>
<td>11.55</td>
<td>11.55</td>
</tr>
<tr>
<td>m₁₂₂UM</td>
<td>5.39</td>
<td>5.39</td>
<td>5.39</td>
<td>(4.6)</td>
<td>5.39</td>
<td>5.39</td>
<td>5.39</td>
</tr>
<tr>
<td>m₂₂₂₂UM</td>
<td>-3.91</td>
<td>-6.81</td>
<td>-3.91</td>
<td>(m₃₃₃₃UM,₁D)</td>
<td>-3.91</td>
<td>-3.91</td>
<td>-3.91</td>
</tr>
<tr>
<td>S for D</td>
<td>1.003 ± 0.002</td>
<td>1.021 ± 0.001</td>
<td>1.008 ± 0.012</td>
<td>(S for T)</td>
<td>0.98 ± 0.002</td>
<td>0.989 ± 0.003</td>
<td>0.99 ± 0.005</td>
</tr>
<tr>
<td>S for I₀</td>
<td>0.430 ± 0.002</td>
<td>0.452 ± 0.002</td>
<td>0.44 ± 0.02</td>
<td>(S for I₁)</td>
<td>0.414 ± 0.002</td>
<td>0.443 ± 0.005</td>
<td>0.429 ± 0.002</td>
</tr>
<tr>
<td>S for U₀</td>
<td>0.089 ± 0.0001</td>
<td>0.11 ± 0.01</td>
<td>0.095 ± 0.005</td>
<td>(S for U₁)</td>
<td>0.081 ± 0.001</td>
<td>0.088 ± 0.005</td>
<td>0.118 ± 0.009</td>
</tr>
<tr>
<td>S for U₂</td>
<td>0.021 ± 0.001</td>
<td>0.032 ± 0.024</td>
<td>0.021 ± 0.001</td>
<td>(S for U₃)</td>
<td>0.0174 ± 0.0002</td>
<td>0.019 ± 0.001</td>
<td>0.025 ± 0.002</td>
</tr>
<tr>
<td>S for U₃</td>
<td>0.0034 ± 0.0002</td>
<td>0.008 ± 0.008</td>
<td>0.0070 ± 0.0007</td>
<td>(0.0042)</td>
<td>0.0049 ± 0.0002</td>
<td>0.006 ± 0.003</td>
<td>0.0139 ± 0.0036</td>
</tr>
</tbody>
</table>
Titration curves for these variants were analyzed in terms of the 4-step, 5-state apoHbA unfolding mechanism described in Fig. 5.7 and eq 5.20. All the m, K_{UC,UM}, and K_{U2,UM} parameters were set to the values determined from the global fittings of apoHbA unfolding data shown in Table 6.1 because the heme pockets of the U_M, U_2, and U_c states are already unfolded and should be little affected by the active site mutations. The K_{2UM,ID}, K_{ID,D}, and folding state signal (S) values were allowed to vary during the curve fitting. The S values remained very similar to those for apoHbA (Table 6.1). The S, K_{2UM,ID}, and K_{ID,D} values obtained from fitting to the data point averages (Fig. 6.1) were used as initial values for fitting each of the individual titration curves in order to estimate experimentally the errors in the fitted values of these key parameters (Table 6.1).

K_{ID,D} for α(H58L/V62F)β(H63L/V67F) apo-rHb is 14,900 ± 2,300, which is ~90-fold larger than the value for refolding from the molten globule to the native apoHbA state (Table 6.1). Thus, the apolar mutations along the E helix in the distal heme pocket markedly increase resistance to unfolding during the first phase to form the I_D state. The K_{ID,D} values for α(H58L/V62F)β(wt) and α(wt)β(H63L/V67F) apo-rHbs are 5,500 ± 150 and 590 ± 110 respectively (Table 6.1). Interestingly, the larger stabilizing effects occur in the α subunits (Table 6.1). These results also suggest that the increases in free energies for heme pocket folding of the individual mutant subunits are roughly additive in the quadruple mutant dimer, i.e. K_{ID,D}(quadruple)/K_{ID,D}(native) ≈ (K_{ID,D}(α-mutant)/K_{ID,D}(native)) • (K_{ID,D}(β-mutant)/K_{ID,D}(native))). As a result, putting these heme pocket mutations in both subunits in the quadruple mutant results in a dramatic shift of first phase of the unfolding curve to very
high values of [GdnHCl], which begins to obscure the second major phase involving dimer dissociation (Fig. 6.1).

Because these E helix mutations are far from the $\alpha_1\beta_1$ interface, the association constant, $K_{2Um,Id}$, for monomer association to the dimer intermediate was expected to remain unchanged. The $K_{2Um,Id}$ values for both apoHb $\alpha$(H58L/V62F)$\beta$(wt) and $\alpha$(wt)$\beta$(H63L/V7F) aporHbs are $6.5 \pm 0.3 \times 10^8$ M$^{-1}$ and $8.6 \pm 0.2 \times 10^8$ M$^{-1}$, respectively, and almost identical to value for native apoHbA, $6.6 \pm 0.3 \times 10^8$ M$^{-1}$. Interestingly however, the $K_{2Um,Id}$ for $\alpha$(H58L/V62F)$\beta$(H63L/V67F) apo-rHb was ~3 fold higher, $19.3 \pm 4.6 \times 10^8$ M$^{-1}$, suggesting that stabilizing the heme pockets of both subunits together has a small synergistic effect on formation of the $\alpha_1\beta_1$ interface in the dimeric intermediate, $I_D$ (Table 6.1).

### 6.3 Unfolding of apo-rHb0.0

apo-rHb0.0 is a wild-type rHb control, in which the subunits have V1M mutations. Unfolding data for dimeric apo-rHb0.0 was obtained in triplicate measurements at $P_0$=12 μM. The data were then fitted to the same 4-step, 5-state model used for native apoHbA and the distal pocket mutants (i.e., Fig. 5.7, and eq 5.20). Again, all the m, $K_{UC,Um}$, and $K_{U2,2Um}$ parameters were fixed to values obtained for apoHbA, while all other parameters were allowed to vary. The V1M mutations cause little effect on the overall stability of apo-rHb0.0 relative to apoHbA. The $K_{Io,D}$ and $K_{2Um,Id}$ values for apo-rHb0.0 were $319 \pm 68$ and $7.99 \pm 0.84 \times 10^8$ M$^{-1}$ respectively, which are within a factor of 2 very similar to the corresponding values for HbA, and the S values were also nearly same for both control proteins (Table 6.1).
6.4 Unfolding of apo-rHb0.1, a cross-linked Hb with a glycine linker between the α subunits

Previously published crystal structures of holo-rHb0.1 have shown that the addition of a glycine linker between the two α subunits does not have a significant effect on the tertiary structures of the hemoglobin subunits nor on the interactions at both the α₁β₁ and α₁β₂ interfaces\textsuperscript{218,219}. However, the glycine crosslink does cause the apo-rHb0.1 to remain a “tetramer” even after heme removal (Figs. 5.3C, 5.3F). In effect, apo-rHb0.1 is really a trimer composed of 3 subunits (2 β chains and 1 di-α chain) and two α₁β₁ interfaces, whereas apoHbA contains only 2 subunits and one α₁β₁ interface. Both previous studies\textsuperscript{44,49,50} and my
Figure 6.2. **Recombinant Hb 0.1 (rHb0.1)** (PDB ID #1O1L) with a one glycine crosslinker between $\alpha$Arg141 and $\alpha$Val143, originally the C-terminus and N-terminus of the $\alpha_1$ and $\alpha_2$ chains. There is a single di-$\alpha$ chain and 2 $\beta$ chains in rHb0,1. rHb0.1 was crystalized as cyanomet-rHb0.1.

gel filtration data (Fig. 5.3) have shown that apoHbA exists as a dimer and that the $\alpha_i\beta_2$ interfaces are disrupted due to unfolding of the F helix following heme removal. A corollary to this idea is that rHb0.1 would also lose its $\alpha_i\beta_2$ interfaces after hemin removal. This statement is supported by the SAXS analysis of rHb0.1 solutions to be described in Chapter 7. These data suggest that although holo-rHb0.1 exists in a compact state, apo-rHb0.1 exists in solution in a more extended state presumably as two dimers held together by the glycine linker, *i.e.*, $\alpha_1\beta_1-\alpha_2\beta_2$.

**Figure 6.3.** **Apo-rHb0.1 equilibrium unfolding mechanism.** Trimeric apo-rHb0.1 (T) unfolds initially into a trimeric molten globule state (I$_T$). I$_T$ then undergoes dissociation into 3 unfolded
monomers ($3U_M$) that still retain a small fraction of $\alpha$-helicity and are composed of 1 di-$\alpha$ monomer ($U_{M(\alpha)}$) and 2 $\beta$ monomers ($2U_{M(\beta)}$). The unfolded monomers then either transiently form a heterotrimERIC unfolded species ($U_3$) through non-specific interactions or completely unfold into monomeric polypeptide chains ($3U_C$).

The scheme shown in Fig. 5.7 for HbA had to be modified to analyze the unfolding of apo-rHb0.1. For the modified model shown in Fig. 6.3, the first phase involves unfolding of the heme pockets in the apo-rHb0.1 trimer (T) to form a trimeric molten globule ($I_T$). During the 2nd phase of unfolding, $I_T$ dissociates into three mostly unfolded monomers ($3U_M$) consisting of one di-$\alpha$ monomer ($U_{M(di-\alpha)}$) and two $\beta$ monomers ($2U_{M(\beta)}$). During the 3rd phase, these monomers either non-specifically associate transiently as a trimer ($U_3$) or each monomer unfolds completely into a peptide chain ($U_C$) that retains no secondary structure. Similar to the apoHb dimer model (Fig. 5.7), the non-specific inter-subunit interactions in the $U_3$ state are disrupted at high [GdnHCl].

