Protein Engineering Studies of Myoglobin

by

Robert Maurus

B.Sc., The University of Guelph, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in
THE FACULTY OF GRADUATE STUDIES
The Department of Biochemistry and Molecular Biology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 1998

© Robert Maurus, 1998
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Biochemistry and Molecular Biology

The University of British Columbia
Vancouver, Canada

Date Aug 5/1988
Abstract

The goals of the work described in this thesis were to gain further insight into structure-function relationships in the oxygen binding protein myoglobin, alter the structure of this protein to enhance a low level latent peroxidase function, and to add a new functionality in the form of a manganese binding site. An initial step in these studies was the elucidation of the high resolution structure of recombinant horse heart myoglobin to confirm that this protein is properly expressed and folded in a manner comparable to horse heart myoglobin obtained from natural sources. The additional variant myoglobins that were studied were organized into three groups. The first was constructed to probe aspects of the role of the distal heme pocket residue His64. Replacement of this histidine with threonine generated a five coordinate heme iron atom in the ferric state, whereas substitution with tyrosine yielded a six coordinate heme iron atom. The lack of peroxidase activity in the His64Thr variant protein appears to be a direct result of an inability to participate effectively in proton transfer during peroxide heterolysis. For the His64Tyr variant protein, the difficulty of dissociating the distal tyrosine ligand from the heme iron center appears to prevent hydrogen peroxide binding to the heme and thus limit peroxidase activity. Additional experiments were designed to facilitate the interpretation and understanding of the structural consequences of ligand binding to myoglobin. In particular, structural characterizations were completed of azide – heme ligation complexes with wild-type recombinant horse heart myoglobin and the His64Thr variant protein. These studies established that there are no large structural rearrangements required for the formation of an azide complex with either protein. However, substitution of His64 with threonine did show two unexpected effects in terms of azide complex formation. These included the observation of two bound azide conformations and on the proximal side of the heme pocket, the presence of a shorter hydrogen bond between Ser92 and the proximal heme iron ligand His93. The second group of variant proteins studied was prepared to probe the influence of amino acid substitutions in the heme binding pocket. These included replacements at the distal heme pocket
residues Leu29 and Val67, and at Leu104 which is located on the proximal side of the heme. These studies showed that substitution of Leu29 for a tyrosine or a lysine can be accommodated in the heme pocket of myoglobin. The peroxidase activity of both variant proteins is higher than that of wild-type myoglobin. This increase in activity appears to be correlated primarily with an increase in the polarity of the distal heme pocket. Surprisingly, the observed increase in heme reduction potential for the Val67Arg variant protein was also accompanied by an increase in peroxidase activity to approximately 8 times that of wild-type myoglobin. The Leu104Asn variant protein also showed an approximate 3-fold increase in activity towards hydrogen peroxide. A third group of variant proteins was designed to build a functional manganese binding site into horse heart myoglobin. Structural studies of the Lys45Glu and Lys45Glu/Lys63Glu variant proteins were conducted and indicate manganese binding occurs at the solvent exposed edge of the heme with ligands being formed by Glu45 and the heme D-propionate group. These residues constitute a site that is comparable to the one found later in structural studies of a manganese peroxidase by other workers. These myoglobin variants show ~3-fold enhancement in the rate of oxidation of Mn(II) to Mn(III) in the presence of hydrogen peroxide. Additional structural studies using Cd(II) ions were also conducted, and in this case two binding sites were located. Overall, the present work demonstrates that amino acid substitutions can influence the latent peroxidase activity of myoglobin and that it is possible to construct a Mn(II) ion binding site at an appropriate position on the surface of myoglobin to promote detectable manganese peroxidase activity.
2.3 The Growth of Protein Crystals

2.3.1 The collection and processing of x-ray diffraction data

2.4 General Aspects of the X-ray Diffraction Experiment

2.4.1 Dependence of X-ray diffraction on structure

2.4.2 The phase problem

2.4.3 Phase determinations for different variant myoglobin crystals

2.5 Refinement of Initial Structural Models

2.5.1 An example of a typical structural refinement

2.5.2 The quality of structural models

2.6 Spectroscopic Characterizations

2.7 Peroxidase Activity Measurements

3.0 Myoglobin as a Structural Framework for Protein Engineering

3.1 Introduction

3.2 Experimental Procedures

3.2.1 Structural Studies

3.2.2 Circular Dichroism

3.2.3 Peroxidase Activity Measurements

3.3 Results

3.3.1 Structural Results

3.3.2 Thermal denaturation studies of recombinant wild-type horse heart myoglobin

3.3.3 Peroxidase activity

3.4 Discussion
6.2.5 Functional studies

6.3 Results

6.3.1 Global comparisons of variant and wild-type myoglobins

6.3.2 The Leu29Lys variant structure

6.3.3 The Leu29Tyr variant structure

6.3.4 The Val67Arg variant structure

6.3.5 The Leu104Asn variant structure

6.3.6 Peroxidase activity assays

6.3.7 Spectroscopic studies of the Leu29Lys variant protein

6.3.8 Thermal denaturation studies of the Leu104Asn variant protein

6.4 Discussion

6.4.1 The effects of substitutions at Leu29

6.4.2 The Val67Arg variant structure

6.4.3 The Leu104Asn variant structure

7.0 Design of a Metal Binding Site into Myoglobin

7.1 Introduction

7.2 Experimental Procedures

7.2.1 Crystallization and diffraction data collection methods

7.2.2 Refinement of the Lys45Glu variant protein structure

7.2.3 Determination of the Lys45Glu/Lys63Glu variant structure

7.2.4 Structural characterization of the complexes of Mn(II) and Cd(II) with the Lys45Glu/Lys63Glu variant myoglobin

7.2.5 Functional Studies

7.3 Results

7.3.1 Structural analyses of the Lys45Glu and Lys45Glu/Lys63Glu variant proteins
7.3.2 The Mn(II) and Cd(II) complexes formed with Lys45Glu/Lys63Glu variant protein ................................................................. 158

7.3.3 Peroxidase activity ............................................................................ 162

7.4 Discussion ............................................................................................ 164

8.0 Summary .............................................................................................. 169

Bibliography .............................................................................................. 173
List of Tables

1.1: Secondary structure elements in horse heart myoglobin ................................................. 17
3.1: Data collection parameters for recombinant wild-type horse heart myoglobin .................. 54
3.2: Refinement results for recombinant wild-type horse heart myoglobin ....................... 56
3.3: Heme geometry of wild-type and recombinant wild-type myoglobins ...................... 63
4.1: Data collection parameters for the His64Thr and His64Tyr variant myoglobins ........... 72
4.2: Refinement statistics for the His64Thr and His64Tyr variant myoglobins .................. 74
4.3: Heme geometry of His64 variant myoglobins ................................................................. 79
5.1: Data collection parameters for the azide forms of recombinant wild-type and 
    His64Thr variant myoglobins ............................................................................... 90
5.2: Refinement statistics for the azide complexed recombinant wild-type and 
    His64Thr variant proteins ................................................................................... 92
5.3: Details of heme iron coordination geometry for met and azide myoglobin forms .......... 96
6.1: Crystal growth conditions for the Leu29Lys, Leu29Tyr, Val67Arg and 
    Leu104Asn variant myoglobins ............................................................................. 112
6.2: Data collection parameters for the Leu29Lys, Leu29Tyr, Val67Arg and 
    Leu104Asn variant myoglobins ............................................................................. 114
6.3: Refinement statistics for the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins .......................................................................................... 117
6.4: Heme geometry of the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins .......................................................................................... 124
7.1: Data collection parameters for variant horse heart myoglobins and associated 
    metal complexes ..................................................................................................... 148
7.2: Refinement statistics for variant horse heart myoglobins and associated 
    metal complexes ..................................................................................................... 150
7.3: Heme geometry in variant proteins and associated metal complexes ....................... 159
7.4: Ligand distances to bound Cd(II) and Mn(II) ions ..................................................... 162
List of Figures

1.1. Schematic illustration of peroxide dependent enzymatic lignin degradation catalyzed by manganese peroxidase .................................................. 4
1.2. Stereo diagram of the active site of cytochrome c peroxidase ........................................... 9
1.3. Ball and stick representation of the protophorhyrin IX heme group .................................. 15
1.4. Ribbon representation of the polypeptide chain of horse heart myoglobin ......................... 16
1.5. Space filling representation of horse heart myoglobin ..................................................... 18

2.1. Schematic representation of the general scheme of experiments ..................................... 24
2.2. Schematic diagram of the hanging drop vapour diffusion technique for the crystallization of proteins .......................................................... 29
2.3. Example of a Wilson plot ................................................................................................. 32
2.4. Omit difference electron density map ............................................................................. 41
2.5. Plot of crystallographic R-factor over the course of restrained parameter least squares structural refinement ...................................................... 42
2.6. Example of a typical Ramachandran plot ....................................................................... 43
2.7. Luzzati plot of crystallographic R-factor as a function of resolution ................................. 44
2.8. An example of a typical assay plot for peroxidase activity ............................................. 48
2.9. A double reciprocal plot prepared from peroxidase rate data ......................................... 49

3.1. Plot of the average positional deviations between the wild-type and recombinant wild-type myoglobin structures ................................................................. 59
3.2. Stereo drawings of the full α-carbon backbones of wild-type and recombinant wild-type horse heart myoglobins ........................................................................ 60
3.3. Average thermal factors of recombinant wild-type horse heart myoglobin ..................... 62
3.4. Circular dichroism spectra of recombinant wild-type horse heart myoglobin ................... 65

4.1. Average deviations of the atoms of the His64Thr and His64Tyr variant structures from that of recombinant wild-type horse heart myoglobin ................................................. 76
4.2. The His64Thr and His64Tyr mutation sites .................................................................... 78
4.3. Alpha carbon backbones of the His64Tyr and His64Thr structures superimposed on that of recombinant wild-type myoglobins.................................80

5.1. Azide binding to recombinant wild-type and His64Thr variant myoglobins..................94

5.2. Average deviations of the atoms of the azide complexes from those of recombinant wild-type horse heart myoglobin and the His64Thr variant structures.....................95

5.3. The minor azide conformation observed in the His64Thr variant.........................97

5.4. Average thermal factors for the main chain atoms in the recombinant wild-type metmyoglobin and its corresponding azide form and metHis64Thr myoglobin and its azide form.................................99

5.5. Stereo diagrams of the azide complexes of recombinant wild-type and His64Thr horse heart myoglobins, together with the related wild-type complex in sperm whale myoglobin...........................................102

5.6. Schematic representation of the two canonical forms of azide bound to the heme iron atom of myoglobin.................................................................105

6.1. Average deviations of the atoms of the Leu29Lys, Leu29Tyr, Val67Arg and variant Leu104Asn structures from wild-type horse heart myoglobin.............120-121

6.2. Alpha carbon backbones of the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn structures superimposed on that of recombinant wild-type myoglobin.........................122

6.3. The Leu29Lys and Leu29Tyr mutation sites..................................................123

6.4. The Val67Arg and Leu104Asn mutation sites..............................................127

6.5. pH dependence of the electronic absorption spectrum of the Leu29Lys variant protein....130

6.6. EPR spectrum of the Leu29Lys variant protein............................................131

6.7. CD spectrum of the Leu29Lys variant protein.............................................132

6.8. The region about residue 104 in the wild-type and Leu104Asn structures................140

7.1. Alpha carbon backbones of the Lys45Glu and Lys45Glu/Lys63Glu structures superimposed on that of recombinant wild-type myoglobin..............................155

7.2. Average deviations of the atoms of the Lys45Glu and Lys45Glu/Lys63Glu structures from wild-type horse heart myoglobin.................................156-157

7.3. The site of Mn(II) binding........................................................................161

7.4. The sites of Cd(II) binding........................................................................163

7.5. The manganese binding site in manganese peroxidase.................................167
List of Abbreviations

Å.......................... Ångstrom unit (1 Å = 0.1 nm)
a,b,c.................. Crystallographic unit cell axis or axis length
α,β,γ.................. Crystallographic unit cell angles
ABTS.................. 2,2'-Azino-di[3-ethyl-benzothialozine-(6)-sulphonic acid]
B......................... Isotropic thermal factor
CcP................... Cytochrome c peroxidase
d......................... Distance
Δd..................... Average positional deviation
EDTA................. Ethylenediamene tetraacetic acid
EPR.................... Electron paramagnetic resonance
F_{obs}, F_{calc}........ Observed and calculated structure factors, respectively
FTIR.................. Fourier transform infrared
h,k,l.................. Miller indices of a reflection
r.m.s................ Root mean squared
Tris-HCl........... Tris(hydroxymethyl)aminomethane hydrochloride
u,v,w................ Positional parameters in Patterson space
x,y,z................ Positional parameters in the crystallographic unit cell

The conventions of the IUPAC-IUB Combined Commissions on Biochemical Nomenclature are followed for both three letter and one letter abbreviations for amino acids [J. Biol. Chem. 241, 527-533 (1966); J. Biol. Chem. 243, 3557-3559 (1968)] and for designating atoms and describing the conformational torsion angles of the polypeptide chain [J. Biol. Chem. 245, 6489-6497 (1970)]. Designations for atoms of the protoheme IX group of myoglobin follow the nomenclature of the Protein Data Bank (Bernstein et al., 1977; see also Figure 1.2). Variant proteins are referred to with an abbreviation consisting of the three letter code for the original wild-type residue, the residue number and the three letter code for the replacement residue.
Acknowledgements

I would like to thank all those who have contributed to the success of my studies. In particular, I would like to acknowledge the assistance of Ralf Bogumil, Nham Nguyen, Dave Burk, Emma Lloyd, Christie Hunter, Dean Hildebrand, Yaoguang Luo, Chris Overall, Harry Tong and Hai-Lun Tang. A special thanks to Grant Mauk and Lawrence McIntosh for their advice as members of my Ph.D supervisory committee and their careful reading of this thesis. I would also like to thank Professors Grant Mauk and Michael Smith for advice and access to equipment in their laboratories. I am also grateful to my research supervisor, Gary Brayer, for his encouragement and guidance throughout my studies. Support for these studies was provided by the Medical Research Council of Canada and the Protein Engineering Network of Centers of Excellence Program. Lastly I would like to thank my family for the support that they have given me over the years.
Chapter 1

Introduction

Of central interest to many areas of chemistry is the development of efficient and highly selective catalysts. A variety of approaches have been taken to explore this possibility with regards to particular reactions. These include the use of organometallic compounds, application of electrochemical techniques and the use of biological materials. In the case of using biological starting materials, three notable strategies have evolved. First, Tramontano et al. (1986) and Schultz (1988) have shown that antibodies raised against compounds that are thought to be transition state analogues of particular reactions can be catalysts of these reactions. Second, for ecological interests, genetic manipulation of bacterial metabolic pathways have permitted the bioconversion and processing of target substrates (Ramos et al., 1987). Third, and of direct relevance to the work in this thesis, is the advent of site directed mutagenesis which permits specific changes in the amino acid composition of proteins to create new enzymes or design novel catalytic activities into already existing proteins (Smith, 1982; Knowles, 1987). This latter approach requires a knowledge of the factors governing protein folding, the functional groups that participate in catalysis and the geometric arrangement of key amino acids.

The characteristics of several proteins have been modified significantly by altering a small number of amino acid residues. For example, the susceptibility of subtilisin to oxidation was decreased by replacing methionine with either alanine or serine (Estell et al., 1985). The pH profile of the same protein was altered with the modification of charged surface residues (Thomas et al., 1985). Efforts to modify the thermal stability of T4 lysozyme (Perry & Wetzel, 1984; Matthews, 1987) and Staphylococcal nuclease (Shortle et al., 1990; Green et al., 1992) have also been
successful. In another example, cofactor specificity of glutathione reductase towards NADP$^+$ has been altered to NAD$^+$ by modifying amino acid residues in the cofactor binding site (Scrutton et al., 1990).

The work in this thesis describes results obtained as part of a protein engineering approach to understand further structure-function relationships in the oxygen storage protein myoglobin, to alter the structure of this protein to enhance a low level latent peroxidase function, and to add a new functionality in the form of a manganese binding site. These latter two goals could potentially result in a catalyst for the $\text{H}_2\text{O}_2$ dependent oxidation of lignin as part of the bleaching process of wood pulp. The following sections of Chapter 1 provide background on the protein engineering effort directed at myoglobin. First, a short account of the current understanding of the mechanism of fungal wood degradation is given. This is followed by the background data available with regards to peroxidase chemistry and structure. A third section reviews our present knowledge of metal binding sites in proteins. Finally a review is given of the structural and functional properties of myoglobin which forms the structural platform for the studies in this thesis directed at the development of an efficient and stable manganese peroxidase for use in wood pulping.

1.1 Naturally Occurring Lignin Degradative Fungal Systems

In terms of dry weight, 15-30% of woody plant cell walls are comprised of a heterologous, highly cross-linked, three dimensional polymer of phenyl propane units, referred to as lignin (Crawford, 1981). This material forms a matrix which surrounds the cellulose in wood. Cellulose is the desired component for paper production and during the pulping process lignin is solubilized through degradation and derivatization so that cellulose fibres are liberated for paper manufacturing (Lundquist et al., 1977). Conventional methods of pulp processing employ a combination of mechanical and chemical approaches to achieve delignification. These chemical processes involve
Chapter 1: Introduction

large amounts of chlorine and sulfur compounds, and the undesirable discharge of the products of these reactions into the environment.

An elegant alternative to delignification would be the use of an enzymatic system capable of either assisting, or even eliminating conventionally used chemical processes. However, the heterologous nature of the lignin polymer renders it highly resistant to enzymatic hydrolysis (Gold & Alic, 1993). Nonetheless, the white rot basidiomycetes *Phanerochaete chrysosporium* and *Trametes versicolor* are capable of degrading lignin by secreting a number of enzymes. These include manganese peroxidase (MnP), lignin peroxidase (LiP), laccase and a $\text{H}_2\text{O}_2$ generating system (Kirk & Farrell, 1987). MnP has been isolated and shown to be a heme containing enzyme (MW ~ 46000 Da) that is capable of catalyzing the Mn(II) dependant oxidation of a variety of model phenolic lignin compounds (Glenn et al., 1986; Wariishi et al., 1991; Tuor et al., 1992). As illustrated in Figure 1.1, MnP oxidizes Mn(II) to Mn(III) with the carboxylic acid chelated form of this latter ion being capable of oxidizing phenolic lignin substructures by diffusing off the enzyme surface into the lignin matrix (Glenn et al., 1986). Mn(II) which diffuses out of the lignin matrix is in turn regenerated to Mn(III) by MnP, thereby closing the enzymatic cycle. The use of manganese ion as a carrier of oxidizing equivalents enables MnP to carry out its degradative function at a distance without having to enter the dense lignin matrix.

1.2 The Peroxidase Mechanism

In general, peroxidases are enzymes that catalyze the oxidation of a variety of substrates using hydroperoxides as a source of oxidizing equivalents. At the start of the work in this thesis, cytochrome $c$ peroxidase was the only enzyme in its class for which a three-dimensional structure was known (Finzel et al., 1984). In most cases peroxidases utilize hydrogen peroxide in their reaction pathways. These enzymes play important roles as antioxidants, in biosynthesis, as part of
Figure 1.1. Schematic illustration of peroxide dependent enzymatic lignin degradation catalyzed by manganese peroxidase (MnP). Degradation is accomplished by oxidizing Mn(II) to Mn(III), which when chelated by organic acids, acts as a diffusible oxidizing agent capable of degrading lignin components [Glenn, 1986 #626]. The use of the diffusible Mn(II)/Mn(III) redox couple enables MnP to oxidize the lignin matrix which is otherwise too dense to be penetrated directly by this enzyme.

degradation pathways and in cellular defense (Campa, 1991). Even though the physiological functions of naturally occurring peroxidases differ considerably, the overall reaction of such enzymes with hydrogen peroxide is similar. This can be described as:

\[
H_2O_2 + 2 HA \rightarrow 2 H_2O + 2 A^+ 
\] (1-1)
where A represents the substrate to be oxidized. For example, in the catalytic cycle of manganese peroxidase, A would represent manganese ions and the reaction would involve the oxidation of two Mn(II) ions to two Mn(III) ions.

The initial step in a peroxidase reaction involves binding a peroxide molecule to the heme reactive centre of the enzyme (E). This event is described in the following expression:

\[ E\text{-Fe}^{III} + \text{H}_2\text{O}_2 \rightarrow \text{HE}\text{-Fe}^{III}\text{-OOH} \]  \hspace{1cm} (1-2)

In the following steps of this process a proton dissociates from hydrogen peroxide thereby yielding a peroxy anion (OOH-) which is bound to the heme iron (Back & Van Wart, 1992; Collins & Loew, 1992). The dissociated proton is transferred to an active site histidine residue. This proton is then transferred to the terminal end of the peroxy anion, and upon cleavage of the original sigma O-O peroxide bond, a water molecule is released and Compound I is formed as illustrated in the following expression:

\[ \text{HE}\text{-Fe}^{III}\text{-OOH} \rightarrow \text{E^*}\text{-Fe}^{IV}=\text{O} + \text{H}_2\text{O} \] \hspace{1cm} (1-3)

One of the electrons required for the reduction of hydrogen peroxide comes from the peroxidase heme Fe(III) center leading to the generation of a ferryl iron (Fe(IV)=O). For most peroxidases, the second electron comes from the porphyrin ring to produce an oxyferryl species with a porphyrin cation radical (Blumberg et al., 1968; Dolphin et al., 1971; Hewson & Hager, 1979). In the case of cytochrome c peroxidase the source of the second oxidizing equivalent was found to be the side chain of Trp191 (Sivaraja et al., 1989).

After formation of Compound I the next steps of the peroxidase mechanism involve a one
electron reduction of Compound I to Compound II, followed by another one electron reduction of Compound II to regenerate the enzyme to its native state. This overall process is shown in the following expression:

\[
E^*-\text{Fe(IV)}=O + 2e^- + 2H^+ \rightarrow E-\text{Fe(III)} + H_2O
\]  

(1-4)

The particular substrate that ultimately provides the two electrons required to restore a peroxidase to its resting state is a function of its physiological role and can range from inorganic compounds to proteins.

In simplified terms, the normal peroxidase cycle may be represented by the following:

\[
E + H_2O_2 \leftrightarrow E-\text{I} + H_2O \quad k_1
\]

\[
E-\text{I} + AH_2 \leftrightarrow E-\text{II} + AH^- \quad k_2
\]

\[
E-\text{II} + AH_2 \leftrightarrow E + AH^- + H_2O \quad k_3
\]

where E-I and E-II are compounds I and II, AH_2 is the reducing substrate and AH^- a free radical product. By defining the initial velocity of the reaction as \(v_0 = -d[AH_2]/dt\) and knowing that for most substrates \(k_2\) is much larger than \(k_3\), the following relationship can be derived (Dunford, 1991).
Chapter I: Introduction

\[
\frac{2[E]}{v_0} = \frac{k_2 + k_3}{k_2 k_3 [AH_2]} + \frac{1}{k_1 [H_2O_2]} \quad (1-5)
\]

In this equation \([E]\) is the concentration of enzyme, and \(k_1, k_2\) and \(k_3\) are the rate constants for the reaction steps outlined previously. This equation provides a method for the direct measurement of enzyme activity and this is facilitated if a colored product is produced by the appropriate substrate \(AH_2\). By this approach, \(k_1\) can be obtained from the slope of a plot of \(2[E]/v_0\) vs. \(1/[H_2O_2]\), again assuming that \(k_2 \gg k_3\) (Dunford, 1991).

1.3 Peroxidase Structure

Yeast cytochrome \(c\) peroxidase (CcP) was the first peroxidase for which a high resolution three-dimensional structure was reported (Finzel et al., 1984). More recently, a number of additional peroxidase structures have been elucidated. These include lignin peroxidase (Poulos et al., 1993), manganese peroxidase (Sundaramoorthy et al., 1994), \textit{Coprinus cinerius} peroxidase (Petersen et al., 1994), a fungal peroxidase from \textit{Arthromyces ramosus} (Kunishima et al., 1994), mammalian myeloperoxidase (Zeng & Fenna, 1992; Fenna et al., 1995), and garden pea ascorbate peroxidase (Patterson & Poulos, 1995). Despite having considerable sequence variation there are strong structural similarities in the active sites of these heme containing enzymes. Although not discussed in detail in this thesis, there are also several known structures of non-heme peroxidases that include the seleno-enzyme glutathione peroxidase (Ladenstein et al., 1979; Epp et al., 1983) and NADH peroxidase (Stehle et al., 1991a; 1991b).

A detailed description of CcP is given below. Notably, much the same structural attributes have been found in other heme containing peroxidases and the following discussion is also
generally applicable to these enzymes. Cytochrome \textit{c} peroxidase consists of a single polypeptide chain of 294 amino acids (34 kDa). Approximately 50\% of the polypeptide chain of CcP forms 10 \(\alpha\)-helices, and a further 12\% is found in a \(\beta\)-sheet conformation. The non-covalently bound heme group is completely buried in the protein matrix between two distinct N and C-terminal domains. A channel of \(~ 5\) Å in diameter and \(~ 10\) Å in length allows solvent and substrate molecules access to one side of the heme pocket.

The active site of cytochrome \textit{c} peroxidase is illustrated in Figure 1.2. The ligation state of the heme iron atom is pentacoordinate in the resting enzyme. The proximal heme ligand is provided by His175, and this residue, along with Trp191 and Asp235, is involved in a hydrogen bond network. One role of hydrogen bonds formed by the carboxylate group of Asp235 is believed to be proper orientation of the His175 and Trp191 side chains. Another role of the His175-Asp235 interaction may be to maintain the pentacoordinate state of the resting enzyme and to stabilize higher oxidation states of the heme iron center generated during the peroxidase catalytic cycle (Valentine et al., 1979; Finzel et al., 1984). However, more recent studies have demonstrated peroxidase activity for CcP variants in which Asp235 is substituted for alanine, asparagine and glutamine, indicating that this residue might be important in maintaining the coordination state and reduction potential of the heme but is not essential in the cleavage of peroxide (Goodin & McRee, 1993, Ferrer et al., 1994). On the other hand, it has been shown that His175 is important in determining the rate of peroxidase activity even though the His175Gly variant exhibits some activity (McRee et al., 1994). Since imidazole bound to the proximal side in the His175Gly variant protein restored only 5\% of wild-type peroxidase activity, it appears a polypeptide chain anchored proximal imidazole needs to be present for optimum activity. Interestingly, it has been determined that high levels of peroxidase activity are retained if the proximal His175 ligand is changed to either
Figure 1.2. A stereo diagram of the active site of cytochrome c peroxidase (Finzel et al., 1984; PDB entry 2CYP). Shown are the heme group and those residues in the heme binding pocket that are implicated in the catalytic function of this enzyme.

a glutamine or glutamate residue (Sundaramoorthy et al., 1994). Also on the proximal side of the heme is Trp191 (Figure 1.2). A notable difference of cytochrome c peroxidase from other peroxidases is that the radical centre during peroxide cleavage has been found to reside on the side chain of Trp191 instead of the porphyrin ring (Sivaraja et al., 1989). Furthermore, replacement of Trp191 with a phenylalanine greatly diminishes peroxidase activity (Mauro et al., 1988).

The distal side of the heme binding pocket in CcP is dominated by the side chains of Arg48, Trp51 and His52. These residues are placed in a manner that is believed to facilitate the cleavage of the peroxide O-O bond through a charge separation process. For example, His52 is positioned to accept a proton from a heme bound peroxide molecule (Poulos & Kraut, 1980). In turn, His52 is also favourably positioned to donate this proton to the terminal end of the peroxy anion (Miller et al., 1994) and thereby promote efficient proton transfer during peroxide cleavage. Recent mutagenesis studies confirm this crucial role for His52 since a $10^5$ fold decrease in the rate of
peroxide reduction is observed when this residue is substituted by a leucine (Erman et al., 1992). The adjacent Trp51 is hydrogen bonded to a distal heme pocket water molecule and this appears to prevent the coordination of Trp51 to the heme iron atom in CcP (Smulevich et al., 1994).

Another highly conserved residue in the distal heme pocket of peroxidases is Arg48 which was originally thought to interact electrostatically with the peroxide anion and assist in peroxide bond cleavage (Poulos & Kraut, 1980). However, replacement of Arg48 with lysine, glutamate or leucine has been shown to have only a small effect on the rate of peroxide cleavage (Vitello et al., 1990; Erman et al., 1992; Bujons et al., 1997). These results suggest a less critical function for this residue and argue that the main role of Arg48 may be to stabilize the iron bound oxene atom (Miller et al., 1994).

1.4 General Features of Metal Ligation in Proteins
Introducing manganese peroxidase characteristics into myoglobin involves two coupled components. The latent peroxidase activity of myoglobin must be increased, and a functional manganese binding site must be created. Creation of such a metal binding site is feasible owing to a considerable background of studies that have characterized metal binding sites in a great number of metalloproteins.

The roles of metal ions in biological systems are diverse and involve both catalytic and structural functions (see reviews by Glusker, 1991; Tainer et al., 1991). The nature of the interactions formed by proteins with metal ions has been characterized by a number of different approaches. These range from analyses of primary sequence distributions of liganding residues (Valee & Auld, 1989; 1990), to three-dimensional structural determinations (Chakrabarti, 1989; 1990; Yamashita et al., 1990). From these studies a number of general features of protein-metal interactions are identifiable. For example, a metal ion is generally ligated by a shell of hydrophilic
groups containing oxygen, nitrogen or sulfur atoms. In particular, hard metal cations like Mn(II) are mostly bound by oxygen containing groups. In this case the predominant liganding residues observed are aspartate and glutamate, although glutamine, serine and threonine are also occasionally utilized. Soft and intermediate metal ions like Cu(I), Cu(II), Zn(I) and Cd(II) are frequently bound to the nitrogen and sulfur atoms of the side chains of histidine and cysteine (Gregory et al., 1993).

In addition to providing direct metal ion ligands, the protein surface plays a role in metal binding recognition and forming the overall binding site. Two general categories of binding sites can be distinguished. The first involves networks of hydrogen bonds, whereas the second is more reliant on bulk electrostatic or through space effects (Thomas et al., 1985; Russell & Fersht, 1987; 1987). With regards to the first group, side chains in metal binding sites are often found hydrogen bonded to other side or main chain atoms to achieve the correct positioning of metal coordinating atoms (Christianson, 1991; Glusker, 1991). In terms of the second type of binding site, these have evolved electrostatic surfaces of highly negative potential in the region of the binding site (Regan, 1993). In general, it has also been observed that in protein metal binding sites the hydrophilic shell binding the metal ion is often embedded within a larger shell of hydrophobic groups. That is, metals bind at centers of high hydrophobicity contrast (Yamashita et al., 1990). While this feature might restrict the flexibility of the side chains that make up the binding site, the surrounding low dielectric about the metal site would be expected to enhance electrostatic interactions between groups within the binding site and with the metal ion.

In the design of new high affinity metal binding sites on proteins considerable difficulty is encountered in properly orienting the functional groups of ligating residues. Additional features that need to be carefully considered are the length of hydrogen bonds, as well as methods to enhance the overall electrostatic and hydrophobic aspects of the metal binding site. Due to the
complexity of these interactions, artificially created protein metal binding sites generally exhibit lower affinities than their naturally occurring counterparts (Regan, 1993). In the construction of a manganese binding site on myoglobin these factors were important considerations, as were insights obtained from the structures of a number of proteins which bind manganese. Some examples of these are superoxide dismutase (Stallings et al., 1984; 1985), different lectins (Einspahr et al., 1986; Rini et al., 1993), phytohemagglutinin-L (PHA-L) (Hamelryck. et al., 1996) and inositol monophosphatase (Bone et al., 1994).

1.5 Protein Engineering of Myoglobin

1.5.1 Why myoglobin?

From a number of vantage points, protein engineering studies of myoglobin have the potential to provide considerable insight into protein structure-function relationships. In terms of this thesis, the three major goals were: to further understand the role of active site residues; to understand the basis for the latent peroxidase activity of this protein; and, to move towards the conversion of this protein into a manganese peroxidase for potential use in industry.

In terms of a manganese peroxidase activity leading to lignin degradation, it might seem that a naturally occurring manganese peroxidase would make a more logical starting point for optimization and use in an industrial pulping application. However, there are a number of disadvantages associated with this approach. These primarily include the inability to obtain this enzyme in sufficiently large quantities and its lack of stability under the conditions required for large scale pulping processes. In addition, when the work in this thesis was initiated there was no structural information available for a manganese peroxidase that might have allowed for modifications to enhance the overall stability of such proteins.

A potential alternative to the use of a natural manganese peroxidase in industrial pulping is the
development of a new enzyme with enhanced functionalities using protein engineering methods and myoglobin as a starting platform. Key advantages to this approach are the latent peroxidase activity of myoglobin (King & Winfield, 1963; Mauk et al., 1973), the fact that it has a heme binding pocket with similarities to that of cytochrome c peroxidase, and that the structure of CcP was available to assist in the design process (Finzel et al., 1984). Additional advantages in using the myoglobin system were the availability of an efficient system for expression and mutagenesis of horse heart myoglobin (Guillemette et al., 1991), the inherent stability of this protein (Hargrove, 1994), and the background of structural and functional information available for myoglobins.

1.5.2 Structural and Functional Properties of Myoglobin
While the main physiological function of myoglobin is in oxygen transport and storage (Theorell, 1934), it has also been shown to catalyze a number of other reactions to a limited degree. These include hydroxylation (Galaris et al., 1988), epoxidation (Oritz de Montellano & Catalano, 1985; Iwahashi et al., 1986; Catalano & Oritz de Montellano, 1987), demethylation (Kedderis et al., 1986), hydrogen peroxide reduction (King & Winfield, 1963; Mauk et al., 1973), uric acid peroxidation (Howell & Wyngaarden, 1960), dopamine decarboxylation (Tate et al., 1972) and heme conversion to bilirubin (Brown et al., 1981). Notably, myoglobin has the same prosthetic group and axial ligands as CcP, horse radish peroxidase and manganese peroxidase, but is lacking key residues that confer strong peroxidase activity to these other proteins (Poulos & Kraut, 1980). Nevertheless, myoglobin has been found to catalyze a variety of peroxide-dependent reactions like the epoxidation of styrene (Oritz de Montellano & Catalano, 1985) and the peroxidation of unsaturated lipids (Galaris et al., 1990).

