Studies in Protein Structure:
The structure and properties of the iron superoxide dismutase from *Methanobacterium thermoautotrophicum*.
The structures of β-lactoglobulin in two new crystal forms.

A dissertation
submitted in partial fulfillment
of the requirements for the degree
of
Doctor of Philosophy
in the
Institute of Fundamental Sciences
at
Massey University
by
Julian James Adams

Massey University

2002

"If the going is tough and the pressure is on; if reserves of strength have been drained and the summit still not in sight, then the quality to seek in a person is neither great strength nor quickness of hand, but rather a resolute mind firmly set on its purpose that refuses to let its body slacken or rest. "

Sir Edmund Hillary.

ABSTRACT

Studies in Protein Structure:

The structures of β-lactoglobulin in two new crystal forms.

The structure and properties of the iron superoxide dismutase from *Methanobacterium thermoautotrophicum*.

By

Julian James Adams

The crystal structure of *Methanobacterium thermoautotrophicum* iron superoxide dismutase (*Mt*-FeSOD) has been determined by X-ray diffraction to a resolution of 2.6 Å. The crystals were grown from PEG 6000 at a pH of 5.5, and the structure was solved by molecular replacement. The structure, in concert with structural and functional data from other Fe and Mn SODs, provides insights into aspects of metal specificity, reactivity of superoxide dismutase towards toward the inhibitor azide and deactivator hydrogen peroxide, and how the primary structure is involved in subtle tuning of these properties. The structure reveals how the protein is designed for thermal and chemical stability, yet retains moderate superoxide dismutase activity at ambient temperature.

Bovine β-lactoglobulin (BLG) has been studied for many decades; numerous X-ray and NMR structures are available. Here we present two

new X-ray structures, one from a crystal grown at very low ionic strength, and at the lowest pH (~5.2) of any X-ray structure. This structure provides validation of the other, high ionic strength X-ray structures. The core elements of BLG, an eight-stranded β-barrel, a three-turn α-helix external to the barrel, and an external β-strand (that forms the dimeric interface), are almost invariant across all structures. Four flexible loops have a variety of positions in the known structures and this represents a set of snapshots of the *in vivo* states of BLG. These flexible loops play an important role in the entropic stabilization of the β-barrel.

ACKNOWLEDGEMENTS

The many staff and students in the Institutes of Fundamental sciences and Molecular BioSciences are to be thanked for their help and input to the success of my studies. The following people deserve special mention:

My supervisor, Professor Geoffrey Jameson, for his unwavering enthusiasm, support and advice. Dr. Bryan Anderson and Dr. Gillian Norris my secondary supervisors, Professor Edward Baker of The University of Auckland also a secondary supervisor, for their invaluable input into my studies. I greatly appreciate the assistance of Dr. Paul Hempstead who guided me through the process of macro-molecular crystallography.

I would like to thank all of my coworkers in the Centre for Structural Biology; Trevor Loo, Dr. Renu Kadirvelraj, Dr. Nandana Ariyaratne, Dr. Stan Moore, Dr. Charlie Matthews and Matthew Bennett. Thanks also to all of my friends, especially Dr's Emily Parker, Rachel Williamson and Gavin Collis.

The efforts of the departmental technical staff is greatly appreciated, especially Udo Von Mulert, and the staff of Electronic Services and Barry Evans, and the staff of Mechanical Services.

I cannot put into words how important the love and support of my wife, Santi, has been in the last two years, thank you. Lastly my parents and brother for their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PUBLICATIONS

Parts or all of chapters 2, 3, 4 and 5 have been submitted to the *Journal of Biological Inorganic Chemistry*: Adams, J. J., Anderson, B. F., Renault, J. P., Vechère-Béaur, C., Morgenstern-Badarau, I., & Jameson, G. B. (2002). Structure and properties of the atypical iron superoxide dismutase from *Methanobacterium thermoautotrophicum, J. Biol. Inorg. Chem.* Accepted.

Parts or all of chapters 9 and 13 have been published in the *International Dairy Journal*: Jameson, G. B., Adams, J. J., Creamer, L. K. (2002). Flexability, functionality and hydrophobicity of bovine β-lactoglobulin, *International Dairy Journal*. **12**, 319.

Manuscript in preparation for *Acta. Crystallographica Section D (Biological Crystallography)*: Adams, J. J., Anderson, B. F., Hempstead, P. D., Creamer, L. K., Jameson G. B. (2002). The structure of bovine β-lactoglobulin A, from crystals grown at very low ionic strength, *Acta. Cryst.* D, manuscript in preparation.

Chapter 1

Superoxide dismutase

1.1 Superoxide dismutase

Superoxide dismutase (SOD) catalyzes the dismutation of the superoxide radical into hydrogen peroxide and dioxygen [Cade *et al.* 1971, Keele *et al.* 1970](Figure 1.1).

$$M^{III} + O_2^- \longrightarrow O_2 + M^{II} \quad (1)$$

$$M^{II} + O_2^- + 2H^+ \longrightarrow H_2O_2 + M^{III} \quad (2)$$

$$2O_2^- + 2H^+ \longrightarrow H_2O_2 + O_2$$

Figure 1.1

Dismutation of superoxide

1. Superoxide is oxidized to oxygen with the loss of one electron to the metal center.
2. Superoxide is reduced to hydrogen peroxide with addition of one electron donated by the metal and addition of two protons.

SOD is an enzyme present in aerobic organisms, having an essential role in the biological defense against superoxide, and therefore against any reactive radical species derived from superoxide. Dismutation of O_2^- by the enzyme occurs at a transition metal-containing active site. The metals Cu, Zn, Mn, Fe and Ni have been found in the active sites of SOD.

Superoxide undergoes spontaneous dismutation in solution but not very rapidly. The enzymatic dismutation has been calculated to be 10^6 times faster than the spontaneous reaction [Fridovich 1983]. The benefits of

SOD in the cellular environment are thus considerable.

Superoxide dismutases can be separated into three main classes based on biochemical and physical properties. The first class contains Cu and Zn at the active site, the second class contains Mn or Fe, and the third class contains Ni. Although the three classes are evolutionarily unrelated and have totally different primary, secondary and tertiary structures, all classes show essentially the same catalytic activity for superoxide *in vitro*.

CuZnSOD (Figure 1.2) was first discovered unintentionally during the isolation of carbonic anhydrase in 1939. As carbonic anhydrase was suspected of containing copper, protein fractions derived from bovine hemolysate were assayed both for the catalysis of the hydration of CO_2 and for copper content. However, carbonic anhydrase in fact contains zinc and not copper, so the fractions isolated on the basis of their high copper content were not active when assayed for carbonic anhydrase activity. Since the activity of the isolated copper protein was not known, the function of copper storage was assigned, and the protein named hemocuprein [Mann & Keilin 1939]. In 1969 an enzyme from bovine erythrocytes was isolated and purified. The enzyme catalyzed the dismutation of superoxide radicals and so was named superoxide dismutase [McCord & Fridovich 1969]. The properties of the enzyme were identical to the previously described copper-containing hemocuprein, and the connection was immediately obvious.

Based on SOD activity, a manganese-containing SOD and an iron-containing SOD were isolated from *E. coli* [Keele *et al.* 1970, Yost & Fridovich 1973](Figure 1.3). The Mn and Fe SODs are described in more detail in section 1.6.

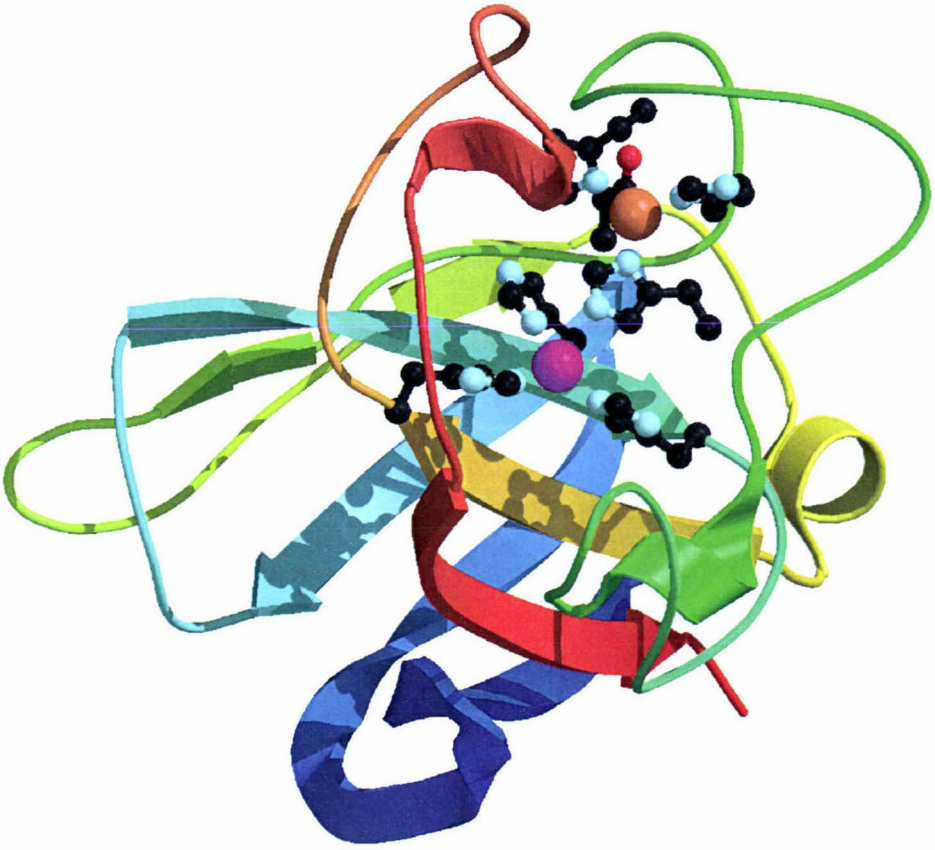


Figure 1.2

Bovine CuZnSOD

CuZnSOD is an antiparallel β-sheet protein with a single domain. The active site is on the surface of the protein. The active site residues (ball and stick, Cu (pink) Zn (orange)) are contributed by many different β-strands form throughout the protein. The molecule is rainbow-coloured: dark blue at the N-terminus through to red at the C-terminus.

All molecular representations unless otherwise indicated were produced with MOLSCRIPT [Kraulis 1991] and rendered with RASTER3D [Merritt & Bacon 1997]

Figure 1.3

MnSOD from *E. coli* [Edwards 1999].

The α-helical domain is shown in blue, the αβ domain in red. The residues of the active site are shown in ball and stick with the active site Mn shown in yellow. The active site is made up of residues from both domains.

Recently, examples of SODs containing only nickel at the active site have been isolated from various *Streptomyces* species [Kim *et al.* 1998, Youn *et al.* 1996]. This class of SODs has no sequence homology with either the CuZn or Fe/Mn class of SOD. As yet no 3D structure of the NiSOD exists. However, EPR, and EXAFS studies of the active site have shown it to contain two Ni atoms as a Ni-S cluster (Figure 1.4)[Choudhury *et al.* 1999].

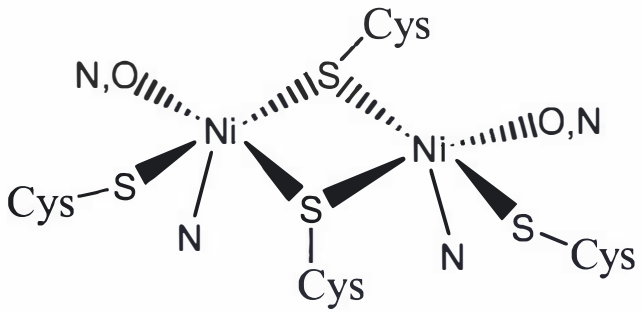


Figure 1.4

The active site of NiSOD from *Streptomyces seoulensis*

The Ni-S cluster as predicted by EPR and EXAFS. S-contributors are assumed to be Cys and the O, N donors are unknown.

1.2 Superoxide

Superoxide, O_2^-, is formed by a single-electron reduction of oxygen in a thermodynamically unfavorable reaction that gives a species of higher energy and reactivity than molecular oxygen. The superoxide radical is capable of acting as either a reducing agent or an oxidizing agent (Figure 1.5).

$$O_2 + e^- \rightleftharpoons O_2^-$$

$$O_2^- + 2H^+ + e^- \rightleftharpoons H_2O_2$$

Figure 1.5

The electrochemistry of superoxide.

The electrode potentials of the half reactions O_2 to O_2^- ($E^{0'}$= -0.137 V) and O_2^- to H_2O_2 ($E^{0'}$= +0.89 V) [Sawyer & Valentine 1981] show that superoxide is a facile redox reagent. The redox reactions are coupled with proton associations in water, conferring a pH dependence on the redox couples. The conjugate acid of O_2^- is the perhydroxyl radical $HO_2^{\bullet}$,

which has a pKa of 4.88 [Behar *et al.* 1970]. The perhydroxyl radical is a more powerful oxidizing species than O_2^-, whereas O_2^- is the better reducing agent than $HO_2^\bullet$ at pH ~7.0 [Ilan *et al.* 1976].

1.3 Superoxide in biochemical systems

Oxygen is for many species essential for life, yet it is a paradox. Some forms of oxygen are toxic and will cause cell death. Oxygen radicals have a dual and seemingly contradictory role in cell metabolism, where they can act as either friend or foe. An example of the protection role is the deliberate production of superoxide radicals during a process called the respiratory burst [Kettle & Winterbourne 1994]. Stimulus of neutrophils (granular white blood cells that participate in the immune response) produces quantities of superoxide radicals, which are intermediate to the formation of more powerful oxidants (eg. HOCl and OCl^-) used by cells as antimicrobial agents [Kettle & Winterbourne 1994].

Non-directed production of oxygen radicals, however, can have a detrimental effect on the cell, oxidizing and modifying essential cell components, leading eventually to cell death [Fridovich 1983, Fridovich 1995]. Superoxide can directly damage cellular components by acting as either a univalent reducing or oxidizing agent. Most cellular damage is due to oxidative damage, although it is thought that O_2^- causes little or no direct cellular damage [Kettle & Winterbourne 1994]. Instead superoxide causes cellular damage indirectly by generating radical products that are even more powerful and reactive cellular toxicants. These radicals can have even greater oxidizing power, and in some cases are better placed within the cell to cause damage to particular cellular components. The perhydroxyl radical is a neutral acid, its ionization involves charge

separation, which is more favorable in solutions with a high dielectric constant. The movement of O_2^- from water into environments within the cell with lower dielectric constants, such as lipid micelles or membranes would be accompanied by its protonation. So it is likely that the reactive species in the lipid environment would be the perhydroxyl radical [Fridovich 1983]. The perhydroxyl radical has been shown to react directly with linoleic, linolenic and arachidonic unsaturated fatty acids. This reaction, which takes place in the absence of metal catalysts or hydrogen peroxide, could initiate oxidative damage to polyunsaturated phospholipids, leading to impairment of cellular membrane function [Gebicki & Bielski 1981].

By far the most important mechanism of indirect damage to cells mediated by superoxide is the formation of the more reactive radical species, the hydroxyl radical ($OH^\bullet$). Any system generating O_2^- will also produce H_2O_2, either by spontaneous dismutation of O_2^-, or by enzymatic dismutation of O_2^-. The combination of O_2^- and H_2O_2 with catalytic amounts of transition metal ions leads to the formation of the extremely reactive and powerful oxidizing radical $HO^\bullet$. Hydroxyl radicals are highly reactive and will react with most cellular components upon collision. Hydroxyl radicals generated in free solution are more likely therefore to react with species found in free solution. However, superoxide radicals are much less reactive and so can diffuse from their formation sites and react with transition-metal ions in more critical parts of the cell. Catalytic metal ions are most likely to be found associated with polyanionic DNA or negatively charged protein surfaces rather than in free solution; thus, the hydroxyl radicals are more likely to be produced adjacent to these sites and react with these more critical cellular components. This sequence of events has been termed the site-specific

Haber-Weiss reaction [Samuni *et al.* 1984, Czapski 1984].

The major site of non-directed production of superoxide is the mitochondrion. The energy generated by electron transport to oxygen is used in the formation of ATP from ADP and phosphate. The aerobic organism benefits from the energy, but must minimize the production of the partially reduced oxygen intermediates. The organism does this in part by using enzymes that mediate two or four-electron reductions without the release of the oxygen intermediates during the reduction [Beyer *et al.* 1991]. However, because the univalent pathway is the lowest energy pathway for the reduction of oxygen, single-electron auto-oxidations of cellular components can still occur, and are in fact the major source of superoxide within the cell. Cellular components such as hemoglobin, myoglobin, reduced cytochrome c, ferredoxins, leukoflavins, and polyhydric phenols can facilitate the production of superoxide [Fridovich 1983].

1.4 The manganese and iron superoxide dismutases

The Fe and Mn SODs make up a family with a common ancestral origin, as they share high amino-acid sequence similarity as well as high structural similarity. They have been isolated from a variety of prokaryotes, eukaryotes and archaea. All have a molecular weight per subunit of approximately 22000 Da. SODs from mesophilic prokaryotes and eukaryotes tend to exist as dimers, whereas those from Mycobacterium and thermophilic prokarya/archaea tend to be tetramers [Kusunose *et al.* 1976].

Iron and manganese SODs are found in eukaryotic mitochondria (only

MnSOD), eubacteria (both Fe and Mn SOD), and archaebacteria (both Fe and Mn SOD). Organisms seem to possess preferentially MnSOD or FeSOD depending on their habitat. It has been hypothesized that in general aerobes and facultative anaerobes possess MnSOD and/or FeSOD while strict anaerobes have no SOD or, in a few cases, FeSOD [Asada *et al.* 1980, Steinman 1982]. The higher redox potential of the Mn^{3+}/Mn^{2+} couple compared with the Fe^{3+}/Fe^{2+} couple may favor MnSOD in an aerobic environment [Takao *et al.* 1990].

1.5 The superoxide dismutase cycle

The superoxide dimutase cycle shown below is referred to as the "5-6-5 mechanism" [Lah *et al.* 1995] and is a structure-based modification of the scheme proposed by Bull & Fee [Bull & Fee 1985](Figure 1.6). Although proposed for FeSOD it appears to be applicable to MnSOD. This mechanism has now been further modified to include recent pulse-radiolysis experiments, which have shown a competing pathway from species 3 returning to species 1 [Hearn *et al.* 2001]:

1. The five-coordinate Fe(III) (or Mn(III)) in the resting oxidized state of the enzyme has an axial solvent ligand, postulated to be OH^- at pH 7, and four protein residues, an aspartate and three histidines, arranged in a trigonal bipyramidal manner.

2. Superoxide anion binds at the coordination site opposite OD2 of Asp156 without substitution of the liganded hydroxide anion, to give a distorted octahedral environment. Electron transfer from O_2^- to Fe(III) then generates the dioxygen leaving group and reduces

the metal to Fe(II). The change in the inner sphere complex is associated with only a small change in the metal-ligand geometry. This is consistent with the low energy characteristics of catalytic intermediates and allows rapid turnover.

3. Protonation of the axial solvent ligand restores the net +1 charge on the metal complex. The reduced enzyme with a trigonal bipyramidal geometry then awaits the second superoxide.

Recent pulse-radiolysis experiments on MnSOD [Hearn *et al.* 2001] have identified two competing pathways in the superoxide reduction reaction. An "inner sphere" reaction that leads to species 4, and an "outer sphere" reaction that proceeds via a concerted mechanism without the formation of an intermediate. As species 4 has a finite lifetime, it can be considered a substrate inhibition intermediate. Alternatively, species 4 can be represented as an Fe^{III}-O_2^{-II} species, which can be considered (Hearne *et al.* 2001) as a product-inhibited species. No reports of the "inner sphere" mechanism have been made with respect to FeSOD. This ability to undergo the "outer sphere" reaction may be a defining functional difference between the Fe and MnSODs; it may also have implications with respect to peroxide and azide sensitivity as will be discussed later.

4. The coordinated superoxide anion is reduced, receives two protons, one is assumed to be provided by the ligated water molecule, and leaves as hydrogen peroxide, with SOD returning to the resting state (1).

Other competing kinetic processes from species 3 and 1 lead to dead-end products 5 and 6.

5. Species 1 gains a ligand bound at the sixth coordination site. EXAFS and X-ray crystallographic studies support the postulation that the sixth ligand is an OH^-. This sixth ligand binds with a pK of ~9. This is consistent with observed competitive inhibition by hydroxide.

6. Species 6 exists at high pH. The Fe(II) is five coordinate. The axial ligand is protonated and Tyr34 is ionized. The negatively charged phenolate ion has the effect of lowering electrostatic attraction for anionic superoxide in the vicinity of the active site.

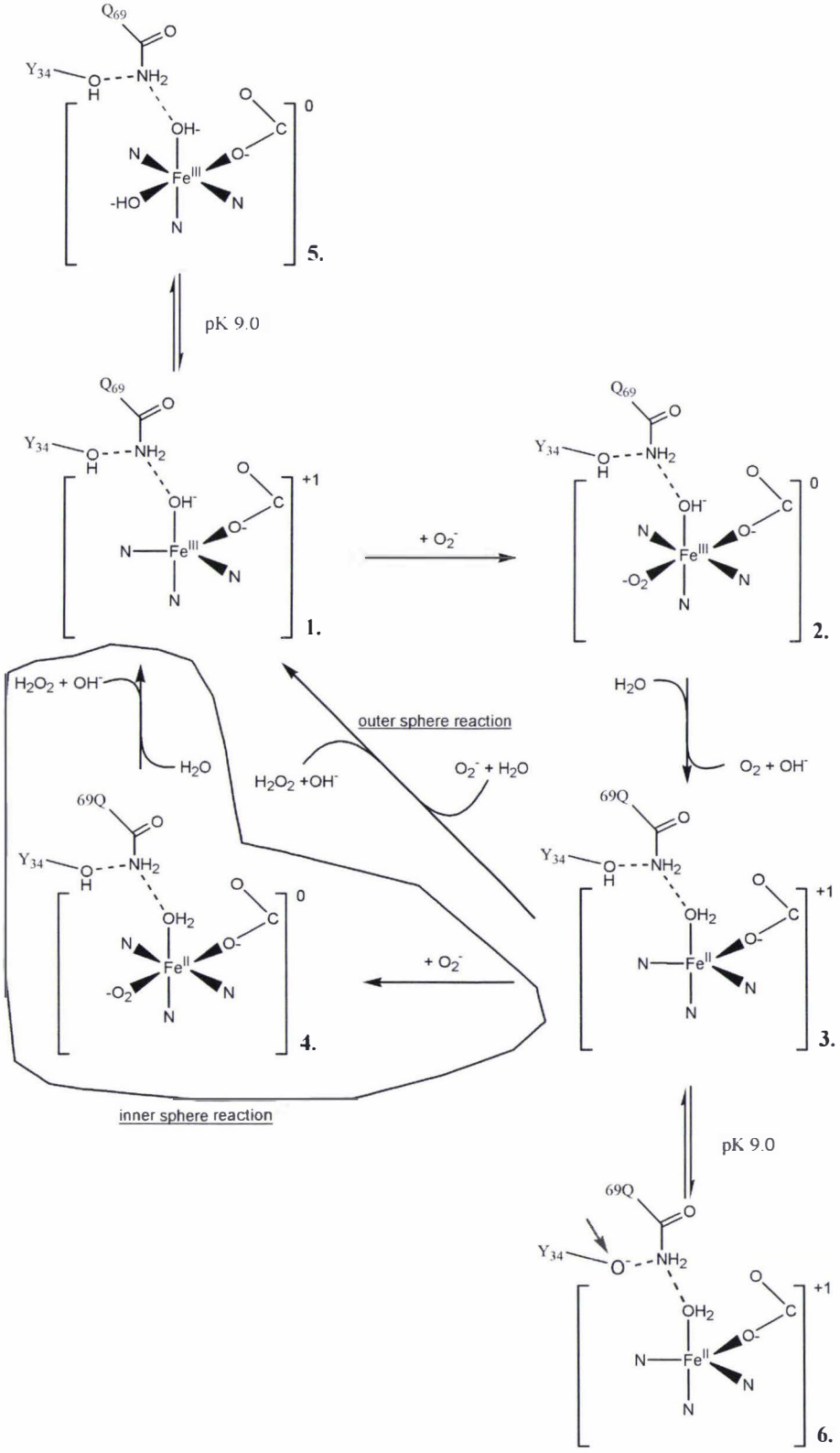


Figure 1.6

The superoxide dismutase cycle [Lah *et al.* 1995]

The original scheme has been modified to include both the outer- and inner-sphere reactions in the superoxide reduction reaction. Key hydrogen bonds are shown with dashed lines.

1.6 Metal-specificity in superoxide dismutases

More than 200 genes have been identified as putative Mn/FeSODs [BLAST 1990] and a substantial proportion of these gene products have been functionally characterized. In addition the SOD's from eleven different species have been structurally characterized (Table 1.1).

Table 1.1. Iron and manganese superoxide dismutase structures.

Iron superoxide dismutase structures	PDB ID	Reference
Pseudomonas putida (*ovalis*)	1dto	Ringe *et al.* 1983 Bond *et al.* 2000
Escherichia coli	1isa	Stallings *et al.* 1983
Mycobacterium tuberculosis	1ids	Cooper *et al.* 1995
Propionibacterium freudenreichii (*shermanii*)	1ar4	Schmidt *et al.* 1996
Aquifex pyrophilus	1coj	Lim *et al.* 1997
Sulfolobus solfataricus	1sss	Ursby *et al.* 1999
Sulfolobus acidocaldarius	1bo6	Knapp *et al.* 1999
Bacteroides (*Porphyromonas*) *gingivalis*	1qnn	Sugio *et al.* 2000
Manganese superoxide dismutase structures		
Thermus thermophilus	3mds	Stallings *et al.* 1985
Bacillus stearothermophilus		Parker & Blake 1988
human mitochondria	1abm	Borgstahl *et al.* 1992
E. coli	1vew	Edwards *et al.* 1998

However, despite these similarities, some members of this family have been shown to have a rigid metal specificity, becoming essentially inactive, at least at pH 7.8, when reconstituted with the wrong metal

(Table 1.2). For the purposes of this discussion metal specificity is defines as: *metal specific*, less than 5% activity after reconstitution with the "other" metal; *metal preferring*, less than 50% activity after reconstitution; and *cambialistic*, greater than 50% activity after reconstitution with the "other" metal. Similarly, among the MnSODs, specificity for Mn has been determined from reconstitution studies for the MnSODs listed in Table 1.3.

Table 1.2. Manganese-substituted FeSOD (inactive).

Manganese-substituted FeSOD	
	Yamakura *et al.* 1998
Ps. ovalis	Yamakura & Suzuki 1986
E. coli	Beyer & Fridovich 1987
Tetrahymena pyriformis	Barra *et al.* 1990
Aquifex pyrophilus	Lim *et al.* 1997
Deinococcus radiodurans	Juan *et al.* 1991
Acidianus ambivalens	Kardinahl *et al.* 2000
Sulfolobus solfataricus	Yamano *et al.* 1999

The MnSODs from mitochondria [Ishikawa *et al.* 1987; Weisiger & Fridovich 1973; Deutsch *et al.* 1991; Ravindranath & Fridovich 1975], for example, pick up only Mn, but lack of activity of an iron-substituted species has not been demonstrated. Similarly, several FeSODs, such as those from *Mycobacterium tuberculosis* [Bunting *et al.* 1998] and *Methanobacterium thermoautotrophicum* [Takao *et al.* 1991], pick up only iron from the growth medium, both in homologous and heterologous expression, even when Fe-depleted and Mn-rich media are used.

Table 1.3. Iron-substituted MnSOD (inactive).

Iron-substituted MnSOD	
E. coli	Whittaker *et al.* 1997
Serratia marcescens	Yamakura *et al.* 1995
Bacillus stearothermophilus	Brock & Harris 1977
Thermus thermophilus	Whittaker & Whittaker 1999
Methylomonas J[a]	Matsumoto *et al.* 1991

[a] Although described as cambialistic the Fe form has less then 5% of the activity of the Mn form, and is classified here as Mn specific.

There is a significant number of metal non-specific (cambialistic) SODs, that show nearly equal activity with Mn or Fe at pH ~7.8 (Table 1.4). Some MnSODs show weak but perceptible activity with Fe; these can be classed as Mn-preferring SODs (Table 1.4). The metal non-specific SODs generally show substantially lower specific activity by the standard assay [McCord & Fridovich 1969] than either the Mn- or Fe-specific SODs, even when fully metallated [Renault *et* al. 2000]. In general, all Fe-substituted MnSODs (Fe_2-(Mn)SOD) show sharply increased activity at a lower pH of ~6.5, whereas the converse is not true for Mn-substituted FeSODs (Mn_2-(Fe)SODs).

Table 1.4. Cambialistic and Mn-preferring SODs.

Cambialistic SODs	
Bacteroides (*Porphyromonas*) *gingivalis*	Hiraoka *et al.* 2000
Bacteroides thetaiotamicron	Pennington & Gregory 1986
Bacteroides fragilis	Gregory & Dapper 1983
Pr. shermanii	Meier & Sehn 1994
Mn-preferring SODs	
Pyrobaculum aerophilum	Whittaker & Whittaker 2000
Aeropyrum pernix	Yamano & Maruyama 1999
Sinorhizobium meliloti	Santos *et al.* 1999
Mycobacterium smegmatis	Yamakura *et al.* 1995
Streptococcus mutans	Martin *et al.* 1986

The origin of this metal specificity has been the subject of much interest [Vance & Miller 1998, Edwards *et al.* 1998, Leveque *et al.* 2001]. Some recent explanations of this metal specificity include pH-dependent metal environments, allowing OH^- to bind to the metal and competitively inhibit the SOD activity [Tierney *et al.* 1995; Meier *et al.* 1998]. Another, not necessarily unrelated, explanation focuses on tuning of the metal's redox potential by subtle, but as yet unidentified, changes in the coordination environment and outer sphere by the protein [Vance & Miller 1998, Leveque *et al.* 2001].

1.7 Inactivation of FeSOD by hydrogen peroxide

In general, FeSOD is inactivated by hydrogen peroxide, whereas MnSOD is not [Asada *et al.* 1975]. CuZnSODs are also inactivated by hydrogen peroxide with the associated disruption caused by oxidation of the side chain of one histidine residue [Bray *et al.* 1974]. *Pseudomonas ovalis* FeSOD has tryptophan and cysteine side chains, as well as histidine,

oxidised, whereas *E. coli* FeSOD loses iron and the tryptophan, but not histidine, side chains are oxidized [Yamakura 1984, Beyer & Fridovich 1987].

The hydrogen peroxide inactivation of FeSOD is believed to be facilitated by an oxidizing species, which is produced by the reaction of hydrogen peroxide with Fe(II). In *E. coli* FeSOD, electron removal from a ligated imidazole followed by electron transfer from a nearby tryptophan is a suspected pathway to the tryptophan destruction [Beyer & Fridovich 1987]. This could explain why tryptophan is destroyed on inactivation, whereas histidine is not.

There is a small group of atypical superoxide dismutases that in their active iron-bound forms are neither inhibited by azide nor rapidly deactivated by hydrogen peroxide, behaviour typically ascribed to Mn-containing SODs. Members of this group include the very well characterised cambialistic SOD from eubacterial *Pr. shermanii* and the partly characterised manganese-preferring SOD from eubacterial *S. meliloti*. The FeSODs from archaeons *Methanosarcina barkerii* [Brioukhanov *et al.* 2000], *A. pernix* [Yamano & Maruyama 1999], and *S. solfataricus* [Yamano *et al.* 1999] show much reduced sensitivity to hydrogen peroxide compared to the *Ec*-FeSOD, as does the unsequenced FeSOD from *Methanobacterium bryantii* [Kirby *et al.* 1981], a close relative of *M. thermoautotrophicum*, which is the topic of this study.

1.8 Inhibition of FeSOD by N_3^-

Azide (N_3^-) is a competitive inhibitor of iron superoxide dismutase activity, with 50% inhibition typically occurring in eubacterial FeSODs at concentrations of 5 mM azide or less. MnSODs on the other hand are

only slightly inhibited by azide, with 50% inhibition requiring concentrations of more than 20 mM azide [Takao *et al.* 1990, Misra & Fridovich 1978]. In the known FeSOD/azide structures, azide binds to the active site in the sixth and vacant coordination site [Lah *et al.* 1995]. In the known MnSOD/azide structure, azide binds to the vacant sixth coordination site but in a different orientation [Lah *et al.* 1995]. These differences will be discussed in detail later.

1.9 The structure of Fe and Mn superoxide dismutases

The structure of Mn and Fe SODs has two domains. The N-terminal domain is an all-alpha domain consisting of two anti-parallel alpha helices connected by a loop region. The C-terminal domain is an alpha-beta domain. The core consists of a three-strand antiparallel β-sheet and four α-helices on one side of the β-sheet. A total of 14 residues is almost invariant among structurally characterised Mn and FeSODs, including the four involved in ligation to the metal ion [Weatherburn 1996]. Two residues, in general, distinguish MnSODs from FeSODs. Residue 77 (*Ec*-MnSOD numbering) is a glycine in MnSODs and a glutamine in FeSODs. Residue 146 is an alanine in FeSODs and either glutamine or histidine in MnSODs. SODs from species that are active with either metal ion at their active site have the MnSOD sequence at these two positions. Structural studies have shown that residues 77 and 146 are in close proximity to the active site and that a glutamine or (rarely) a histidine residue interacts with the strictly conserved Tyr34 and with the coordinated solvent species.

The active site containing the metal ion lies between the two domains. Two ligands are provided by each domain. Helices α1 and α2 each

provide a histidine ligand coordinated through the ring nitrogen (NE2). The end of the beta-sheet provides a histidine and an aspartate residue. The metal ion is five coordinate with the fifth ligand being an axially bound water or hydroxide.

1.10 Methanobacterium thermoautotrophicum iron superoxide dismutase

The discovery of a SOD gene in *M. thermoautotrophicum* was the first example of a SOD from a strict anaerobe. The deduced amino acid sequence of the gene product resembles eubacterial and eukaryotic MnSODs rather than the eubacterial FeSODs, based on the analysis of Parker & Blake [Parker & Blake 1988] (Table 1.5). However, the gene product is observed to contain Fe and not Mn [Takao 1990]. Since 1988 nearly 200 SOD primary sequences have become available, including archaebacterial SOD genes. These new data have overtaken some of the conclusions of Parker & Blake, although their work is still very well cited. Since methanogens live under extremely anaerobic conditions, it is intriguing in itself as to the existence of the gene and the active SOD which results.

Table 1.5. Primary residues that distinguish Fe and Mn SODs [Parker & Blake 1988]

Residue number	MnSOD[a]	FeSOD[b]	*Mt*-FeSOD
76	Gly	Ala	Gly
77	Gly	Gln	Gly
84	Phe	Tyr	Phe
154	Gln	Ala	His
155	Asp	Gly	Asn

[a] *Bacillus stearothermophilus*

[b] *Escherichia coli*

M. thermoautotrophicum iron superoxide dismutase *(Mt*-FeSOD) has a tetrameric structure similar to the two known archaebacterial FeSODs [Kirby *et al.* 1981, Searcy & Searcy 1981]. *Mt*-FeSOD's enzymatic characteristics, such as partial inhibition by azide and resistance against hydrogen peroxide disruption, are quite different from other FeSODs but resemble MnSOD [Misra & Fridovich 1978; Beyer & Fridovich 1987].

Mt-FeSOD is a highly charged protein, considerably more so than either *E. coli* Mn or FeSOD. EPR studies show that, although the protein sequence appears to code for an MnSOD, the active site is 100 % iron(III), with a typical FeSOD coordination environment [Renault *et al.* 2000]. The EPR studies reveal that the electronic structure of the active site primary coordination sphere is almost identical to that of *E. coli* FeSOD with the effective *g* values being only slightly different, and the rhombic parameters being 0.22 for *E. coli* FeSOD and 0.23 for *Mt*-FeSOD. The rhombic parameter is significantly lower than 1/3, indicating a distorted non-planar trigonal-bipyramidal coordination environment about the metal center [Renault *et al.* 2000].

MnSOD from *E. coli* and *T. thermophilus* undergo a low temperature thermochromism (indicative of conformational change) [Whittaker & Whittaker 1996; Whittaker & Whittaker 1997] with transition temperatures of 220 K and 305 K respectively. *Mt*-FeSOD also undergoes this conformational change, which confirms the assignment of an Mn-type primary sequence [Renault *et al.* 1999] with a transition temperature of 286 K. This value falls between the values observed for the *E. coli* MnSOD a mesophile, and the *T. thermophilus* MnSOD, an obligate thermophile, and may indicate that the T_{opt} for *Mt*-FeSOD may be lower than the T_{opt} for the organism. One explanation for a lower T_{opt} of the enzyme (and possibly an explanation for the existence of SOD in a thermophilic anaerobe) is that the SOD's function in the organism is to protect the cell from oxidative damage if the thermophilic environment (hot pool) is quenched with cold oxygenated water (monsoon rain).

1.11 Goals

The goals of this research project were to provide structural information about *M. thermoautotrophicum* iron superoxide dismutase, with specific aims to answer the questions of active-site coordination number and geometry, secondary coordination-sphere geometry, metal specificity, inhibitor response, and surface properties and solubility. To this end, three structures were solved and answers to many of these questions have been found.

Chapter 2

The structure of *Methanobacterium thermoautotrophicum* iron superoxide dismutase at 3.0 Å resolution

2.1 Protein expression and purification

This section (2.1 only) encompasses work done at the Institut de Chimie Moleculaire d'Orsay, Paris, France, by Jean Philippe Renault and Irène Morgenstern-Badarau, with whom we have been collaborating on this project. The work is included so as to provide a full picture of the structural determination.

The plasmid containing the *Mt*-FeSOD gene was transformed into the BL21 λ DE5 strain of *Escherichia coli*. The cells were grown in LB media containing both ampicillin (100 μg mL^{-1}) and $FeSO_4$ (100 μM). The cells were lysed by freeze-thaw in 50 mM Tris-HCl, 1mM EDTA, 1 mM DTT, and 0.2 mM phenylmethylsulfonyl fluoride. The suspension was centrifuged at 16000 g, and the clear supernatant fractionated with successive $(NH_4)_2SO_4$ cuts, with the majority of the *Mt*-FeSOD precipitating between 60 and 95 % $(NH_4)_2SO_4$ saturation. The precipitate was redissolved in 50 mM Tris-HCl/1mM EDTA and the solution was dialyzed overnight against this buffer [Takao *et al.* 1991]. The dialyzed protein solution was loaded onto a Resource Q anion-exchange column and eluted with a linear gradient of NaCl. Final purification was achieved on a Sephadex G200 size-exclusion column. The protein was dialyzed against $(NH_4)_2CO_3$ buffer and lyophilized [Renault *et al.* 2000].

2.2 Crystallization

The *Mt*-FeSOD solid (as received) was dissolved in ultra purified-water (distilled from activated charcoal) at a concentration of 20 mg mL^{-1}. Standard SDS-PAGE (0.01% SDS) of the solution showed two bands (Figure 2.1), one at 22 kDa consistent with the monomer and another much larger band at higher molecular weight, consistent with an extremely stable oligomer. With more extreme denaturing conditions (0.1% SDS) only monomer was observed. Native-PAGE showed the protein to be homogeneous.

Crystals were grown by the hanging drop method at 4°C: each drop consisted of 1 μL protein (20 mg mL^{-1} in water) and 1 μL of well solution. The well solutions contained either 100 mM $NH_4CH_3CO_2$, pH 5.5 and 20 % PEG 6000, or 100 mM $NH_4CH_3CO_2$, pH 5.5 and 18 % PEG 5000MME, or 100 mM $NH_4CH_3CO_2$, pH 5.5 and 15 % PEG 20000. Crystals grew to a size of 200x100x50 microns in approximately one week (Figure 2.2). Crystals grown from PEG 6000 and PEG 5000MME were used for data collection.

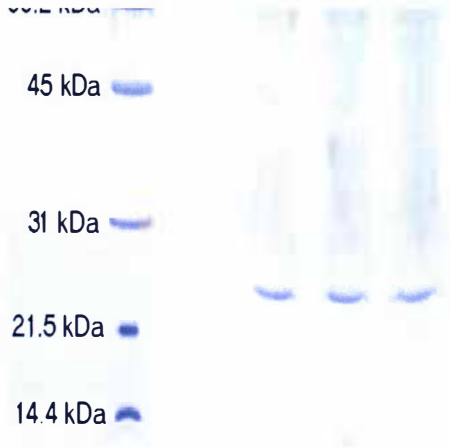


Figure 2.1
SDS-PAGE of *Mt*-FeSOD.
Note, the oligomeric protein band
at the top of the gel.

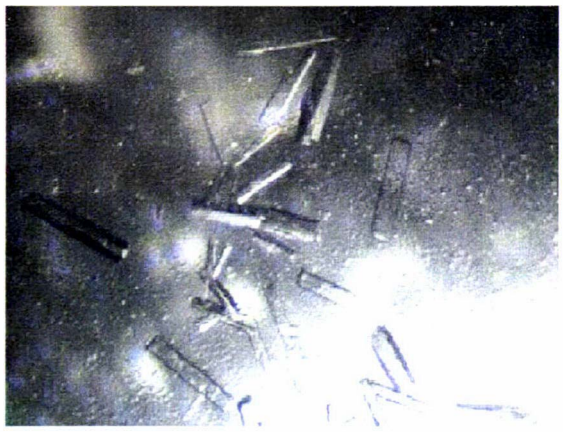


Figure 2.2
Mt-FeSOD crystals.

2.3 Data collection

Three crystals of *Mt*-FeSOD were used for data collection. In each case, a single crystal was mounted in a nylon loop, with a cryoprotecting solution consisting of the well solution from which the crystal was grown and 15-25 % glycerol. The crystal was soaked in the cryoprotectant for approximately 30-50 seconds then snap frozen in the cryo-stream (Oxford Cryosystems). Data were collected on a Rigaku RU200 rotating-anode generator running at 1.1 kW Cu-Kα, 0.1 mm focus, fitted with an AXCO SPA 50 mono-capillary X-ray optic coupled with a Rigaku RAxis IIC image-plate detector. X-ray frames were taken of each crystal (30 minute, 0.5° oscillation), until the crystal suffered radiation damage as described in Table 2.1.

Table 2.1. Data collection for *Mt*-FeSOD at 3.0 Å resolution.

Crystal	Start phi	End phi	% data	Resolution
ja70D3	0	45	55	3.0
ja76B1	0	30	30	2.8
ja76D3	40	80	45	3.1
TOTAL			98	3.0

Table 2.2. Data collection statistics for *Mt*-FeSOD at 3.0 Å resolution.

Space group	*C*2	
Unit cell	$a = 189.44$ Å	$\alpha = 90°$
	$b = 115.06$ Å	$\beta = 94.838°$
	$c = 58.49$ Å	$\gamma = 90°$
Data collection		
Observed reflections	153964	
Unique reflections	24553	
Completeness	98.3 %	(98.9 %)[a]
R_{merge} on I	0.123	(0.312)[a]
Overall I/σ	6.4	(2.0)[a]
Wilson plot B (Å^2)	45.8	
Redundancy	6.2	(3.0)[a]

[a] Outermost shell (3.11 - 3.0 Å)

Each set of frames was indexed separately, with the cell and orientation matrix determined by the auto-index procedure in DENZO [Otwinoski & Minor 1997]. The resulting data files were scaled and merged together with SCALEPACK [Otwinoski & Minor 1997]. Unit cell dimensions were averaged over all data. Relevant statistics are in Table 2.2. Measured intensities were converted to amplitudes using the program TRUNCATE [CCP4 1994].

2.4 Solution

The structure was solved with AMORE [Navaza 1994], using a truncated model from *Sulfolobus solfataricus* FeSOD (PDB ID: 1sss)[Ursby *et al.* 1999]. In the search model, side chains of identical residues were retained, side chains of non-identical residues were truncated to Ala (or Gly), side chains of residues coordinated to the metal were truncated to Ala, the metal atom was removed and water molecules deleted. Rotation and translation searches were carried out on data in the resolution range 12-4.0 Å. The rotation function (Table 2.3) showed three solutions to be superior to the others. The translation search (Table 2.4) revealed three dimers (six monomers) in the asymmetric unit. Note: each solution has an inverse (180° rotation) pair in the translation function. These pairs correspond to a dimer A→B, B→A rotation; either solution is valid. Rigid-body refinement of the six monomers using data in the range 12-3.5 Å resolution resulted in an *R* factor of 0.464 and a correlation coefficient of 45.9% (Table 2.5). Each solution was applied to the search model structure and the resultant coordinates generated (Figure 2.3).

Table 2.3 Solution to the rotation function for *Mt*-FeSOD at 3.0 Å resolution.

α	β	χ	Tx	Ty	Tz	Corr	Rfac	Solution
79.33	53.78	195.41	0.0000	0.0000	0.0000	9.5	0.0	1
154.00	70.39	94.32	0.0000	0.0000	0.0000	9.5	0.0	2
76.80	55.16	17.04	0.0000	0.0000	0.0000	9.1	0.0	3
153.96	67.91	274.60	0.0000	0.0000	0.0000	8.8	0.0	4
-1.50	65.96	116.37	0.0000	0.0000	0.0000	8.8	0.0	5
357.41	63.61	296.81	0.0000	0.0000	0.0000	8.5	0.0	6
267.25	67.26	120.17	0.0000	0.0000	0.0000	8.3	0.0	7
244.08	76.79	231.43	0.0000	0.0000	0.0000	8.0	0.0	8
242.50	78.81	51.83	0.0000	0.0000	0.0000	7.8	0.0	9
102.77	64.12	155.04	0.0000	0.0000	0.0000	7.7	0.0	10
265.80	65.14	300.44	0.0000	0.0000	0.0000	7.6	0.0	11
16.42	86.24	75.53	0.0000	0.0000	0.0000	7.6	0.0	12
194.84	51.56	184.44	0.0000	0.0000	0.0000	7.6	0.0	13
303.26	27.20	136.48	0.0000	0.0000	0.0000	7.5	0.0	14
192.30	52.19	5.67	0.0000	0.0000	0.0000	7.4	0.0	15
16.56	83.93	254.74	0.0000	0.0000	0.0000	7.4	0.0	16
306.62	83.18	83.79	0.0000	0.0000	0.0000	7.3	0.0	17
301.00	26.00	319.00	0.0000	0.0000	0.0000	7.3	0.0	18
246.65	49.66	220.47	0.0000	0.0000	0.0000	7.3	0.0	19
182.66	12.51	182.39	0.0000	0.0000	0.0000	7.3	0.0	20

Table 2.4. Solution to the translation function for *Mt*-FeSOD at 3.0 Å resolution.