In order derive a relatively simple expression for the unfolding curve of apo-rHb0.1, the di-$\alpha$ and $\beta$ monomers are considered to be equivalent unfolded species. The equilibrium constants for folding into apo-rHb0.1 trimers are then defined as:

$$K_{I_T,T} = \frac{[T]}{[I_T]} = K_{I_T,T}^0 \exp \left( \frac{-m_{I_T,T}[GdnHCl]}{RT} \right)$$

$$K_{3U_M,I_T} = \frac{[I_T]}{[U_M]^3} = K_{3U_M,I_T}^0 \exp \left( \frac{-3m_{3U_M,I_T}[GdnHCl]}{RT} \right)$$

$$K_{U_C,U_M} = \frac{[U_M]}{[U_C]} = K_{U_C,U_M}^0 \exp \left( \frac{-m_{U_C,U_M}[GdnHCl]}{RT} \right)$$

$$K_{U_3,3U_M} = \frac{[U_M]^3}{[U_3]} = K_{U_3,3U_M}^0 \exp \left( \frac{-m_{U_3,3U_M}[GdnHCl]}{RT} \right)$$

\[ \text{eq 6.1} \]
\( P_0 \) at each [GdnHCl] is defined in terms of the equilibrium constants and \([U_M]\) as shown in eq 6.2.

\[
P_0 = 3[T] + 3[I_T] + 3[U_3] + [U_c] + [U_M]
\]

\[
0 = 3(K_{I_T,3}K_{3U_M,3T} + K_{3U_M,3T} + \frac{1}{K_{3U_M,3T}})[U_M]^3 + (1 + \frac{1}{K_{U_c,U_M}})[U_M] - P_0 \quad \text{eq 6.2}
\]

The absolute value of \([U_M]\) is computed from the cubic root of this expression (eq 6.2) using a Newton-Rhapson algorithm written in MATLAB R2016a (The MathWorks®, Inc., Natick, MA). The value of \([U_M]\) at a given [GdnHCl] is then used to compute population fractions of the different folding states (Y values) with eq 6.3, which in turn is used to define the overall CD signal (S) in eq 6.4.

\[
Y_T = \frac{3[T]}{3[T] + 3[I_T] + 3[U_3] + [U_c] + [U_M]}
\]

\[
Y_{I_T} = \frac{3[I_T]}{3[T] + 3[I_T] + 3[U_3] + [U_c] + [U_M]}
\]

\[
Y_{U_3} = \frac{3[U_3]}{3[T] + 3[I_T] + 3[U_3] + [U_c] + [U_M]}
\]

\[
Y_{U_M} = \frac{[U_M]}{3[T] + 3[I_T] + 3[U_3] + [U_c] + [U_M]}
\]

\[
Y_{U_C} = \frac{[U_C]}{3[T] + 3[I_T] + 3[U_3] + [U_c] + [U_M]}
\]

\[
S_{\text{GdnHCl}} = S_T Y_T + S_{I_T} Y_{I_T} + S_{U_3} Y_{U_3} + S_{U_C} Y_{U_C} + S_{U_M} Y_{U_M}
\]

\[
= 3(S_K_{I_T,3}K_{3U_M,3T} + S_I_{I_T}K_{3U_M,3T} + \frac{S_U_{U_3}}{K_{3U_M,3T}})[U_M]^3 + \frac{S_{U_C}}{K_{U_c,U_M}} + S_{U_M} \quad \text{eq 6.4}
\]
Using this model, the dependence of apo-rHb0.1 unfolding on total protein concentration, ranging from 0.85 μM to 50 μM, was examined (Fig. 6.4). Unfolding of apo-rHb0.1 still occurs in two major phases and a third residual phase (Figs. 6.4 and 6.5), similar to apoHbA and the other rHbs examined. Remarkably, apo-rHb0.1 unfolding showed virtually no dependence on P₀ in contrast to apoHbA. Global fitting of the 5 titration curves in Fig. 6.4 to eq 6.4 was done in MATLAB by allowing all the K, m, and S values to vary, and the results are shown in Table 6.1. A large K₃UM,fr value of 6.8×10^{16} M^{-2} was obtained, consistent with the observed high resistance of the molten globule Iₜ state to dissociation and unfolding (Fig. 6.5, Table 6.1).
Figure 6.4. **GdnHCl induced equilibrium unfolding of apo-rHb0.1 as a factor protein concentration.** The circles, squares, and triangles are the observed data, and the solid lines are the fitted curves. All five sets of unfolding measurements for apo-rHb 0.1 were fitted globally to the 4-step apo-rHb0.1 unfolding mechanism (Fig. 6.3) using eq 6.4. All the unfolding titrations were done in 200 mM potassium phosphate, 1 mM DTT, pH 7 at 10°C.

Lack of dependence on protein concentration for the second phase of unfolding was due to the small value of $K_{U_3,3M}$, $3.6 \times 10^{-16}$ M$^2$, which describes the dissociation of the $U_3$ state (Table 6.1). Thus, although increasing protein concentration inhibits $I_T$ dissociation, this inhibition is compensated by promotion of transient aggregation of the partially unfolded
monomers to the U_3 state. These results suggest that the di-α linkage not only stabilizes the oligomeric molten globule, but also promotes re-association of the mostly unfolded di-α and β subunits (U_M states), inhibiting their complete unfolding into chains with no secondary structure. Genetic crosslinking has been used in the past to optimize oligomer stability for the gene ν protein and the arc repressor. The resulting enhanced stability in these cross-linked proteins was attributed to the linker’s role in driving the association of the subunits by keeping them at high local concentration and reducing the entropy of the unfolded subunits^{213,220-222}. This type of mechanism seems to be consistent with the small fitted $K_{U_3,3U_M}$ parameter obtained for the association of apo-rHb0.1 monomers.

Remarkably, the fitted $K_{IT,T}$ value of 158 for apo-rHb0.1 is virtually identical to that for the heme pocket stabilities of apoHbA and apo-rHb0.0 dimers (Table 6.1). This result quantitatively supports our mechanistic interpretation that the initial phase of apoHb unfolding involves unimolecular "melting" of the heme pockets and not dimer or tetramer dissociation. This result also shows that crosslinking has little effect on the average stabilities of the α and β heme pockets (Fig. 6.5, Table 6.1).
Figure 6.5. **Defining human apoHb unfolding transitions through mutagenesis.**

The unfolding measurements were done at 12 μM protein subunits for apoHbA, apo-rHb0.1, α(H58L/V62F)β(wt) apo-rHb, and α(wt)β(H63L/V67F) apo-rHb. The circles are the observed data and the solid lines are the fitted curves. The unfolding fitting for apo-rHb 0.1 was from the global fitting to the 4-step apo-rHb0.1 unfolding mechanism (Fig. 6.3) using eq 6.4. The unfolding measurements for the apoHbs without the genetic cross-linking were fitted to the 4-step apoHb heterodimer unfolding mechanism (Fig. 5.7) using eq 5.16. All measurements were done in 200 mM potassium phosphate, 1mM DTT, pH 7 at 10°C.

The value of $K_{3UM,1}$ for the formation of the trimeric intermediate state from unfolded monomers is determined by the strength of the two $\alpha_i\beta_i$ interactions that are formed. Work done in this study further suggests that $\alpha_i\beta_2$ and $\alpha_2\beta_i$ interfaces do not exist in this apo-rHb0.1 intermediate because the heme has been removed and its pockets are unfolded. This idea is strongly supported by noting that the square root of $K_{3UM,1}$ is $2.6 \times 10^8$ M$^{-1}$, which is very
similar to the values of $K_{2UM,1b}$ for formation of a single dimer interface in apoHbA, apo-rHb0.0, and the distal pocket apo-mutants (all $\sim 6 \times 10^8 \text{ M}^{-1}$, Table 6.1). Similar considerations apply when comparing the formation of the $U_3$ and $U_2$ states. The square root of the fitted value of $K_{U_3,3UM}$ is $1.9 \times 10^{-8} \text{ M}^{-1}$ and similar in magnitude to $K_{U_2,2UM}$ ($0.5 \times 10^{-8} \text{ M}^{-1}$, Table 6.1) for apoHbA, which supports the idea that two interface interactions have to be disrupted during the dissociation of $U_3$ into monomers.

As described above, roughly twice as much free energy (i.e. more GdnHCl) is required to dissociate apo-rHb0.1 into its three monomeric units than for native apoHbA to dissociate into two units as demonstrated by comparing -$RT\ln(K_{3UM,1I})$ to -$RT\ln(K_{2UM,1b})$ (Table 6.1). As shown in Figure 6.5, the net result of this effect is that the second phase of unfolding is significantly shifted to the right towards higher [GdnHCl], demonstrating that the molten intermediate for crosslinked apoglobin is much more resistant to dissociation (Fig. 6.5). Additionally, the trimeric molten globule ($I_T$) and the partially unfolded monomers both show a higher fraction of secondary structure compared to apoHbA and apo-rHb0.0 (Table 6.1, Fig. 6.5), with $S_{I_T}$ and $S_{UM}$ equal to $\sim 0.7$ and $\sim 0.3$ respectively, whereas the other $S$ values were comparable to those for apoHbA.

The large right-shift of the second unfolding phase due to crosslinking verifies that the intermediate state for apoHb represents an oligomeric molten globule, which in the second phase dissociates into monomeric units. The unfolding curves for $\alpha(H58L/V62F)\beta(wt)$ and $\alpha(wt)\beta(H63L/V67F)$ apo-rHbs in Fig. 6.5 emphasize the effect of stabilizing the heme pockets and, shown together with the apo-rHb0.1 curve, provide experimental verification of our
structural interpretations of the two major steps in apohemoglobin unfolding shown in Figs. 5.7 and 6.3.