The importance of the physiological role of myoglobin has meant that it has been, and continues to be, intensively studied (see reviews by Springer et al., 1994; Olson & Phillips, 1996).
Chapter 1: Introduction

Mammalian myoglobins generally consist of 153 amino acids and an iron protoporphyrin IX heme prosthetic group (Figure 1.3). The best characterized myoglobins are those from sperm whale and horse. Sperm whale myoglobin has the distinction of being the first protein for which a three dimensional structure was determined at atomic resolution (Kendrew et al., 1958; 1960). This original work led to the solution of the structures of various forms of sperm whale myoglobin. For example, both the met and deoxy sperm whale myoglobin structures have been determined (Takano, 1977; 1977), as have the structures of the oxy form (Phillips, 1978; 1980) and a recombinant form of the same protein (Phillips, 1990).

The overall protein fold of horse heart myoglobin is similar to that of sperm whale myoglobin (Evans & Brayer, 1988; 1990) and forms a compact structure of approximate dimensions 45 x 35 x 25 Å (Figure 1.4). About 75% of the polypeptide chain of horse heart myoglobin is folded into 8 \( \alpha \)-helixes that have been designated A through H (Table 1.1). The majority of the remainder of the polypeptide chain forms turns of varying lengths between these \( \alpha \)-helical sections (Figure 1.4). Helical segments are found to pack so as to form a hydrophobic pocket in which the heme prosthetic group is non-covalently bound. Of particular importance in the formation of the heme crevice are helices B, C, E, F and G. The degree of heme solvent exposure in horse heart myoglobin can be seen in the space filling representation shown in Figure 1.5.

Of the six coordination sites of the heme iron atom of myoglobin, four are occupied by pyrrole nitrogen atoms and a fifth ligand is provided through interaction with the NE2 atom of the imidazole ring of the side chain of His93. This latter amino acid is commonly referred to as the proximal histidine ligand (Figure 1.4). The sixth coordination site of the heme iron atom on the distal side of the heme is the site of oxygen binding. During the oxygen storage and transport functions of myoglobin the heme iron atom is in the ferrous (+2) oxidation state and reversibly binds dimolecular oxygen and other diatomic gases like CO and NO (Antonini & Brunori, 1971).
In addition to the ferrous form, under physiological conditions the heme iron slowly auto-oxidizes to the ferric (+3) form which can bind a water molecule (metmyoglobin) or alternatively a number of different anions including fluoride, cyanide and azide (Antonini & Brunori, 1971).

Considerable effort has been directed at understanding the role the protein matrix has on determining the chemistry of the heme of myoglobin. This has been aided by the advent of site
Chapter 1: Introduction

Figure 1.4. A ribbon representation of the polypeptide chain of horse heart myoglobin. A ball and stick representation is also used to show groups that are of particular interest to the work conducted in this thesis. These include Leu29, Lys45, Lys63, His64, Val67, Leu104 and the heme ligands His93 and Wat156. The top panel provides the labelled positions of these groups, while their spatial displacement is more accurately displayed in the stereo image below.
Chapter 1: Introduction

Table 1.1. Secondary structure elements in horse heart myoglobin.

<table>
<thead>
<tr>
<th>Residues</th>
<th>Secondary Structure</th>
<th>Helix Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 - 18</td>
<td>α-helix</td>
<td>A</td>
</tr>
<tr>
<td>20 - 36</td>
<td>α-helix</td>
<td>B</td>
</tr>
<tr>
<td>36 - 43</td>
<td>$3_{10}$-helix</td>
<td>C</td>
</tr>
<tr>
<td>43 - 46</td>
<td>$3_{10}$-type turn</td>
<td></td>
</tr>
<tr>
<td>46 - 49</td>
<td>$3_{10}$-type turn</td>
<td></td>
</tr>
<tr>
<td>51 - 58</td>
<td>α-helix</td>
<td>D</td>
</tr>
<tr>
<td>58 - 78</td>
<td>α-helix</td>
<td>E</td>
</tr>
<tr>
<td>78 - 81</td>
<td>Type I turn</td>
<td></td>
</tr>
<tr>
<td>82 - 97</td>
<td>α-helix</td>
<td>F</td>
</tr>
<tr>
<td>100 - 119</td>
<td>α-helix</td>
<td>G</td>
</tr>
<tr>
<td>119 - 122</td>
<td>Type II turn</td>
<td></td>
</tr>
<tr>
<td>124 - 150</td>
<td>α-helix</td>
<td>H</td>
</tr>
</tbody>
</table>

directed mutagenesis methods which have allowed the roles of a number of residues in the heme
binding pocket to be studied by replacing them with other amino acids. For example, in wild-type
metmyoglobin, the presence of the distal heme pocket residue His64 results in formation of a
hydrogen bond between this residue and the water molecule coordinated to the heme iron atom.
The consequence of replacing His64 with other amino acids has been found to be a significant
alteration in the axial coordination of the ferric heme iron (see review: Springer et al., 1994). Thus
substitution of His64 with tyrosine results in His-Fe-Tyr ligation as demonstrated by spectroscopic
results (Egeberg et al., 1990a; Hargrove et al., 1994; Tang et al., 1994). Furthermore, resonance
Raman studies have shown that an arginine at position 64 leads to a six-coordinate low-spin heme
iron, and it has been suggested that the side chain of Arg64 might coordinate to the ferric heme iron
atom (Morikis et al., 1990).
Figure 1.5. A space filling representation of horse heart myoglobin. Main and side chain atoms are shown with light and medium shaded spheres, respectively. Heme atoms are drawn more heavily shaded. The majority of heme atoms are buried in a hydrophobic pocket. The direction of view presented is similar to that of Figure 1.4.
Chapter 1: Introduction

On the proximal side of the heme, a key structural element is the heme liganding residue His93 and without this interaction myoglobin is unable to bind dioxygen reversibly (Springer et al., 1994). Furthermore, with removal of His93 oxygen binding results immediately in auto-oxidation of the heme iron atom to the Fe(III) state in the presence of oxygen. An important conserved element believed to be involved in maintaining the orientation of the side chain of His93 when it is liganded to the heme iron atom, is a hydrogen bond formed to the side chain of Ser92 (Shiro et al., 1994). Another conserved proximal heme pocket residue is His97 which is located adjacent to the proximal heme ligand. This residue appears to act as a barrier to solvent entry to the proximal side of the heme and may be important in preventing heme dissociation (Hargrove et al., 1996).

A number of other conserved residues are found in the heme pocket. For example, Leu 29 is located near His64 (Figure 1.4). Amino acid substitutions at Leu29 suggest that steric hinderance at this position is not an important factor in influencing ligand binding affinities (Carver, 1992). The same was found for the distal heme pocket residue Val68 which is located on the exterior of the protein (Egeberg, 1990b). Other conserved residues in the heme pocket include Phe43, Val67 and Leu104. These later residues are all in van der Waals contact with the heme prosthetic group, provide stabilizing heme-protein hydrophobic interactions, and may have a role in ligand binding properties (Hargrove et al, 1994a).

1.6 Research Objectives
The goals of the work in this thesis were to gain further insight into structure-function relationships in myoglobin, alter the structure of this protein to enhance a low level latent peroxidase activity, and to add a new functionality in the form of a manganese binding site to promote a potential manganese peroxidase activity. To address all three of these objectives a series of carefully designed myoglobin variants were constructed and analyzed. In addition to three dimensional
Chapter 1: Introduction

structural analyses of these variants using x-ray diffraction techniques, which is a major part of the work described in this thesis, site directed mutagenesis approaches, as well as functional and spectroscopic analyses, were conducted. General aspects of experimental procedures utilized, including x-ray structural analyses, spectroscopic measurements and activity assays, are discussed in Chapter 2, while subsequent chapters describe the details pertaining to specific myoglobin variants. Additional discussion of the use of molecular genetics methods to produce the Leu29Lys variant Chapter 6

Initial research required a thorough structural characterization of recombinant wild-type horse heart myoglobin which serves as the starting point and reference for all subsequent work. In Chapter 3 a description is given of the determination of the high resolution structure of this protein and a comparison with its naturally occurring counterpart. Chapters 4 and 6 involve an investigation of the latent peroxidase function of myoglobin based on the knowledge of the structure of cytochrome c peroxidase (Finzel et al., 1984). The amino acids studied as part of this work included Leu29, His64, Val67 and Leu104.

Chapter 5 discusses a part of the overall project directed at understanding the structural and functional consequences of ligand binding to myoglobin with the focus being on the complexation of linear ligands such as azide. The data obtained was used to assess the ability of this protein to accommodate ligands of various sizes.

In Chapter 7 a discussion of successful efforts to introduce a manganese binding site near the heme pocket of myoglobin is presented. Some of the requirements considered in building this binding site included optimizing ligand interactions, having a close proximity to the heme active center, making changes that serve to assist in the enhancement of a peroxidase function, and minimizing the number of amino acid substitutions needed so as to preserve the overall folding of myoglobin. This chapter further documents the characterization of some of the functional
Chapter 1: Introduction

properties of this manganese binding site. Finally, Chapter 8 provides a concise summary of the experiments described in this thesis and the results obtained.

1.7 Collaborations and Publications

Beyond the x-ray diffraction studies that form a major part of the work described in this thesis, aspects of site directed mutagenesis approaches, as well as functional and spectroscopic analyses were conducted as part of a collaborative effort with the laboratories of Dr. A. G. Mauk (R. Bogumil, C. Hunter, D. Hildebrand and E. Lloyd) and Dr. M. Smith (C. Overall and H. Tang). I am particularly grateful to these colleagues for providing various proteins and also for teaching me the relevant site directed mutagenesis techniques to allow me to produce and isolate the Leu29Lys variant protein, and the use of spectroscopic and assay methods to analyze this and other variants.

Several portions of the work presented in this thesis have already been published in refereed journals as outlined below.


   This work is discussed in Chapters 3 and 6.


   This work is presented in Chapter 4.


   This work is presented in Chapter 4.

Chapter 1: Introduction

This work is presented in Chapter 5.

Also as part of collaborative studies not discussed in this thesis, I have been an author on the following publications related to horse heart myoglobin.


Chapter 2

General Overview of Experimental Methods

This chapter serves to give an overview of the experimental techniques used to study both wild-type and variant recombinant horse heart myoglobins. A general discussion of experimental approach is first presented, and then followed by a description of the techniques of mutagenesis, protein crystallization, diffraction data collection, structural refinement, refinement model quality assessment and peroxidase activity measurements. The precise details of these methods as used for different variant and wild-type myoglobins are presented in following chapters along with the descriptions of the structures involved.

2.1. General Experimental Approach

As described in Chapter 1, the work in this thesis has three major objectives. These include gaining a better understanding of structure-function relationships in myoglobin and to design variant proteins that have enhanced peroxidase activities. The final goal was to design a stable and efficient manganese peroxidase. The general strategy used in this protein engineering project is shown in Figure 2.1.

Overall, the design of myoglobin variant proteins was based on three main sources of information. The first was the three dimensional structures of related myoglobins and other heme proteins, especially that of cytochrome c peroxidase (Finzel et al., 1984) which was the only available peroxidase structure at the time these studies were initiated. A second source of information involved previous studies of the factors affecting heme chemistry. A third source of information was the available structural data on metal binding sites in other proteins. Please
Figure 2.1: A schematic representation of the general scheme of experiments conducted to characterize the structural and functional properties of variant and wild-type horse heart myoglobins, as well as promote enhanced peroxidase activity and introduce a manganese binding site into this protein. Shaded boxes indicate areas of research carried out as part of this thesis.
Chapter 2. General Overview of Experimental Methods

note that the specific rationale for the design of each individual variant protein is left for discussion in subsequent chapters.

An essential element in the design of variant proteins was the use of molecular graphics modelling based on the high resolution structure of wild-type horse heart myoglobin (Evans & Brayer, 1990). The majority of such modelling work employed the program MMS (Dempsey, 1986) running on a Silicon Graphics 4D/340 workstation. In these studies the goal was to choose amino acid substitutions that would address a particular objective while at the same time were not likely to induce substantial polypeptide refolding. Such refolding events could complicate interpretation of functional data once designed variant proteins were expressed, purified and their three dimensional structures determined using x-ray diffraction techniques. In addition to structural analyses, each variant protein was subjected to a series of spectroscopic and functional studies as outlined in Figure 2.1. A major effort was then directed towards using the collected functional and structural data to assess the value of replacement amino acids with respect to modifying functionality. Insight gained from initial rounds of these investigations were then used in the design of subsequent variant myoglobins in order to further enhance activities.

2.2 Protein Preparation

A synthetic gene coding for horse heart myoglobin (Guillemette et al., 1991) was used in the preparation of myoglobin variants. Mutagenic oligonucleotides were synthesized with a modified Applied Biosystems 380A DNA Synthesizer at the UBC Nucleic Acid and Protein Synthesis laboratory, and purified using a Millipore C18 Sep Pak column. The procedure used for oligonucleotide-directed mutagenesis has been described by Zoller & Smith (1983; 1984). The method of Kunkel (1985) was used to select the variant protein genes. The required uracil containing DNA was produced by a RZ1032 strain of E. coli that is deficient in
Chapter 2. General Overview of Experimental Methods

uracil-N-glycosylase and dUTPase. Necessary transformations were carried out using the CaCl$_2$ method (Maniatis et al., 1989). Single strand DNA sequencing was used to select for colonies containing the desired mutation and to ensure that no other mutations were introduced by the mutagenesis procedure (Maniatis et al., 1989).

Expression of wild-type and variant protein synthetic genes was done in *E. coli* (strain LE392) using 10 or 20 L fermenters with superbroth medium (tryptone 10 g/L, yeast extract 8 g/L, NaCl 5 g/l and ampicillin 100 μg/L). Cells were harvested using a tangential flow centrifuge at the UBC Biotechnology Fermentation facility. With the exception of position 29 variant proteins, purifications were carried out as described by Lloyd & Mauk (1994). Specific procedures used for the position 29 variant proteins are described in detail in Chapter 6. In brief, bacteria were washed and suspended in 10 volumes of 100 mM NaCl, 1 mM EDTA, 50 mM Tris, pH 8.0 before cell disruption with a CD 30 APV Gualin homogenizer. The cell homogenate was clarified by centrifugation at 8,000 x g for 20 min at 4 °C and the 65-100% saturation ammonium sulfate precipitate was dissolved in 5 mM Tris-HCl (pH 8.4), 1 mM EDTA, and exhaustively dialysed against the same buffer. After centrifugation, the supernatant was loaded onto a DEAE-Sepharose column (4.5 x 22 cm) in equilibration buffer. The bound myoglobin was eluted with 50 mM Tris/HCl (pH 8.4) and then concentrated by ultrafiltration (Amicon YM-10 membrane). This concentrate was chromatographed on a G75 Sepharose column (2.5 x 92 cm) in equilibration buffer. The yield of recombinant myoglobin was ~1.5 mg/g cells (wet weight).

Molar absorption extinction coefficients of variant myoglobins were determined using the pyridine-hemochromagen method (De Duve, 1948) which allows for the determination of variant holomyoglobin protein concentrations. Except for position 29 recombinant proteins, all other variant proteins isolated from *E. coli* were reconstituted with fresh heme to avoid the formation of sulfmyoglobin which could potentially occur during protein expression (Lloyd & Mauk, 1994).
Chapter 2. General Overview of Experimental Methods

For this, apomyoglobin was prepared from purified protein by extracting the heme using the acid-butanolone method (Teale, 1959). Heme reconstitution was accomplished with 1.2 equivalents of protoporphyrin IX (10 mg/ml in 0.2 M NaOH) obtained from Porphyrin Products (Logan, Utah). Excess heme was removed from holomyoglobin using a Sephadex G-75 Superfine column (2.5 x 75 cm). Note that position 29 variant myoglobins could not be reconstituted with heme and were therefore used as purified from the *E. coli* expression system.

2.3 The Growth of Protein Crystals

An essential requirement for structure determination using x-ray diffraction methods is the growth of suitable crystals of the protein of interest. Such crystals are precisely ordered, three dimensional arrays containing a large number of individual protein molecules. Each protein molecule is held in the crystalline lattice by non-covalent intermolecular interactions formed with neighbouring molecules. The fundamental building block of this repeating and periodic array is referred to as the asymmetric unit. Symmetry operations often relate such asymmetric units to one another. A unit cell can be defined which encloses a full set of asymmetric units which are related by a unique set of symmetric elements. The unit cell is the smallest and simplest volume element that is representative of the whole crystal. A discussion and compilation of possible protein crystal unit cells can be found in Blundell & Johnson (1976) and Drenth (1994).

Obtaining suitable single crystals is mainly a trial and error procedure to find the best conditions under which to manipulate a supersaturated solution of the protein being studied. Protein that comes out of solution under these conditions can either form a precipitate or the desired result, single crystals. In this regard, a variety of techniques have been developed to bring a protein solution to supersaturation in a manner that favours crystal formation (eg.: McPherson, 1982; 1990). These techniques share a common theme in that protein is dissolved in a suitable solution
Chapter 2. General Overview of Experimental Methods

from which it will be induced to form crystals. This solution is chosen such that the conditions used will not lead to protein precipitation. This protein solution is then brought to a state of supersaturation through some mode of diffusion, usually with a solution containing a higher concentration of crystallizing agent. To achieve optimal crystal growth, protein supersaturation must be approached slowly to avoid formation of too many nuclei.

All of the myoglobin crystals examined in this thesis were produced using the hanging drop vapour diffusion technique (Figure 2.2). In this method, a drop of protein solution (5-10 µl) is prepared on a siliconized (Sigmacote; Sigma Chemical Co., St. Louis, Missouri) microscope glass cover slip. This is mixed with a portion of crystallizing solution (generally 1:1 ratio by volume) and the cover slip is then put upside down over a well containing only the crystallizing solution (typically 1 mL). This well is sealed by the previous application of high vacuum silicon grease (Dow Corning Co., Midland Michigan) to its rim before the cover slip is put into place. Over time through vapour diffusion the protein solution will gradually attain the same concentration of crystallizing agent as that of the solution at the bottom of the well. In this way the protein solution in the hanging droplet will at some point become supersaturated and hopefully this will induce crystal growth.

A typical hanging drop crystallization experiment with horse heart myoglobin was carried out using a twenty four well Linbro plate (Flow Laboratories, McLean VA). Such plates allow one to conveniently scan different crystallization variables although it is generally possible to screen only two of these on a given trial plate. For horse heart myoglobin and its variants the parameters found most important to inducing crystallization included pH, buffer composition, concentration of crystallizing agent in both the hanging drop and crystallization well solutions, concentration of protein in the hanging drop, and the effect of additives. Initial scans to determine crystallization conditions were based on the known conditions that produced crystals of diffraction quality for
natural horse heart myoglobin (Sherwood et al., 1987).

Early in these studies it was observed that crystal formation tended to occur rapidly leading to needle like crystals that were poor candidates for high resolution diffraction analyses. To overcome this deficiency two strategies were employed. First, the difference in concentration of crystallizing agent between the hanging drop and the crystallization well solutions was minimized to keep the degree of supersaturation low and the approach to this state slow. Second, to supply a limited number of nucleation sites a hair seeding technique was used in which micro crystals (obtained by crushing already existing myoglobin crystals) were introduced into the protein drop (Leung et al., 1989). Together these two strategies were able to produce adequate numbers of wild-type and variant protein crystals of suitable morphology for x-ray diffraction structure studies.

It is important to emphasize that the growth of diffraction quality crystals was neither straightforward nor guaranteed, despite knowledge of the general crystal growth conditions. Crystals in some cases appeared in 2-3 days and grew to full size in 1-2 weeks. However, the highest quality crystals were those that appeared over the course of a month and reached full size in 2-3 months.
Chapter 2. General Overview of Experimental Methods

Crystallization of each variant protein also proved to be unique and obtaining crystals with a suitable morphology for x-ray diffraction studies required a large number of scans of conditions. Considerable variability in crystal quality and multiple space groups were observed for the various variant proteins. The precise conditions used to crystallize each variant myoglobin are discussed in following chapters.

2.3.1 The collection and processing of x-ray diffraction data

All x-ray diffraction data used in this thesis was collected with a Rigaku R-Axis II area detector. This instrument uses an approach to collect diffraction data that is analogous to the screenless oscillation method described by Arndt & Wonacott (1977), but instead of using photographic film to record and measure diffraction intensities, two radiation sensitive phosphor plates are used. During data collection each crystal was mounted on a goniometer head in a thin walled glass capillary (1.0 mm diameter), with short columns of mounting buffer placed on both sides to maintain crystal hydration during data collection. This glass capillary was then sealed with dental wax.

The x-ray radiation used (CuKα; λ=1.54178 Å) was generated from a rotating anode fitted with a monochromator and operated at 90 mA and 59 kV. Exposure time for each image plate varied from 15 to 45 minutes depending on the size and quality of the crystal available. To maximize the amount of diffraction data obtained, the glass capillary was first mounted vertically and then offset 45°. Each of these diffraction data sets were collected with an oscillation angle in the range of 1.2° and over a φ axis rotation range of 90°. For all diffraction data collections the crystal to detector distance was set between 54 to 78 mm depending upon the resolution of the diffraction pattern.

All diffraction data sets were processed using the Rigaku R-AXIS software (Higashi, 1990;
Chapter 2. General Overview of Experimental Methods

Sato et al., 1992) which is based on a technique described by Rossmann (1979). In this procedure diffraction intensities are integrated via a profile fitting routine. The level of background radiation was determined by measuring the observed intensity directly surrounding each measured and integrated peak intensity. Partial diffraction spots that were spread over two data collection frames, but which belonged to one diffraction intensity, were added together after the peak integration step was completed. Corrections for background, Lorentz and polarization effects were initially applied to the individual diffraction intensities measured on each frame of data. Following these corrections, the data collection frames were scaled to account for crystal decay and absorption effects in a resolution dependent manner. After scaling, multiple measurements of intensities were merged and all intensities were then reduced to structure factor amplitudes.

An estimate of the absolute scale for each diffraction data set was obtained using the method described by Wilson (1942). This is a statistical approach that compares the intensities of observed structure factors with those predicted for a crystal composed of a comparable number of atoms in a random distribution. A representative Wilson plot is shown in Figure 2.3 using diffraction data obtained for recombinant wild-type horse heart myoglobin. The y-intercept of a least squares fit line to the points in Figure 2.3 represents an initial estimate of the overall scale of this diffraction data. The slope of the fit line in Figure 2.3 serves as a useful indicator of the overall isotropic thermal factor for all atoms in the crystallized protein. In Figure 2.3 the poor fit of the plotted line at low resolution likely arises from the fact that the contents of the unit cell are not randomly distributed. At high resolution a similar deviation is likely due to the weakness of the diffraction intensities. Due to these types of deviations, resolution cutoffs are routinely applied before a line is fit to data such as that in Figure 2.3.
Figure 2.3. An example of a Wilson plot which is based on the diffraction data set collected for recombinant wild-type horse heart myoglobin (see Chapter 3 for further details). For this analysis reciprocal space was divided into resolution bins so that each of these contained 100 reflections. The absolute scale was determined by a least squares fit (solid line) to those data points falling between 1.7 and 4 Å resolution as indicated by the vertical dashed lines. This analysis suggests that the scale factor for this data set is 1.4 based on the y-intercept.

2.4 General Aspects of the X-ray Diffraction Experiment

In this section, a general discussion of the fundamentals of x-ray diffraction theory is presented as it has been applied in this thesis. The intend is to provide a reader unfamiliar with these methods the background necessary to follow the crystallographic aspects of the experiments described. A
more advanced discussion of protein structure determination by x-ray crystallography can be found in a number of excellent texts (for example: Blundell & Johnson (1976); Stout & Jenson (1989); McRee (1993); and, Drenth (1994)).

2.4.1 Dependence of x-ray diffraction on structure

X-rays are a form of electromagnetic radiation that is able to propagate through matter. Their relatively short wavelength ( ~ 0.1 - 1000 Å) makes the observed interference or diffraction of x-rays sensitive to the atomic structure within molecules. However, images of molecules cannot be formed directly with x-rays. In contrast to a conventional microscope where visible light scattered from an object can be recombined by a lens to produce a magnified image, there are no lens materials that can bend and focus x-rays. For this reason x-ray images must be constructed computationally from the experimentally recorded intensities found in the diffraction pattern of the molecule of interest.

X-rays that interact with atoms are scattered in a discrete manner. This scattering depends on both the kind of atom involved and its position in space. For a continuous electron distribution, the structure factor F(S) can be defined as an integral in which the total scattered wave F in the direction of the reciprocal space vector S, is the summation of individually scattered waves by the continuous electron density ρ, at position r, integrated over the volume V. This can be expressed as follows:

\[ F(S) = \int V \rho(r) e^{i2\pi r \cdot S} dV \]  

(2-1)

Periodic arrays of atoms such as found in a crystal, restrict the observation of scattered diffraction intensities to a discrete set of periodic reflections which are the result of the scattered waves meeting the conditions for constructive interference. The periodicity of the crystalline lattice
allows for the replacement of the scattering vector $S$ in Equation 2-1 with a discrete direction specified by the Miller indices of a reflection. Note that under these conditions the diffraction of x-rays can be regarded as occurring from planes in the crystal lattice. These planes can be drawn through lattice points and are designated by a set of three integers, defined as the Miller indices $h$, $k$ and $l$. Each set of parallel and equidistant planes is considered an independent diffractor and produces a single reflection. This, together with the assumption that electron density is localized at atomic centres specified by real space coordinates $(x, y, z)$, gives an equation that relates the electron density present in the unit cell of a crystal, $\rho(x,y,z)$, to structure factors, $F_{hkl}$, which are composed of both phase and amplitude information.

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx + ky + lz)}$$ (2-2)

### 2.4.2 The phase problem

The electron density Equation (2-2) is extremely useful, but in itself is not sufficient to determine an atomic structure. The complete structure factor, $F_{hkl}$, consists of an amplitude, $|F_{hkl}|$, which is proportional to the square root of the measurable diffraction intensity, $I_{hkl}$, as well as a non-measurable phase term, $\alpha_{hkl}$. This can be expressed as follows:

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{i\alpha_{hkl}} e^{-2\pi i(hx + ky + lz)}$$ (2-3)

The inability to measure the phase angles of individual diffraction intensities poses a serious problem for macromolecular crystallography. To overcome this, a number of methods have been derived to estimate the necessary phase information so as to allow for the calculation of electron
density maps.

2.4.3 Phase determinations for different variant myoglobin crystals

Phase information for crystallized proteins studied in this thesis was determined by molecular replacement methods. The complexity of the methods used in each case depended on whether the unit cell dimensions and symmetry of the variant protein crystal were similar (isomorphous) to those of the wild-type protein or significantly different (non-isomorphous). Most of the variant protein crystals studied were of the isomorphous type and this significantly facilitated the process of obtaining phase estimates for observed structure factor amplitudes. For the case of isomorphous crystals it was assumed that the wild-type and variant protein structures were closely related and therefore that the phases of the variant protein structure factor amplitudes were similar to those of the wild-type structure. Therefore, for an isomorphous crystal form, the starting model of the variant protein consisted of the wild-type myoglobin structure placed in the unit cell of the variant protein crystal. Following this placement structure factor amplitudes and phases were calculated from the positional and thermal parameters of the atoms within the unit cell in the following manner.

\[
F(h,k,l) = C \sum_j f_j(s_{hkl}) e^{-B_j(s_{hkl})^2} e^{2\pi i(hx_j + ky_j + lz_j)}
\]  

(2-4)

In this expression, each of the atoms (j) in the model protein structure is represented by its atomic position \((x_j, y_j, z_j)\), an atomic scattering factor \(f_j\), and a thermal parameter \(B_j\), which is related to the mean square displacement of the atom about its average position. The constant C, represents a scale factor, and \(s_{hkl}\) is equal to the quantity \(\sin \theta / \lambda\), for the reflection hkl. Phases determined in this manner can be used as estimates for the experimentally measured structure factor amplitudes.
from the variant protein crystal.

In contrast, if a variant protein crystal form was found to be novel and therefore non-isomorphous with respect to wild-type myoglobin crystals, a more involved version of the molecular replacement method (Rossmann, 1972) was required to obtain phase information and solve this structure. A general discussion of this approach is presented in this section, while details of the specific analyses of this type carried out for various protein crystals, are presented in Chapters 6 and 7.

For such cases, where there is a reasonably good preliminary model for a variant protein (i.e. wild-type myoglobin), this can be used in the process of determining phase information. This approach utilizes the Patterson function which can be expressed as follows:

$$ P(u,v,w) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}|^2 e^{-2\pi i (hu + kv + lw)} $$

(2-5)

Patterson map peaks correspond to all the interatomic vectors within the unit cell. Therefore there will be a peak in the Patterson map at position $u$, $v$, $w$, for each pair of atoms in the unit cell at positions $x_1$, $y_1$, $z_1$ and $x_2$, $y_2$, $z_2$, such that $u=x_1-x_2$, $v=y_1-y_2$ and $w=z_1-z_2$. Note that calculation of a Patterson map only requires the availability of experimentally measured structure factor amplitudes, and that there is no requirement for phase information.

A Patterson map can also be made starting with a model structure. By this approach, a comparison of the Patterson map obtained using the observed variant protein crystal structure factor amplitudes can be made compared with that of the preliminary search model (wild-type myoglobin) to obtain the positioning of the variant protein structure in its unique unit cell. This is carried out in two separate stages. First, a search for the best rotational orientation is performed and involves systematically rotating the search model Patterson map about three mutually
orthogonal axes and comparing it to the Patterson map calculated from the observed structure factor amplitudes from the variant protein crystal. During this procedure, each orientation is evaluated by calculating a correlation factor between the two Patterson maps with this factor being maximal when the orientation of the search model coincides with that of the variant protein molecule in the new unit cell.

A rotation search can also be conducted to determine whether more than one copy of the variant protein molecule is present in the asymmetric unit. This search is identical to that just described except that the Patterson map calculated from the variant protein crystal structure factor amplitudes is used as both the search and target maps. This technique is known as the self-rotation function search.

The second part of the orientational search for the best position of the variant protein involves determining the optimal translation of the search model with respect to the unique variant protein unit cell axes. This is again accomplished by a systematic search, with the assessment of the fit obtained based on the calculation of a correlation coefficient. Once the best rotational and translational parameters for the search model have been determined, phase estimates for the experimentally observed structure factor amplitudes of the variant protein crystal can be calculated (Equation 2-4). These phase estimates can then be used to calculate electron density maps for use in optimizing the original search model to more closely reflect the variant protein structure and allow for the start of a more detailed structural refinement as described in following sections.

2.5 Refinement of Initial Structural Models
The initial structural model obtained upon solving the phase problem can be improved through an iterative process in which a series of alternating cycles are carried out involving computational least squares fitting of atomic coordinates to the observed diffraction data and manual adjustments to the
polypeptide chain. The progress of structural refinement can be monitored by calculating the
crystallographic R-factor, a measure of the agreement between the experimentally observed and
calculated structure factor amplitudes. This can be expressed as follows:

$$R - \text{factor} = \frac{\sum \|F_{o,hkl} - |F_{c,hkl}|\|}{\sum |F_{o,hkl}|} \quad (2-6)$$

In this equation $|F_{o,hkl}|$ is the observed structure factor amplitude of a reflection and $|F_{c,hkl}|$ is the
structure factor amplitude calculated from the model refinement structure. A low R-factor implies
that there is good agreement between the experimental data and the model structure. In general, for
well defined protein structures the crystallographic R-factor falls in the range of 0.10 to 0.20.

In this thesis a restrained parameter least squares technique (Hendrickson, 1985) was used in
structural refinement. This method seeks to minimize the differences between the observed
structure factor amplitudes and those calculated from the model structure. In this process the
stereochemical parameters of the protein model are maintained within an acceptable range of values
by applying geometric restraints to bond distances and angles, main and side chain planar groups,
chiral centers, torsion angles, hydrogen bonds, nonbonded contacts and thermal parameters.

A major limitation of the restrained parameter least squares approach to structural refinement is
its relatively small radius of convergence. This can be overcome through manual interventions that
use molecular graphics assisted inspections of electron density maps to allow for the best fitting of
portions of structure that are unable to refine due to conformational barriers. Manual fitting of
structure is generally achieved by reference to a difference electron density map. This can be
calculated through an expansion of the basic form of Equation 2-3 to include both observed and
calculated structure factors:
Chapter 2. General Overview of Experimental Methods

\[ \Delta \rho(x,y,z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} (|F_{o,hkl}| - |F_{c,hkl}|) e^{i\alpha_{c,hkl}} e^{-2\pi i(hx + ky + lz)} \]  

(2-7)

In this equation \( |F_{o,hkl}| \) represents structure factor amplitudes obtained from the x-ray diffraction experiment, and \( |F_{c,hkl}| \) and \( \alpha_{c,hkl} \) are the structure factor amplitudes and phases calculated from the model structure. In this type of difference electron density map, positive peaks would represent electron density that is unaccounted for in the current refinement model, whereas negative electron density peaks indicate electron density in the refinement model that is not supported by the experimental data.

Another commonly used type of difference electron density map which provides a representation of the quality of the fit of the electron density derived from experimental data to the model refinement structure, is a \( 2F_o - F_c \) map calculated with Equation 2-7, but where \( 2F_{o,hkl} \) is substituted in place of \( F_{o,hkl} \). This form of a difference electron density map superimposes differences in electron density on top of the electron density representing the current refinement model. To confirm particular structural features, additional omit difference electron density maps proved very useful. This type of map is calculated by using Equation 2-7 and replacing \( F_{c,hkl} \) and \( \alpha_{c,hkl} \), with \( F_{\text{omit},hkl} \) and \( \alpha_{\text{omit},hkl} \) which are the structure factor amplitudes and phases calculated from the model refinement structure where the structural features of interest have been omitted. Atoms omitted in this way appear as strong peaks in the resulting difference electron density map and this facilitates the fitting of the electron density within these regions of the refinement model.

### 2.5.1 An example of a typical structural refinement

In this thesis all structures were refined by restrained parameter least-squares methods (Hendrickson, 1985). In this section the general approach to such refinements is discussed, while the specifics as they pertain to individual structural refinements are described in following chapters.
Chapter 2. General Overview of Experimental Methods

To illustrate a typical refinement, the determination of the Leu29Lys variant myoglobin structure is used as an example. In this case the starting refinement model was composed of the coordinate set for recombinant wild-type myoglobin (this structure is described in Chapter 3) with residue 29 truncated to an alanine.