Solution 1

α	β	χ	Tx	Ty	Tz	Corr	Rfac
265.80	65.14	300.44	0.3170	0.0000	0.3365	10.2	56.8
267.25	67.26	120.17	0.3185	0.0000	0.3365	10.2	56.6
91.81	67.47	110.54	0.1830	0.0000	0.3125	9.4	57.2
90.44	65.51	291.00	0.1830	0.0000	0.3125	9.1	57.2
357.41	63.61	296.81	0.1280	0.0000	0.1538	9.1	55.2
-1.50	65.96	116.37	0.1280	0.0000	0.1490	8.9	55.3
244.08	76.79	231.43	0.3661	0.0000	0.0192	6.6	56.8
242.50	78.81	51.83	0.0565	0.0000	0.1490	6.3	56.7
51.95	12.31	356.06	0.2842	0.0000	0.1346	5.9	57.5
154.00	70.39	94.32	0.3333	0.0000	0.2212	5.9	56.7
278.81	84.21	84.47	0.3393	0.0000	0.4663	5.5	57.7
269.82	90.00	63.74	0.4911	0.0000	0.1202	5.4	62.7
279.43	79.83	262.74	0.3423	0.0000	0.4760	5.4	57.7
303.26	27.20	136.48	0.0997	0.0000	0.4663	5.4	57.3

Solution 2

α	β	χ	Tx	Ty	Tz	Corr	Rfac
91.81	67.47	110.54	0.1838	0.7010	0.8116	18.4	55.1
90.44	65.51	291.00	0.1840	0.7019	0.8114	18.3	55.1
357.41	63.61	296.81	0.0627	0.0923	0.5884	17.7	54.1
-1.50	65.96	116.37	0.0627	0.0934	0.5888	17.4	54.2
25.65	61.96	111.67	0.7572	0.6294	0.1429	12.0	55.4
154.00	70.39	94.32	0.1268	0.6161	0.8473	11.8	55.4
246.65	49.66	220.47	0.2627	0.9929	0.2801	11.7	56.0
16.56	83.93	254.74	0.0576	0.9933	0.7592	11.6	56.1
16.42	86.24	75.53	0.0574	0.9955	0.7614	11.6	56.1
51.95	12.31	356.06	0.4370	0.6259	0.0618	10.8	57.0
153.96	67.91	274.60	0.9287	0.4769	0.5303	10.8	56.4
21.32	48.40	297.49	0.6923	0.4164	0.2266	10.8	56.2
246.00	51.54	41.03	0.7703	0.7862	0.1470	10.8	55.9
242.50	78.81	51.83	0.7929	0.8300	0.2314	10.7	56.4

Solution 3

α	β	χ	Tx	Ty	Tz	Corr	Rfac
357.41	63.61	296.81	0.0625	0.0892	0.5902	25.5	52.2
-1.50	65.96	116.37	0.5626	0.5901	0.5900	24.9	52.4
154.00	70.39	94.32	0.8467	0.4713	0.5435	17.4	54.6
153.96	67.91	274.60	0.2396	0.2037	0.6190	17.3	54.3
265.80	65.14	300.44	0.8178	0.4995	0.3385	17.2	55.2
244.08	76.79	231.43	0.1993	0.5493	0.3850	17.0	55.2
267.25	67.26	120.17	0.8179	0.5000	0.3383	17.0	55.1
242.50	78.81	51.83	0.1988	0.5494	0.3852	16.9	55.1
79.33	53.78	195.41	0.4273	0.0003	0.9682	16.7	54.8
76.80	55.16	17.04	0.0986	0.6214	0.9856	16.7	54.7
16.42	86.24	75.53	0.5577	0.4986	0.7673	16.6	55.0
102.77	64.12	155.04	0.7923	0.5291	0.1714	15.9	55.5

Table 2.5. Final solutions for *Mt*-FeSOD at 3.0 Å resolution.

α	β	χ	Tx	Ty	Tz	Corr	Rfac	
270.12	69.13	118.90	0.3189	-0.0044	0.3434	45.9	46.4	1
91.61	68.57	108.41	0.1841	0.7027	0.8095	45.9	46.4	2
359.46	63.31	294.04	0.0624	0.0854	0.5958	45.9	46.4	3

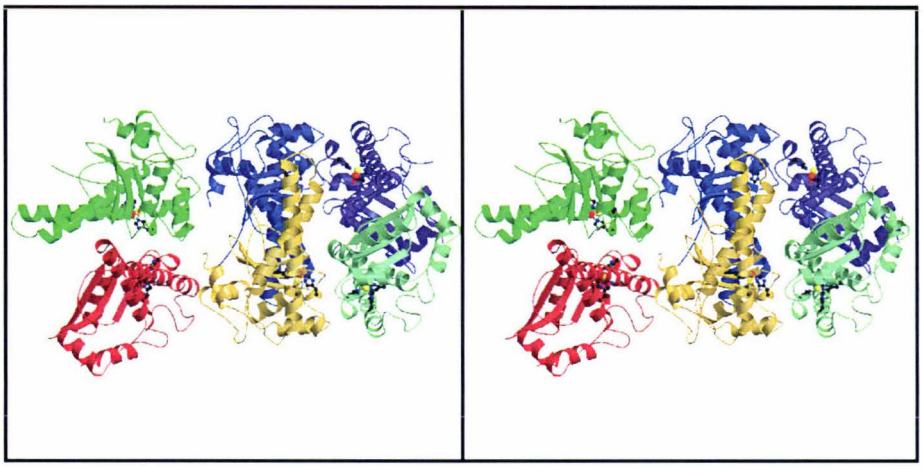


Figure 2.3

Stereo figure of the asymmetric unit of *Mt*-FeSOD at 3.0 Å resolution. Canonical (functional) dimers are A (red) and B (dark green), C (gold) and D (light green), and E (blue) and F (purple). The asymmetric unit contains one tetramer and ½ a tetramer, the complete tetramer ABA′B′, is formed by the crystallographic C_2 axis.

2.5 Refinement

From the initial truncated model a further round of rigid-body refinement and positional *xyz* refinement with CNS [Brünger *et al.* 1998] allowed electron density maps to be calculated. Subsequent model rebuilding with TURBO-FRODO [Roussel & Cambillau 1989] revealed most of the amino-acid side chains. Crystallographic simulated-annealing with CNS to remove model bias achieved R = 0.317, R_{free} = 0.348 with data in the range 50 to 3.0 Å resolution. The standard parameters used here and elsewhere for simulated annealing include: non-crystallographic symmetry (NCS) set to "on", anisotropic bulk-solvent correction applied to data in the range 6-2.6 Å, 100 steps for initial minimisation, 100 steps for geometry minimisation, "slow cool" molecular dynamics scheme, starting temperature of 3500 K, cooling rate of 25 K per step, 100 steps for final dynamics, target "MLF", and X-ray weight set to auto. Several

rounds of model building followed by *xyz* and grouped *B*-factor refinement brought the *R* and R_{free} to 0.266 and 0.326 respectively. A total of 251 water molecules was added with the "water-pick" procedure of CNS based on peak heights of 2σ in a $2F_o$-F_c map. All water molecules were *xyz*-fixed and their *B*-factors only were refined. Water molecules with poor density or unfavorable protein contacts were deleted. In the final cycle of refinement the peptide bond planarity restraint was greatly relaxed to allow the peptide bonds to find their minima in the electron density. The restraints were gradually retightened to produce the final model with individually refined but tightly restrained *B*-factors. The final values for *R* and R_{free} are 0.255 and 0.295. Considerable attention was paid to determining the optimum grouping of monomers for application of strict non-crystallographic symmetry (NCS) restraints. Representative electron density, real space R-factors (which define the fit to the electron density) are shown in Figures 2.5 and 2.6. All refinement cycles had 2 groups of strict three-fold NCS restraints applied to the whole molecule. The RMS deviations for bond lengths and angles are 0.009 Å and 1.5° respectively. A summary of refinement statistics is given in Table 2.6.

Other software used: MTZDUMP and MTZUTILS for inspection and modification of MTZ files, IPDISP for display of image-plate frames, HKLVIEW for display of processed data, BAVERAGE for calculation of average B values, MTZ2VARIOUS for conversion of file formats, RASMOL for display of molecules, XLOGGRAPH, PDBSET [CCP4 1994], LSQMAN [Kleywegt 1996] for calculation of molecular superpositions, MOLEMAN [Kleywegt 1999] and DATAMAN [Kleywegt & Jones 1999] for manipulation of coordinate files, and MAPPAGE [Roussel & Cambillau 1989] for preparation of electron density maps for TURBO-FRODO.

Table 2.6. Refinement statistics for *Mt*-FeSOD at 3.0 Å resolution.

Reflections	23319
Free *R* factor reflections	1234
Resolution limits	50-3.0 Å
R factor	0.255
Free *R* factor	0.295
Total number non-H atoms	10187
Number water molecules	251
R.m.s deviations from ideals	
Bond lengths	0.009 Å
Bond angles	1.4°
Dihedral angles	21.4°
Improper angles	0.9°
Average *B*-factors ($Å^2$)	
Main chain	22.7
Side chain	23.9
Water molecules	16.3
Ramachandran plot	Figure 2.4
Most favoured (%)	83.5
Allowed (%)	14.8
Generously allowed (%)	1.1
Disallowed (%)	0.6

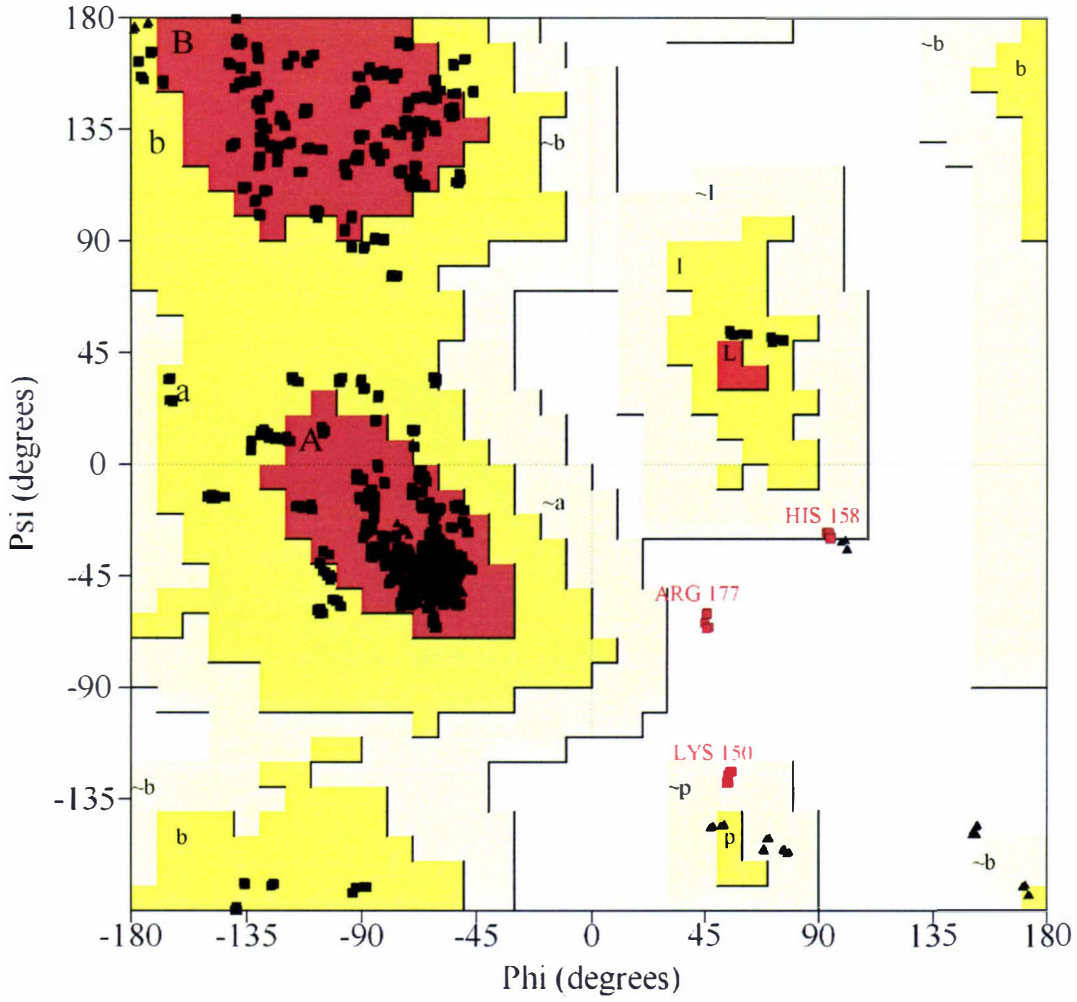


Figure 2.4

Ramachandran plot for *Mt*-FeSOD at 3.0 Å resolution after final refinement.

Labeled residues are K150, H158 and R177 (these outliers are discussed in full in section 5.3).

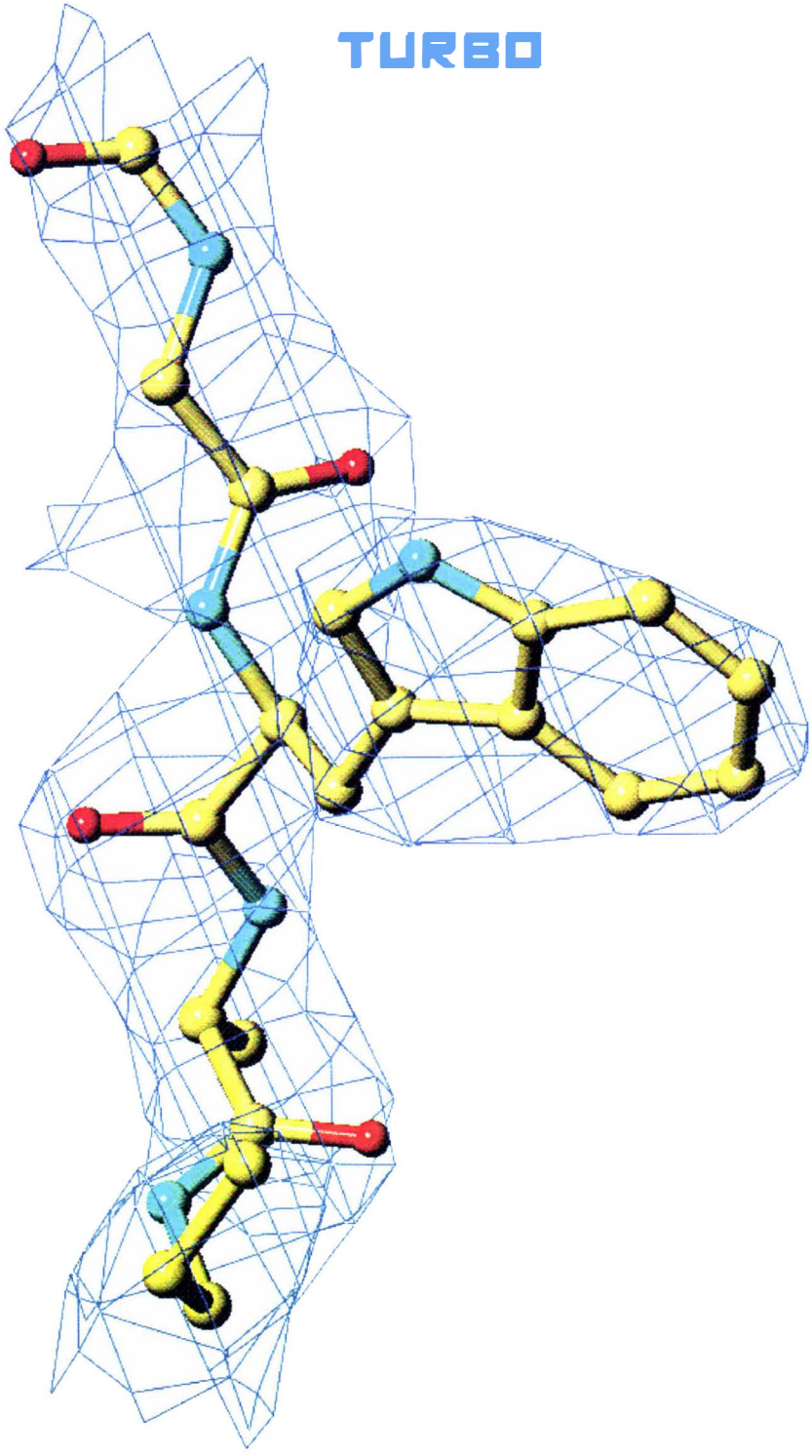


Figure 2.5

Typical electron density ($2F_o$-F_c) around Trp 121 of *Mt*-FeSOD. The electron density is contoured at 1.2 σ.

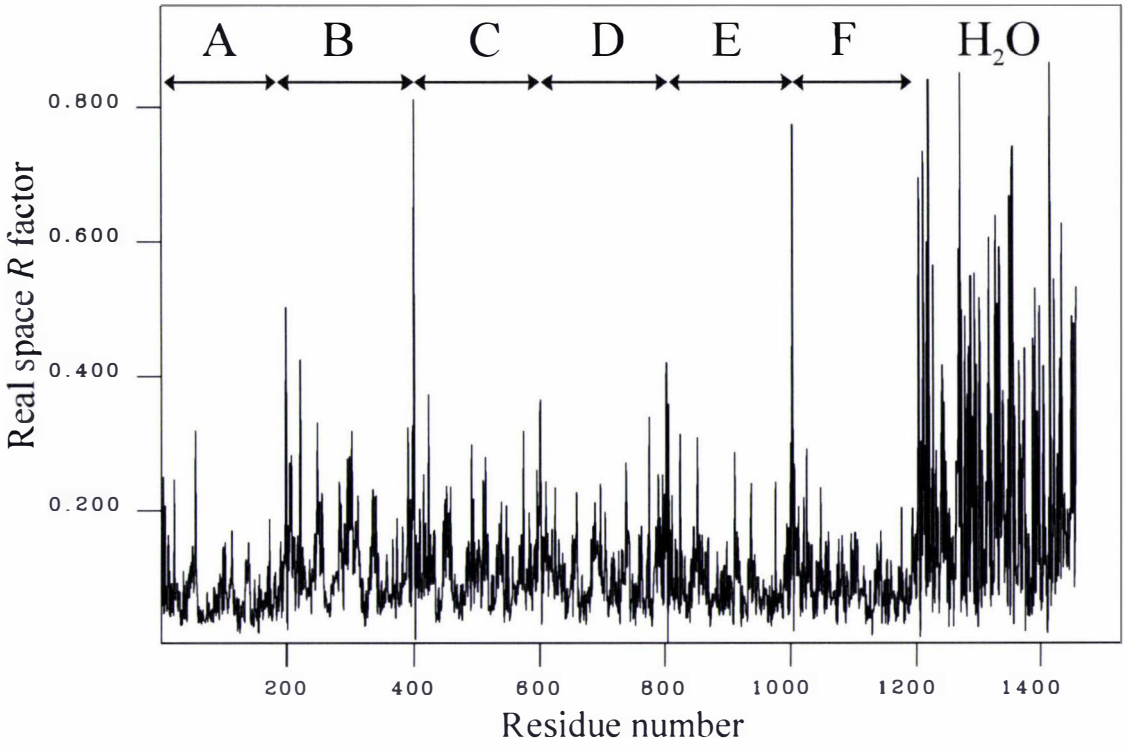


Figure 2.6

Real space *R*-factors per residue for *Mt*-FeSOD at 3.0 Å resolution, covering chains A-F and water molecules.

Chapter 3

The structure of *Methanobacterium thermoautotrophicum* iron superoxide dismutase at 2.6 Å resolution from crystals grown with 100 mM NaN_3

3.1 Crystallization.

The protein used in this crystallization was from the same preparation as in Chapter 2.

Crystals were grown by the hanging drop method at 4°C; each drop consisted of 1 μL protein (20 mg mL^{-1} in water) and 1 μL of well solution. The well solutions contained either 100 mM $NH_4CH_3CO_2$, pH 5.5, 100 mM NaN_3, and 20 % PEG 6000, or 100 mM $NH_4CH_3CO_2$, pH 5.5, 100 mM NaN_3, and 18 % PEG 5000MME, or 100 mM $NH_4CH_3CO_2$, pH 5.5, 100 mM NaN_3, and 15 % PEG 20000. Crystals grew to a size of 200x100x50 microns in approximately three weeks. Crystals grown from PEG 6000 were used for data collection.

3.2 Data collection.

One crystal of *Mt*-FeSOD was mounted in a nylon loop, with a cryoprotecting solution consisting of the well solution from which the crystal was grown and 15-25 % glycerol. The crystal was soaked in the cryoprotectant for approximately 30-50 seconds then snap frozen in the cryo-stream (Oxford Cryosystems). Data were collected on a Rigaku RU200 rotating-anode generator running at 1.1 kW Cu-Kα, 0.1 mm focus, fitted with an AXCO SPA 50 mono-capillary X-ray optic coupled

with a Rigaku RAxis IIC image-plate detector. X-ray frames were taken of the crystal (40 minute, 0.5° oscillation).

Table 3.1. Data collection statistics for *Mt*-FeSOD grown from 100 mM azide.

Space group	*C*2	
Unit cell	a = 189.28 Å	$\alpha = 90°$
	b = 114.01Å	$\beta = 94.58°$
	c = 57.98Å	$\gamma = 90°$
Data collection		
Observed reflections	283465	
Unique reflections	33418	
Completeness	88.1 %	(88.8 %)[a]
R_{merge} on I	0.122	(0.342)[a]
Overall I/σ	7.1	(2.3)[a]
Redundancy	8.4	(2.0)[a]

[a] Outermost shell (2.69 – 2.6 Å)

The cell and orientation matrix were determined by the auto-index procedure in DENZO [Otwinoski & Minor 1997]. The resulting data files were scaled and merged together with SCALEPACK [Otwinoski & Minor 1997]. Unit cell dimensions were averaged over all data. Relevant statistics are in Table 3.1. Measured intensities were converted to amplitudes using the program TRUNCATE [CCP4 1994]. The R_{free} set from the original structure (ja076) was transferred to the new data set with the program DATAMAN [Kleywegt & Jones 1999]; the transfer ensured that the R_{free} data set was unbiased with respect to the original model.

3.3 Solution

The previously determined structure with the side chains of residues coordinated to the metal truncated to Ala, the metal atom removed and water molecules deleted was subjected to rigid body refinement (CNS) [Brunger *et al.* 1998], leading to *R*-factors of $R = 0.386$ and $R_{free} = 0.392$.

3.4 Refinement

The initial model was refined with simulated-annealing (CNS) to remove model bias and achieved $R = 0.329$ and $R_{free} = 0.370$ with data in the resolution range 50 to 3.0 Å. The F_o-F_c map showed no density for the active site metal ion or coordinated water. Strict six-fold NCS constraints were applied. A summary of refinement statistics is given in Table 3.2.

Figure 3.1 shows the active site electron density after the active site amino acid residues have been built. There is no new density for the active site metal in the electron density; it is therefore concluded that the N_3^- has removed the metal from the active site. The structure was abandoned at this point.

Other software used: MTZDUMP and MTZUTILS for inspection and modification of MTZ files, IPDISP for display of image-plate frames, HKLVIEW for display of processed data, BAVERAGE for calculation of average B values, MTZ2VARIOUS for conversion of file formats, RASMOL for display of molecules, XLOGGRAPH, PDBSET [CCP4 1994], LSQMAN [Kleywegt 1996] for calculation of molecular superpositions, MOLEMAN [Kleywegt 1999] and DATAMAN [Kleywegt & Jones 1999] for manipulation of coordinate and data files,

and MAPPAGE [Roussel & Cambillau 1989] for preparation of electron density maps for TURBO-FRODO.

Table 3.2. Refinement statistics for *Mt*-FeSOD grown from 100 mM azide

Reflections	33314
Free *R* factor reflections	1663
Resolution limits	50-2.6 Å
R factor	0.329
Free *R* factor	0.340
Total number non-H atoms	9930
Number water molecules	0

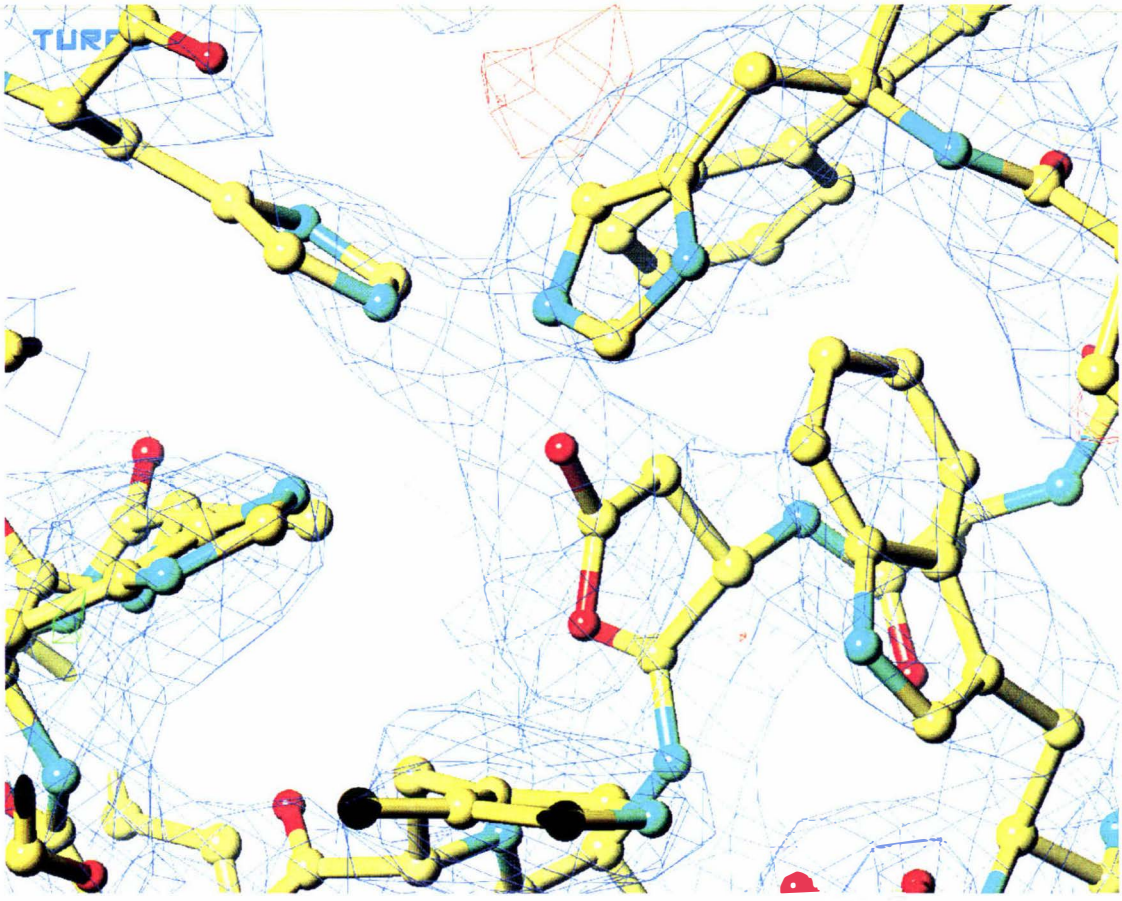


Fig 3.1

Active site electron density, (Blue $2F_o$-F_c at 1.2 σ)

(Red F_o-F_c at -3σ) (Green F_o-F_c at 3σ)

Note the absence of any difference electron density at the iron site.

Chapter 4

The structure of *Methanobacterium thermoautotrophicum* iron superoxide dismutase at 2.6 Å resolution, crystals grown from a solution containing 20 mM NaN_3 and 5 mM $FeCl_3$

4.1 Crystallization.

The protein used in this crystallization was from the same preparation as in chapter 2.

Crystals were grown by the hanging drop method at 4°C; each drop consisted of 1 μL protein (20 mg mL^{-1} in water) and 1 μL of well solution. The well solutions contained either 100 mM $NH_4CH_3CO_2$, pH 5.5, 20 mM NaN_3, 5 mM $FeCl_3$, and 20 % PEG 6000, or 100 mM $NH_4CH_3CO_2$, pH 5.5, 20 mM NaN_3, 5 mM $FeCl_3$, and 18 % PEG 5000MME. Crystals grew to a size of 200x100x50 microns in approximately three weeks. Crystals grown from PEG 6000 were used for data collection.

4.2 Data collection.

A single crystal of *Mt*-FeSOD was mounted in a nylon loop, with a cryoprotecting solution consisting of the well solution from which the crystal was grown and 15-25 % glycerol. The crystal was soaked in the cryoprotectant for approximately 30-50 seconds then snap frozen in the cryo-stream (Oxford Cryosystems). Data were collected on a Rigaku RU200 rotating-anode generator running at 1.1 kW Cu-Kα, 0.1 mm focus, fitted with an AXCO SPA 50 mono-capillary X-ray optic coupled

with a Rigaku RAxis IIC image-plate detector. X-ray frames were taken of the crystal (30 minute, 0.5° oscillation).

Table 4.1. Data collection statistics for *Mt*-FeSOD at 2.6 Å resolution.

Space group	*C*2	
Unit cell	$a = 187.62$ Å	$\alpha = 90°$
	$b = 114.08$Å	$\beta = 94.95°$
	$c = 58.09$ Å	$\gamma = 90°$
Data collection		
Observed reflections	320991	
Unique reflections	34876	
Completeness	92.9 %	(93.2 %)[a]
R_{merge} on I	0.128	(0.547)[a]
Overall I/σ	7.25	(1.7)[a]
Wilson B (Å^2)	34.925	
Redundancy	9.2	(3.0)[a]

[a] Outermost shell (2.69 - 2.60 Å)

The cell and orientation matrix were determined by the auto-index procedure in DENZO [Otwinoski & Minor 1997]. The resulting data files were scaled and merged together with SCALEPACK [Otwinoski & Minor 1997]. Unit cell dimensions were averaged over all data. Measured intensities were converted to amplitudes using the program TRUNCATE [CCP4 1994]. The R_{free} set from the original structure (ja076) was transferred to the new data set with the program DATAMAN [Kleywegt & Jones 1999] to ensure that the R_{free} data set was unbiased with respect to the original model. As the resolution of the new data set was higher than the old set, the R_{free} set was completed by random assignment of 5%

of the outermost shells to the R_{free} set. Relevant data collection statistics are in Table 4.1.

4.3 Solution

The previously determined structure with the side chains of residues coordinated to the metal truncated to Ala, the metal atom removed and water molecules deleted was subjected to rigid body refinement (CNS) [Brunger *et al.* 1998], leading to *R*-factors of $R = 0.279$ and $R_{free} = 0.281$.

4.4 Refinement

The model was rebuilt with TURBO-FRODO [Roussel & Cambillau 1989] Then simulated-annealing (as described in chapter 2) with CNS achieved $R = 0.249$, $R_{free} = 0.256$ with data in the resolution range 50 to 2.6 Å. Several rounds of model building followed by *xyz* and individual *B*-factor refinement brought the R and R_{free} to 0.240 and 0.251 respectively. A total of 360 water molecules was added with the "water-pick" procedure of CNS based on peak heights of 2σ in a $2F_o$-F_c map. All water molecules were *xyz*-fixed and *B*-factors only were refined. Water molecules with poor density or unfavourable protein contacts were deleted. In the final cycle of refinement the peptide pond planarity restraint was greatly relaxed to allow the peptide bonds to find their minima in the electron density. The final model with individually refined but tightly restrained *B*-factors brought R and R_{free} to values of 0.221 and 0.235. During all stages of refinement, high-energy six-fold NCS restraints were applied to the positional and thermal parameters of main chain atoms, and low-energy six-fold NCS restraints were similarly applied to the side chain atoms. The form of the restraint was

$$\Sigma_i^{\text{reference molecule}} \Sigma_j^{NCS} w_i(x_{ij} - \langle x_i \rangle)$$

where x_{ij} is the parameter I in molecule j, $\langle x_i \rangle$ is the mean parameter after superposition of the NCS-related molecules (here NCS=6), and w_i is the weight applied.

Representative electron density and real space R-factors are shown in Figures 4.2 and 4.3, respectively. The RMS deviations for bond lengths and angles are 0.008 Å and 0.9° respectively. A summary of refinement statistics is given in Table 4.2.

Other software used: MTZDUMP and MTZUTILS for inspection and modification of MTZ files, IPDISP for display of image-plate frames, HKLVIEW for display of processed data, BAVERAGE for calculation of average B values, MTZ2VARIOUS for conversion of data file formats, RASMOL for display of molecules, XLOGGRAPH, PDBSET [CCP4 1994], LSQMAN [Kleywegt 1996] for calculation of molecular superpositions, MOLEMAN [Kleywegt 1999] and DATAMAN [Kleywegt & Jones 1999] for manipulation of coordinate and data files, and MAPPAGE [Roussel & Cambillau 1989] for preparation of electron density maps for TURBO-FRODO.

Table 4.2. Refinement statistics for *Mt*-FeSOD at 2.6 Å resolution.

Reflections	33107
Free *R* factor reflections	1769
Resolution limits	50-2.6 Å
R factor	0.221
Free *R* factor	0.235
Total number non-H atoms	10404
Number water molecules	360
R.m.s deviations from ideals	
Bond lengths	0.008 Å
Bond angles	0.96°
Dihedral angles	21.6°
Improper angles	0.7°
Average *B*-factors (Å^2)	
Main chain	24.9
Side chain	26.8
Water molecules	28.0
Ramachandran plot	Figure 4.1
Most favoured (%)	87.8
Allowed (%)	10.6
Generously allowed (%)	1.7
Disallowed (%)	0.0

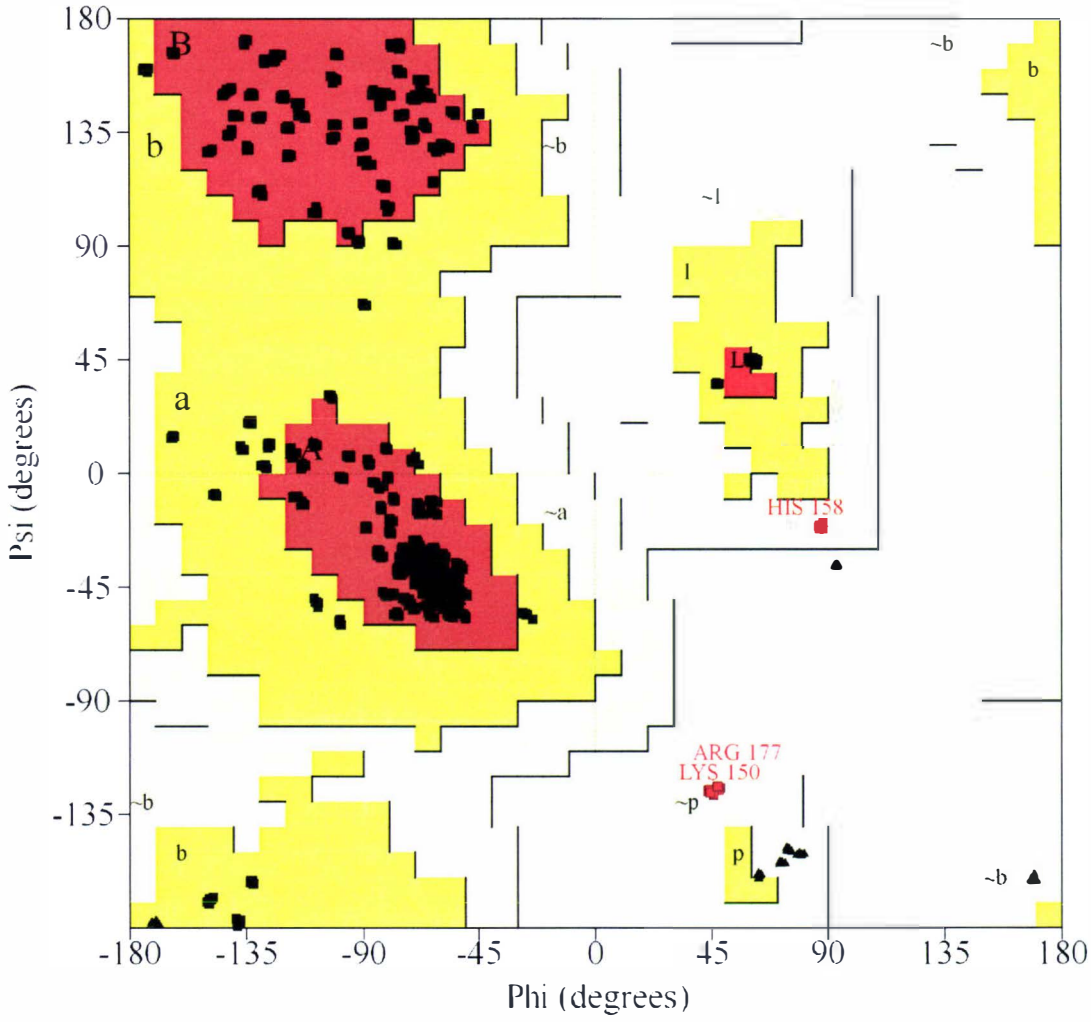


Figure 4.1

Ramachandran plot for *Mt*-FeSOD at 2.6 Å resolution after final refinement.

Labelled residues are K150, H158 and R177 (these outliers are discussed in full in section 5.3). Glycines are shown as triangles.

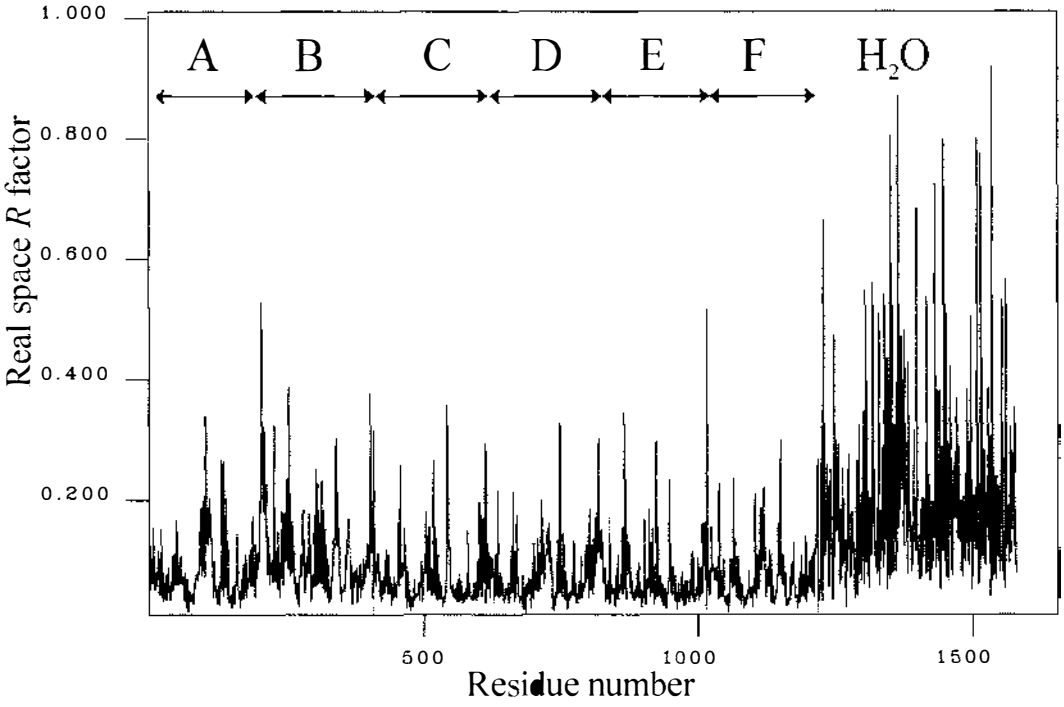


Figure 4.2

Real space *R*-factors per residue for *Mt*-FeSOD at 2.6 Å resolution, covering chains A-F and water molecules.

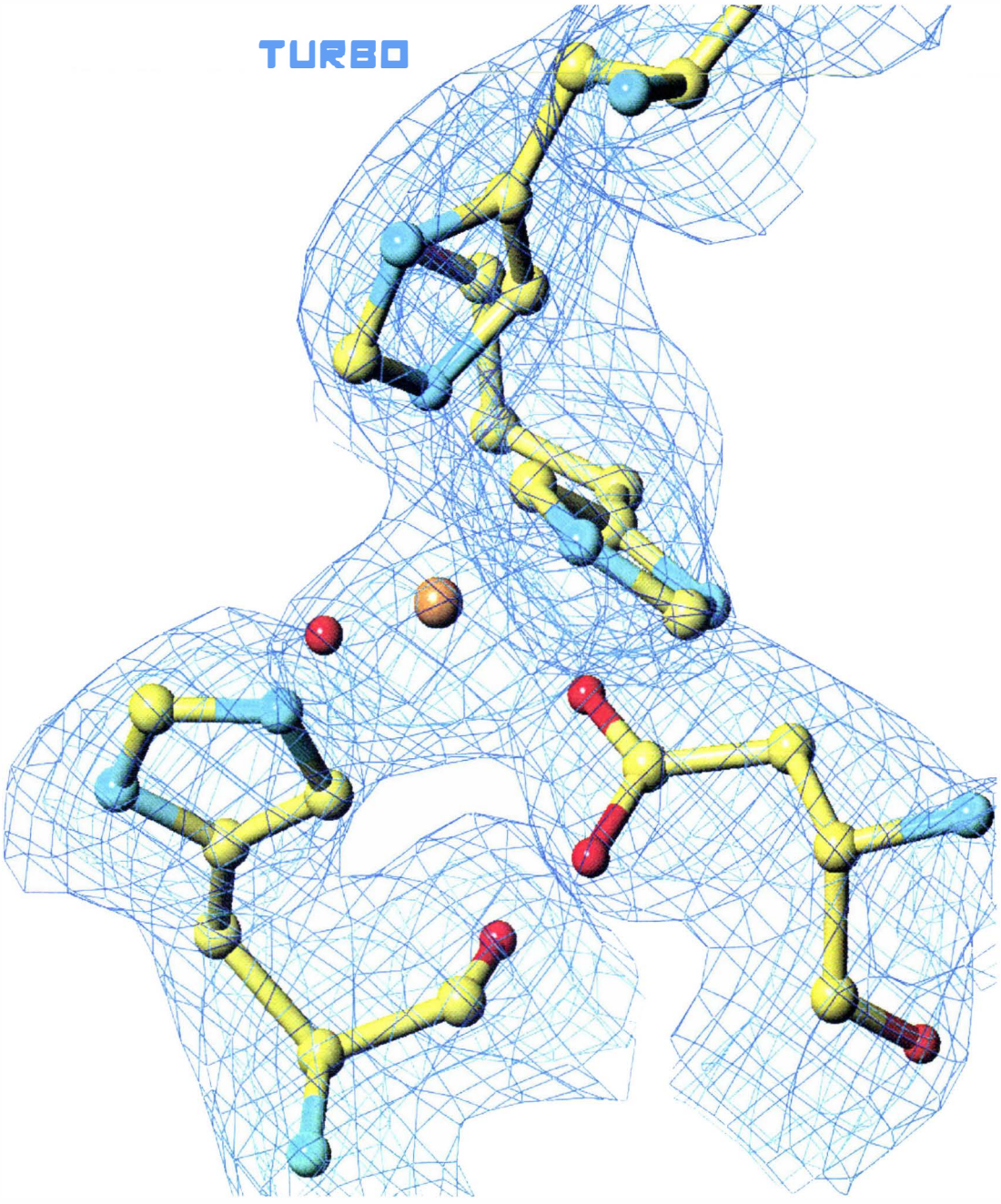


Figure 4.3

Typical electron density ($2F_o$-F_c) around the active site of *Mt*-FeSOD. The electron density is contoured at 1.2 σ (dark blue) and 3.0 σ (light blue).

The iron atom is shown as an orange sphere and the coordinated solvent species as a smaller red sphere.

Chapter 5

Results and Discussion

5.1 Structure determinations of *Mt*-FeSOD

Several challenges were overcome to successfully complete the structure of *Mt*-FeSOD. Crystallization of the protein proved to be difficult due to slight inhomogeneity of the protein sample. Although the protein was homogenous by SDS-PAGE, by reverse phase (C18 Phonominix Jupiter, linear gradient (60 Min.) H_2O 0.1% TFA – Acetonitrile 0.08% TFA) HPLC a small amount of suspected denatured material was detected. This material could not be removed by either anion exchange chromatography nor by hydrophobic interaction chromatography. Therefore, the protein sample had to be used as-is. After a number of attempts a suitable crystallization condition at low temperature was found. However, only small crystals could be grown. These small crystals were extremely sensitive to X-rays. The first crystals trialed in the X-ray beam lasted less than three hours, and diffracted poorly to approximately 8 Å resolution. Further crystallization trials with sodium acetate as the buffer were undertaken. These crystals diffracted to higher resolution. The resolution achieved and data collection quality were aided greatly by an upgrade to the X-ray optics, in which a toroidal glass capillary optic was installed. This optic produces a very small (0.1 mm), intense, highly collimated beam of X-rays. The crystals were still very sensitive to X-rays and suffered free radical damage within twenty-four hours. The very intense, small beam allowed faster collection times, which maximised the amount of data collected before the crystal was destroyed (Chapter 2). After the original model was refined, crystallizations were undertaken to co-

crystallize *Mt*-FeSOD with azide so we had a more complete picture of the biophysical properties associated with azide insensitivity. Crystallizations with 100 mM NaN_3 yielded crystals that diffracted to higher resolution (2.6 Å), and no longer showed sensitivity to X-ray radiation. This was found to be due to the total removal of the metal from the active site. Therefore, in the absence of metal there was source from which electrons could be stripped to produce radical species. It became apparent that the equilibrium constant for the iron azide complex ($Fe(OH_2)_5(N_3)^{2+}$) in solution ($K_1 \sim 10^5$)[Neves & De Andrade 1986; Avsar 1980], was higher than the binding constant for the metal to the protein (Chapter 3), and that the metal was labile on the time scale of crystal growth (several weeks). To counteract this equilibrium, a third set of crystallizations was undertaken with 20 mM NaN_3 and 5 mM $FeCl_3$. These crystals also diffracted to 2.6 Å resolution, and showed little sensitivity to X-rays. Fortunately the reasons for the improved diffraction and lack of sensitivity toward the X-rays was the exact opposite of the previous data collection. Now the active site was completely metalated (Chapter 4). The homogeneity of the active site reduced the susceptibility toward X-rays perhaps by having the normal redox recovery pathways for the enzyme, facilitated by metal-metal communication across the canonical dimer interface still available (Section 1.8 The superoxide dismutase cycle). Azide was not bound, consistent with prior observations that azide does not bind [Takao 1990]. Careful comparison of the *Mt*-FeSOD structure with the Fe-SOD-azide and Mn-SOD-azide structures, detailed later, showed that steric crowding around the active site is responsible for the very low azide affinity, which is unusual for FeSODs. The last data set proved to contain excellent data and a well-defined model resulted.