6.5 Unfolding of fetal apohemoglobin

I compared the apoglobin stabilities of adult and recombinant fetal human Hbs in order to test the general applicability of the mechanism in Fig. 5.7. For this part of my thesis work, I mentored and worked together with a Rice University undergraduate student, William Ou, who worked on the unfolding mechanism of HbF for his Senior Honors Research (Bioc 401/402). Triplicate measurements of GdnHCl induced apoHbF unfolding curves were made at three different $P_0$ values (Fig. 6.6). Global fitting of all three sets of curves to the 4-step apoHbA unfolding model (Fig. 5.7) was done by optimizing the S, K, and m values as described for apoHbA unfolding data.

Again two major phases and a minor third phase are observed, but, in the case of apoHbF, the dependence of the second phase of unfolding on $P_0$ is more dampened than for apoHbA. A direct comparison of apoHbA and apoHbF unfolding curves at $P_0 = 12 \mu$M is shown in Fig. 6.7. Relative to apoHbA, the small shift of the second phase for apoHbF to higher [GdnHCl] implies that the $\alpha_1\gamma_1$ interface is more resistant to dissociation than the $\alpha_1\beta_1$ interface in the apo-dimer intermediate (Fig. 6.7). The S coefficients and m values for the various transitions of apoHbF were very similar to those for apoHbA (Table 6.1).
Figure 6.6. GdnHCl induced equilibrium unfolding of apoHbF as function of total protein concentration. The solid circles are the observed data, and the solid lines are the fitted curves to the 4-step apoHb heterodimer unfolding mechanism (Fig. 5.7) obtained from global fittings across varying apoHb protein concentrations using eq 5.20. Unfolding measurements of apoHbF were done in 200 mM potassium phosphate, 1 mM DTT, pH 7 at 10° C.

The $K_{I_D}$ of 128 ± 11 for apoHbF indicates similar stabilities for the heme pockets in $\gamma$ and $\beta$ subunits when compared with apoHbA’s $K_{I_D}$ value (Fig. 6.7, Table 6.1). The $K_{2UM,JD}$ value for apoHbF is, however, ~10-fold larger, $84 \pm 22 \times 10^8 \text{ M}^{-1}$, compared to the value of this equilibrium constant for apoHbA (Table 6.1). Previous studies have suggested that the $\gamma_1\gamma_2$ and $\alpha_1\gamma_1$ interfaces in $\gamma$ homo-oligomers and HbF respectively are stronger than the $\beta_1\beta_2$ and $\alpha_1\beta_1$ interfaces in $\beta$ homo-oligomers and HbA, respectively, due to the increased hydrophobicity of $\gamma$Ile116 compared to $\beta$His116 at these interfaces.\cite{133,215,216} Interestingly, the value of $K_{U_2U_2}$ for
dissociation of the apoHbF $U_2$ species also decreased ~30 fold to $0.16 \times 10^{-7}$ M compared to $5.5 \times 10^{-7}$ M for apoHbA (Table 6.1). Thus, the results in Figs. 6.6 and 6.7 suggest that the increased apolar character of $\gamma$ chains stabilizes the $\alpha_1\gamma_1$ interface in the $I_D$ state of apoHbF. This increase in surface hydrophobicity of $\gamma$ chains also appears to stabilize the unfolded $U_2$ state, which partially compensates for the favorable effect on $\alpha_1\gamma_1$ dimer formation and decreases the dependence of the second phase of unfolding on protein concentration (Figs. 6.6 and 6.7). The smaller $K_{U_2,2U}$ value for apoHbF keeps the monomeric, partially unfolded subunits together and reduces the entropy of the unfolded subunits\textsuperscript{213,220-222}, facilitating reformation of the apoHbF molten globule.

Figure 6.7. Comparison between GdnHCl induced equilibrium unfolding curves for apoHbA and apoHbF. Protein concentration was 12 $\mu$M total protein subunits for both proteins. The solid circles are the observed data, and the solid lines are the fitted curves to the 4-step apoHb heterodimer unfolding mechanism (Fig. 5.7) obtained from global fittings of unfolding measurements to eq 5.20. Unfolding measurements of apoHbF and apoHbA were done in 200 mM potassium phosphate, 1 mM DTT, pH 7 at 10°C.
The data in Figs. 6.6 and 6.7 represent the first attempt in literature to compare directly the equilibrium association constants for apoHbF and apoHbA heterodimer formation. Past experiments in literature focused on measuring holo-α and holo-β/γ subunit assembly and/or dissociation rate constants, which had the added complexity of competing reactions that form homo-oligomers, \( \beta_4 \) and \( \gamma_2/\gamma_4 \) formation\textsuperscript{215,216,223}. In the latter case for disassociation rate constants, exact interpretations required computation of the fractional amount of heterodimer present in the sample\textsuperscript{214,223}. Our results provide a more direct comparison for \( \alpha_1\beta_1 \) and \( \alpha_1\gamma_1 \) apodimer formation which is probably the most relevant assembly reaction during erythropoiesis.

### 6.6 Conclusions

The three-dimensional structural homology between human Hb subunits and monomeric mammalian myoglobin finds a parallel in their folding mechanisms. The results in both Chapters 5 and 6 suggest strongly that the pathway for human apoHb unfolding is very similar to that for mammalian apoMb.

ApoMb unfolds at neutral pH by a 3-state, 2-step mechanism involving a molten globule intermediate, which retains \( \sim40\% \) of the helical content of the native structure\textsuperscript{9,13,27,39,41}. The intermediate (I) folding state for sperm whale Mb has been characterized by NMR and consists of an unfolded heme pocket and a folded hydrophobic core of A (N-termini), G, and H (C-termini) helices with evidence for another minor intermediate with additional partial folding of the B helix\textsuperscript{73,74}. Results from the Olson laboratory\textsuperscript{9,13,41,71} as well as that of of Wright and Baldwin\textsuperscript{15,27,73,74,217} established that the initial phase in apoMb unfolding involves “melting” of
heme pocket to form a molten globule state. The introduction of large, apolar amino acids at the 7th and 11th positions of the E helix (i.e., His(E7)Phe, His(E7)Leu, and Val(E11)Phe) on the distal side of the heme pocket markedly inhibits the first phase of unfolding by stabilizing native apoMb. This stabilization also dramatically increases overall holoMb expression in animal muscle, *E. coli*, and cell-free translation systems.$^9,13,41,71,217$

In the case of human apoHb, the initial step is formation of a dimeric molten globule intermediate via the “melting” of the heme pockets and is followed by concerted dissociation and formation of unfolded monomeric α and β or γ chains. As with apoMb, the introduction of His(E7)Leu and Val(E11)Phe mutations on the distal side of the heme pocket in the α and β subunits of HbA significantly increases the stability of the native apoHb dimer (Table 6.1, Fig. 6.1). The heme pocket mutagenesis results for apoHb, combined with previous studies for apoMb, leads to the conclusion that unfolding of the heme pocket precedes formation of the molten globule heterodimer.

For both apoHbA and apoHbF, the helical content of the molten globule dimeric state ($I_ν$) is approximately ~40% relative to the original apoHb folded dimer (D) (Table 6.1). The unfolding of the $I_ν$ intermediate shows a dependence on protein concentration, requiring higher [GdnHCl] at higher $P_0$ (Figs. 5.8 and 6.6). This result suggests strongly that the intermediate is still an $α_1β_1$ heterodimer. Bioinformatics studies by Ptitsyn$^7$ suggested strongly that the AGH helices act as a nucleus for globin folding, as seen experimentally with the molten globule of apomyoglobin$^{73,74}$. Therefore, retention of A, G, and H helical secondary structure is likely needed for the strong hydrophobic surface interactions to form at the $α_1β_1$ interface. The idea that the second phase involves hetero-dimer dissociation into monomers is further strengthened by the unfolding results for apo-rHb0.1. The di-α linker in rHb0.1 markedly shifts this second
phase towards much higher [GdnHCl] concentrations due to the presence of two $\alpha_1\beta_1$ interfaces (Figs. 6.4 and 6.5).

The *in vitro* unfolding results for human apoHb also provide insight into some key physiological observations relevant to the folding and assembly of human holoHb *in vitro* and during erythropoiesis in pre-erythroid cells. Isolated $\alpha$ monomers are unstable in solution, even in the presence of bound heme. The absence of a protein oligomer partner leads quickly to denaturation, a hallmark of the more serious $\beta$-thalassemia diseases, in which one or more of the $\beta$ genes are defective in the patient. In contrast, excess $\gamma$ and $\beta$ chains can self-assemble into more stable homo-dimers and tetramers, accounting for why $\alpha$-thalassemia is often less severe in terms of anemia. In effect, the formation of these homodimers and tetramers helps inhibit complete unfolding of $\beta$ and $\gamma$ subunits in the same way the $U_2$ and $U_3$ states inhibit complete unfolding of the monomeric subunits in the mechanisms shown in Figs. 5.7 and 6.3. Thus, the $U_2$ or $U_3$ states probably also represent non-specific binding that could occur to form transient homo-$\gamma$ or $\beta$ oligomers. The subunits in these partially folded forms are preserved until $\alpha$ subunits are present to form stable heterodimers and tetramers. Previous work has shown that $\gamma$ homo-oligomers are much more stable than $\beta$ oligomers, again due to increased hydrophobicity at their dimer interfaces. The $\sim$30 fold decrease in the $U_2$ dissociation equilibrium constant, $K_{U_2\rightarrow U_M}$, for apoHbF relative to apoHbA is consistent with this idea that $\gamma$ subunits can form more stable oligomeric interactions with each other (Table 6.1).