To begin refinement 10 cycles of restrained least-squares were performed using lower resolution data in the 6.0 - 2.0 Å resolution range. This was followed by fitting the side chain of Lys29 in an omit difference electron density map (Figure 2.4) and including these additional atoms in the refinement model. A further 15 cycles of refinement were then carried out. Additional cycles of refinement followed the inclusion of higher resolution data (6.0 - 1.6 Å), the adjustment of scale factor, and rounds of manual manipulations of the model that included establishment of the solvent structure (see Figure 2.5). In general, manual interventions consisted of the sequential examination of the entire polypeptide chain using omit and 2F_0 - F_c difference electron density maps. The criteria used for inclusion of water molecules in the refinement model were the presence of a hydrogen bond (2.5 - 3.6 Å in length) to a donor or acceptor protein atom and the refinement of an isotropic thermal factor to less than 60 Å^2. Potential water molecules were generally identified from peak searches of F_0 - F_c difference electron density maps. Refinement was continued until positional shifts became small (< ~ 0.03 Å per cycle), an indication that convergence had been reached.

2.5.2 The quality of structural models

During refinement the stereochemistry of each myoglobin structural model was periodically checked using the program PROCHECK (Laskowski et al., 1993). Of particular interest in these analyses were Ramachandran plots (for example see Figure 2.6; (Ramachandran et al., 1963)), peptide bond planarity, alpha carbon chirality, side chain \( \chi_1 \) angles, as well as contact distances
Figure 2.4. An omit difference electron density map in the vicinity of Lys29 in the Leu29Lys variant of horse heart myoglobin. This map was calculated with the omission of all the atoms of Lys29 and contours are shown at the 2.5 σ level. Superimposed on this map is the final refined structure of the Leu29Lys variant protein (see Chapter 6 for details).

between non-bonded atoms. In the case of significant deviations from ideality the residues involved were closely examined in electron density maps and adjusted accordingly.

Error estimates for the atomic coordinates of refined myoglobin structures were determined by two methods. One method has been described by Cruickshank (1949; 1954; 1985) and provides an estimate of the coordinate error of each atom based on the refined thermal factor of that atom, the type of atom, and the overall fit between observed and calculated structure factor amplitudes. In a second method (Luzzati, 1952) an overall coordinate error estimate is based on a plot of the dependence of the crystallographic R-factor on resolution. An example of such a Luzzati plot is illustrated in Figure 2.7 and shows the results obtained for the recombinant
Figure 2.5. A plot of crystallographic R-factor over the course of restrained parameter least squares structural refinement of Leu29Lys horse heart myoglobin (see Chapter 6 for details). Specific refinement cycles are indicated during which manual interventions and water searches were conducted using $F_o - F_c$, $2F_o - F_c$ and omit difference electron density maps covering the course of the entire polypeptide chain. The full side chain of Lys29 was added to the refinement model at cycle 11 based on an examination of difference electron density maps. Refinement was concluded when the overall r.m.s. positional shifts for all atoms had become < 0.03 Å.
Chapter 2. General Overview of Experimental Methods

Figure 2.6. An example of a typical Ramachandran plot obtained for structures completed as part of this thesis. The data shown is from the final refined structure of Leu104Asn horse heart myoglobin (see Chapter 6). The degree of shading shows the most favoured (dark grey), allowed (medium grey), generously allowed (light grey) and disallowed (white) regions of Psi/Phi space based on an analysis of 118 high resolution protein structures. Glycine residues are indicated as triangles, while all other residues are represented as squares. The program PROCHECK [Laskowski, 1993] was used to generate the plot drawn.
Chapter 2. General Overview of Experimental Methods

Figure 2.7. Luzzati (1952) plot of crystallographic R-factor as a function of resolution for the recombinant wild-type horse heart myoglobin structure determination (see Chapter 3). The theoretical dependence of R-factor on resolution for various levels of coordinate error is shown with broken lines (error levels printed to the right). Examination of this plot suggests an r.m.s. coordinate error of ~0.18 Å.

wild-type myoglobin structure determination (see Chapter 3). Generally the error value estimates obtained by these two methods were found to be comparable. However, both these methods assume that a complete diffraction data set is being used in the structure determination and that the only errors present are in the form of coordinate errors. Neither of these assumptions is completely valid and care must therefore be taken when using such error estimates as a measure of the individual atomic coordinate error of a particular atom.

2.6 Spectroscopic Characterizations

Spectroscopic methods can also be very useful in the study of the functional effects of amino acid
Chapter 2. General Overview of Experimental Methods

substitutions. One advantage of these techniques is the ability to examine a protein under a wider range of conditions and temperatures than possible for crystallographic methods. In particular, these techniques can provide additional information about the electronic structure of the heme iron atom of myoglobin. Spectroscopic methods employed in this work include electronic absorption spectroscopy (Smith & Williams, 1970) and electron paramagnetic resonance spectroscopy (Palmer, 1985). All spectroscopic results were obtained in close collaboration with Dr. Ralf Bogumil using instrumentation in the laboratory of Prof. Grant Mauk (U.B.C.).

The basic properties of heme iron centers are as follows. The electronic configuration of the neutral iron atom is \([\text{Ar}]\)\(3d^6\)\(4s^2\). The two common oxidation states of iron are the ferrous Fe(II) and ferric Fe(III) states which have the electronic configurations \([\text{Ar}]\)\(3d^6\)\(4s^0\) and \([\text{Ar}]\)\(3d^5\)\(4s^0\), respectively. The electron distribution within the five 3d-orbitals of an iron atom is determined by bonding interactions of the iron atom with coordinating ligands and the geometry of these interactions. As discussed in Chapter 1, for a heme bound iron atom, four of the coordination bonds are filled by the heme pyrrole nitrogen atoms. Addition of one or two axial ligands provides a system with tetragonal symmetry. In myoglobin one of these axial ligands is provided by the NE2 nitrogen atom of the His93 side chain.

Electron absorption spectroscopy provides information about transitions between energy levels of different electronic states. The intensity and shape of visible absorption bands give insight into both the oxidation and spin states of the heme iron centre. A typical spectrum of the heme chromophore consists of an intense band at \(~400-420\) nm (also called the Soret band) and two absorption bands of lower intensity with maxima between \(500-640\) nm (Makinen & Churg, 1983 and references therein). These bands are characteristic of \(\pi-\pi^*\) transitions of the porphyrin aromatic system (Longuet-Higgins et al., 1950). Although the electronic spectra are primarily attributable to the porphyrin aromatic system, they are also sensitive to the electronic configuration.
Chapter 2. General Overview of Experimental Methods

of the heme iron centre, and this feature can be used to infer information regarding coordination environment.

Electron paramagnetic resonance spectroscopy (EPR) allows detection and measurement of transitions involving unpaired electrons. In terms of an iron atom this is limited to the ferric form. In this method a strong externally applied magnetic field induces normally degenerate energy levels of unpaired electrons to become non-degenerate. The resulting resonance condition is described by the following equation:

\[ \Delta E = h \nu = g \beta H \]  

(2-8)

In this expression \( \Delta E \) is the difference in energy between the two spin states, \( h \) is Planck's constant, \( \nu \) is the operating frequency of the spectrometer, \( g \) is the dimensionless g-factor measured in an EPR experiment, \( \beta \) is the Bohr magneton, and \( H \) is the resonance magnetic field. The g-factor relates resonance frequency and external magnetic field. The orientational dependence of this relationship governs the shape of the EPR spectrum and together with the g-factor provides insight into the local environment of the unpaired electron.

For proteins with heme iron centres, paramagnetic Fe(III) derivatives can be studied with EPR spectroscopy (Brudvig, 1995). Depending on the strength of the ligand field, a variety of spin states can be observed for the ferric heme iron atom (Scheidt & Reed, 1981). Generally at room temperature, ferric heme proteins exist either in high spin or low spin states, or are in a thermal equilibrium of these states. Axial ligands of moderate field strength can produce a heme iron system that is close to a high-spin/low-spin crossover. A system of this type is the hydroxide bound form of myoglobin. Electronic configurations of a ferric heme iron can be arranged in the following order of field strength: intermediate \( (s = 3/2) \) > high spin \( (S = 5/2) \) > low spin \( (S = 1/2) \).
Chapter 2. General Overview of Experimental Methods

In addition, the degree of axial or rhombic distortion can be directly calculated from g values (Palmer, 1985) and provide further information concerning coordination environments. Blumberg and Peisach (1971) related the degree of rhombicity to the axial ligand field strength through the study of a large number of low spin ferric heme proteins. This compilation of data is of value for the assignment of coordination environments.

2.7 Peroxidase Activity Measurements

Most assays of peroxidase enzymatic activity employ a reducing substrate that yields a colored product. A number of aromatic amines have been used for this, but are undesirable due to their carcinogenic nature. An alternative is ABTS (2,2’-Azido-di(3-ethyl-benzthiazoline-6-sulphonic acid)) which is stable and has a well defined visible absorption spectrum. As discussed in Section 1.2 and evident from equation 1-5, if the ABTS concentration is fixed and initial rates of product formation at various peroxide concentrations are measured, than \( k_j \) can be obtained from the slope of a plot of \( 2[E]/v_0 \) vs. \( 1/[H_2O_2] \), assuming that \( k_2 \gg k_3 \) (Dunford, 1991).

Using this approach peroxidase activity measurements were made with ABTS for all myoglobin variant proteins described in this thesis. These assays were carried out in 0.1 M MES, 0.2 mM ABTS, at pH 6.0 and 25 °C, and with a protein concentration of 0.2 μM (total volume 1 ml). Each variant protein was assayed using peroxide concentrations of 20, 5.0, 2.0, 1.5, 1.0 and 0.5 mM. Reactions were directly carried out in 1 ml cuvettes and initiated by adding protein. The progress of a given set of reactions was monitored at 414 nm (Childs & Bardsley, 1975) using a Cary 3 spectrophotometer fitted with a thermostatted circulating water bath. Individual assays at different peroxide concentrations were plotted and initial rate data derived from the beginning linear section of plots which generally corresponded to the first 15 to 30 seconds of the recorded
Chapter 2. General Overview of Experimental Methods

Figure 2.8. An example of a typical assay for peroxidase activity using recombinant wild-type horse heart myoglobin. The lines represent the increase in absorbance at 414 nm resulting from the generation of ABTS cation for different concentrations of hydrogen peroxide (0.5, 1.0, 1.5, 2.0, 5.0, and 20 mM) in the reaction mixture. The initial rate of each reaction was estimated from the linear sections of these lines which were generally restricted to the first 30 seconds of each assay. The initial rate values obtained were then used to generate a plot from which a $k_f$ value was determined (see Figure 2.9).

measurements. An example of kinetic data obtained in this way for recombinant wild-type myoglobin is presented in Figure 2.8. Such initial rate data were then used with the corresponding protein and hydrogen peroxide concentrations, to generate a double reciprocal plot as shown in Figure 2.9. The slope of the resulting line corresponds to $1/k_f$ for the peroxidase reaction and in
Figure 2.9. A double reciprocal plot prepared from the initial peroxidase rate data presented in Figure 2.8. The hollow triangles indicate the observed initial rates at corresponding hydrogen peroxide concentrations. From the slope of a least squares fit line to these points it can be determined that the value of $k_f$ for recombinant wild-type horse heart myoglobin is 13 mM$^{-1}$ min$^{-1}$.

The case of recombinant wild-type myoglobin is found to be 13 mM$^{-1}$ min$^{-1}$. 
Chapter 3

Myoglobin as a Structural Framework for Protein Engineering

3.1 Introduction

The advent of genetic manipulation techniques and the ability to synthesize genes has allowed the introduction of specific substitutions into the amino acid sequence of a protein and facilitated the expression of such variants. These techniques make accessible the exploration of fundamental questions directed at protein structure-function relationships and the possibility of constructing new proteins with novel features for use in industrial and medical applications. As outlined in Chapter 1, the goal of the work in this thesis is to use these techniques, in conjunction with modelling and structure determination methods, to increase our understanding of the mechanism of myoglobin, and the origin of this protein's latent peroxidase activity. The ultimate objective is to use myoglobin as a structural platform for the creation of an industrially useful manganese peroxidase.

An excellent starting point for studies of myoglobin is the availability of the high resolution structure of the horse heart protein (Evans & Brayer, 1990) and the fact that a synthetic gene encoding horse heart myoglobin has been constructed and can be expressed in E. coli (Guillemette et al., 1991). This provides the necessary basis for producing not only the wild-type protein, but also for obtaining specifically designed variant proteins to probe structure-function relationships and to begin the process of redesigning the active site of this protein to produce novel functions. Another distinct advantage of this system is the comprehensive background knowledge that is available for horse heart and other myoglobins (for example see review by Springer et al., 1994) that provides the groundwork for further structure-function studies and sets the stage for more
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

challenging protein engineering efforts. In addition, these studies can draw on the known three-dimensional structures of other myoglobins including those from the sperm whale (Phillips, 1990), pig (Smerdon et al., 1990), human (Hubbard et al., 1990), elephant (Bisig et al., 1995), seal (Scouloudi & Baker, 1978), sea turtle (Nardini et al., 1995), yellow fin tuna (Birnbaum et al., 1994) and mollusc (Conti, 1993).

Earlier studies have shown horse heart myoglobin is a globular protein (Figure 1.2) composed of a single polypeptide chain of 153 amino acids with a molecular weight of approximately 17.8 kDa (Dautrevaux et al., 1969; Romero-Herrera & Lehmenn, 1974). This protein consists of eight α-helices (79% of the total polypeptide chain) that are connected by short surface loops (Figure 1.3). These helices pack such that they form a hydrophobic pocket in which the heme prosthetic group in noncovalently bound to the protein. This heme group is almost completely buried in myoglobin with only the two propionate groups being exposed to solvent at the protein surface (Figure 1.2). Four of the six coordination sites of the heme iron atom are occupied by the nitrogen atoms of the four heme pyrrole rings. At its fifth coordination site, the heme iron atom is bound to the NE2 nitrogen atom of the proximal heme pocket residue His93. The sixth coordination site, located on the opposite or distal side of the heme plane, is the site of binding for exchangeable ligands such as molecular oxygen and carbon monoxide. The distal heme pocket residue, His64, is implicated as a factor in ligand entry and recognition processes (Springer et al., 1994). When serving in the role of oxygen storage or translocation, the heme iron atom of myoglobin is in the ferrous (II) form. When this heme iron atom is oxidized to its ferric (III) form it binds a water molecule at the sixth coordination site. This latter form of the protein is referred to as metmyoglobin.

Beyond gaining insight into structure-function relationships in myoglobin, a further goal of the work in this thesis was to use protein engineering approaches in which horse heart myoglobin
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

is used as a structural template to design a new enzyme exhibiting manganese peroxidase activity. This latter aspect of the project was divided into two streams of study. The first was to enhance the latent peroxidase activity (Mauk et al., 1973) of myoglobin using the heme prosthetic group as the reaction center. Here, the strategy was to remodel the active site of myoglobin by making substitutions at residues having positions related to those in cytochrome c peroxidase and other heme enzymes. In this way an assessment of the importance of these residues in contributing to a peroxidase function could be evaluated. Work in this area is discussed in Chapters 4, 5 and 6. The second area of study was to create a manganese binding site at a suitable location on myoglobin and this work is discussed in Chapter 7.

The focus of the current chapter is to establish that the structure of horse heart myoglobin, as expressed from a synthetic gene in *E. coli*, is properly folded and has a comparable structure to that of the natural enzyme obtained from horse heart tissue. This is an important aspect of studies described in following chapters where both the recombinant wild-type and specifically designed variant myoglobins are compared in order to understand the functional consequences of structural alterations. Also, in following chapters the recombinant wild-type protein has been used as the basis to enhance the latent peroxidase activity of myoglobin and it is therefore crucial to know the structural starting point represented by the recombinant wild-type horse heart myoglobin in order to help gauge the success in these studies.

### 3.2 Experimental Procedures

#### 3.2.1 Structural Studies

Crystals of recombinant wild-type horse heart myoglobin in the oxidized met-form were grown at room temperature (25 °C) using the hanging drop vapour diffusion technique. Each 10 μL hanging droplet (pH 8.0) contained 8 mg/mL protein, 60-62% saturated ammonium sulfate, 20 mM Tris
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

HCl and 1 mM EDTA. The corresponding hanging drop well solution contained 1.0 mL (pH 8.0) of 68-70% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA. Crystals grew to a maximum size of 0.2 x 0.2 x 0.1 mm in about 4 weeks. Recombinant wild-type metmyoglobin crystals were found to be isomorphous with those grown for wild-type horse heart metmyoglobin (Sherwood et al., 1987; Evans & Brayer, 1988; 1990) as detailed in Table 3.1.

Diffraction data for recombinant wild-type horse heart myoglobin was collected on a Rigaku R-Axis II imaging plate area detector system to 1.7 Å resolution. This area detector used CuKα radiation generated from a Rigaku RU300 rotating anode fitted with a monochromator and operated at 90 mA and 59 kV. The exposure time used per image plate was 30 minutes. In order to maximize the amount of data that could be obtained diffraction intensities were collected in two separate sets. Following the collection of the first data set where the glass capillary containing the crystal was mounted vertically on a goniometer head, a second set was collected by adjusting the glass capillary 45° off vertical. For each diffraction data set reflections were collected over a rotation range of 90° about the spindle axis using an oscillation angle of 1.2° and a crystal to detector distance of 61 mm. The R-Axis II data processing software (Higashi, 1990; Sato et al., 1992) was used to convert diffraction intensities into structure factors. This process is based on techniques developed by Rossmann (1979) and is described in greater detail in Chapter 2. The structure factor data set obtained was put on an absolute scale using the method of Wilson (1942) and final data processing statistics are presented in Table 3.1.

The starting atomic coordinates used for refinement of the structure of recombinant wild-type horse heart myoglobin were those of the 1.9 Å resolution structure of its naturally obtained counterpart (Evans & Brayer, 1990). Included in the recombinant wild-type myoglobin refinement model were a sulfate ion and well defined water molecules having thermal factors of less than 35 Å², as well as an internal water molecule bound to the heme iron atom. Structural refinement
Table 3.1: Data collection parameters for recombinant wild-type horse heart myoglobin.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Recombinant(^{\ddagger}) wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2(_1) (P2(_1))</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>64.3 (64.3)</td>
</tr>
<tr>
<td>b</td>
<td>29.0 (28.9)</td>
</tr>
<tr>
<td>c</td>
<td>35.9 (35.9)</td>
</tr>
<tr>
<td>β</td>
<td>107.2° (107.1°)</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>64359</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>10426</td>
</tr>
<tr>
<td>Merging R-factor (%)(^{\dagger})</td>
<td>8.3</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>(\infty - 1.7)</td>
</tr>
</tbody>
</table>

\(^{\ddagger}\) Corresponding crystallographic unit cell parameters for naturally occurring wild-type horse heart myoglobin are shown in parentheses (Evans & Brayer, 1990).

\(^{\dagger}\) \(R_{merge} = \frac{\sum_{hkl} \sum_{i=0}^{\infty} |I_{hkl} - \bar{I}_{hkl}|}{\sum_{hkl} \sum_{i=0}^{\infty} I_{hkl}}\)

utilized 10106 structure factors in the resolution range of 6.0 to 1.7 Å. During the course of refinement, omit, \(F_o - F_c\) and \(2F_o - F_c\) difference electron density maps covering the entire polypeptide chain of recombinant wild-type myoglobin were examined at three different times. This allowed for the manual adjustment of a number of main and side chain conformations during the refinement process. Additional water molecules were identified and added to the refinement...
model using an automated solvent search method (Tong et al., 1994). The presence of such water molecules were manually confirmed using $F_o - F_c$ difference electron density maps. Water molecules were only retained in the refinement model if consistently observed in $F_o - F_c$ and $2F_o - F_c$ maps, were found to form reasonable hydrogen bonds to protein atoms and refined to thermal factors $< 60 \text{Å}^2$. All water molecules were refined as neutral oxygen atoms having full occupancy.

Refinement was concluded when overall r.m.s. positional shifts became small ($\leq 0.03 \text{Å}$), indicating that convergence had been reached. Final refinement and stereochemical statistics for the structure of recombinant wild-type myoglobin are tabulated in Table 3.2. Two separate estimates of coordinate errors were made for the final refinement model. Examination of a plot of the crystallographic R-factor as a function of resolution, along with theoretical plots calculated by Luzzati (1952) as described in Chapter 2, indicated an overall r.m.s. coordinate error of 0.18 Å. A method developed by Cruickshank (1949, 1954, 1985) based on estimates of individual atomic errors gives an overall r.m.s. coordinate error of 0.14 Å.

3.2.2 Circular Dichroism

Circular dichroism measurements of recombinant wild-type horse heart myoglobin and its heme extracted apo form were performed with a Jasco Model J-720 spectropolarimeter calibrated with ammonium-d-camphor-10-sulfonate (ACS). Spectra (190-250 nm) were recorded with samples placed in a water-jacketed, cylindrical quartz cuvette (pathlength 0.1 cm), and temperature was regulated with a NESLAB Model RT 110 circulating water bath operated under computer control. The cuvette temperature was measured with a NESLAB RS-2 remote sensor interfaced to a computerized data acquisition system. Protein solutions ($\approx 1 \times 10^{-5} \text{M}$) were prepared in 10 mM potassium phosphate buffer (pH 7.0 and pH 8.0). Apoprotein was prepared by acidification to pH 2.0 and 2-butanone extraction (Teale, 1959). Following heme extraction the apomyoglobin
Table 3.2: Refinement results for recombinant wild-type horse heart myoglobin.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Recombinant wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Refinement results</strong></td>
<td></td>
</tr>
<tr>
<td>No. of reflections</td>
<td>10106</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>6.0-1.7</td>
</tr>
<tr>
<td>Completeness in range (%)</td>
<td>74</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1242</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>73</td>
</tr>
<tr>
<td>Average thermal factors (Å²)</td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>14.2</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>30.0</td>
</tr>
<tr>
<td>Final refinement R-factor (%)†</td>
<td>16.0</td>
</tr>
<tr>
<td><strong>2. Stereochemistry of final model</strong></td>
<td></td>
</tr>
<tr>
<td>Bond distances (Å)</td>
<td></td>
</tr>
<tr>
<td>Bond (1-2)</td>
<td>0.020</td>
</tr>
<tr>
<td>Angle (1-3)</td>
<td>0.038</td>
</tr>
<tr>
<td>Planar (1-4)</td>
<td>0.055</td>
</tr>
<tr>
<td>Planes (Å)</td>
<td>0.015</td>
</tr>
<tr>
<td>Chiral volumes (Å³)</td>
<td>0.029</td>
</tr>
<tr>
<td>Non-bonded contacts (Å) †</td>
<td></td>
</tr>
<tr>
<td>Single torsion</td>
<td>0.214</td>
</tr>
<tr>
<td>Multiple torsion</td>
<td>0.177</td>
</tr>
<tr>
<td>Possible hydrogen bonds</td>
<td>0.187</td>
</tr>
<tr>
<td>Torsion angles (°)</td>
<td></td>
</tr>
<tr>
<td>Planar (0° or 180°)</td>
<td>1.8</td>
</tr>
<tr>
<td>Staggered (±60°, 180°)</td>
<td>19.2</td>
</tr>
</tbody>
</table>

\[
\hat{R} \text{-factor} = \frac{\sum_{w} | | F_{i} | - | F_{i}^\prime | |}{\sum_{w} | F_{i} |}
\]

† The r.m.s. deviations from ideality for this class of restraint incorporate a reduction of 0.2 Å from the radius of each atom involved in a contact.
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

solution was first dialysed against 0.6 mM NaHCO₃ containing 1 mM EDTA and then against 10 mM sodium phosphate buffer (pH 7.0 and 8.0) using a microdialysis setup (Overall, 1987).

Thermal denaturation studies were performed by monitoring ellipticity at 222 nm. Temperature was increased with a heating rate of 50 °C/h from 30 to 85 °C. The average of three measurements was used to determine the $t_m$. The measured melting curves were smoothed using the filter function of the Jasco software, and the first derivative of the melting curve was calculated for a more accurate determination of $t_m$.

3.2.3 Peroxidase Activity Measurements

The detailed methodology used to determine the peroxidase activity of recombinant wild-type horse heart myoglobin is outlined in Chapter 2. In short, these assays were carried out in a total volume of 1 mL (pH 6.0, 25 °C) containing 0.1 M MES and 0.2 mM ABTS. The recombinant wild-type protein was assayed using hydrogen peroxide concentrations of 20, 5, 2, 1.5 and 0.5 mM. Reactions were directly carried out in 1 mL cuvettes and initiated by adding 2 µL of protein solution to prepared reaction mixtures. The concentration of protein in each assay sample was 0.2 µM. The progress of each assay was monitored at 414 nm (Childs & Bardsley, 1975) using a Cary 3 spectrophotometer fitted with a circulating water bath set to 25 °C.

3.3 Results

3.3.1 Structural Results

To facilitate a comparison of the three dimensional atomic structure determined for recombinant horse heart myoglobin when it is expressed in E. coli using a synthetic gene, with the structure of wild-type myoglobin from horse heart tissue, a least squares best fit superposition of the two polypeptide chain backbones was performed. A plot of observed main and side chain positional
differences is presented in Figure 3.1. Excluding the disordered terminal residues 1 and 152-153, the overall average positional deviation found between main and side chain atoms is 0.21 and 0.53 Å, respectively. A stereo drawing of the superimposed structures of wild-type and recombinant wild-type horse heart myoglobins showing the course of both polypeptide backbones is presented in Figure 3.2. Large positional differences in main chain placement between these structures are found at the amino terminus (residue 1; Δd = 1.1 Å), the carboxy terminus (residues 152-153; Δd = 3.35 Å), as well as at Gly 23 (Δd = 0.48 Å), Leu49 (Δd = 0.53 Å) and Asp122 (Δd = 0.63 Å). The C and N terminal regions are particularly poorly defined in both myoglobin structures, suggesting the presence of substantial positional disorder. Thus, the large positional deviations observed at both termini are likely the result of differential fitting to poor electron density, rather than a reflection of the different routes of expression of the two myoglobins. Gly23 is part of the first N-terminal turn of the B-helix and the positional shift seen for this residue is probably the result of a better fit to the available electron density as represented at higher resolution in the recombinant wild-type horse heart myoglobin structure determination. Both Leu49 (CD loop) and Asp122 (GH loop) are located in surface exposed regions, have high observed thermal factors (Figure 3.3), and are poorly resolved in electron density maps. As with the shifts observed for N and C termini residues, it would seem the positional differences observed at Leu49 and Asp122 are the result of high polypeptide chain mobility and their resultant poor definition in electron density maps. Despite the positional shifts discussed above, it is notable that all hydrogen bond interactions formed between main chain atoms are similar in both the wild-type and recombinant wild-type horse heart myoglobin structures.

When recombinant proteins are expressed in *E. coli*, there is often incomplete removal of the initiation methionine (Ben-Basset & Bauer, 1986). This situation has been observed for
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

Figure 3.1. Plot of the average positional deviations between main (thick lines) and side (thin lines) chain atoms of the wild-type and recombinant wild-type myoglobin structures. The horizontal dashed line represents the overall average deviation between all main chain atoms, excluding disordered amino acids at the two polypeptide chain termini (residues 1 and 152-153). The filled circle at position 154 represents the average positional deviation between atoms of the two heme groups.

recombinant horse heart myoglobin (Guillemette et al., 1991) with ~75% retention of an extra N-terminal methionine. In the case of sperm whale myoglobin the equivalent N-terminal methionine was observed in the three dimensional structural analysis of this protein using x-ray diffraction methods (Phillips, 1990). In contrast, structural analyses of recombinant wild-type horse heart myoglobin showed no indication of an additional methionine residue at the N-terminal end of the polypeptide chain. Examination of electron density maps indicates such an additional residue would have sufficient room to pack on the protein surface and in the crystal lattice. It is
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

Figure 3.2. a) Stereo drawings of the full α-carbon backbones of wild-type (thin lines; PDB entry 1YMB) and recombinant wild-type (thick lines) horse heart myoglobins. The heme group, the proximal (His93) and distal (Wat156) heme ligands, as well as the distal His64 residue have also been drawn. For clarity every 20th residue along the polypeptide chain has been labelled with its one-letter amino acid designation and sequence number. b) Stereo drawings showing a detailed view of the superposition of the heme group and surrounding residues of wild-type (thin lines) and recombinant wild-type (thick lines) horse heart myoglobins.
possible this N-terminal methionine is sufficiently disordered that it can not be defined in such maps. However, this residue is also not observed in any related variant myoglobins, some of which pack in different space groups. This latter observation suggests the recombinant wild-type and variant horse heart myoglobins being crystallized are completely processed and do not have a N-terminal methionine residue.

Between the two horse heart myoglobin structures several side chain positional differences are observed (> 1.0 Å; Figure 3.1). However, with one exception, these all involve surface exposed residues (Gln9, Lys16, Lys56, Lys118 and Asn140) which have considerable positional uncertainty as indicated by the high thermal factors of these side chains and their poor definition in electron density maps. It is notable that, other than Ile101, no substantial positional differences are observed for buried side chains between the wild-type and recombinant wild-type myoglobins. The shift observed for Ile101 appears to be the result of its more precise definition in the higher resolution electron density maps available for recombinant wild-type horse heart myoglobin.

A comparison of main chain thermal factors was also carried out between the wild-type and wild-type recombinant horse heart myoglobins and this is illustrated in Figure 3.3. The overall trend of average thermal factor values along the polypeptide chain is similar in both proteins. In total, there are three regions of polypeptide chain with average main chain temperature factors greater than 25 Å$^2$. These include both terminal ends and a region centered around residue 49. The latter portion of polypeptide chain is part of a surface loop between the C and D helices (residues 44-50) and also exhibits noticeable main chain positional differences (Figure 3.1). A superposition of the detailed structures of the heme pockets of wild-type and recombinant wild-type horse heart myoglobins is illustrated in Figure 3.2. Heme group positioning in both structures is similar with an average positional deviation of 0.24 Å for all 43 heme atoms when the superpositioning process
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

Figure 3.3. The average thermal factors of the four main chain atoms of each amino acid in wild-type (thin line) and recombinant wild-type (thick line) horse heart myoglobins are plotted as a function of residue number. Differences in these thermal factor profiles are plotted above. For these plots the overall average main chain thermal factor of wild-type myoglobin has been normalized to that of the recombinant wild-type myoglobin structure.

is based on a fit of all main chain atoms (residues 2-151). This value is 0.22 Å if only heme atoms are used to obtain the same superpositioning of structures. An assessment of heme planarity and geometry for both structures is also summarized in Table 3.3. Comparable heme iron coordinate distances are found between the wild-type and recombinant wild-type myoglobins (see Table 3.3).
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

**Table 3.3. Heme geometry of wild-type and recombinant wild-type myoglobins.**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Wild-type</th>
<th>Recombinant wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>I. Angular deviations (°) of pyrrole ring plane normals from the porphyrin ring plane normal.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>5.6</td>
<td>4.3</td>
</tr>
<tr>
<td>B</td>
<td>3.8</td>
<td>4.4</td>
</tr>
<tr>
<td>C</td>
<td>5.4</td>
<td>3.6</td>
</tr>
<tr>
<td>D</td>
<td>7.6</td>
<td>3.0</td>
</tr>
<tr>
<td>II. Angular deviations (°) of pyrrole ring plane normals from the pyrrole nitrogen plane normal.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>8.9</td>
<td>5.9</td>
</tr>
<tr>
<td>B</td>
<td>8.5</td>
<td>4.1</td>
</tr>
<tr>
<td>C</td>
<td>8.8</td>
<td>6.6</td>
</tr>
<tr>
<td>D</td>
<td>11.5</td>
<td>5.0</td>
</tr>
<tr>
<td>III. Heme iron displacements from the pyrrole nitrogen and porphyrin ring planes (Å).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyrrole N</td>
<td>0.03</td>
<td>0.09</td>
</tr>
<tr>
<td>porphyrin</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>IV. Heme iron ligand bond distances (Å).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His93 NE2</td>
<td>2.26</td>
<td>2.12</td>
</tr>
<tr>
<td>Wat156 O</td>
<td>2.29</td>
<td>2.17</td>
</tr>
<tr>
<td>Heme NA</td>
<td>2.00</td>
<td>2.03</td>
</tr>
<tr>
<td>Heme NB</td>
<td>2.05</td>
<td>2.02</td>
</tr>
<tr>
<td>Heme NC</td>
<td>2.01</td>
<td>2.00</td>
</tr>
<tr>
<td>Heme ND</td>
<td>1.98</td>
<td>1.98</td>
</tr>
</tbody>
</table>

The pyrrole nitrogen plane is defined by the four pyrrole nitrogen atoms of the heme group. The four pyrrole ring planes are each defined by the five atoms of the ring and the first carbon atom attached to each of the four carbons of the ring. The porphyrin ring is defined by the five atoms in each of the four pyrrole rings, the four bridging methine carbon atoms, the first carbon atom of each of the eight side chains of the heme and the central iron atom of the heme (33 atoms in total). The heme atom nomenclature used in this table follows the conventions of the Protein Data Bank (Bernstein, 1977) and is diagrammatically illustrated in Figure 1.2 of this thesis.
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

As can be seen in Figure 3.2, the placements of the heme iron atom ligands His93 and Wat156 are also similar. In addition, the His93 NE2-Fe-O (Wat156) angle remains similar being 168.6° in the wild-type structure and 171.3° in recombinant wild-type horse heart myoglobin. The largest difference observed is the 0.14 Å shorter heme iron atom to His93 ligand bond distance in the recombinant wild-type protein.

One difference between these two myoglobin structures concerns the hydrogen bonding network about the two heme propionates. Two water molecules hydrogen bonded to the heme propionate groups in the wild-type protein are not seen in the recombinant wild-type protein. Of these, one is hydrogen bonded to the O2D atom of the pyrrole D propionate group and the other is hydrogen bonded to the O2A atom of the pyrrole A propionate group as well as the side chains of residues Ser92 and His97. It is possible that these differences are related to differing crystallization conditions, as well as the higher resolution data collected for the recombinant protein with a resultant improvement of structural definition in this region.

3.3.2 Thermal denaturation studies of recombinant wild-type horse heart myoglobin

The circular dichroism spectra (190-250 nm) of recombinant wild-type myoglobin is shown in Figure 3.4, as is the change in ellipticity with increasing temperature. At pH 7, recombinant wild-type horse heart myoglobin exhibited a \( t_m \) value of 81.3 ± 1°C. This compares to 79.0 ± 1°C observed for wild-type myoglobin (C.L. Hunter, personal communication). At pH 8.0 this value was 78.2 ± 1°C. At pH 7.0 apo-recombinant wild-type horse heart myoglobin exhibited a \( t_m \) of 60.4 ± 1°C and at pH 8.0 this value was 57.9 ± 1°C.
Chapter 3: Myoglobin as a Structural Framework for Protein Engineering

3.3.3 Peroxidase activity

Measurements of the peroxidase activity of recombinant wild-type horse heart myoglobin were conducted as described in Section 3.2.3 and in more detail in Chapter 2. Individual assays at different peroxide concentrations were plotted (Figure 2.4) and initial rate data derived from the initial linear sections of individual activity lines which corresponded to the first 15 to 30 seconds of the reaction. These initial rate data were then used to generate the double reciprocal plot shown in Figure 2.5. The slope of the line in this plot was taken as the rate constant $k_1$ and the value

Figure 3.4. Circular dichroism spectra of recombinant wild-type horse heart myoglobin (10 mM potassium phosphate buffer, pH 7.0, 25 °C). The first derivative of the CD melting curve at 222 nm is shown in the inset.
obtained was 13 mM\(^{-1}\) min\(^{-1}\). Peroxidase activity for recombinant wild-type horse heart myoglobin showed linear behaviour for reactions containing 0.5 - 20 mM H\(_2\)O\(_2\).