Funding for this project was provided in part by a Marsden grant.

5.2 Description of the asymmetric unit

The asymmetric unit contains six monomers. Four of the six subunits in the asymmetric unit (C, D, E, and F) form a tight tetramer with 222 NCS symmetry. The remaining two subunits (A and B) form a dimer (the canonical dimer formed in, for example, the dimeric MnSOD from *E. coli* [Edwards *et al.* 1998]). This dimer forms a tetramer (similar to the CDEF tetramer) about a crystallographic two-fold axis. The CDEF tetramer superimposes on the ABA′B′ with an RMS difference of 0.174 Å. The AB dimer superimposes on the CD dimer, and on the EF dimer with RMS deviations of 0.11 Å and 0.12 Å, respectively.

5.3 Quality of the model

The $2F_o$-F_c map, contoured at 1σ level, is well defined across all subunits of the asymmetric unit. Final F_o-F_c maps showed few uninterpretable holes or peaks in the electron density. The six-fold NCS restraints yield substantial improvements to electron density maps, effectively giving 3.0 Å resolution maps a 2.4 Å resolution appearance (Figure 4.3). This is achieved by a reduction in the number of parameters being refined. As each monomer is tightly restrained to be the same as the other monomers, this effectively means only one monomer is being refined. When electron density maps are being calculated the map is averaged over all six subunits, which improves the appearance of the map. The final values for R and R_{free}, are acceptable when considered in terms of the NCS and resolution of the data. The Ramachandran plot [Ramachandran & Sasisekharan 1968] showed 87.8% of the residues in the most favoured regions, as defined by PROCHECK [CCP4 1994]. Three residues (Lys

150, His 158 and Arg 178) lie near or just outside the border separating the generously allowed from the disallowed regions. Typical of other Fe and MnSOD structures, Lys 150 is the residue immediately prior to the residue whose side chain hydrogen-bonds to the coordinated solvent species. Arg 178 is similarly structurally conserved and Ramachandranly unfavoured in other Fe and MnSOD structures. His 158, as described in the next section, forms part of the dimer-dimer interface, as also do Lys 150 and Arg 178.

5.4 Overall structure

Mt-FeSOD has a classical Mn/Fe SOD secondary and tertiary structure, as shown in Figure 5.1, where *Mt*-FeSOD is superimposed on the mesophilic *E. coli* FeSOD structure (*Ec*-FeSOD, PDB ID: 1isa) [Lah *et al.* 1995]. The dimeric quaternary structure of *Ec*-FeSOD and other Mn/Fe SODs is also conserved in *Mt*-FeSOD. Among structurally characterised Mn/Fe SODs, *Mt*-FeSOD is most closely similar in primary, secondary, tertiary and homotetrameric quaternary structure to the FeSODs from *S. solfataricus* (PDB ID: 1sss) [Ursby *et al.* 1999], and *S. acidocaldarius* (PDB ID: 1b06) [Knapp *et al.* 1999]. In comparison to the dimeric *Ec*-FeSOD, *Mt*-FeSOD has a two-and-a-half turn extension at the C-terminal end of helix α1 and a one-turn extension at the N-terminal end of helix α2. The C-terminal extension of helix α1 is common to all tetrameric Fe/MnSODs, both thermophilic and mesophilic, with the exception of the mitochondrial MnSOD [Borgstahl *et al.* 1992], which adopts an alternative tetrameric association. The extension of helix α2, on the other hand, appears to be common only to Fe and Mn SODs from thermophiles and hyperthermophiles. These extensions to helices α1 and α2 remove the loop-α-loop motif, which is common to the mesophilic

SODs. The loop-α-loop motif blocks the tetramerization site in these SODs. The other main point of difference between the tertiary structures of thermophilic and mesophilic SODs occurs in the extended loop between β2 and β3, which lacks a single helical turn in *Mt*-FeSOD. This loop, which is rather variable in sequence and structure among Fe and Mn SODs, forms part of the hydrophobic region of the dimer-dimer interface. From β3 to the C-terminus the tertiary structure is very well conserved among all Fe and Mn SODs (Figure 5.1).

The Ramachandranly unfavoured Lys 150, which has phi ≈ 45° and psi ≈ -135°, is the first residue of an extremely tight two-residue hairpin turn, which accommodates His 151, one of the outer sphere residues. His 151 makes a number of hydrogen bonds to active site ligands and other outer-sphere residues, the most important being to the active-site solvent-derived species (H151_CE1-WAT205_O (2.94Å)). In *Mt*-FeSOD, unlike other Fe and MnSODs, the peptide following this tight turn hydrogen-bonds back to the C-terminal end of the β2 strand. Lys 150 also makes up part of the dimer-dimer interface, whereby K150_NZ makes a bifurcated salt bridge with E149_OE1 and E149_OE2 (across the crystallographic two-fold axis for the AB dimer). E149_O makes a bifurcated hydrogen bond to N152_N and also to Y153_N, as part of this hairpin β-turn. K150_N makes a hydrogen bond to S129_O; this association locks the β hairpin turn down onto the rest of the protein. The above hydrogen-bonding network constitutes a new motif for this region of the SOD structure. The equivalent residue to Lys150 in *Ec*-FeSOD [Lah *et al.* 1995]), Asn 140, has a similar disallowed Ramachandran conformation; however, the hydrogen bonding network for this residue is different, as there is not a tight turn associated with this residue. K150_O makes a hydrogen bond with W131_NE1, one of the conserved outer sphere

tryptophan residues. W131_NE1 itself has a bifurcated hydrogen-bonding environment, also hydrogen bonding to H151_ND1, which hydrogen bonds to the coordinated solvent species through atom H151_CE1. This is a non-classical hydrogen bond and will be discussed in section 5.6. This arrangement at W131_NE1 is a conserved feature of Fe and MnSODs, with the equivalent residues in *Ec*-FeSOD forming a similar network.

His 158 falls in the generously allowed region of the reverse α-helix conformation of Ramachandran space. His 158 is a surface residue, which through atoms H158_ND1 and H158_O makes hydrogen bonds across the dimer interface with R160_NH1 and R160_NH2. In other Fe and MnSOD structures the residue equivalent His 158 falls in the allowed regions of the Ramachandran plot.

Arg 177 falls into the disallowed region of the Ramachandran plot (phi ≈ 45°, psi ≈ -135°). Arg 177 is the C-terminal residue of the α6 helix in the C-terminal α-β region of the tertiary structure. Arg 177 forms a tight turn and has some unusual interactions: R177_O makes a π interaction with the phenyl moiety of Tyr 173. Structurally equivalent residues in both Fe and MnSODs also fall into a similar "forbidden" region of the Ramachandran plot.

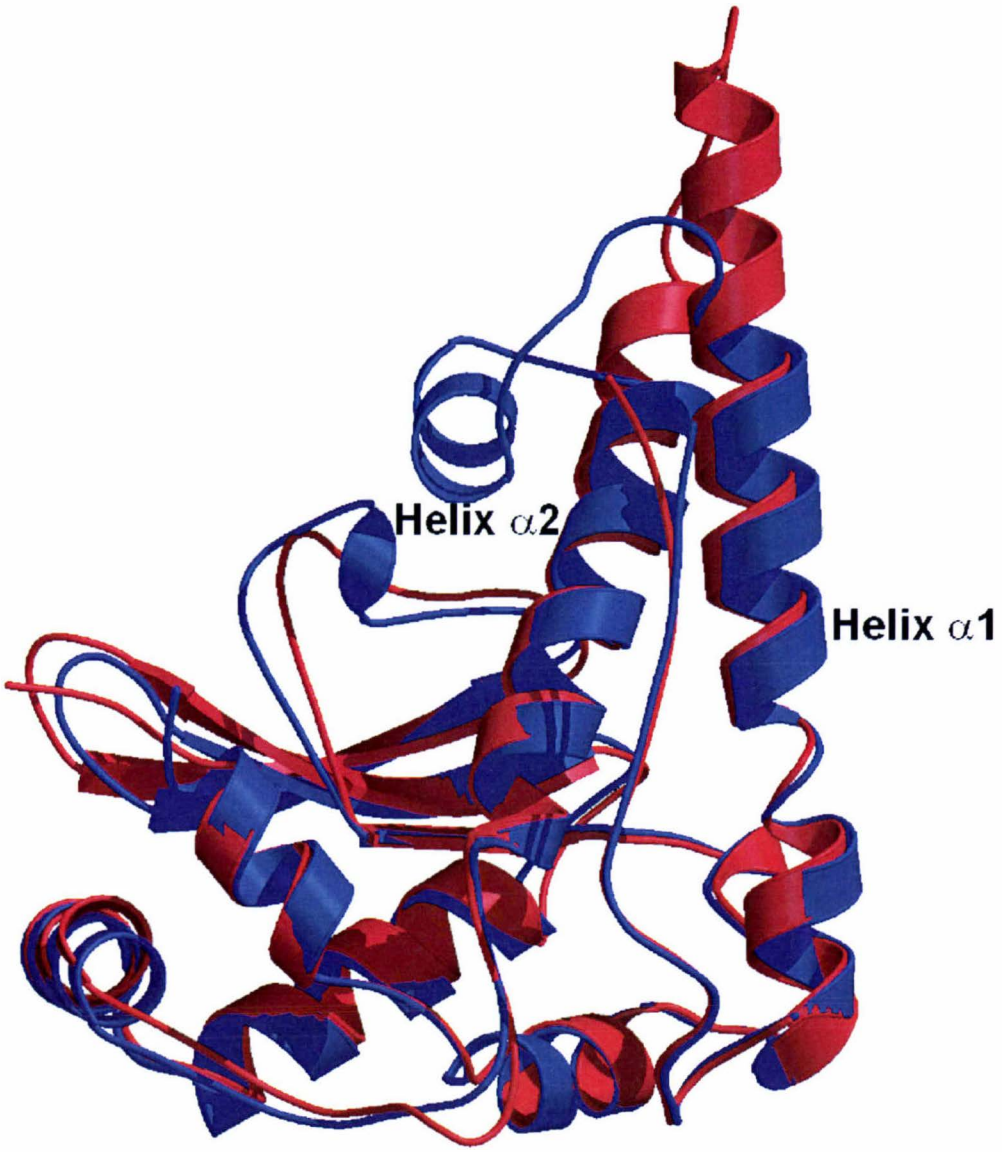


Figure 5.1

Structural alignment of an *E. coli* FeSOD (1isa) monomer (BLUE), and *Mt*-FeSOD monomer (RED). Note the extended helices α1 and α2 in *Mt*-FeSOD which removes the loop-α-loop motif which blocks the tetramerization site in *Ec*-FeSOD

5.5 Surface properties

The surface of the protein is almost entirely negatively charged, with only a few positively charged spots, and almost no aliphatic patches. This is unusual for Mn or Fe SODs, as most SODs have a somewhat mottled surface charge density of positive and negative charges and aliphatic patches. The high charge density of *Mt*-FeSOD is also evident in the pI of the protein (pI = 5.2) whereas most SODs have a pI that is above 7.0. This negatively charged exterior may offer protection to the protein from non-specific chemical or enzymatic attack, as the high negative charge density could repel the attacking species.

Mt-FeSOD has a smaller than usual interface between the subunits that form the functional dimer. This interface has a surface area of ~560 $Å^2$, which corresponds to only 6.3% of the protein surface as measured with GRASP [Nicholls *et al.* 1991]. Eubacterial Fe and Mn SODs more typically have corresponding dimer interfaces constituting ~10% of the total surface area [Lah *et al.* 1995]. The functional dimer has two buried salt bridges. The first, H170A-E169B, is a highly conserved feature of the functional dimer in all Fe and Mn SODs. Indeed, in only four putative Fe or Mn SODs out of more than 200 sequences is a residue other than glutamate seen at the position equivalent to E169: *Oryza sativa* (Rice) (SwissProt entry: Q9SNQ0, E→Q and Q9LWS3, E→S), *Schizosaccharomyces pombe* (Fission yeast) (O42919, E→N and O74379, E→Q). This salt bridge may facilitate redox communication between the active sites in the functional dimer [Edwards *et al.* 1998]. The other salt bridge, E127A-K150B is remote from active site and appears to provide stabilization only. Lastly there is an interdimer hydrogen bond H36A-Y173B which is a well-conserved structural feature in other SODs. Disruption of this hydrogen bond by mutation of either

residue has been shown to greatly affect the enzyme activity and/or the dimer interface [Edwards *et al.* 2001]. The dimer-dimer interface for *Mt*-FeSOD has five salt bridges: R142-D52, R55-E118, E69-K64, K64-E127, E127-K68 and a hydrogen bond C137-Y8. Each of these bonding pairs is present four times in the dimer-dimer interface. These interactions contribute to a much larger 855 $Å^2$ hydrophobic contact zone per monomer. The total surface area buried is 560 + 855 = 1415$Å^2$, with a total buried area of 5660$Å^2$ per tetramer. As shown in Figure 5.2, the tetramer unit is essentially a square box of protein with the substrate access funnels being the only major invaginations of the surface. This compact design and large hydrophobic interface affords the tetrameric protein its thermal stability.

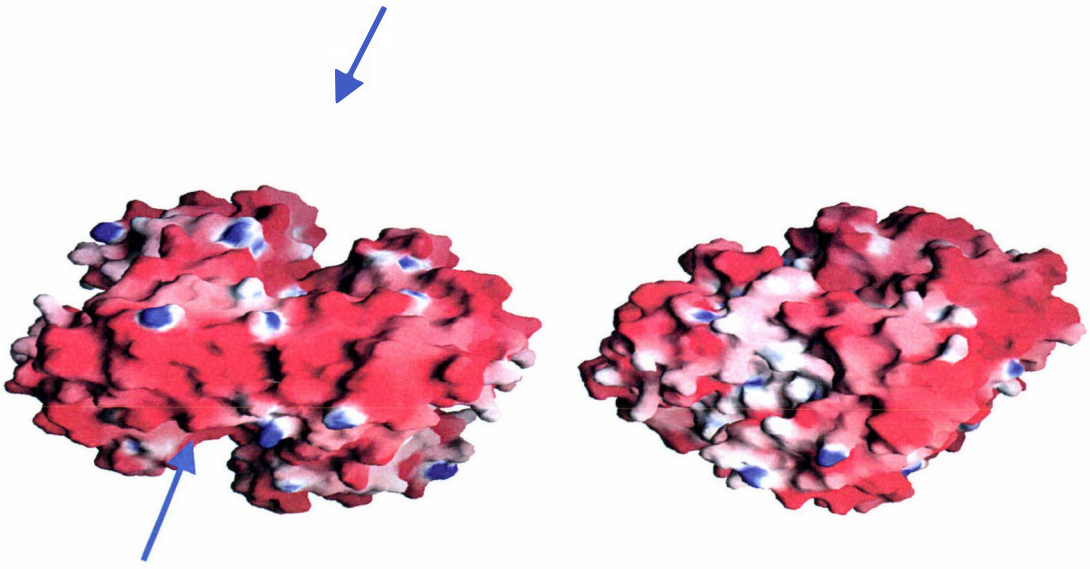


Figure 5.2

Orthogonal views of the electrostatic potential of *Mt*-FeSOD, the colours are contoured at +15 (blue, positive) to -15 kT (red, negative). The substrate access funnels are arrowed. This figure was produced with GRASP [Nicholls *et al.* 1991]

5.6 Active site and its outer sphere

The active site and outer sphere regions of all twelve Fe and Mn SODs that have been structurally characterised to date, including *Mt*-FeSOD (Figure 5.3), are closely isostructural. The Fe (or Mn) center is in a distorted trigonal bipyramidal geometry with His80_NE2, His_170_NE2 and Asp166_OD2 forming the trigonal plane, and His32_NE2 and a coordinated solvent species (here labeled HOH205) forming the trigonal axis.

The coordinated solvent species, generally assumed to be hydroxide, at least in the M^{III} form of the SOD, forms a non-classical C-H...OH hydrogen bond with His 151 (His151_CE1...HOH205, 3.1 Å). Atom NE2 of His 151 hydrogen bonds to Tyr40_OH, but with poor geometry

(Y40_OH...H151_NE2, 3.3 Å; Y40_CZ—Y40_OH...NE2_H151, 71°). Atom ND1 of His 151 hydrogen bonds to Trp 131 with good geometry (H151_ND1...NE1_W131, 3.0 Å; H151_ND1...NE1_W131—CD2_W131, 139°). In most MnSODs, glutamine, rather than histidine, hydrogen bonds through the NE2 amine moiety to the coordinated solvent species and to the structural equivalent of Tyr 40. The carbonyl moiety, OE1, of the glutamine hydrogen bonds to the structural equivalent of Trp 131. Tyrosine at position 40 is very strongly conserved across all Mn and Fe SODs, with exceptions found for the probable SODs from *Borrelia burgdorferi* (SODM_BORBU, Y→F), *Alcaligenes eutrophus* (*Ralstonia eutropha*) and *Ralstonia metallidurans* (SODM_ALCEU and CAC42412, Y→A) and *Agaricus bisporus* (Common mushroom) (SODM_AGABI, Y→L), and putative SODs from *Schizosaccharomyces pombe* (O42919, Y→V and O74379, Y→L) [Altschul *et al.* 1997; Bairoch & Apweiler 2000]. Tryptophan at position 131 is also close to being absolutely conserved across all Mn and Fe SODs: exceptions are found for probable superoxide dismutases from *Trypanosoma cruzi* (Q27791 and O02616, W→L) and from pea (SODM_PEA, four-residue deletion) [Altschul *et al.* 1997; Bairoch & Apweiler 2000].

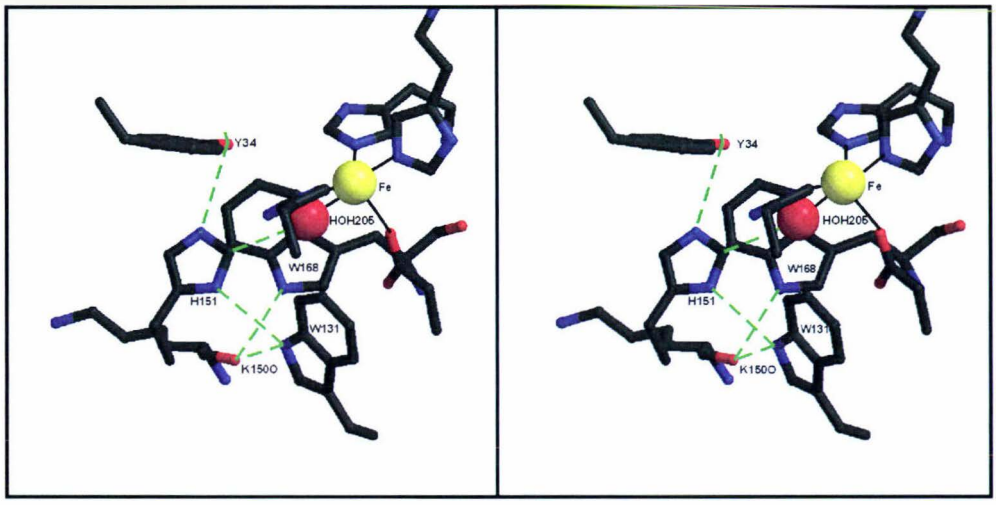


Figure 5.3

Stereo diagram of the active site and outer-sphere of *Mt*-FeSOD. Note the hydrogen-bonding network (GREEN) around H151 and the distorted trigonal bipyramidal geometry of the iron (yellow)

5.7 Peroxide resistance

Typically FeSODs, but not MnSODs, are rapidly and irreversibly inactivated by the presence of sub-millimolar concentrations of hydrogen peroxide, one of the products of superoxide dismutation [Gregory & Dapper 1980; Asada *et al.* 1975; Yamakura & Suzuki 1986]. The FeSOD from *E. coli* can be used as a model for a peroxide-sensitive FeSOD and the MnSOD from *E. coli* as the model for a peroxide-insensitive MnSOD. Fenton chemistry (Figure 5.4) at the active-site metal is believed to produce the reactive species, but the propagation of those reactive species is, in this case, a more important factor. Discussions of peroxide sensitivity have identified a surface tryptophan residue (equivalent to Trp 71 in *Ec*-FeSOD) [Beyer & Fridovich 1987], which is conserved across some eubacterial FeSODs, including all structurally characterized FeSODs, but typically is not present in MnSODs [Parker & Blake 1988]. For the FeSODs from *E. coli* and *Ps. ovalis*, H_2O_2-induced loss of activity

is associated with damage to tryptophan residues near the metal site, Trp 122 and Trp 158 in the *Ec*-FeSOD numbering scheme. Trp 122 is buried, and although Trp 158 is partly solvent-exposed, Trp 158 lies at the base of the substrate-access funnel and is part of the dimer interface. As a consequence, Trp 158 is not accessible to *inter*molecular contact by FeSOD molecules. However, in *Ec*-FeSOD and other H_2O_2-sensitive FeSODs, Trp 122 forms part of a potential radical cascade that extends from the active site to the exposed accessible surface of the protein, where radicals can be passed on intermolecularly to other SOD molecules. This cascade involves Trp 122, which lies in close proximity (~4 Å) of the metal and is in hydrogen-bonding contact *via* Q69 (His 145 in some species) with the coordinated solvent species. Trp 122 is also in close contact with Trp 124 (W124_NE1...CB_W122, 3.5 Å). In turn, Trp 124 is in close proximity to the surface-exposed tryptophan, Trp 71 (W71_CZ3...CZ3_W124, 4.3 Å). With a small torsional twist about the CA-CB bond of Trp 71, this contact can be reduced to ~3.0 Å. Table 5.1 summarizes for a range of Fe and Mn SODs the residues of this putative radical cascade, and peroxide sensitivity of these SODs.

$$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+}\text{-}OH^- + OH\bullet$$

Figure 5.4

Fenton chemistry.

A reaction to produce the hydroxyl radical.

Table 5.1 Radical pathway residues for select SODs

SOD	Fe	Mn	71	74	75	124
Fe-SOD	Sensitive ?	Sensitive ?				
Fe_Ec	Y		W^{71}	T^{74}	F^{75}	W^{124}
Fe_Tpyr	Y		V^{68}	W^{71}	I^{72}	W^{124}
Fe_Ssol	~Y		K^{82}	A^{85}	L^{86}	V^{137}
Fe_Mtub	Y		V^{73}	T^{76}	I^{77}	A^{126}
Fe_Apyr	Y		V^{79}	E^{82}	L^{83}	I^{130}
Fe_Mthe	N		V^{78}	E^{82}	L^{83}	V^{133}
Cambialistic						
Fe_Psher	N	N	V^{73}	S^{76}	V^{77}	S^{128}
V73W	Y	N	W^{73}	S^{76}	V^{77}	S^{128}
Fe_Bgin	Y	N	L^{71}	N^{74}	L^{75}	W^{124}
Mn-SOD						
Mn_Ec		N	A^{79}	S^{82}	L^{83}	W^{130}
Mn_Bste		N	A^{79}	S^{82}	L^{83}	W^{132}

Fe_Ec: FeSOD from *Escherichia coli*; PDB ID: 1isa [Beyer & Fridovich 1986; Lah *et al.* 1995]

Fe_Tpyr: FeSOD from *Tetrahymena pyriformis*; [Barra *et al.* 1990]

Fe_Ssol: FeSOD from *Sulfolobus solfataricus*; PDB ID: 1sss [Ursby *et al.* 1999; Yamano *et al.* 1999; Dello Russo *et al.* 1997]

Fe_Mtub: FeSOD from *Mycobacterium tuberculosis*; PDB ID: 1ids [Kusunose *et al.* 1976; Kang *et al.* 1998; Cooper *et al.* 1995]

Fe_Apyr: FeSOD from *Aquifex pyrophilus*; PDB ID: 1coj [Lim *et al.* 1997]

Fe_Mthe: *Mt*-FeSOD; This work

Fe_Psher: Cambialistic SOD from *Propionibacterium freudenreichii shermanii* [Schmidt *et al.* 1996; Meier *et al.* 1994b; Meier *et al.* 1998]

V73W: Mutant of the cambialistic SOD from *Propionibacterium freudenreichii shermanii* [Gabbianelli *et al.* 1997]

Fe_Bgin: Cambialistic SOD from *Porphyromonas* (*Bacteroides*) *gingivalis* [Hiraoka *et al.* 2000]

Mn_Ec: MnSOD from *Escherichia coli*; PDB ID: 1vew [Edwards *et al.* 1998; Keele *et al.* 1968; Ose *et al.* 1976; Misra *et al.* 1978]

Mn_Bste: MnSOD from *Bacillus stearothermophilus*; PDB ID: not deposited [Yamakura *et al.* 1995; Brock *et al.* 1977; McAdam *et al.* 1977]

Unusually, the activity of *Mt*-FeSOD is relatively unaffected by the presence of hydrogen peroxide [Takao *et al.* 1991], but *Mt*-FeSOD, as previously discussed, has an MnSOD-like primary structure, and correspondingly, the residues at the positions equivalent to Trp 71 and Trp124 are both valines, Val 78 and Val 133. Similar behavior is exhibited by the FeSODs from *S. meliloti* and *Pr. shermanii*, which are also missing at least one tryptophan from the radical cascade. The presence of valines instead of tryptophans in *Mt*-FeSOD has two consequences: firstly valine is an aliphatic amino-acid and is not particularly reactive toward radical attack, and, secondly, valine is considerably smaller then tryptophan, and therefore, the closest approach of the two valine residues is 8.8 Å, thus providing no pathway for radical species to escape the immediate vicinity of the active site (Figure 5.4). The V73W mutant of the FeSOD from *Pr. shermanii* shows much greater sensitivity to H_2O_2 than does the wild-type enzyme, consistent with the role of this residue, equivalent to Trp 71 of *Ec*-FeSOD, in the radical cascade.

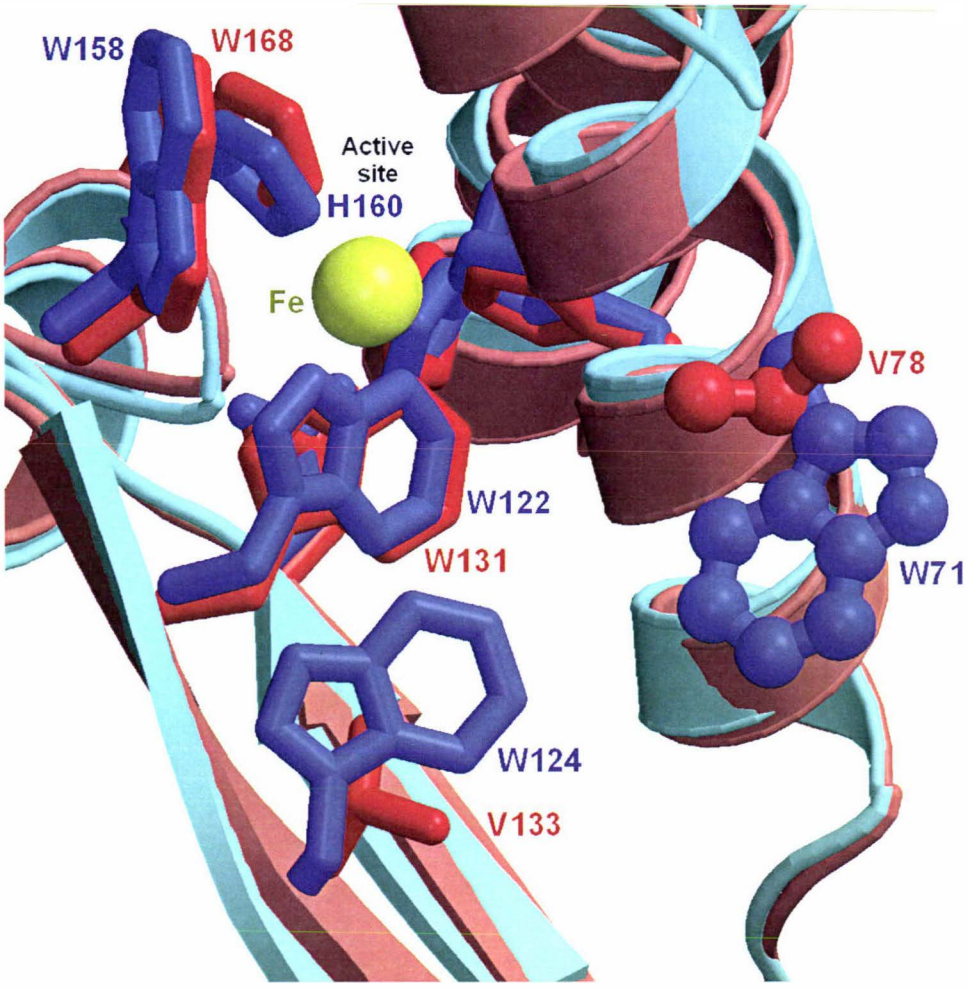


Figure 5.4

Structural alignment of the *E. coli* FeSOD (1isa) radical cascade (BLUE), and *Mt*-FeSOD (ORANGE). Note the absence in *Mt*-FeSOD of the last two links in the tryptophan radical cascade of *Ec*-FeSOD, from the active site to the protein surface.

Why, then, are the FeSODs from *Tetrahymena pyriformis*, *S. solfataricus*, *M. tuberculosis*, *A. pyrophilus*, *P.* (*Bacteroides*) *gingivalis*, *Bacteroides fragilis* and *Streptococcus mutans*, which all lack tryptophan at the position equivalent to Trp 71 of *Ec*-FeSOD, still susceptible to H_2O_2? Model building and study of the available structures have shown that there may be alternative radical cascades that would allow the propagation of the radical species to the surface of the protein. These

cascades involve tyrosine and tryptophan [Adams *et al.* 2002](Figure 5.5 and 5.6).

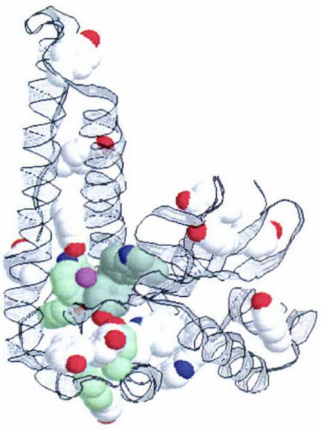


Figure 5.5

Radical cascade (green) for FeSOD from *A. pyrophilus*

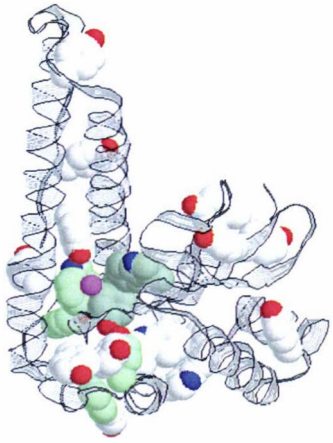


Figure 5.6

Radical cascade (green) for FeSOD from *Sulfolobus acidocaldarius*

All Trp and Tyr residues are shown in space filling mode and the active-site metal ion is shown in purple. Residues shown in green form putative alternative radical cascades from the active site to the surface of the protein.

Other explanations of peroxide sensitivity or insensitivity have focused on the redox tuning at the active site [Santos *et al.* 1999; Yamakura *et al.* 1998]. The measured redox potentials of Fe and Mn SODs have been shown to be almost identical in the resting state [Vance & Miller 1998; Barrette *et al.* 1983; Leveque *et al.* 2001]. There is no thermodynamic reason why Mn-SOD should not undergo Fenton chemistry. However, there are no reported examples of Mn enzymes being involved in Fenton chemistry. One explanation for this may be found in the superoxide reduction step of the SOD cycle (Species 3-4-1 or species 3-1, Figure 1.6, Chapter 1), where Mn-SODs have been shown to proceed by an inner-sphere mechanism at high substrate loadings (ie. The superoxide reduction reaction is less labile and substrate inhibition can occur) [Hearn

et al. 2001; McAdam *et al.* 1977; Pick *et al.* 1974]. Fe-SOD on the other hand does not appear to undergo substrate inhibition; therefore, the reaction only proceeds via the "outer sphere" concerted mechanism [McAdam *et al.* 1977; Bull & Fee 1985]. This implies that the electrons on the Fe center are more labile. With higher electron lability, Fenton chemistry may proceed more readily [Adams *et al.* 2002]. This is demonstrated by the cambialistic SOD from *B. gingivalis*, which is peroxide-insensitive with Mn at the active site and peroxide-sensitive with Fe at the active site [Hiraoka *et al.* 2000].

5.8 Azide resistance

Azide (N_3^-) is a competitive inhibitor of many iron superoxide dismutases' activity, with 50% inhibition occurring at concentrations of 5 mM azide or less. MnSODs on the other hand are only slightly inhibited by azide, with 50% inhibition requiring concentrations of more than 20 mM azide. *Mt*-FeSOD has a weak MnSOD-like response to azide [Takao *et al.* 1991].

In the FeSOD/azide structures (from E. coli (1isc) [Lah *et al.* 1995] and *Pr. shermanii* (1avm) [Schmidt *et al.* 1997]), azide binds to the active site in the sixth and vacant coordination site, with N1 interacting with the metal (N1-Fe 2.1 Å), and N2 and N3 interacting with the π cloud of His 73 (a ligand histidine) (N2-, N3-His 2.3-3.5Å). N3 also interacts with H31N (3.04Å), but with very poor geometry for a hydrogen bond. In the MnSOD/azide structurally characterised (*T. thermophilus* (1mng)[Lah *et al.* 1995]), azide binds in a different way. N1 interacts with the metal, and

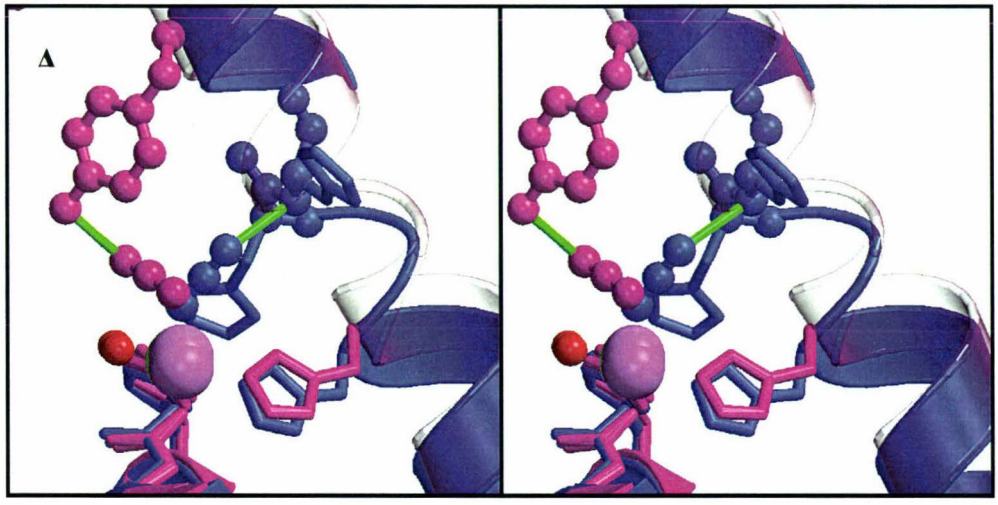
binds in the same position as in FeSOD. However N2 and N3 are positioned 90° with respect to N2 and N3 in FeSOD/azide structures. N2 does not interact with the protein, and N3 makes a hydrogen bond with Y36_OH (2.60Å) (isostructural to Y34 in *Ec*-FeSOD)(Figure 5.7A). As described in section 5.5, eubacterial FeSODs and MnSODs have one important difference in the active-site environment: the position of the residue that hydrogen bonds to the outer-sphere tyrosine. Geometrically this difference affects the angle that the hydrogen of Y34_OH makes with the arene ring of Y34. In FeSOD, where Q69_NE2 makes a hydrogen bond (3.0 Å) with Y34_OH, the hydrogen-bond is perpendicular to the arene ring. In MnSOD, Q146_NE2 makes the hydrogen bond with Y34_OH (3.0 Å), and this hydrogen bond is almost parallel with the arene ring. This has the effect of making slightly more room in the vicinity of Y34, allowing enough space to accommodate the N3 of the azide. Moreover, it also improves the geometry of the Y34_OH for hydrogen bonding to N3. These interactions are less strong than the π interaction that azide makes in FeSOD, and, hence, are consistent with the lower affinity with which azide binds to MnSODs. MnSOD does not bind azide in the pocket where FeSOD does due to steric interactions between the azide N3 and H32_CB. Modelling azide binding to *Mt*-FeSOD leads to a similar sterically forbidden interaction N3…H36_CA, and therefore *Mt*-FeSOD does not bind azide in the typical FeSOD pocket. *Mt*-FeSOD also cannot bind azide in the typical MnSOD binding pocket as the pocket has been blocked off by Y40_OH (Figure 5.7B). Tyr 40 cannot rotate about CA-CB bond to make enough room for the azide to bind as Tyr 40 is sterically constrained by the side chains of Phe 77 and Trp 168, which are in a rigid hydrophobic part of the structure.

Some FeSODs are only weakly inhibited by azide at standard assay

conditions of pH 7.8. A study of the inhibition of the cambialistic *Pr. shermanii* FeSOD as a function of pH and azide concentration concluded that the insensitivity of the enzyme to azide was due to competitive inhibition by hydroxide. At low pH OH^- does not bind and the enzyme becomes sensitive to azide [Meier *et al.* 1998].

The SODs from *A. pernix* and *S. solfataricus* require azide concentrations of at least 20 mM for 50% inhibition [Yamano & Maruyama 1999; Yamano *et al.* 1999]. The Q146H mutant of *Ec*-MnSOD does not bind azide at all, where as the wild type does [Edwards *et al.* 2001]. All these SODs have a common trait in which histidine is observed at the position equivalent to 146 (*Ec*-MnSOD numbering). *Mt*-FeSOD also has His at this position (151). This histidine makes a nonclasical hydrogen bond C-H..O to the water-derived ligand species. This hydrogen-bonding arrangement applies subtle electronic tuning to the metal environment and may lower the binding affinity for azide.

Mt-FeSOD has both a geometrical and electronic impediment to the binding of azide and, therefore, has an atypical response to azide.



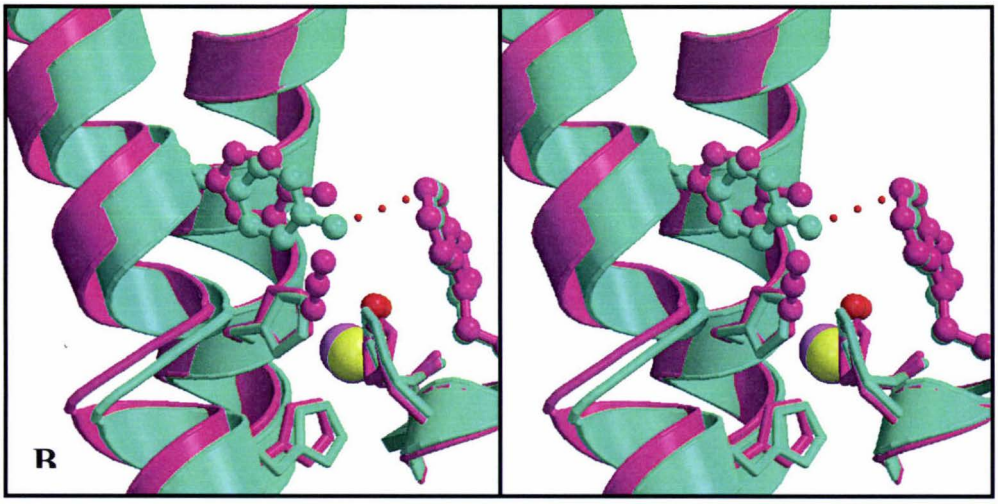


Figure 5.7

(A) Structural alignment of E. *coli* FeSOD/azide (1isc) monomer (BLUE), and T. *thermophilus* MnSOD/azide monomer (pink). Note the different orientation and hydrogen bonding of the azide.

(B) Structural alignment of *T. thermophilus* MnSOD/azide monomer (pink) and *Mt*-FeSOD (green). Note that helix α1 of *Mt*-FeSOD jumps out of register after the kink and forces a different position onto Y40 that prevents it moving to accommodate an azide. .

5.9 Concluding remarks

Mt-FeSOD is an atypical FeSOD. It is a highly expressed radical defence protein expressed in a strict anaerobe [Takao *et al.* 1991]: *Mt*-FeSOD is an iron-containing SOD, with an MnSOD-like primary structure, and it has MnSOD-like biophysical properties. *Mt*-FeSOD has an unusually highly charged exterior surface, an extremely tight dimer-dimer interface, and very tight tertiary structure, all designed to afford the protein both thermal and chemical stability. *Mt*-FeSOD is resistant against the common Fe and Mn SOD inhibitors, azide and hydrogen peroxide. From the X-ray structure we have determined, we conclude that protection against azide and hydrogen peroxide is a function of the primary structure and not of the metal cofactor bound. Other insights form this structure include the metal specificity of the whole class of Fe, Mn SODs. This work is published in Adams *et al.* 2002, and was not discussed in this thesis, as the work was a collaborative effort, and therefore did not constitute wholly my own work.

REFERENCES

Adams, J. J., Anderson, B. F., Renault, J. P., Vechère-Béaur, C., Morgenstern-Badarau, I., & Jameson, G. B. (2002). *J. Biol. Inorg. Chem.* submitted.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389.

Asada, K., Yoshikawa, K., Takahashi, M., Maeda, Y., & Enmanji, K. (1975). *J. Biol. Chem.* **250**, 2801.

Asada, K., Kanematsu, S., Okada, S. & Hayahawa, T. (1980). *Chemical and Biological Aspects of Superoxide Dismutase,* pp136-153, Elsevier. Amsterdam (NEL).

Avsar, E. (1980). *Acta Chem. Scand. Ser.* **A34**, 405.

Bairoch, A. & Apweiler, R. (2000). *Nucleic Acids Res.* **28**, 45.

Barra, D., Schinina, M. E., Bossa, F., Puget, K., Durosay, P., Guissani, A., & Michelson, A. M. (1990).). *J. Biol. Chem.* **265**, 17680.

Barrette, W. C., Sawyer, D. T., Fee, J. A., & Asada, K. (1983). *Biochemistry* **22**, 624.

Behar, D., Czapski, C., Rabani, J., Dorfman, L. M., & Schwarz, H. A. (1970). *J. Phys. Chem.* **74**, 3209.

Beyer, W. F., Jr., & Fridovich, I. (1987). *Biochemistry* , **26**, 1251.

Beyer, W., Imlay, J., & Fridovich, I. (1991). *Prog. Nucl Acid Res. Mol. Biol.* **40**, 221.

BLAST: Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403.

Bond, C. J., Huang, J-Y., Hajduk, R., Flick, K. E., Heath, P. J., & Stoddard, B. L. (2000). *Acta Cryst.* **D56**, 1359.

Borgstahl, G. E. O., Parge, H. E., Hickey, M. J., Beyer, W. F., Jr., Hallewell, R. A., & Tainer, J. A. (1992). *Cell* **71**, 107.

Bray, R. C., Cockle, S. A., Fielden, E. M., Roberts, P. B., Rotilio, G., & Calabrese, L. (1974). *Biochem. J.* **139**, 43.

Brock, C. J. & Harris, J. I. (1977). *Biochem. Soc. Trans.* **5**, 1537.

Brünger, A., Clore, M., Gros, P., Nalges, M., & Read, R. (1998). *Acta Cryst.* **D54**, 905.

Bull, A. & Fee, J. A. (1985). *J. Am. Chem. Soc.* **107**, 3295.

Bunting, K., Cooper, J. B., Badasso, M. O., Tickle, I. J., Newton, M., Wood, S. P., Zhang, Y., & Young, D. (1998). *Eur. J. Biochem.* **251**, 795.

Cade, P. E., Bader, R. F. W., & Pelletier, J. (1971). *J. Chem. Phys.* **54**, 3517.

Choudhury, S. B., Lee, J. W., Davidson, G., Yim, Y. I., Bose, K., Sharma, M. L., Kang, S. O., Cabelli, D. E., & Maroney, M. J. (1999). *Biochemistry* **38**, 3744.

Cooper, J. B., McIntyre, K., Badasso, M. O., Wood, S. P., Zhang, Y., Garbe, T. R., & Young, D. (1995). *J. Mol. Biol.* **246**, 531.

Czapski G. (1984). *Israel J. Chem.* **24**, 29.

Dello Russo, A., Rullo, R., Nitti, G., Masullo, M., & Bocchini, V. (1997). *Biochim. Biophys. Acta* **1343**, 23.

Deutsch, H. F., Hoshi, S., Matsuda, Y., Suzuki, K., Kawano, K., Kitagawa, Y., Katsube, Y., & Taniguchi, N. (1991). *J. Mol. Biol.* **219**, 103.