Finally, the requirement of globin heterodimerization before formation of the apoHb intermediate state accounts for why the $\alpha$ and $\beta$ chains have to be translated in equal amounts for efficient expression of holoHb in *E. coli* and during erythropoiesis. This observation
suggests that the partially folded heterodimer has to be formed before hemin can be bound. Isolated apo-α and β subunits do not form stable molten globule states by themselves and, in the unfolded states, are unlikely to bind heme specifically before precipitating.
Chapter 7

Human met- holohemoglobin unfolding

7.1 Introduction

The unfolding mechanism of ferric holohemoglobin A involves an on-pathway hemichrome intermediate, which is formed through the reversible binding of hemin to the molten globule apo-dimer. By incorporating the previously determined 5-state model for apohemoglobin\(^{16}\) (Chapters 5 and 6), I have derived a simple model for human methemoglobin assembly based on GdnHCl-induced unfolding curves that were obtained from analyzing simultaneously changes in CD, fluorescence, and visible absorbance. CD spectra measure helical secondary structure, fluorescence emission intensity measures the loss of hemin, which quenches protein tryptophan fluorescence when bound to the protein, and absorbance spectra can be used to determine the fraction of metHb remaining, the appearance of hemichrome intermediate, and the amount of dissociated free hemin.

Culbertson and Olson\(^{31}\) proposed a similar mechanism for the folding of tetrameric HbA based on their mechanism for the folding of the monomeric holoMb, but they did not test their ideas quantitatively with either apo- and holoHb unfolding experiments. According to their original mechanism, the holo-heterotetramer either loses hemin (H) to form an apotetramer, which then dissociates into \(\alpha,\beta_1\) apodimers, or the holotetramer dissociates into holodimers, which then lose hemin. Hemin can bind reversibly with either apoglobin species but has a lower affinity for dimers\(^{31}\). They also suggested that the resulting dimers undergo a
further simple 2-step unfolding mechanism with hemin binding reversibly to both a dimeric molten globule intermediate and an unfolded state. This 8-step model involved complex quartic equations, and they did not pursue simulations or experimental analysis.

My thesis work and past studies have shown that hemin removal results in formation of an apohemoglobin dimer with no evidence of an apo-tetramer unless the protein is genetically crosslinked. The results in Chapter 5 show that hemin binding to the proximal histidine on the F helix in the apoHb dimer stabilizes the proximal side of the heme pocket which concurrently enables the \( \alpha_1\beta_2 \) interface interaction leading to tetramer formation at high concentrations of hemoglobin.

Like oxyhemoglobin, methemoglobin is in the R quaternary state, and past work has shown that holoHb tetramers in the R state readily dissociate into the \( \alpha_1\beta_1 \) dimers in the 1 to 10 \( \mu \)M concentration region due to weak contacts at the \( \alpha_1\beta_2 \) interface. The equilibrium tetramer to dimer dissociation constants, \( K_{4,2} \), for methemoglobin and oxyhemoglobin are \( \sim 10^{-5} \) and \( \sim 10^{-6} \) M, respectively. In contrast, deoxyhemoglobin, is in the T state; the extent of interaction at the \( \alpha_1\beta_2 \) interface increases markedly; and \( K_{4,2} \) decreases dramatically to the \( 10^{-11} - 10^{-12} \) M range, resulting in little dimer formation even in very dilute solutions. Thus, in contrast to deoxyHb, the metHb form readily dissociates into dimers in dilute solution in the micromolar region. During unfolding, dissociation into dimers almost certainly occurs before hemin dissociation, which exhibits equilibrium dissociation constants in the nanomolar region.

HoloHb equilibrium unfolding experiments done in this work and various past Hb denaturation studies have noted the formation of low spin Fe (III) hemichromes intermediates. Dithionite can reduce the hemichrome directly to a deoxyHb state.
with pentacoordinate heme, and the hemichrome intermediate is thought to be one of the sub-populations of partially-folded hemoglobin that is on the pathway for assembly of the holoprotein. However, when the addition of dithionite leads to formation of a hexacoordinate hemochrome, the original hemichrome species reflects extensive structural changes. These changes appear irreversible when following the removal of the denaturing reagents or conditions, and the native state of the hemoglobin is not recovered.

7.2 Tetramer-dimer (α₁β₂) interface interaction as a factor governing hemin binding

In order to determine whether hemin disassociation directly disrupts the α₁β₂ interface, SAXS analysis was done on rHb0.1 as a model system. Human apohemoglobin exists as a dimer even at high protein concentrations. In contrast, the K₄₂ range for dissociation of α₁β₂ interface in holo-methemoglobin is 1-10 µM so that at ≥ 100 µM almost of the protein is a tetramer. NMR structural characterization of sperm whale apomyoglobin has shown loss of helicity in the F helix, which is where the proximal histidine coordinates to the hemin iron atom. The F helix is critical for formation of the α₁β₂ tetramer interface in Hb. Thus, the complete disassociation into apoHb dimers following hemin disassociation is most likely due to disorder of the F helix resulting from the loss of the HisF8-iron coordination. Unfortunately, human apoHb is unstable at room temperature, making structural characterization of the apo-form extremely difficult, and, as result, no work in the literature discusses or validates this idea.
In our analytical gel filtration analysis\textsuperscript{16}, genetically crosslinked apo-rHb0.1 exists as a tetramer, but this protein elutes somewhat earlier than either tetrameric holo-HbA or holo-rHb0.1 (Fig. 5.3). This result suggests a non-compact structure for apo-rHb0.1 analogous to what would be expected for two independent $\alpha_1\beta_1$ dimer units covalently held together by the glycine linker. In our analysis of apo-rHb0.1 unfolding, we assumed that the $\alpha_1\beta_1$ tetramer interface is disrupted\textsuperscript{16}. If verified, this idea for apo-rHb0.1 would support our structural interpretation for why native apoHbA is always a dimer, \textit{i.e.}, the $\alpha_1\beta_2$ interface requires bound hemin. Thus, we performed small angle X-ray scattering experiments (SAXS) to examine the radius of gyration and globular shape of various forms of rHb0.1. The goal was to show that apo-rHb0.1 is opened up into structure resembling two dimers held together by the glycine linker.

### 7.3 SAXS analysis of rHb0.1

When the scattering curves obtained from the SAXS measurements for apo-rHb0.1 and holo-rHb0.1 were fitted to the Guiner plot equation, $\ln(I(q)) = \ln(I(0)) - q^2R_g^2/3$, the radius of gyration ($R_g$) obtained for apo-rHb0.1 is 31 Å, which is significantly higher than for holo-rHb0.1, which is 27 Å (Fig. 7.1A). CRYSOL software predicted an even more smaller $R_g$ of 25 Å for holorHb0.1 from an analysis of theoretical scattering curves simulated for the deposited 1011 PDB structure. Regardless of this discrepancy, the electron pair distance function ($P(r)$) profile was also much broader for apo-rHb0.1 than for holo-rHb0.1, indicating a more extended state for holo-rHb0.1 (Fig. 7.1B).
Previously published SAXS work\textsuperscript{227} showed a more Gaussian like P(r) profile for holo-HbA in comparison to our results for holo-rHb0.1. Thus, we probably need to obtain SAXS measurements for holo-rHb0.0 as a control to see if the V1M mutation at the N-termini of the subunits in the holo-recombinant hemoglobins is playing a role. However, the slight deviation of Rg and P(r) holo-rHb0.1 from an ideal globular shape could also be attributed to hemin disassociation during sample collection due to the longer exposure times of the samples to X-ray radiation at the home source. Rigid body modeling attempted in CORAL based on the experimental scattering curves also suggested a significantly more compact structure for holo-rHb0.1, compared to apo-rHb0.1 which appears to be significantly more elongated.

The net result of these observations is that our idea that apo-rHb0.1 can be considered as two $\alpha_i\beta_1$ dimers strung together with a glycine linker is appears to be correct. The larger radius gyration correlates at smaller elution volume for apo-rHb0.1 than HbA tetramers or holo-rHb0.1. Both observations support the idea that hemin removal completely disrupts the $\alpha_i\beta_1$ interface, which simplifies both the apo- and holo- hemoglobin A mechanisms for unfolding by ruling out apoHb tetramer formation\textsuperscript{16}. 
7.2 HoloHbA assembly model including hemin binding

Our first proposed model for the assembly (or unfolding) of holoHb is shown in Fig. 7.2 and is based on the 4-step, 5-state apoHb assembly model shown in Fig. 5.7. The equilibrium constants are defined in direction of forming the native holoprotein dimer (DH$_2$) or tetramer (TH$_4$). According to this model, each of the $\alpha$ or $\beta$ unfolded monomeric polypeptide chain ($U_C$) that come off the ribosome partially folds into $U_M$ that contains some small fraction of helicity. This species can then either transiently bind hemin (H) to form $U_H$ or bind other globin chains to form a dimeric unfolded species ($U_2$). The $\alpha$ $U_M$ species and $\beta$ $U_M$ species can also interact to form the $\alpha_1\beta_1$ interface generating the heterodimeric molten globule (I$_D$) state.
which can bind H reversibly to form a hemichrome species \((\text{IH}_2)\). Further folding of the apo-I\(_D\) state leads to formation of the folded apoHbA dimer (D). The binding of heme to the apo-D state, leads to the formation of holoHbA dimer \((\text{DH}_2)\) dimer, containing a fully folded F helix as well as an intact \(\alpha_1\beta_2\) interface. Two DH2 metHb species can form a tetramer at high protein concentrations leading to the final tetramer holoHb \((\text{TH}_4)\).
Values for the hemin binding constants (K_{DH2}, K_{IH2}, and K_{UH}) can, in principle, be obtained by a combined analysis of holo- and apo-Hb unfolding curves as was done for myoglobin in Chapter 4. This analysis can be performed by analyzing holoprotein unfolding curves and fitting for the hemin binding parameters, K_{DH2}, K_{IH2}, and K_{UH}, K_{ID,D}, K_{2UM,ID}, and the other apoglobin parameters can be fixed to values obtained from independent analyses of the apoHb unfolding curves, as described in Chapters 5 and 6.