### 3.4 Discussion

The elucidation of the high resolution structure of recombinant horse heart myoglobin presented in this chapter had as its primary goal to confirm that this protein is properly expressed, and folded in a manner comparable to horse heart myoglobin obtained from natural sources. This is important in terms of using the recombinant protein as a valid substitution for the naturally occurring myoglobin and being able to use earlier studies as basis for new research planned for its recombinant counterpart. Studies of modified recombinant proteins are discussed in Chapters 4-7.

One major advantage in using the horse heart myoglobin system for detailed structure-function studies is the ability to crystallize the wild-type and recombinant wild-type proteins isomorphously. This feature has not been possible in other related systems (Phillips, 1990). As described in later chapters, the ability to grow isomorphous crystals also extends to specifically designed recombinant variant forms of horse heart myoglobin. This similarity in crystallization approach, coupled with comparable methods of refinements for both the wild-type and variant horse heart myoglobins, allows for the examination of these structures for subtle differences that might otherwise be obscured by alternative structure determination techniques and strategies.

As detailed in this chapter, the structures of the wild-type and recombinant wild-type myoglobins are comparable, if not identical, when considered in light of estimated atomic coordinate errors and differences in the resolutions of their structure determinations. In particular, it is clear that the polypeptide chain fold (Figure 3.2), as well as heme group placement and geometry (Table 3.3) of both myoglobins are similar. Differences noted in main and side chain placement can be traced to regions of high thermal mobility which have poor definition in electron
density maps. Thus, such apparent positional shifts can be ascribed to uncertainty in polypeptide chain fitting and refinement, as opposed to utilization of an E. coli expression system. In agreement with these structural results are studies showing the functional and spectroscopic properties of wild-type and recombinant wild-type horse heart myoglobins are identical (Guillemette et al., 1991). Collectively, these data confirm that recombinant horse heart myoglobin can be used as being representative of its naturally occurring counterpart.
4.1 Introduction

The major objectives of the studies conducted in this chapter were to examine the role of the distal heme pocket residue His64 in determining the ligation state of the heme iron atom in myoglobin and potentially isolate design possibilities to enhance the latent peroxidase activity of this protein. Despite a wealth of structural and functional information available for heme proteins, the present understanding of the detailed molecular factors governing heme ligation is limited (Springer et al., 1994). This means the derivation of new structural designs to elicit differential ligation states remain problematic. In this chapter structural and functional studies were pursued to clarify these issues by examining the His64Thr and His64Tyr variants of horse heart myoglobin.

Peroxidases, which also possess a distal histidyl residue, appear to be pentacoordinate in their resting state (Poulos et al., 1993; Kunishima et al., 1994; Sundaramoorthy et al., 1994). In the case of myoglobin the heme iron atom of the met form is six-coordinate and high spin, with a water molecule bound as a dissociable sixth or distal ligand. The presence of the distal heme pocket residue His64 leads to the formation of a hydrogen bond from the side chain of this residue to the water molecule coordinated to the heme iron atom at the distal coordination position. In the reduced form of myoglobin this sixth ligand site is either unoccupied (deoxy form) or bound by dioxygen (oxy form).

Recent spectroscopic (Morikis et al., 1990; Egeberg et al., 1990a; Ikeda-Saito et al., 1992; Rajarathnam et al., 1992; Balasubramanian et al., 1993; Bogumil et al., 1994; Tang et al., 1994), structural (Quillin et al., 1993; Rizzi et al., 1993) and kinetic (Braunstein et al., 1988; Springer et
Chapter 4: The Role of the Distal Residue His64

al., 1989; Carver et al., 1990; Cutruzzola et al., 1991; Sakan et al., 1993; Brancaccio et al., 1994) studies of myoglobin variants in which His64 has been replaced with various amino acid residues have assisted in determining the involvement of this distal histidine in ligand binding properties in both oxidation states. These studies show that such substitutions can lead to alterations in the axial coordination of the ferric heme iron of myoglobin. For example, spectroscopic results suggest substitution of His64 with tyrosine results in His-Fe-Tyr ligation (Egeberg et al., 1990a; Hargrove et al., 1994; Tang et al., 1994). Furthermore, resonance Raman studies have shown that an arginine at position 64 leads to a six-coordinate low-spin heme iron and this residue might coordinate to the ferric heme iron atom (Morikis et al., 1990).

Other spectroscopic studies have established that replacement of His64 with non-polar residues like leucine, phenylalanine and valine leads to formation of pentacoordinate metmyoglobin derivatives which lack a water molecule bound as a sixth heme iron ligand at the distal coordination site (Morikis et al., 1990; Ikeda-Saito et al., 1992; Biram et al., 1993; Qin et al., 1993). Loss of a coordinated water molecule is attributed to the additional hydrophobicity of the replacement residues and to the elimination of the stabilizing hydrogen bond formed by bound water to the distal His64 (Quillin et al., 1993; Bogumil et al., 1995). This heme ligation state would be similar to that found in peroxidases. Lack of a water molecule at the distal coordination position is also observed in some other naturally occurring globins with hydrophobic residues located at the position corresponding to that of the distal His64. Aplysia myoglobin (distal Val) and Glycera hemoglobin (distal Leu) (Mintorovitch & Satterlee, 1988; Bolognesi et al., 1989) are examples of such proteins.

In contrast, substitution of His64 with glycine does not cause loss of heme iron atom coordinated water since the normal hydrogen bond formed by His64 is replaced by a similar interaction with a second, well ordered water molecule (Quillin et al., 1993). Also, spectroscopic studies show replacement of His64 with polar residues such as aspartate and glutamine permits
retention of the coordinated water molecule (Morikis et al., 1990; Ikeda-Saito et al., 1992; Qin et al., 1992; Biram et al., 1993; Bogumil et al., 1995). These findings have been confirmed by high resolution x-ray analyses of several His64 variants of sperm whale myoglobin (Quillin et al., 1993).

In this chapter, the high resolution structures of the His64Tyr and His64Thr variants of horse heart myoglobin were elucidated to provide for a better understanding of the role of His64 in ligand binding and to obtain a structural basis for reported spectroscopic data. This in turn provides a basis for gaining further insight into the nature of the latent peroxidase activity exhibited by myoglobin.

4.2 Experimental Procedures

4.2.1 Structural studies

Crystals of the His64Thr and His64Tyr horse heart myoglobin variants were grown using the hanging drop vapour diffusion method as discussed in Chapter 2. For His64Thr myoglobin a 10 µL hanging droplet (pH 8.4) containing 8 mg/mL protein, 60-62% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA was suspended over a hanging drop well containing 1.0 mL (pH 8.4) of 62-65% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA. For His64Tyr myoglobin a 10 µL hanging droplet was constructed as for the His64Thr variant but was suspended over a well containing 1.0 mL (pH 8.4) of 68-70% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA. Crystals of both variant proteins grew to a maximal size of 0.2 x 0.1 x 0.05 mm in about 10 weeks and were isomorphous with those grown for wild-type (Evans & Brayer, 1988; 1990) and recombinant wild-type (see Chapter 3) horse heart metmyoglobins. Space group and unit cell parameters for these crystals are tabulated in Table 4.1.

Diffraction data were collected to 1.9 Å and 2.0 Å resolution, respectively, from crystals of the His64Thr and His64Tyr variants of myoglobin. This was accomplished with a Rigaku R-Axis II
imaging plate area detector system using CuK\textsubscript{\alpha} radiation from a RU-300 rotating anode generator fitted with a graphite monochromator and operated at 90 mA and 59 kV. For each data frame, the crystal was oscillated through a \( \phi \) angle of 1.2° and exposed for 25 - 45 minutes. The crystal to detector distance was set to 78 mm. X-ray intensity data were processed using the R-Axis II software (Higashi, 1990; Sato et al., 1992) which is based on techniques described by Rossmann et al. (1979) and statistics related to this procedure are summarized in Table 4.1. Note that crystals of the His64Tyr variant protein were especially small, diffracted weakly and were found to decay rapidly in the x-ray beam. Despite extensive crystallization trails over a wide range of conditions it was not possible to grow higher quality crystals in this case. Furthermore, a large number of His64Tyr variant protein crystals had to be screened on the area detector before a usable data set was finally obtained. Thus, the low completeness of the diffraction data set for the His64Tyr variant is a reflection of the small size of these crystals and the relatively low diffraction intensities present. While crystals obtained for the His64Thr variant protein were also small these proved to have greater stability in the x-ray beam allowing for longer data collection exposure times and thereby providing a more complete data set.

A restrained parameter least squares approach (Hendrickson, 1981; 1985) was employed to refine the His64Thr variant structure using the 1.7 Å resolution structure of recombinant wild-type horse heart metmyoglobin (see Chapter 3) as the starting model. For the His64Tyr variant, the 1.9 Å structure of wild-type horse heart metmyoglobin (Evans & Brayer, 1990) was used as the starting model since at the time the structure of recombinant wild-type horse heart metmyoglobin had not yet been completed. Included in initial refinements were well defined water molecules (\( B < 35 \text{ Å}^2 \)) and a sulfate ion found in the wild-type and recombinant wild-type myoglobin structures. For both variants, residue 64 was initially modelled as an alanine. Omit, \( F_o - F_c \) and \( 2F_o - F_c \)
Chapter 4: The Role of the Distal Residue His64

Table 4.1: Data collection parameters for the His64Thr and His64Tyr variant myoglobins.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>His64Thr</th>
<th>His64Tyr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>64.4</td>
<td>63.7</td>
</tr>
<tr>
<td>b</td>
<td>28.9</td>
<td>28.8</td>
</tr>
<tr>
<td>c</td>
<td>35.8</td>
<td>35.7</td>
</tr>
<tr>
<td>β</td>
<td>107.2°</td>
<td>106.6°</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>86121</td>
<td>22024</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>7270</td>
<td>3453</td>
</tr>
<tr>
<td>Merging R-factor (%)†</td>
<td>8.6</td>
<td>15.5</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>∞ - 1.9</td>
<td>∞ - 2.0</td>
</tr>
</tbody>
</table>

\[
R_{merge} = \frac{\sum_{i,j} \sum_{\mu} |I_{\mu i} - \overline{I}_{\mu}|}{\sum_{\mu} \sum_{i,j} I_{\mu i}}
\]

difference electron density maps covering the entire polypeptide chain of both variant proteins were examined periodically during refinement. These allowed for the manual adjustment of a number of surface side chain conformations and the placement of the remaining side chain atoms of residue 64.

The C-terminal region (residues 152-153) of both His64 variant proteins was found to be disordered and modelled by finding a geometrically reasonable configuration that maximized the use of the electron density present. Water molecule positions were determined by an iterative
procedure (Tong et al., 1994) and retained in the structure if they were found to form reasonable hydrogen bonds to protein atoms and refined to a thermal factor of less than 60 Å$^2$. Refinements were concluded when positional shifts became small (overall r.m.s. shifts < 0.03 Å) indicating that convergence had been reached. Final refinement results are tabulated in Table 4.2, along with statistics related to the stereochemistry of the final refinement models.

Atomic coordinate errors for these two refined structures have been estimated using two methods. Inspection of a Luzzati (1952) plot indicates overall coordinate errors of 0.18 Å and 0.21 Å, whereas the corresponding values using Cruickshank’s (1949; 1954; 1985) method are 0.17 Å and 0.21 Å, for the His64Thr and His64Tyr variant structures, respectively.

4.2.2 Peroxidase activity measurements

Peroxidase activity measurements were performed on the His64Thr and His64Tyr variant proteins using the methodology outlined in Chapter 2. These assays were carried out in a total volume of 1 mL (pH 6.0, 25 °C) containing 0.1 M MES and 0.2 mM ABTS. Each protein was assayed using hydrogen peroxide concentrations of 20, 5, 2, 1.5 and 0.5 mM. Reactions were directly carried out in 1 mL cuvettes and initiated by adding 2 μL of protein solution to prepared reaction mixtures. The concentration of protein in each assay sample was 0.2 μM. The progress of each assay was monitored at 414 nm (Childs & Bardsley, 1975) using a Cary 3 spectrophotometer fitted with a circulating water bath set at 25 °C. Peroxidase activity measurements for both of the His64Thr and His64Tyr variant proteins showed no measurable activity at any of the peroxide concentrations used.
Chapter 4: The Role of the Distal Residue His64

Table 4.2: Refinement statistics for the His64Thr and His64Tyr variant myoglobins.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>His64Thr</th>
<th>His64Tyr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Refinement Results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections</td>
<td>6916</td>
<td>3324</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>6.0-1.9</td>
<td>6.0-2.0</td>
</tr>
<tr>
<td>Completeness in Range (%)</td>
<td>58</td>
<td>43</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1240</td>
<td>1243</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>53</td>
<td>126</td>
</tr>
<tr>
<td>Average thermal factors (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>19.9</td>
<td>13.2</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>31.0</td>
<td>24.3</td>
</tr>
<tr>
<td>Final refinement R-factor (%) ‡</td>
<td>16.3</td>
<td>16.9</td>
</tr>
<tr>
<td><strong>2. Stereochemistry of final models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond distances (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond (1-2)</td>
<td>0.013</td>
<td>0.009</td>
</tr>
<tr>
<td>Angle (1-3)</td>
<td>0.040</td>
<td>0.022</td>
</tr>
<tr>
<td>Planar (1-4)</td>
<td>0.062</td>
<td>0.028</td>
</tr>
<tr>
<td>Planes (Å)</td>
<td>0.015</td>
<td>0.007</td>
</tr>
<tr>
<td>Chiral volumes (Å³)</td>
<td>0.013</td>
<td>0.146</td>
</tr>
<tr>
<td>Non-bonded contacts (Å) †</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single torsion</td>
<td>0.203</td>
<td>0.209</td>
</tr>
<tr>
<td>Multiple torsion</td>
<td>0.167</td>
<td>0.198</td>
</tr>
<tr>
<td>Possible hydrogen bonds</td>
<td>0.188</td>
<td>0.189</td>
</tr>
<tr>
<td>Torsion angles (°)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planar (0° or 180°)</td>
<td>1.9</td>
<td>1.1</td>
</tr>
<tr>
<td>Staggered (±60°, 180°)</td>
<td>19.4</td>
<td>24.3</td>
</tr>
</tbody>
</table>

‡ $R_{factor} = \frac{\sum_{i} |F_{o,i}|-|F_{c,i}|}{\sum_{i} |F_{o,i}|}$

† The r.m.s. deviations from ideality for this class of restraint incorporate a reduction of 0.2 Å from the radius of each atom involved in a contact.

74
Chapter 4: The Role of the Distal Residue His64

4.3. Results

4.3.1. The His64Thr variant structure

To facilitate comparison of the recombinant wild-type and His64Thr variant myoglobin structures, a least squares best fit superposition of the two polypeptide chain backbones was performed. A plot of the observed main and side chain positional differences observed is presented in Figure 4.1. These results indicate an overall average deviation of 0.15 Å for all main chain atoms. The largest displacements are found at the polypeptide chain termini (residue 1 and 152-153; Δd = 0.37 Å and 0.56 Å, respectively) and at residues 42-48 (Δd = 0.33 Å). As described in Chapter 3, both terminal regions are poorly defined in the wild-type and recombinant wild-type myoglobin structures, suggesting the presence of substantial positional disorder. Thus, the large positional deviations observed for the His64Thr variant at both termini are likely the result of differential fitting to poor electron density, rather than a result of the amino acid substitution introduced at position 64. The shifts observed at residues 42-48 represent a displacement of the loop region between the B and C helices and are likely related to the replacement of His64 by a threonine residue. Significant side chain positional differences (> 0.8 Å) all involve charged and surface exposed residues with the exceptions of the side chain atoms of Leu61 (Δd = 0.95 Å) and Ile101 (Δd = 0.82 Å). Close proximity to residue 64 appears to cause the observed reorientation of the side chain of Leu61 in the variant protein. Ile101, on the other hand, is poorly defined in electron density maps for both wild-type and variant proteins, and the observed differences are likely due to positional uncertainty.

A detailed comparison of the His64Thr and recombinant wild-type structures in the region of residue 64 is shown in Figure 4.2. The heme iron atom is notably drawn out of the porphyrin plane towards His93 in the His64Thr variant, with the distance between this iron atom and the porphyrin ring plane increasing from 0.08 Å in recombinant wild-type myoglobin to 0.28 Å (Table
Figure 4.1. Plots of the average positional deviations between the main (thick lines) and side (thin lines) chain atoms of a) recombinant wild-type and His64Thr variant and b) recombinant wild-type and His64Tyr variant horse heart myoglobins. The horizontal dashed line on each plot represents the overall average positional deviation found for main chain atoms when the disordered terminal residues 1 and 152-153 are excluded from least squares superpositioning. The filled circle at residue position 154 represents the overall average deviation found for all 43 heme atoms.
4.3). The shorter His93-Fe ligand bond observed may contribute to the loss of the distal water molecule (Wat156) which in the wild-type protein also forms a hydrogen bond to His64. Despite these differences, the overall conformations of the heme groups in the wild-type and His64Thr variant proteins are comparable and the observed average deviation in the positions of all 43 heme atoms is 0.22 Å. This similarity extends to heme planarity (Table 4.3) and placement of the proximal heme ligand His93, as well as other residues that constitute the heme binding pocket (Figure 4.2). However, substantial alterations are observed in the hydrogen bonding network about the mutation site near the heme in the His64Thr variant. For example, a new hydrogen bond between the side chains of Thr64 and Lys45 is formed (Figure 4.2). It appears the observed positional shift of residues 42-48 (Figure 4.1) is likely related to this newly introduced hydrogen bond. Thr64 also forms a link to Asp60 through a further hydrogen bond to Wat169. This water molecule is also present in the recombinant wild type myoglobin structure.

4.3.2 The His64Tyr variant structure
A plot of the observed main and side chain positional differences observed upon comparison of the His64Tyr and recombinant wild-type myoglobins is presented in Figure 4.1. The overall average displacement in main and side chain atoms is 0.30 and 0.55 Å, respectively. With the exception of the disordered N and C-terminal regions of the polypeptide chain (residues 1; Δd = 2.5 Å and 152-153; Δd = 1.6 Å) no substantial alterations in main chain atom positions are observed. The two largest displacements involve regions about residue 49 and 89 (Figure 4.1). As with the shifts observed for N and C termini residues, it would seem the positional differences observed at Leu49 (Δd = 0.60 Å) is the result of high polypeptide chain mobility. This region of the polypeptide chain lies between the C and D helices and has high thermal factors with the highest average main chain atomic thermal factors of any non-terminal region in recombinant wild-type (Figure 3.3) and
Chapter 4: The Role of the Distal Residue His64

Figure 4.2. a) Stereo diagrams showing the region about residue 64 in the recombinant wild-type (thin lines) and His64Thr variant (thick lines) horse heart myoglobins. Heme group conformation is comparable between these two proteins with the exception of the heme iron atom being drawn out of the porphyrin plane toward His93 in the His64Thr variant protein (see Table 3.3 and 4.3). Note the absence of Wat214 and the distal heme ligand Wat156 in the His64Thr variant protein. Both of these waters form a hydrogen bond to the side chain of His64 in wild-type myoglobin. The OG hydroxyl group of the Thr64 side chain in the His64Thr variant protein hydrogen bonds to Lys45 and Wat169. b) A comparison of the region about residue 64 in the His64Thr variant of horse heart (thick line) and sperm whale (thin line) myoglobins (Quillin et al., 1993). Sequence differences at residues 45 and 67 are indicated. Neither Wat155, nor Wat303 are observed in horse heart metmyoglobin. Wat169 is not observed in His64Thr sperm whale metmyoglobin. Hydrogen bonds in all diagrams are indicated with dashed lines.
Table 4.3: Heme geometry of His64 variant myoglobins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>His64Thr</th>
<th>His64Tyr</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Angular deviations (°) of the pyrrole ring plane normals from the porphyrin ring plane normal.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>B</td>
<td>2.6</td>
<td>4.5</td>
</tr>
<tr>
<td>C</td>
<td>3.5</td>
<td>7.0</td>
</tr>
<tr>
<td>D</td>
<td>3.1</td>
<td>11.9</td>
</tr>
<tr>
<td>II. Angular deviations (°) of the pyrrole ring plane normals from the pyrrole nitrogen plane normal.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1.5</td>
<td>4.4</td>
</tr>
<tr>
<td>B</td>
<td>3.5</td>
<td>11.5</td>
</tr>
<tr>
<td>C</td>
<td>4.3</td>
<td>3.1</td>
</tr>
<tr>
<td>D</td>
<td>4.8</td>
<td>18.7</td>
</tr>
<tr>
<td>III. Heme iron displacements from the pyrrole nitrogen and porphyrin ring planes (Å).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>porphyrin plane</td>
<td>-0.28</td>
<td>+0.38</td>
</tr>
<tr>
<td>pyrrole nitrogen plane</td>
<td>-0.28</td>
<td>+0.39</td>
</tr>
<tr>
<td>IV. Heme iron ligand bond distances (Å).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His93 NE2</td>
<td>1.93</td>
<td>2.29</td>
</tr>
<tr>
<td>Tyr64 OH</td>
<td>—</td>
<td>2.18</td>
</tr>
<tr>
<td>Heme-NA</td>
<td>2.04</td>
<td>2.05</td>
</tr>
<tr>
<td>Heme-NB</td>
<td>2.01</td>
<td>2.01</td>
</tr>
<tr>
<td>Heme-NC</td>
<td>1.99</td>
<td>2.05</td>
</tr>
<tr>
<td>Heme-ND</td>
<td>2.00</td>
<td>1.93</td>
</tr>
</tbody>
</table>

The pyrrole nitrogen plane is defined by the four pyrrole nitrogens of the heme group. The four pyrrole ring planes are each defined by the five atoms of the ring and the first carbon atom attached to each of the four carbons of the ring. The porphyrin ring is defined by the five atoms in each of the four pyrrole rings, the four bridging methine carbon atoms, the first carbon atom of each of the eight side chains of the heme and the central iron atom of the heme (33 atoms in total). The heme atom nomenclature used in this table follows the conventions of the Protein Data Bank (Bernstein, 1977) and is diagrammatically illustrated in Figure 1.2. The direction of heme iron displacements to the distal or proximal side of the heme group are indicated by + and - signs, respectively.
Chapter 4: The Role of the Distal Residue His64

Figure 4.3. a) Stereo drawing of the α-carbon backbones of the recombinant wild-type (thin lines) and His64Tyr variant (thick lines) horse heart myoglobins. The heme group, the heme ligand His93 and the side chains of Leu29, Leu104 and residue 64, have also been drawn. Every 20th α-carbon atom has been labelled with its one-letter amino acid designation and sequence number. b) Stereo drawing showing heme ligation in the recombinant wild-type (thin lines) and His64Tyr variant (thick lines) horse heart myoglobins. In the wild-type met form of myoglobin a water molecule is found to simultaneously interact with the heme iron atom and His64 in the distal heme pocket. In the His64Tyr variant protein the hydroxyl group of Tyr64 directly ligates to the heme iron atom.
variants myoglobins. The positional deviation observed for Leu89 ($\Delta d = 0.80 \, \text{Å}$) which is located within the F-helix is possibly the result of fitting to lower quality and completeness data rather than due to the substitution of His64. Generally, the larger variability in atomic positions observed between the His64Tyr and wild-type myoglobins is likely due to both the lower resolution and poorer quality of data that could be collected for the His64Tyr variant myoglobin with the resultant loss of clarity in the electron density maps of this protein.

As Figure 4.3 shows, within the heme pocket of the His64Tyr variant myoglobin, the side chain of Tyr64 forms an axial ligand to the heme iron atom. This confirms an earlier spectroscopic analysis of this variant protein (Tang et al., 1994) and that of the analogous variant sperm whale myoglobin (Egeberg et al., 1990a). One conformational change that appears to be directly associated with the His64Tyr substitution involves Leu 29. For this residue the average side chain atom displacements is 1.1 Å (see Figures 4.1 and 4.3). The change in conformation observed ($\Delta \chi_1 = 11^\circ; \Delta \chi_2 = 30^\circ$) allows for the larger spatial requirements of a tyrosyl side chain at residue position 64.

Although occupying a comparable spatial positioning to that of the wild-type protein, considerable planar distortion is evident in the heme group of the His64Tyr variant protein (Table 4.3). The overall average positional deviation for all 43 heme atoms between these two proteins is 0.46 Å. Major heme distortions are localized in pyrrole ring D and about the CHC carbon atom that links pyrrole rings B and C. In contrast to the wild-type protein where the heme iron atom is found in the plane of the pyrrole nitrogen atoms, in the His64Tyr variant the heme iron atom is displaced 0.39 Å from this plane in the direction of the new distal heme pocket ligand Tyr64. The proximal His93 ligand also shifts by a similar amount towards the distal side of the heme group. With these movements, the heme iron bond distance to His93 is longer ($\Delta d = 0.17 \, \text{Å}$) in the His64Tyr protein whereas the heme iron atom to pyrrole ring nitrogen bond distances are comparable in both the wild-type and variant myoglobins (Table 3.3 and 4.3). It is notable that the 2.17 Å distance from
Chapter 4: The Role of the Distal Residue His64

the heme iron atom to the associated water molecule in wild-type metmyoglobin is similar to the tyrosine hydroxyl to heme iron atom distance of 2.18 Å in the His64Tyr variant protein.

4.4 Discussion

4.4.1 Structure-function implications of His64Thr substitution

The structural studies presented here establish that the coordination environment of the heme iron atom in the His64Thr variant protein of horse heart metmyoglobin is pentacoordinate. Despite having a heme coordination environment in common with many other peroxidases, this variant protein has no such enzymatic activity (Section 4.2.2). This is in contrast to the low level latent peroxidase activity observed in recombinant wild-type horse heart myoglobin (see Chapter 3). There are two potential reasons for this situation. First, removal of the distal heme pocket His64 may prevent initial peroxide orientation adjacent to the heme group and binding to the iron atom of this moiety since the substituted Thr64 can not form a stabilizing hydrogen bond to bound heme ligands like that observed for His64 (Springer et al., 1994). Second, the distal heme pocket histidine residue found in peroxidases (eg. His52 in cytochrome c peroxidase) also appear to actively stabilize reaction intermediates during catalysis (Poulos & Kraut, 1980; Vitello et al., 1990; Erman et al., 1992; Miller et al., 1994). As with initial peroxide binding, this feature is not provided for the His64Thr variant protein given the orientation and smaller size of the side chain of Thr64 (Figure 4.2).

The pentacoordinated state of the heme group of the His64Thr variant protein was anticipated from the results of a related substitution involving isoleucine in sperm whale myoglobin (Morikis et al., 1990; Ikeda-Saito et al., 1992; Rajarathnam et al., 1992). However, as shown in Figure 4.2 this result contrasts with the recently reported structure of the corresponding His64Thr variant of sperm whale myoglobin where a water molecule (Wat155) having an occupancy of 72% was found to be present as the sixth ligand to the heme iron atom (Quillin et al., 1993). In sperm whale myoglobin
this water molecule also forms an interaction with the carbonyl oxygen atom of Thr64 through a hydrogen bond with an additional new water molecule (Wat303).

Differences in distal heme ligation between His64Thr variant horse heart and sperm whale myoglobins appear to result from the combined effects of several relatively subtle differences in their structures. At least four such contributions can be identified. First, a lysine is present at position 45 in the horse heart protein, while an arginine is located at this position in sperm whale myoglobin (see Figure 4.2). As a result, the hydrogen bond interactions formed by residue 45 with heme propionate D, Thr64, and Asp60 are different in the two proteins. Most notably, the NH2 atom of Arg45 in sperm whale myoglobin is located near the position that Watl69 occupies in the horse heart protein and as a result Watl69 is not present in the sperm whale protein. In the His64Thr variant of horse heart myoglobin Wat 169 is hydrogen bonded to Thr64. Second, the conformation of Phe46 is shifted between the horse heart and sperm whale variant proteins. In His64Thr sperm whale myoglobin, the side chain of this group is located further from the site of axial water coordination making this region of the distal heme pocket more hydrophilic in character. Third, in His64Thr horse heart myoglobin, a valine is present at residue position 67, while a threonine is present at this position in sperm whale myoglobin (Figure 4.2). The greater polarity of the threonyl residue would favour the presence of water molecules in the distal heme binding pocket of the sperm whale protein. Fourth, the position of the backbone carbonyl group of residue 64 in horse heart myoglobin is located ~ 0.5 Å further from the heme iron than it is in the sperm whale protein. In His64Thr sperm whale myoglobin this feature would facilitate formation of a hydrogen bond to Wat303 which appears to help stabilize the observed distal water ligand (Wat 155; Figure 4.2)

In addition, at least two differences can be identified between the His64Thr variant and wild-type horse heart myoglobin structures that may contribute to the unique titration behaviours of their proximal His93 ligands (Bogumil et al., 1995). First, in the variant protein the ferric high-spin heme iron atom moves out of the heme plane towards His93 (Table 4.3) and the resultant bond

83
length (1.93 Å) is shorter than that observed in the wild-type protein (2.12 Å). This feature could contribute to the observed decrease in the pK$_a$ of the coordinated imidazole in the variant protein to 9.49 (Bogumil et al., 1995) from ~10 found for wild-type horse heart myoglobin (Morishima et al., 1980). Second, the trans effect of coordination of an anionic hydroxide ligand to the wild-type protein at alkaline pH should weaken Fe(III)-His93 interaction and decrease the effect of heme iron coordination on the pK$_a$ of the His93 imidazole group (Decatur et al., 1996).

4.4.2 Structure-function implications of His64Tyr substitution

The ligation bond length of 2.18 Å observed between the distal heme pocket Tyr64 and the heme iron atom in the His64Tyr variant of horse heart myoglobin (Table 4.3) is larger than the bond lengths of 1.84 Å and 1.89 Å found for the proximal tyrosinate ligands of beef liver catalase (Fita & Rossmann, 1985) and the catalase from Proteus mirabilis (Gouet et al., 1995), respectively. In addition, extended x-ray absorption fine structure analysis of the His64Tyr variant protein ligation determined a bond length of 1.88 Å for the Tyr64 to heme iron atom interaction. This apparent ligation bond length discrepancy between extended x-ray absorption fine structure analyses and x-ray crystallographic results may represent a difference in heme iron ligation between the solution and crystalline states. This possibility was assessed by examination of EPR spectra for His64Tyr horse heart myoglobin (Maurus et al., 1994). In solution, the EPR spectrum of the His64Tyr variant protein was shown to be independent of pH between values of 7 and 11, indicating that Tyr64 remains coordinated to the heme iron atom over this range of pH. In the crystalline form two unique axial high spin signals are observed for the His64Tyr variant protein. These could originate as the result of some variant protein having His93-Fe(III)-H$_2$O ligation similar to that observed for wild-type metmyoglobin. This interpretation implies that in the crystalline state an equilibrium exists between a form of the protein with Tyr64 coordinated to the heme iron and a form in which a water molecule provides the sixth heme iron ligand. However, molecular modelling shows that this is unlikely due to a lack of space in the distal heme pocket to accommodate both simultaneous
Chapter 4: The Role of the Distal Residue His64

water ligation and the presence of the bulky Tyr64 side chain. Another possibility is that two different forms of His93-Fe(III)-Tyr64 ligation coexist in the crystalline state. In this case, an axial EPR signal would arise from a form of iron ligation with a longer iron-tyrosine bond length and a less distorted heme plane. Comparison of the thermal factors between the wild-type and variant proteins reveals comparable values for the heme moiety in both, but significantly higher values are observed for the main and side chain atoms of residue 64 in the variant structure (average is 11 Å² and 17 Å² for His64 and Tyr64, respectively). The apparent flexibility of Tyr64 and lack of space for water ligation provide strong circumstantial evidence that the additional axial high spin EPR signal observed in the crystalline form of the His64Tyr variant protein results from a second species having a different heme iron to tyrosine ligation bond length and a less distorted heme plane (Maurus et al., 1994).

4.4.3 His64 substitution and latent peroxidase activity

Heme containing peroxidases are thought to have essential residues in the distal heme pocket which facilitate proton transfer and stabilize the charge separation process (see Chapter 1; Poulos et al., 1980; Goodin et al., 1993) The results obtained in this chapter serve to underline the importance of the distal His64 in facilitating the presence of latent peroxidase activity in myoglobin, since neither of the His64Thr and His64Tyr variant myoglobins showed measurable peroxidase activity (see Section 4.2.2). Based on these results it seems that replacement of His64 in horse heart myoglobin is unlikely to have a positive role in enhancing its peroxidase activity.

After concluding these studies a sperm whale myoglobin variant was reported wherein His64 was replaced with a leucine residue and in order to create a peroxidase like active site, Phe43 was substituted with histidine (Ozaki et al., 1997). This strategy of moving the distal histidine further away from the heme iron atom proved successful as the peroxide dependent oxidation of styrene is found to occur at a rate that is 300 times that of wild-type myoglobin. These results suggest that a similar strategy might prove helpful in improving the manganese peroxidase activity of myoglobin.
Chapter 5

Azide as a Probe of Heme Protein Ligand Interactions

5.1 Introduction

The experiments described in this chapter were designed to facilitate the interpretation and understanding of the structural consequences of ligand binding to myoglobin with the focus being on complexation of linear ligands such as hydrogen peroxide and azide. In this regard, azide was chosen for detailed study since such heme protein complexes are structurally stable (Stryer et al., 1964; Sage et al., 1991; Rizzi et al., 1993). Three dimensional structural information obtained for azide binding, in conjunction with joint spectroscopic work, has the potential to provide critically important information concerning the ability of myoglobin to accommodate ligands of various sizes. It was hoped these results would also assist in the effort to understand the basis for the latent peroxidase activity of myoglobin and suggest amino acid modifications that might enhance this activity.