Edwards, R. A., Whittaker, M. M., Whittaker, J. W., Baker, E. N., & Jameson, G. B. (2001). *Biochemistry* **40**, 15.

Edwards, R. A. (1999). *Ph. D. dissertation.*

Edwards, R. A., Baker, H. M., Whittaker, M. M., Whittaker, J. W., Jameson, G. B., & Baker, E. N. (1998). *J. Biol. Inorg. Chem.* **3**, 161.

Fridovich, I. (1983). *Ann. Rev. Pharmacology and Toxicology* **23**, 239.

Fridovich, I. (1995). *Ann. Rev. Biochem.* **64**, 97.

Gabbianelli, R., Battistoni, A., Capo, C., Polticelli, F., Rotilio, G., Meier, B., & Desideri, A. (1997). *Arch. Biochem. Biophys*. **345**, 156.

Gebicki, J. M. & Bielski, B. I. J. (1981). *J. Am. Chem. Soc*. **103**, 7020.

Gregory, E. M. & Dapper, C. H. (1983). *Arch. Biochem. Biophys*. **220**, 293.

Hearn, A. S., Stroupe, M. E., Cabelli, D. E., Lepock, J. R., Tainer, J. A., Nick, H. S., & Silverman, D. N. (2001). *Biochemistry* **40**, 12051.

Hiraoka, B. Y., Yamakura, F., Sugio, S., & Nakayama, K. (2000). *Biochem. J.* **345**, 345.

Ilan, Y. A., Czapski, C., & Meisel D. (1976). *Biochim. Biophys. Acta* **430**, 209.

Ishikawa, T., Hunaiti, A. R., Piechot, G., & Wolf, B. (1987). *Eur. J. Biochem*. **170**, 317.

Juan, J. Y., Keeney, S. N., & Gregory, E. M. (1991). *Arch. Biochem. Biophys*. **286**, 257.

Kang, S. K., Jung, Y. J., Kim, C. H., & Song, C. Y. (1998). *Clinical and Diagnostic Laboratory Immunology* **5**, 784.

Kardinahl, S., Anemuller, S., & Schafer, G. (2000). *Biol. Chem.* **381**, 1089.

Keele, B. B., McCord, J. M., & Fridovich, I. (1970). *J. Biol. Chem.* **245**, 5753.

Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* **D55**, 941.

Kleywegt, G. J. (1996). *Acta Cryst*. **D52**, 842.

Kleywegt, G. J. (1999). *J. Mol. Biol*. **273**, 371.

Kettle, A. J. & Winterbourn, C. C. (1994). *J. Biol. Chem*. **269**, 17146.

Kirby, T. W., Lancaster, J. R., & Fridovich, I. (1981). *Arch. Biochem. Biophys*. **210**, 140.

Kim, E. J., Chung, H. J., Suh, B., Hah, Y. C., & Roe, J. H. (1998). *Mol. Microbiol*. **27**, 187.

Knapp, S., Kardinahl, S., Hellgren, N., Tibbelin, G., Schafer, G., & Ladenstein, R. (1999). *J. Mol. Biol.* **285**, 689.

Kraulis, P. J. (1991). *J. Appl. Crystallogr.* **24**, 945.

Kusunose, E., Ichihara, K., Noda, Y., & Kusunose, M. (1976). *J. Biochem.* **80**, 1343.

Lah, M. S., Dixon, M. M., Pattridge, K. A., Stallings, W. C., Fee, J. A., & Ludwig, M. L. (1995). *Biochemistry* **34**, 1646.

Leveque, V. J., Vance, C. K., Nick, H. S., & Silverman, D. N. (2001). *Biochemistry* **40**, 10586.

Lim, J-H., Yu, Y. G., Han, Y. S., Cho, S-J., Ahn, B-Y., Kim, S-H., & Cho, Y. (1997). *J. Mol. Biol.* **270**, 259.

Mann, T. & Keilin, D. (1939). *Proc. Royal Soc. Lond.* **B126**, 303.

Martin, M. E., Byers, B. R., Olson, M. O., Salin, M. L., Arceneaux, J. E., & Tolbert, C. (1986). *J. Biol. Chem.* **261**, 9361.

Matsumoto, T., Yamakura, F., & Terauchi, K. (1991). *Free Radical Res. Commun.* **12-13(Pt. 1)**, 329.

McAdam, M. E., Fox, R. A., Lavelle, F., & Fielden, E. M. (1977). *Biochem. J.* **165**, 71.

McCord, J. M. & Fridovich, I. (1969). *J. Biol. Chem.* **244**, 6049.

Meier, B., Scherk, C., Schmidt, M., & Parak, F. (1998). *Biochem. J.* **331**, 403.

Meier, B. & Sehn, A. P. (1994). *Biochem. J.* **304**, 803.

Meier, B., Sehn, A. P., Michel, C., & Saran, M. (1994b). *Arch. Biochem. Biophys*. **313**, 296.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol*. **277B**, 505.

Misra, H. P. & Fridovich, I. (1978). *Arch. Biochem. Biophys*. **189**, 317.

Navaza, J. (1994). *Acta Cryst*. **A50**, 157.

Neves, E. A. & De Andrade, J. F. (1986). *Polyhedron* **5**, 717.

Nicholls, A., Sharp, K., & Honig B. (1991). *PROTEINS: Struct. Funct. Gen.* **11(4)**, 281.

Ose, D. E. & Fridovich, I. (1979). *Arch. Biochem. Biophys*. **194**, 360.

Otwinoski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307.

Parker, M. W. & Blake, C. C. F. (1988). *J. Mol. Biol*. **199**, 649.

Pennington, C. D. & Gregory, E. M. (1986). *J. Bacteriol.* **166**, 528.

Pick, M., Rabani, J., Yost, F., Fridovich, I. (1974). *J. Am. Chem. Soc.* **96**, 7329.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Prot. Chem.* **23**, 283-437.

Ravindranath, S. D. & Fridovich, I. (1975). *J. Biol. Chem.* **250**, 6107.

Renault, J. P., Verchère-Béaur, C., Morgenstern-Badarau, I., Yamakura, F., & Gerloch, M. (2000). *Inorg. Chem.* **39**, 2666.

Renault, J. P., Morgenstern-Badarau, I., & Piccioli, M. (1999). *Inorg. Chem.* **38**, 614.

Ringe, D., Petsko, G. A., Yamakura, F., Suzuki, K., & Ohmori, D. (1983). *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3879.

Roussel, A. & Cambillau, C. (1989) *Silicon Graphics Geometry partners Directory,* pp 72-78. SGI, (USA)

Samuni, A., Chevion, M., & Czapski, C. (1984). *Radiation Research* **99**, 562.

Santos, R., Bocquet, S., Puppo, A., & Touati, D. (1999). *J. Bacteriol.* **181**, 4509.

Sawyer, D. T. & Valentine, J. S. (1981). *Acc. Chem. Res.* **14**, 393.

Searcy, K. B. & Searcy, D. G. (1981). *Biochim. Biophys. Acta* **670**, 39.

Schmidt, M., Meier, B., & Parak, F. (1996). *J. Biol. Inorg. Chem.* **1**, 532.

Schmidt, M., Scherk, C., Iakovleva, O., Nolting, H. F., Meier, B., & Parak, F. PDB: 1AVM.

Steinman, H. M. (1982). *Superoxide Dismutase,* pp 11-68. CRC, (USA).

Stallings, W. C., Powers, T. B., Pattridge, K. A., Fee, J. A., & Ludwig, M. L. (1983). *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3884.

Stallings, W. C., Pattridge, K. A., Strong, R. K., & Ludwig, M L. (1985). *J. Biol. Chem.* **260**, 16424.

Sugio, S., Hiraoka, B. Y., & Yamakura, F. (2000) *Eur. J. Biochem.* **267**, 3487.

Takao, M., Yasui, A., & Oikawa, A. (1991). *J. Biol. Chem.* **266**, 14151.

Takao, M., Oikawa, A., & Yasui, A. (1990). *Arch. Biochem. Biophys.* **283**, 210.

Tierney, D. L., Fee, J. A., Ludwig, M. L., & Penner-Hahn, J. E. (1995). *Biochemistry* **34**, 1661.

The CCP4 Suite: Programs for Protein Crystallography. (1994). *Acta Cryst.* **D50**, 760-763.

Ursby, T., Adinolfi, B. S., Al-Karadaghi, S., De Vendittis, E., & Bocchini, V. (1999). *J. Mol. Biol.* **286**, 189.

Vance, C. K. & Miller, A. F. (1998). *J. Am. Chem. Soc.* **120**, 461.

Whittaker, M. M., & Whittaker, J. W. (2000*). J. Biol. Inorg. Chem.* **5**, 402.

Whittaker, M. M. & Whittaker, J. W. (1999). *J. Biol. Chem.* **274**, 34751.

Whittaker, M. M. & Whittaker, J. W. (1997). *J. Biol. Inorg. Chem.* **2**, 667.

Whittaker, M. M. & Whittaker, J. W. (1996). *Biochemistry* **35**, 6762.

Whittaker, M. M., Ekberg, C. A., Meier, A. E., & Whittaker, J. W. (1997). *J. Inorg. Biochem.* **67**, 204.

Weisiger, R. A. & Fridovich, I. (1973). *J. Biol. Chem.* **248**, 3582.

Weatherburn, D. C., (1996). *Perspec. Bioinorganic Chem.* **3**, 1.

Yamakura, F., Rardin, R. L., Petsko, G. A., Ringe, D., Hiraoka, B. Y., Nakayama, K., Fujimura, T., Taka, H., & Murayama, K. (1998). *Eur. J. Biochem.* **253**, 49.

Yamakura, F., Kobayashi, K., Ue, H., & Konno, M. (1995). *Eur. J. Biochem.* **227**, 700.

Yamakura, F., & Suzuki, K. (1986). *Biochim. Biophys. Acta* **874**, 23.

Yamakura, F. (1984). *Biochem. Biophys. Res. Comm.* **122**, 635.

Yamano, S., Sako, Y., Nomura, N., & Maruyama, T. (1999). *J. Biochem.* **126**, 218.

Yamano, S., & Maruyama, T. (1999). *J. Biochem.* **125**, 186.

Yost, F. J., Jr. & Fridovich, I. (1973). *J. Biol. Chem.* **248**, 4905.

Youn, H-D., Kim, E-J., Roe, J-H., Hah, Y. C., & Kang, S-O. (1996). *Biochem. J.* **318**, 889.

Chapter 7.

β-Lactoglobulin

7.1 Early studies of bovine β-lactoglobulin

The main whey proteins of cow's (bovine) milk are β-lactoglobulin (BLG) (Figure 7.1) (~0.3 mg/mL), α-lactalbumin (~0.1 mg/mL), serum albumin (~0.04 mg/mL) and immunoglobulin (~0.08 mg/mL) [Bell & McKenzie 1964]. β-Lactoglobulin is, therefore, the major protein component of whey. The composition of milk was unclear until the 1930s: albumin was recognized as the major component of the skim milk of cow and was the general subject of much research. β-Lactoglobulin was in 1934, when efforts to follow Sjögren and Svedberg's procedure to isolate albumin [Sjögren & Svedberg 1930] resulted instead in the isolatation of a new crystalline globular protein [Palmer 1934]. Seven years later, Palmer's protein was named "β-lactoglobulin" (BLG) [Cannan *et al.* 1942].

β-Lactoglobulin (BLG) was for a long time considered to be a pure protein sample, because the protein was purified by the method of crystallization, and the electrophoresis experiments showed protein homogeneity [Pederson 1936]. However, different values had been reported for the molecular weight of BLG, as well as different amino acid compositions. This pointed to more than one component being present in the sample. High-resolution electrophoreses in the 1940's finally proved that there were at least two components to bovine BLG [Li 1946].

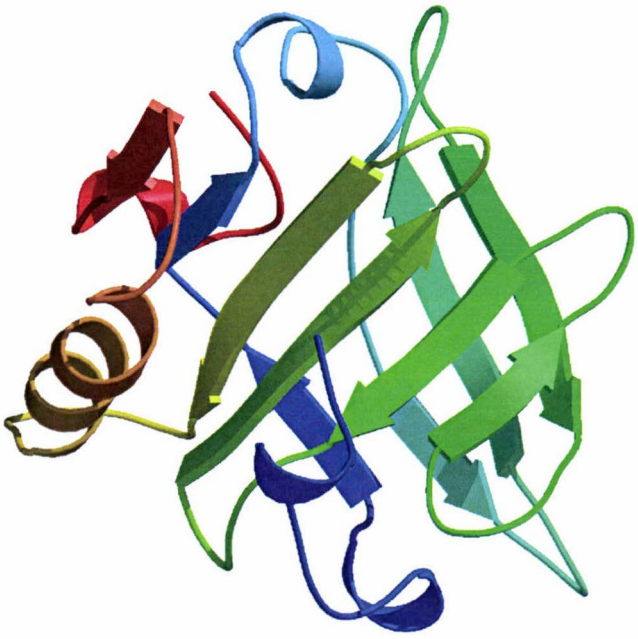


Figure 7.1

BLG monomer

A number of variants of BLG have been found in the bovine gene pool. Bell reported a new variant of bovine BLG in Australia cattle [Bell 1962]. Bell denoted β1-lactoglobulin as variant A, β2-lactoglobulin as variant B and his new discovery as variant C. Subsequent studies have revealed that BLG variants occur frequently in different bovine sub-species; cow has five and possibly six variants: A, B, C, D, W and possibly H. Each variant is identified by mutations at specific sites (Table 7.1).

Table 7.1. Mutation sites for bovine BLG variants [SRS ExPASY]

Variant	Site / Amino Acid				
	45	56	59	64	118
A	E	I	Q	D	V
B	E	I	Q	G	A
C	E	I	H	G	A
D	Q	I	Q	G	A
W	E	L	Q	G	A

7.2 Distribution of β-lactoglobulin

The presence of β-lactoglobulin in species other than cow has been studied [Lyster *et al.* 1966, Liberatori *et al.* 1979, Pervaiz & Brew 1986]. The β-lactoglobulin protein is found in the milk of a number of mammals, in particular ruminants (Table 7.2). Most β-lactoglobulins occur as dimers, although monomeric β-lactoglobulins exist in the milk of some non-ruminant species such as pig, horse, cat, etc. No similar protein is present or at least expressed in the milk of rodents. It is believed that no BLG exists in human milk [Bull & Currie 1946]. Reports [Liberatori *et al.* 1979b; Brignon *et al.* 1985] of BLG in human milk have been the subject of much debate. A number of studies have identified the reported protein as bovine BLG or a truncated bovine BLG (13.5 KDa) [Fukushima *et al.* 1997; Conti & Godovac-Zimmermann 1980; Perez & Calvo 1995], yet other studies find no antibody response in human milk to BLG antibodies [Restani *et al.* 2000]. A number of lipocalin genes are found in the human genome. The one with highest sequence identity (43%) is glycodelin. This protein is expressed in the placenta [Murakami *et al.* 1998]. None of the other identified lipocalins has been observed in

human milk. [(BLAST) Altschul *et al.* 1990, (ClustalW) Pearson 1990, Venter *et al.* 2001].

Table 7.2. Species where BLG is present in the milk

Species	Oligomer	Reference
Cow	Dimer	Bull & Currie 1946
Oxen	Dimer	Bull & Currie 1964
Yak	Dimer	Grosclaude *et al.* 1966
Zebra	Dimer	Lyster *et al.* 1966
Buffalo	Dimer	Lyster *et al.* 1966
Bison	Dimer	Lyster *et al.* 1966
Musk ox	Dimer	Lyster *et al.* 1966
Eland	Dimer	Lyster *et al.* 1966
Goat	Dimer	Askonas 1954
Sheep	Dimer	Lyster *et al.* 1966
Mouflon	Dimer	Liberatori *et al.* 1979b
Red deer	Dimer	Lyster *et al.* 1966
European elk	Dimer	Lyster *et al.* 1966
Reindeer	Dimer	Lyster *et al.* 1966
White-tailed deer	Dimer	Lyster *et al.* 1966
Fallow deer	Dimer	Lyster *et al.* 1966
Caribou	Dimer	Lyster *et al.* 1966
Giraffe	Dimer	Lyster *et al.* 1966
Okapi	Dimer	Lyster *et al.* 1966
Prong-horn antelope	Dimer	Lyster *et al.* 1966
Camel	Monomer	Liberatori *et al.* 1979b
Peccary	Monomer	Lyster *et al.* 1966
Pig	Monomer	Liberatori *et al.* 1979b
Horse	Monomer	Liberatori *et al.* 1979b
Zebra	Monomer	Lyster *et al.* 1966
Rhinoceros	Monomer	Lyster *et al.* 1966
Donkey	Monomer	
Dolphin	Monomer	Pervaiz & Brew 1986
Manatee	Monomer	Pervaiz & Brew 1986
Beagle	Monomer	Pervaiz & Brew 1986
Cat	Monomer	Halliday *et al.* 1991
Grey kangaroo	Monomer	McKenzie & Muller 1983
Red kangaroo	Monomer	McKenzie & Muller 1983

The ruminant animals have dimeric β-lactoglobulin and show high sequence similarity with bovine BLG. Cow has at least five variants of BLG, donkey possesses two variants, sheep have variants A, B and C, which are different from the A, B and C variants for cows. Cow contains only one gene for BLG: this gene has either heterozygous or homozygous combinations from the animal's parents giving rise to combinations AA, BB and AB, with the AB combination expressing both variants in the milk. Some other animals have multiple BLG genes; cat, for instance, has two BLG genes.

7.3 The lipocalin super-family

BLG belongs to the lipocalin super-family of proteins [Pervaiz & Brew 1987]. Members of this family have a molecular mass of around 20 kDa (~160 amino acids) and are able to bind small hydrophobic molecules. More than sixty individual lipocalins have been identified [SRS ExPASY] and studied. The lipocalins are defined by the fact that they share a similar structural motif of an eight-stranded anti-parallel β-barrel and a single three-turn α-helix external to the barrel (Figure 7.1). The name lipocalin refers to the cup-shaped (Greek: calyx) molecule, which has a central hydrophobic pocket into which non-polar lipophilic moieties bind. Most lipocalins bind specific small hydrophobic molecules. For example, odorant-binding protein and mouse major urinary protein are involved in binding pheromones. However, no specific binding site or function has been found for β-lactoglobulin. The structures of a number of lipocalins have been determined by X-ray crystallography. The similarity in three-dimensional structure confirms that they belong to the same family of proteins. In contrast to the Fe and Mn SODs, the lipocalins have low sequence similarity, sharing only a GxW motif near

residue 20 in the first β strand, even though the tertiary structures are very similar. For example odorant-binding protein and BLG share less than 10% identity.

7.4 Ligand-binding properties of BLG

The lipocalin family to which BLG belongs is defined by the binding of small hydrophobic, biological molecules. Unlike other lipocalins, BLG appears to be a non-specific binder of physiologically important hydrophobic molecules (Table 7.3). Several biologically important ligands in this table are the free fatty acids, which are present in milk. Fatty acids may bind to BLG in a reversible manner, at a single site with a binding constant of ~10^5 similar to SDS [Spector & Fletcher 1970]. Other biologically important molecules, retinol, hemin and protoporphyrin IX, all bind to bovine BLG with relatively high affinity. A variety of other small hydrophobic molecules can be bound to BLG with lower affinities (see Table 7.3).

Retinol binds to BLG with the highest affinity (5.0×10^7) [Fugate & Song 1980]. Retinol-binding protein, which is a specific retinol carrier *in vivo*, has a retinol binding constant of only 5.9×10^6 [Malpeli *et al.* 1995], an order of magnitude lower than that of BLG. This suggests that BLG's biological function may be involved in transport of retinal, an important molecule in the ocular vision system [Blomhoff *et al.* 1990]. It is now believed retinal plays an important role in the regulation of cell growth [Schubert *et al.* 1986]. Though a great deal of research in this area has been done and is on going, the function of BLG is still not certain.

Table 7.3. Ligand binding constants to BLG [Perez & Calvo 1995]

Ligand	Binding constant	Reference
Retinol	5.0×10^{7}	Fugate & Song 1980
β-Ionone	1.47×10^{6}	Dufour & Haertle 1990
Stearate	1.7×10^{5}	Spector & Fletcher 1970
Palmitate	6.8×10^{5}	Spector & Fletcher 1970
SDS	3.1×10^{5}	O'Neill & Kinsella 1987
Hemin	4.0×10^{6}	Dufour & Haertle 1990
Protoporphyrin IX	2.5×10^{6}	Dufour & Haertle 1990
Toluene	4.5×10^{2}	Robillard & Wishnia 1972a
p-Nitrophenol	1.9×10^{5}	Farrell *et al.* 1987
2–Heptanone	2.0×10^{2}	O'Neill & Kinsella 1987
Laurate	0.5×10^{5}	Spector & Fletcher 1970
Oleate	0.4×10^{5}	Spector & Fletcher 1970
Butane	1.7×10^{3}	Wishnia & Pinder 1966
Pentane	7.1×10^{3}	Wishnia & Pinder 1966
Iodobutane	2.8×10^{3}	Wishnia & Pinder 1966
2,6-MANS	3.4×10^{5}	Lovrien & Anderson 1969
Methyl orange	0.2×10^{4}	O'Neill & Kinsella 1987
n-Octylbenzene-*p*-sulphonate	6.3×10^{4}	O'Neill & Kinsella 1987
p-Nitrophenol	1.9×10^{4}	Farrell *et al.* 1987
p-Nitrophenylacetate	3.0×10^{4}	Farrell *et al.* 1987
p-Nitrophenyl-β-glucuronide	1.6×10^{4}	Farrell *et al.* 1987
p-Nitrophenyl sulfate	2.0×10^{3}	Farrell *et al.* 1987
p-Nitrophenyl pyridoxal phosphate	3.1×10^{3}	Farrell *et al.* 1987
2-Octanone	0.5×10^{3}	O'Neill & Kinsella 1987
2-Nonanone	2.4×10^{3}	O'Neill & Kinsella 1987
Trifluorotoluene	4.2×10^{2}	Robillard & Wishnia 1972b
Hexafluorobenzene	1.6×10^{3}	Robillard & Wishnia 1972b

Ligand-binding studies must be viewed with reference to their limitations. Most of the studies are done with fluorescence probes, where the fluorescence is quenched on binding. The amount of quenching is not consistent between different experimental conditions, and therefore some results may not be reliable. Moreover, the longer chain amphipathic molecules exist as micelles in aqueous environments; intrinsic solubility is very low [L. Sawyer, personal communication].

7.5 Possible physiological function of BLG

BLG is a nutritious component of milk; all the essential amino acids are included in BLG. Nutrition alone does not explain its lipocalin fold, acidic stability and resistance to gastric proteolysis [Futterman & Heller 1972, Hemley *et al.* 1979]. Other biological functions must be responsible for the existence of BLG. Since BLG is a member of the lipocalin super-family and since BLG binds *in* vitro a variety of hydrophobic substrates, the simple explanation for the function of BLG is that it protects and transports small hydrophobic molecules *in vivo.* A more sophisticated physiological function for BLG may involve the transfer of a passive immunity [Fugate & Song 1980], or activation of a gastrointestinal lipase enzyme [Perez *et al.* 1992]. Very recently, some evidence of bactericidal activity for BLG has been published [Pellegrini *et al.* 2001].

7.6 Crystal forms of BLG

Since the 1950s, X-ray diffraction techniques have become the most widely used method to determine the three-dimensional structure of proteins. BLG crystallizes readily, and was the subject of X-ray diffraction experiments as early as 1938. Long before phase-determination methods were developed for the solution of protein crystal structures, two crystal forms of BLG had been identified by Crowfoot and Riley [Crowfoot & Riley 1938]. One of crystal forms had the space group $P2_12_12_1$ with unit cell dimensions close to lattices now identified as Q or N.

Nearly three decades after Crowfoot and Riley's pioneering study, Aschaffenburg published the unit cell parameters and space groups for a number of BLG crystal forms [Aschaffenburg *et al.* 1965], spanning a pH range from 3.5 to 7.6. All seven crystal systems except cubic are represented. Table 7.4 shows all the known crystal forms. Subsequent structure determinations have involved primarily lattices X, Y and Z, due to the reproducibility of the crystallization conditions and the stability of the crystallizations to minute differences in conditions.

Table 7.4. Crystal forms of β-lactoglobulin [Aschaffenberg *et al.* 1965]

Variant	pH	Lattice	Space Group	*a* (Å)	*b* (Å)	*c* (Å)	α (°)	β (°)	γ (°)
A+B	3.5	P	$P6_3$	67	67	141	90	90	90
B	5.2	R	$P2_1$	36.1	127.5	36	90	106.1	90
B	5.2	S	$P2_1$	36.4	127.6	36.4	90	98.2	90
A	5.2	W	$P2_1$	36.4	68.2	72.4	90	92.2	90
A	5.2	N	$P2_12_12_1$	68.8	70.2	137.8	90	90	90
A/B	5.2	Q	$P2_12_12_1$	69.2	70.7	157.5	90	90	90
B	5.2	T	$P4_12_12$	69.2	69.2	138.8	90	90	90
A+B	5.2	U	$P4_22_12$	67.5	67.5	133.5	90	90	90
A	6.9	X	$P1$	31.8	49.7	56.6	122.8	97.5	104.1
A/B	7.6	Y	$C222_1$	55.7	81.2	67.2	90	90	90
A/B	7.6	Z	$P3_121$	54.3	54.3	117	90	90	120

7.7 The structure of bovine BLG

The structural determination of bovine BLG was begun in the 1950s and first reported at a low resolution of 6 Å in 1979 for lattices X, Y and Z [Green *et al.* 1979]. were solved. However, it is impossible to trace even the secondary structure at such low resolution. The electron density of lattice X did reveal the existence of a molecular two-fold axis relating pairs of molecules of the dimer. This axis is not necessarily coincident with a crystallographic symmetry axis. It was assumed that the lattice X (pH 6.5), lattice Y (pH 7.5), and lattice Z (pH 8.1) structures corresponded to the pH-dependent conformational change known as the Tanford transition [Tanford *et al.* 1959]. Green *et al.* also reported that a

free sulfhydryl (Cys 121) site, the presumed site of heavy metal binding, in lattice X differed in position by 3 Å from that for lattices Y and Z. This conclusion was based on the heavy-atom positions which can be deduced relatively accurately even at low resolution.

In 1986 a lattice Y structure at 2.8 Å resolution was reported for variant A of bovine BLG (BLGA) [Papiz *et al.* 1986]. This structure showed the lipocalin fold and the conformational similarity of BLGA with retinol-binding protein [Winter *et al.* 1993]. The molecule consists of an anti-parallel β-sheet, formed by 8 strands wrapped round to form a flattened cone or calyx. A three-turn α-helix and an external β-strand are located towards the C-terminus. Although this model described the basic structure of BLGA, threading errors occurred because the loop regions were generally ill-defined.

A second medium-resolution structure of BLGA at 2.5 Å resolution was reported [Monaco *et al.* 1987], solved by molecular replacement in the trigonal lattice Z. The structures of the Y- and Z-lattice crystal forms showed no great difference in the overall β-barrel topology, although structure differences in several areas were reported. All residues were identified and the two disulfide bonds (66-160 and 106-119) were clearly visible, but again serious irregularities with the data and model call the solution into question. The structure was refined without an *R*-free monitor, and the average RMS deviation of bond lengths from their ideal values is very high at 0.08 Å. The most significant and major issue was the fact that molecules are packed as monomers not as the expected dimers in the crystal lattice.

High-resolution models for BLGA in lattice Y [Bewley *et al.* 1997] and

lattice X [Brownlow *et al.* 1997] have been determined. The model at 1.8 Å resolution in lattice Y (space group $C222_1$) contains residues 1-62, 66-110, 115-154 and 10 water molecules, R_f=0.27, R=0.22. One disulfide bridge between Cys 106 and Cys 119 is clearly defined; the other disulfide bond between Cys 66 and Cys 160 is poorly defined. BLGA in lattice X was determined at 1.9 Å resolution. The C and N-termini are not visible and the loop CD (amino acids 61-65) is not well defined according to the authors. These two structures convincingly correct the errors of earlier models.

The structures of BLGA in lattice Z at pH 6.2, 7.1 and 8.2 revealed that the Tanford transition is a pH-induced opening and closing of a "gate" over the open end of the calyx. This "gate" is comprised of residues 85-90 [Qin *et al.* 1998]. At low pH Glu 87 is buried and protonated; at high pH it is exposed and deprotonated. As a result, the pK_a of this glutamic acid is anomalously high at 7.3. This opening and closing of access to the calyx has functional implications for the binding and release of small hydrophobic molecules.

Two independent NMR structures have been recently published [Kuwata *et al.* 1999; Uhrinova *et al.* 2000]. Both conclude that the solution structure is very similar to the X-ray structures. NMR structures have limitations when dealing with unusual structural features such as γ-turns, which BLG has, as the minimization functions have no terms restraining the feature.

7.8 Binding of fatty acids, the structures

Studies of ligand binding to BLGA have resulted in three structures. Qin *et al.* [1998b] reported the structure of BLGA with a substituted fatty acid 12-bromododecanoic acid (BRC) bound in the central pocket of BLGA. This structure proved that fatty acids do bind in the calyx of the protein and not on the surface [Qin *et al.* 1998b]. The group of Sawyer also published a structure with palmitic acid bound inside the calyx [Wu *et al.* 1999]. Sawyer has an unpublished structure of BLGA with retinol bound only part-way into the calyx. The central pocket of BLGA is very hydrophobic (Figure 7.2). In the case of binding 12-bromododecanoic acid, numerous van der Waals contacts are made between aliphatic side chains and the aliphatic tail of the fatty acid. These intermolecular contacts are summarised in Table 7.5.

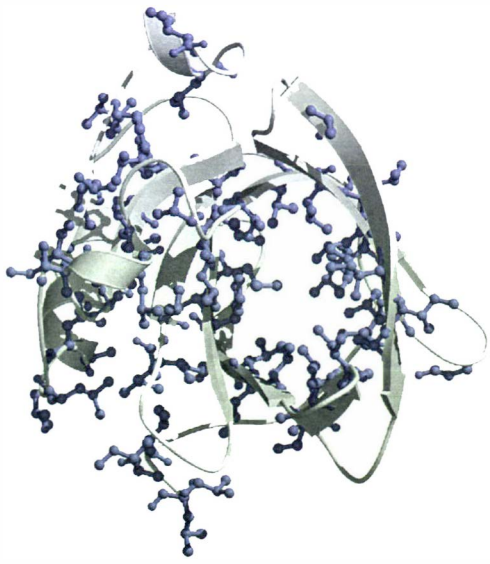


Figure 7.2

Hydrophobic core of BLGA

View looking down the calyx of BLG. Aliphatic amino acids are shown in ball and stick.

Several lysine residues (Lys 60 and Lys 69) at the rim of the calyx are in position to hydrogen bond to the carboxylate head group of the fatty acid. One of the most important contacts is BRC_OE2...NZ_Lys60 NZ (3.10 Å). This hydrogen-bond would appear to be one of the driving forces behind the binding of fatty acids, as it is a specific interaction. Porcine BLG which has the substitution K60E does not bind fatty acids. Fatty-acid type molecules of different sizes have been shown to hydrogen-bond to K60_NZ bond, regardless of size. The longer chain molecules wrap around themselves so as to fit into the calyx and allow the hydrogen bond [L Sawyer *personal communication*]. Lys 60 also makes a hydrogen bond with a water molecule (HOH 187 3.47 Å), and Lys 69 makes an intermolecular hydrogen-bond with Glu 62. Figure 7.3 shows 12-bromododecanoic acid (red) and most of the residues that make the intermolecular contacts shown in Table 7.5.

Table 7.5. Molecular contacts between BLGA and 12-bromododecanoic acid

RESIDUE	ATOM	RESIDUE	ATOM	DISTANCE (Å)
K60	NZ	BRC	OE2	3.1
			OE1	4.07
		HOH187	O	3.47
K69	NZ	E62	OE1	3.24
			OE2	3.54
I71	CD1	BRC	C4	3.75
		BRC	C5	3.8
M107	SD	BRC	C5	3.54
	SD	BRC	C6	3.79
	SD	BRC	C7	3.68
V41	CG2	BRC	C6	3.9
I71	CD1	BRC	C4	3.75
	CD1	BRC	C5	3.8
I56	CD1	BRC	C9	3.57
F105	CE1	BRC	C10	3.68
	CE2	BRC	C10	3.93
	CD1	BRC	C10	4.22
	CD2	BRC	C10	4.43
	CZ	BRC	C10	3.52
	CG2	BRC	C10	4.57

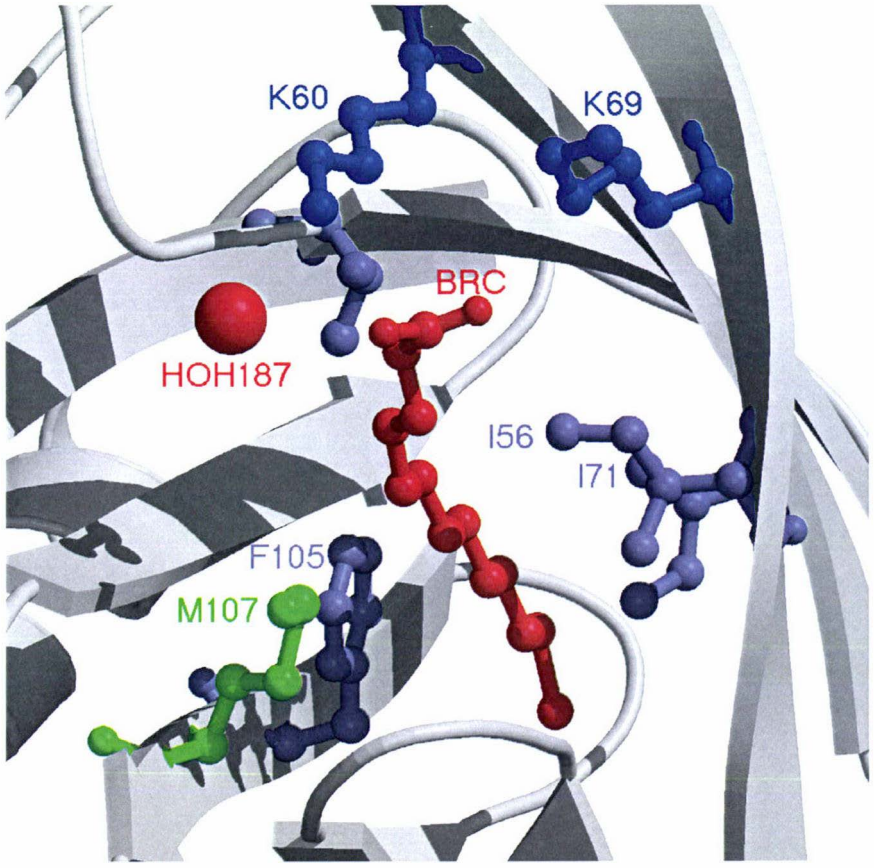


Figure 7.3

12-bromododecanoic acid bound in the calyx of BLGA

Aliphatic amino acids are shown in grey (ball and stick) the 12-bromododecanoic acid in red, and the lysine side-chains in blue.

7.9 Recombinant bovine β-lactoglobulin

The production of bovine BLG by recombinant DNA methods has long been a goal, as a reliable source of pure variants of BLG, as well as mutant proteins with easy purifications, is necessary for further mechanistic and structural studies of BLG. There have been many attempts to produce recombinant bovine BLG reported to the literature (Table 7.6).

An early attempt to express bovine BLG in *Escherichia coli* in its mature form [Batt *et al.* 1990] met with limited success. Met-BLG was expressed by using a *tac*-promoter vector, pTTQ18 and accounted for approximately 15% of the total cellular protein. The majority of the met-BLG was found in an insoluble denatured form. By using guanidinium hydrochloride, the recombinant protein was renatured. In an attempt to enhance the thermal stability of BLG, two mutants, L104C and A132C, were designed in the hope that an additional disulfide bridge would be formed with the free cysteine, Cys 121 [Cho *et al.* 1994]. These mutants were cloned into a *tac* expression vector. However, only insoluble, denatured protein was expressed. Chatel *et al.* [1999] also attempted to express BLG in *E. coli*. The BLG gene was inserted into the pET26 vector and transformed into COS-7 cells. BLG was again expressed in denatured form.

Expression in insect and mammalian cells has also been attempted. In 1992, Mizumachi reported that the expression of recombinant BLG in insect cells produced properly folded recombinant protein, at a yield of 5 mg/L of culture [Mizumachi *et al.* 1993]. More recently, expression in mouse tibialis muscle cells yielded correctly folded protein, but at very low levels [Chatel *et al.* 1999]. Hyttinen introduced the BLG gene from a genomic library into a murine germline [Hyttinen *et al.* 1998]. Bovine BLG was expressed efficiently in the mammary gland of the transgenic mice. Expression levels of bovine BLG in the milk exceeded 1 mg mL. However, similar purification and isolation issues exist for transgenic mice as exist for bovine milk, which makes a bacterial or yeast expression system more desirable.

Several attempts have been made to express BLG in yeast. The clone for expression was constructed by inserting the entire cDNA encoding pro-

BLG into a yeast expression vector (pYG100). BLG was expressed in *Saccharomyces cereviseae* at a modest concentration of 1.1 mg/L of culture [Tostuka *et al.* 1990]. The most successful synthesis of recombinant BLG thus far, has been the expression of BLGA in *Pichia pastoris* [Denton *et al.* 1997]. Expression levels of more than 1 g/L of culture have been achieved. Although the recombinant BLG was reported to be indistinguishable from native BLG there is evidence that post-translational modification of residues, in particular lysines, has occurred [Denton *et al.* 1997].

Table 7.6. Reported expression systems for recombinant BLG

Organism	Type	Yield	Folding	Reference
E. coli	Cytoplasm	~1 mg L^{-1}	Insoluble unfolded	Batt *et al.* 1990
		Low	Insoluble unfolded	Chatel *et al.* 1999
P. pastoris	Secreted	1 g L^{-1}	Soluble folded	Denton *et al.* 1997
S. cerevisiae	Secreted	1.1 mg L^{-1}	Soluble folded	Totsuka *et al.* 1990
Muscle		ng	Soluble folded	Chatel *et al.* 1999
Insect	Secreted	5 mg L^{-1}	Soluble folded	Mizumachi *et al.* 1993
Milk (mouse)	Secreted	~1 mg mL^{-1}	Soluble folded	Hyttinen *et al.* 1998

7.10 Goals

The goals of this research project are to add to the known structural information about β-lactoglobulin, with specific aims to compare the structure of BLG at high and low ionic strength, and attempt to co-crystallize at low ionic strength BLG with a number of small hydrophobic molecules. The X-ray structure of BLG at low ionic strength will help answer the question, whether the structure of BLG at high ionic strength (> 2 M) is in any way different to the structure at physiological ionic strength (~150 mM), and what those differences might be. The structures that will be detailed in the next few chapters have given insight to the nature of the flexible loops that are not well defined in most structures, and have shown that the core elements of BLG are well conserved over a variety of crystallization and NMR experimental conditions. Our laboratory is involved in a substantial project to produce recombinant BLG for functional and structural studies. My contribution to this project will focus on ligand binding, and in turn I will gain the skills and knowledge required to produce recombinant proteins.

Chapter 8

The structure of BLGA Y′ lattice a 2.6 Å resolution

8.1 Crystallization

Crystals were grown by the hanging drop method at 22°C; each drop consisted of 1 μL protein (20 mg mL^{-1} in 0.1 M Tris pH 7.0) and 1 μL of well solution. The well solutions contained 0.2 M Tris pH 7.5, and 2.5 M $(NH_4)_2SO_4$. Crystals (Figure 8.1) grew to a size of 500x400x300 microns in approximately two weeks.

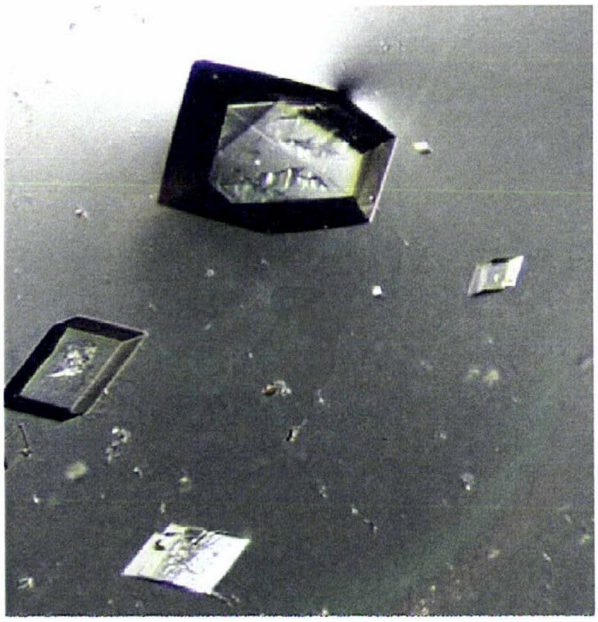


Figure 8.1

BLGA Y lattice crystals

8.2 Data collection and processing

A single crystal (Figure 8.1) was mounted in a nylon loop, with a cryoprotecting solution consisting of the well solution from which the crystal was grown and 15% glycerol. The crystal was soaked in the cryoprotectant for approximately 30-50 seconds then snap frozen in the cryo-stream (Oxford Cryosystems). Data were collected on a Rigaku

RU200 rotating-anode generator running at 5.4 kW Cu-Kα, 0.3 mm focus, fitted with a graphite monochromator coupled with a Rigaku RAxis IIC image-plate detector. X-ray frames were taken of each crystal (50 minute, 1.0° oscillation).

The cell and orientation matrix were determined by the auto-index procedure in DENZO [Otwinoski & Minor 1997]. The resulting data files were scaled and merged together with SCALEPACK [Otwinoski & Minor 1997]. Unit cell dimensions were averaged over all data. Relevant statistics are in Table 8.1. Measured intensities were converted to amplitudes using the program TRUNCATE [CCP4 1994].

The data collected to 1.8 Å resolution were easily identified as the Y lattice. However, the processing of the data in this C-centred orthorhombic lattice, space group $C222_1$, with DENZO [Otwinoski & Minor 1997] did not adequately predict all the X-ray diffraction spots in the pattern recorded (Figures 8.2 and 8.3). There was a systematic missing of lines of spots that corresponded to an underlying primitive lattice and the space group $P2_12_12_1$, with two molecules in the asymmetric unit, and eight molecules in the unit cell. The two molecules in the asymmetric unit are related to each other by non-crystallographic symmetry. The tables below describe the data merging statistics in $P2_12_12_1$, and on both the C-centred lattice ($h+k = 2n$) and primitive data only ($h+k = 2n+1$) (Table 8.1). The data for which $h+k = 2n+1$ did not have significant intensity beyond ~2.7 Å resolution.

Table 8.1. Data merging statistics for BLGA in the Y′ lattice.

(a) All data

Shell Lower	Upper Limit (Å)	I	Average error	stat.	Norm. Chi^2	Linear R-fac	Square R-fac
99.00	3.88	7340.6	205.9	119.4	0.994	0.018	0.017
3.88	3.08	3611.1	111.6	72.6	1.004	0.029	0.023
3.08	2.69	1239.9	69.5	50.7	1.095	0.064	0.043
2.69	2.44	662.8	55.2	46.8	1.098	0.111	0.072
2.44	2.27	458.9	53.4	48.3	0.991	0.157	0.098
2.27	2.13	323.4	53.7	50.8	1.012	0.226	0.139
2.13	2.03	224.9	54.1	52.5	1.038	0.333	0.200
2.03	1.94	147.4	53.1	52.3	1.042	0.476	0.271
1.94	1.86	84.8	55.1	54.8	1.076	0.716	0.425
1.86	1.80	49.8	56.5	56.0	1.291	0.000	0.611
All reflections		1489.5	78.2	60.9	1.047	0.052	0.024

(b) Data for which $h+k = 2n$ (even)

Shell Lower	Upper Limit (Å)	I	Average error	stat.	Norm. Chi^2	Linear R-fac	Square R-fac
99.00	3.88	14688.1	289.2	218.4	0.949	0.015	0.017
3.88	3.08	7153.3	172.1	110.1	0.962	0.022	0.022
3.08	2.69	2424.7	87.1	60.2	1.114	0.041	0.037
2.69	2.44	1272.2	72.8	50.8	1.018	0.068	0.060
2.44	2.27	885.9	63.9	50.3	1.008	0.094	0.080
2.27	2.13	616.4	59.7	52.0	1.068	0.134	0.114
2.13	2.03	425.1	57.7	53.3	1.091	0.190	0.154
2.03	1.94	275.5	54.5	52.5	1.126	0.271	0.207
1.94	1.86	152.4	55.7	54.8	1.143	0.417	0.322
1.86	1.80	83.8	73.0	55.6	1.027	0.688	0.481
All reflections		2945.6	100.6	77.2	1.048	0.035	0.021

(c) Data for which $h+k = 2n+1$ (odd)

Shell Lower Upper Limit (Å)		I	Average error	stat.	Norm. Chi^2	Linear R-fac	Square R-fac
99.00	3.88	100.5	28.6	26.2	1.108	0.198	0.149
3.88	3.08	106.5	46.2	45.2	1.053	0.315	0.259
3.08	2.69	54.6	64.0	55.2	1.052	0.806	0.611
2.69	2.44	30.4	62.4	62.3	1.038	0.000	0.890
2.44	2.27	25.4	70.1	70.0	0.866	0.000	0.000
2.27	2.13	22.8	82.0	81.9	0.948	0.000	0.000
2.13	2.03	21.1	88.9	88.8	0.996	0.000	0.000
2.03	1.94	15.8	96.3	96.2	0.964	0.000	0.000
1.94	1.86	17.1	100.4	100.3	1.124	0.000	0.000
1.86	1.80	9.4	91.2	90.9	1.057	0.000	0.000
All reflections		43.8	70.3	68.9	1.010	0.366	0.343

(d) Comparison of data

	$h+k$ = even	$h+k$ = odd	all data
# data	13523	9691	27047
data(I/σI)	29	0.61	19
resolution limit (Å)	1.8	2.7	1.8
Redundancy	3.3 (2)	1.6 (0)	3.5 (2)
R_{merge}	0.035	0.36	0.052

Summary of reflections intensities and R-factors by shells

R linear = SUM (ABS(I - <I>)) / SUM (I)

R square = SUM ((I - <I>) ^{2}) / SUM (I^2)

Chi^2 = SUM (I - <I>) ^{2}) / (Error ^{2} × N / (N-1))

In all sums single measurements are excluded

Table 8.2. Data collection statistics for BLGA in the Y′ lattice.