The equilibrium hemin binding and dimer-tetramer assembly constants are defined in eq 7.1. The total protein concentration, P_0, (eqs 7.2 and 7.3) can be derived in terms of these constants (eq 7.1) and the apoHb folding constants described in eqs 5.3, 5.7, and 5.12. The result is a complex equation containing two independent variables of [U_M] and [H] and terms containing these variables taken to the fourth power (eq 7.3).

\[
K_{2,a} = \frac{[TH_4]}{[DH_2]^2} = K_{2,a}^0 \exp \left( \frac{-m_{2,a}[\text{GdnHCl}]}{RT} \right)
\]
\[
K_{DH2} = \frac{[DH_2]}{[D][H]^2} = K_{DH2}^0 \exp \left( \frac{-m_{DH2}[\text{GdnHCl}]}{RT} \right)
\]
\[
K_{IH2} = \frac{[IH_2]}{[I][H]^2} = K_{IH2}^0 \exp \left( \frac{-m_{IH2}[\text{GdnHCl}]}{RT} \right)
\]
\[
K_{UH} = \frac{[UH]}{[U][H]} = K_{UH}^0 \exp \left( \frac{-m_{UH}[\text{GdnHCl}]}{RT} \right)
\]

**eq 7.1**

\[
P_0 = [\text{UH}] + 4[\text{TH}_4] + 2[\text{DH}_2] + 2[\text{IH}_2] + 2[\text{D}] + 2[\text{I}] + 2[\text{U}_2] + [\text{U}_M] + [\text{U}_C]
\]

**eq 7.2**

\[
P_0 = K_{U_M,UH}[\text{U}_M][\text{H}] + \left( K_{2,4}^2 K_{DH}^2 K_{1,D,D}^2 K_{2UM} K_{1D} [\text{U}_M] [\text{H}]^2 + K_{DH} K_{1D,D} + K_{UH} \right) K_{2UM} K_{1D} [\text{U}_M] [\text{H}]^2 + 2 K_{1,D,D} K_{2UM} + \frac{1}{K_{U_2,U_M}} [\text{U}_M]^2 + \left( 1 + \frac{1}{K_{U_2,U_M}} \right) [\text{U}_M]
\]
Free hemin concentration \([H]\) is defined as the total apoHbA concentration in solution (eq 7.4) as was previously done for holoMb assembly model by Culbertson and Olson\(^9\). Given that myoglobin is a monomer, Culbertson and Olson\(^9\) were able to solve for \([U_M]\) in their Mb analyses using a quadratic equation by inserting \([H]\) into total \(P_0\). However, for the holoHb tetrameric model, inserting eq 7.4 into eq 7.3, results in a much more complex equation with \([U_M]\) taken to the 8\(^{th}\) power. Assuming that roots for \([U_M]\) and \([H]\) can be obtained, the fractions of the various folding intermediates are defined in eq. 7.5.
7.3 HoloHb unfolding measurements

Following the approach taken for apoHb in Chapters 5 and 6, unfolding curves were measured for: metHbA, metHbF, met-rHb0.1 and the holo-met forms of the distal pocket mutants: α(H58L/V62F)β(wt), α(wt)β(H63L/V67F), and α(H58L/V62F)β(H63L/V67F). The experimental conditions were similar to those used for the apoHb experiments, with the samples being incubated at 10 °C for 1 hour with GdnHCl of varying concentrations in 200 mM potassium phosphate, pH 7. Unfolding measurements for metHbA, metHbF, and met-rHb0.1 were done at three different protein concentrations. The unfolding was followed by
CD, UV-Vis, and tryptophan fluorescence spectroscopy. However, in order to obtain a global human holoHb unfolding model to enable the unfolding data analysis of the different holoHb variants later, the unfolding measurements of holoHbA was analyzed extensively first and is the focus of this chapter.

Initially, experimental attempts were also made to measure unfolding of native metHbA with 1 mM dithiothreitol (DTT), as was done with apoHbA, but DTT slowly reduces the ferric iron atom in hemin, resulting in mixture of various oxidation states and the generation of reactive oxygen species due to autooxidation of oxygenated forms. I also tried doing unfolding experiments in 1 mM DTT with ferrous samples bound with either CO or oxygen in order to prevent reduction by DTT. However, the oxygen-bound sample started undergoing cyclic auto-oxidation and reduction processes, whereas the CO started disassociating when the protein unfolded. In order for each GdnHCl titration point to be consistent, these experiments needed to be done under constant flow of oxygen or CO. I then attempted the unfolding titrations with deoxyHb samples by preparing all my buffers under nitrogen gas, using gas-tight Hamilton syringes, and flushing my cuvettes with nitrogen in order to prevent oxidation of the protein samples. These manipulations proved to be too difficult to do rapidly and routinely in manner that would allow screening of the various hemoglobin variants.

Because the methemoglobin experiments could not be done in the presence of DTT, I was worried about the issue of formation of disulfide crosslinks as was observed during the preparation of apoHb (see Chapter 5). I tested for this problem by diluting samples incubated with ~6M GdnHCl. Roughly 90% folded metHbA with normal CD and absorbance spectral characteristics and little precipitation was recovered after incubation for several hours in buffer without denaturant.
7.4 Spectral deconvolution

Visible spectra for the holo-metHb samples were collected in the 350-660 nm wavelength region and then deconvoluted into the spectra of fully folded native metHb (NH), hemichrome intermediates (IH), and free hemin (H). I wrote a simple program in MATLAB R2016a (The MathWorks®, Inc., Natick, Massachusetts, United States), which used a linear regression algorithm to fit the raw observed data to the sum of the basis spectra for each species multiplied by its corresponding population fraction (Y). The NH species is assumed to be composed of heme bound dimers (DH₂) and tetramers (TH₄) that are in rapid equilibrium. I also attempted to deconvolute the Trp fluorescence emission spectra collected at the range between 310-380 nm into components. However, I was not able to obtain meaningful basis spectra because little or no fluorescence is observed for the NH states, the Trp fluorescence for the IH state is unknown, and only the apoHb states show significant emission intensity.

7.5 Analysis of holoHbA unfolding curves

As an initial test of the model in Fig. 7.2, I examined the unfolding of native holoHbA at three different total protein concentrations, P₀. Triplicate spectroscopic measurements was taken at P₀ = 1.1, 12, and 60 μM. The average of the data points at each [GdnHCl] was analyzed globally for all three P₀ values. When the visible absorbance spectral measurements were deconvoluted, the hemichrome and free heme population fractions are dependent on protein concentration but Y_NH (the sum of the population fractions of TH₄ and DH₂ states) was not (Figure 7.3). These observations correlate with the normalized CD signal measurements at
222 nm, which show that the first phase for the loss of helicity does not show a dependence on protein concentration (Figure 7.4).

Figure 7.3 Deconvolution of visible absorbance spectral measurements of holoHbA unfolding. Three spectral species were observed: folded holoHb (NH), hemichrome (IH), and free hemin (H). Each variant’s population (Y) as a factor of protein and denaturant concentration is visualized above. All the unfolding titrations were done in 200 mM potassium phosphate, pH 7 at 10°C using GdnHCl denaturant. The raw spectra used for deconvolution at each protein concentration were averaged over 3 sample measurements.
Figure 7.4 Normalized CD unfolding curves for metHbA All the unfolding titrations were performed in 200 mM potassium phosphate, pH 7 at 10°C using GdnHCl denaturant. The fitted lines were obtained as described in the text.

From Figs. 7.3 and 7.4, it can be seen the first phase of unfolding (between 0 M to ~1.7 M GdnHCl) involves the transition from native aquometHb species (NH), both TH₄ and DH₂, to the intermediates IH₂ and I. This conclusion is emphasized in the deconvolution of the visible absorbance spectra shown in Fig. 7.3 where both the decrease of the NH species and the gradual increase of hemichromes occur during this phase. At ~ 1.5 M GdnHCl the hemichrome species, IH₂ species starts to dominate and then from ~1.7 M to ~ 3.5 M GdnHCl, the second
major phase occurs leading to loss of hemin, and appearance of free hemin. The visible absorbance change associated with loss of the hemichrome species correlates with the loss of helicity in the second phase of unfolding seen in the CD measurements. The free hemin spectrum dominates at the end of the unfolding titration at $[\text{GdnHCl}] \geq 3 \text{ M}$. At the lowest protein concentration, background noise has a significant effect on the observed spectra, particularly for the final fractions of free hemin (see dashed line in Fig. 7.3).

In this model, the equilibrium unfolding parameters and intrinsic CD signals ($S$) for all the apo-folding species of HbA were fixed at the parameters previously determined from the independent analysis and experiments for apoHbA unfolding (Table 6.1). This step emphasizes the decoupling of the apoHb folding process from the heme binding steps and is critical in reducing the complexity of holoHb folding mechanism. The CD signals for the apo-species were renormalized in term of the fraction of negative ellipticity at 222 nm of apo- to holoHbA as done previously by Culbertson and Olson for the holomyoglobin model\(^9\).