It has been established that the heme binding pocket exerts a substantial influence on the reactivity of the heme group and its ability to bind various ligands, and as a result attention has focused on determining which amino acids adjacent to the heme may be important factors in this regard (Adachi et al., 1992; Allocatelli et al., 1993; Huang & Boxer, 1994; Lambright et al., 1994; Theriault et al., 1994; Smerdon et al., 1995). Of particular interest are His93 and His64, the proximal and distal histidine residues, respectively. These two histidines are the two closest residues to the heme iron atom (Figure 3.2). The proximal His93 forms a direct heme iron axial ligand while the distal His64 is positioned in the ligand binding site where it is ~4.5 Å from the heme iron atom. The proximal His93 ligand is thought to play a crucial role in controlling the
binding energy of the active site (Egeberg et al., 1990; Adachi et al., 1993). On the other hand, distal residues such as His64 are thought to control accessibility to the heme iron atom and directly influence the orientation and conformational freedom of bound ligands (Springer et al., 1994).

The role of the distal His64 residue in modulating myoglobin ligand binding has received particular attention through the analysis of variants in which it has been replaced by a variety of other amino acids (see review: Springer et al., 1994). These recent studies have shown that replacement of His64 significantly alters the axial coordination of the ferric heme iron. For example, residues of an aliphatic nature like valine, leucine, phenylalanine and isoleucine lead to pentacoordinated ferric heme proteins lacking a water molecule as the sixth ligand in the distal heme pocket site (Morikis et al., 1989; Rohlfis, 1990; Mattevi et al., 1991; Ikeda-Saito et al., 1992; Biram et al., 1993; Quillin et al., 1993; 1994; Bogumil et al., 1995). The loss of this bound water molecule has been attributed to two factors. The first of these is the additional hydrophobicity of the nonpolar residues introduced into the distal heme pocket and the second involves the disruption of the stabilizing hydrogen bond between the distal His64 side chain and the normally coordinated water molecule.

In the ferric state myoglobin can also bind a variety of anions such as $N_3^-$, CN$, SCN$, and F$^-$ (Makinen & Churg, 1983). The azide anion has been widely used to prepare a stable low-spin iron Fe(III) heme protein derivative with the bound azide anion preferring a bent conformation when coordinated to the heme iron in both myoglobin (Sage et al., 1991; Stryer et al., 1964; Rizzi et al., 1993) and hemoglobin (Perutz & Mathews, 1966; Deatherage et al., 1979). Some of the factors governing azide and cyanide binding have been examined and it has been suggested that steric hinderance by His64 is a key factor in the regulation of the association rate of azide (Baker, 1984; Rizzi et al., 1993; Brancaccio et al., 1994). For example, larger association rates are observed when smaller amino acids like glycine, alanine and threonine are substituted for His64. However,
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

the azide affinity of wild-type myoglobin is greater than most position 64 variants as indicated by the much higher azide dissociation rates of these variants. The lower dissociation rate of wild-type myoglobin may be the result of a stabilizing hydrogen bonding interaction between His64 and the azide anion (Bogumil et al., 1994; Brancaccio et al., 1994).

An analysis by Fourier transform infrared (FTIR) spectroscopy of the interaction of azide with horse heart myoglobin variants has revealed that the electrostatic potential surrounding the bound ligand is one important factor governing the $v_{\text{max}}$ of the asymmetric azide stretch (Bogumil et al., 1994). The introduction of positively or negatively charged surface residues shifts $v_{\text{max}}$ to higher or lower energy, respectively, without a significant change in bandwidth. However, the largest changes were observed in the FTIR spectra of His64 variants and this was proposed to result from disruption of interactions between the distal His64 and the bound azide ligand.

In this chapter, the high resolution structures of the azide bound forms of wild-type and His64Thr horse heart myoglobins were elucidated to provide for a better understanding of the interaction of azide with myoglobin and the structural basis of reported spectroscopic data. Through this approach it was hoped to gain insight into the structural and functional characteristics of linear di- and tri-atomic ligands bound to myoglobin in general. This in turn provides a basis for understanding the nature of peroxide binding to this protein and the source of the latent peroxidase activity exhibited by myoglobin.

5.2 Experimental Procedures

Crystals of the ferric azide complexed forms of both the recombinant wild-type and His64Thr variant proteins were grown at room temperature (25 °C) using the hanging drop vapour diffusion method. For recombinant wild-type myoglobin, a 5 μL hanging droplet (pH 7.5) containing 15 mg/mL protein, 60% saturated ammonium sulfate, 20 mM Tris HCl, 5 mM NaN₃ and 1 mM
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

EDTA was suspended over a well containing 1.0 mL (pH 7.5) of 65% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA. For His64Thr myoglobin the 5 μL hanging droplet (pH 8.4) contained 15 mg/mL protein, 65% saturated ammonium sulfate, 20 mM Tris HCl, 25 mM NaN₃ and 1 mM EDTA and was suspended over a well (pH 8.4) containing 1.0 mL of 70% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA. For both proteins crystals grew to a maximal size of 0.2 x 0.1 x 0.05 mm in about 10 weeks and were isomorphous with those grown for wild-type (Evans & Brayer, 1988; 1990) and recombinant wild-type (see Chapter 3) horse heart metmyoglobins. Space group and unit cell parameters for azide complexed crystals are tabulated in Table 5.1, whereas the related values for wild-type and recombinant wild-type metmyoglobins can be found in Chapter 3 (Table 3.1).

Diffraction data sets for the azide forms of the recombinant wild-type and His64Thr myoglobins were collected on a Rigaku R-Axis II imaging plate area detector system to 2.0 Å and 1.8 Å resolution, respectively. This area detector used CuKα radiation generated from a rotating anode fitted with a monochrometer and operated at 90 mA and 59 kV. For each data collection frame, the crystal was oscillated through an angle of 1.5° and exposed to the X-ray beam for 30-45 minutes. The crystal to detector distance was set to 78 mm. X-ray intensity data was processed to structure factors using the R-AXIS II data processing software (Higashi, 1990; Sato et al., 1992) which is based on techniques described by Rossman et al. (1979). See Chapter 2 for a detailed description of these procedures. Statistics related to data processing are summarized in Table 5.1.

Given the isomorphous unit cells involved, a restrained parameter least squares approach (Hendrickson, 1981; 1985) was employed to refine the azide form of recombinant wild-type myoglobin using the structure of recombinant wild-type horse heart metmyoglobin (presented in Chapter 3) as the starting model. For the azide-His64Thr protein the structure of
Table 5.1: Data collection parameters for the azide forms of recombinant wild-type and His64Thr variant myoglobins.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>azide wild-type</th>
<th>azide His64Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2,</td>
<td>P2,</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>64.3</td>
<td>64.0</td>
</tr>
<tr>
<td>b</td>
<td>28.9</td>
<td>28.8</td>
</tr>
<tr>
<td>c</td>
<td>35.9</td>
<td>35.8</td>
</tr>
<tr>
<td>β</td>
<td>107.2°</td>
<td>107.0°</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>14776</td>
<td>19842</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>5531</td>
<td>8046</td>
</tr>
<tr>
<td>Merging R-factor (%)†</td>
<td>10.2</td>
<td>10.5</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>∞ - 2.0</td>
<td>∞ - 1.8</td>
</tr>
</tbody>
</table>

† \( R_{merge} = \frac{\sum_{\mathbf{h}} \sum_{\mathbf{w}=0}^{\mathbf{w}} |I_{\mathbf{h}} - \bar{I}_{\mathbf{h}}|}{\sum_{\mathbf{h}} \sum_{\mathbf{w}=0}^{\mathbf{w}} I_{\mathbf{h}}^2} \)

His64Thr myoglobin (presented in Chapter 4) was used as the starting model. For both proteins, inspection of initial \( F_o - F_c \) difference electron density maps unambiguously showed the presence of the azide ligands. Thus initial refinements of both structures started with the inclusion of an azide ligand, modelled as a linear triatomic molecule (Dori & Ziola, 1973). Also included in the initial refinement models were well defined water molecules (B < 35 Å²) and a sulfate ion found in the original recombinant wild-type myoglobin structure. Care was taken so that water molecules within 6 Å of the bound azide ligand were excluded from the starting refinement model.
During the course of refinements, omit, $F_o - F_c$ and $2F_o - F_c$ difference electron density maps covering the entire polypeptide chain of both proteins were examined at two different times. This allowed for the manual adjustment of a number of surface main and side chain conformations as well as that of the azide molecule. For both structures, the C-termini region (residues 152-153) was found to be disordered in electron density maps and modelled by finding a geometrically reasonable configuration that maximized the use of the electron density present. Additional water molecules were found by searching $F_o - F_c$ difference electron density maps. These were restricted to those having reasonable hydrogen bonds to protein atoms and thermal factors which refined to less than 60 Å². Refinement was concluded when positional shifts became small (overall r.m.s. shifts < 0.03 Å) indicating that convergence had been reached. Final refinement parameters are tabulated in Table 5.2, along with statistics related to the stereochemistry of the final refinement models.

Atomic coordinate errors for the azide complexed recombinant wild-type and His64Thr myoglobin structures were estimated using two methods. Inspection of a Luzzati (1952) plot suggests overall r.m.s. coordinate errors of 0.18 Å and 0.20 Å for these proteins, respectively. The corresponding estimates obtained using the method of Cruickshank (1949; 1954; 1985) are 0.12 and 0.16 Å.

5.3 Results

As can be seen in Figure 5.1, bound azide is coordinated to the heme iron atom in recombinant wild-type horse heart myoglobin. It is accommodated in the distal heme pocket with its free end pointing in the direction of Leu29 and Ile107. An analysis of solvent accessibility shows this azide ligand is completely buried in the heme crevice. It is also near the side chain of His64 and the NE2
## Table 5.2: Refinement statistics for the azide complexed recombinant wild-type and His64Thr variant proteins.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>azide wild-type</th>
<th>azide His64Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Refinement results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections used</td>
<td>5260</td>
<td>7631</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>6.0-2.0</td>
<td>6.0-1.8</td>
</tr>
<tr>
<td>Completeness in range (%)</td>
<td>60</td>
<td>64</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1242</td>
<td>1239</td>
</tr>
<tr>
<td>No. of ligand atoms</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>59</td>
<td>51</td>
</tr>
<tr>
<td>Average thermal factors (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>21.0</td>
<td>17.4</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>29.8</td>
<td>26.3</td>
</tr>
<tr>
<td>Final refinement R-factor (%)</td>
<td>17.3</td>
<td>17.9</td>
</tr>
<tr>
<td>2. Stereochemistry of models</td>
<td>r.m.s. deviations from ideal values</td>
<td></td>
</tr>
<tr>
<td>Bond distances (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond (1-2)</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>Angle (1-3)</td>
<td>0.040</td>
<td>0.036</td>
</tr>
<tr>
<td>Planar (1-4)</td>
<td>0.056</td>
<td>0.054</td>
</tr>
<tr>
<td>Planes (Å)</td>
<td>0.014</td>
<td>0.012</td>
</tr>
<tr>
<td>Chiral volumes (Å³)</td>
<td>0.056</td>
<td>0.024</td>
</tr>
<tr>
<td>Non-bonded contacts (Å) †</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single torsion</td>
<td>0.189</td>
<td>0.216</td>
</tr>
<tr>
<td>Multiple torsion</td>
<td>0.170</td>
<td>0.199</td>
</tr>
<tr>
<td>Possible hydrogen bonds</td>
<td>0.179</td>
<td>0.184</td>
</tr>
<tr>
<td>Torsion angles (°)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planar (0° or 180°)</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Staggered (±60°, 180°)</td>
<td>22.5</td>
<td>19.3</td>
</tr>
</tbody>
</table>

R-factor = \( \frac{\sum_{i} w |F_{o}|-|F_{c}|}{\sum_{i} w |F_{o}|} \)

† The r.m.s deviations from ideality for this class of restraint incorporate a reduction of 0.2 Å from the radius of each atom involved in a contact.
nitrogen atom of this residue is within hydrogen bonding distance (2.8 Å) of the ligated N1 nitrogen atom of the azide group. The observed heme iron to azide ligation distance is 2.11 Å and the corresponding bonding angle (Fe-N1-N3) is 119°. The azide ligand was found to refine to full occupancy with thermal factors of 22, 23 and 21 Å² for its N1, N2 and N3 atoms, respectively. These values are comparable to the average thermal factor of 21 Å² found for all other protein atoms in the recombinant wild-type myoglobin structure (Table 5.2).

To facilitate comparison of the azide and met recombinant wild-type myoglobin forms, these structures were superimposed using a least squares fit based on all main chain atoms. A plot of the resultant main and side chain positional differences observed is presented in Figure 5.2. These results indicate an overall average deviation of 0.25 Å for all main chain atoms. The largest displacements observed occur at the two polypeptide chain termini (residues 1 and 151-153, Δd = 0.55 Å and 0.60 Å, respectively), Gly15 (Δd = 0.48 Å), Lys50 (Δd = 0.60 Å), Ala57 (Δd = 0.60 Å), His81 to Glu83 (Δd = 0.43 Å) and Gly148 (Δd = 0.51 Å).

Comparison of the residues that make up the heme binding pocket show these are similarly positioned in both the azide and metmyoglobin forms. The only notable exception is the side chain of Leu 29 (Δd = 0.40 Å for side chain atoms) which is in van der Waals contact with the azide ligand (Figure 5.1). However, it is also of interest that the overall positioning of the heme groups in the azide and met forms of recombinant wild-type myoglobin appear to be shifted with respect to one another. If only the positions of heme atoms are superimposed, the overall average positional difference is found to be 0.18 Å. As can be seen in Figure 5.1, these shifts involve a movement of the heme of azide complexed recombinant wild-type myoglobin in the direction of Leu104. The most notable deviations in individual heme atoms involve the placement of heme propionate D. Within this group the largest displacement affects the CBD atom which moves 1.1 Å towards the heme propionate A group.
Figure 5.1. Stereo plots of the heme and surrounding residues in the a) met (thin lines) and azide (thick lines) recombinant wild-type and b) met (thin lines) and azide (thick lines) His64Thr variant myoglobins. In the recombinant wild-type metmyoglobin structure Wat156 (marked by a cross) forms the 6th ligand to the heme iron atom (see details in Chapter 3). As discussed in Chapter 4 the heme iron of metHis64Thr myoglobin does not have a water ligand. The dashed lines indicate hydrogen bond interactions.
Figure 5.2. Plots of the average positional deviations between the main (thick lines) and side (thin lines) chain atoms of a) recombinant wild-type and azide recombinant wild-type and b) His64Thr and azide His64Thr horse heart myoglobins. The horizontal dashed line on each plot represents the overall average positional deviation found between all main chain atoms. The filled circle at residue position 154 represents the overall average deviation found for all 43 heme atoms.
Table 5.3: Details of heme iron coordination geometry for met and azide myoglobin forms.

<table>
<thead>
<tr>
<th>Protein</th>
<th>wild-type met-form</th>
<th>His64Thr met-form</th>
<th>wild-type azide form</th>
<th>His64Thr azide form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe - His93 NE2</td>
<td>2.12</td>
<td>1.93</td>
<td>2.06</td>
<td>2.14</td>
</tr>
<tr>
<td>Fe - Wat156 O</td>
<td>2.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fe - Azide N1</td>
<td>-</td>
<td>-</td>
<td>2.11</td>
<td>2.11</td>
</tr>
<tr>
<td>Fe - Heme NA</td>
<td>2.03</td>
<td>2.04</td>
<td>2.01</td>
<td>2.03</td>
</tr>
<tr>
<td>Fe - Heme NB</td>
<td>2.02</td>
<td>2.01</td>
<td>2.03</td>
<td>1.99</td>
</tr>
<tr>
<td>Fe - Heme NC</td>
<td>2.00</td>
<td>1.99</td>
<td>2.00</td>
<td>2.01</td>
</tr>
<tr>
<td>Fe - Heme ND</td>
<td>1.98</td>
<td>2.00</td>
<td>1.97</td>
<td>1.99</td>
</tr>
<tr>
<td>Fe out of plane</td>
<td>0.09</td>
<td>0.28</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>Angle (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe - ligand</td>
<td>-</td>
<td>-</td>
<td>119</td>
<td>121</td>
</tr>
</tbody>
</table>

‡Distance of the heme iron atom from the mean plane calculated using all atoms of the porphyrin ring.

†Angle between the heme iron azide coordination bond and a corresponding linear ligand direction.

In the His64Thr variant protein bound azide is also coordinated to the heme iron atom (see Figure 5.1) with a similar ligand geometry to that observed for azide complexed recombinant wild-type myoglobin (Table 5.3). However, in the azide His64Thr complex two conformations of the bound azide are observed. The most prominent of these refined to 90% occupancy and thermal factors of 22, 28, and 28 Å² for the N1, N2 and N3 atoms, respectively. Comparison of this major
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

**Figure 5.3.** Stereo diagram of the minor azide conformation observed in the His64Thr variant.

Azide complex form with that of the met His64Thr structure shows that the proximal ligand (His93) to heme iron distance is 0.21 Å longer in the azide form (Table 5.3). This appears to be a direct result of heme iron movement into the heme porphyrin plane upon binding of the azide ligand. In addition, two new water molecules, Wat157 and Wat158 (Figure 5.1), were found in the distal heme pocket and these form a link from the azide ligand to solvent at the surface of myoglobin. Wat157 is closest to the azide molecule and has a refined occupancy of 64% and a thermal factor of 44 Å², whereas Wat158, which is located further away, refines to full occupany with a thermal factor of 30 Å².

In the azide complexed His64Thr variant structure the presence of low level electron density directed from the ligated N1 nitrogen atom of bound azide towards the protein surface in an omit difference electron density map, together with the incomplete occupancies of this group and Wat157, suggested a second minor conformation for bound azide was present. This observation is in agreement with results from electron paramagnetic resonance spectroscopy measurements (R.
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

Bogumil, personal communication). From the perspective of the available space in the distal heme pocket a second azide conformation would require exclusion of the low occupancy water molecule, Wat157. However, refinement of an azide molecule at this alternative position did not prove successful likely as a result of the low occupancy of the N2 and N3 azide atoms in this orientation. This second azide conformation was therefore modelled manually by fitting the available observed electron density and this is illustrated in Figure 5.3.

Based on a least-squares superposition of all main chain atoms, the overall average value for main chain atom positional differences between the structures of the His64Thr variant with and without bound azide is 0.28 Å (Figure 5.2). The two largest regions of positional shifts are at the amino (residue 1, Δd = 0.66 Å) and carboxy (residues 152-153, Δd = 0.80 Å) termini of the polypeptide chain. As can be seen in Figure 5.1, heme group positioning in these structures is comparable with an average positional shift of 0.24 Å for all 43 heme atoms based on a fit of all main chain atoms in the polypeptide chain. This value is only 0.12 Å if only heme atoms are used to obtain an overlap fit. Groups in the heme pocket that undergo substantial positional shifts between the azide bound and unbound forms of His64Thr myoglobin include the side chains of Ser92 and His93, as well as the propionate D group of the heme. In particular, differences in side chain torsional angles are observed in χ_1 for Ser92 (Δ = 22°) and χ_2 of His93 (Δ = 15°), which result in a shortening of the Ser92 OG to His93 NE2 hydrogen bond by 0.45 Å. This is not observed in recombinant wild-type myoglobin where hydrogen bond distances are comparable in the azide bound and unbound forms (Figure 5.1).

As Figure 5.4 shows, part of the flexible CD connecting loop (residues 44-50) and part of the D helix (residues 51-58) have higher observed thermal factors in the azide bound forms of recombinant wild-type and His64Thr myoglobins than in the corresponding met-forms. In contrast, another region in the protein centered around position 80, a part of the EF connecting
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

Figure 5.4. Plots of the average thermal factors of main chain atoms in a) recombinant wild-type metmyoglobin (thin lines) and its corresponding azide form (thick lines) and b) methHis64Thr myoglobin (thin lines) and its azide form (thick lines). In each diagram the upper plot shows the observed differences between the mean main chain temperature factors of the met and azide forms. Other than the relatively disordered N and C-terminal regions, residues 50-56 show a substantial increase in thermal factors between met and azide forms, although this is less pronounced in His64Thr myoglobin. The largest decrease in thermal factor values is observed for residues 79-82 in wild-type myoglobin. For each plot, the thermal factors of the azide complexed form have been normalized so that the overall mean thermal factor values for all main chain atoms of the met and azide myoglobin structures are equal.
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

loop, shows a marked decrease in thermal factors for the azide complexed proteins. All these affected regions are distant from the heme and more than 15 Å from the site of azide binding.

5.4 Discussion

A detailed structural analysis of azide interactions with myoglobin is of interest to understand the balance between the affinity of the heme group of this protein for ligands and sterical hinderance imposed by nearby amino acid groups. Of particular interest is to what extent movement of the heme iron atom is tracked by the proximal histidine residue upon ligand binding and how ligand binding impacts on other heme protein interactions. The structural studies discussed in this chapter have allowed for a comprehensive characterization of azide ligation geometry in horse heart myoglobin. The azide anion is found to be in a bent configuration, with the observed Fe-N1-N3 angle being 119° and 121° for the recombinant wild-type and His64Thr variant (major form) myoglobins, respectively (Figure 5.1 and Table 5.3). For both proteins the free terminal end of bound azide points into the distal heme pocket in the direction of Ile107. For wild-type myoglobin, the proximity of the His64 NE2 atom to the N1 atom of the bound azide (d = 2.8 Å) suggests the presence of ligand stabilization through a hydrogen bonding interaction (Figure 5.1). In contrast to the aqua form of wild-type myoglobin where the water ligand to the heme iron atom is a hydrogen bond donor in its interaction with the side chain of His64, when azide anion is bound the His64 NE2 atom becomes the hydrogen bond donor. In case of azide bound to the His64Thr protein no such direct hydrogen bond to the protein is present and the ligand interacts with a new water molecule found in the distal heme pocket.

It was expected from initial molecular modelling studies, that accommodating an azide molecule would require a number of rearrangements in the residues that make up the distal heme pocket. However, for the azide bound form of recombinant wild-type myoglobin only the side chain of
Leu29 and the heme group show displacements that appear to be a direct response to the spatial and bonding requirements of the azide ligand (Figures 5.1 and 5.2). In contrast, accommodating an azide molecule in the His64Thr variant does not require any substantial rearrangements of residues in the distal heme pocket (Figures 5.1 and 5.2). However, on the proximal side of the heme of this variant protein the side chains of Ser92 and His93, as well as the propionate D group of the heme show significant positional shifts. These movements result in a 0.45 Å decrease in the length of the hydrogen bond from Ser92 OG to the NE2 nitrogen atom of the proximal His93 heme ligand (Figure 5.1).

Although not observed for heme pocket residues, upon azide binding distant portions of the structure of horse heart myoglobin, such as the CD turn, part of the D helix and the EF turn of the protein (see Table 1.1 for helix/loop residue designations), show an increase in mobility in their backbone atoms in both wild-type and His64Thr myoglobins (Figure 5.4). This greater backbone flexibility appears to be a result of accommodating an azide ligand into these proteins, although a mechanism that might explain this observation is not apparent due to the distant location of these polypeptide chain segments from the ligand binding site. This phenomena does not appear to arise from differential crystal packing forces since the crystal forms used are all isomorphous (Tables 3.1, 4.1 and 5.1) and the interactions observed in these regions are identical.

While these studies of the structure of azide complexation with horse heart myoglobin were in progress, two related structures involving different myoglobins were reported. These are the azide-myoglobin complexes of the Aplysia limacina (Mattevi et al., 1991) and sperm whale (Rizzi et al., 1993) myoglobins. The active site of Aplysia myoglobin lacks a distal His64 residue. Here, bound azide is oriented towards the surface exposed portion of the distal heme crevice and is hydrogen bonded to Arg66, which is directed into the heme pocket in the presence of anionic ligands (Qin et al., 1992; Conti, 1993). The bound azide conformation in the sperm whale
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

Figure 5.5. Stereo diagrams of the azide complexes of recombinant wild-type (thin lines) and His64Thr (medium lines) horse heart myoglobins, together with the related wild-type complex in sperm whale myoglobin (thick lines; Protein Data Bank entry: 2MBA).

The conformation of azide binding observed in both horse heart and sperm whale myoglobins appears to represent two aspects of ligand binding. In the first instance an effort is made to maximize the overlap of the heme iron atom d orbitals with the azide nitrogen π orbital. Secondly, the observed azide binding conformation is also a steric solution designed to minimize van der
Waals repulsions with heme group atoms and those of adjacent side chains, particularly Leu29, Phe43, His64, Val68 and Ile107. It is of interest that the observed azide binding angles (117°, 119°) and bonding distances (2.1 Å) in both myoglobins are similar to that observed in azide bound to free heme (Adams et al., 1976). Due to its approximate four-fold nature one might have anticipated azide to bind to free heme in the direction of any of the four methine bridges. In wild-type horse heart and sperm whale myoglobins three of these four potential positional orientations are blocked by Phe43, His64 and Val68 (see Figure 5.5). One might expect that removal of one of these residues would allow for an alternative azide configuration. Indeed, structural studies and electron paramagnetic resonance measurements (R. Bogumil, personal communication) indicate the presence of two azide conformations in the His64Thr variant of horse heart myoglobin (Figure 5.3). In this variant protein, the major form of bound azide (~ 90% occupancy) has this ligand oriented into the heme pocket as in wild-type horse heart myoglobin while in its minor form bound azide is directed towards solvent at the protein surface. In His64Thr myoglobin replacement of His64 by a smaller threonine residue opens up a solvent channel which, in the major conformation of bound azide is occupied by two new water molecules (Figure 5.1; Wat 157 and 158). In its alternative minor conformation bound azide is oriented into this newly created solvent channel and is within hydrogen bonding distance of Wat158 (Figure 5.1). One limiting factor on the occupancy of this minor azide conformation may be steric overlap with the nearby CG2 atom of the substituted Thr64 which is ~ 3.2 Å from the bound ligand.

The EPR spectrum of azide complexed His64Thr horse heart myoglobin shows an equal ratio of two different low spin forms that are also indicative of two different ligand conformations (R. Bogumil, personal communication). Since EPR measurements are performed at low temperature (10 K) it is possible that the relative ratio of the two observed forms is temperature dependent and would be more like that observed in structural studies if conducted at 298 K. Temperature
dependent changes in ligation state have been recently observed in a Val68His variant of horse
heart myoglobin (Lloyd et al., 1995) and in a variant of cytochrome c peroxidase (Ferrer et al.,
1994). The low occupancy of the second azide conformation found in His64Thr variant structural
studies would also agree with a FTIR analysis conducted at room temperature (293 K; R.
Bogumil, personal communication) where the spectrum obtained for the His64Thr - azide complex
could be fitted with only one low-spin band indicating a single bound form for the ligand. If a
substantial amount of a second ligand conformation were present it would be expected to be easily
resolvable given the sensitivity of the FTIR technique (Bogumil et al., 1994).

A comparison of the three dimensional structures of the azide complexes of the His64Thr and
wild-type horse heart myoglobins reveals that the coordination geometry of bound azide (see Table
5.3) and the orientation of the azide anion (major form) in the distal pocket is comparable in both
proteins (Figure 5.5). Despite a similar coordination geometry, significant changes are observed
between the FTIR spectra of the wild-type and His64Thr myoglobin - azide complexes (R.
Bogumil, personal communication). For example, the $v_{\text{max}}$ of the asymmetric azide stretch is at
higher energy in wild-type myoglobin than in the His64Thr variant (2023.7 versus 2018.6 cm$^{-1}$).
In studies of inorganic azide complexes it has been found that the larger the differences between the
two N-N distances of the ligand in azide complexes, the larger the $v_{\text{max}}$ of the asymmetric azide
stretch (Dori & Ziola, 1973). Taking this into account, differences in $v_{\text{max}}$ between wild-type and
His64Thr horse heart myoglobins might be explained by referring to the two resonances structures
depicted in Figure 5.6. In both proteins it is possible that a mixture of these two resonances
structures is present. However, in wild-type myoglobin the dominant form is likely to be that with
the greater negative charge on the N1 atom (form A) since a strong hydrogen bonding interaction
can be formed with the distal His64 side chain. A predominance of this form would also result in
relatively large differences in the two N1-N2 and N2-N3 bonding distances and an increase in the
Figure 5.6. A schematic representation of the two canonical forms of azide bound to the heme iron atom of myoglobin. In both the wild-type and His64Thr myoglobins there is likely a mixture of these two resonance structures. In the wild-type protein the dominant form is expected to be that with the greater negative charge on the N1 atom (form A) due to a strong hydrogen bond interaction formed with His64. In His64Thr myoglobin this hydrogen bond interaction is not formed and therefore form A is probably not as favoured as in wild-type myoglobin leading to a more equal population of forms A and B.

infrared frequency observed. In the His64Thr variant no equivalent hydrogen bonding interaction can be formed by the azide ligand. Thus, in this case form A in Figure 5.6 is probably not as favoured as in wild-type myoglobin and a more distributed population of both forms A and B is likely. In other studies, redistribution of different resonance structures has been successfully used
Chapter 5: Azide as a Probe of Heme Protein Ligand Interactions

to explain changes observed in myoglobin-variants studied by FTIR (myoglobin - CO complexes: Li et al., 1994) and resonance Raman spectroscopy (myoglobin - azide complexes: Biram & Hester, 1994).

Another factor which could influence the observed $v_{\text{max}}$ is the increased polarity of the heme pocket in the His64Thr variant. The existence of enhanced polar interactions is apparent from the presence of additional water molecules in the distal heme cavity (Wat157 and Wat158, Figure 5.1). It has been shown that the electrostatic potential surrounding the bound ligand is an important factor governing $v_{\text{max}}$ in myoglobin - CO and azide complexes (Bogumil et al., 1994; Li et al., 1994).

In addition to a shift of $v_{\text{max}}$, a broadening of infrared bands is observed for azide complexed His64Thr horse heart myoglobin. Increased space in the heme cavity due to Thr64 substitution, combined with the lack of a stabilizing hydrogen bond, would result in more conformational freedom for the bound azide ligand. Reference to crystallographically determined thermal factors in order to obtain an indication of atomic mobility shows similar values for the N1 atom in both the wild-type and His64Thr variant myoglobins. However, increased thermal factors are observed for the N2 and N3 atoms in the His64Thr variant protein (major form) indicating greater positional flexibility at the nonligated end of the azide molecule in this case. The observed broadening of infrared bands can therefore be attributable to several potential factors including: increased mobility of the ligated azide; increased solvent accessibility of the heme pocket; and potentially, a contribution from the second low occupancy azide conformation.

In the vicinity of the ligand binding site the only substantial difference observed is a shorter Ser92 OG to His93 NE2 hydrogen bond in the His64Thr - azide complex. In the wild-type myoglobin - azide complex the partial positive ionic nature of the azide anion is probably stabilized by the $\pi$ electron system of the distal His64. Removal of this distal histidine would shift the partial
positive charge of bound azide anion towards the proximal histidine and thereby may be responsible for the shorter hydrogen bond observed.

In conclusion, the detailed characterization of azide - heme ligation geometry in wild-type horse heart myoglobin establishes that there are no large structural rearrangements required to accommodate formation of this complex. The largest positional shift observed is a movement of the Leu29 side chain away from the ligand. However, substitution of the distal His64 with threonine has two unexpected effects in terms of azide complex formation. First, this substitution allows for two bound azide conformations (major and minor) forms and secondly, on the proximal heme side, a shorter hydrogen bond is observed between Ser92 and the proximal heme iron ligand His93.
Chapter 6

Electrostatic Modification of the Heme Binding Pocket of Myoglobin

6.1 Introduction

Earlier studies of heme proteins have shown that a key factor in specifying the chemistry of heme prosthetic groups is the dielectric constant of the binding pocket in which this group is found (Moore et al., 1986). To address this issue in terms of myoglobin structure-function relationships, a number of substitutions of the heme pocket residues Leu29, Val67, and Leu104 were designed and their three dimensional structures determined. Another factor in the choice of substitutions made was to investigate the possibility of increasing the latent peroxidase activity of myoglobin.

Residue 29 is a leucine in most species of myoglobins (Bashford et al., 1987). This residue is positioned on the distal side of the heme binding pocket and in close proximity to the heme group. The location of residue 29, and in particular its side chain, makes this residue an ideal candidate for replacements that would increase the hydrophilicity of the heme pocket. This could provide insight into heme functional parameters, the ability of the heme pocket to stabilize higher oxidation states of the heme iron atom, and the possibility of altering heme ligand affinity. To explore these issues, preliminary modelling studies suggested the most interesting replacements of Leu29 would involve lysine and tyrosine. Tyrosine was chosen because this replacement was expected to alter heme ligand binding characteristics. The lysine replacement was designed to mimic the functional group of Arg48 in CcP (Figure 1.2) and thereby assist in increasing the latent peroxidase activity of myoglobin. In addition to structural determinations, the Leu29Lys and Leu29Tyr variant proteins were further characterized using electron paramagnetic resonance (EPR) and electronic absorption spectroscopy in order to study the coordination structure of the heme iron over a wider pH range than is possible with X-ray crystallographic techniques.
In horse heart myoglobin, the apolar, but solvent exposed side chain of Val67 is positioned adjacent to the distal heme pocket residue His64. In myoglobins from other sources residue 67 is often found to be threonine. Polar substitutions at position 67 have been shown by other workers to produce large changes in heme azide binding parameters as a result of both steric and electrostatic effects (Bogumil et al., 1994; Brancaccio et al., 1994). In particular, such replacements cause altered heme water coordination interactions and in this way influence the rate of anion movement into the distal heme pocket as well as the stability of the Fe-N3 heme azide bond. Modelling studies suggested the close proximity of Val67 to the heme active site, coupled with a substitution of an arginine at this position, might assist in a peroxidase charge separation step. It was expected that the side chain of this replacement arginine would have sufficient conformational freedom to reach close to the heme iron centre in order to carry out this function. Overall this hypothesis was based on the related Arg48 found in the active site of cytochrome c peroxidase (See Figure 1.2).

In most myoglobins residue 104 is conserved as a leucine and is part of a grouping of hydrophobic amino acids located on the proximal side of the heme pocket adjacent to the heme group. The highly conserved nature of Leu104 suggests an important role in either the functionality or stability of myoglobin. Modelling studies suggest that replacement of Leu104 with an asparagine allows for the placement of a hydrophilic residue at this position with the expectation that this should not lead to major steric conflicts. Thus this replacement seemed a good prospect to probe the nature of heme function by firstly altering the dielectric character of the heme pocket and secondly as a means to investigate the role of Leu104 and its neighbouring cluster of hydrophobic side chains in the folding and stability of the heme pocket.