Space group	$P2_12_12_1$	
Unit cell	a = 54.79 Å	α = 90°
	b = 80.20 Å	β = 90°
	c = 66.13 Å	γ = 90°
Data collection		
Observed reflections	307846	
Unique reflections	27047	
Completeness	97.5 %	(81.9 %)[a]
R_{merge} on I	0.052	(0.611)[a]
Overall I/σ	19.0	(0.9)[a]
Wilson B (Å^2)	31.50	
Redundancy	11.3	(2.0)[a]

[a] Outermost shell (1.86 – 1.8 Å)

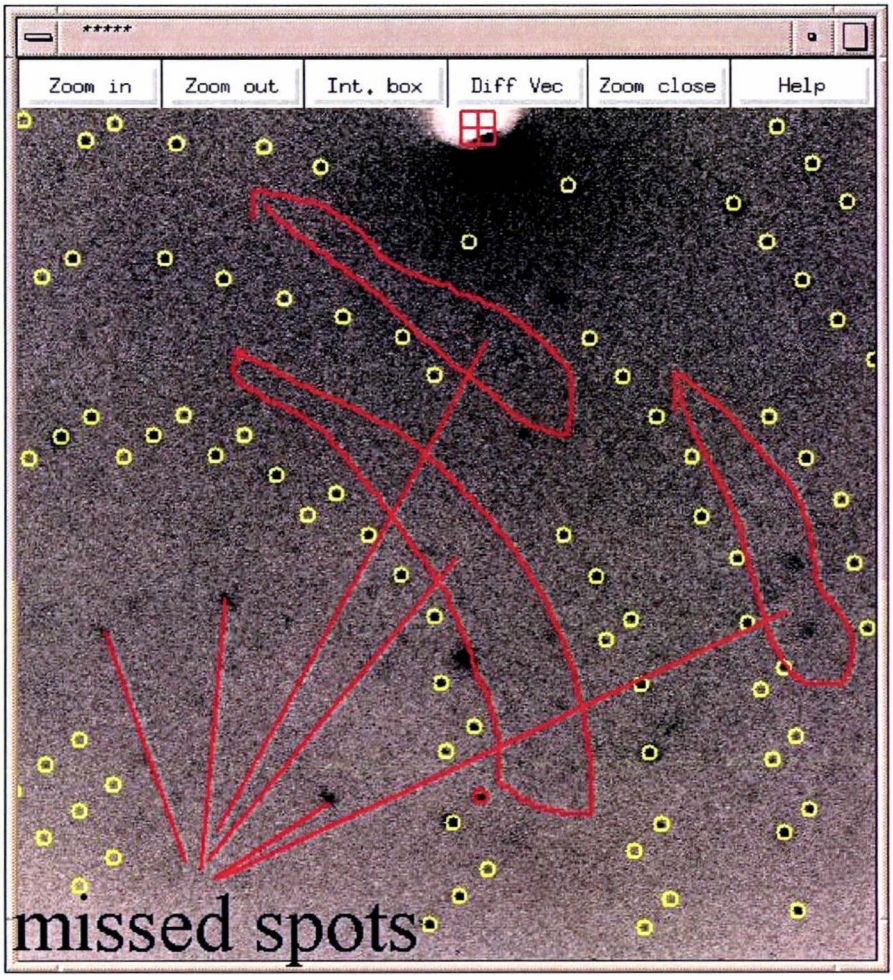


Figure 8.2

BLGA Y′ lattice diffraction image

Zoom area of a diffraction frame showing the predicted spots [yellow (non-overlapped) and red (overlapped) circles] for space group $C222_1$. The highlighted areas contain diffraction spots missed by the by assuming space group $C222_1$.

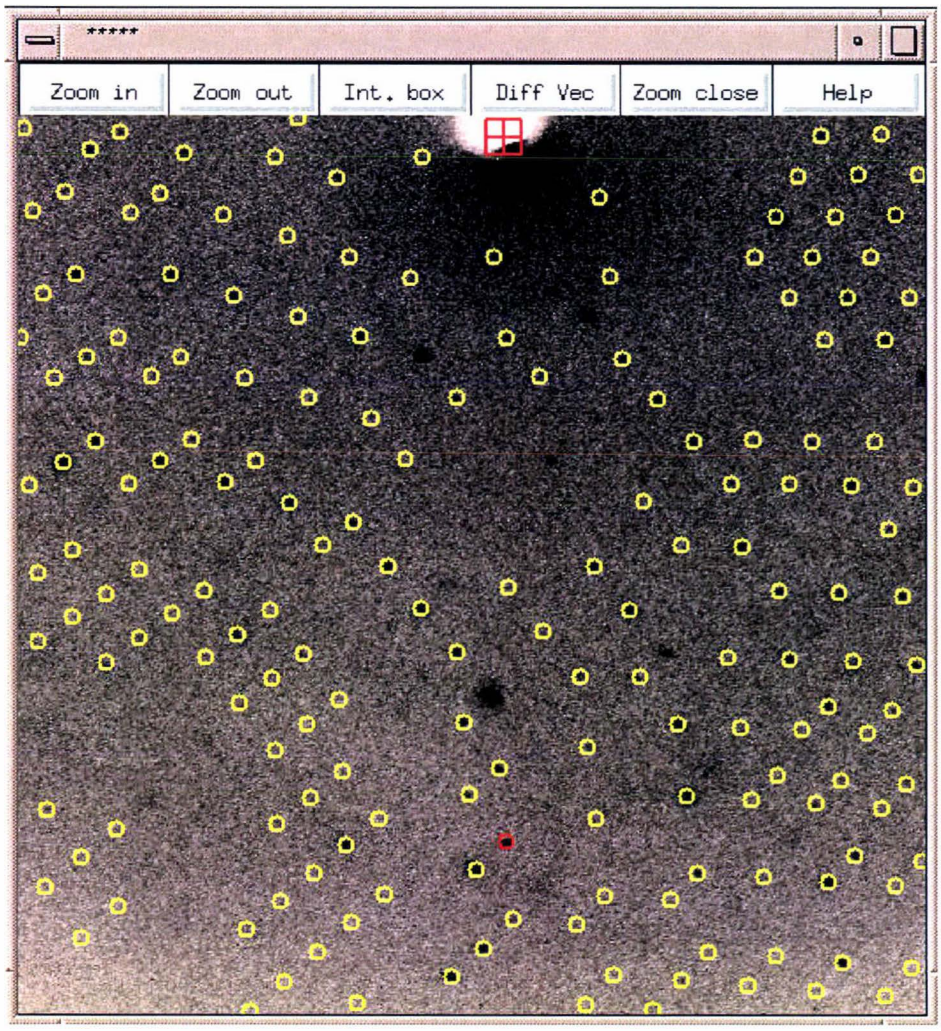


Figure 8.3

BLGA Y′ lattice diffraction image 2

Zoom area of a diffraction frame showing the predicted spots [yellow (non-overlapped) and red (overlapped) circles] for space group $P2_12_12_1$. Note that no diffraction spots are missed by indexing in space group $P2_12_12_1$.

8.3 Solution

The structure was solved with AMORE [Navaza 1994], using the BLGA model (PDB ID: 1bso). Rotation and translation searches were carried out on data in the resolution range 12-4.0 Å (Tables 8.3 and 8.4). The translation search revealed two monomers in the asymmetric unit. Rigid-body refinement of the two monomers using data in the resolution range 12-3.0 Å resulted in an *R* factor of 41 % and correlation coefficient of 64.6 (Table 8.5).

Table 8.3. Solution to the rotation function for BLGA in the Y′ lattice.

α	β	χ	Tx	Ty	Tz	Corr	Rfac	Solution
43.39	87.47	34.35	0.0000	0.0000	0.0000	23.7	0.0	1
136.61	92.53	214.35	0.0000	0.0000	0.0000	23.7	0.0	2
70.85	121.00	220.38	0.0000	0.0000	0.0000	16.4	0.0	3
109.15	59.00	40.38	0.0000	0.0000	0.0000	16.4	0.0	4
15.13	156.51	136.50	0.0000	0.0000	0.0000	14.2	0.0	5
164.87	23.49	316.50	0.0000	0.0000	0.0000	14.2	0.0	6
65.93	49.54	112.40	0.0000	0.0000	0.0000	14.1	0.0	7
114.07	130.46	292.40	0.0000	0.0000	0.0000	14.1	0.0	8
0.78	89.91	24.52	0.0000	0.0000	0.0000	13.8	0.0	9
-0.78	90.09	204.52	0.0000	0.0000	0.0000	13.8	0.0	10
98.82	54.62	159.94	0.0000	0.0000	0.0000	13.0	0.0	11
81.18	125.38	339.94	0.0000	0.0000	0.0000	13.0	0.0	12
57.76	116.50	98.50	0.0000	0.0000	0.0000	13.0	0.0	13
122.24	63.50	278.50	0.0000	0.0000	0.0000	13.0	0.0	14
77.43	120.35	344.60	0.0000	0.0000	0.0000	13.0	0.0	15

The top fifteen solutions were input to the translation function.

Table 8.4. Solution to the translation function for BLGA in the Y′ lattice.

α	β	χ	Tx	Ty	Tz	Corr	Rfac
136.61	92.53	214.35	0.2460	0.3449	0.1407	39.8	47.9
43.39	87.47	34.35	0.2551	0.3454	0.3571	39.7	47.9
109.15	59.00	40.38	0.4199	0.1837	0.3725	16.4	54.8
164.87	23.49	316.50	0.2837	0.2903	0.0290	15.9	55.3
167.61	50.43	141.84	0.2084	0.3394	0.2790	15.6	54.1
48.50	44.10	136.50	0.1419	0.4403	0.1017	15.5	55.4
131.50	135.90	316.50	0.3447	0.3143	0.3270	15.4	55.3
-0.78	90.09	204.52	0.1182	0.2668	0.0839	15.2	55.7
0.78	89.91	24.52	0.1182	0.2332	0.4161	15.2	55.7
118.40	76.73	346.25	0.4595	0.2727	0.1107	15.2	53.7
15.13	156.51	136.50	0.2312	0.4310	0.3313	14.7	55.5
65.93	49.54	112.40	0.0685	0.3806	0.2932	14.7	55.0
70.85	121.00	220.38	0.3339	0.4773	0.4310	14.2	55.7
61.60	103.27	166.25	0.1679	0.4714	0.0943	14.0	55.0
122.24	63.50	278.50	0.0315	0.1325	0.3266	13.4	55.3

The best solutions from the translation function were refined as rigid bodies.

Table 8.5. Final solutions for BLGA in the Y′ lattice.

α	β	χ	Tx	Ty	Tz	Corr	Rfac	Solution
45.96	85.69	31.76	0.2530	0.3453	0.3580	64.6	41.0	1
134.03	94.33	211.77	0.2471	0.3453	0.1420	64.6	41.0	2

Each solution was applied to the search model structure and the resultant coordinates generated. Then all the symmetry equivalents were generated and the two molecules closest to the origin were chosen for further refinement.

8.4 Refinement

The model was initially refined by simulated-annealing with CNS [Brünger *et al.* 1998], achieving R = 0.382, R_{free} = 0.416. The standard parameters used here and for other BLG structures when utilizing simulated annealing include: non-crystallographic symmetry (NCS) set to "on", anisotropic bulk-solvent correction applied to data in the range 6-3.0 Å, 100 steps for initial minimisation, 100 steps for geometry minimisation, "slow cool" molecular dynamics scheme, starting temperature of 1800 K, cooling rate of 25 K per step, 100 steps for final dynamics, target "MLF", and X-ray weight set to auto. Several rounds of model building with TURBO-FRODO [Roussel & Cambillau 1989] followed by *xyz* and grouped *B*-factor refinement brought the R and R_{free} to 0.308 and 0.3682 respectively. A total of 18 water molecules was added with the "water-pick" procedure of CNS based on peak heights of 2σ in a $2F_o$-F_c map. All water molecules were *xyz*-fixed and *B*-factors only were refined. Water molecules with poor density or unfavourable protein contacts were deleted. The final model with individually refined but tightly restrained *B*-factors brought R and R_{free} to values of 0.267 and 0.293. All refinements were done with strict (high energy) two-fold NCS restraints applied to the whole molecule. Two sections of each molecule could not be defined in the electron density. Amino acids 61-65 and 111-115, were removed for final refinement and defined as breaks in the structure. The RMS deviations for bond lengths and angles are 0.015 Å

and 1.6°, respectively. A summary of refinement statistics is given in Table 8.6. Figures 8.4 and 8.5 show the real-space *R* factors per residue and representative electron density in the β-barrel.

Other software used: MTZDUMP and MTZUTILS for inspection and modification of MTZ files, IPDISP for display of image-plate frames, HKLVIEW for display of processed data, BAVERAGE for calculation of average B values, MTZ2VARIOUS for conversion of file formats, RASMOL for display of molecules, XLOGGRAPH, PDBSET [CCP4 1994], LSQMAN [Kleywegt 1996] for calculation of molecular superpositions, MOLEMAN [Kleywegt 1999] and DATAMAN [Kleywegt & Jones 1999] for manipulation of coordinate files, and MAPPAGE [Roussel & Cambillau 1989] for preparation of electron density maps for TURBO-FRODO.

Table 8.6. Refinement statistics for BLGA in the Y′ lattice

Reflections	9398
Free *R* factor reflections	346
Resolution limits	50-2.6 Å
R factor	0.267
Free *R* factor	0.293
Total number non-H atoms	2323
Number water molecules	18
R.m.s deviations from ideals	
Bond lengths	0.015 Å
Bond angles	1.6°
Dihedral angles	25.9°
Improper angles	1.1°
Average *B*-factors (Å^2)	
Main chain	34.1
Side chain	42.8
Water molecules	34.7
Ramachandran plot	Figure 8.6
Most favoured (%)	85.6
Allowed (%)	11.7
Generously allowed (%)	1.5
Disallowed (%)	1.1

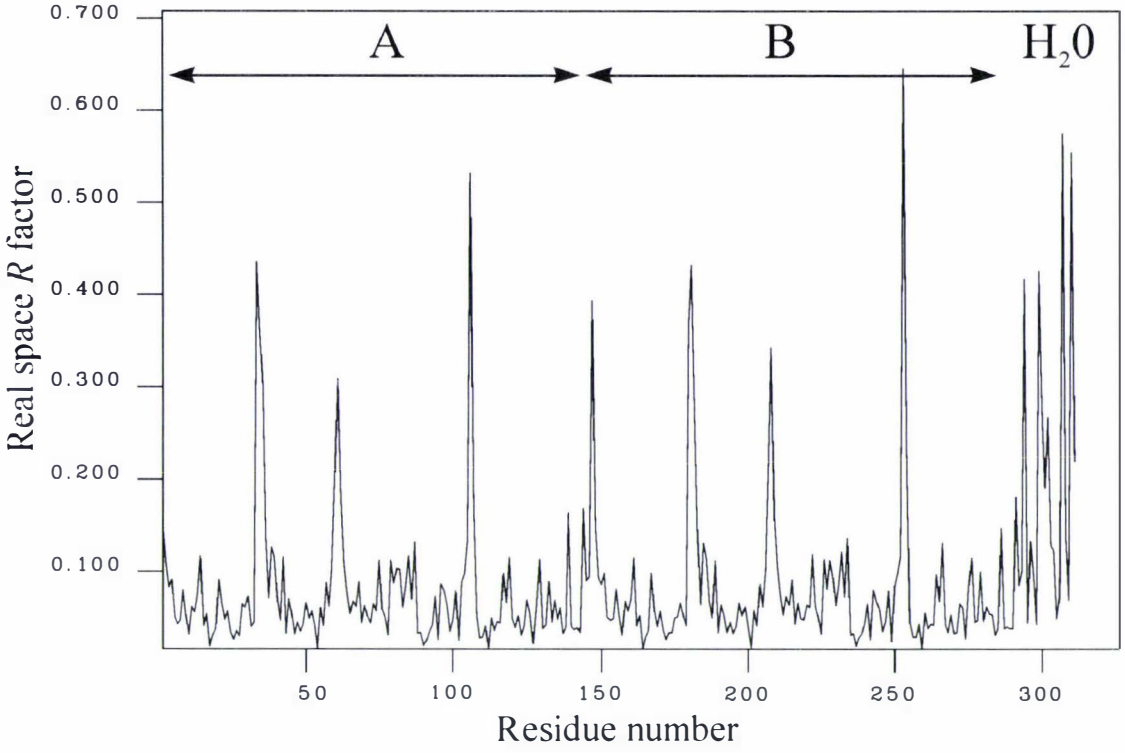


Figure 8.4

Real space R-factor for each residue of BLGA in the Y′ lattice, covering chains A-B and water molecules.

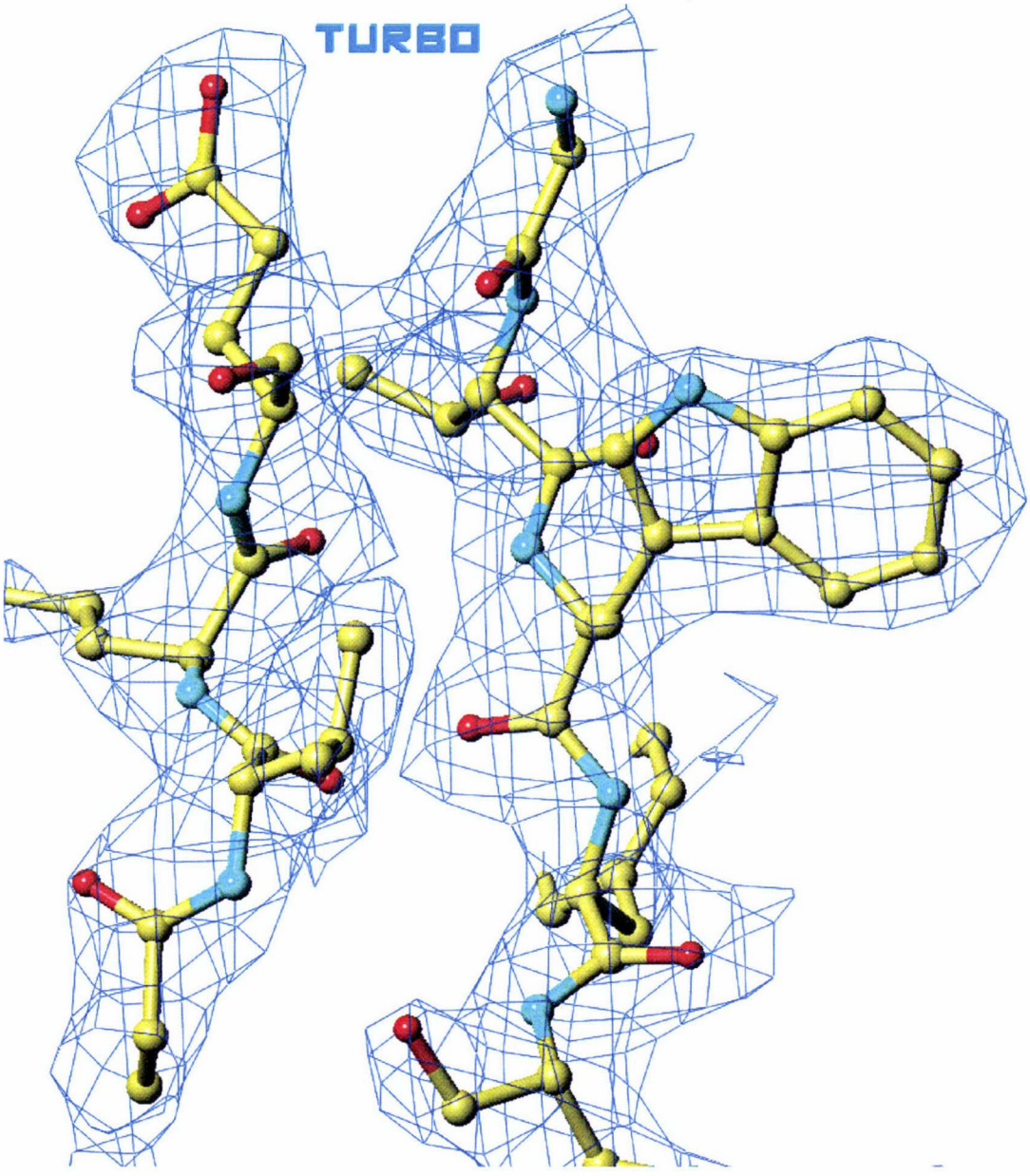


Figure 8.5

Typical electron density of the β-barrel about Trp 19 of BLGA in the Y′ lattice. The electron density is contoured at 1.2 σ.

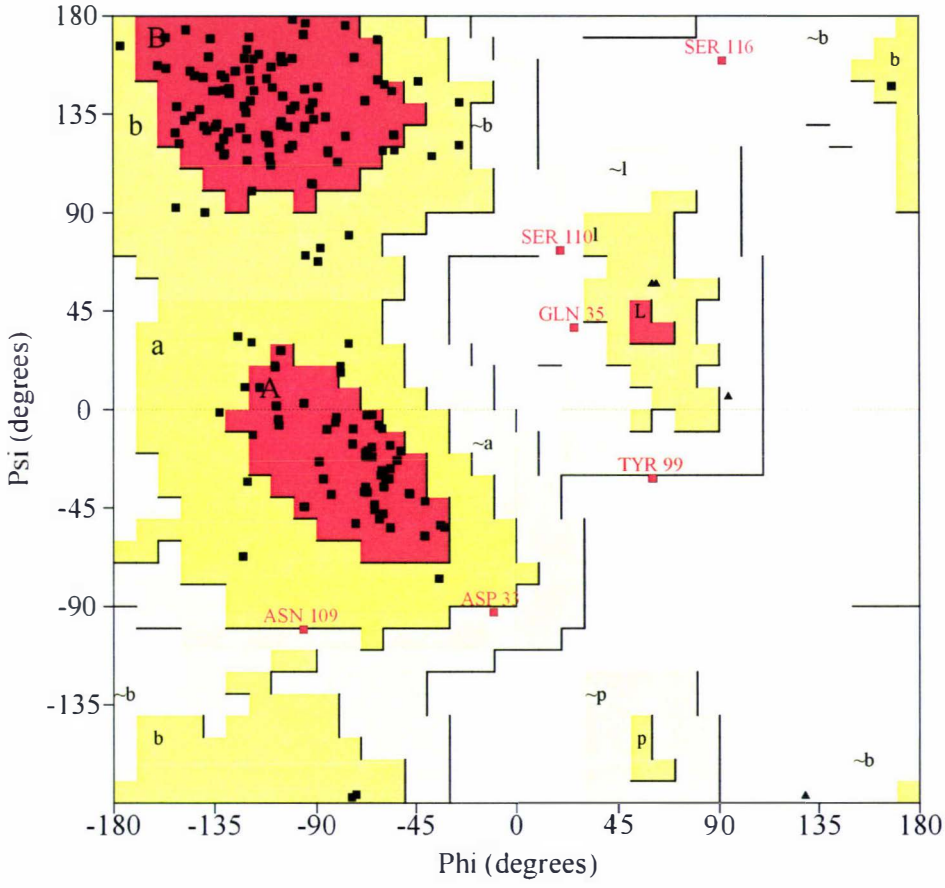


Figure 8.6

Ramachandran plot of BLGA residues from the Y′ lattice. Labelled residues are D33, Q35, N109, S110, S116 and Y99. D33 and Q35 on the AB loop form part of the dimer interface and will be discussed later. N109, S110 and S116 are the terminal residues of the break in electron density, and should be disregarded. Y99 forms a gamma turn, a conserved feature of BLG.

Chapter 9
The structure of BLGA U′ lattice at 3.0 Å resolution

9.1 Crystallisation

Crystals were grown by micro dialysis (Hampton Research, Dialysis Buttons) of a solution of BLGA (20 mg mL^{-1} in 0.1 M Tris pH 7.0) against ultra purified water. After one week the water was aspirated away and replaced with new ultra purified water. Crystals grew to a size of 300x100x100 microns in approximately four weeks.

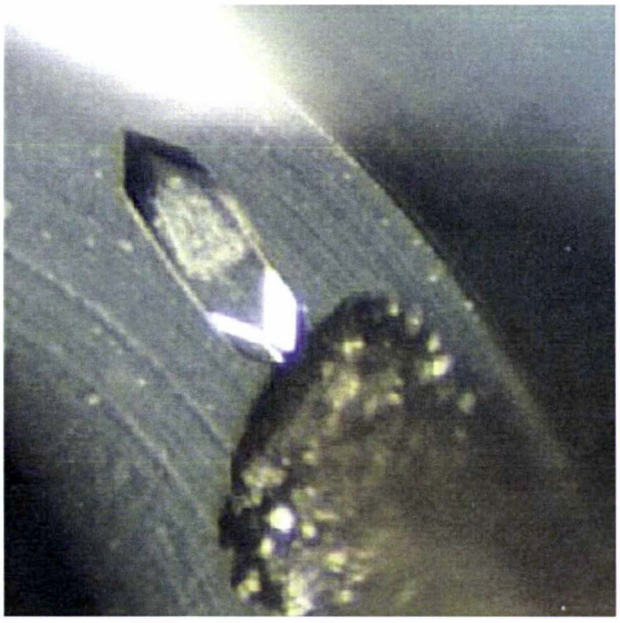


Figure 9.1

BLGA U lattice crystals

9.2 Data collection and processing

A single crystal (Figure 9.1) was mounted in a nylon loop, with a cryoprotecting solution consisting of the well solution from which the crystal was grown and 40% glycerol. The crystal was soaked in the cryoprotectant for approximately 30-50 seconds then snap frozen in the cryo-stream (Oxford Cryosystems). Three crystals were used. The first crystal had data collected from it with a tetragonal data collection strategy; this crystal was unintentionally damaged during cryo-recovery for permanent storage. After the orthorhombic lattice was discovered, a second crystal was mounted for data collection. Only a small fraction of the required data were collected from this crystal when failure of the cryostream destroyed the crystal. A third crystal was mounted and the data collection finished. Data were collected on a Rigaku RU200 rotating-anode generator running at 1.1 kW Cu-Kα, 0.1 mm focus, fitted with an AXCO SPA 50 mono-capillary X-ray optic coupled with a Rigaku RAxis IIC image-plate detector. X-ray frames were taken of each crystal (30 minute, 0.5° oscillation).

Each set of frames was indexed separately, with the cell and orientation matrix determined by the auto-index procedure in DENZO [Otwinoski & Minor 1997]. The resulting data files were scaled and merged together with SCALEPACK [Otwinoski & Minor 1997]. Unit cell dimensions were averaged over all data. Relevant statistics are in Table 9.1. Measured intensities were converted to amplitudes using the program TRUNCATE [CCP4 1994].

The data collected to 3.0 Å resolution were originally identified as the tetragonal U lattice. The data were merged and scaled into the space group $P4_22_12$. However, no solution could be found in this space group. After trials with other tetragonal space groups, it was decided to try the orthorhombic lattice and space group $P2_12_12_1$ The data were re-merged and scaled into $P2_12_12_1$ and the solution reattempted.

Table 9.1. Data collection statistics for BLGA U′ lattice.

Space group	$P2_12_12_1$	
Unit cell	$a = 68.19$ Å	$\alpha = 90°$
	$b = 68.23$ Å	$\beta = 90°$
	$c = 131.46$ Å	$\gamma = 90°$
Data collection		
Observed reflections	151099	
Unique reflections	12501	
Completeness	97.4 %	(96.8 %)[a]
R_{merge} on I	0.097	(0.298)[a]
Overall I/σ	10.3	(3.0)[a]
Wilson B (Å^2)	33.50	
Redundancy	12.0	(3.0)[a]

[a] Outermost shell (3.11 – 3.0 Å)

9.3 Solution

The structure was solved with AMORE [Navaza 1994], using the BLGA model (PDB ID: 1beb). Rotation and translation searches were carried out on data in the resolution range 12-4.0 Å (Tables 9.2 and 9.3). The translation search revealed four monomers in the asymmetric unit. Rigid-body refinement of the four monomers using data in the resolution range

12-3.5 Å resulted in an *R* factor of 39.8 % and correlation coefficient of 54.4 (Table 9.4).

Table 9.2. Solutions to the rotation function for BLGA U′ lattice.

α	β	χ	Tx	Ty	Tz	Corr	Rfac	Solution
28.81	63.11	167.04	0.0000	0.0000	0.0000	11.8	51.5	1
117.55	63.36	166.74	0.0000	0.0000	0.0000	10.0	52.1	2
56.19	71.54	279.76	0.0000	0.0000	0.0000	9.8	51.9	3
145.85	72.82	282.24	0.0000	0.0000	0.0000	10.9	51.5	4
103.63	80.97	255.07	0.0000	0.0000	0.0000	8.9	52.3	5
76.71	87.43	77.73	0.0000	0.0000	0.0000	8.2	52.4	6
71.00	46.40	96.18	0.0000	0.0000	0.0000	7.4	52.7	7
14.29	80.40	256.37	0.0000	0.0000	0.0000	9.6	52.1	8
104.22	90.00	255.53	0.0000	0.0000	0.0000	7.6	52.4	9
42.60	73.20	147.37	0.0000	0.0000	0.0000	8.5	52.2	10
164.40	87.75	78.50	0.0000	0.0000	0.0000	8.3	52.3	11
132.99	73.13	146.73	0.0000	0.0000	0.0000	8.0	52.3	12
17.33	90.00	255.32	0.0000	0.0000	0.0000	7.6	52.5	13
54.19	21.76	111.08	0.0000	0.0000	0.0000	7.0	52.6	14
164.47	67.07	221.31	0.0000	0.0000	0.0000	8.3	52.2	15
168.81	10.74	185.24	0.0000	0.0000	0.0000	6.9	53.0	16
159.61	46.53	96.39	0.0000	0.0000	0.0000	8.0	52.5	17
111.38	56.02	301.51	0.0000	0.0000	0.0000	5.8	53.2	18
18.74	77.03	124.50	0.0000	0.0000	0.0000	8.6	52.3	19
101.83	76.28	307.41	0.0000	0.0000	0.0000	7.3	52.7	20

Table 9.3. Solutions to the translation function for BLGA U′ lattice.

Solution 1

α	β	χ	Tx	Ty	Tz	Corr	R fac
28.81	63.11	167.04	0.4389	0.4713	0.2656	18.2	51.2
56.19	71.54	279.76	0.2545	0.3610	0.4853	17.9	50.4
145.85	72.82	282.24	0.3891	0.0040	0.2352	17.4	50.8
42.60	73.20	147.37	0.4910	0.0880	0.3282	15.1	52.0
117.55	63.36	166.74	0.1658	0.1075	0.0564	14.2	51.8
132.99	73.13	146.73	0.1611	0.2405	0.4132	13.9	52.5
14.29	80.40	256.37	0.1965	0.1260	0.2343	13.8	51.5
18.74	77.03	124.50	0.1372	0.1976	0.0000	13.2	52.3
17.33	90.00	255.32	0.0130	0.1076	0.1721	13.2	51.9
159.61	46.53	96.39	0.3491	0.1087	0.0457	13.1	52.1
71.00	46.40	96.18	0.3786	0.2501	0.4852	13.0	52.2
103.63	80.97	255.07	0.3309	0.4460	0.4567	13.0	51.8
164.40	87.75	78.50	0.4011	0.2214	0.0265	12.6	52.4
164.47	67.07	221.31	0.0264	0.2278	0.1500	12.6	52.0
76.71	87.43	77.73	0.3487	0.1500	0.1319	12.5	52.0
104.22	90.00	255.53	0.2369	0.0313	0.1900	12.4	51.9
54.19	21.76	111.08	0.1399	0.3040	0.4395	11.7	52.5
101.83	76.28	307.41	0.0015	0.3037	0.4852	11.6	52.8
168.81	10.74	185.24	0.0716	0.1825	0.2235	10.6	52.7
111.38	56.02	301.51	0.2421	0.4086	0.3420	10.5	52.6

Solution 2

α	β	χ	Tx	Ty	Tz	Corr	Rfac
145.85	72.82	282.24	0.8902	0.5259	0.2348	28.8	47.5
56.19	71.54	279.76	0.2652	0.8597	0.4856	27.1	48.0
117.55	63.36	166.74	0.2790	0.6861	0.0185	25.5	49.1
76.71	87.43	77.73	0.5019	0.0202	0.3132	19.8	50.6
14.29	80.40	256.37	0.5886	0.4563	0.2435	19.5	50.7
17.33	90.00	255.32	0.8923	0.9553	0.7977	19.4	50.2
103.63	80.97	255.07	0.1031	0.3420	0.7063	19.3	50.6
18.74	77.03	124.50	0.5276	0.2988	0.3271	19.3	50.4
164.47	67.07	221.31	0.8981	0.4385	0.7199	19.3	50.4
132.99	73.13	146.73	0.5677	0.3629	0.2575	19.0	51.0
42.60	73.20	147.37	0.1423	0.0585	0.7283	18.9	50.7
71.00	46.40	96.18	0.9715	0.6611	0.6794	18.7	51.2
164.40	87.75	78.50	0.5030	0.7257	0.1600	18.5	51.0
159.61	46.53	96.39	0.5229	0.4093	0.2352	18.5	50.8
168.81	10.74	185.24	0.4756	0.9896	0.2225	18.3	51.3
104.22	90.00	255.53	0.3131	0.4705	0.7181	18.2	51.2
101.83	76.28	307.41	0.2229	0.5343	0.7103	18.1	51.3
111.38	56.02	301.51	0.4023	0.4188	0.2006	17.7	51.2
54.19	21.76	111.08	0.9742	0.3233	0.7661	17.6	51.4

Solution 3

α	β	χ	Tx	Ty	Tz	Corr	Rfac
117.55	63.36	166.74	0.2823	0.6860	0.0184	35.7	45.9
76.71	87.43	77.73	0.6404	0.4053	0.7229	27.5	48.4
103.63	80.97	255.07	0.1223	0.1479	0.3040	27.3	48.5
132.99	73.13	146.73	0.9761	0.5589	0.4715	27.2	48.0
164.47	67.07	221.31	0.8543	0.4854	0.7544	27.1	48.4
104.22	90.00	255.53	0.3783	0.9800	0.7850	27.1	48.4
101.83	76.28	307.41	0.4580	0.8407	0.2226	26.6	48.7
17.33	90.00	255.32	0.1732	0.7576	0.8287	26.5	48.2
18.74	77.03	124.50	0.9467	0.4740	0.7844	26.4	49.4
159.61	46.53	96.39	0.9384	0.3705	0.7739	26.4	48.6
42.60	73.20	147.37	0.9810	0.4632	0.7772	26.4	48.6
164.40	87.75	78.50	0.5039	0.7232	0.1619	26.3	48.8
14.29	80.40	256.37	0.3906	0.5467	0.2814	26.0	48.9
71.00	46.40	96.18	0.7533	0.4007	0.6854	25.9	49.1
54.19	21.76	111.08	0.0466	0.0284	0.7036	25.7	49.2
111.38	56.02	301.51	0.5880	0.4590	0.2431	25.4	48.9
168.81	10.74	185.24	0.4742	0.9903	0.2218	25.3	49.0

Solution 4

α	β	χ	Tx	Ty	Tz	Corr	Rfac
56.19	71.54	279.76	0.2652	0.8622	0.4850	44.4	43.1
14.29	80.40	256.37	0.1784	0.4660	0.8415	33.3	47.1
76.71	87.43	77.73	0.6410	0.4042	0.7234	33.3	46.9
111.38	56.02	301.51	0.5852	0.9212	0.2448	33.1	46.7
103.63	80.97	255.07	0.3130	0.8503	0.8079	33.1	46.5
42.60	73.20	147.37	0.4114	0.2596	0.4387	33.0	47.0
104.22	90.00	255.53	0.7474	0.3855	0.6827	32.9	46.9
18.74	77.03	124.50	0.1801	0.9486	0.2912	32.9	46.8
117.55	63.36	166.74	0.1307	0.6856	0.0203	32.8	48.1
17.33	90.00	255.32	0.9949	0.2229	0.2890	32.8	47.1
164.47	67.07	221.31	0.5481	0.6321	0.2512	32.7	47.2
168.81	10.74	185.24	0.4749	0.9905	0.2223	32.6	47.1
164.40	87.75	78.50	0.5138	0.8806	0.1508	32.6	46.8
71.00	46.40	96.18	0.7280	0.4431	0.6964	32.5	46.8
159.61	46.53	96.39	0.2556	0.8940	0.1580	32.4	47.4
132.99	73.13	146.73	0.5669	0.3622	0.2568	32.4	47.3
54.19	21.76	111.08	0.2131	0.8006	0.9449	31.9	47.7
101.83	76.28	307.41	0.4594	0.8398	0.2224	31.8	47.1

The best solutions from the translation function from each set were refined as rigid bodies.

Table 9.4. Final solutions for BLGA U′ lattice.

α	β	χ	Tx	Ty	Tz	Corr	Rfac	Solution
30.32	62.21	168.48	0.4431	0.4698	0.2677	54.4	39.8	1
146.06	70.90	281.77	0.8943	0.5301	0.2347	54.4	39.8	2
115.83	61.79	170.27	0.2916	0.6883	0.0200	54.4	39.8	3
56.56	69.04	279.41	0.2595	0.8621	0.4854	54.4	39.8	4

Each solution was applied to the search model structure and the resultant coordinates generated. Then all the symmetry equivalents were generated and the two dimers closest to the origin were chosen for further refinement (Figure 9.2).

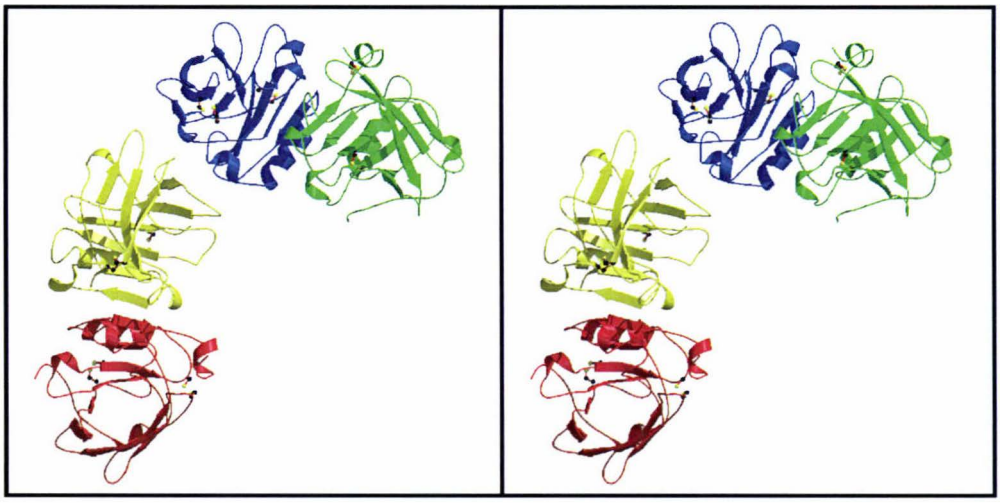


Figure 9.2

Stereo view of the asymmetric unit contents of BLGA U′ lattice.

9.4 Refinement

The model was refined by simulated annealing with CNS [Brunger *et al.* 1998] (using the standard parameters described in Chapter 8), achieving R = 0.322 and R_{free} = 0.398 for data in the resolution range 50 to 3.0 Å. Several rounds of model building with TURBO-FRODO [Roussel &

Cambillau 1989] followed by *xyz* and grouped *B*-factor refinement brought the *R* and R_{free} to 0.304 and 0.321, respectively. A total of 55 water molecules were added with the "water-pick" procedure of CNS based on peak heights of 2σ in a $2F_o$-F_c map. All water molecules were *xyz*-fixed and *B*-factors only were refined. Water molecules with poor density or unfavourable protein contacts were deleted. The final model with individually refined but tightly restrained *B*-factors brought *R* and R_{free} to values of 0.293 and 0.313. All refinements were done with strict (high energy) four-fold NCS restraints applied to the whole molecule. The RMS deviations for bond lengths and angles are 0.012 Å and 1.6° respectively. A summary of refinement statistics is given in Table 9.5. Figures 9.3 and 9.4 show the real-space *R* factors per residue and representative electron density in the β-barrel. The real space *R* factors being between 0 and 0.4 over the four polypeptide chains is indicative of very good fit of the atoms to the electron density.

Other software used: MTZDUMP and MTZUTILS for inspection and modification of MTZ files, IPDISP for display of image-plate frames, HKLVIEW for display of processed data, BAVERAGE for calculation of average B values, MTZ2VARIOUS for conversion of file formats, RASMOL for display of molecules, XLOGGRAPH, PDBSET [CCP4 1994], LSQMAN [Kleywegt 1996] for calculation of molecular superpositions, MOLEMAN [Kleywegt 1999] and DATAMAN [Kleywegt & Jones 1999] for manipulation of coordinate files, and MAPPAGE [Roussel & Cambillau 1989] for preparation of electron density maps for TURBO-FRODO.

Table 9.5. Refinement statistics for BLGA U′ lattice.

Reflections	12239
Free *R* factor reflections	619
Resolution limits	50-3.0 Å
R factor	0.293
Free *R* factor	0.313
Total number non-H atoms	4991
Number water molecules	55
R.m.s deviations from ideals	
Bond lengths	0.012 Å
Bond angles	1.6°
Dihedral angles	25.4°
Improper angles	0.9°
Average *B*-factors (Å^2)	
Main chain	28.3
Side chain	30.2
Water molecules	18.4
Ramachandran plot	Figure 9.5
Most favoured (%)	78.2
Allowed (%)	19.7
Generously allowed (%)	0.7
Disallowed (%)	1.4

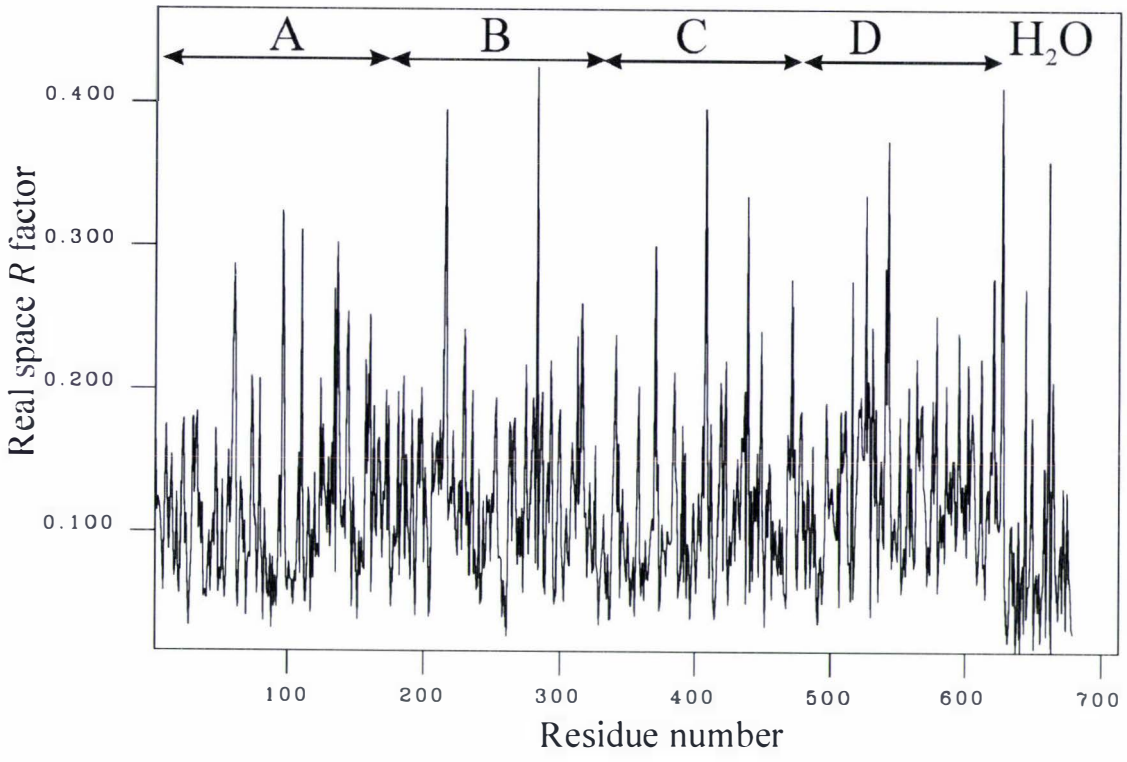


Figure 9.3

Real space R-factor for each residue for BLGA U′ lattice, covering chains A-D and water molecules.