The definition of the measured CD signal at given $[\text{GdnHCl}]$ ($S_{\text{GdnHCl}}$) is similar to that derived for apoHb and was obtained by taking the sum of the product of the population fraction of each folding species and its intrinsic CD signal (eq 7.6).

\[
S_{\text{GdnHCl}} = S_D Y_D + S_I Y_I + S_U Y_U + S_{U_2} Y_{U_2} + S_{U_c} Y_{U_c} + S_{U_H} Y_{U_H} + S_{T_H^4} Y_{T_H^4} + S_{D_H} Y_{D_H} + S_{I_H^2} Y_{I_H^2}
\]

(eq 7.6)

For the visible absorbance measurements, fitting was done to a matrix ($Y_{\text{Abs}}$) composed of the population fractions of the different heme bound species and free heme (eq 7.7), which were determined independently from deconvolution of the visible absorbance spectra at...
different $[P_0]$ and [GdnHCl] into fractions of holo-DH$_2$ and holo-TH$_4$ remaining, IH$_2$
(hemichrome) present, and free H appearing at the end of the reaction. These fractions are
defined by the model described in eq 7.5.

$$Y_{Abs} = [Y_{NH}, Y_{IH}, Y_H] \text{ (eq 7.7)}$$

The absorbance data in Fig. 7.3 and the fraction CD changes in Fig. 7.4 were fit
simultaneously to the model in Fig. 7.2, using eqs. 7.1-7.7. I used the Newton-Rhapson
numerical method$^{228}$ to solve for the $[U_M]$ and $[H]$ roots of the two non-linear eqs 7.3 and 7.4
simultaneously in MATLAB R2016a (The MathWorks®, Inc., Natick, Massachusetts, United
States). The numerical analysis algorithm for solving these equations is followed by a call for
MATLAB’s lsqcurvefit function for using non-linear least squares trust region reflective
algorithm$^{229,230}$. This function is called to globally fit the holoHb unfolding model equations in
term of the folding population fractions (eq 7.5) needed to fit simultaneously both the
fractional CD changes and visible absorbance fractions, $Y_{NH}, Y_{IH},$ and $Y_H$ across various total
protein concentrations, $P_0$ (eqs 7.6 and 7.7). The resultant folding parameters obtained from the
global fit were then iteratively fed back into the next loop to obtain new estimates of $[U]$ and
$[H]$, and the looping process continued until a minimum chi-square value of the global fitting
was reached and did not change. The algorithm can potentially be modified in the future for
describing the folding of other complex oligomeric proteins as well as for examining the effect
cofactors and ligand binding to the complex protein systems.
Figure 7.5. Comparison of fluorescence unfolding curves for metHbA with total predicted population fractions of partially and completely unfolded states. A. Measured fluorescence intensities for metHbA normalized to the limiting value at high [GdnHCl]. The conditions are described in Figs. 7.3 and 7.5. These data were not used in the global fitting analyses because it was not clear how to specify signals for the various intermediates. However, it is clear that the increased intensity at either 355 nm or 347 nm seems to follow the appearance of free hemin
shown in Fig. 7.3, which in turns reflects the apoprotein states that have lost their prosthetic group. B. A plot of total population fraction of IH$_2$, I$_D$, UH, U$_M$, U$_2$, and U$_C$ states computed using the fitted parameters obtained from quantitatively analyses of the CD and visible absorbance curves for native holoHbA unfolding

I also recorded fluorescence emission spectra for all the samples. The emission intensity increases markedly with increasing [GdnHCl], reaching a maximum for the completely unfolded state at high [GdnHCl] (Figure 7.5A). The emission peak occurs at 347 nm for P$_0$ 12 μM and 60 μM, but at 1.1 μM, the peak at the end of the titration was at 355 nm. Thus, normalized intensities at both wavelengths are plotted in Figure 7.5A. The emission of free tryptophan in solution is at 355 nm, but completely denatured spectrin, a heterodimer, showed fluorescence emission peak at 347 nm$^{231}$. Thus, it is not clear why the fluorescence peak shifts with protein concentration. Regardless, the unfolding curves normalized at either 347 nm or 355 nm did not show protein concentration dependence. Thus, fluorescence emission curves (Fig. 7.5B) seem to represent the total population fractions of partially and completely unfolded states, including the hemichrome IH$_2$ species when the heme is not longer rigidly bound. In Fig. 7.5B, the predicted sum of the population fractions of the IH$_2$, I$_D$, UH, U$_M$, U$_2$, and U$_C$ states is plotted versus [GdnHCl] at three different P$_0$ values, calculated using the global fitting parameters obtained from fits to the CD and visible absorbance data. These theoretical curves look similar to the measured fluorescence curves both in terms of dependence on [GdnHCl] and a lack of effect of protein concentration. Folded holoHbs (NH states) do not have significant fluorescence emission, and therefore the initial increase in fluorescence emission at intermediate [GdnHCl] argues strongly that even the IH$_2$ hemichrome species, which is only partially unfolded, has increased tryptophan fluorescence.
The simple model for hemin binding to the dimeric species assumes that there are no differences between the hemin affinities of the subunits and that hemin binding is cooperative; that is either both subunits have hemin bound in the IH₂ state or both do not have hemin bound, as in the I (or I₀) state. Both these assumptions are clearly approximations. Hemin loss is known to be faster from the β chains than α subunits based on kinetics experiments¹¹⁵, and there is no experimental evidence that hemin binding in dimers is cooperative. However, because hemin binding promotes folding of the heme pocket and facilitates formation of the α₁β₂ interface in tetramers and almost certainly strengthens the α₁β₁ dimer interface, cooperative hemin binding would seem logical for apodimers. Regardless, in the absence of these simplifying assumptions, the scheme for holoHbA folding would become overwhelmingly complex with DH, DH₂ and IH, IH₂ species compounded by each of the DH and IH species being composed of α₁Hβ₂ and αβ₁H intermediates. Because no one has tried to analyze holoHb unfolding quantitatively before, I chose to use the simple cooperative heme binding model shown in Fig. 7.2 as a first attempt.
Figure 7.6 **Predicted population fractions of heme variants during holoHbA unfolding.** Population fractions are shown for folded holoHb (NH), hemichrome (IH₂), and free hemin (H), and each variant’s population (Y) as a factor of protein and denaturant concentration is visualized above. These fractions are predicted from holoHbA unfolding measurements fitting to the holoHbA assembly model in Fig. 7.3. All the unfolding titrations were done in 200 mM potassium phosphate, pH 7 at 10°C using GdnHCl denaturant.

In order to achieve reasonable fits to this model, the $K_{\text{metHb}}$ value for metHb was fixed at $10^5$ M⁻¹ based on past experimental data and our analytical gel filtration analysis^{21}. Global fitting of the CD signals in Figure 7.4 and the raw deconvoluted population fractions in Figure 7.3 yielded $K_{\text{NH}}$, $K_{\text{IH}}$, and $K_{\text{UH}}$ values of $1.61 \times 10^{17}$ M⁻², $1 \times 10^{16}$ M⁻², and $1.1 \times 10^5$ M⁻¹, respectively, for HbA. All the parameters for apoHb unfolding were fixed to the values reported for apoHbA in Chapters 5 and 6. The values for $m_{\text{NH}}$, $m_{\text{DH}}$, $m_{\text{IH}}$, and $m_{\text{UH}}$ were determined to be 19.8 kJ mol⁻¹M⁻¹, 2 kJ mol⁻¹M⁻¹, 5.3 kJ mol⁻¹M⁻¹, and 0.00001 kJ mol⁻¹M⁻¹ respectively. The
extremely small value for \( m_{\text{UH}} \) predicts that GdnHCl does affect non-specific heme binding to the unfolded monomeric states (eq 7.1), as was also seen with U₂ formation for apoHb (Chapter 5). The \( S_{CS} \) signals for TH, DH, IH, and UH were either assigned or fitted to be 1,1, 0.24, and 0.07 respectively.

The population fractions were calculated using these fitted parameters and the results are shown in Fig. 7.6. The observed trends follow the trends observed for fractions of metHb, hemichrome, and free hemin that were obtained from deconvolution of the observed absorbance changes (Fig. 7.3). Some discrepancies do occur at low protein concentration due both to the background noise in the measurements and simplifying assumptions about heme binding to the various U states. A basis spectrum for the UH species should probably be included for deconvolution of the spectra at high [GdnHCl], but it is not clear how to obtain a "pure" spectrum for hemin non-specifically bound to an unfolded globin chain.

During the first phase of unfolding, no dependence on protein concentration was observed, although dissociation of the holotetramer into a holodimer must be occurring at low protein concentrations. The reason for this lack of dependence is shown in the population fraction curves shown in Fig. 7.7, which were computed from the fitted parameters for holoHbA. The \( \alpha_1 \beta_2 \) interface is weak and \( K_{4,2} \) is in the \( \mu \text{M} \) region. As result, it takes less than 0.5 \( \mu \text{M} \) GdnHCl to completely disassociate the tetramer into dimers, even at 60 \( \mu \text{M} \) total protein (Fig. 7.7). As result, the dominant process in the first phase is still unfolding of the heme pockets in the holodimer to create the IH₂ state.
Figure 7.7. Population fractions of holo-tetrameric and dimeric metHbA predicted from fittings of metHbA unfolding measurements. All the unfolding titrations were done in 200 mM potassium phosphate, pH 7 at 10°C using GdnHCl denaturant. The fitted lines were obtained as described in the text.

Because the $K_{DH}$ and $K_{IH}$ values represent the overall binding constant for simultaneous binding of two hemin groups, an estimate of the apparent association constant for hemin binding to an individual subunit in the D or I state is given by the square root of these values, $\sim 10^9$ M$^{-1}$ and $10^8$ M$^{-1}$, respectively, which correspond to equilibrium dissociation constants ($K_d$).
of 1 nM and 10 nM, respectively. These $K_d$ values are roughly 1,000-fold higher than the corresponding values for hemin dissociation from holoMb and are consistent with the much higher rates of hemin dissociation from metHb dimers (~2 to 40 h$^{-1}$) compared to that for meMb ($\leq 0.01$ h$^{-1}$) under physiological conditions.