In summary, this thesis chapter describes the structural characterization of the Leu29Lys, Leu29Tyr, Leu104Asn and Val67Arg variants of myoglobin and the interpretation of these results in conjunction with observed functional parameters for these proteins.
Chapter 6: Electrostatic Modification of the Heme Binding Pocket of Myoglobin

6.2 Experimental Procedures

6.2.1 Site directed mutagenesis, expression and purification of the Leu29Lys variant

The Leu29Lys variant of horse heart myoglobin was obtained by site directed mutagenesis of a synthetic gene and expression of this modified gene in *E. coli* under similar conditions to those used for the wild-type protein (see Chapter 2 for background details). For this, a 600 mL inoculation culture of the transformed cells was used to establish a 10 L culture in superbroth (10 g/L tryptone, 8 g/L yeast extract, 5 g/L NaCl) in a 10 L Chemap fermenter. During growth, temperature was maintained at 37°C, air flow was 7.5 L/min, air saturation was controlled at 5% (v/v) and mixer agitation was 400 rpm. After 18 hours, bacteria (~60 gm) were harvested with a Sharples T-1P tubular bowl centrifuge and frozen at −70°C until they were further processed.

Variant protein was purified (see Chapter 2) using a modified method based on the work of Guillemette et al. (1991). Bacteria were thawed, washed and suspended in 10 volumes of 100 mM NaCl, 1 mM EDTA, 50 mM Tris (pH 8.0), before cell disruption with a CD 30 APV Gualin homogenizer. The cell homogenate was clarified by centrifugation at 8,000 × g for 20 min at 4°C. The supernatant was loaded onto a DEAE-Sepharose column (4.5 x 20 cm) in an equilibration buffer of 50 mM Tris/HCl at pH 8.0. Myoglobin was eluted from this column with the same equilibrium buffer and applied directly to a zinc chelating affinity column (4.5 x 10 cm). The zinc chelating Sepharose (Pharmacia) used was initially charged with 250 mL of 0.1 M ZnSO₄, 25 mM acetic acid, and then equilibrated with 50 mM Tris (pH 8.5) and 1 mM PMSF. Protein bound to this column was eluted with a solution of 50 mM Tris (pH 8.5) and 50 mM imidazole, and was then concentrated and exchanged into a buffer containing 20 mM Tris (pH 8.0) and 1 mM EDTA (pH 8.0) by ultrafiltration (Amicon YM-10 membrane). This concentrate was chromatographed on a G75 Sepharose column (2.5 x 92 cm) in equilibration buffer (50 mM Tris/HCl (pH 8.0)). The yield of recombinant Leu29Lys myoglobin was about 1.0 mg/gm of cells (wet weight).
Chapter 6: Electrostatic Modification of the Heme Binding Pocket of Myoglobin

outlined in Chapter 1, samples of the three other variant proteins studied in this chapter were isolated using the same techniques and obtained from Dr. Chris Overall (Leu29Tyr and Leu104Asn) and Dr. Emma Lloyd (Val67Arg).

6.2.2 Crystallizations and X-ray diffraction data collections

The Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins were crystallized in the oxidized met-form at room temperature (25°C) using the hanging drop vapour diffusion method as described in Chapter 2. The optimal crystal growth conditions for each of these proteins are outlined in Table 6.1. Crystals obtained for the Leu29Lys, Val67Arg and Leu104Asn variant proteins were of the space group P2₁ and isomorphous with those grown for wild-type and recombinant wild-type horse heart metmyoglobins (Evans et al., 1988; see Chapter 3). Crystals formed by the Leu29Tyr variant protein were of quite different morphology and found to belong to the orthorhombic space group P2₁2₁2₁. Space group and unit cell parameters for all crystal forms are tabulated in Table 6.2.

Diffraction data sets from crystals of the variant myoglobins were collected on a Rigaku R-Axis II imaging plate area detector system. This area detector used CuKα radiation generated from a rotating anode fitted with a monochromator and operated at 90 mA and 59 kV. Crystals were oscillated through a φ angle of 1.2° - 1.8° for each data collection frame. In order to maximize the data that could be collected, two separate series of data frame scans were collected for each variant protein crystal. In one series the crystal was mounted vertically on the goniometer head, while in the second series an adapter was used to offset the spindle axis of the capillary holding the crystal by 45°. X-ray intensity data was processed to structure factors using R-Axis II data processing software (Higashi, 1990; Sato et al., 1992) which is based on techniques described by Rossman et al. (1979). See Chapter 2 for a detailed description of these procedures. Statistics related to data processing are summarized in Table 6.2.
Table 6.1: Crystal growth conditions for the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Optimal crystal growth conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu29Lys</td>
<td>20 mM Tris/HCl, 1 mM EDTA, pH 7.5, 15 mg/mL protein, 65% (60%) A.S.†</td>
</tr>
<tr>
<td>Leu29Tyr</td>
<td>20 mM Tris/HCl, 1 mM EDTA, pH 7.0, 8 mg/mL protein, 69% (65%) A.S.†</td>
</tr>
<tr>
<td>Val67Arg</td>
<td>20 mM Tris/HCl, 1 mM EDTA, pH 8.3, 15 mg/mL protein, 65% (60%) A.S.†</td>
</tr>
<tr>
<td>Leu104Asn</td>
<td>20 mM Tris/HCl, 1 mM EDTA, pH 8.4, 8 mg/mL protein, 69% (60-62%) A.S.†</td>
</tr>
</tbody>
</table>

† A.S. refers to the hanging drop reservoir ammonium sulfate concentration in per cent saturation at 20°C (100% = 4.06 M) and the associated value in parenthesis corresponds to the initial ammonium sulfate concentration in the hanging droplet.

6.2.3 Molecular replacement procedures for the Leu29Tyr variant protein

Crystallization of the Leu29Tyr variant protein in an alternative space group required the use of the molecular replacement approach to solve the structure of this protein. The general aspects of the molecular replacement method are outlined in Chapter 2, whereas details pertaining directly to the solution of the Leu29Tyr variant structure are described in this section.

The molecular replacement software used was that of AMoRe (Navaza, 1994) as it is implemented in the CCP4 package (Collaborative Computing Project, 1994). The search model consisted of the structure of recombinant wild-type horse heart metmyoglobin (described in Chapter 3) in which the side chain of residue 29 was modeled as an alanine and all solvent molecules were omitted. This search model was placed in a P1 unit cell with cell parameters of a=b=c=100 Å and α=β=γ=90°. Using calculated structure factors from the search model and normalized observed structure factor amplitudes, a rotation function analysis (25 Å sphere of integration) was calculated at 2° intervals using 8.0 - 3.5 Å data. The highest peak obtained was
9.7 σ above the mean, which was significantly higher than the next peak (4.6 σ above the mean). A translation function search, using data between 8.0 and 3.5 Å resolution was equally clear, with the highest peak being 11.3 σ above the mean. The R-factor calculated for this preliminary rotational and translational orientation of the search model was 26.7 % for data in the resolution range of 8.0 - 3.5 Å.

6.2.4 Structural refinements

A restrained parameter least squares approach (Hendrickson, 1981; 1985) was employed to refine each variant protein structure. For the Leu29Lys, Val67Arg and Leu104Asn proteins the structure of recombinant wild-type horse heart metmyoglobin (see Chapter 3) served as the starting model with substituted residues being truncated to alanines. For the Leu29Tyr variant, the initially oriented search model obtained from the molecular replacement procedure (Section 6.2.3) was used. Examination of $F_0 - F_c$ difference electron density maps clearly indicated the conformations of substituted side chains for the Leu29Lys, Leu29Tyr and Leu104Asn variants. In contrast, the replacement arginine side chain of the Val67Arg variant protein was poorly defined beyond its CG atom.

For all these variants well defined water molecules ($B < 35 \AA^2$) and a sulfate ion found in the original wild-type protein structure were included in initial refinements. However, care was taken to exclude those water molecules falling within 6 Å of substituted residues. Omit, $F_o - F_c$ and $2F_o - F_c$ difference electron density maps covering the entire course of each variant polypeptide chain were examined periodically during refinement. These allowed for a number of manual structural corrections and the placement of substituted side chains. The C-terminal end of the polypeptide chain (residues 152-153) in all variant structures was found to be substantially disordered in electron density maps and modelled by finding a geometrically reasonable
Table 6.2: Data collection parameters for the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Leu29Tyr</th>
<th>Leu29Lys</th>
<th>Val67Arg</th>
<th>Leu104Asn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>31.1</td>
<td>64.0</td>
<td>64.3</td>
<td>64.6</td>
</tr>
<tr>
<td>b</td>
<td>36.2</td>
<td>28.8</td>
<td>30.9</td>
<td>28.9</td>
</tr>
<tr>
<td>c</td>
<td>124.4</td>
<td>35.8</td>
<td>35.1</td>
<td>35.9</td>
</tr>
<tr>
<td>β</td>
<td>α=β=γ=90°</td>
<td>107.0°</td>
<td>105.7°</td>
<td>107.2°</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>18946</td>
<td>78645</td>
<td>36342</td>
<td>47230</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>6742</td>
<td>14228</td>
<td>10598</td>
<td>12079</td>
</tr>
<tr>
<td>Merging R-factor (%)†</td>
<td>7.6</td>
<td>5.4</td>
<td>8.3</td>
<td>8.4</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>∞ - 2.0</td>
<td>∞ - 1.6</td>
<td>∞ - 1.8</td>
<td>∞ - 1.7</td>
</tr>
</tbody>
</table>

† \( R_{merge} = \frac{\sum_{i} \sum_{n \neq 0} |I_{i,n} - \bar{I}_{i}|}{\sum_{i} \sum_{n} I_{i,n}} \)

configuration. Additional water molecules were found by manually searching for peaks in \( F_o - F_c \) difference electron density maps. Potential water molecule positions were only included if found to form reasonable hydrogen bonds to protein atoms and then retained only if refined thermal factors were < 60 Å². Refinements were concluded when positional shifts became small (overall r.m.s. shifts < 0.03 Å) indicating that convergence had been reached. Final refinement parameters for all variant proteins are tabulated in Table 6.3, along with statistics related to the structural stereochemistry.
Atomic coordinate errors for all structures were estimated using two methods. Inspection of a Luzzati (1952) plot indicates overall r.m.s. coordinate errors of 0.19 Å, 0.18 Å, 0.21 Å and 0.20 Å for the Leu29Lys, Leu29Tyr Val67Arg and Leu104Asn variant proteins, respectively. The corresponding estimates obtained using the Cruickshank method (1949; 1954; 1985) are 0.14 Å, 0.16 Å, 0.18 and 0.16 Å.

6.2.5 Functional studies
The detailed methodology used to determine peroxidase activity levels for wild-type and variant myoglobins is discussed in Chapter 2. These assays were carried out in a total volume of 1 mL (pH 6.0, 25 °C) containing 0.1 M MES and 0.2 mM ABTS. Each variant protein was assayed using hydrogen peroxide concentrations of 20, 5, 2, 1.5 and 0.5 mM. Reactions were directly carried out in 1 mL cuvettes and initiated by adding 2 µL of protein solution to prepared reaction mixtures. The concentration of protein in each assay sample was 0.2 µM. The progress of each assay was monitored at 414 nm (Childs & Bardsley, 1975) using a Cary 3 spectrophotometer fitted with a circulating water bath set at 25 °C.

Additional spectroscopic characterization of the Leu29Lys variant protein was done in collaboration with Dr. Ralf Bogumil (Laboratory of Professor Grant Mauk, Department of Biochemistry and Molecular Biology, UBC). General aspects of these methods are outlined in Chapter 2 whereas the details of data acquisition for the Leu29Tyr variant are discussed below.

Electronic absorption spectra were recorded with a Cary 219 spectrophotometer interfaced to a microcomputer (On-Line Instruments-System, Bogart, Georgia) and equipped with a thermostatted circulating water bath (Lauda Model RM6). Protein samples (~5 x 10^{-6} M) were prepared in 0.1 M potassium phosphate (pH 6.5) or 50 mM CHES (pH 10) buffers. For spectrophotometric pH titrations, the Leu29Lys variant protein was exchanged into 0.1 M NaCl and the pH was changed by stepwise addition of 0.1 M NaOH to a 3 ml quartz cuvette. The pK_a for the acid-alkaline...
transition was obtained from absorbance changes at 583 nm. This data was fitted to a single proton titration function (Minsq, version 4.02, MicroMath, Inc).

EPR spectra were obtained at X-band frequencies (ca 9.5 GHz) and 10 K, with a Bruker Model ESP 300E spectrometer equipped with an Oxford Instruments liquid helium cryostat. The magnetic field was calibrated with an NMR gaussmeter (Bruker Model EP 035 M) and the microwave frequency was measured with a frequency counter (Hewlett Packard, Model 5352B). Spectra were obtained under two different pH conditions with protein samples (1 mM) exchanged into 50 mM Hepes (pH 7) or 50 mM glycine (pH 11).

Circular dichroism measurements were performed with a Jasco Model J-720 spectropolarimeter calibrated with ammonium-d-camphor-10-sulfonate (ACS). Spectra (190-250 nm) were recorded with samples placed in a water-jacketed, cylindrical quartz cuvette (path length 0.1 cm). Temperature was regulated with a NESLAB Model RT 110 circulating water bath operated under computer control. The cuvette temperature was measured with a NESLAB RS-2 remote sensor interfaced to a computerized data acquisition system. Protein solutions (~ 1 x 10⁻⁵ M) were prepared in 10 mM potassium phosphate buffer (pH 7.0 or 8.0). Apoprotein was prepared by acidification to pH 2.0, followed by 2-butanone extraction (Teale, 1959). Upon heme removal the apomyoglobin solution was dialysed against 0.6 mM NaHCO₃ containing 1 mM EDTA, and then against 10 mM sodium phosphate buffer (pH 7.0 or 8.0) using a microdialysis setup (Overall, 1987).

Thermal denaturation studies were performed by monitoring elipticity at 222 nm. Temperature was increased at a heating rate of 50 °C/h from 30 to 85 °C. The average of three measurements was used to establish a melting curve for each variant protein. These melting curves were smoothed using a filter function and the first derivative of each calculated to obtain a determination of \( t_m \).
Table 6.3: Refinement statistics for the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Leu29Tyr</th>
<th>Leu29Lys</th>
<th>Val67Arg</th>
<th>Leu104Asn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Refinement results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections</td>
<td>6305</td>
<td>13477</td>
<td>10272</td>
<td>11726</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>6.0-2.0</td>
<td>6.0-1.6</td>
<td>6.0-1.8</td>
<td>6.0-1.7</td>
</tr>
<tr>
<td>Completeness within range (%)</td>
<td>80</td>
<td>62</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1246</td>
<td>1243</td>
<td>1246</td>
<td>1242</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>37</td>
<td>59</td>
<td>54</td>
<td>62</td>
</tr>
<tr>
<td>Average thermal factors (Å²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>19.4</td>
<td>20.6</td>
<td>13.0</td>
<td>17.8</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>29.7</td>
<td>31.6</td>
<td>27.1</td>
<td>30.8</td>
</tr>
<tr>
<td>Final refinement R-factor (%)‡</td>
<td>18.9</td>
<td>17.3</td>
<td>19.7</td>
<td>17.9</td>
</tr>
<tr>
<td><strong>2. Stereochemistry of final models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond distances (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond (1-2)</td>
<td>0.019</td>
<td>0.021</td>
<td>0.019</td>
<td>0.014</td>
</tr>
<tr>
<td>Angle (1-3)</td>
<td>0.041</td>
<td>0.038</td>
<td>0.038</td>
<td>0.037</td>
</tr>
<tr>
<td>Planar (1-4)</td>
<td>0.053</td>
<td>0.055</td>
<td>0.058</td>
<td>0.055</td>
</tr>
<tr>
<td>Planes (Å)</td>
<td>0.013</td>
<td>0.015</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>Chiral volumes (Å³)</td>
<td>0.058</td>
<td>0.032</td>
<td>0.064</td>
<td>0.013</td>
</tr>
<tr>
<td>Non-bonded contacts (Å) †</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single torsion</td>
<td>0.195</td>
<td>0.198</td>
<td>0.202</td>
<td>0.199</td>
</tr>
<tr>
<td>Multiple torsion</td>
<td>0.174</td>
<td>0.157</td>
<td>0.178</td>
<td>0.173</td>
</tr>
<tr>
<td>Possible hydrogen bonds</td>
<td>0.200</td>
<td>0.150</td>
<td>0.155</td>
<td>0.167</td>
</tr>
<tr>
<td>Torsion angles (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planar (0° or 180°)</td>
<td>1.9</td>
<td>2.0</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Staggered (±60°, 180°)</td>
<td>21.3</td>
<td>18.4</td>
<td>19.2</td>
<td>17.1</td>
</tr>
</tbody>
</table>

‡ \( R_{factor} = \frac{\sum_{w} |F_o| - |F_c|}{\sum_{w} |F_o|} \)

† The rms deviations from ideality for this class of restraint incorporate a reduction of 0.2 Å from the radius of each atom involved in a contact.
6.3  Results

6.3.1 Global comparisons of variant and wild-type myoglobins

To facilitate structural comparisons, each of the four variant protein structures determined in this chapter was superimposed onto the polypeptide chain backbone of recombinant wild-type myoglobin using a least squares fit of \( \alpha \)-carbon atoms. In each case, the global fold of the polypeptide chain of these variant proteins was found to be similar to that of the wild-type protein. These results are plotted in Figure 6.1 and illustrated in the stereo drawings of Figure 6.2. As with the wild-type (Evans & Brayer, 1990) and recombinant wild-type (Chapter 3) horse heart myoglobins, the N-terminal and C-terminal ends of the polypeptide chain (residues 1, 152-153) are disordered in the variant proteins. The following sections present detailed analyses of the four variant protein structures.

6.3.2 The Leu29Lys variant structure

With the exclusion of disordered terminal residues, the overall average value for main chain atom positional differences between the recombinant wild-type and Leu29Lys variant structures is 0.11 Å. As Figures 6.1 and 6.2 show, it is clear that substitution of Leu29 for a lysine causes minimal perturbation in the structure of myoglobin. As illustrated in Figure 6.3 the detailed structure in the heme pocket is also conserved in the Leu29Lys variant protein. The most notable change here is the formation of a direct hydrogen bond between the side chain of Lys29 and the distal heme pocket water molecule (Wat156) liganded to the heme iron atom. This new hydrogen bond does not affect heme structure and positioning. The average positional difference for all 43 heme atoms between the Leu29Lys and recombinant wild-type myoglobins is 0.13 Å, with the retention of heme planarity parameters and coordinate bond lengths (Table 6.4).

6.3.3 The Leu29Tyr variant structure

There is the potential for positional shifts observed in the structure of the Leu29Tyr variant protein to arise from either substitution of Leu29 for tyrosine or changes in the way adjacent myoglobin
molecules interact in the crystal lattice of the alternative crystal packing arrangement formed by this variant protein (Table 6.2). Comparison of the Leu29Tyr variant myoglobin structure with its wild-type counterpart indicates an overall average positional deviation of 0.44 Å for all main chain atoms. The largest displacements occur at the two polypeptide chain termini (Gly1, Δd = 2.4 Å; Gln152 and Gly153, Δd = 1.2 Å; Figures 6.1 and 6.2). As discussed earlier, these two regions of polypeptide chain are locally disordered, have high thermal parameters and are distant from the site of substitution of tyrosine at residue 29. It is therefore unlikely these shifts are related to the presence of Tyr29. Other substantial structural changes in main chain conformation occur at residues 48-51 (Δd = 1.2 Å), 56-57 (Δd = 1.1 Å) and 121-122 (Δd = 1.3 Å). The conformational shifts occurring in these regions can all be related to alterations in lattice interactions between the different crystal packing modes of the Leu29Tyr variant and wild-type myoglobins. In terms of side chain atoms the overall average positional deviation observed is 0.83 Å. In particular, a number of surface exposed side chains showed significant positional deviations, including Ile21, Lys50, Lys56, Asp60, Lys87 and Lys96.

As evident from Figure 6.3, the replacement of the side chain Leu29 by the planar and larger side chain of tyrosine occurs within the heme pocket directly adjacent to the heme porphyrin ring. Omit and Fo - Fc difference electron density maps about residue 29 were well defined and clearly indicated the positioning of the side chain of Tyr29. Within the distal heme pocket a number of nearby side chains have been displaced to accommodate the side chain of Tyr29. These include Phe43 (Δd = 0.91 Å), Lys45 (Δd = 1.3 Å), Phe46 (Δd = 0.85 Å), Leu61 (Δd = 1.6 Å) and His64 (Δd = 0.57 Å). Notably, the conformations of the heme groups in the Leu29Tyr variant and wild-type myoglobins are quite different with the observed average deviation in the positions of all 43 heme atoms being 0.64 Å. In general, the largest shifts observed are a displacement of a portion of the heme towards the distal side of the heme pocket and a rearrangement of the propionate D group (Figure 6.3 and Table 6.4). Also, the heme iron atom is notably drawn out of the porphyrin plane towards His93 in the Leu29Tyr variant, with the distance between this iron atom
Figure 6.1. continued on next page
Figure 6.1. Plots of the average positional deviations between the main (thick lines) and side (thin lines) chain atoms of recombinant wild-type horse heart myoglobin and the a) Leu29Lys, b) Leu29Tyr, c) Val67Arg and d) Leu104Asn variant proteins. The horizontal dashed line drawn on each plot represents the overall average positional deviation found between all main chain atoms with the exclusion of the disordered terminal residues 1 and 152-153. For the plot involving the Val67Arg variant, residues 118-122 were also excluded from this calculation. In each diagram the filled circle at residue number 154 represents the average positional deviation of the atoms of the heme group.
Figure 6.2. Stereo diagrams of the α-carbon backbones of a) Leu29Lys (medium lines) and Leu104Asn (thick lines), b) Leu29Tyr and c) Val67Arg variant myoglobins. The α-carbon backbone of recombinant wild-type horse heart myoglobin is shown in thin lines in all three drawings. Also drawn are the sites of amino acid substitutions (designated with three letter amino acid codes), the heme group, the proximal His93 and distal His64 residues, and for clarity the location of approximately every 20th amino acid is labelled with its residue number and one letter amino acid designation.
Figure 6.3. Stereo drawings showing the region about the heme groups in the recombinant wild-type (thin lines) and a) Leu29Lys (thick lines) and b) Leu29Tyr (thick lines) variant myoglobins. Dashed lines are used to indicate hydrogen bond interactions between His64, residue 29, heme propionate D and the water ligand (Wat156). Note that Wat156 is displaced in the Leu29Tyr variant protein.
Table 6.4: Heme geometry of the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant myoglobins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Leu29Lys</th>
<th>Leu29Tyr</th>
<th>Val67Arg</th>
<th>Leu104Asn</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Angular deviations (°) of pyrrole ring plane normals from the porphyrin ring plane normal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4.8</td>
<td>6.1</td>
<td>5.4</td>
<td>3.3</td>
</tr>
<tr>
<td>B</td>
<td>4.2</td>
<td>3.9</td>
<td>2.9</td>
<td>4.5</td>
</tr>
<tr>
<td>C</td>
<td>2.2</td>
<td>0.7</td>
<td>3.9</td>
<td>5.7</td>
</tr>
<tr>
<td>D</td>
<td>1.6</td>
<td>2.1</td>
<td>5.1</td>
<td>2.0</td>
</tr>
<tr>
<td>II. Angular deviations (°) of pyrrole ring plane normals from the pyrrole nitrogen plane normal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>7.4</td>
<td>9.6</td>
<td>8.8</td>
<td>5.4</td>
</tr>
<tr>
<td>B</td>
<td>5.0</td>
<td>8.7</td>
<td>5.3</td>
<td>1.9</td>
</tr>
<tr>
<td>C</td>
<td>4.6</td>
<td>3.3</td>
<td>5.7</td>
<td>8.6</td>
</tr>
<tr>
<td>D</td>
<td>1.7</td>
<td>3.0</td>
<td>3.6</td>
<td>3.1</td>
</tr>
<tr>
<td>III. Heme iron displacements from the pyrrole nitrogen and porphyrin ring planes (Å).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyrrole N</td>
<td>+0.02</td>
<td>-0.23</td>
<td>+0.03</td>
<td>+0.06</td>
</tr>
<tr>
<td>porphyrin</td>
<td>+0.07</td>
<td>-0.34</td>
<td>+0.06</td>
<td>+0.11</td>
</tr>
<tr>
<td>IV. Heme iron ligand bond distances (Å).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His93 NE2</td>
<td>2.19</td>
<td>2.01</td>
<td>2.14</td>
<td>2.14</td>
</tr>
<tr>
<td>Wat156 O</td>
<td>2.18</td>
<td></td>
<td>2.17</td>
<td>2.29</td>
</tr>
<tr>
<td>Heme NA</td>
<td>2.03</td>
<td>1.98</td>
<td>1.98</td>
<td>2.00</td>
</tr>
<tr>
<td>Heme NB</td>
<td>1.98</td>
<td>2.01</td>
<td>2.00</td>
<td>1.99</td>
</tr>
<tr>
<td>Heme NC</td>
<td>2.00</td>
<td>1.97</td>
<td>2.01</td>
<td>2.02</td>
</tr>
<tr>
<td>Heme ND</td>
<td>2.05</td>
<td>2.03</td>
<td>1.96</td>
<td>2.01</td>
</tr>
</tbody>
</table>

The pyrrole nitrogen plane is defined by the four pyrrole nitrogens of the heme group. The four pyrrole ring planes are each defined by the five atoms of the ring and the first carbon atom attached to each of the four carbons of the ring. The porphyrin ring is defined by the five atoms in each of the four pyrrole rings, the four bridging methine carbon atoms, the first carbon atom of each of the eight side chains of the heme and the central iron atom of the heme (33 atoms in total). The heme atom nomenclature used in this table follows the conventions of the Protein Data Bank (Bernstein, 1977) and is diagrammatically illustrated in Figure 1.2 of this thesis. The direction of heme iron displacements to the distal or proximal side of the heme group are indicated by + and - signs, respectively.
and the porphyrin ring plane increasing from 0.08 Å in recombinant wild-type myoglobin to 0.34 Å (Table 6.4). In addition, a shorter His93-Fe ligand bond is observed. Despite these shifts, the planarity of the heme group in the Leu29Tyr variant protein is comparable to that of wild-type myoglobin (see Tables 3.3 and 6.4). As can be seen in Figure 6.3 the distal heme ligation site is not occupied in the Leu29Tyr variant protein making the heme iron atom five coordinate. This is likely due to limited space for a water molecule to ligate to the heme iron atom in the Leu29Tyr variant protein. The tyrosine side chain is in hydrogen bonding distance (d = 3.0 Å) to His64 which in turn interacts (d = 3.3 Å) with the heme D propionate group (Figure 6.3).

### 6.3.4 The Val67Arg variant structure

An overall average displacement of main chain atoms of 0.38 Å is observed for the Val67Arg variant protein when compared to wild-type myoglobin (Figures 6.1 and 6.2). This comparison excludes disordered terminal residues (Gly1, Gln152 and Gly153) and the region about residues 118-122 which shows a large positional shift. As discussed earlier, the large main chain positional differences at both termini are associated with locally disordered residues. Structural changes observed at residues 118-122 (Δd = 1.9 Å) would appear to be related to an alternative set of crystal lattice interactions which are reflected in adjusted unit cell dimensions, particularly along the b axis. This is likely due to the surface location of the charged side chain of Arg67 which is in the vicinity of packing interactions involving residues 118-122. For a number of other surface exposed side chains significant positional deviations are observed and with one notable exception, these also exhibit high thermal factors. These include Lys16, Glu59, Val68, Gln91, Lys96, Pro120 and Asp122. The exception is Val68, which is adjacent to the site of the Val67Arg substitution to which the displacement of Val68 appears to be directly linked.

The side chain of Arg67 was found to be poorly defined beyond its CG atom in electron density maps. Despite this, the refined thermal factor values observed for this side chain are
reasonable and in the range of 30-35 Å². No other alternative positioning of this side chain could be detected in electron density maps prepared at the end of refinement.

As illustrated in Figure 6.4 the structure of the heme pocket in wild-type myoglobin is conserved in the Val67Arg variant protein. The Val67Arg substitution does introduce an ionizable group at the solvent surface entrance of the distal portion of the heme pocket. With this, the most notable feature is a newly created hydrogen bond network between the side chains of Lys45, Asp60, Wat161, Arg67, His64 and Wat156 (see Figure 6.4). To participate in this interaction the side chain of Lys45 (Δd = 1.0 Å) and Asp60 (Δd = 1.0 Å) shift positions (see Figure 6.4). As a result, the hydrogen bond interaction between the heme propionate D group and Lys45 that is present in wild-type myoglobin, is not observed in the Val67Arg variant.

Based on a least squares superposition of all 43 heme atoms, the overall average observed positional difference in atomic positions between the heme structures of Val67Arg and recombinant wild-type myoglobins is 0.17 Å. Furthermore, there is close similarity in overall positioning of the heme group within its binding pocket in both proteins as the overall average value for heme atom differences is only 0.20 Å based on a superimposition of only their polypeptide chains. Heme planarity parameters and coordinate bond lengths of the Val67Arg variant are also comparable to recombinant wild-type myoglobin (Tables 3.3 and 6.4).

6.3.5 The Leu104Asn variant structure

Comparison of the recombinant wild-type and Leu104Asn variant myoglobins shows an overall average deviation of 0.14 Å for all main chain atoms, excluding the disordered terminal regions where the largest displacements occur (Gly1; Δd = 0.32 Å and Gln152-Gly153; Δd = 0.64 Å). These two terminal regions are distant from the site of amino acid substitution at residue 104. As evident from Figure 6.4, replacement of the hydrophobic side chain of Leu104 by the polar side chain of asparagine occurs on the proximal side of the heme pocket and directly adjacent to the
heme porphyrin ring. Omit and $F_o - F_c$ difference electron density maps about residue 104 were of exceptionally high quality allowing for assignments of the positions of the OD1 and ND2 atoms of Asn104 with a high degree of confidence. It is notable that there are no hydrogen bond partners present within the heme pocket adjacent to the Asn104 side chain. Further analysis shows the region around Asn104 contains a small cavity (16 Å$^3$) positioned between this residue and the side chain of His93. A comparable cavity is present in wild-type myoglobin (15 Å$^3$) at a similar
position. However, cavities of such small volume do not normally allow for the presence of a water molecule and none is observed in the electron density maps of either myoglobin.

Based on a least squares superpositioning of all protein atoms, the conformations of the heme groups in the recombinant wild-type and Leu104Asn variant proteins are comparable, with the observed average deviation in the positions of all 43 heme atoms being 0.16 Å (Figure 6.2). Also similar are the placement of heme ligands, propionates and the heme iron atom (Figure 6.4 and Table 6.4). Although the structural perturbations involved in the substitution of asparagine for leucine at position 104 are minimal, some groups do exhibit substantially higher average thermal parameters. These include the heme D propionate group and the side chains of Phe138, Ile142 and Tyr146. These latter three residues all form part of the H-helix and are in van der Waals contact with Asn104. The thermal parameters of the side chain atoms of Asn104 also increases correspondingly when compared to these of the normally present leucine residue.

6.3.6 Peroxidase activity assays
Peroxidase activity measurements for the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant proteins were obtained by using the slopes of initial rate measurements and double reciprical plots as described in detail in Chapter 2. The $k_1$ values obtained for the Leu29Lys, Leu29Tyr, Val67Arg and Leu104Asn variant proteins were 19.4, 19.7, 87.7 and 42.2 mM$^{-1}$ min$^{-1}$, respectively. This compares to a value of 13 mM$^{-1}$ min$^{-1}$ obtained for recombinant wild-type horse heart myoglobin (Chapter 3). The Leu29Lys variant protein showed non-linear behaviour when measured for peroxidase activity in reactions containing 5 mM and 20 mM H$_2$O$_2$. Under these conditions peroxidase activity substantially decreased after 15-20 seconds. After approximately 1 minute, all peroxidase activity ceased for the reaction mixture containing 20 mM H$_2$O$_2$. 

128
6.3.7 Spectroscopic studies of the Leu29Lys variant protein

For the Leu29Lys variant of myoglobin the observed maxima in electronic absorption spectra are comparable to those of wild-type myoglobin (Scheler et al., 1957; Antonini, 1965). However, at pH 6.0 the Soret band of the ferric form of the Leu29Lys variant is slightly red shifted (~ 2.5 nm) compared to that reported for horse heart, sperm whale and human myoglobins, although otherwise this spectrum is similar and characteristic of a six coordinate heme protein. At pH 10.5, the electronic absorption spectra of wild-type metmyoglobin is characteristic for the hydroxo-bound ferriheme derivative with the Soret band shifted to lower energy and visible absorption bands at 539 and 585 nm. The pKₐ for this transition in wild-type myoglobin is 8.93 (Antonini & Brunori, 1971). A spectrophotometric pH titration of the Leu29Lys variant indicates little change up to pH 9.0 (Figure 6.5). At higher pH, absorption signals for a low spin form of this protein appear and are similar to the wild-type metmyoglobin hydroxide form. A pKₐ of 10.23 (standard deviation 0.067) was determined for this transition by fitting the absorbance change at 589 nm to a one-proton titration (Figure 6.5 inset).

Low temperature EPR (10 K, pH 7.0) spectra of the Leu29Lys variant (Figure 6.6) and wild-type myoglobins (Bogumil et al., 1995) exhibit comparable signals that are characteristic of a high spin H₂O bound form (g=5.92, 1.99). As Figure 6.6 shows, at pH 11 a mixture of high and low spin spectra is observed for the Leu29Lys variant protein. The low spin signals are characteristic of a metmyoglobin hydroxide form (g=2.58, 2.17, 1.85) and are similar to those observed for recombinant wild-type myoglobin (Bogumil et al., 1995). Note that the EPR data agree with the results from the electronic absorption data as shown in Figure 6.5. Specifically, the higher pKₐ of the alkaline transition of the Leu29Lys protein means there is a reasonable amount of ferric high spin form even at pH 11. Under these conditions the amount of high spin component is much smaller in the wild type protein (R. Bogumil; personal communication) confirming an alkaline transition at a higher pH for the Leu29Lys variant.
Figure 6.5. Plots showing the pH dependence of the electronic absorption spectrum of the Leu29Lys variant protein. In the inset the absorbance values at 589 nm are plotted vs. change in pH. This data was fit to a one proton titration as indicated by the plotted continuous line.