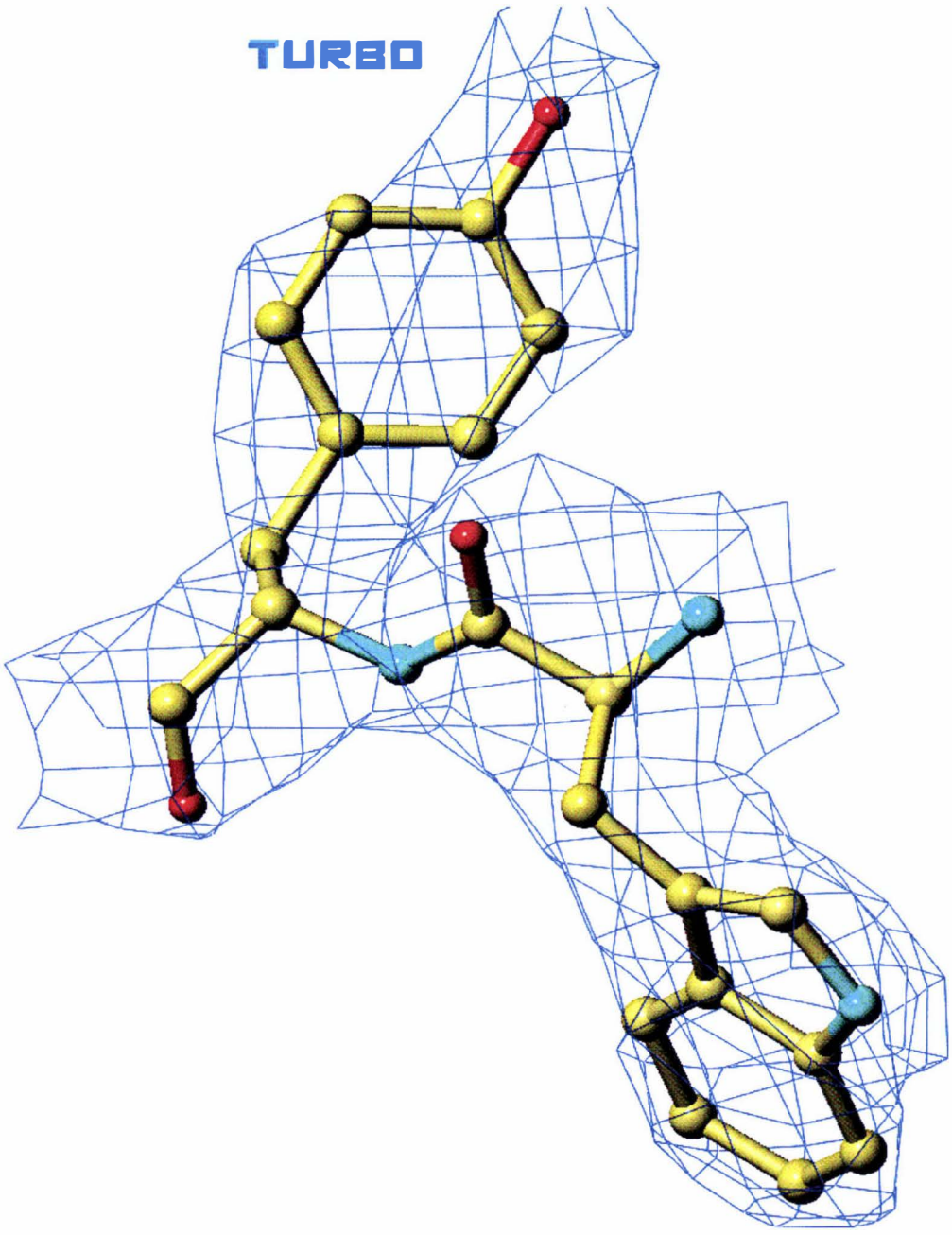


Figure 9.4

Typical electron density of the β-barrel about Trp19 for BLGA U′ lattice.

The electron density is contoured at 1.2 σ.

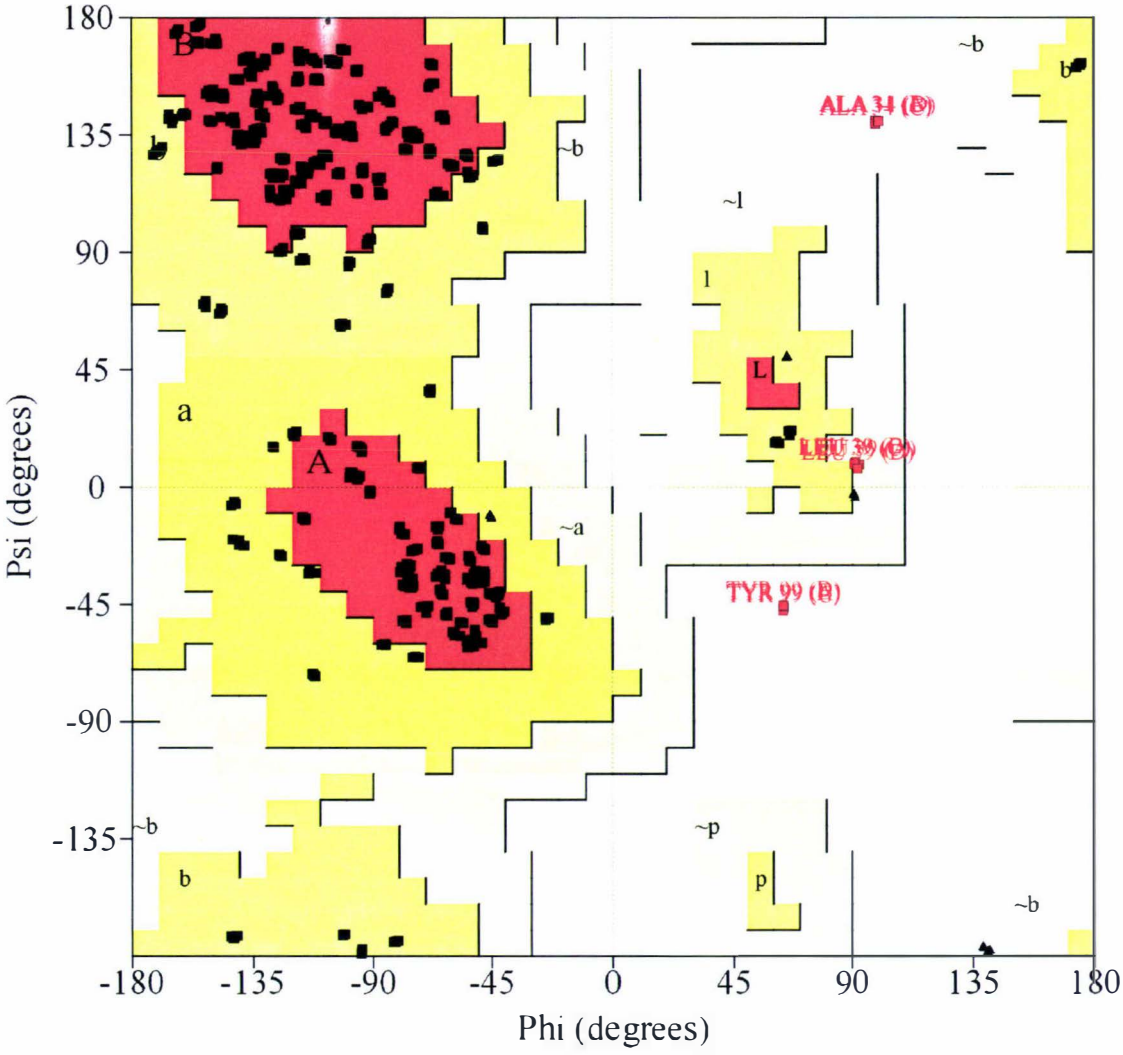


Figure 9.5

Ramachandran plot of BLGA residues for BLGA U′ lattice.

Highlighted are residues Y99, which is part of a γ turn, A34, which is part of the dimer interface and is geometrically strained, and L39, which lies just in the generously allowed region.

Chapter 10

The preliminary X-ray analysis of BLGA J lattice at 3.0 Å resolution

10.1 Crystallization

The protein used is from the same batch used previously.

Crystals were grown by micro dialysis (Hampton Research, Dialysis Buttons) of a solution of BLGA (20 mg mL^{-1} in 0.1 M Tris pH 7.0) against ultra purified water spiked with retinoic acid solution. After one week the water was aspirated away and replaced with new ultra purified water and again spiked with retinoic acid solution. Crystals grew to a size of 800x800x300 microns in approximately four weeks. The crystallization was protected from light by wrapping the box in foil.

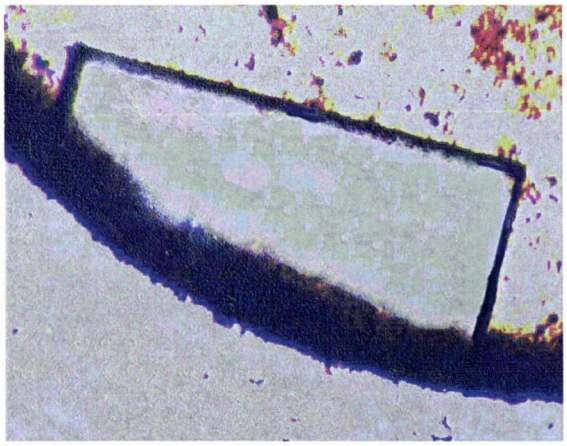


Figure 10.1

BLGA J lattice crystals

10.2 Data collection and processing

A single crystal (Figure 10.1) was mounted in a nylon loop, with a cryoprotecting solution consisting of the well solution from which the

crystal was grown and 40% glycerol. The crystal was soaked in the cryoprotectant for approximately 30-50 seconds then snap frozen in the cryo-stream (Oxford Cryosystems). Data were collected on a Rigaku RU200 rotating-anode generator running at 1.1 kW Cu-Kα, 0.1 mm focus, fitted with an AXCO SPA 50 mono-capillary X-ray optic coupled with a Rigaku RAxis IIC image-plate detector. X-ray frames were taken of each crystal (30 minute, 0.5° oscillation).

The cell and orientation matrix were determined by the auto-index procedure in DENZO [Otwinoski & Minor 1997]. The resulting data files were scaled and merged together with SCALEPACK [Otwinoski & Minor 1997].

Table 10.1. Data collection statistics for BLGA J lattice.

Space group	?	
Unit cell	$a = 36.0$ Å	$\alpha = 90°$
	$b = 36.0$ Å	$\beta = 90°$
	$c = 229.7$ Å	$\gamma = 90°$

Due to the long cell edge the data and high mosaicity (~1°) the space group has not been determined. The data set is also incomplete due to the long cell edge causing overlaps. A new attachment for the goniometer head has been manufactured and the process of completing the data set is ongoing. Many different solutions have been attempted. However, none of them led to a viable solution. Further data collections to minimize overlap of data spots have not resolved the space group and the structure-solution difficulties. Recollection of the data using a finer slicing (0.1-0.2°) may help to minimize spot overlap.

Chapter 11

Site-directed mutagenesis to produce the K69E mutant of bovine BLGA.

11.1 Point mutations in recombinant BLGA

It was intended to make 3 mutants, K60L, K69E and E62L. It is anticipated that K60L and K69E will lower the binding constant for ligand binding and render the protein inactive with respect to fatty acid binding. The E62L may have the opposite effect and raise the binding constant considerably, perhaps to the point where ligand binding is irreversible. Due to time constraints only K69E was made.

11.2 Expression System

The expression system being used is the ThioFusion™ expression system (Invitrogen), with the GI724 strain of *E. coli*. The target protein is expressed at the C-terminus of thioredoxin, an 11.7 kD protein that is extremely soluble in *E. coli*. The properly folded thioredoxin is intended to aid correct folding of the target protein.

The plasmid used to create thioredoxin fusions, pTrxFus, uses the P_L promoter from bacteriophage to drive expression. This promoter is one of the most efficient promoters for bacterial expression. It is also tightly regulated. pTrxFus is propagated in *E. coli* cells (either GI698 or GI724) where the *c*I repressor gene is under control of the *trp* promoter. The *trp* repressor and attenuation regulate expression of the *c*I repressor. When cells are grown in tryptophan-free medium, the *c*I repressor gene is transcribed, and the *c*I repressor protein binds to the P_L promoter

preventing transcription. Expression is induced by adding tryptophan to the medium which shuts down *c*I repressor synthesis.

11.3 Site-directed mutagenesis.

The ThioFusion™ (Invitrogen) expression system is not amenable to a one-step site-directed mutagenesis (SDM) of the circular plasmid [Kunkel 1985]. The mutagenesis must be done on the isolated gene of interest. The mutagenesis is achieved by three-step polymerase chain reaction (PCR) [Grandori *et al.* 1997] (Figure 11.1).

Step 1

Half-gene PCR mutagenesis using a forward primer with the mutation and a reverse primer from 5' end.

Step 2

Half-gene PCR mutagenesis using a forward primer from 5' end and a reverse primer with the mutation.

Step 3

PCR using the purified ½ genes as the template with forward and reverse end primers.

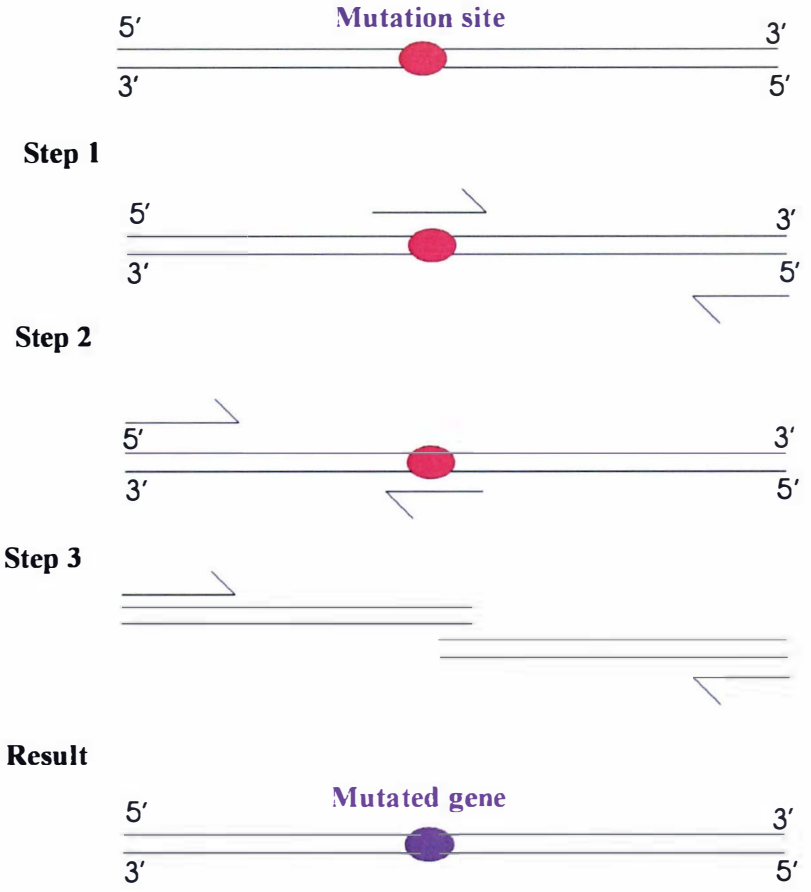


Figure 11.1

The three-step PCR process for site-directed mutagenesis.

A forward primer BLGA_K69E_F and reverse primer BLGA_K69E_R spanning the mutation site were designed. End primers KPN1FP and BAMH1RP were supplied by Dr K A N S Ariyaratne.

Cells (GI724) containing the BLGA_pTrxFus plasmid were cultured overnight and the plasmid removed from the culture by mini-prep (Biorad). The plasmid was then sequentially digested with KpnI (GibcoBRL) and BamHI (Roche) DNA restriction enzymes to yield both double-digested pTrxFus plasmid and the BLGA gene to be used as the PCR template.

The three-step SDM procedure divides the gene to two parts. K69E_1

(BLGA_K69E_F, BAMH1RP) and K69E_2 (BLGA_K69E_R, KPN1FP). For each part K96E_1 and K69E_2, a PCR reaction mixture was prepared containing the primers, 100 pmol L^{-1} (2 μL), mixed with ~2 ng (2 μL) of DNA template (BLGA), 5 mL 10 × *Pfx* amplification buffer, 1 μL 0.05 M $MgSO_4$, 300 μM dNTP's (1 μL), 1 Unit *Pfx* Platinum DNA polymerase (GibcoBRL) (1 μL), H_2O (37 μL). The samples were subjected to 25 PCR cycles of denaturation at 94° C (1 minute), annealing at 55° C (1 minute), and extension at 72° C (1 minute) in a DNA thermocycler (Perkin Elmer, Cetus).

The PCR product was examined by agarose gel electrophoresis [Maniatis *et al.* 1989] (ethidium-bromide stained) and showed a 150 bp band for K69E_1 and a 250 bp band for K69E_2 (Figure 11.2). K69E_1 was purified by agarose gel electrophoresis [Maniatis *et al.* 1989] and extracted with a Quiagen PCR purification kit.

The full gene PCR consisted of 2 μL each of K69E_1 and K69E_2, 100 pmol L^{-1} (2 μL) each of primers BAMH1RP, and KPN1FP, 5 μL 10 × *Pfx* amplification buffer, 1 μL 0.05 M $MgSO_4$, 300 μM dNTP's (1 μL), 1 Unit *Pfx* Platinum DNA polymerase (1 μL), H_2O (37 μL). The PCR product was purified by agarose gel electrophoresis [Maniatis *et al.* 1989] (Figure 11.3), extracted from the gel by Quiagen PCR purification kit, and analysed by DNA sequencing (ABI systems, big-dye terminator; MuSeq sequencing service of Massey University).

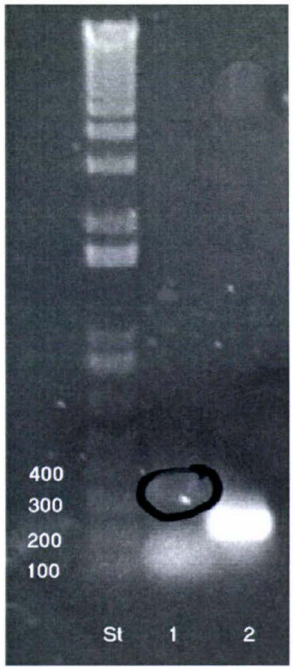


Figure 11.2
PCR products K69E_1 (lane 1) and K69E_2 (lane 2). (St) DNA ladder.

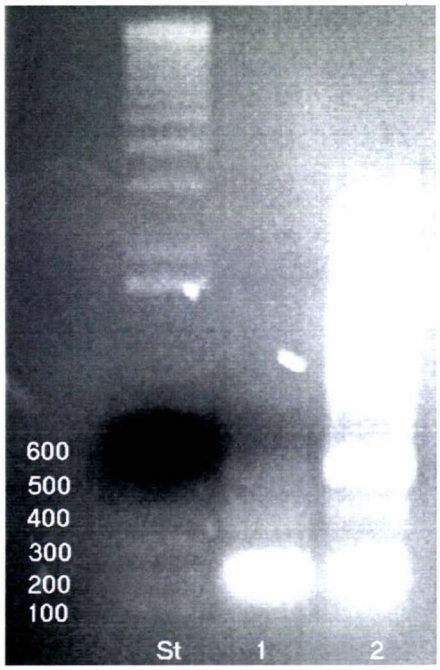


Figure 11.3
Full PCR product (lane 2) and failed PCR (lane 1). (St) DNA ladder.

Bands from 600 Bp to 300 Bp are obscured by the loading dye which travels at approximately ¾ the speed of the 100 Bp DNA fragment.

11.4 Cloning

The double-digested pTrxFus vector isolated above was purified by agarose gel electrophoresis [Maniatis *et al.* 1989] and the mutated synthetic gene was sequentially digested with KpnI (GibcoBRL) and BamHI (Roche) and purified with the Quiagen PCR purification kit. The ligation of the vector (20 ng) and the insert (35 ng) was carried out overnight at 16 °C using 1 unit of T4 DNA ligase in a total volume of 10 µL.

A 110 µL aliquot of G1724 *E. coli* competent cells (prepared following

the standard protocol for preparation of chemically competent cells, as set out in the ThioFusion™ (Invitrogen) expression system manual. No cell density is known or calculated.) was transformed with 10 μL of the ligation mixture by heat shock at 42 °C for 90 seconds. Then, 900 μL of SOC media (2% (w/v) Tryptone, 0.5% (w/v) yeast extract, 0.05 % (w/v) NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) was added to the transformation mixture and incubated at 30 °C for one hour.

A 100 μL aliquot of the transformation mixture was plated on RMG-agar (0.6 % (w/v) Na_2HPO4, 0.3 % (w/v) KH_2PO_4, 0.05 % (w/v) NaCl, 0.1% (w/v) NH_4Cl, 2% (w/v) Casamino acids, 0.0095 % (w/v) $MgCl_2$, 1.5 % (w/v) agar) and incubated at 30 °C for 36 hours. Five colonies were selected at random and grown individually in RM media (0.6 % (w/v) Na_2HPO_4, 0.3 % (w/v) KH_2PO_4, 0.05 % (w/v) NaCl, 0.1% (w/v) NH_4Cl, 2% (w/v) Casamino acids, 0.0095 % (w/v) $MgCl_2$) at 30 °C with shaking (200 – 225 rpm.) for 20 hours.

The plasmid was removed from the cultured cells by mini-prep (Biorad)(Figure 11.5). The plasmid was then sequentially digested with KpnI (GibcoBRL) and BamHI (Roche) DNA restriction enzymes. The presence of the mutated gene is confirmed (Figure 11.6) by the loss of a ~500 Bp fragment in the second digestion. One colony was chosen at random and PCR amplified. The PCR mixture consisted of 2 μL DNA template, 100 pmol L^{-1} (2 μL) each of primers BAMH1RP, and KPN1FP, 5 μL 10 × *Pfx* amplification buffer, 1 μL 0.05 M $MgSO_4$, 300 μM dNTP's (1 μL), 1 Unit *Pfx* Platinum DNA polymerase (1 μL), H_2O (37 μL). The PCR product was purified by agarose gel electrophoresis [Maniatis *et al.* 1989], and DNA sequenced to confirm the mutation and reading frame (Figures 11.4).

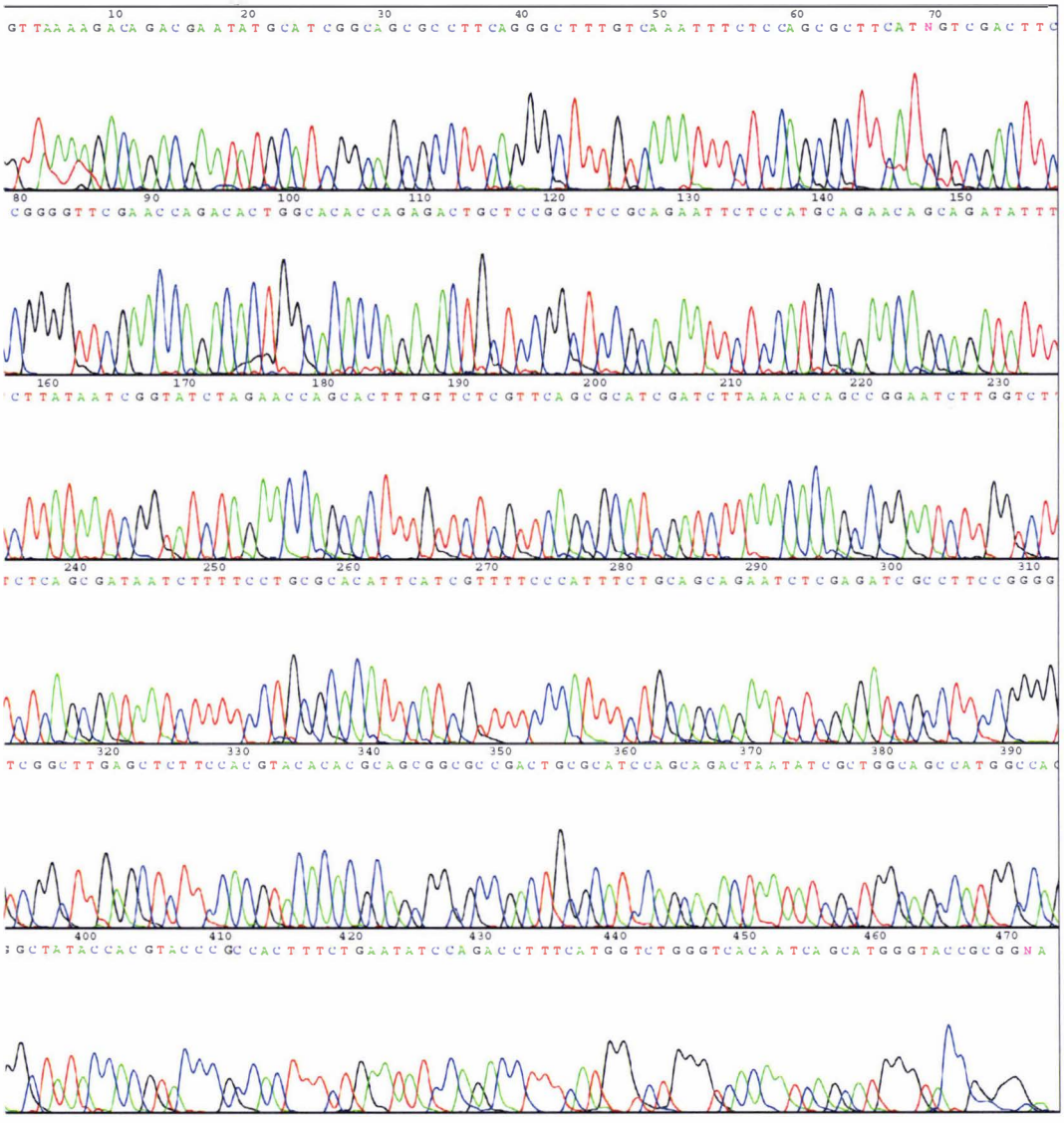


Figure 11.4

DNA sequence of the cloned gene with the K69E point mutation

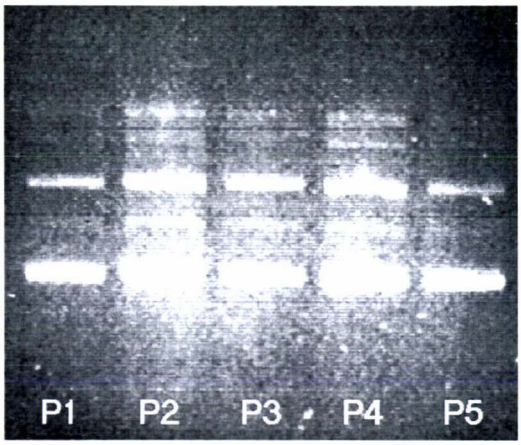


Figure 11.5

Plasmid isolated from 5 randomly chosen transformants.

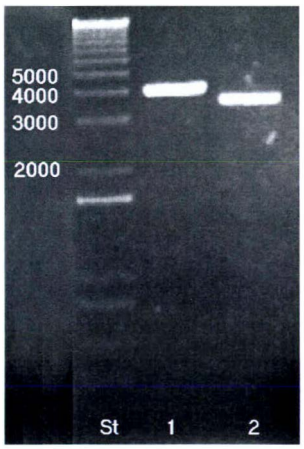


Figure 11.6

Single (1) and double digest (2) of colony 4.

(St) DNA ladder.

11.5 Protein Expression

100 mL of induction media inoculated with 1 mL of the liquid RM clone culture was incubated at 30 °C with shaking until the optical density at 550 nm (OD_{550}) reached 0.5 absorbance units (AU). 1 mL of culture media was removed as a control sample before the cells were induced with DL-tryptophan at a final concentration of 400 μg/mL. The culture was incubated for a further 17 hours at 30 °C with shaking. The culture growth was monitored at intervals 1 hr, 2 hr, 3 hr, and 17 hr by OD_{550} and samples were saved for SDS-PAGE [Laemmli 1970] analysis of protein expression. For whole cell analysis, 500 μL of culture media was centrifuged at 14000 rpm [Eppendorf mini centrifuge] for 2 minutes to pellet the cells. The pelleted cells were then resuspended in 10 μL of deionized water and 10 μL of 2× sample buffer (0.5 M Tris-HCl (pH 6.8), 20% glycerol, 4% SDS, 0.1 % DTT, 0.02 % bromophenol blue). Then the samples were heated at 100 °C for 5 minutes and analysed by SDS-PAGE using 12 % gel [Laemmli 1970] (Figure 11.7). To test for the solubility of the BLG-fusion protein (BLG*), the cells were pelleted by centrifugation,

resuspended in 15 mL of 0.01 M Tris buffer pH 7.0 0.1 M NaCl, containing Complete™ protease inhibitor (Roche) (½ tablet). Cells were lysed by sonication over ice using three 15-second bursts at 110 Watts (Virsonic Digital 550 Ultrasonic Cell Disrupter). Soluble cell lysate was separated from insoluble cell debris by centrifugation at 6000 g (Sorvall RC2B centrifuge / SS34 rotor). Cell debris were resuspended in 15 mL of 0.01 M Tris buffer pH 7.0 0.1 M NaCl buffer and analysed together with the soluble fraction by SDS-PAGE as above (Figure 11.7).

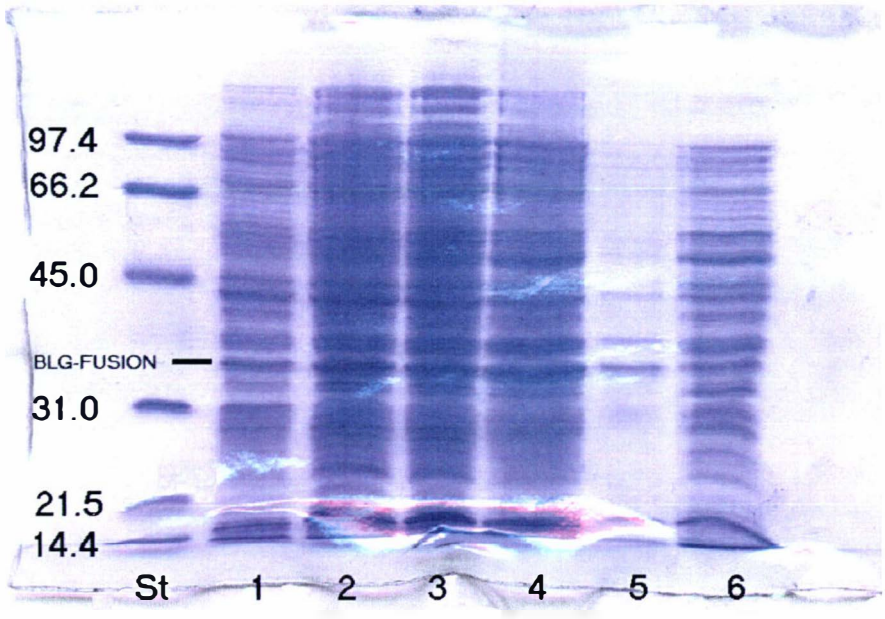


Figure 11.7

SDS-PAGE following induction.

Lanes 1-4 BLG* induction t = 0, 1, 3 and 17 hrs (whole cell).

Lane 5 cell debris at 10X concentration; lane 6 soluble fraction; St molecular mass markers. Coomassie Brilliant Blue staining.

11.6 Protein purification (1)

The soluble cell lysate was diluted until the optical density at 280 nm was 1.5. The pH of the solution was then adjusted to pH 2.7 with a 50% acetic acid solution. 7% NaCl solid (w/v) was added and stirred for 2 hours at room temperature. The solution was then centrifuged at 16000 g (Sorvall RC2B centrifuge / SS34 rotor) to remove the precipitate that had formed. The clear supernatant solution then had NaCl solid added to bring the NaCl percentage (w/v) up to 30%. The precipitated BLG* was then removed from the solution by centrifugation. The precipitate was redissolved in a minimum volume of 0.1 M Tris-HCl (pH 6.8) and loaded onto a Superdex G200 column and eluted with Ekmax buffer (50 mM Tris-HCl, 10 mM $CaCl_2$). Fractions were monitored by optical detection at 280 nm and 114 nm. Fractions containing protein were assayed by SDS-PAGE [Laemmli 1970], and fractions containing BLG* were pooled (Figure 11.8).

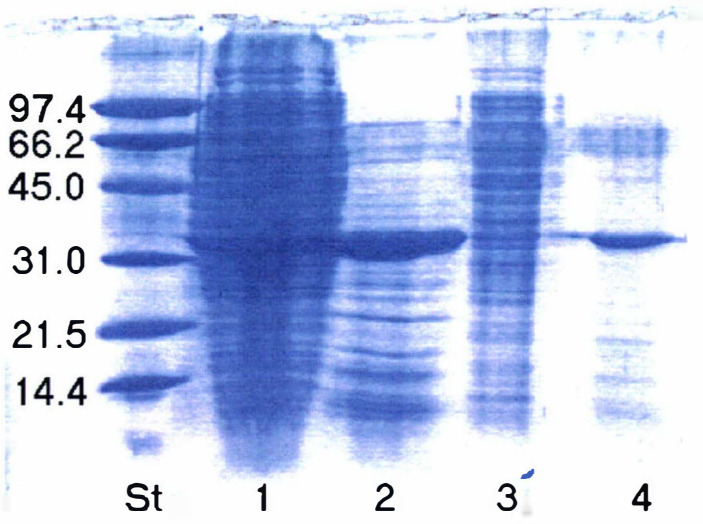


Figure 11.8

SDS-PAGE showing the purification scheme.

Lane 1, soluble protein. Lane 2, soluble protein after pH cut and addition of 7% (w/v) NaCl. Lane 3, insoluble protein after pH cut and addition of 7% NaCl at 10X concentration. Lane 4, soluble protein after G200 column. St, molecular mass markers.

The pooled fractions were concentrated by ultrafiltration (Centricon 10 (Millipore)) until the protein concentration was approximately 4 mg mL^{-1}. The Enterokinase enzyme (Invitrogen) was added (1 U 1 μL) and the solution incubated at 25° C overnight. After incubation overnight, 10 μL of each sample was mixed with 1 μL of SDS sample buffer, and analysed by SDS-PAGE (Figure 11.9). The SDS-PAGE analysis showed the thioredoxin had been cleaved from the BLG*, but the BLG* had been also destroyed by proteolysis. The control sample of BLG* without enterokinase had not degraded.

11.7 Protein purification (2)

A second purification protocol was developed without the acid precipitation step, as it was a possibility that this step was partially unfolding the protein.

The soluble cell lysate was diluted 20:1 with 10 mm Bis-tris (pH 6.5) and loaded onto a 7 mL Source-Q column and eluted with a linear gradient of 0 – 1 M NaCl in 10 mM Bis-tris at pH 6.5. The fractions containing BLG* as detected by SDS-PAGE were pooled and concentrated by ultrafiltration (Vivaspin) and loaded onto a Superdex G200 column and eluted with Ekmax buffer (50 mM Tris-HCl, 10 mM $CaCl_2$). Fractionation was followed by optical detection at 280 nm and 114 nm. Fractions containing protein were assayed by SDS-PAGE [Laemmli 1970]. One fraction was chosen for enterokinase digestion. To a 100 μL sample enterokinase (Invitrogen) was added (0.2 U, 0.2 μL) and the solution incubated at 25° C overnight. After incubation 10 μL of each sample was mixed with 1 μL of SDS sample buffer, and analysed by SDS-PAGE. The SDS-PAGE analysis showed that the thioredoxin had been cleaved from the BLG*, but that the BLG* had again been destroyed by proteolysis into three fragments, as observed with protein purification protocol (1) section 11.6 (Figure 11.9). The control sample of BLG* without enterokinase had not degraded.

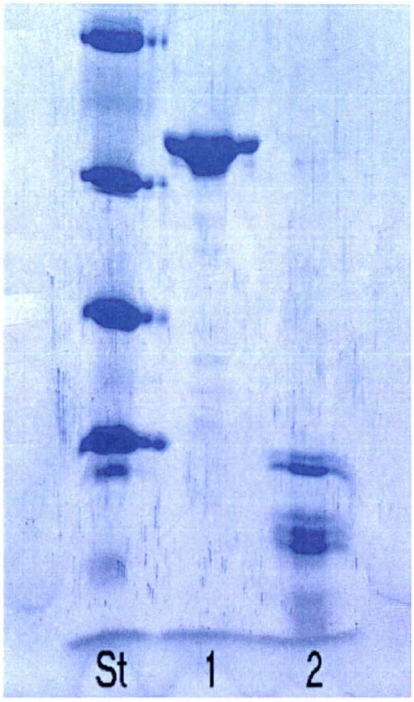


Figure 11.9

SDS-PAGE enterokinase digestion of K69E-fusion purified section 11.6.

Lane 1, control sample of BLG* without enterokinase. Lane 2, enterokinase digest.

The faint band just above the lane numbers is a small amount of the loading dye that persists after destaining.

A sample of BLGA isolated and purified from cow's milk was incubated for 8 hours with Enterokinase (Invitrogen) and showed exactly the same degradation pattern as the recombinant sample. The specificity of these batches of enterokinase is being looked into by the manufacturer.

11.8 Protein analysis

A sample of the BLG* was purified and denatured by reverse phase HPLC on a phonominex C-18 column. The protein was loaded in water and eluted with a linear gradient (60 min) of (water 0.1 % TFA) –

(acetonitrile 0.08 % TFA). The protein eluted from the column at 59.4% acetonitrile. The sample was collected and analysed by mass spectrometry on a Micromass ZMD electro-spray mass spectrometer. The analysis showed the calculated and experimental mass to be the same (Figure 11.10).

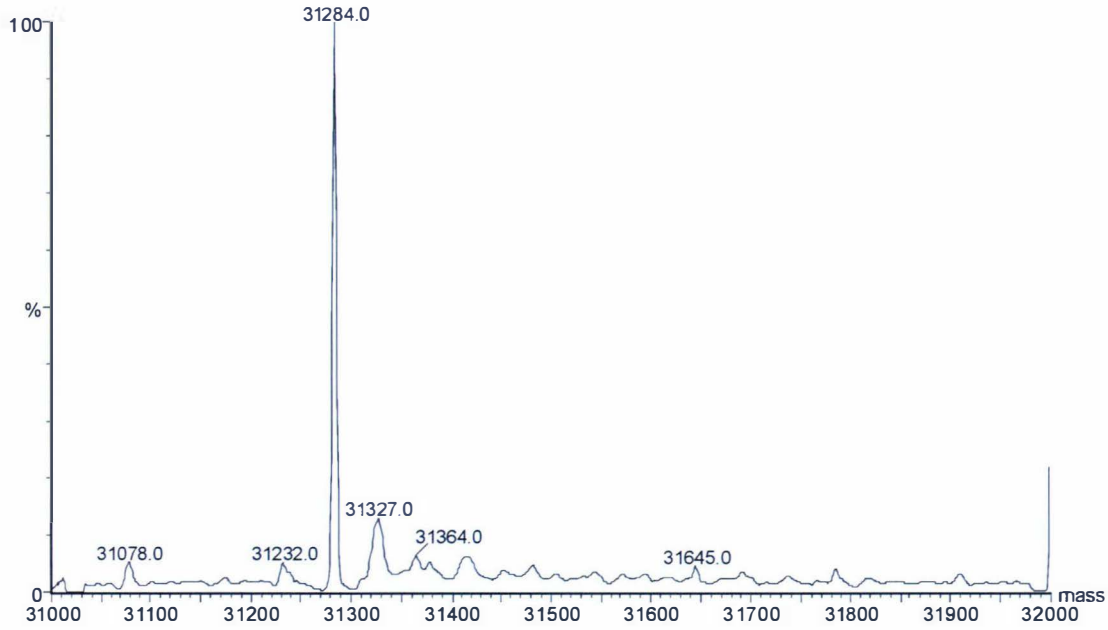


Figure 11.10

Mass spectrum of BLG*

A sample of BLG* was buffer exchanged (Vivaspin) into 10 mM phosphate buffer at pH 7.0 and concentrated to 1 mg mL^{-1} for circular dichroism (CD) spectroscopy. Circular dichroism spectroscopy was carried out on a Jasco J790 circular dichroism spectrometer. The CD spectrum showed the characteristic BLG peak at 287 nm. This peak is overlayed with the peaks from the thioredoxin fusion on the N-terminus of the BLG, but is nevertheless a strongly recognisable feature of the CD spectrum (Figure 11.11). The peak is generated by Trp 19, which is held in a very rigid conformation in the structure. The peak gives strong evidence that the protein is folded correctly.

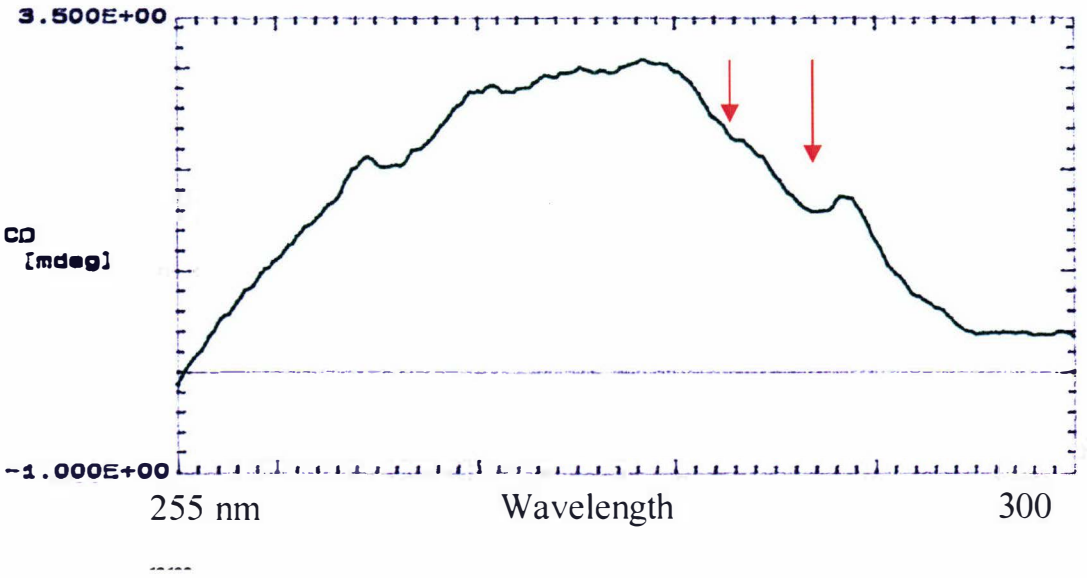


Figure 11.11

CD spectrum of the BLG-thioredoxin fusion protein. The arrows show the pair of peaks at 287 and 282 nm, which is characteristic of BLG in solution.

Chapter 12

Results and Discussion

12.1 Structure determinations of BLG

BLG, although easy to crystallize, has proved to be a challenge for structural determinations. The very first attempt at crystallization at low-salt concentrations by micro dialysis was successful, and data were collected, to a modest 3.6 Å resolution. These data, however, consisted of only half of the required data, and were inadvertently corrupted due to a number of technical failures. Repeat crystallizations were undertaken; these yielded not a single crystal. Neither did a number of repeat experiments using identical conditions (as far as was technically possible). After many attempts, the purity of the Milli-Q water supply to the laboratory came under suspicion. Conductivity tests confirmed the water did not contain ionic material. Mass spectra, however, revealed a considerable quantity of organic matter, with concentrations approaching that of the protein sample itself. To remove the organics, a system of column chromatography through activated carbon, followed by distillation from activated carbon (using a Vigreux column) was set up. This ultra purified water was analyzed by mass spectrometry and was found to be almost totally pure. The crystallizations were again trialed and crystals grew after one change of the well solution. Although crystallizations were now reproducible, the crystal form seemed to be somewhat random. This unpredictability is perhaps one of the reasons these low-salt crystal forms have not been successfully analyzed by crystallography previously.

The structure of BLGA co-crystallized with retinol or retinoic acid, has the retinoid species bound half way inside the calyx [Kontopidis et al. 2002]. Retinol has been shown to have a two-step binding to BLG [Fugate & Song 1980]. It was hoped that at low ionic strength the retinol would bind fully inside the calyx. Many attempts to co-crystallize BLGA with a variety of fatty-acid molecules, under either high or low salt conditions, have met with limited success. Either crystals did not grow or if crystals were obtained, after data were collected, no electron density for bound ligands was found inside the calyx of BLG. As no significant results can be drawn from these structures they were abandoned and have not been reported in this thesis.

All the protein used for BLG crystallography was prepared in the laboratory of Dr Lawrence Creamer, New Zealand Dairy Research Institute (now the Fonterra Research Centre). Dr Creamer is one of our collaborators in the BLG projects. Funding for this project comes as a FRST grant to the New Zealand Dairy Research Institute.

12.2 BLGA Y′ structure at 2.6 Å resolution

12.2.1 Description of the asymmetric unit and structure

The asymmetric unit contains two BLGA molecules related by 2-fold NCS. Both molecules have been truncated, with the removal of amino acid residues 61-65 (loop C-D) and 111-115 (loop G-H) and the C-termini 157-162, as these sections showed no reliable electron density as a result of disorder. This disorder will be discussed later. The overall structure of the BLGA monomer and dimer is essentially the same as the Y lattice $C222_1$ structure (PDB ID: 1qg5) [Oliveira *et al.* 2001]. A small angular difference in the dimer interface when the two structures are

compared is the most likely reason for the breakdown of the crystallographic 2-fold axis at the dimer interface, and therefore of the higher symmetry.

12.2.2 Quality of the model

The $2F_o$-F_c map, contoured at 1σ level, is well defined across the two subunits of the asymmetric unit, except about loops C-D, G-H and the C-terminus where the electron density abruptly becomes very ill defined. Final F_o-F_c maps showed few uninterpretable holes or peaks in the electron density. The two-fold NCS restraints yield small improvements in the electron density maps. The final values for R and R_{free}, 0.267 and 0.293 are somewhat higher than desirable, and are a consequence of the disorder and weak $h+k$ odd data of the primitive cell. The Ramachandran plot [Ramachandran & Sasisekharan 1968] showed 85.6 % of the residues in the most favoured regions, as defined by PROCHECK [CCP4 1994]. This compares most favourably with other BLG structures. Outlying residues in the Ramachandran plot 110 and 116 are the end residues where the break in the electron density occurs and therefore cannot be considered in the Ramachandran calculation. Residues 33 and 35 are part of the dimer interface, the Y lattice dimer interface about residue 33 is sterically crowded and, therefore, the residues adopt geometrically unfavourable conformations. Y99 is the γ-turn, which is a well-known structural feature of BLG and other lipocalins. Due to the considerable proportion of the molecule that is disordered and to the limited resolution of the data available, very few well-ordered water molecules were included in the structure.

12.3 Structure of BLGA in lattice U′ at 3.0 Å resolution

12.3.1 Description of the asymmetric unit and structure

The asymmetric unit contains four BLGA molecules, making two dimers. Four-fold NCS was applied across all four molecules. For refinement purposes the "C" and "D" molecules were chosen from different asymmetric units, so as to minimise the box that contained the asymmetric unit contents. This has the effect of speeding up the refinement calculations. For analysis and presentation, the dimers were reassembled with the unit cell and symmetry elements. The overall structure of the BLGA monomer and dimer is very similar to the X and Z lattice structures (PDB IDs: 1beb [Brownlow *et al.* 1997] and 3blg [Qin *et al.* 1998]).