Nonspecific heme binding was only modeled for the $U_m$ state and not for $U_c$ and $U_2$ species, under the assumption that heme binding is promoted by residual helicity. The value of $K_{UH}$ is very small and similar to the values estimated for $H$ binding the $U$ state of myoglobin, and the equilibrium dissociation constant for hemin dimer formation. Thus, the fraction of UH states is very small and at the end of the titration most of the hemin is free in solution. This conclusion is verified by the data in Fig. 7.5, where large increases in tryptophan fluorescence are observed during the last phase of unfolding and reflect the loss of quenching due to hemin dissociation.

### 7.6 Conclusions and future directions

The key conclusions from this chapter are similar to those obtained when looking at hemin binding to apoMb. In this case, the heme group stabilizes both the heme pocket of apodimers and facilitates the formation of native tetramers (Fig. 7.8). Given the complexity of holoHb model due to the various folding intermediates, the analysis of human Hb is never complete, but I have been able to show that equilibrium dissociation constants for hemin binding can be obtained by analyzing holoHbA unfolding curves using the parameters derived in Chapters 5 and 6 for apoHbA unfolding. As in the case of Mb, hemin can bind to the molten
globule state of apoHb dimers to generate hemichrome spectral intermediates that are on the assembly pathway.

Figure 7.8 Comparison between apo- and holo HbA and HbF. The unfolding titrations were done in 200 mM potassium phosphate, pH 7 at 10°C using GdnHCl denaturant for 12 μM protein; and for apoHb unfolding, 1mM DTT was also in buffer.

When combined, the work in this thesis argues strongly that the mechanism of globin assembly is by reversible binding of heme to folded and partially folded apoprotein in both native and molten globule states. In Chapter 4, I had shown that this mechanism, which was initially determined by Culbertson and Olson for sw holoMb, is conserved in 3 other mammalian myoglobins of diverging amino acid sequences, which evolved during the 200
million years of mammalian transition from land back to water\textsuperscript{42}. Perhaps even more remarkable is that the mechanism is also conserved in hemoglobins even with the differences in oligomeric state and sequences between myoglobins and the subunits of hemoglobin. In both cases, folding occurs by initial formation of a molten globule (likely involving the AGH helices), which in the case of hemoglobin requires formation of an $\alpha_i\beta_1$ heterodimeric interface. The final step involves unimolecular folding of the heme pocket active sites, which can occur naturally or be facilitated by heme binding.

As shown in Fig. 7.8 for holoHbA and holoHbF, hemin binding stabilizes both proteins, requiring higher [GdnHCl] to unfold the globin compared to the corresponding apoglobins. Using the model determined for metHbA assembly in this chapter, I plan to analyze the unfolding data for metHbF to further investigate the relationship between hemichrome formation and the strength of the $\alpha_i\beta_1$ ($\alpha_i\gamma_i$) interface. My preliminary analyses have shown that the unfolding of holoHbF leads to a significantly higher occurrence of hemichromes, which is consistent with the higher stability of the molten globule state, as was seen in the apoHbF unfolding work in Chapter 6. However, more extensive modeling is needed to quantitate these effects.
I have also collected data for the unfolding of the holoHb distal pocket mutants α(H58L/V62F)β(⁠wt⁠) , α(⁠wt⁠)β(H63L/V67F) , and α(H58L/V62F)β(H63L/V67F). As described in Chapter 6, these mutations were designed to decrease heme affinity at the expense of enhancing heme pocket folding. In the case of apoHb unfolding, these mutations shifted the initial unfolding transition to much higher [GdnHCl] because they stabilized the heme pocket. In effect, they are doing the same thing as hemin binding. This effect can be seen in the preliminary data shown in Fig. 7.9, where the unfolding curve of the α(H58L/V62F)β(⁠wt⁠) metHb mutant is similar to that of HbA indicating that the distal pocket mutation does not have much of an effect when the hemin is bound in the α subunit, in contrast to what is seen for
apoHb unfolding (compare Figs 6.5 and 7.9). However, the effect of these mutations in the
holoHb is more complex when introduced in both subunits together, and further analysis of the
data is needed.

Genetic crosslinking does stabilize the dimeric intermediate (rHb0.1 curve in Fig. 7.9). However, I have to expand the holoHb unfolding model to take into account its tetrameric structure and dissociation into three subunits, which will make the model even more complex (see Fig. 6.3) Nevertheless, qualitatively the results in Fig. 7.8 and 7.9 do support the overall mechanism of Hb assembly shown in Fig. 7.2 and indicate that my conclusion about the conserved nature of hemoglobin and myoglobin folding is valid.
Chapter 8

Future Directions

In this thesis work, quantitative models for the monomeric mammalian holoMb expression and overall tetrameric human holoHb assembly were derived by decoupling the apoglobin folding process from the heme binding steps. Culbertson and Olson initially introduced this approach for studying holoMb folding\textsuperscript{9,30,31}. Despite their differences in amino acid sequence and quaternary structure, mammalian Mb and human Hb, follow a similar folding and assembly trajectory. The three key conclusions drawn from my thesis work are: a) apoMb stability, and not heme affinity, determines mammalian holoMb expression; b) human apoHb unfolds via a dimeric molten globule intermediate formed by the initial unfolding of the heme pockets in both subunits; and c) heme binding markedly stabilizes the heme pockets, can occur to the molten globule, generating reversible hemichrome dimers, and facilitates assembly into Hb tetramers\textsuperscript{13,16}.

There are limitations to my apo- and holo-Hb folding models, which, although already complex, do make some simplifying assumptions. They do not allow for the differences in stability between the $\alpha$ and $\beta$ subunits, which is due, in part, to the spectroscopic limitations of the experimental methods. Similarly, these models do not resolve early folding events involving transient helix formation (i.e. differences between $U_C$, $U_2$, and $U_M$ in Fig. 5.7). Both these issues could be addressed using solution state NMR methods that could structurally resolve transient helical content in unfolded globins or even non-specific subunit interactions,
as well as the difference in helical content of the α and β subunits as folding of Hb proceeds\textsuperscript{191,232}.

Figure 8.1 **Assembly and misfolding of human hemoglobin A**

Despite these limitations, the models derived in this study are useful for understanding the altered stability and assembly of human Hb variants that give rise to clinically relevant blood disorders (Fig. 8.1), as well as for understanding differences between embryonic, fetal, and adult Hbs. The strong analogy between the simple Mb assembly mechanism and the more complex Hb assembly, particularly in terms of similar molten globule intermediates, suggests that these idea and approaches should apply to the folding of various members of the globin superfamily including both the single domain and multi-domain proteins.
Finally, my conclusion that apogobin stability is the key to holoprotein production can be applied to optimize expression of other recombinant globins in vivo in either E. coli or transgenic animals. The importance of the $\alpha_1\beta_1$ interface and the AGH helices can be used to engineer more stable globins, and these contributions\(^7\) could potentially be critical for producing blood substitutes with long shelf lives and in vivo half-lives.

8.1 Early folding events during human Hb assembly

According to solution state NMR studies by Wright et. al.\(^{233}\), unfolded apoMb at pH 2.3 still shows transient formation of the secondary structure in the A and H helical regions. At low pH, unfolded apoMb also demonstrated propensity to form transient hydrophobic interactions initially between G-H helical regions, followed by a collapse with the A helix region\(^{232,233}\). My 4-step, 5-state apoHb folding model (Fig. 5.7) also suggest the presence of transient helical content in the unfolded monomeric subunits. Thus, it is likely that real time NMR studies\(^{234}\) on dimeric apohemoglobin at low temperatures could yield information on protein dynamics and specific helical content of the helical segments in the different folding states proposed in my mechanism. In-cell NMR\(^{235}\) is another method to probe in vivo folding of Hb, and Fetler K., Simplaceanu V, and Ho C.\(^{236}\) have previously used 1H-NMR to probe oxygenation of HbA in intact red blood cells.

Pappu and Weaver\(^{237}\) proposed a diffusion-collision-coalescence model for the initial folding steps of apoMb. They showed that, if unfolded apoMb contains a certain degree of secondary structure in the A and H helices, then folding would be promoted by the initial coalescence of the G-H segments leading to the molten globular AGH complex followed by
BAGH complex formation. Presumably, as unfolded Hb subunits start to exhibit transient local domains of helical structure or hydrophobic clusters, these domains likely could coalescence in the individual $\alpha$ and $\beta/\gamma$ monomers followed by intermolecular collisions to form the heterodimer molten globule. Understanding these events is critical in ascertaining the factors that favor globin folding over aggregation. In particular, it is important to understand which specific point mutations disrupt or enhance initial helical formations and then switch the balance between folding and globin aggregation.

8.2 Expression and assembly studies of human embryonic Hb variants

Primitive erythropoiesis occurs in the yolk sack until 9th week of gestation, and the initial Hb variant synthesized in the yolk sack erythroblasts is Hb Gower I ($\zeta_2\varepsilon_2$) which is followed by the synthesis of Hb Gower II ($\alpha_2\varepsilon_2$). During the 6th week, well-defined erythropoiesis starts to occur in the liver with a gradual increase in $\gamma$ subunit synthesis, enabling switching from expression of embryonic Hbs to fetal Hb. Hb Portland I ($\zeta_2\gamma_2$) synthesis is observed during this period. Another embryonic Hb variant, Hb Portland II, ($\zeta_2\beta_2$), is rarely detected because only trace amounts of the $\beta$ subunits are expressed during initial embryonic development.