6.3.8 Thermal denaturation studies of the Leu104Asn variant protein

The circular dichroism spectra (190-250 nm) of the recombinant wild-type and Leu104Asn horse heart myoglobins are shown in Figure 6.7, along with the change in ellipticity with increasing temperature. As shown in Figure 6.7, at pH 7 the Leu104Asn variant protein exhibits a significantly lower \( t_m \) value (71.8 ± 1°C) than that of recombinant wild-type myoglobin (81.3 ± 1°C). A similar trend was observed at pH 8.0 (data not shown) where \( t_m \) was measured as
Figure 6.6. X-band EPR spectra of the Leu29Lys variant horse heart myoglobin at a) pH 7 and b) pH 11. Spectra were measured at 4 K. At pH 7 the spectra is characteristic for a high spin H$_2$O bound form (g=5.92, 1.99), while at pH 11 a low spin signal characteristic of a metmyoglobin hydroxide form (g=2.58, 2.17, 1.85) is present.

67.8 ±1°C and 78.2 ± 1°C for the variant and recombinant wild-type proteins, respectively. To distinguish between the effects of heme affinity and apomyoglobin unfolding, thermal stability measurements were also performed on the apo forms of the recombinant wild-type and Leu104Asn myoglobins (data not shown). At pH 7.0, apo-recombinant wild-type myoglobin exhibits a $t_m$ of 60.4 ± 1°C, whereas the Leu104Asn variant has a $t_m$ of 52.0 ± 1°C. At pH 8.0 the measured $t_m$ values for these proteins were 57.9 ± 1°C and 51.3 ± 1°C, respectively.
Figure 6.7. Circular dichroism spectra of the recombinant wild-type (thin line) and Leu104Asn variant (thick line) myoglobins. Sample conditions were 10 mM potassium phosphate buffer, pH 7.0 and 25 °C. The first derivative of the CD melting curve at 222 nm for each protein is shown in the inset plot.

6.4 Discussion

While myoglobin has the same prosthetic group as both cytochrome c peroxidase and manganese peroxidase, as well as proximal and distal heme pocket histidines, it is lacking other key residues in the heme pocket that preclude the expression of high peroxidase activity. The work in this chapter concentrates on an examination of the structural and functional effects of substituting charged or polar residues in the heme binding pocket of myoglobin. This provides the opportunity to move towards the dual goals of gaining further insight into myoglobin structure-function
relationships and the requirements necessary to increase the latent peroxidase activity of this protein.

6.4.1 The effects of substitutions at Leu29

Replacements at this position were designed to evaluate the impact of altered polarity and spatial interactions in the distal heme binding pocket on overall myoglobin structure, heme ligand binding properties and the latent peroxidase activity of this protein. With regards to increasing the peroxidase activity of myoglobin the focus was to introduce amino acids that might assist in the charge separation process of heterolytic peroxide bond cleavage and in the stabilization of higher oxidation states of the heme reactive centre. In particular, for the Leu29Lys variant protein this constituted an attempt to mimic the role of Arg48 in cytochrome c peroxidase (see Chapter 1, section 1.3).

The introduction of a charged lysine into the confined and tightly packed hydrophobic interior of myoglobin would be expected to lead to polypeptide chain structural rearrangements. It is therefore surprising that there are no substantial changes in polypeptide chain or heme conformation in the Leu29Lys variant protein (Figures 6.1, 6.2 and 6.3). This unique feature allows for the direct correlation of observed changes in heme chemistry to the introduction of the long charged lysine side chain without the need to consider the additional effects of structural shifts.

The observed electronic absorption spectra of the ferric and ferrous forms of both the Leu29Lys variant and wild-type myoglobins are characteristic for a six coordinate heme protein (Figure 6.5; Antonini & Brunori, 1971). Nonetheless, the introduction of a polar lysine side chain has altered the heme electronic structure as evident from a small red shift in the Soret band (2.5 nm). It has been proposed that point charges buried in the interior of a protein are able to induce red shifts in the absorption maxima of a protein bound chromophore via electrostatic interactions (Eccles & Honig, 1983). The three-dimensional structure of the Leu29Lys variant protein shows
such an electrostatic interaction does occur between the lysyl side chain and the heme iron centre via the aqua ligand Wat156 (Figure 6.3).

Lys29 also substantially changes the titration behaviour of the Wat156 heme ligand. This is expressed in a significant increase in the pK$_a$ of Wat156 from 8.9 in recombinant wild type myoglobin to 10.2 in the Leu29Lys variant protein (Figure 6.5; Antonini & Brunori, 1971). These results agree well with EPR data (Figure 6.6). For example, due to the high pK of the alkaline transition of the Leu29Lys variant protein there is still a significant amount of ferric high spin form even at pH 11. A close examination of the structure of the distal heme pocket shows that the NZ atom of the substituted lysine residue at position 29 is in an ideal hydrogen bonding conformation with respect to Wat156 (d = 2.8 Å; Figure 6.3). A further hydrogen bond is made to this aqua ligand (d= 2.8 Å) from the NE2 atom of His64. This hydrogen bonding configuration is unique for heme proteins and could explain the substantial increase in heme pK$_a$ value of the aqua ligand. In this model it is assumed that the NZ amino group of Lys29 is predominantly deprotonated under the crystallization conditions used (pH 7.5), having a pK$_a$ some 2-3 units lower than that of a lysine residue in a polar environment with the dielectric constant of water. This might reasonably be expected given the pK$_a$ values of amino acid residues located in the interior of a protein can vary considerably from those normally observed. It has been previously demonstrated that dielectric constant, neighbouring charges, hydrophobic interactions and hydrogen bonds can all exert pronounced effects on the pK$_a$ of an amino acid. A number of examples of lysine residues with pK$_a$ values of around 7 have been found in other proteins (Schmidt & Westheimer, 1971; Chan et al., 1981; Kietz et al., 1991; Kim & Churchich, 1991).

The FTIR spectra of the azide complex formed by the Leu29Lys variant protein is similar to that of the wild-type myoglobin - azide complex in terms $v_{\text{max}}$, half-bandwidth, and high-spin/low-spin equilibrium (R. Bogumil, personal communication). Therefore a similar binding conformation of the azide anion to the heme iron atom is expected with a hydrogen bond being formed between bound azide and the distal His64 residue. Nevertheless, azide affinity is
significantly reduced in the Leu29Lys variant protein and this might be explained by two effects. First, stabilization of the aqua ligand by two hydrogen bonds (one from each of Lys29 and His64) would be expected to reduce the $K_{off}$ rate of Wat156. This would result in a lower azide binding constant due to increased competition between the azide ion and the aqua ligand for the heme ligand binding site. The second and possibly more important effect is the steric hindrance imposed by the large side chain of Lys29 which is positioned within the distal heme pocket. As can be seen in Figure 6.3 there is insufficient space for an azide ion to bind to the heme iron atom without a significant reorientation of the side chain of Lys29. Furthermore, examination of the structure of the distal heme pocket shows that reorientation of Lys29 would result in sterically unfavourable interactions with neighbouring residues. A comparable situation is found in the molluscan myoglobins where His64 is replaced by a valine and an arginine at residue 67 serves as an alternative hydrogen bond donor to bound ligands (Mattevi et al., 1991). In recent $^1$H-NMR and EPR studies of met-azido and met-imidazole complexes it was shown that the bulky side chain of Arg67 has steric conflicts with heme iron atom bound azide or imidazole, and that these ligands are not as stable as those of vertebrate myoglobins with a distal His64 (Yamamoto et al., 1995).

The peroxidase activity observed for the Leu29Lys variant protein shows a small increase (from 13 to 19 mM$^{-1}$ min$^{-1}$) over that measured for wild-type myoglobin. This increased activity may be correlated to the increased hydrophilic nature of the distal heme pocket in the Leu29Lys variant protein and the additional ability of the introduced lysine side chain to stabilize higher oxidation states of the heme iron atom. However as discussed earlier, the bulky Lys29 side chain likely interferes sterically with ligand exchange as seen in complexes of azide with the heme iron atom and this could be a key factor in limiting peroxidase activity in the Leu29Lys variant protein. Furthermore, as outlined in Chapter 1, heterolysis of the peroxide bond requires both charge separation and a proton (Poulos & Kraut, 1980). It is unlikely that Lys29 can assist in these processes since its side chain would need to be in the ionized form for these roles. As discussed
earlier, spectroscopic and structural results indicate that the side chain of Lys29 is likely to be in the deprotonated state under the conditions studied.

In contrast to the Leu29Lys variant protein, there are substantial structural changes present in the distal heme pocket of the Leu29Tyr variant protein when compared to wild-type myoglobin. In particular, the substitution of a large planar tyrosine side chain requires adjustments to the side chains of residues Phe43, Lys45, Phe46, Leu61, His64 and the heme D propionate group. The structure of the Leu29Tyr variant protein also reveals that the OH1 oxygen atom of Tyr29 is within hydrogen bonding distance (d = 3.0 Å) to the NE2 atom of His64 (Figure 6.3). To optimize this interaction the side chain of His64 shifts towards Tyr29. In addition, structural studies indicate that the heme iron atom is five coordinate. This is likely due to the additional steric constraints the Tyr29 side chain imposes on the distal ligand binding site. For example, the OH1 oxygen atom of Tyr29 is too close to the heme iron atom (4.2 Å) to allow for water ligation.

Studies of the Leu29Tyr variant protein using EPR (10 K) suggest the presence of multiple conformational states (R. Bogumil, personal communication). In particular, a mixture of five and six coordinate heme irons (pH 7) is observed suggesting an equilibrium between these two conformations at low temperature. The five coordinate heme species would seem to be the same as that observed in the x-ray diffraction structure at room temperature (298 K), whereas the other six coordinate heme species appears to be an aqua bound form. Since both, the six-coordinate aqua-bound form of Leu29Tyr metmyoglobin and its alternate five coordinate species would be expected to show acid-alkaline transitions one should be able to observe four different species at high pH. These include the six coordinate aqua-bound form (high spin), the five coordinate species (high spin), the deprotonated six-coordinate aqua-bound form (low spin), and the species resulting from deprotonation of the proximal heme ligand His93 which occurs with a pK_a of about 9.3-9.5 (high spin; Bogumil et al, 1995). Indeed, all four of these species can be identified by EPR spectroscopy. Two of the three high spin species observed at pH 11 are similar to two high spin species observed at pH 7.0 whereas the third likely originates from deprotonation of the proximal
heme ligand His93. The observed low spin species is characteristic of the deprotonated six coordinate aqua-bound form (His-Fe(III)-OH').

The rhombicity of the five coordinate form of the Leu29Tyr variant protein is significantly higher at low pH than five coordinate metmyoglobin variants lacking His64 in the distal heme pocket. A good example of this is the His64Thr variant discussed in Chapter 4. Comparison of the structures of the heme groups of the Leu29Tyr and His64Thr variant proteins reveals similar heme geometry and heme iron displacements (Tables 4.3 and 6.3). Nonetheless, the observed differences in heme electronic structure indicated by the higher degree of rhombicity in the EPR spectra of the Leu29Tyr variant protein would suggest greater heme distortions are present in this protein. However, these distortions are evidently more subtle than can be observed at the current 2.0 Å resolution of the Leu29Tyr variant protein structure determination.

It is believed that the polarity of the heme environment is an important factor in modulating the redox properties of heme proteins (Kassner, 1972; 1973; Churg & Warshel, 1986). In general, a more hydrophilic heme environment is expected to stabilize higher oxidation states of the heme iron atom. The presence of the OH1 hydroxyl group of Tyr29 does increase the hydrophilic nature of the distal heme pocket and as such could potentially lead to an increased level of peroxidase activity. In fact, the peroxidase activity of the Leu29Tyr variant protein (20 mM⁻¹ min⁻¹) is higher than that of wild-type myoglobin (13 mM⁻¹ min⁻¹). This might also be understood through a proposed mechanistic model of the role of Arg48 in cytochrome c peroxidase which involves moving towards the heme iron bound oxene atom and stabilizing it through a hydrogen bond interaction (Miller et al., 1994). Tyr29, although not ideally positioned for this purpose, might stabilize such a reaction intermediate.

6.4.2 The Val67Arg variant structure

Replacement of Val67 by an arginine was designed to evaluate the impact of altering the polarity of the distal heme pocket on heme ligand binding properties and the latent peroxidase activity of this
protein. In particular it was expected that this substitution might mimic the role of Arg48 in cytochrome c peroxidase. As Figure 6.4 shows, the Val67Arg replacement in myoglobin is positioned at the solvent surface entrance of the distal heme pocket and leads to the formation of a hydrogen bond network between Arg67, Lys45, His64, Asp60, the heme D-propionate group, Wat161 and Wat156. From a structural perspective, the Val67Arg substitution does not substantially change folding in the heme binding pocket. Functionally, the heme iron remains capable of binding numerous ligands including CO, CN\(^-\) and N\(_3\)^- in both the ferric and ferrous forms (E. Lloyd, personal communication). The absorption spectra of such derivatives exhibit essentially the same features as the wild type protein (Scheler et al., 1957; Antonini, 1965), with only slight changes in the wavelength maxima being apparent (E. Lloyd, personal communication). The titration behaviour of the distal water ligand also showed similar behaviour (pK\(_a\) = 8.0) to this group in the wild-type protein (pK\(_a\) = 8.93; Antonini & Brunori, 1971). Thus, the results of structural, functional and spectroscopic observations all indicate that the overall integrity of horse heart myoglobin is retained in the presence of the Val67Arg substitution.

From electrostatic considerations, the introduction of a positively charged group such as arginine near the heme would be expected to favour the reduced state of the heme group (Kassner, 1972; 1973; Churg & Warshel, 1986). In agreement with this theoretical consideration, the measured heme reduction potential for the Val67Arg variant protein is 106 mV (E. Lloyd, personal communication), which is higher than that of wild-type myoglobin (45 mV; Lim, 1990).

Despite its higher reduction potential, the Val67Arg variant protein is observed to have a 7.7 times increase in peroxidase activity over that of wild-type myoglobin. As discussed in Chapter 1, peroxidases are found to have low reduction potentials (~ -200 mV). Therefore it is surprising that the Val67Arg variant protein exhibits higher peroxidase activity despite the fact that the higher oxidation states of the heme iron reactive centre are likely to be destabilized relative to those of wild-type myoglobin. Key features likely to be a major factor in determining the reactivity of this variant are the increased hydrophilicity of the distal side of the heme pocket and the presence of the
ionizable arginine side chain which might serve to provide the protons required for peroxide bond cleavage. As outlined in Chapter 1, heterolysis of the peroxide bond requires charge separation and a proton. In cytochrome c peroxidase this mechanistic feature is provided by Arg48 (Poulos & Kraut, 1980; Finzel et al., 1984). There is the potential for the side chain of Arg67 to assist in both of these processes in the Val67Arg modified horse heart myoglobin.

6.4.3 The Leu104Asn variant structure
The Leu104 side chain is in intimate association with the heme group on the proximal side of the heme binding pocket (Figure 6.4). Substitution of leucine by asparagine at this position was designed to evaluate the impact of altered polarity and spatial interactions on myoglobin structure and stability, spectroscopic properties, and the latent peroxidase activity of this protein. Although this substitution does not cause large polypeptide chain or heme group conformational shifts, spectroscopic analyses and other structural parameters indicate that there are significant differences between the Leu104Asn variant and wild-type myoglobins. As shown in Figures 6.3 and 6.8, despite the side chain of Asn104 being of equivalent size to that of the normally present leucine, its planar nature prevents it from packing as efficiently with adjacent side chains. This feature, coupled with the fact that the polar side chain of Asn104 can not form hydrogen bond interactions in the heme pocket, appears to lead to the substantial increases in side chain mobilities observed for Phe138, Ile142 and Tyr146 (Figure 6.8). All three of these residues are at the packing interface between helices G and H and evidently their increased mobility leads to a decrease in the global thermal stability of the Leu104Asn variant as reflected by a $t_m$ $\sim$10 °C lower than that of the recombinant wild-type protein (Figure 6.7). Interestingly, a reverse replacement of Asn52 in yeast cytochrome c with isoleucine, which removes a polar residue from a largely hydrophobic environment, increases the $t_m$ of this protein by $\sim$10 °C (Hickey et al., 1991).

The decreased stability of the Leu104Asn variant can potentially be understood through the principles involved in protein denaturation. In general, the process of protein denaturation can be
regarded as a phase transfer wherein groups from the interior of the protein are brought into contact with water upon unfolding. Comparison of the free energies of amino acid transfers from octanol to water (Fauchere & Pliska, 1983; Eisenberg & McLachlan, 1986) and ethanol to water (Tanford, 1962) place leucine and asparagine on opposite sides of the hydrophobic scale, with leucine being very hydrophobic and asparagine highly polar. Therefore, the Leu104Asn substitution imposes a penalty in free energy for removal of water and burial of the polar asparagine side chain into the highly hydrophobic environment within the core of myoglobin on protein folding. In addition, an asparagine side chain in the unfolded state of the Leu104Asn variant would exist within an aqueous environment with the potential to form up to three hydrogen bonds. The unfulfilled hydrogen bonding potential of asparagine in the folded state (Figure 6.4) would also be expected to contribute to the thermal destabilization of the folded variant protein. Although this is difficult to evaluate quantitatively, an estimate of the free energy for each of these hydrogen bonds is between 8-12 kJ mol$^{-1}$ (Privalov, 1979) and the loss of energy derived from the free energies of transferring
asparagine side chains from water to organic solvents such as ethanol or octanol, is in a similar range.

As can be seen in Figure 6.8, residue 104 forms part of the G-helix in myoglobin and a contributing factor to the decreased thermal stability of the Leu104Asn variant may be related to amino acid preferences for α-helix positions. It is notable that leucine exhibits significant helix forming potential whereas asparagine is a helix breaker (Chou & Fasman, 1974; Richardson, 1988). In addition, Leu104 is located at the fifth position from the N-terminal end of the G-helix (Figure 6.8), a position in α-helices where leucine is preferred (normalized ratio 2.6:1). In contrast, none of the 215 naturally occurring α-helices examined has an asparagine (normalized ratio 0:1) at the fifth N-terminal α-helix position (Richardson, 1988). This may be a reflection of the hydrogen bonding potential of the asparagine side chain which could potentially compete with helix stabilizing hydrogen bonds resulting in helix destabilization. However, as evident from Figure 6.8, steric constraints in myoglobin restrict the side chain of Asn104 from not forming hydrogen bonds with either its own or neighbouring peptide groups. Nonetheless, during initial peptide chain folding of this variant myoglobin hydrogen bonds might occur and modify the stability of folding intermediates.

Another factor that may be related to stability in the Leu104Asn variant protein, is the steric constraints of packing interactions and the fact that internal amino acid side chains are required to interact cooperatively with surrounding groups. In this manner, well ordered internal hydrophobic amino acid residues, such as the highly conserved Leu104 in myoglobin, can make significant contributions to protein thermal stability. Thus, while the entropy of such well ordered groups is reduced in the folded state, a greater fraction of the enthalpy of hydrophobic interaction is expressed in the free energy of stabilization. However, the rigidity of the interactions between buried residues in the folded conformation of a protein result in that part of the molecule being least able to show structural relaxation in compensation for amino acid substitutions. Thus, the change in hydrophobicity resulting from the exchange of Leu104 for asparagine leads to disruption of
cooperative hydrophobic interactions in the hydrophobic core of myoglobin. One direct consequence is less favourable interactions with neighbouring residues and for Asn104 this includes Phe138, Ile142 and Tyr146 whose increased side chain flexibilities may in turn disrupt hydrophobic G-helix to H-helix interactions. As shown in Figure 6.8, Leu104 is an important component of the interface between helices G and H and therefore will also have a direct effect on the stability of this interaction. Indeed, a subdomain of myoglobin consisting of helices A, G and H, which retains its structure in initial unfolding steps, where the remainder of the protein is essentially unfolded, has been demonstrated by hydrogen exchange studies (Hammes & Wu, 1971; Hughson et al., 1990; 1991). Destabilization of this subdomain structure by Asn104 may be another factor in the lower t_m observed for the Leu104Asn variant protein.

An additional factor to consider is that the stability of holomyoglobin is dependent not only on the intrinsic stability of the apoprotein, but also on the strength of interactions between this protein and its heme prosthetic group (Hargrove et al., 1994b). It is possible that a change in this latter interaction may be a contributing factor in any lowered stability of myoglobin. The major forces responsible for retaining the heme in position include the proximal and distal ligands, van der Waals interactions between the heme and residues lining the heme pocket, and hydrogen bonding interactions of the heme propionates with surface amino acid residues (Hargrove, 1994a; 1994b). It has been determined that the rate constant for heme dissociation from the Leu104Asn variant protein is 9 times greater than that for wild-type myoglobin (C.L. Hunter and A.G. Mauk, unpublished data). This is consistent with qualitative predictions of Hargrove et al. (1994b) who suggested the possibility of increased rates of heme dissociation upon increased hydrophilicity of the heme binding pocket. Although it is difficult to discriminate between the relative importance of this effect compared with the altered intrinsic stability of the apoprotein as a contributing factor in the thermal destabilization of the Leu104Asn variant, it highlights the need to test many parameters to identify the source of alterations in protein stability.
In terms of functional properties, substitution of asparagine at residue 104 leads to a 3 fold increase in activity towards hydrogen peroxide when compared to wild-type myoglobin. This increased activity is likely related to the increased polarity of the heme pocket in the Leu104Asn variant protein which could allow for faster binding of hydrogen peroxide molecules to the heme iron atom. Also, while axial ligation of the heme iron atom appears to be the predominant determinant of heme reactivity (Moore & Williams, 1977), the dielectric character of surrounding residues and the degree of solvent exposure also has a strong influence on the chemistry of the heme group (Moore et al., 1986). In this regard, the Asn104 side chain is in intimate association with the heme group and potentially assists in stabilizing higher oxidation states of the iron centre.

In summary, substitution of Leu104 for asparagine in the hydrophobic core of horse heart myoglobin has served to illustrate the complex interplay of factors that contribute to the global stability of myoglobin. Although the immediate environment of the hydrophobic interface between helices G and H was perturbed by substituting a nonpolar residue with a polar residue of equivalent size at position 104, there were no major disruptions in the backbone of the protein. The conclusion that replacement of leucine with asparagine at this site destabilizes the protein both through loss of favourable helix-helix packing interactions and destabilization of helix G supports the concept that the protein core in this region is efficiently packed, that this packing results in significant stabilization of native myoglobin, and that packing disruptions at sites in such regions can lead to significant alterations in structural stability even in the absence of structural perturbations.
Chapter 7

Introduction of a Metal Binding Site into Myoglobin

7.1 Introduction

The role of metal ions in biological systems is diverse and includes involvement in catalytic activities, electron transfer, signalling, substrate binding and transport and structural roles in protein domain assembly and stability. Williams (1985) has pointed out that the environment of a metal ion in a protein frequently differs from that found in an aqueous medium. These special protein environments confer unique properties on the metal ion to facilitate its particular catalytic or structural role in the protein. Interactions of proteins with metal ions have been characterized by a number of different approaches. These include methods as diverse as determining the primary sequence distributions of liganding residues (Valee & Auld, 1989; 1990), to three dimensional structural analyses of metal binding sites by x-ray crystallography (Chakrabarti, 1989; 1990; Yamashita et al., 1990). From such studies, the molecular aspects of metal recognition and complexation by proteins is now understood in sufficient detail to allow for experiments to design new metal binding sites into proteins (Arnold & Haymore, 1991; Tainer et al., 1991; Regan, 1993; Matthews, 1995). For example, new metal binding sites have been incorporated into proteins in order to regulate catalytic activity (Corey & Schultz, 1989; Higaki et al., 1990; 1992; Willett et al., 1995), to enhance structural stability (Pantoliano et al., 1988; Kuroki et al., 1989; Braxton & Wells, 1992), and to facilitate protein purifications (Smith et al., 1988; Arnold & Haymore, 1991).

It has been generally observed that naturally occurring metal binding sites possess features that result in affinities for metals much higher than measured for artificially created ones. For example, in many of these cases residues far removed from a natural metal binding site may affect metal
Chapter 7: Design of a Metal Binding Site into Myoglobin

recognition and specificity. Thus, for many natural metal binding sites, beyond the direct metal ligands formed, the entire protein plays a role in metal binding. Overall there appear to be two general categories of natural binding sites. In one, hydrogen bond networks that involve metal ligation are of primary importance, while in the other, bulk electrostatic or through space effects are of greater importance (Thomas et al., 1985; Russell et al., 1987, 1987). Another important observation is that amino acid side chains that are part of metal binding sites are often found hydrogen bonded to adjacent side chains, or carbonyl or amide groups on the protein to achieve optimal positioning. Also, some natural metal binding sites have evolved electrostatic surfaces with high negative potential to enhance their affinity for positively charged metal ions. Another aspect of natural binding sites is that hydrophilic ligands are often found surrounded by a shell of hydrophobic residues (Yamashita et al., 1990). This feature likely serves to restrict the flexibility of the side chains that make up the binding site, as well as enhance electrostatic interactions between groups within it.

The objective of the work described in this chapter was to create a functional manganese binding site into horse heart myoglobin and to examine how such a site influences structure-function relationships in this protein. Also of interest was whether myoglobin could be modified to oxidize Mn(II) to Mn(III) in the presence of hydrogen peroxide and thereby mimic the functionality of the enzyme manganese peroxidase. As discussed in Chapters 3 and 6, myoglobin provides a good starting point for such an approach since it already has some peroxidase activity.

A number of design criteria had to be considered in constructing a manganese binding site on myoglobin in association with corresponding peroxidase activity enhancing substitutions. For example, since manganese ions prefer oxygen ligands, the most appropriate amino acids to provide such functional groups would be aspartate and glutamate. Also, to allow for fast on/off binding of manganese it was important that the manganese binding site designed should be freely accessible to
solvent. Another critical aspect was to ensure sufficiently tight binding of manganese ions and that this occurs close to the heme prothetic group to allow for optimal electron transfer. Furthermore, the manganese binding site had to be constructed in such a manner that metal binding did not interfere with substitutions introduced to promote the latent peroxidase activity of myoglobin. It is of interest that after completion of this work, the structure of a manganese peroxidase became available (Sundaramoorthy et al., 1994). This provided a unique opportunity to compare the manganese binding site designed on horse heart myoglobin with a naturally occurring one and thereby gain further insight into the best approaches to utilize in metal binding site construction.

7.2 Experimental Procedures

7.2.1 Crystallization and diffraction data collection methods

Crystals of the oxidized met-forms of the Lys45Glu and Lys45Glu/Lys63Glu variant myoglobins were grown at 25°C using the hanging drop vapour diffusion method. For the Lys45Glu variant protein each 10 µl hanging droplet (pH 7.3) contained 15 mg/ml protein, 60% saturated ammonium sulfate, 20 mM Tris HCl, 1 mM EDTA, and was suspended over a well (pH 7.3) containing 1 ml of 67% saturated ammonium sulfate, 20 mM Tris HCl and 1 mM EDTA. The composition of the hanging droplets used for the Lys45Glu/Lys63Glu variant protein was similar except for being adjusted to pH 8.3, and the well solution in this case was 65% saturated with ammonium sulfate. For both variant proteins, crystals grew to a maximum size of 0.4 mm x 0.4 mm x 0.2 mm after about 1 month. Lys45Glu variant protein crystals were isomorphous with those grown for wild-type and recombinant wild-type horse heart myoglobin (Evans et al., 1988; 1990; see Chapter 3). In contrast, the morphology of crystals formed by the Lys45Glu/Lys63Glu variant protein was distinctively different and these crystals were found to be of a different space
group. Table 7.1 contains a complete tabulation of space group and unit cell parameters determined for both variant protein crystal forms.

Diffraction data sets were collected on a Rigaku R-Axis II imaging plate area detector system using CuK$_\alpha$ radiation from a Rigaku RU 300 rotating anode generator fitted with a monochromator and operated at 90 mA and 59 kV. The exposure time used per image plate measurement was 15 - 40 minutes depending on crystal size and quality. In order to maximize the amount of data obtained from each variant protein crystal, diffraction intensities were collected in two separate sets. Following the collection of the first data set, where the glass capillary containing the crystal was mounted vertically on a goniometer head, a second set was collected by adjusting the glass capillary to be 45° off vertical. Each of these diffraction data sets were collected over a rotation range of 90° about the spindle axis using an oscillation angle of 1.2° and a crystal to detector distance of 61 mm. X-ray intensity data were processed to structure factors (summarized in Table 7.1) using procedures described by Higashi (1990) and Sato et al. (1992), as implemented in the R-Axis II data processing software. Each complete data set was put on an absolute scale using the method described by Wilson (1942).

7.2.2 Refinement of the Lys45Glu variant protein structure

Refinement of the Lys45Glu variant structure was initiated using a truncated version of the structure of recombinant wild-type horse heart myoglobin (Chapter 3) as a starting model in which the side chain of residue 45 was represented as an alanine. Well defined water molecules bound to the wild-type structure and having isotropic thermal factors < 35 Å$^2$ were included in this starting model, as was the sulfate ion found at the N-terminal end of the E-helix (residues
Table 7.1: Data collection parameters for variant horse heart myoglobins and associated metal complexes.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Lys45Glu</th>
<th>Lys45Glu/ Lys63Glu</th>
<th>Lys45Glu/ Lys63Glu/Cd</th>
<th>Lys45Glu/ Lys63Glu/Mn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>64.2</td>
<td>29.0</td>
<td>28.9</td>
<td>29.3</td>
</tr>
<tr>
<td>b</td>
<td>28.8</td>
<td>35.8</td>
<td>35.6</td>
<td>35.8</td>
</tr>
<tr>
<td>c</td>
<td>35.9</td>
<td>125.8</td>
<td>125.1</td>
<td>125.2</td>
</tr>
<tr>
<td>β</td>
<td>107.1°</td>
<td>α=β=γ=90°</td>
<td>α=β=γ=90°</td>
<td>α=β=γ=90°</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>31298</td>
<td>100676</td>
<td>125272</td>
<td>155914</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>10978</td>
<td>14180</td>
<td>11392</td>
<td>11123</td>
</tr>
<tr>
<td>Merging R-factor (%)†</td>
<td>5.7</td>
<td>9.7</td>
<td>8.3</td>
<td>7.3</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>∞ - 1.7</td>
<td>∞ - 1.6</td>
<td>∞ - 1.7</td>
<td>∞ - 1.8</td>
</tr>
</tbody>
</table>

† \( R_{merge} = \frac{\sum_{i=0}^{n} \sum_{d=0}^{n} |I_{i,d} - \bar{I}_{d}|}{\sum_{i=0}^{n} \sum_{d=0}^{n} I_{i,d}} \)

58-79). Note that the starting variant protein model also excluded all water molecules associated with the wild-type enzyme that were within 8 Å of residue 45.

The Lys45Glu starting model was refined using a restrained parameter least squares approach as implemented in the program PROLSQ (Hendrickson, 1985). Following 10 cycles of refinement a \( F_o - F_c \) difference electron density map failed to indicate a clear position for the side chain of
Glu45 which was introduced into the refinement model at this point based on a best fit to the observed electron density. During further PROLSQ refinement, omit, $F_o - F_c$ and $2F_o - F_c$ difference electron density maps covering the course of the entire polypeptide chain of the refinement model were examined periodically. In this process, the C-terminal end of the polypeptide chain (residues 152-153) was found to be substantially disordered in electron density maps and modelled by finding a geometrically reasonable configuration. In conjunction with further refinement, water molecules were identified by a manual search of $F_o - F_c$ difference electron density maps. Water molecules were confirmed with omit and $2F_o - F_c$ difference electron density maps, and retained in the refinement model only if participating in reasonable hydrogen bond interactions with protein atoms and found to refine with thermal factors of $< 60 \, \AA^2$. All water molecules were refined as neutral oxygen atoms with full occupancy. Refinement was continued until the average shifts in atomic positions became small (r.m.s. $< 0.03 \, \AA$), indicating that this procedure had converged. Final refinement statistics for the Lys45Glu variant structure are presented in Table 7.2.

Atomic coordinate error for the refined Lys45Glu variant structure was estimated using two methods. Inspection of a Luzzati (1952) plot indicates an overall r.m.s. coordinate error of 0.17 \, \AA, while the corresponding estimate using the method of Cruickshank (1949, 1954, 1985) is 0.14 \, \AA.

### 7.2.3 Determination of the Lys45Glu/Lys63Glu variant structure

Crystallization of the Lys45Glu/Lys63Glu variant of myoglobin in an alternative space group necessitated a molecular replacement approach for the structure solution of this protein. In this process XPLOR software (Brunger, 1990) was used and the search model consisted of the
Table 7.2: Refinement statistics for variant horse heart myoglobins and associated metal complexes.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Lys45Glu</th>
<th>Lys45Glu/ Lys63Glu</th>
<th>Lys45Glu/ Lys63Glu/Cd</th>
<th>Lys45Glu/ Lys63Glu/Mn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Refinement results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections</td>
<td>9812</td>
<td>13801</td>
<td>11136</td>
<td>10698</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>6.0-1.7</td>
<td>6.0-1.6</td>
<td>6.0-1.7</td>
<td>6.0-1.8</td>
</tr>
<tr>
<td>Completeness within range (%)</td>
<td>77</td>
<td>79</td>
<td>77</td>
<td>87</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1242</td>
<td>1242</td>
<td>1242</td>
<td>1242</td>
</tr>
<tr>
<td>No. of solvent molecules</td>
<td>68</td>
<td>43</td>
<td>55</td>
<td>49</td>
</tr>
<tr>
<td>Average thermal factors (Å²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>17.8</td>
<td>18.0</td>
<td>19.3</td>
<td>18.3</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>31.9</td>
<td>28.0</td>
<td>31.3</td>
<td>30.8</td>
</tr>
<tr>
<td>Final refinement R-factor (%)‡</td>
<td>16.5</td>
<td>18.1</td>
<td>17.5</td>
<td>19.9</td>
</tr>
<tr>
<td>2. Stereochemistry of final models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond (1-2)</td>
<td>0.021</td>
<td>0.021</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td>Angle (1-3)</td>
<td>0.039</td>
<td>0.041</td>
<td>0.041</td>
<td>0.038</td>
</tr>
<tr>
<td>Planar (1-4)</td>
<td>0.056</td>
<td>0.057</td>
<td>0.055</td>
<td>0.054</td>
</tr>
<tr>
<td>Planes (Å)</td>
<td>0.016</td>
<td>0.014</td>
<td>0.015</td>
<td>0.013</td>
</tr>
<tr>
<td>Chiral volumes (Å³)‡</td>
<td>0.047</td>
<td>0.035</td>
<td>0.042</td>
<td>0.046</td>
</tr>
<tr>
<td>Non-bonded contacts (Å)‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single torsion</td>
<td>0.202</td>
<td>0.195</td>
<td>0.194</td>
<td>0.194</td>
</tr>
<tr>
<td>Multiple torsion</td>
<td>0.160</td>
<td>0.162</td>
<td>0.177</td>
<td>0.164</td>
</tr>
<tr>
<td>Possible hydrogen bonds</td>
<td>0.159</td>
<td>0.144</td>
<td>0.153</td>
<td>0.166</td>
</tr>
<tr>
<td>Torsion angles (°)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planar (0° or 180°)</td>
<td>1.9</td>
<td>1.9</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Staggered (±60°, 180°)</td>
<td>20.0</td>
<td>19.2</td>
<td>20.0</td>
<td>19.2</td>
</tr>
</tbody>
</table>

‡ R-factor = \( \frac{\sum w_i |F_o| - |F_i|}{\sum w_i |F_o|} \)

† The rms deviations from ideality for this class of restraint incorporate a reduction of 0.2 Å from the radius of each atom involved in a contact.
Chapter 7: Design of a Metal Binding Site into Myoglobin

structure of recombinant wild-type horse heart myoglobin (see Chapter 3) having the side chains of residues 45 and 63 truncated to that of alanine. A rotation function search was initiated by placing the starting model in a cube having dimensions (100 Å) which were more than twice the longest dimension of the search model (~ 45 Å). A Patterson map representing the search model was calculated using a resolution range of 15 - 3.0 Å and a grid size of 0.25 Å. A fast rotation function was evaluated at 2° intervals (25 Å sphere of integration), over a region of rotation space composed of $0^\circ \leq \theta_\perp \leq 180^\circ$, $0^\circ \leq \theta_2 \leq 90^\circ$ and $0^\circ \leq \theta_\parallel \leq 720^\circ$ (Rao et al., 1980) using normalized diffraction data between 10 Å and 3.5 Å. Overall the rotation function search yielded one solution that was significantly better than others having a correlation coefficient of 0.18 (3.4 σ). The next highest peak had a correlation coefficient of 0.10 (1.8 σ).