The pH at which the crystal was grown is the lowest pH thus far for an X-ray structure. As the high concentration protein solution was dialysed against unbuffered ultra purified water, the pH in the dialysis button will have been approaching the pI of BLG at 5.2; this is due to the BLG buffering itself. The pH of the structure is the closest pH for an X-ray structure when compared to the pH conditions used for the NMR structural determinations. The structure has the E-F loop in the closed position, as expected at acid pH. The resolution of the flexible loops in the structure is most likely a consequence of inter molecular hydrophobic interactions, promoted by the fact that there are so few ions in solution for the protein side chains or backbone to interact with. Due to the limited resolution of the data available, only a few well-ordered water molecules were included in the structure, so as not to introduce errors into the model.

12.3.2 Quality of the model

The $2F_o$-F_c map, contoured at 1σ level, is well defined across all subunits of the asymmetric unit, including the often disordered flexible loop regions. Final F_o-F_c maps showed few uninterpretable holes or peaks in the electron density. The four-fold NCS restraints yield substantial improvements to electron density maps, effectively giving 3.0 Å resolution maps a 2.6 Å appearance. The final values for R and R_{free}, 0.293 and 0.313 are disappointingly high; this is most likely a consequence of weak data and the problems with the data collection. The Ramachandran plot [Ramachandran & Sasisekharan 1968] showed 78.2 % of the residues in the most favoured regions, as defined by PROCHECK [CCP4 1994]. This is somewhat lower than in other BLG structures. Other than the γ-turn at Tyr 99, only Ala 34 lies in an unfavourable geometry. Ala 34 makes up part of the dimer interface, and there are two hydrogen bonds that hold Ala 34 in an unfavourable geometry (A34_O(A)-A37_N(A) 2.80 Å, and A34_N(A)-D33_OD1(B) 3.21 Å).

12.4 Comparison of BLGA structures

The X-ray structures of the three most common BLG variants A, B and C, have been determined. Variant A has been characterised in three different crystal lattices. This dissertation presents one completely new lattice and a minor symmetry change in another lattice most likely a consequence of freezing the crystal. Variant A has also been subject to two independent NMR determinations [Uhrinova *et al.* 2000; Kuwata *et al.* 1999]. These independent structures cover an extremely wide range of experimental conditions (Table 12.1). X-ray structures of proteins have often been

criticised as being "unnatural", not relevant to the proteins *in vivo* environment, and therefore not physiologically relevant. BLG is in a relatively unique position of now having the structures of four different crystal lattices and two NMR experiments available to compare. We can compare these structures and look for differences that could be caused by the crystal packing or the environment in the crystal or NMR solution and postulate whether the structure is dependent on external factors, or is structurally consistent and independent of experimental conditions.

In the comparison we shall first consider the superposition of a monomer from one experiment onto a monomer from another. For this comparison the program TURBO was used for the fitting and examination of the results graphically. Pairs of molecules were superimposed by calculation of a reorientation matrix by least-squares fit of three Cα positions from identical residues in the different structures. Then this reorientation matrix was refined by including in the least-squares fit all pairs of Cα atoms lying within 0.5 Å of each other. On successive refinement of the reorientation matrix, more pairs of Cα atoms were included in the calculation. The iterations were continued until no more pairs of atoms met the 0.5 Å criterion for inclusion and the reorientation matrix was stable. The final number of pairs of Cα atoms included in the least-squares superposition was recorded (Table 12.1). As the distance cut-off was set to the same value for each superposition, all the RMSD values were approximately equal (0.32 Å).

Table 12.1. Pairwise structural alignment of BLG monomers. Figures indicate the number of Cα atoms from model "A" within 0.5 Å of the identical Cα atom in model "B" after the orientation matrix has stabilised.

	Y	Z	U′	Y′	Z_{BR}	Z_{PA}	B-Y	C-Y
X	86	104	112	71	97	87	98	89
Y	_	73	88	118	72	70	126	120
Z		_	89	65	156	110	81	78
U′			_	77	99	79	93	92
Y′				_	65	65	110	101
Z_{BR}					_	124	82	78
Z_{PA}						_	79	70
B-Y							_	126
C-Y								_

X) BLGA X lattice (PDB ID: 1beb) [Brownlow *et al.* 1997]

Y) BLGA Y lattice (PDB ID: 1qg5) [Oliveira *et al.* 2001]

Z) BLGA Z lattice pH 6.2 (PDB ID: 3blg) [Qin *et al.* 1998]

U′) BLGA U′ lattice

Y′) BLGA Y′ lattice

Z_{BR}) BLGA Z lattice with 12-bromododecanoic acid bound (PDB ID: 1bso) [Qin *et al.* 1998b]

Z_{PA}) BLGA Z lattice with palmitate bound (PDB ID: 1b0o) [Wu *et al.* 1999]

B-Y) BLGB Y lattice [G. Jameson *personal communication*]

C-Y) BLGC Y lattice [G. Jameson *personal communication*]

The above comparison shows that independent of experiment or even variant of BLG the structures superimpose very well, with at worst 50%

of the structure in good alignment. Average alignments of 90-120 (60-75 %) residues demonstrate that the core elements of BLG, the eight-strand β-barrel, the three-turn α-helix and external β-strand "I", are structurally independent of experimental conditions.

Comparison of the X-ray structures with the NMR structures shows a much lower number of residues in alignment. NMR structures consist of an ensemble structure comprising 20-30 of the lowest energy models consistent with the NOE NMR data. This makes comparison of models considerably more difficult. One qualitative method of comparing the structures is comparison of the Ramachandran plots, in particular two features. Firstly the γ-turn at Tyr 99 which is a conserved feature in all X-ray structures, and is conserved in many other lipocalin superfamily members, is missing in one of the two NMR models available (PDB ID: 1cj5) [Kuwata 1999]. In the other NMR structure (PDB ID: 1dv9) [Uhrinova *et al.* 2000], this feature was specifically built back into the model, after the original solution was refined. The other feature to note is the clustering of the residues involved in α-helical secondary structural elements. The X-ray structures generally have a significant cluster of residues fall in the lower-right quadrant of the most-favoured region of the α-helix region of the Ramachandran plot. The NMR structures conversely have a somewhat more random coverage of the α-helix region of the Ramachandran plot. As the Ramachandran plot represents graphically the main chain Psi (ψ) and Phi (φ) angles it can be seen that the NMR structures in general are less well ordered than the X-ray structures (Figures 12.1, 12.2 and 12.3).

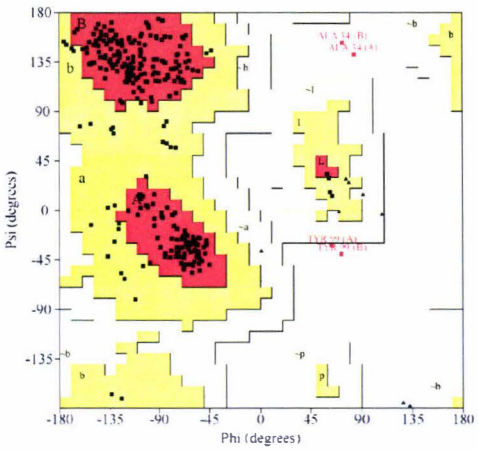


Figure 12.1

Ramachandran plot from the X-ray structure 1beb

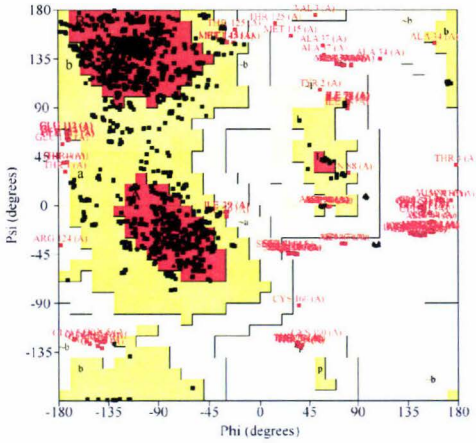


Figure 12.2

Ramachandran plot from the NMR structure 1cj5

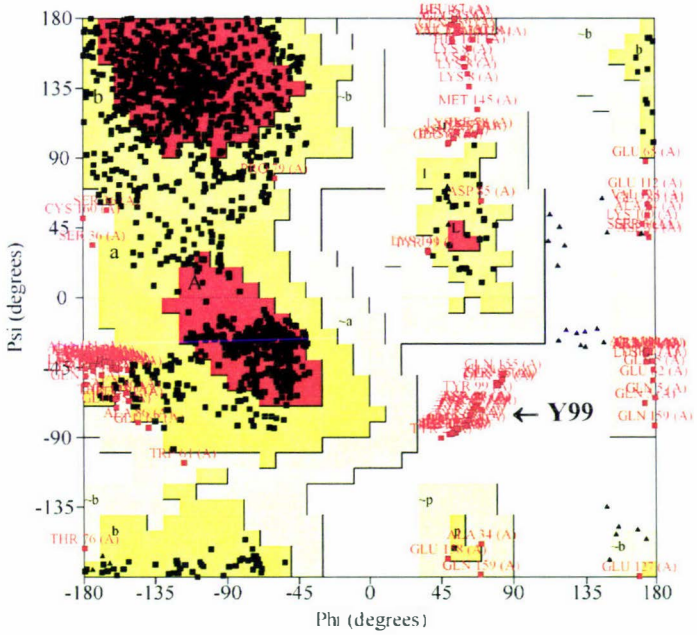


Figure 12.3

Ramachandran plot from the NMR structure 1dv9.
Note Tyr 99 falls into the forbidden region

At pH > 4 BLG is a dimer in solution and in the crystal. In lattices Y and Z the dimer two-fold axis falls on a crystallographic two-fold axis. In both cases the asymmetric unit contains only one monomer with the dimer completed by the symmetry. X lattice has no symmetry in the unit

cell and contains a dimer; U′ contains two independent dimers in the asymmetric unit; and Y′ has the same unit cell as Y but with the two-fold axis that related the monomers removed, and therefore has a dimer in the asymmetric unit. For comparison of the dimer structures, the "B" subinit of the dimer for the Y and Z lattices was generated from the unit cell dimensions and space group symmetry. The dimer structures were compared by least-squares superposition of the "A" subunits to each other as described above and then by looking at the differences in the position and orientation of the "B" subunits. The dimer interface consists of the I strand and a small interface at residues 33 and 34. The I strand interface consists of four hydrogen bonds, (H146_O(A)..S150_N(B), R148_N(A)..R148_O(B), R148_O(A).. R148_N(B), and S150_N(A)-H146_O(B)). In all structures these bonds are conserved (Figure 12.4). In isolation the interface seems invariant. However, when the dimer structures are aligned, the interface has a considerable degree of variability, achieved mainly through twist of the interface. The differences in dimer interface for molecules in lattices X, Z, U′ are quite small, with angular differences of 6-7° for molecules in the Z, and U′ lattices, with respect to the X lattice. With respect to X lattice, molecules in the Y lattice and Y′ lattice, on the other hand, show a considerable difference in the dimer interface compared to molecules in the X, Z, and U′ lattices, with an angular difference of -14° for molecules in the Y lattice and -11° for the Y′ lattice. There is, therefore, a difference of more than 20° in the orientation of the B subunit in lattice Z or U' compared to lattce3 Y. The variability of the dimer interface can be seen in the differing positions of the "B" molecule in Figure 12.5.

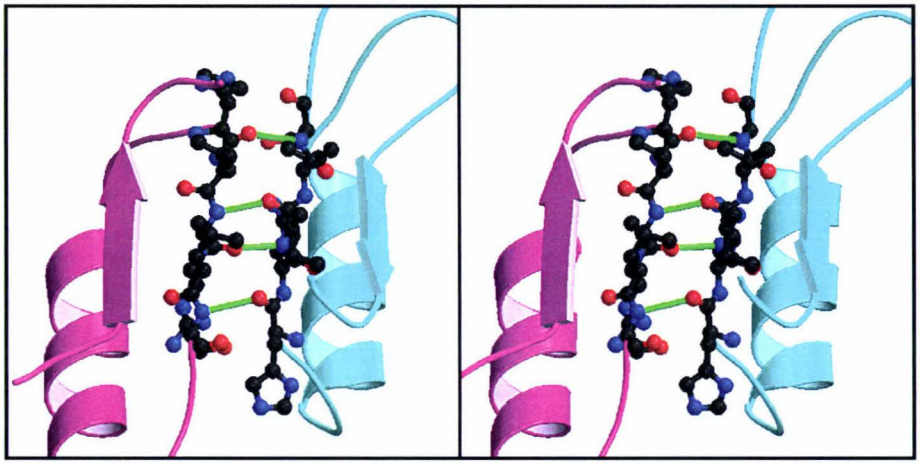


Figure 12.4

Stereo figure of the I strand dimer interface.
The four hydrogen bonds (green) are invariant in all BLGA structures.
The A strand and α helix 2 are shown in a cartoon representation.

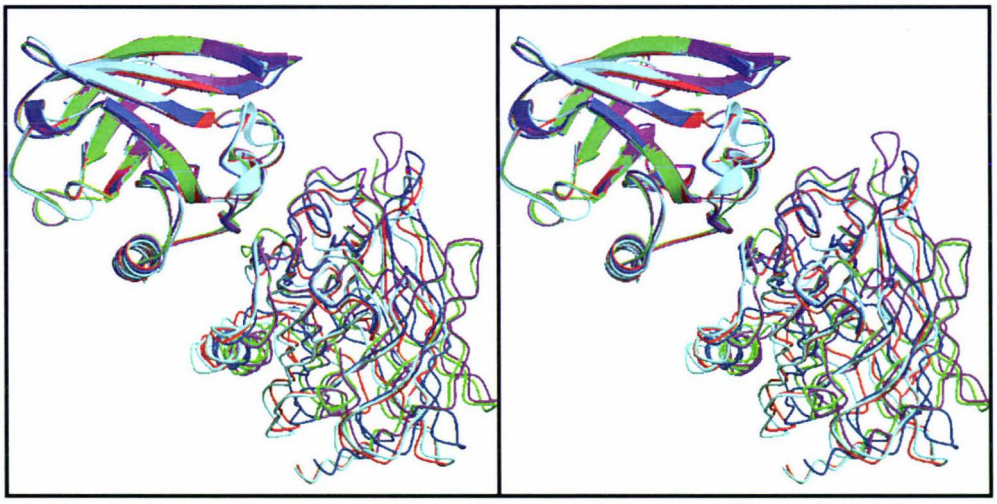


Figure 12.5

Stereo figure of the structural alignment of the dimers of BLGA in different lattices, based on the superposition of the A monomer (Ribbon). Note the considerable variability of the position of the B monomer (smoothed Cα trace), U′ lattice (red), Y′ lattice (green), X lattice (blue), Y lattice (purple) and
Z lattice (light blue).

The differences and similarities of both the monomeric and dimeric structures give us a clear indication of what effects experimental method and experimental conditions have on the structure of BLG. The structures of BLGA in the two new lattices presented here are similar to the previously published BLGA structures and can be considered alongside them as adding to the knowledge base of structural information on BLG. We can now be very confident in the structure of the core elements of the BLG. Where the structure of the conformationally labile elements in the structure can be defined, we can also have confidence that this structure represents a snapshot of a natural conformation. If all the structures are viewed together, then we can build a clear representative picture of BLG's structure in its *in vivo* environment. Specifically, BLG maintains a similar dimeric structure at both high and low salt conditions at pH > 4. Both secondary and tertiary core structure (the 8-stranded β barrel, three-turn α helix external to the barrel, and external β strand) are essentially invariant over the pH range 2.5 to 8.2, and are independent of monomeric or dimeric association of BLG subunits.

12.5 BLG functionality (stability) and structure

BLG is one of the most exhaustively studied proteins known. Yet some of the most basic questions of functionality remain speculative or unanswered. The molecular structure of BLG has played a role in the functional studies. The structure has revealed a very rigid eight-stranded β-barrel that supports a large hydrophobic pocket (calyx). Figure 12.6 gives a representation of the secondary structure and hydrogen bonding of BLGA [Qin *et al.* 1998]. One end of this pocket is mostly occluded with a very tightly controlled structure comprising loops B-C, D-E and F-G. These short loops have a very well defined structure in both X-ray and

NMR structure determinations. At the other end of the pocket four, long, poorly defined loop structures serve several functions (Figure 12.7). Firstly they allow the opening and closing of the pocket through movements of the E-F loop, which are pH dependent. This pH-dependent conformational change is referred to as the Tanford transition [Tanford *et al.* 1959, Qin *et al.* 1998] and has been well studied by X-ray crystallography and a variety of spectroscopic and titrimetric techniques. At low pH the E-F loop is closed and the calyx is apparently empty, and at pH above 6.8 the E-F loop is open and the cavity fills with water.

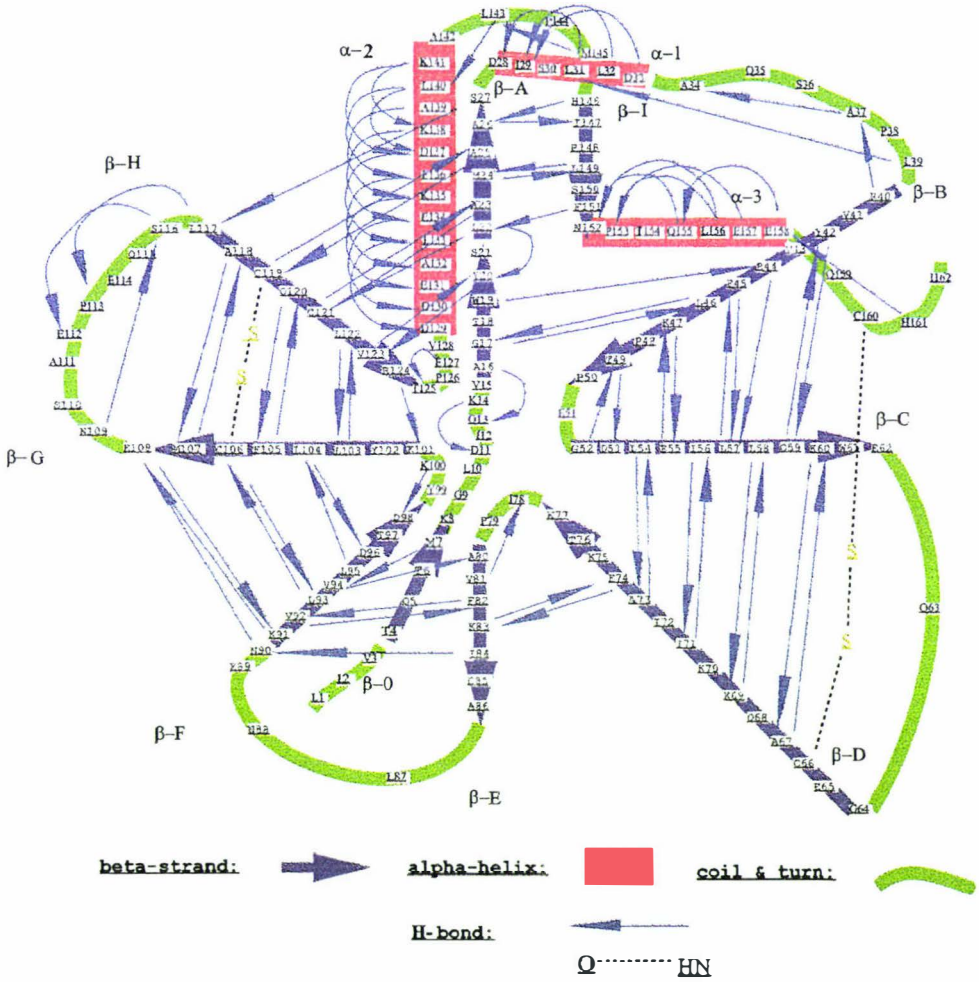


Figure 12.6

Secondary structure and hydrogen bonding in BLGA [Qin *et al.* 1998].

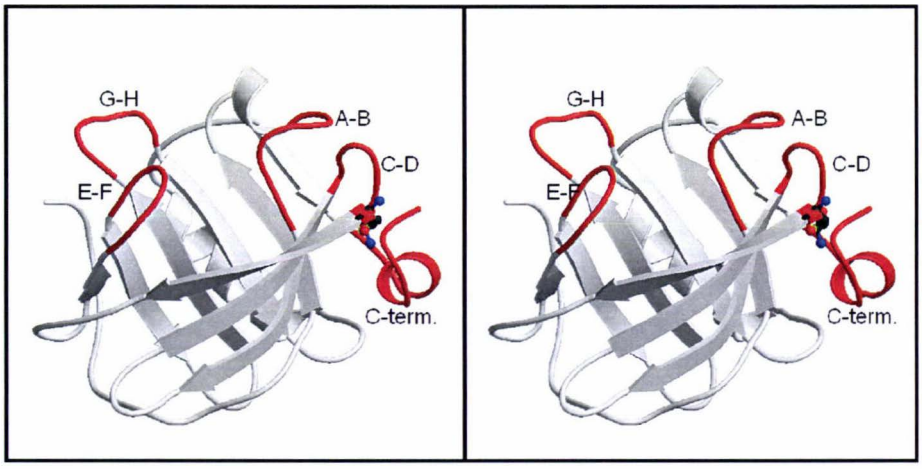


Figure 12.7

Stereo figure showing the conformationally labile loops (red) in the BLG structure.

The other loops at this end of the pocket, loops A-B, C-D, G-H and the C-terminus show remarkable flexibility, their electron density in all but a few X-ray structures is very weak, and few NOE contacts can be found in the NMR structures. These approximately 60 amino acid residues have a conformationally labile structure, with each residue able to reside in multiple conformations and positions in 3D space. As discussed above (Section 12.4 Comparison of BLGA structures) the multitude of structures of BLGA with wildly different experimental conditions portrays a set of snapshots of BLGA's structure. From this information some of the functional/structural characteristics of BLG can be investigated.

To expose hydrophobic side chain moieties (Leu, Ile, Val, Phe, Ala, Met and Pro) to the bulk solvent has associated with it an energy cost brought about by the surface tension between the two phases [Jameson *et al.* 2002]. The surface area of the internal hydrophobic pocket can be calculated by calculating the surface area of the liganded and unliganded BLG Z lattice structure [Qin *et al.* 1998], taking the difference and

compensating for the surface area of the head group of the ligand 12-bromododecanoic acid. Using the program GRASP [Nicholls *et al.* 1991] unliganded BLGA has a surface area of 7200Å^2, liganded BLGA has a surface area of 6650 Å^2 and the head group ~50 Å^2, giving a surface area for the pocket of ~600 Å^2. From the difference in surface area of BLGA at pH 6.2 (E-F loop closed conformation) and pH 8.2 (E-F loop open conformation), the volume of the pocket can also be calculated (by GRASP) and is approximately 1000 Å^3. The free energy cost to the molecule associated with the pocket can be described by Helmholtz free-energy $\Delta A=(\Delta\gamma)(\Delta\text{Area})$. $\Delta\gamma$ is the difference in surface tension between water (72.8 mN m^{-1} with respect to air) and the hydrophobic interior (the surface tension of the hydrophobic surface was approximated to that of benzene 28.9 mN m^{-1} with respect to air); therefore, $\Delta\gamma$ = 44 mN m^{-1}. ΔArea = 600 Å^2 as described above. Therefore $\Delta A \approx 160$ kJ mol^{-1} of destabilization energy [Jameson *et al.* 2002]. To counter this destabilization, stabilization energy must be found. Entropic energy associated with multiple conformations of the flexible loops can counter the destabilization of exposing a hydrophobic surface to the bulk solvent. If each residue has two conformations (over simplified, as it treats each residue as a rigid body ignoring side-chain rotamers, and many residues may have more than two stable conformations), entropy may be calculated as $S = NR\ln 2$ and (R = 8.314 J K^{-1}mol^{-1}, N is the number of disordered residues). There are ~50 residues that typically show high thermal motion in X-ray structures and conformational disorder in NMR structures, leading to S ≈ 290 J K^{-1}mol^{-1}. In terms of free energy, this corresponds to ~90 kJ mol^{-1} at 20°C (298 K) [Jameson *et al.* 2002]. The flexible loops A-B, C-D, E-F and G-H, therefore, contribute substantial stabilization energy to the very rigid β-barrel structure, and allow the maintenance of a large solvent-exposed hydrophobic pocket (see Figure

7.2). The importance of conformational entropy has been discussed by Stone [Stone 2001]. Examples include, extended poly-glycine loops that enhanced the stability of helix bundles [Simons *et al.* 2001] and increased back bone rigidity on ligand binding [Zidek *et al.* 1999].

12.6 Industrial importance of BLG

Most milk products are subject to some form of heat treatment during the manufacturing process, the milk may also be subjected to elevated pressures in pumps and pipes as it is processed. Heat and pressure are known to cause subtle to gross changes in the structure of proteins and other components of milk. Some of these changes are beneficial as they develop the flavour or texture of the milk product; however, many of these changes are detrimental, such as fouling of processing equipment, gelling during processing, or not gelling when it is desired. Study of the changes in the structure of those components will provide the dairy industry with fundamental knowledge, which will aid in the design of processes and process equipment.

BLG has been identified as one of the major contributing components in the fouling of milk processing plants. It forms insoluble aggregate gels with β-casein on heating surfaces in heat exchangers [Murray *et al.* 2000]. The cleaning and removal of these gels are a significant cost to the dairy industry. Ongoing study of milk components will offer remedies to these and other processing difficulties in the milk industry [Sawyer *et al.* 2002].

12.7 BLG folding and unfolding

The mechanism of folding of BLG either *in vivo* or *in vitro* is difficult to study. Much effort has been put into study of the unfolding of BLG; these studies include chemical, pressure and thermal denaturation of BLG. Techniques such as CD [Sawyer *et al.* 1971; Lapanje & Poklar 1989; Manderson *et al.* 1999a], UV-VIS and IR spectroscopy [Boye *et al.* 1996; Casal *et al.* 1988], fluorescence [Mills 1976; Manderson *et al.* 1999b], PAGE [Sawyer 1968; Manderson *et al.* 1998], and 1D-NMR [Griko *et al.* 1994; Li *et al.* 1994] have been used; these techniques follow bulk events during the unfolding. Developments in NMR have opened the door to 2D-NMR and higher-order NMR techniques for studying proteins; these techniques can be used to follow individual residues during the unfolding events. The three most common variants of BLG (A, B and C) have substantially different thermal denaturation curves (Figure 12.8) [Manderson *et al.* 1999a]. A recent study [Edwards *et al.* 2002] has provided an unfolding mechanism for BLGA. BLGA unfolds in a sequential manner over a temperature range of 55°C to 80°C. First the D-E interface is disrupted, followed by the C-D interface. Helix α2 which melts at 60°C, then the A-B, E-F and A-I interfaces disappear at 65°C. At 70°C the A-H interface is broken, and lastly the F-G and G-H strands disassociate at 75-80°C. BLGB on the other hand shows a much more cooperative unfolding mechanism as demonstrated by the much steeper denaturation curve (Figure 12.8). BLGC shows a cooperative denaturation curve, with cooperativity similar to that observed for BLGB, but temperature at which 50% denaturation has occurred is higher temperature than that observed for variants A and B.

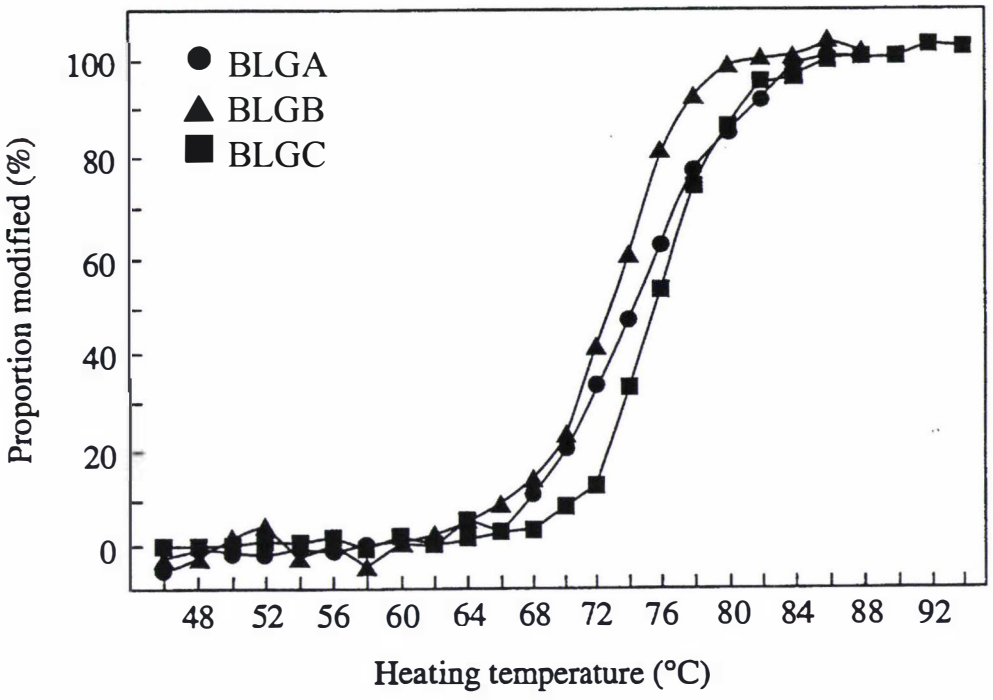


Figure 12.8
Thermal denaturation curves of the common BLG variants [Manderson *et al.* 1999a].

CD and NMR spectra of BLG that has been treated with per-fluorinated alcohols clearly show that the protein has refolded into a predominantly α-helix structure [Gast *et al.* 2001]. This α-helical structure may be a folding intermediate structure, and this structure may be able to be stabilized with mutation of the polypeptide sequence by recombinant synthesis, which is discussed later (12.8 Recombinant protein studies).

X-ray crystallographic studies of the unfolding intermediates or fully denatured protein is not practical, as large quantities of homogeneous protein is required. However, NMR and other spectroscopic techniques where the protein is denatured *in situ* have given numerous structural insights into the folding and unfolding of BLG. For example CD spectroscopy has shown the cooperativity of BLG unfolding [Manderson

et al. 1999a].

12.8 Recombinant protein studies

Section 7.12 (Recombinant bovine β-lactoglobulin) chronicles the numerous failed attempts to produce BLG from recombinant genes in *E. coli*. Our lab has taken on a somewhat ambitious project to synthesize BLG in *E. coli* by tagging the protein with a thioredoxin N-terminal fusion. We undertook the project to overcome some of the issues with production of recombinant proteins in *Pichia pastoris* [Denton *et al.* 1998]. This expression system developed in Scotland by our collaborators Lindsay Sawyer *et al.* has one major draw back: *Pichia pastoris* has a tendency to incorporate sugars into the protein, although this can be overcome to some extent by starving the *Pichia pastoris* [L. Sawyer and C. Holt, *Personal communication*]. In any event the homogeneity of the recombinant protein is compromised. The protein samples show a marked inhomogeneity [P. Barlow, *Personal communication*]. *E. coli* does not incorporate sugars into proteins as they are synthesised, so therefore we have expended considerable effort to optimise an *E. coli* expression system [Ariyaratne *et al* 2002].

Thus far, studies have been confined to purified BLG from genetically homozygous cows expressing either A, B or C variants of BLG. Recombinant protein production opens the prospects of studying the functionality, reactivity, quaternary structure, and folding and unfolding of BLG by mutation of the protein sequence to engineer in or out particular properties.

BLGA has now been successfully synthesised recombinantly in *E. coli* with methods developed in our laboratory by Dr K A N S Ariyaratne [Ariyaratne *et al.* 2002]. This methodology has been used for mutational studies by Dr Ariyaratne, myself and several others in our laboratory. The following paragraphs chronicle, firstly, my synthesis of BLGA mutant K69E and, then, discussion of other mutational studies ongoing in our laboratory and their functional, structural significance or motivation. Lastly I look at mutational studies that may propose answers to some of the key questions outstanding with respect to BLG.

Functional mutants, K60E, K69E and E62L and others may radically perturb the binding of hydrophobic molecules. It is anticipated that K60E and K69E will lower the binding constant of ligands with hydrophilic end groups and render the protein inactive with respect to fatty acid binding. The E62L may have the opposite effect and raise the binding constant considerably for ligands with non-hydrophilic end groups. The K69E mutant has been synthesised and, once cleaved, is ready for crystallographic and functional studies. As discussed in Chapter 11, the processing of the fusion protein to produce the recombinant BLG was not proceeding well. Incorporating a TEV protease site in the BLG gene before insertion into the PtrxFus plasmid is under consideration, BLG was not degraded by TEV protease.

The well-studied genetic variants of BLG show remarkably different physical properties with respect to denaturation, despite the point mutations between variants being reasonably conservative. Their three dimensional structures are extremely similar, but do offer some explanation of the different properties [Qin *et al.* 1999]. Mutants D64G and V118A (KANS Ariyaratne) make up the "half A, half B" mutant set. These are being studied to gauge each mutation's contribution to the

unfolding temperature and mechanism difference between the A and B variants.

R148P (KANS Ariyaratne) is an interfacial mutant, designed to produce monomeric BLG at pH 7.0. Monomeric BLG will be invaluable for both crystallographic and NMR study of folding, refolding and ligand binding. C121S (KANS Ariyaratne) is another mutant that will aid in the unfolding and aggregation studies, as it will no longer have a free thiol and will not be able to aggregate by disulphide cross-linkage.

Secondary-structure prediction algorithms show that β-strand A has a high propensity to form an α-helix structure as shown in Figure 12.9. Point mutations of key hydrogen bonding residues in β-strand A may allow the isolation of a stable α-helical folding intermediate structure. This will aid in the elucidation of the folding/unfolding mechanism.

Other structural mutations that could provide details about the folding and unfolding of BLG are mutations that change the length of the A-B, E-F and G-H loops. These loops as discussed above (Section 12.5 BLG functionality and structure) contribute entropic stabilization to the protein fold, which stabilises the solvent exposed surface of the calyx. By either shortening or extending these loops the calyx may be opened or closed and the functionality and structure possibly changed considerably.

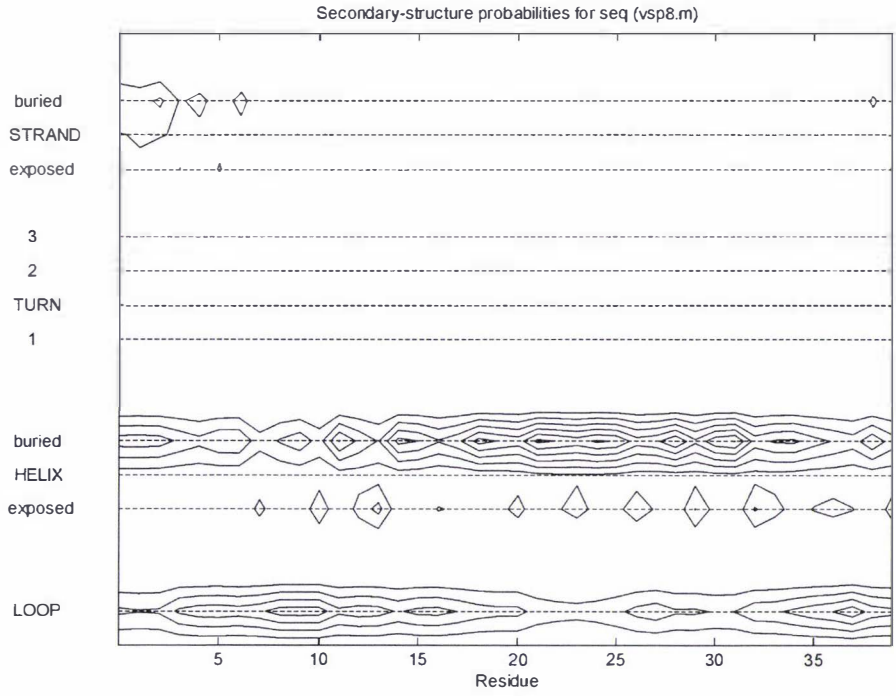


Figure 12.9

Secondary structure prediction of the first 40 residues of BLGA (http://bmerc-www.bu.edu/psa/request.htm, using the standard weighting scheme [Stultz *et al.* 1993]). The probability of a residue having a particular type of secondary structure is indicated by the height of the contours. Note that from residues 12 to 33, a portion of the structure observed to be a β strand flanked by α-helical turns, there are very high contours for the "buried α-helix" type secondary structure.

The *in vivo* physiological role of BLG has been studied for sixty years or more, yet the function in both lactating cow and calf remains as much a mystery as ever. Molecular biology may now be able to answer this question with a "knock out" BLG cow that is unable to express BLG. This may finally answer the question of why some mammalian species express BLG and others do not.

12.9 Concluding remarks

Two new structures of BLG have been completed and have provided new structural information about β-lactoglobulin. The comparison of the structures of BLG at high and low ionic strength have provided structural validation to the already considerable body of structural work in the literature. Specifically, the dimer interface observed under high-salt conditions is preserved under low-salt conditions. Moreover, under low-salt conditions, the large hydrophobic pocket seen in structures at high-salt conditions is also preserved. The EF loop, whose conformation is sensitive to pH, is found in the "closed" position in the low-salt pH 5.5 structure, similar to its conformation in the high-salt structure at pH 6.2 and to the relatively low-salt NMR structures at pH ~2.6. The attempts to co-crystallize BLG with a number of small hydrophobic molecules at low ionic strength may have been successful; however, the data set from the crystal that has given indications of containing a bound ligand refuses to solve and further work will be required. Recombinant BLG K69E has been synthesized and will be used for further work on functional aspects of BLG.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403.

Ariyaratne, K. A. N. S., Brown, R., Dasgupta, A., de Jonge, J., Jameson, G. B., Loo, T. S., Weinberg, C., & Norris, G. E. (2002). *Int'l. Dairy J.* **12**, 311.

Aschaffenburg, R., Green, D. W., & Simons, R. M. (1965). *J. Mol. Biol.* **13**, 194.

Askonas, B. A. (1954). *Biochem. J.* **58**, 332.

Batt, C. A., Rabson, L. D., Wong, D. W. S., & Kinsella, J. E. (1990). *Agric. Biol. Chem.* **54**, 949.

Bell, K. (1962). *Nature* **195**, 705.

Bell, K. & McKenzie, H. A. (1964). *Nature* **204**, 1275.

Bewley, M. C., Qin, B. Y., Jameson, G. B., Sawyer, L. K., & Baker, E. N. (1997). *Int'l. Dairy Fed. Bull.* **Special Issue. "Milk Protein Polymorphism"**. 100.

Blomhoff, R., Green, M. H., Berg, T., & Norum, K. R. (1990). *Science* **250**, 399.

Boye, J. I., Ismail, A. A., & Alli, I. (1996). *J. Dairy Res.* **63**, 97.

Brignon, G., Chtourou, A., & Ribadeau-Dumas, B. (1985). *J. Dairy Res.* **52**, 249.

Brownlow, S., Cabral, J. H. M., Cooper, R., Flower, D. R., Yewdall, S. J., Polikarpov, I., North, A. C. T., & Sawyer, L. (1997). *Structure* **5**, 481.

Brünger, A., Clore, M., Gros, P., Nalges, M., & Read, R. (1998). *Acta Cryst.* **D54**, 905.

Bull, H.B. & Currie, B.T. (1946). *J. Am. Chem. Soc.* **68**, 45.

Cannan, R. K., Palmer, A. H., & Kibrick, A. C. (1942). *J. Biol. Chem.* **142**, 803.

Casal, H. L., Kohler, U., & Mantsch, H. H. (1988). *Biochim. Biophys. Acta* **957**, 11.

The CCP4 Suite: Programs for Protein Crystallography. (1994). *Acta Cryst.* **D50**, 760.

Cho, Y., Gu, W., Watkins, S., Lee, S. P., Kim, T. R., Brady, J. W., & Batt, C. A. (1994). *Protein Eng.* **7**, 263.

Chatel, J.-M., Adel-Patient, K., Creminon, C., & Wal, J.-M. (1999). *Protein Expression Purif.* **16**, 70.

Conti, A., & Godovac-Zimmermann, J. (1990). *Biol. Chem. Hoppe Seyler,* **371**, 261.

Crowfoot, D. & Riley, D. (1938). *Nature* **141**, 521.

Denton, H., Kim, T-R., Goto, Y., Hirota, N., Kuwata, K., Wu, S-Y., Sawyer, L., & Batt, C. A. (1997). *Protein Eng.* **10**, 1339.

Dufour, E. & Haertle, T. (1990). *Protein Eng.* **4**, 185.

Edwards, P. J. B., Jameson, G. B., Palmano, K. P., & Creamer, L. K. (2002). *Int'l. Dairy J.* **12**, 331.

Farrell, H. M., Bede, M. J., & Enyeart, J. A. (1987). *J. Dairy Sci.* **70**, 252.

Fugate, R. D. & Song, P. S. (1980). *Biochim. Biophys. Acta* **625**, 28.

Fukushima, Y., Kawata, Y., Onda, T., & Kitagawa, M. (1997). *J. Nutr. Sci. Vitaminol.* **43**, 673.

Futterman, S. & Heller, J. (1972). *J. Biol. Chem.* **247**, 5168.

Gast, K., Siemer, A., Zirwer, D., & Damaschun, G. (2001). *Eur. Biophys J.* **30**, 273.

Godovac-Zimmermann, J., Conti, A., James, L., & Napolitano, L. (1988). *Biol. Chem. Hoppe-Seyler* **369**, 171.

Grandori, R., Struck, K., Giovanielli, K., & Carey, J. (1997). *Protein Eng.* **10**, 1099.

Green, D.W., Aschaffenburg, R., Camerman, A., Coppola, J. C., Dunnil, P., Simmons, R. M., Komorowski, E. S., Sawyer, L., Turne, E. M. C., & Woods, K. F. (1979). *J. Mol. Biol.* **131**, 375.

Griko, Y. V. & Kutyshenko, V. P. (1994). *Biophys. J.* **67**, 356.

Grosclaude, F., Pujolle, J., Garnier, J., & Ribadeau-Dumas, B. (1966). *Ann. Biol. Anim. Biochim. Biophys.* **6**, 215.

Halliday, J. A., Bell, K., & Shaw, D. C. (1991). *Biochim. Biophys. Acta* **1077**, 25.

Hemley, R., Kohler, B. E., & Siviski, P. (1979). *Biophys. J.* **28**, 447.

Hyttinen, J. M., Korhonen, V. P., Hiltunen, M. O., Myohanen, S. & Janne, J. (1998). *J. Biotech.* **61**, 191.

Kleywegt, G. J. & Jones, T. A. (1999). *Acta. Cryst.* **D55**, 941.

Kleywegt, G. J. (1996). *Acta. Cryst.* **D52**, 842.

Kleywegt, G. J. (1999). *J. Mol. Biol.* **273**, 371.

Kontopidis, G., Holt, C., & Sawyer, L. (2002). *J. Mol. Biol.* **318**, 1043.

Kunkel, T. A. (1985). *Proc. Natl. Acad. Sci. U. S. A.* **82**, 488.

Kuwata, K., Hoshino, M., Forge, V., Era, S., Batt, C. A., & Goto, Y. (1999). *Prot. Sci.* **8**, 2541.

Jameson, G. B., Adams, J. J., & Creamer, L. K. (2002). *Int'l. Dairy J.* **12**, 319.

Laemmli, U. K. (1970). *Nature* **227**, 680.

Lapanje, S. & Poklar, N. (1989). *Biophys. Chem.* **34**, 155.

Li, C-H. (1946). *J. Am. Chem. Soc.* **68**, 2746.

Li, H., Hardin, C. C., & Foegeding, E. A. (1994). *J. Agr. Food Chem.* **42**, 2411.

Liberatori, J., Guetti, L. N., & Conti, A. (1979). *Boll. Soc. It. Biol. Spar.* **55**, 822.

Liberatori J., Morisio G. L., Conti A., & Napolitano L. (1979). *Boll. Soc. It. Biol. Spar.* **55**, 1369

Lovrien, R. & Anderson, W. F. (1969). *Arch. Biochem. Biophys.* **131**, 139.

Lyster, R. L. J., Jenness, R., Phillips, N. I., & Slone, R. E. (1966). *Biochem Physiol.* **17**, 967.

Malpeli, G., Stoppini, M., Zapponi, M. C., Folli, C., & Berni, R. (1995). *Eur. J. Biochem.* **229,** 486.

Maniatis, T., Sambrook, J., & Fritsch, E. J. (1989). *Molecular cloning, a laboratory manual, Second edition*, Cold Spring Harbour Laboratory Press.

Manderson, G. A., Creamer, L. K., & Hardman, M. J. (1999a). *J. Agri. Food Chem.* **47**, 4557.

Manderson, G. A., Creamer, L. K., & Hardman, M. J. (1999b). *J. Agri. Food Chem.* **47**, 3617.

Manderson, G. A., Creamer, L. K., & Hardman, M. J. (1998). *J. Agri. Food Chem.* **46**, 5052.

McKenzie, H. A. & Muller, V. J. (1983). *Comp. Biochem. Physiol.* **B74**, 259.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277B**, 505.

Mills, O. E. (1976). *Biochemistry*, **11**, 150.

Mizumachi, K., Kurisaki, J. & Tsuji, N. M. (1993). *Anim. Cell Technol.: Basic Appl. Aspects, Proc. Int. Meet. Jpn. Assoc. Anim. Cell Technol., 5th, MeetingDate 1992,* 115.

Monaco, H. L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L., & Eliopoulos, E. E. (1987). *J. Mol. Biol.* **197**, 695.

Murakami, K., Lagarde, M., & Yuki, Y. (1998). *Electrophoresis* **19**, 2521.

Murray, B. S. & Deshaires, C. (2000). *J. Colloid Interface Sci.* **227**, 32.

Navaza, J. (1994). *Acta Cryst.* **A50**, 157.