Manning et al. have attempted to examine heterodimer assembly of embryonic Hbs in vitro; however, no direct measurements of the equilibrium association of heterogeneous monomers to form dimers were made because of the strength of this interface interaction, which leads to equilibrium dissociation constants in the picomolar range. Similarly, affinity of apoHb for hemin has not been measured in a directly because the $K_d$ values are on the order
of $10^{-13}$ to $10^{-11}$ M. However, some past studies have suggested that heme affinities of the $\varepsilon$ and $\gamma$ subunits are 20-fold and 10-old higher than that of $\beta$ subunits. A stability study on these variants did not describe intermediate folding states, but interestingly this work suggested that Hb Gower II has higher stability than either Hb Gower I or Hb Portland I. However, all three variants much less stable than HbA. A key question to address in future studies is determine why embryonic Hbs are so much less stable than either HbF and HbA and whether this observation relates to the expression switching that occurs during early development. Hb folding model developed in Chapters 5-7 could be used to examine the assembly of embryonic Hbs and quantitatively assess the cause of their instabilities. The results might help to understand the factors governing erythropoiesis during early development and, at least, define quantitatively the folding parameters and heme affinity of the embryonic Hbs for comparison with HbF and HbA.

8.3 Understanding human blood disorders arising from inefficient globin expression or globin misfolding

The clinical relevance of Hb folding studies was introduced in Section 1.7, and also described in section 6.6. Gene clusters expressing $\alpha$ globins and $\beta$ globins are located on human chromosomes 16 and 11 respectively. Genetic point mutations, or multiple codon insertions or deletions can lead to disorders of globin synthesis, Hb misfolding, or Hb instability. Most of these instability mutations are catalogued in the Globin gene database server (http://globin.bx.psu.edu/). The Hb folding model can be used to understand the mechanisms behind these diseases and to predict unstable Hb mutants that have not yet been
reported clinically. In principle, genetic editing can then be used to treat potentially fatal hemoglobinopathies. For example, Traxler et al.\textsuperscript{244} had shown CRISPR-Cas9 can be used to edit the promoter of the $\gamma$ subunits in order to maintain its expression and prevent sickle cell disease in adults.

$\alpha$ subunits are expressed from 4 globin genes, 2 in each chromosome 16, and the phenotype of $\alpha$-thalassemia is only severe when 3 of these genes are deleted. In this condition, $\beta$ chains form the holo-tetramer HbH, which is unstable despite being a soluble tetrameric protein. If all 4 genes are deleted, then during the switching to HbF in fetus, the $\gamma$ subunits form the tetrameric Hb Bart and the newborn infant or fetus will experience fatal hypoxia due to infective oxygen delivery by the $\gamma$ subunits as a result of their high oxygen affinity\textsuperscript{245}. Again the mechanism in Fig. 5.7 could be used to probe the specific biophysical characteristics of HbH that lead to its instability, whether it is due to unstable heme pocket or weak dimer-monomer interface or due to low heme affinity of the subunits. Point mutations that lead to $\alpha$- and $\beta$-thalassemias could also be investigated in terms of whether these mutations disrupt formation of the molten apodimer or even switch the balance to monomer aggregation.

8.4 The unfolding pathways of the members of globin superfamily

Members of the globin superfamily are extensive, and spread across the 3 kingdoms of life. Based on phylogeny, these globins can globally be categorized into the following 3 groups: the 3-over-3 (3/3) $\alpha$-helical fold single domain globins and the multi-domain flavohemoglobins; the 3/3 globin coupled sensors and protoglobins; and the 2-over-2 (2/2) $\alpha$-helical single domain globins. The 3/3 single domain globins and flavohemoglobins exist in
both bacteria and eukaryotes, the 3/3 globin coupled sensors and protoglobins exist in bacteria and archaea, while the 2/2 globins exist in all 3 kingdoms of life\textsuperscript{246}.

Vertebrates globins, which are 3/3 single domain globins, likely evolved from a common ancestor that differentiated into neuroglobins and cellular globins about 600 million years ago. After myoglobins differentiated from hemoglobins, another ancestral chromosomal duplication event was thought to have led to the evolutionary split between cytoglobins and myoglobins\textsuperscript{247}.

Neuroglobins are expressed mainly in the brain, retina and various other nerve tissues. With a high affinity for oxygen, neuroglobins act as an oxygen storage protein similar to Mbs, as well as a neuroprotective factor, by reducing neuronal damage during hypoxia. Human neuroglobin is a hexa-coordinated monomeric protein (Fig. 8.2) that coordinates with proximal histidine on the F helix and the distal histidine on the E helix, in the absence of ligand, both in the ferrous and ferric state\textsuperscript{248,249}. Human neuroglobin does have a high affinity for oxygen, with increasing affinity at lower pH\textsuperscript{249}. The disordered and flexible CD-D region in neuroglobin is presumed to enable conformational changes that allow for oxygen binding, following displacement of the distal histidine, in the heme pocket. A large 120 Å\textsuperscript{3} cavity, lined by hydrophobic residues and which is not present in Mbs and Hbs, exist between the distal side of the heme pocket and the EF loop, and this cavity is thought to play a role in ligand diffusion into the heme cavity\textsuperscript{248}. 
Surprisingly, despite the presence of an energetically unfavorable cavity$^{248}$, apo- and holo (hexa-coordinated)- human neuroglobins have shown to be more resistant to both acid and urea mediated unfolding relative to myoglobins$^{25}$. Given that Picotti et al.$^{25}$ have shown in 2004 that it is experimentally possible to perform apo- and holo- unfolding experiments independently for neuroglobins, the next direction for stability studies on neuroglobins should be to obtain a quantitative model for folding describing spectroscopic measurements as we and
others have done for Mbs\textsuperscript{9} and Hbs\textsuperscript{16}. Picotti et al.’s\textsuperscript{25} study did not look for molten globule intermediate during the unfolding of neuroglobin, even though it seems to be common for all other globins\textsuperscript{16,39,74}. Holomyoglobins unfold via a hemichrome intermediate state, whereas neuroglobins are essentially hemichromes in the folded state\textsuperscript{248,249} and this altered structural difference may change the folding pathway from that observed for Mbs and Hbs.

Hexacoordination with the endogenous histidine ligand also increased the resistance of neuroglobin to heme loss\textsuperscript{25}, whereas hemichromes folding intermediates for Mbs\textsuperscript{13} and Hbs (Chapter 7) have much lower heme affinity. However like Mb, neuroglobin loses ~20\% helical structure when heme is extracted\textsuperscript{249}. In apoMb this loss of helical structure is due to F helix becoming disordered following disruption of coordination with the proximal histidine on the F helix. In contrast, the helical content of F helix in aponeuroglobin is conserved, which was attributed to tyrosine at residue 88 on the F helix\textsuperscript{250}, and instead other helical regions appear to partially unfold.

In cytoglobin, the heme Fe is also coordinate to both proximal and distal histidines in the native state\textsuperscript{251}. Human cytoglobin is expressed ubiquitously in human tissues and has been suggested to be involved in oxygen content regulation, nitric oxide dioxygenation, and scavenging of reactive oxygen species in cells, although no definitive function has been determined experimentally\textsuperscript{252}. Human cytoglobin is a larger monomeric protein containing 190 amino acids, with greatly extended N- and C-terminals regions compared to mammalian Mb\textsuperscript{253}.

Interestingly, thermal unfolding studies showed that ferric hexacoordinated human cytoglobin is also more thermally stable than either ferric horse heart Mb or remarkably CO bound HbA. The melting temperatures for holo-cytoglobin and neuroglobin were 100 °C and 95 °C respectively, while the ferric Mb and CO-HbA, respectively, had melting temperatures of 81
190

°C and 71 °C. Presently, there have been no studies found in literature on apocytoglobin unfolding. Resolving the folding and heme binding mechanism of cytoglobin will be critical in not only understanding the folding mechanism and assembly mechanisms of hexacoordinated globins, but will contribute towards the physiological functional studies of cytoglobin. Clearly, the heme hexaacordination in cytoglobin and neuroglobin is playing a role in stabilizing these proteins, presumably by increasing their affinity for heme. Finally, the approach of decoupling apoglobin stability from heme binding through combined analysis of apo- and hologlobin unfolding curves can be applied to any member of the growing globin superfamily including multidomain globins, where the individual domains would likely unfold in distinct transitions with and without heme.
REFERENCES


biological inorganic chemistry: JBIC: a publication of the Society of Biological
Inorganic Chemistry 12, 101-117.
kinase by reversible formation of an internal disulfide bond induced by the fungal toxin
in a batch system using wheat germ extract, Biosci Biotechnol Biochem 58, 1119-1123.
from apoprotein and heme, monitored by stopped-flow absorption, fluorescence and
Biochemistry international 21, 849-855.
formation in acetate solution. Parallels between hemozoin formation and
biomineralization processes, Bioconjugate Chemistry 40, 204-213.
on polymerization, Biochim Biophys Acta 9, 323-327.
[156] Itoh, T., Yamada, T., Kodera, Y., Matsushima, A., Hiroto, M., Sakurai, K., Nishimura, H.,
and Inada, Y. (2001) Hemin (Fe3+) and heme (Fe2+)-smectite conjugates as a model of
hemoprotein based on spectrophotometry, Bioconjugate Chemistry 12, 3-6.
immune colloidal gold strip for cetacean meat restraining illegal trade and consumption:
structure and function of myoglobin, Biochimica et biophysica acta 1794, 103-109.
cells using tangential flow filtration and Immobilized metal ion affinity chromatography,
Journal of chromatography. B, Analytical technologies in the biomedical and life
sciences 879, 131-138.
technology in NMR high-throughput structure determination, Methods Mol Biol 607,
127-147.
structure of the ferric heme iron in engineered distal histidine myoglobin mutants,
Journal of Biological Chemistry 267, 22843-22852.
expression, FASEB J 3, 2360-2370.