The most significant peak found in the rotation function search was used as a basis for a translation function search. The strongest peak in the translation function search, calculated with data between 10.0 and 3.5 Å resolution, had a correlation coefficient of 0.34 which was significantly better than the second highest peak of 0.16 and therefore a positive indication of the appropriate translational location of the molecule. The starting R-factor for the search model oriented with the best rotational and translational parameters was 43% in the 8.0 - 3.0 Å resolution range. Additional refinement of overall structural positioning was undertaken with a rigid body approach (Brunger, 1990) using data between 8.0 and 3.0 Å resolution. After 40 cycles of refinement, the fit of positional parameters had substantially improved, resulting in an R-factor of 27% in this resolution range.

Further refinement of the Lys45Glu/Lys63Glu structure used a restrained parameter least squares approach (Hendrickson, 1981; 1985), with both Lys45 and Lys63 being initially modelled as alanine residues. The methods used in refinement for this variant protein structure were similar to those described for the Lys45Glu variant protein in Section 7.2.2, with the exception that the
Chapter 7: Design of a Metal Binding Site into Myoglobin

sulfate ion found in wild-type myoglobin was excluded from the refinement model. Omit difference electron density maps failed to indicate a clear position for the side chain of Glu45 which was therefore modelled by a best fit to the observed electron density. Final refinement statistics are summarized in Table 7.2. The overall r.m.s. atomic coordinate error for the refined Lys45Glu/Lys63Glu variant structure was estimated to be 0.18 Å by inspection of a Luzzati (1952) plot. The corresponding estimate using the method of Cruickshank (1949, 1954, 1985) was 0.16 Å.

7.2.4 Structural characterization of the complexes of Mn(II) and Cd(II) with the Lys45Glu/Lys63Glu variant myoglobin

Soaking experiments to produce Mn(II) and Cd(II) complexes with the Lys45Glu/Lys63Glu variant of horse heart myoglobin were carried out by adding saturating quantities of MnSO$_4$ and CdSO$_4$ into 10 μl hanging drops containing crystals of this protein. Diffraction data collection and processing then proceeded using the methods described in Section 7.2.1. Restrained parameter least squares refinement (Hendrickson, 1981; 1985) of these complexes was initiated using the previously refined Lys45Glu/Lys63Glu variant myoglobin structure (minus solvent molecules) as the starting model. For the Cd(II) complexed Lys45Glu/Lys63Glu variant structure, F$_o$ - F$_c$ maps clearly indicated the presence of two bound Cd(II) atoms. The strongest electron density was observed for Cd(II) binding near Glu45 and the heme D-propionate group. The next highest electron density peak corresponded to Cd(II) binding between the side chains of His119 and Asp122.

For the Mn(II) complexed Lys45Glu/Lys63Glu variant structure, F$_o$ - F$_c$ maps indicated only one electron density peak that could be attributed to a Mn(II) atom. The bound position of Mn(II) corresponded to the site observed for Cd(II) binding involving Glu45 and the heme D-propionate
Chapter 7: Design of a Metal Binding Site into Myoglobin

The detailed methodology used to determine the peroxidase activities of the Lys45Glu and Lys45Glu/Lys63Glu variant horse heart myoglobins is outlined in Chapter 2. In short, these assays were carried out in a total volume of 1 mL (pH 6.0) containing 0.1 M MES and 0.2 mM ABTS. Each variant protein was assayed at a number of hydrogen peroxide concentrations (20, 5, 2, 1.5, 1.0 and 0.5 mM). Assays were carried out in 1 mL cuvettes and initiated by adding 2 μL of protein solution to prepared reaction mixtures. The concentration of protein in each assay sample was 0.2 μM. The progress of each assay was monitored at 414 nm (Childs & Bardsley, 1975) using a Cary 3 spectrophotometer fitted with a circulating water bath set to 25 °C.

7.3 Results
7.3.1 Structural analyses of the Lys45Glu and Lys45Glu/Lys63Glu variant proteins

To facilitate structural comparisons, each of these variant protein structures was superimposed onto the polypeptide chain backbone of recombinant wild-type horse heart myoglobin using a least squares fit of α-carbon atoms. As shown in Figure 7.1, the global fold of the polypeptide chains of the Lys45Glu and Lys45Glu/Lys63Glu variant proteins is similar to that of wild-type myoglobin.
Plots of the average positional deviations of residues along the polypeptide chain are shown in Figure 7.2.

Comparison of the Lys45Glu variant protein with wild-type myoglobin indicates an overall average positional deviation of 0.11 Å for main chain atoms. This value was calculated with the exclusion of the disordered N-terminal residue Gly1 (Δd = 2.4 Å) and the C-terminal residues Gln152 - Gly153 (Δd = 0.3 Å). The disorder observed at these terminal ends is similar to that found in wild-type recombinant horse heart myoglobin (Chapter 3) and therefore appears unrelated to the amino acid substitution introduced at Lys45. The largest side chain positional differences involve surface exposed residues and include Asp4, Gln9, Lys56, Leu86 and Thr95. Despite the loss of the normally present hydrogen bond interaction between Lys45 and the heme D-propionate, the heme pocket conformations of the Lys45Glu variant and wild-type myoglobins are comparable. This similarity extends to the positioning of the heme group for which an average positional deviation of 0.11 Å is observed for all 43 heme atoms. A full tabulation of heme geometry and coordinate bonds is presented in Table 7.3.

The overall average value for main chain atom differences between the Lys45Glu/Lys63Glu variant and wild-type myoglobin structures is 0.20 Å (excluding residues 1 and 152-153). The largest positional deviations are at the N (residue 1, Δd = 2.5 Å) and C (residue 153, Δd = 1.3 Å) terminal ends, as well as for residues 41-59 (Δd = 0.62 Å; Figure 7.1). As found in recombinant wild-type myoglobin, the large deviations observed for terminal residues reflect their disordered conformations in electron density maps. The conformational difference observed for residues 41-59 represents a displacement of the surface exposed C-D loop region. This is likely in response to a change in crystal lattice interactions since this region is at a crystal packing interface unique to the orthorhombic crystal form of the Lys45Glu/Lys63Glu variant protein. For this variant protein the largest shifts in side chain groups occur for Gln8, Lys50, Lys56, Asp60, Thr70, Lys87,
Figure 7.1. Stereo diagrams of the α-carbon backbones of the a) Lys45Glu (thick lines) and b) Lys45Glu/Lys63Glu (thick lines) variant myoglobins. The α-carbon backbone of recombinant wild-type horse heart myoglobin is also shown with thin lines in both drawings. Also indicated are the sites of amino acid substitutions, the heme group, the proximal His93 and distal His64 histidines, and the location of every 20th amino acid along the polypeptide chain.
Figure 7.2. continued on next page
Figure 7.2. Plots of the average positional deviations between the main (thick lines) and side (thin lines) chain atoms of recombinant wild-type horse heart myoglobin and the a) Lys45Glu and b) Lys45Glu/Lys63Glu variant proteins. Additional plots show the average positional deviations between the main (thick lines) and side (thin lines) chain atoms of the Lys45Glu/Lys63Glu variant horse heart myoglobin and its related c) Cd(II) and d) Mn(II) complexed forms. On each plot a horizontal dashed line indicates the overall average positional deviation found between all main chain atoms, with the exclusion of the disordered terminal residues 1 and 152-153. The filled circle at residue number 154 in each plot represents the average positional deviations of atoms of the heme group.
Chapter 7: Design of a Metal Binding Site into Myoglobin

Gln91 and Asp122, all of which are located in solvent exposed positions. As Table 7.3 shows, heme conformation, ligation and planarity are preserved in the Lys45Glu/Lys63Glu variant horse heart myoglobin. For all heme atoms the overall positional difference is 0.18 Å.

For both of the Lys45Glu and Lys45Glu/Lys63Glu variant proteins the side chain atoms of Glu45 were poorly defined beyond the CB atom in omit and F_o - F_c difference electron density maps. This suggests multiple side chain conformations are present for this solvent exposed surface residue. The positioning of the side chain of Glu45 presented in Figure 7.1 represents the most populated conformation of this residue based on an interpretation of the available electron density. Not surprisingly, this side chain has high average thermal factors for both the Lys45Glu (B = 45 Å^2) and Lys45Glu/Lys63Glu (B = 50 Å^2) variant myoglobins. These values are significantly higher than the thermal factors found for their related main chain atoms (27 and 24 Å^2, respectively) and the overall thermal factors found for all protein atoms in these structures (18 and 19 Å^2, respectively). In contrast, the Glu63 side chain in the Lys45Glu/Lys63Glu variant myoglobin was well defined in electron density maps and refined to a thermal factor of about 16 Å^2.

7.3.2 The Mn(II) and Cd(II) complexes formed with Lys45Glu/Lys63Glu variant protein

Initial difference electron density maps showed strong positive peaks close to the heme D-propionate and Glu45 in both the Cd(II) and Mn(II) bound Lys45Glu/Lys63Glu variant myoglobins. A second strong positive electron density peak was found in the Cd(II) bound form near His119 and Asp122. No Mn(II) binding was observed at this second site. Bound Cd(II) ions were found to refine to nearly full occupancy (site 1 = 90%, site 2 = 93%) with thermal factors of 29 and 52 Å^2 for sites 1 and 2, respectively. A Mn(II) ion refined to full occupancy in site 1 with
Table 7.3. Heme geometry in variant proteins and associated metal complexes.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Lys45Glu</th>
<th>Lys45Glu/ Lys63Glu</th>
<th>Lys45Glu/ Lys63Glu/Cd</th>
<th>Lys45Glu/ Lys63Glu/Mn</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Angular deviations (°) of pyrrole ring plane normals from the porphyrin ring plane normal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4.8</td>
<td>6.5</td>
<td>6.3</td>
<td>7.2</td>
</tr>
<tr>
<td>B</td>
<td>3.2</td>
<td>5.2</td>
<td>6.4</td>
<td>4.7</td>
</tr>
<tr>
<td>C</td>
<td>2.5</td>
<td>3.0</td>
<td>4.1</td>
<td>2.3</td>
</tr>
<tr>
<td>D</td>
<td>3.2</td>
<td>3.6</td>
<td>4.6</td>
<td>4.8</td>
</tr>
<tr>
<td>II. Angular deviations (°) of pyrrole ring plane normals from the pyrrole nitrogen plane normal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>7.4</td>
<td>9.8</td>
<td>8.7</td>
<td>9.4</td>
</tr>
<tr>
<td>B</td>
<td>3.2</td>
<td>5.7</td>
<td>7.9</td>
<td>4.7</td>
</tr>
<tr>
<td>C</td>
<td>4.8</td>
<td>5.0</td>
<td>3.8</td>
<td>2.7</td>
</tr>
<tr>
<td>D</td>
<td>1.1</td>
<td>2.5</td>
<td>2.0</td>
<td>3.7</td>
</tr>
<tr>
<td>III. Heme iron displacements from the pyrrole nitrogen and porphyrin ring planes (Å).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyrrole N</td>
<td>+0.03</td>
<td>+0.05</td>
<td>+0.04</td>
<td>+0.06</td>
</tr>
<tr>
<td>porphyrin</td>
<td>+0.05</td>
<td>+0.08</td>
<td>+0.07</td>
<td>+0.08</td>
</tr>
<tr>
<td>IV. Heme iron ligand bond distances (Å).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His93 NE2</td>
<td>2.13</td>
<td>2.16</td>
<td>2.15</td>
<td>2.23</td>
</tr>
<tr>
<td>Wat156 O</td>
<td>2.16</td>
<td>2.16</td>
<td>2.22</td>
<td>2.22</td>
</tr>
<tr>
<td>Heme NA</td>
<td>2.04</td>
<td>2.01</td>
<td>1.99</td>
<td>2.00</td>
</tr>
<tr>
<td>Heme NB</td>
<td>1.97</td>
<td>1.99</td>
<td>1.99</td>
<td>1.95</td>
</tr>
<tr>
<td>Heme NC</td>
<td>2.03</td>
<td>2.03</td>
<td>2.00</td>
<td>1.98</td>
</tr>
<tr>
<td>Heme ND</td>
<td>2.01</td>
<td>2.02</td>
<td>1.98</td>
<td>2.01</td>
</tr>
</tbody>
</table>

The pyrrole nitrogen plane is defined by the four pyrrole nitrogens of the heme group. The four pyrrole ring planes are each defined by the five atoms of the ring and the first carbon atom attached to each of the four carbons of the ring. The porphyrin ring is defined by the five atoms in each of the four pyrrole rings, the four bridging methine carbon atoms, the first carbon atom of each of the eight side chains of the heme and the central iron atom of the heme (33 atoms in total). The heme atom nomenclature used in this table follows the conventions of the Protein Data Bank (Bernstein, 1977) and is diagrammatically illustrated in Figure 1.3.
a thermal factor of 43 Å².

As shown in Figure 7.3 and summarized in Table 7.4, ligands are formed to the bound Mn(II) ion by the heme D-propionate, the side chain of Glu45, two water molecules and the side chain of His113. This latter residue originates from a symmetry related molecule in the crystalline lattice. When compared to its most populated position in the unbound variant protein form, the side chain of Glu45 is reoriented in the Mn(II) complex (Δd = 1.0 Å; Figure 7.3).

The ligands involved in binding Cd(II) ions at site 1 are the same as for Mn(II) ions (Figure 7.4). In this case Cd(II) binding also leads to reorientation of the side chain of Glu45 (Δd = 1.1 Å; Figure 7.4). In site 2 the bound Cd(II) ion is found ligated to the side chains of His119 and Asp122, as well as to two water molecules (Figure 7.4). In this site a small reorientation of the Asp122 side chain occurs upon binding Cd(II) ion. The ligand bond distances occurring at site 2 are presented in Table 7.4.

The overall polypeptide chain conformations of the free and metal bound forms of the Lys45Glu/Lys63Glu variant protein were found to be similar (Figures 7.2, 7.3 and 7.4; Table 7.3). This is reflected in an overall average value of 0.13 Å observed for main chain atom differences between the Mn(II) and Cd(II) bound and unbound forms of this variant protein. However, a comparison of the bound positions of the Mn(II) and Cd(II) ions shows significant differences. These include the positioning of these ions in site 1 (Δd = 0.84 Å; Figure 7.4) and the conformations of the liganding groups Glu45 (Δd = 2.0 Å) and heme D-propionate (Δd = 0.80 Å). Also considerably different are the positions of liganding water molecules. On the other hand, His113 remains comparably positioned (Δd = 0.20 Å) when either Mn(II) or Cd(II) ions are bound at this site.
Figure 7.3. a) Stereo drawing of the α-carbon backbones of the Mn(II) free (thin lines) and Mn(II) complexed (thick lines) forms of Lys45Glu/Lys63Glu horse heart myoglobin. Also drawn are the heme group, the proximal His93 and distal His64 histidines, the distal ligand Wat156, Glu45 and Glu63, as well as the location of bound Mn(II) ion. For clarity every 20th α-carbon along the polypeptide chain has been labelled. b) Stereo drawing showing a detailed view of the residues surrounding the Mn(II) binding site in the free (thin lines) and complexed (thick lines) forms of Lys45Glu/Lys63Glu horse heart myoglobin. His113 is from a related molecule in the crystalline lattice and ligand water molecules are indicated with crosses. c) Stereo drawing showing the superposition of the heme group and residues surrounding the introduced Mn(II) binding site in the Mn(II) (thick lines) and Cd(II) (thin lines) complexed forms of Lys45Glu/Lys63Glu horse heart myoglobin.
Table 7.4. Ligand distances to bound Cd(II) and Mn(II) ions.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Cd(II)</th>
<th>Mn(II)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduced binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heme O1D</td>
<td>2.7</td>
<td>3.0</td>
</tr>
<tr>
<td>O2D</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Glu45 OE1</td>
<td>2.3</td>
<td>2.5</td>
</tr>
<tr>
<td>OE2</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Water molecule</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Water molecule</td>
<td>3.0</td>
<td>2.9</td>
</tr>
<tr>
<td>His113 NE2 (symmetry related)</td>
<td>2.5</td>
<td>2.3</td>
</tr>
<tr>
<td>II. Natural binding site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp122 OD1</td>
<td>2.5</td>
<td>-</td>
</tr>
<tr>
<td>His119 ND1</td>
<td>3.1</td>
<td>-</td>
</tr>
<tr>
<td>Water molecule</td>
<td>3.3</td>
<td>-</td>
</tr>
<tr>
<td>Water molecule</td>
<td>3.3</td>
<td>-</td>
</tr>
</tbody>
</table>

7.3.3 Peroxidase activity

Measurements of the peroxidase activity of the Lys45Glu and Lys45Glu/Lys63Glu variant myoglobins in the absence of bound metals were conducted as described in Section 7.2.5 and in more detail in Chapter 2. Individual assays at different peroxide concentrations were plotted and rate data derived from the initial linear sections of individual activity lines which corresponded to the first 15 to 30 seconds of the reaction. This initial rate data was then used to generate a double reciprocal plot like the one shown in Figure 2.5. The slope of the line in this plot was taken as the rate constant $k_j$ and the values obtained were 16.5 and 16.1 mM$^{-1}$min$^{-1}$ for the Lys45Glu and
Figure 7.4. a) Stereo drawing of the α-carbon backbones of the Cd(II) free (thin lines) and Cd(II) complexed (thick lines) forms of Lys45Glu/Lys63Glu horse heart myoglobin. Also drawn are the heme group, the proximal His93 and distal His64 histidines, the distal ligand Wat156, Glu45, Glu63, His119 and Asp122, as well as the location of the two bound Cd(II) ions. For clarity every 20th α-carbon along the polypeptide chain has been labelled. b) Stereo drawing showing a detailed view of the residues surrounding the specifically developed metal binding site in the Cd(II) free (thin lines) and complexed (thick lines) forms of Lys45Glu/Lys63Glu horse heart myoglobin. His113 is from a symmetry related molecule in the crystalline lattice and Cd(II) ligand water molecules are indicated with crosses. c) Stereo drawing showing the natural metal binding site in myoglobin in the Cd(II) free (thin lines) and Cd(II) complexed (thick lines) forms of Lys45Glu/Lys63Glu myoglobin.
Chapter 7: Design of a Metal Binding Site into Myoglobin

Lys45Glu/Lys63Glu variant myoglobins, respectively. Peroxidase activity for both variant myoglobins showed linear behaviour for reactions containing 0.5 - 20 mM concentrations of H$_2$O$_2$.

7.4 Discussion

Several design criteria were considered in the development of the metal binding studied in this work: ligand interactions that would bind Mn(II) ions as part of the catalytic cycle; a site that was freely accessible to solvent and sufficiently close to the heme active center to allow for efficient electron transfer; site placement such that metal binding would not adversely affect the accessibility of peroxide to the heme or disrupt the latent peroxidase function of myoglobin; and, minimization of the number of amino acid substitutions required for site construction.

Based on the required design criteria, modelling studies were able to determine a number of potential binding sites. Due to its proximity to the heme group and its location on the protein surface, a binding site involving Lys45 and the heme group was judged to be the most promising. In addition, this site only required the single substitution of glutamate for lysine at residue 45 and this replacement was expected to have little effect on the overall folded structure of myoglobin. As subsequent structural analyses have shown, introduction of the Lys45Glu substitution did not substantially alter the structure of myoglobin or the conformation of its heme prosthetic group (Figures 7.1, 7.2 and Table 7.3). However, while the Lys45Glu variant protein retains a high degree of thermal stability, a small increase in the rate of heme dissociation is observed and may be due to the loss of the stabilizing hydrogen bond interaction normally found between Lys45 and the heme D-propionate group (Hunter et al., 1997). The side chain of Glu45 is found positioned at the solvent surface entrance of the distal heme pocket and takes on multiple conformations.

Subsequent crystal soaking experiments have shown that Mn(II) ion binding at the designed site
Chapter 7: Design of a Metal Binding Site into Myoglobin

is weak and it is not possible to observe bound Mn(II) ions in structural studies due to low site occupancy. As determined by EPR spectroscopy, the equilibrium binding constant for Mn(II) ion binding to the Lys45Glu variant protein is 370 M$^{-1}$. This is an increase over the value of 180 M$^{-1}$ observed for wild-type myoglobin (C.L. Hunter, Ph.D. Thesis). Despite the weak binding observed, NMR spectroscopy experiments have been able to confirm the binding of Mn(II) ions in the designed manganese binding site by the specific broadening of heme D-pyrrole resonances in the presence of Mn(II) ions. Similar NMR methods were employed to identify the site of Mn(II) ion binding near the heme edge in the manganese peroxidase from *Phanerocheate chrysosporium* (Banci et al., 1993).

To increase the binding affinity of the designed myoglobin manganese binding site, a second substitution was introduced in which Lys63 was replaced by a glutamate. Lys63 is on the surface of myoglobin in close proximity to the heme D-propionate group and residue 45. Based on modelling studies it was expected that the side chain of a glutamate at residue 63 would be able to form part of the manganese binding site and in this way increase the affinity of this site for Mn(II) ions. Modelling studies also suggested that since both the Lys45Glu and Lys63Glu substitutions are on the protein surface, that these changes would be unlikely to have a significant effect on the overall fold of myoglobin. Subsequent structural studies have shown the side chains of both Glu45 and Glu63 are located at the solvent surface entrance of the distal side of the heme pocket (Figures 7.1, 7.2 and Table 7.3). The side chain of Glu45 was found to be poorly defined in electron density maps and likely takes on multiple conformations. In contrast, the side chain of Glu63 is well defined in electron density maps.

As crystal soaking experiments have shown, the designed binding site in the Lys45Glu/Lys63Glu variant protein of myoglobin has a much higher affinity for Mn(II) ions and it was therefore possible to characterize binding interactions occurring in this site using x-ray
diffraction techniques. As illustrated in Figure 7.3, the ligands involved in binding a Mn(II) ion are the heme D-propionate group, the side chain of Glu45, two water molecules, and the side chain of His113 from a symmetry related molecule in the crystalline lattice. In solution it would be expected that the His113 ligand would be replaced with a water molecule. Surprisingly, the substituted Glu63 side chain does not form a part of the manganese binding pocket and seems to serve to enhance manganese binding affinity by generally increasing the negative electrostatic potential of this region of the protein surface. EPR measurements indicate a binding constant of $1730 \text{ M}^{-1}$ for the Lys45Glu/Lys63Glu variant protein, confirming that Glu63 considerably enhances Mn(II) ion binding (C.L. Hunter, Ph.D. Thesis). As with the Lys45Glu variant protein, NMR data also indicates Mn(II) ion binding involves the heme D-propionate group.

Although manganese binding has been considerably enhanced in the Lys45Glu/Lys63Glu variant myoglobin, the binding affinity of the related site in the manganese peroxidase from *Phanerochaete chrysosporium* is $\sim 500$ fold greater (Wariishi et al., 1992). In this manganese peroxidase the ligands to a bound Mn(II) ion include two glutamates, an aspartate, the heme A-propionate and two water molecules (Figure 7.5; Sundaramoorthy et al., 1994). The lower affinity of the designed myoglobin binding site seems likely to be a result of the smaller number of negatively charged ligands. This is supported by experiments that show removal of one of the binding site ligands (Asp179) in manganese peroxidase results in a drastic decrease in Mn(II) ion binding affinity (Kusters-van Someron et al., 1995). This suggests the binding affinity of the designed manganese binding site in myoglobin could be substantially enhanced by the addition of further carboxyl group ligands.

Potentiometric titrations of the Lys45Glu/Lys63Glu variant have identified two manganese binding sites on the protein surface. One of these corresponds to the designed site and the other is a much weaker naturally occurring site (C.L. Hunter, Ph.D. Thesis). This second natural site has
Figure 7.5. Stereo drawing of the manganese binding site in the manganese peroxidase isolated from \emph{P. chrysosporium}. In this enzyme the Mn(II) ion is bound to glutamates 35 and 39, Asp179, the heme A-propionate and two water molecules. Also drawn are the proximal His173 and distal His46 residues, and the distal Wat556 heme ligand.

also been reported for sperm whale myoglobin, where Mn(II), Cu(II) and Zn(II) ions were found to bind between His119, Asn122 and Lys16 (Banaszak et al., 1965; Gersond & Netter, 1966). To characterize the corresponding site on horse heart myoglobin, soaking experiments with Cd(II) ions were carried out using Lys45Glu/Lys63Glu variant myoglobin. This choice of ion was made since earlier studies had shown that the binding affinity at the natural site was too weak to allow visualization of a bound Mn(II) ion using diffraction methods. On the other hand, Cd(II) ions are of larger atomic size and have less stringent requirements on ligand geometry, allowing them to be more easily localized in electron density maps. In this way structural studies were able to determine that the natural metal binding site on horse heart myoglobin also involves the side chains of His119 and Asp122 (Figure 7.3). Note that this natural site is located $\sim 35 \, \text{Å}$ from the designed manganese binding site. It is also $\sim 25 \, \text{Å}$ from the heme active center and therefore not likely to have a direct
Chapter 7: Design of a Metal Binding Site into Myoglobin

effect on heme chemistry.

In addition to successfully introducing a Mn(II) ion binding site into myoglobin, the functionality of the Lys45Glu and Lys45Glu/Lys63Glu variant proteins in terms of a manganese peroxidase activity was assessed in the presence of the organic Mn(II) chelator malonate (C.L. Hunter, Ph.D. Thesis). Both variant proteins show approximate 3-fold increases over wild-type myoglobin in the rate of Mn(II) oxidation with hydrogen peroxide. It is likely this enhanced activity depends in large part on the presence of the introduced manganese binding site. Another factor may be the decreased heme reduction potential resulting from the substitutions of Lys45 and Lys63 for glutamates. For the Lys45Glu/Lys63Glu variant protein the observed reduction potential is 0 mV. This is 52 mV lower than wild-type myoglobin and more in line with those of peroxidases like CcP (Conroy et al., 1978) and the manganese peroxidases H3 and H4 (Millis et al., 1989), which have reduction potentials of -194, -88 and -93 mV, respectively.
Chapter 8

Summary

As described in Chapter 3, the elucidation of the high resolution structure of recombinant horse heart myoglobin was required to confirm that this protein is folded in a manner comparable to horse heart myoglobin obtained from natural sources. A major advantage in using the horse heart myoglobin system for detailed structure-function studies is the ability to isomorphously crystallize the naturally occurring wild-type, recombinant wild-type, and most recombinant variant proteins. As detailed in Chapter 3, the structures of the wild-type and recombinant wild-type myoglobins are comparable, if not identical (Figure 3.2, Table 3.3).

The results obtained from the studies conducted in Chapter 4 demonstrate the importance of the distal heme pocket residue His64 in conferring latent peroxidase activity in myoglobin. Neither the His64Thr nor His64Tyr variants exhibited measurable peroxidase activity. A primary goal in constructing the His64Thr variant was to generate a five coordinate heme iron atom in the ferric state and thereby mimic the comparable coordination state seen in natural peroxidases. While a five coordinate heme environment was created, the lack of peroxidase activity for this variant appears to be a direct result of the inability of Thr64 to assist effectively this process in proton transfer during peroxide heterolysis. For the His64Tyr variant protein, the prior requirement for dissociation of the distal tyrosine ligand from the heme iron center appears to prevent hydrogen peroxide binding and thus peroxidase activity. Based on these results it seems that a single replacement of His64 in horse heart myoglobin is unlikely to have a positive role in enhancing its peroxidase activity. In recently reported work an alternative strategy that included replacing His64 with leucine and introducing a new histidine further away from the heme iron atom considerably increased the level of the peroxide dependent oxidation of styrene (Ozaki et al., 1997). These results suggest that such a
strategy might prove helpful in improving the manganese peroxidase activity of myoglobin.

In Chapter 5 detailed structural analyses of azide interactions with horse heart myoglobin were conducted to understand the balance between the affinity of the heme group of this protein for ligands and sterical hinderance imposed by nearby amino acid groups. Of particular interest was the extent to which movement of the heme iron atom is tracked by the proximal histidine residue upon ligand binding and how ligand binding impacts on other heme protein interactions. This study establishes that there are no large structural rearrangements required to accommodate formation of an azide complex in horse heart myoglobin (Figures 5.1, 5.2, 5.3, and Table 5.3). It is likely that the conformation hydrogen peroxide assumes when bound to myoglobin is related to that seen for azide. It is notable that substitution of the distal heme pocket residue His64 with threonine shows unexpected effects in terms of azide complex formation. For example, this substitution allows for two bound azide conformations (Figure 5.3). Furthermore, on the proximal heme side, a shorter hydrogen bond is observed between Ser92 and the proximal heme iron ligand His93.

Beyond the key role of His64, the dielectric constant of the heme pocket is likely to be a major factor in defining the chemistry of the heme prosthetic group. To address this issue, the studies conducted in Chapter 6 were designed to examine the role of several highly conserved heme pocket residues. These included the distal heme pocket residues Leu29 and Val67, as well as Leu104 which is located on the proximal heme side. In addition, while myoglobin has the same prosthetic group as peroxidases and a similar ligating histidine on the proximal side of the heme, it is lacking other essential residues in the heme pocket that are believed to be necessary for the expression of a high rate of peroxidase activity. Therefore, the work in Chapter 6 also concentrated on an examination of the structural and functional effects of substituting some of these missing functional groups in the heme binding pocket of myoglobin at residue positions 29, 67 and 104. This provided the opportunity to move towards the dual goals of gaining further insight into the roles of conserved myoglobin residues and the requirements necessary to increase the latent peroxidase
activity of this protein.

The increased hydrophilic nature of the distal heme pocket resulting with the substitution of tyrosine for leucine at position 29 was expected to assist in stabilizing higher oxidation states of the heme iron atom and therefore contribute to an increased level of peroxidase activity. Indeed, the observed peroxidase activity of the Leu29Tyr variant protein is higher than that of wild-type myoglobin and this increase does appear to be correlated with an increase in the polarity of the distal heme pocket. In addition, a proposed mechanistic model that may be relevant to this myoglobin variant involves the role of Arg48 in cytochrome c peroxidase, a residue which may move toward the iron bound oxene atom and stabilize it through a hydrogen bond interaction. Although not ideally positioned, Tyr29 might be acting to stabilize such a reaction intermediate. However, it is clear Tyr29 by itself is not adequate to confer a strong peroxidase activity on myoglobin and additional substitutions would have to be added to the Leu29Tyr variant protein to enhance this activity.

In general, the reduction potentials of peroxidases are relatively low. As such, it is of interest that despite an increase in heme reduction potential, the Val67Arg variant myoglobin also shows an increase in peroxidase activity that is ~ 8 times greater than that of wild-type myoglobin. This is particularly surprising in light of the fact that higher oxidation states of the heme iron reactive centre are likely to be destabilized in this variant protein relative to those of wild-type myoglobin. A feature likely to be a major factor in determining the reactivity of the Val67Arg variant is the increased polarity of the distal side of the heme pocket. In further studies discussed in Chapter 6 it was found that substitution of asparagine for leucine at residue 104 in the proximal heme pocket leads to a 3-fold increase in activity towards hydrogen peroxide. Like the results obtained for the Val67Arg variant protein this additional activity appears related to the increased polarity of the heme pocket.

The studies in Chapter 7 were more directly related to the goal of introducing a manganese peroxidase activity into myoglobin by constructing a manganese binding site on the surface of this
protein. The requirements for this binding site were stringent and included a significant degree of solvent accessibility, sufficiently tight manganese binding, close proximity to the heme group, and the need to accommodate other amino acid substitutions introduced to promote the latent peroxidase activity of myoglobin. The resulting manganese binding site was located near the heme D-propionate group and it is of interest that it turns out to be comparably positioned to a related site in manganese peroxidase (Sundaramoorthy et al., 1994). It should be emphasized that consideration of the placement of a manganese binding site on myoglobin preceded the eventual determination of the location of this binding site in manganese peroxidase. Functional characterizations show an ~3-fold enhancement of the oxidation of Mn(II) to Mn(III) in the presence of hydrogen peroxide for both variant myoglobins having the designed manganese binding site.

It is clear from the present work that amino acid substitutions can influence the latent peroxidase activity of horse heart myoglobin. In addition it has been possible to construct a Mn(II) binding site at an appropriate position on myoglobin and demonstrate a measurable manganese peroxidase activity. Thus the preliminary building blocks that could potentially lead to an efficient manganese peroxidase activity are in place, although it is also clear that additional protein engineering work will be required to optimize this catalytic function. Overall it is evident from the studies in this thesis that protein engineering techniques provide us with the tools necessary to design new enzymes, but that the complex interplay of protein structure-function relationships at the atomic level means that this remains a formidable challenge.


U.S.A. 68, 614-618.


Tong, H., Berghuis, A. M., Chen, J., Luo, Y., Guss, J. M., Freeman, H. C., & Brayer, G. D.