Nicholls, A., Sharp, K., & Honig B. (1991). *PROTEINS: Struct. Funct. Gen.* **11(4)**, 281

Oliveira, K. M. G., Valente-Mesquita, V. L., Botelho, M. M., Sawyer, L., Ferreira, S. T., & Polikarpov, I. (2001). *Eur. J. Biochem.* **268**, 477.

O'Neill, T. E. & Kinsella, J. E. (1987). *J. Agric. Food. Chem.* **35**, 770.

Otwinoski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307-326.

Palmer, A. H. (1934). *J. Biol. Chem.* **104**, 359.

Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findley, J. B. C., Sivaprasadarao, R., Jones, T. A., Nowcomer, M. E., & Kraulis, P. J. (1986). *Nature* **324**, 383.

Pearson, W. R. (1990). *Methods Enzymol.* **183**, 63.

Pedersen, K. O. (1936). *Biochem. J.* **30**, 961.

Pellegrini, A., Dettling, C., Thomas, U., & Hunziker, P. (2001). *Biochim. Biophys. Acta* **1526**, 131.

Perez, M. D. & Calvo, M. (1995). *J. Dairy. Sci.* **78**, 978.

Perez, M. D., Sanchez, L., Aranda, P., Ena, J. M., Oria, R.. & Calvo, M.

(1992). *Biochim. Biophys. Acta* **1123**, 151.

Pervaiz, S. & Brew, K. (1986). *Arch. Biochem. Biophys.* **246**, 846.

Pervaiz, S. & Brew, K. (1987). *FASEB. J.* **1**, 209.

Qin, B. Y., Bewley, M. C., Creamer, L. K., Baker, E. N., & Jameson, G. B. (1999). *Prot. Sci.* **8**, 75.

Qin, B. Y., Bewley, M. C., Creamer, L. K., Baker, H. M., Baker, E. N., & Jameson, G. B. (1998). *Biochemistry* **37**, 14014.

Qin, B. Y., Creamer, L. K., Baker, E. N., Jameson, G. B. (1998b). *FEBS Lett.* **438**, 272.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Prot. Chem.* **23**, 283-437.

Restani, P., Gaiaschi, A., Plebani, A., Beretta, B., Velona, T., Cavagni, G., Poiesi, C., Ugazio, A. G., & Galli, C. L. (2000). *Ann. Allergy. Asthma Immunol.* **84**, 353.

Robillard, K. A. & Wishnia, A. (1972a). *Biochemistry* **11**, 3835.

Robillard, K. A. & Wishnia, A. (1972b). *Biochemistry* **11**, 3841.

Roussel, A. & Cambillau, C. (1989) *Silicon Graphics Geometry partners Directory,* SGI, U.S.A, pp 72-78.

Sawyer, L., Barlow, P. N., Boland, M. J., Creamer, L. K., Denton, H., Edwards, P. J. B., Holt, C., Jameson, G. B., Kontopidis, G., Norris, G. E., Uhrinova, S., & Wu, S-Y. (2002). *Int'l Dairy J.* **12**, 299.

Sawyer, W. H., Norton, R. S., Nichol, L. W., & McKenzie, G. H. (1971). *Biochim. Biophys. Acta* **243**, 19.

Sawyer, W. H. (1968). *J. Dairy Sci.* **51**, 323.

Sheubert, D., LaCorbiere, M., & Esch, F. (1986). *J. Cell Biol.* **102**, 2295.

Simons, B., Scholl, D., Cyr, T., & Hefford, M. A. (2001). *Prot. Pept. Lett.* **8**, 89.

Stultz, C. M., White, J. V., & Smith, T. F. (1993). *Prot. Sci.* **2**, 305.

Sjogren, B. & Svedberg, T. (1930). *J. Am. Chem. Soc.* **52**, 3650.

Spector, A. A. & Fletcher, J.E. (1970). *Lipids* **5**, 403.

Tanford, C., Bunville, L. G., & Nozaki, Y. (1959). *J. Am. Chem. Soc.* **81**, 4032.

Totsuka, M., Katakura, Y., Shimizu, M., Kumagai, I., Miura, K., & Kaminogawa, S. (1990). *Agric. Biol. Chem.* **54**, 3111.

Uhrinova, S., Smith, M., Jameson, G. B., Uhrin, D., Sawyer, L., & Barlow, P. N. (2000). *Biochemistry* **38**, 3565.

Venter, C. J. *et al.* (2001). *Science* **291**, 1304.

Winter, N. S., Bratt, J. M., & Banaszak, L. J. (1993). *J. Mol. Biol.* **230**, 1247.

Wishnia, A. & Pinder Jr., T. W. (1966). *Biochemistry* **5**, 1534.

Wu, S. Y., Perez, M. D., Puyol, P., & Sawyer, L. (1999). *J. Biol. Chem.* **274**, 170.

Zidek, l., Novotny, M. V., & Stone, M. J. (1999). *Nature: Struct. Biol.* **6**, 1118.

Appendix A

BLAST search based on sequence identity, followed by a CLUSTAL-W multiple alignment [www.expasy.ch], using the default alignment parameters. Edited to take into account the available structures of superoxide dismutase.

```
CLUSTAL W (1.81) multiple sequence alignment

SODF_BACFR      ------------------------------------------------------------
SODF_PORGI      ------------------------------------------------------------
AAK88724        ------------------------------------------------------------MM 2
SODG_PSEPU      ------------------------------------------------------------
Q9AIX5          ------------------------------------------------------------
Q9WWG8          ------------------------------------------------------------
SODF_PSEAE      ------------------------------------------------------------
Q9ZGN1          ------------------------------------------------------------
BAB35788        ------------------------------------------------------------
AAG56645        ------------------------------------------------------------
SODF_ECOLI      ------------------------------------------------------------
Q9RQQ7          ------------------------------------------------------------
Q9RCF0          ------------------------------------------------------------
Q9KQF3          ------------------------------------------------------------
SODF_PHOLE      ------------------------------------------------------------
Q9JUW9          ------------------------------------------------------------
Q9JZV6          ------------------------------------------------------------
Q9F4F5          ------------------------------------------------------------
SODF_BORPE      ------------------------------------------------------------
SODF_LEGPN      ------------------------------------------------------------
SODF_COXBU      ------------------------------------------------------------
Q9AXR7          ------------------------------------------------------------
O15904          ------------------------------------------------------------
SODF_BABBO      ------------------------------------------------------------
Q27740          ------------------------------------------------------------
O77071          ------------------------------------------------------------
Q27791          ------------------------------------------------------------
O02616          ------------------------------------------------------------
AAK52814        ------------------------------------------------------------
O15640          ------------------------------------------------------------
AAL03316        ------------------------------------------------------------
SODF_RICPR      ------------------------------------------------------------
Q9RM31          ------------------------------------------------------------
SODF_HELPY      ------------------------------------------------------------
Q9R2E7          ------------------------------------------------------------
Q9S6R8          ------------------------------------------------------------
SODF_HELPJ      ------------------------------------------------------------
Q9R2E8          ------------------------------------------------------------
Q9R2E6          ------------------------------------------------------------
SODF_CAMCO      ------------------------------------------------------------
SODF_CAMJE      ------------------------------------------------------------
SODF_ENTHI      ------------------------------------------------------------
Q9A2K4          ------------------------------------------------------------
Q9M7R2          ----------------------------------------------------MASLG- 5
SODF_SOYBN      ----------------------------------------------------MASLGG 6
SODF_NICPL      ------------------------------------------------------------
AAK62615        ------------------------------------------------------------
SODF_ARATH      ----------------------------------------------RFHWRVLSSLHS 12
Q9FE21          ------------------------------------------------------------
O65327          ------------------------------------------------------------
Q9LU64          ----------------------------------MMNVAVTATPSSLLYSPLLLPSQGP 25
O82583          ---------------------------------MAGLAVLPSTS-----ALPLPFPGC 20
Q42683          ------------------------------------------------MALAMKAQAS 10
Q9ZWM8          ---------------------------------MAAFASALR-------VLPSPPAAV 18
Q9SNQ0          ---------------------------------MAAFASALR-------VLPSPPAAV 18
Q9FMX0          ---------------------------------MSSCVVTTS-------CFYTISDSS 18
Q9LWS3          ----------------------------------------------------MDEDAE 6
SODF_SYNP7      ------------------------------------------------------------
SODF_SYNY3      ------------------------------------------------------------
```

```
O50257          ------------------------------------------------------------
Q9RFQ4          ------------------------------------------------------------
Q9R6Y6          ------------------------------------------------------------
SODF_PLEBO      ------------------------------------------------------------
AAK86683        ------------------------------------------------------------
Q985K3          ------------------------------------------------------------
CAC45545        ------------------------------------------------------------
SODF_RHIME      ------------------------------------------------------------
AAK88862        ----------------------------------------MDLPPKNRLTST 12
O30970          ------------------------------------------------------------
O15641          ------------------------------------------------MFRRV 5
O02615          ------------------------------------------------MLRRA 5
Q95051          ------------------------------------------------------------
Q95052          ------------------------------------------------------------
O43957          ------------------------------------------------------------
O84924          ------------------------------------------------------------
SODM_LACLA      ------------------------------------------------------------
AAK99478        ------------------------------------------------------------
SODM_STRPN      ------------------------------------------------------------
AAK74904        ------------------------------------------------------------
Q9AGW1          ------------------------------------------------------------
SODM_STRAG      ------------------------------------------------------------
SODM_STRPY      ------------------------------------------------------------
SODM_STRMU      ------------------------------------------------------------
AAF64074        ------------------------------------------------------------
SODM_BACCA      ------------------------------------------------------------
Q9LBF6          ------------------------------------------------------------
SODM_BACST      ------------------------------------------------------------
SODM_BACSU      ------------------------------------------------------------
O86168          ------------------------------------------------------------
Q9ZF38          ------------------------------------------------------------
Q9KD10          ------------------------------------------------------------
SODM_LISMO      ------------------------------------------------------------
SODM_LISIV      ------------------------------------------------------------
BAB57715        ------------------------------------------------------------
Q9Z5W5          ------------------------------------------------------------
Q9K4V3          ------------------------------------------------------------
Q9F326          ------------------------------------------------------------
BAB56295        ------------------------------------------------------------
Q99X82          ------------------------------------------------------------
Q9EZZ2          ------------------------------------------------------------
SOD2_PLEBO      ------------------------------------------------------------
SOD3_PLEBO      ------------------------------------------------------------A 1
SOD1_PLEBO      ----------------------------------------MQTTFRRILILF 12
SODM_PASHA      ------------------------------------------------------------
Q59133          ------------------------------------------------------------
SODM_HAEDU      ------------------------------------------------------------
SODM_PASMU      ------------------------------------------------------------
SODM_HAEIN      ------------------------------------------------------------
BAB38257        ------------------------------------------------------------
AAG59102        ------------------------------------------------------------
SODM_ECOLI      ------------------------------------------------------------
SODM_SALTY      ------------------------------------------------------------
SODM_YEREN      ------------------------------------------------------------
Q9RUV2          ------------------------------------------------------------
SODM_BUCAI      ------------------------------------------------------------
SODM_XANCP      ------------------------------------------------------------
SODF_METJ       ------------------------------------------------------------
Q9PAA4          ------------------------------------------------------------
SODM_PSEPU      ------------------------------------------------------------
Q9WWG7          ------------------------------------------------------------
```

```
SODM_PSEAE      ------------------------------------------------------------
SODM_BORPE      ------------------------------------------------------------
SODM_CHLRE      ------------------------------------------------------------
SODM_ACICA      ---------------------------------------------------MTRSL 5
Q9PC60          ------------------------------------------------MFLLRQG 7
Q9A7E6          ---------------------------------------------MQRLL 5
Q9KW87          ------------------------------------------------------------
Q9KW85          ------------------------------------------------------------
AAK53166        ------------------------------------------------------------
Q9KNN7          ------------------------------------------------------------
AAK53165        ------------------------------------------------------------
AAK53164        ------------------------------------------------------------
AAK53167        ------------------------------------------------------------
AAK53163        ------------------------------------------------------------
AAK53547        ------------------------------------------------------------
Q9KW86          ------------------------------------------------------------
SODM_THETH      ------------------------------------------------------------
SODM_THEAQ      ------------------------------------------------------------
SODM_BORBU      ------------------------------------------------------------
P96201          ------------------------------------------------------------
SODF_METTH      ------------------------------------------------------------
SODF_METTM      ------------------------------------------------------------
Q9P9I6          ------------------------------------------------------------
SODF_PYRAE      ------------------------------------------------------------
SODF_AERPE      ------------------------------------------------------------
SODF_SULSO      ------------------------------------------------------------
BAB67393        ------------------------------------------------------------
SODF_SULAC      ------------------------------------------------------------
SODF_ACIAM      ------------------------------------------------------------
Q9KCK8          MSIEWGQLHAWVNRVEEAWNQNKALLRSESVDTFLFEETLERLKEQLQNDGEPTEIGELA 60
SODF_BACSU      MKRES-YQAEMFNWCEALKDQ---IQKRGQLD--QFEDQIDKMIEALEDDQ------TTE 48
BAB62412        ------------------------------------------------------------
Q9APY3          ------------------------------------------------------------
Q9F9R1          ------------------------------------------------------------
Q9AM00          ------------------------------------------------------------
SODM_MYCAV      ------------------------------------------------------------
SODM_MYCLP      ------------------------------------------------------------
SODM_MYCLE      ------------------------------------------------------------
SODM_MYCFO      ------------------------------------------------------------
SODM_MYCSM      ------------------------------------------------------------
SODM_NOCAS      ------------------------------------------------------------
SODF_MYCTU      ------------------------------------------------------------
SODM_PROFR      ------------------------------------------------------------
Q9X469          ------------------------------------------------------------
SODF_STRCO      ------------------------------------------------------------
Q9X6N3          ------------------------------------------------------------
Q9HQF1          ------------------------------------------------------------
SOD1_HALCU      ------------------------------------------------------------
SOD1_HALSG      ------------------------------------------------------------
Q9HQ47          ------------------------------------------------------------
SODM_HALHA      ------------------------------------------------------------
SOD2_HALSG      ------------------------------------------------------------
SOD2_HALVO      ------------------------------------------------------------
SOD1_HALVO      ------------------------------------------------------------
SODM_HALMA      ------------------------------------------------------------
SODF_AQUAE      ------------------------------------------------------------
SODF_AQUPY      ------------------------------------------------------------
BAB59201        ------------------------------------------------------------
Q9HM56          ------------------------------------------------------------
CAC42412        ------------------------------------------------------------
SODM_ALCEU      ------------------------------------------------------------
```

```
AAH10548        ----------------------------------------MLCRAACS--TGRRL-- 13
BAB22095        ----------------------------------------MLCRAACS--TGRRL-- 13
SODM_MOUSE      ----------------------------------------MLCRAACS--TGRRL-- 13
SODM_RABIT      ----------------------------------------------------------
AAH12423        ----------------------------------------MLSRAVCG--TSRQL-- 13
SODM_HUMAN      ----------------------------------------MLSRAVCG--TSRQL-- 13
SODM_CAVPO      ----------------------------------------MLCRAVCS--ASRRL-- 13
SODM_HORSE      ----------------------------------------MLCRAACS--TSRKL-- 13
AAK97214        ----------------------------------------MLCRLASAGRSRAAL-- 15
Q9NB66          ----------------------------------------MLLARAFL--PARSL-- 13
SODM_CHAFE      ----------------------------------------MLLARAF----ARRS-- 11
SODM_CAEEL      ----------------------------------------MLQNTVRC--VSKLV-- 13
SODN_CAEEL      ----------------------------------------MLQSTART--ASKLV-- 13
SODM_ONCVO      ----------------------------------------MNLIIGVAG--RLLV-- 13
SODM_DROME      ----------------------------------------MFVARKIS--------- 8
SODM_CHLMU      ----------------------------------------------------------
SODM_CHLTR      ----------------------------------------------------------
SODM_CHLPN      ----------------------------------------------------------
Q9M532          ----------------------------------------MALRSLVTRRTLGLA-- 15
SODM_HEVBR      ----------------------------------------MALRSLVTRKNL----- 12
SODM_NICPL      ----------------------------------------MALRTLVSRRTL----- 12
Q9SM64          ----------------------------------------MALRTLVSRRTL----- 12
SODM_CAPAN      ----------------------------------------MALRNLMTKKPF----- 12
SODM_PEA        ----------------------------------------MAARTLLCRKTL----- 12
O82584          ----------------------------------------MAFQTLLAKKAL----- 12
Q9SRK3          ----------------------------------------MAIRCVASRKTL----- 12
SODM_ARATH      ----------------------------------------MAIRCVASRKTL----- 12
O65324          ----------------------------------------MAIRSVASRRTL----- 12
Q42672          ----------------------------------------MALRNLATRKTLS---- 13
Q96185          ----------------------------------------MALRTLAAKKTLGLA-- 15
O82571          ----------------------------------------MALRTLAAKKTLGLA-- 15
P93606          ----------------------------------------MALRTLAAKKTLGLA-- 15
Q43121          ----------------------------------------MALRTLASRKTL----- 12
SODM_ORYSA      ----------------------------------------MALRTLASRKTL----- 12
Q43803          ----------------------------------------MALRCVASRKTL----- 12
SODP_MAIZE      ----------------------------------------MALRTLASKNALSFA-- 15
SODN_MAIZE      ----------------------------------------MALRTLASKNALSFA-- 15
SODO_MAIZE      ----------------------------------------MALRTLASKNALSFA-- 15
SODM_MAIZE      ----------------------------------------MALRTLASKKVLSFP-- 15
Q43273          ----------------------------------------MALRTLASKKVLSFP-- 15
Q9LYK8          ----------------------------------------MTTTVIIIIFVAIFA-- 15
SODM_CANAL      ----------------------------MFSIRSSSRVLLKASSATTRATLNA 25
P79022          ----------------------------------MLSSAIKRSAVGIARRSVVSS 21
SODM_YEAST      ----------------------------------MFAKTAAANLTKKGGLSLLST 21
Q9P945          ----------------------------------MASLIRTSLRTGLRASSSS 19
Q9UQX0          ----------------------------------MLRFLSKNSVAAIRNVSIA 19
O74200          ----------------------------------MLKSFRDLVFKKTIKN 16
Q9P921          ----------------------------------------------------------
AAK82369        ----------------------------------------------MSGQ------- 4
SODM_GANMI      ----------------------------------------------MA---------- 2
Q9Y773          ----------------------------------------------MSGQ------- 4
SODM_AGABI      ----------------------------------------------MA---------- 2
SODM_ASPFU      ----------------------------------------------MSQQ------- 4
SODM_PENCH      ----------------------------------------------MTSQT------- 5
SODM_NEUCR      ----------------------------MVNLGSIWQNLLASQAPLQSMTG 23
SODF_TETPY      ----------------------------------------------------------
O42919          ----------------------------------------------MLNT 4
O74379          ----------------------------------------------------------
AAK80516        ----------------------------------------------------------
```

```
SODF_BACFR      ---------------------------MTYEMPKLPYANNALEPV-ISQQTIDYHYGK 30
SODF_PORGI      ---------------------------MTHELISLPYAVDALAPV-ISKETVEFHHGK 30
AAK88724        DITRRTILTAGAATATVAIFPRPHVAQAALPLTQPKLPFAEADLAPV-ISARTVGLHYGK 61
SODG_PSEPU      ---------------------------AFELPPLPYAHDALQPH-ISKETLEFHHDK 29
Q9AIX5          ---------------------------MAFELPPLPYAHDALQPH-ISKETLEFHHDK 30
Q9WWG8          ---------------------------MAFELPPLPYEKTALAPH-ISAETLEYHHDK 30
SODF_PSEAE      ---------------------------AFELPPLPYEKNALEPH-ISAETLEYHHDK 29
Q9ZGN1          ---------------------------MAFELPPLPYEKNALEPY-ISTETLEYHYGK 30
BAB35788        ---------------------------MSFELPALPYAKDALAPH-ISAETIEYHYGK 30
AAG56645        ---------------------------MSFELPALPYAKDALAPH-ISAETIEYHYGK 30
SODF_ECOLI      ---------------------------SFELPALPYAKDALAPH-ISAETIEYHYGK 29
Q9RQQ7          ----------------------MENRVMAFELPALPYAKDALEPH-ISAETLDYHHGK 35
Q9RCF0          ---------------------------MAFELPALPYAKDALEPH-ISAETLEYHYGK 30
Q9KQF3          ---------------------------MAFELPALPYAKDALEPH-ISAETLDFHHGK 30
SODF_PHOLE      ---------------------------AFELPALPFAMNALEPH-ISQETLEYHYGK 29
Q9JUW9          ---------------------------MEHKLPQLPYELDALSPH-LSKETLEFHYGK 30
Q9JZV6          ---------------------------MEHKLPQLPYELDALSPH-LSKETLEFHYGK 30
Q9F4F5          ---------------------------MEHKLPQLPYEPDALSPH-LSKETLEFHYGK 30
SODF_BORPE      ---------------------------MAHTLPPLPYALDALAPR-ISKETLEFHYGK 30
SODF_LEGPN      ---------------------------MTFTLPQLPYALDALAPH-VSKETLEYHYGK 30
SODF_COXBU      ---------------------------MAFELPDLPYKLNALEPH-ISQETLEYHHGK 30
Q9AXR7          -----MFAAAVSPIAAARASIVPRPHTRIMALQLPELPYSKDALAPH-INQQTLEFHYGK 54
O15904          ---------------------------MAFKLPALPYGMRELIPH-ISEETLSFHYGK 30
SODF_BABBO      ---------------------------MAFKLPALPYGMRELIPH-ISEETLSFHYGK 30
Q27740          ---------------------------MVITLPKLKYALNALSPH-ISEETLNFHYNK 30
O77071          ---------------------------MVFTLPPLPYAHDALAPH-ISSETLQFHHGK 30
Q27791          ---------------------------MVFSIPPLPWGYDGLAAKGLSKQQVTLHYDK 31
O02616          ---------------------------MVFSIPPLPWGYDGLAAKGLSKQQVTLHYDK 31
AAK52814        ---------------------------MTFSIPPLPWGYDGLAAKGISKEQVTFHYDK 31
O15640          ---------------------------MPFAVQPLPYPHDALASKGMSKEQVTFHHEK 31
AAL03316        ----------------------MTYCNKSNQTSYPFILPDLPYEKESFKPH-FTLETFDYHHGK 41
SODF_RICPR      ----------------------MTYCSKANQPSYPFILPDLPYDKESFKPH-FTRETFDYHHGK 41
Q9RM31          ---------------------------MFTLRELPFAKDSMGDF-LSPVAFDFHHGK 29
SODF_HELPY      ---------------------------MFTLRELPFAKDSMGDF-LSPVAFDFHHGK 29
Q9R2E7          ---------------------------MFTLRELPFAKDSMGDF-LSPVAFDFHHGK 29
Q9S6R8          ---------------------------MFTLRELPFATDSMGDF-LSPVAFDFHHGK 29
SODF_HELPJ      ---------------------------MFTLRELPFAKDSMGDF-LSPVAFDFHHGK 29
Q9R2E8          ---------------------------MFTLRELPFAKDSMGDF-LSPVAFDFHHGK 29
Q9R2E6          ---------------------------MFTLRELPFAKDSMGDF-LSPVAFDFHHGK 29
SODF_CAMCO      ---------------------------MFELRKLPYDTNAFGDF-LSAETFSYHHGK 29
SODF_CAMJE      ---------------------------MFELRKLPYDTNAFGDF-LSAETFSYHHGK 29
SODF_ENTHI      ---------------------------MSFQLPQLPYAYNALEPH-ISKETLEFHHDK 30
Q9A2K4          ---------------------------MTYTLPDLPYAYDALEPT-ISANTLRFHHDK 30
Q9M7R2          LQNVAG-------INLLFKEGP----KVNAKFELKPPPYPLNGLEPV-MSQQTLEFHWGK 53
SODF_SOYBN      LQNVSG-------INFLIKEGP----KVNAKFELKPPPYPLNGLEPV-MSQQTLEFHWGK 54
SODF_NICPL      ---------------------------KFELQPPPYPMDALEPH-MSSRTFEFHWGK 29
AAK62615        ----------------MAASS----AVTANYVLKPPPFALDALEPH-MSKQTLEFHWGK 38
SODF_ARATH      SLKLEL-------QLQRMAASS----AVTANYVLKPPPFALDALEPH-MSKQTLEFHWGK 60
Q9FE21          ----------------MAASS----AVTANYVLKPPPFALDALEPH-MSKQTLEFHWGK 38
O65327          ---------------MAASA----AVTANYVLKPPPYPLDALEPH-MSKQTLEFHWGK 38
Q9LU64          NRRMQWKRNGKRRLGTKVAVSG---VITAGFELKPPPYPLDALEPH-MSRETLDYHWGK 80
O82583          SRRAGSLR----RIAPKVGRSGP--LTIVAQLELKPPPYSLNALEPY-MSQETLEFHWGK 73
Q42683          SLVAGQRR----AVRPASGRRA---VITRAALELKSPPYALDALEPH-MSKQTLEFHWGK 62
Q9ZWM8          PRR-------LRSREQRQGCRSRRYSKVVAYYALTTPPYKLDALEPY-ISKRTVELHWGK 70
Q9SNQ0          PRR-------LRSREQRQGCRSRRYSKVVAYYALTTPPYKLDALEPY-ISKRTVELHWGK 70
Q9FMX0          IRLKSPKLLNLSNQQRRRSLRSRGGLKVEAYYGLKTPPYPLDALEPY-MSRRTLEVHWGK 77
Q9LWS3          ANG-----------DESSGTDED---ASVSWIEQQPLPYPSDALEPY-ISKETVEQHWGV 51
SODF_SYNP7      ---------------------------MSYELPALPFDYTALAPY-ITKETLEFHHDK 30
SODF_SYNY3      ---------------------------AYALPNLPYDYTALEPC-ISKSTLEFHHDK 29
```

```
O50257          ----------------------------MPPLPYPDTALEPA-ISARTISFHYGK 26
Q9RFQ4          --------------------------MAFVQDPLPFDINALDPYGMKAETFEYHYGK 31
Q9R6Y6          ----------------------------MAFVQEPLPYDFNALEQYGMKGETFEYHYGK 31
SODF_PLEBO      ----------------------------MAFTQPPLPFPKDALEPYGMKAETFDYHYGK 31
AAK86683        ----------------------------MAFELPELPYDYDALAPY-MSRETLEFHHDK 30
Q985K3          ----------------------------MAFELPALPYDYEALQPY-MSKETLEYHHDK 30
CAC45545        ----------------------------MAFELPNLPYDYDALAPY-MSRETLEYHHDK 30
SODF_RHIME      ----------------------------AFELPNLPYDYDALAPY-MSRETLEYHHDK 29
AAK88862        VIEVLEVLSPCFRKALSMLLSEQLTKEYPMAFQLPDLPYTHDALAPSGMSEETLKFHHDL 72
O30970          ----------------------------MAFELPALPYAHDALASLGMSKETLEYHHDL 31
O15641          SMKAATATAPVGFAFLCYHTLP-LLR---YPAELPTLGFNYKDGIQPVMSSRQLELHYKK 61
O02615          ---VNISIARGRMALMSYATLPDLLKPSGAPAELPKLGFNWKDGCAPVFSPRQMELHYTK 62
Q95051          --------------MFSIAPIP-----------------YMETGISGFLTKHAVEIHVTK 29
Q95052          --------------MFSIEPVS-----------------FLESGLPNFLTPHAVQIHVTK 29
O43957          ---------------MFTMEHPA-----------------YLKTGLPGFLTQHAVEVHVTK 29
O84924          ----------------------------MTYTLPDLPYAYDALEPY-IDEETMHLHHDK 30
SODM_LACLA      ----------------------------MAFTLPELPYAPNALEPF-FDEATMRLHHGK 30
AAK99478        ----------------------------MAIILPELPYAYDALEPY-IDAETMHLHHDK 30
SODM_STRPN      ----------------------------AIILPELPYAYDALEPY-IDAETMHLHHDK 29
AAK74904        ----------------------------MAIILPELPYAYDALEPY-IDAETMHLHHDK 30
Q9AGW1          ----------------------------MAIILPDLPYAYDALEPY-IDEETMHLHHDK 30
SODM_STRAG      ----------------------------AIILPDLPYAYDALEPH-IDAETMTLHHDK 29
SODM_STRPY      ----------------------------AIILPELPYAYDALEPQ-FDAETMTLHHDK 29
SODM_STRMU      ----------------------------AILLPDLPYAYDALEPY-IDAETMTLHHDK 29
AAF64074        ----------------------------MPFELPALPYPYDALEPH-IDKETMNIHHTK 30
SODM_BACCA      ----------------------------PFELPALPYPYDALEPH-IDKETMNIHHTK 29
Q9LBF6          ----------------------------MPFELPALPYPYDALEPH-IDKETMNIHHTK 30
SODM_BACST      ----------------------------PFELPALPYPYDALEPH-IDKETMNIHHTK 29
SODM_BACSU      ----------------------------AYELPELPYAYDALEPH-IDKETMTIHHTK 29
O86168          ----------------------------MAYELPELPYAYDALEPH-IDKETMTIHHTK 30
Q9ZF38          ----------------------------MAYKLPELPYAYDALEPH-IDKETMNIHHTK 30
Q9KD10          ----------------------------MAFELPKLPYPANALEPH-IDEATMNIHHGK 30
SODM_LISMO      ----------------------------MTYELPKLPYTYDALEPN-FDKETMEIHYTK 30
SODM_LISIV      ----------------------------MTYELPKLPYTYDALEPN-FDKETMEIHYTK 30
BAB57715        ----------------------------MAFELPKLPYAFDALEPH-FDKETMEIHHDR 30
Q9Z5W5          ----------------------------MAFELPKLPYAFDALEPH-FDKETMEIHHDR 30
Q9K4V3          ----------------------------MAFELPNLPYGFDALEPH-IDQQTMEIHHGK 30
Q9F326          ----------------------------MAFELPNLPYEFDALEPY-IDKETMEIHHDK 30
BAB56295        ----------------------------MAFKLPNLPYAYDALEPY-IDQRTMEFHHDK 30
Q99X82          ----------------------------MAFKLPNLPYAYDALEPY-IDQRTMEFHHDK 30
Q9EZZ2          ----------------------------MAFKLPNLPYAYDALEPY-IDQRTMEFHHDK 30
SOD2_PLEBO      ----------------------------MAFELKPLPYAYDALEPY-IDATTMQLHHDK 30
SOD3_PLEBO      STQQTPAQSPTASPTVSTPVAYVDRPLTASPAQLPPLPYDAGALSKA-IDAETMRIHHDR 60
SOD1_PLEBO      VGLLVPLFFACQSNSQVDAAPSAAPQLSASPAKLDPLPYDYAALEPY-IDAQTMRLHHDK 71
SODM_PASHA      ----------------------------MAYTLPELGYAYDALEPH-FDAKTMEIHHSK 30
Q59133          ----------------------------MAYQLPELGYAYDALEPH-FDKATMEIHHTK 30
SODM_HAEDU      ----------------------------TYQLPELGYAYDALEPY-FDKETMEIHHSK 29
SODM_PASMU      ----------------------------MAYTLPELGYAYDALEPH-FDAMTMEIHHSK 30
SODM_HAEIN      ----------------------------SYTLPELGYAYNALEPH-FDAQTMEIHHSK 29
BAB38257        ----------------------------MSYTLPSLPYAYDALEPH-FDKQTMEIHHTK 30
AAG59102        ----------------------------MSYTLPSLPYAYDALEPH-FDKQTMEIHHTK 30
SODM_ECOLI      ----------------------------SYTLPSLPYAYDALEPH-FDKQTMEIHHTK 29
SODM_SALTY      ----------------------------SYTLPSLPYAYDALEPH-FDKQTMEIHHTK 29
SODM_YEREN      ----------------------------MSYSLPSLPYAYDALEPH-FDKQTMEIHHTK 30
Q9RUV2          ----------------------------MAYTLPQLPYAYDALEPH-IDARTMEIHHTK 30
SODM_BUCAI      ----------------------------MSYVLPSLPYSYNALEPF-FDEETMKIHHTK 30
SODM_XANCP      ----------------------------MAYTLPQLPYAYDALEPN-IDAQTMEIHHTK 30
SODF_METJ       ----------------------------AYTLPPLDYAYTALEPH-IDAQTMEIHHTK 29
Q9PAA4          ----------------------------MAYTLPILPYAYDALQPH-IDAQTMEIHYTK 30
SODM_PSEPU      ----------------------------MPHTLPALPYAYDALEPH-IDAQTMEIHHTK 30
Q9WWG7          ----------------------------MPYTLPALPYAYDALEPH-IDAQTMEIHYTK 30
```

```
SODM_PSEAE      ----------------------------MPHALPPLPYAYDALEPH-IDALTMEIHHSK 30
SODM_BORPE      ----------------------------MPYVLPALSYAYDALEPH-IDARTMEIHHTR 30
SODM_CHLRE      ----------------------------MAQALPPLPYDYGSLEPH-VDATTMNIHHTK 30
SODM_ACICA      KTTLILLAS---SVISMS----ALAE-----FKQAPLPYATNALQPA-IDQQTMEIHYGK 52
Q9PC60          HLTLCLIAALLASFLSFASTQPQIAP-----FTLPPLPYAVSALEPA-IDTQTMTLHHDF 61
Q9A7E6          TAIAISAFGITTPVLAQTSPAPQVAAAPQPAFTLAPLPYVYEALSPV-IDTETMRIHHGR 64
Q9KW87          ----------------------------MSHTFPALPYPYDALEPY-IDAKTMEVHYSK 30
Q9KW85          ----------------------------MSHTFPELPYSYDALEPY-IDAKTMEVHYSK 30
AAK53166        ----------------------------MPHLFPDLPYAYDALEPY-IDTKTMEVHYSK 30
Q9KNN7          ------------MLSALLIAKRYSSKEDTMPHLFPDLPYAYDALEPY-IDTKTMEVHYSK 47
AAK53165        ----------------------------MPHLFPDLPYAYDALEPY-IDTKTMEVHYSK 30
AAK53164        ----------------------------MPHLFPDLPYAYDALEPY-IDTKTMEVHYSK 30
AAK53167        ----------------------------MPHLFPDLPYAYDALEPY-IDAKTMEVHYSK 30
AAK53163        ----------------------------MPHLFPDLPYAYDALEPY-IDAKTMEVHYSK 30
AAK53547        ----------------------------MPHLFPDLPYAYDALEPY-IDAKTMEVHYSK 30
Q9KW86          ----------------------------MPHLFPDLPYAYDALEPY-IDAKTMEVHYSK 30
SODM_THETH      ----------------------------PYPFKLPDLGYPYEALEPH-IDAKTMEIHHQK 31
SODM_THEAQ      ----------------------------PYPFKLPELGYPYEALEPH-IDARTMEIHHQK 31
SODM_BORBU      ----------------------------MFKLPELGYDYDAVEPY-IDAKTMEIHHSK 29
P96201          ---MKNNFLKHKKSPLMYPDYCGSSSTKGEGFKLKPLDYPYDALEPS-IDAETVKIHHDK 56
SODF_METTH      ----------------------MNDLEKKFYELPELPYPYDALEPHISREQ-LTIHHQK 36
SODF_METTM      ----------------------MEKKFYELPELPYPYDALEPYISEEQ-LRIHHEK 33
Q9P9I6          ----------------------MAKELYKLPPLKFGYGDLAPYISEEQ-LKLHHDK 33
SODF_PYRAE      ----------------------MVTTKRYTLPPLPYAYNALEPYISAEI-MQLHHQK 34
SODF_AERPE      ----------------------MVSFKRYELPPLPYNYNALEPYIIEEI-MKLHHQK 34
SODF_SULSO      ----------------------TLQIQFKKYELPPLPYKIDALEPYISKDI-IDVHYNG 36
BAB67393        ----------------------MAIQIQFKKYELPPLPYKVDALEPYISKDI-IDVHYNG 37
SODF_SULAC      ----------------------TQVIQLKRYEFPQLPYKVDALEPYISKDI-IDVHYNG 36
SODF_ACIAM      ----------------------MSSLTYLKKYELPPLPYNLDALEPYISKEI-IDVHYNG 37
Q9KCK8          EHLYEEWARLFSEQTEAGGSNGMRRPVPIGGHRLPPLPYPYEALEPYIDREI-MRLHHQK 119
SODF_BACSU      EDWYKQAAALYRDITESDDTSERRAYVPIGKHVLPKLPYKYSALEPYISRDI-MILHHTK 107
BAB62412        ----------------------------MAVYELPELDYAYDALEPHIAAEI-MELHHSK 31
Q9APY3          ----------------------------MAVYELPELDYAYDALEPHIAAEI-MELHHSK 31
Q9F9R1          ----------------------------MAEYTLPDLDWDYAALEPHISGQI-NEIHHTK 31
Q9AM00          ----------------------------MAEYTLPDLDWDYAALEPHISGQI-NEIHHTK 31
SODM_MYCAV      ----------------------------AEYTLPDLDWDYAALEPHISGQI-NEIHHTK 30
SODM_MYCLP      ----------------------------AEYTLPDLDWDYEALEPHISGQI-NEIHHTK 30
SODM_MYCLE      ----------------------------AEYTLPDLDWDYAALEPHISGEI-NEIHHTK 30
SODM_MYCFO      ----------------------------AEYTLPDLDYDYGALEPHISGQI-NELHHSK 30
SODM_MYCSM      ----------------------------AEYTLPDLDYDYGALEPHISGQI-NELHHSK 30
SODM_NOCAS      ----------------------------AEYTLPDLDYDYSALEPHISGQI-NELHHSK 30
SODF_MYCTU      ----------------------------MAEYTLPDLDWDYGALEPHISGQI-NELHHSK 31
SODM_PROFR      ----------------------------AVYTLPELPYDYSALEPYISGEI-MELHHDK 30
Q9X469          ----------------------------MALYTLPELPYDYSALAPVISPEI-IELHHDK 31
SODF_STRCO      ----------------------------SVYTLPELPYDYSALAPVISPEI-IELHHDK 30
Q9X6N3          ----------------------------MATYTLPELPYDYAALEPVINPQI-IELHHDK 31
Q9HQF1          ----------------------------MSEYELPPLPYDYDALEPHISEQV-LTWHHDT 31
SOD1_HALCU      ----------------------------SEYELPPLPYDYDALEPHISEQV-LTWHHDT 30
SOD1_HALSG      ----------------------------MSEYELPPLPYDYDALEPHISEQV-LTWHHDT 31
Q9HQ47          ----------------------------MSQHELPSLPYDYDALEPHISEQV-VTWHHDT 31
SODM_HALHA      ----------------------------MSQHELPSLPYDYDALEPHISEQV-VTWHHDT 31
SOD2_HALSG      ----------------------------MSQHELPSLPYDYDALEPHISEQA-VTWHHDT 31
SOD2_HALVO      ----------------------------MS-YELDPLPYEYDALEPHISEQV-LTWHHDT 30
SOD1_HALVO      ----------------------------MSDYELDPLPYEYDALEPHISEQV-LTWHHDT 31
SODM_HALMA      ----------------------MSEHSNPELPPLPYDYDALEPHISEQV-LTWHHDT 34
SODF_AQUAE      ----------------------------GVHKIQPKDHLKPQNLEGISNEQ-IEPHFEA 30
SODF_AQUPY      ----------------------------GVHKLEPKDHLKPQNLEGISNEQ-IEPHFEA 30
BAB59201        ----------------------------MAETWEIKEKLKPRGLEGISDVQ-IDNHFDV 30
Q9HM56          ----------------------------MAETWEIKEKLKPRGLEGISDVQ-IDNHFDV 30
CAC42412        ----------------------------MLYEMKPLG-CEPAKLTGLSEKL-IFSHYEN 29
SODM_ALCEU      ----------------------------MLYEMKPLG-CEPAKLTGLSEKL-IFSHYEN 29
```

```
AAH10548        ----GPVAG---------------AAGSRHKHSLPDLPYDYGALEPHINAQI-MQLHHSK 53
BAB22095        ----GPVAG---------------AAGSRHKHSLPDLPYDYGALEPHINAQI-MQLHHSK 53
SODM_MOUSE      ----GPVAG---------------AAGSRHKHSLPDLPYDYGALEPHINAQI-MQLHHSK 53
SODM_RABIT      ---------------------------HGRGMKHSLPDLPYDYGALEPHINAQI-MELHHSK 34
AAH12423        ----APVLG---------------YLGSRQKHSLPDLPYDYGALEPHINAQI-MQLHHSK 53
SODM_HUMAN      ----APALG---------------YLGSRQKHSLPDLPYDYGALEPHINAQI-MQLHHSK 53
SODM_CAVPO      ----APALG----------------ILGVRQKHSLPDLPYDYGALQPHINAEI-MQLHHSK 53
SODM_HORSE      ----VPALG----------------SLGSRQKHSLPDLQYDYGALEPYINAQI-MQLHHSK 53
AAK97214        ----VAPLG---------------CLVARQKHTLPDLPYDYGALEPHISAEI-MQLHHSK 55
Q9NB66          ----AAAG----------------LWCRQKHTLPDLPYDYGALEPTISAEI-MQLHHSK 51
SODM_CHAFE      ----LRAG----------------LWCRQKHTLPDLPYDYGALEPTISAEI-MQLHHSK 49
SODM_CAEEL      ----QPITG---------------VAAVRSKHSLPDLPYDYADLEPVISHEI-MQLHHQK 53
SODN_CAEEL      ----QPVAG---------------VLAVRSKHTLPDLPFDYADLEPVISHEI-MQLHHQK 53
SODM_ONCVO      ----GKNYC----------------LNTQRLKHVLPDLPYDYGALEPILSAEI-MQVHHGK 53
SODM_DROME      ------PNC----------------KPGVRGKHTLPKLPYDYAALEPIICREI-MELHHQK 46
SODM_CHLMU      ----------------------------MVFSSYKLPDLPYDYDALEPVISAEI-MHLHHQK 33
SODM_CHLTR      ---------------------------MVFSSYMLPALPYDYDALEPVISAEI-MQLHHQK 33
SODM_CHLPN      ---------------------------MSFVPYSLPELPYDYDALEPVISSEI-MILHHQK 33
Q9M532          ----SNSAKLVSGS-------AVAQLRGFKTFSLPDLPYDYGALEPAISGEI-MQLHHQK 63
SODM_HEVBR      ----PSAFKAATG---------LGQLRGLQTFSLPDLPYDYGALEPAISGEI-MQLHHQK 58
SODM_NICPL      ----ATGLG------------FRQQLRGLQTFSLPDLPYDYGALEPAISGDI-MQLHHQN 55
Q9SM64          ----ATGLG------------FRQQLRGLQTFSLPDLPYNYGALEPAISGDI-MQLHHQN 55
SODM_CAPAN      ----AGILT------------FRQQLRCVQTFSLPDLSYDYGALEPAISGEI-MQLHHQK 55
SODM_PEA        ----SSVLRNDAKPIGAAIAAASTQSRGLHVFTLPDLAYDYGALEPVISGEI-MQIHHQK 67
O82584          ----GTALRNGAAELGLAPALGLCQARKLQTFSLPDLPYDYGSLEPAISGEI-MRIHHQK 67
Q9SRK3          -----AGLKETSSR--------LLRIRGIQTFTLPDLPYDYGALEPAISGEI-MQIHHQK 58
SODM_ARATH      -----AGLKETSSR--------LLRIRGIQTFTLPDLPYDYGALEPAISGEI-MQIHHQK 58
O65324          -----AGLKETSSR--------LLGSRSIQTFTLPDLPYDYSALEPAISGEI-MQIHHQK 58
Q42672          ----ATGKR-------------SGSTCRLADATLPDLPYDYAPWS-AISGEI-MQLHHQK 54
Q96185          ----LGGARPP------------AAARGVATFTLPDLPYDYGALEPAVSGEI-MRLHHQK 58
O82571          ----LGGAR-------------------GVATFTLPDLPYDYGALEPAVSGEI-MRLHHQK 52
P93606          ----LGGARPL------------AAARGVATFTLPDLPYDFGALEPAVSGEI-MRLHHQK 58
Q43121          -----AAAALPLAA--------AAAARGVTTVALPDLPYDYGALEPAISGEI-MRLHHQK 58
SODM_ORYSA      -----AAAALPLAA--------AAAARGVTTVALPDLPYDYGALEPAISGEI-MRLHHQK 58
Q43803          -----AAAALPLAA--------AAAARGVTTVALPDLPYDYGALEPAISGEI-MRLHHQK 58
SODP_MAIZE      ----LGGAARP--S--------AASARGVTTVALPDLSYDFGALEPAISGEI-MRLHHQK 60
SODN_MAIZE      ----LGGAARP--S--------AASARGVTTVALPDLSYDFGALEPVISGEI-MRLHHQK 60
SODO_MAIZE      ----LGGAARP--S--------AESARGVTTVALPDLSYDFGALEPVISGEI-MRLHHQK 60
SODM_MAIZE      ----FGGAGRPLAA--------AASARGVTTVTLPDLSYDFGALEPAISGEI-MRLHHQK 62
Q43273          ----FGGAGRPLAA--------AASARGVTTVTLPDLSYDFGALEPAISGEI-MRLHHQK 62
Q9LYK8          ----TTLHDARGAT-------MEPCLESMKTASLPDLPYAYDALEPAISEEI-MRLHHQK 63
SODM_CANAL      AASK--------------------TFTRSKYSLPELDYEFSATEPYISGQI-NEIHYTK 63
P79022          SVG----------------------AVRTKVSLPDLDWDFGALEPHISGQI-NEIHYTK 57
SODM_YEAST      TAR-----------------------RTKVTLPDLKWDFGALEPYISGQI-NELHYTK 55
Q9P945          AAPL--------------------TFTRGKATLPDLAYDYGALEPAISGKI-MELHHKN 57
Q9UQX0          RG-----------------------VHTKATLPPLPYAYNALEPALSETI-MKLHHDK 53
O74200          SRF----------------------FSHSKHVLPSLPYDYQALEPYLSADL-IELHYNQ 52
Q9P921          -----------------------------------------ALEPYLSHDL-LELHYNK 17
AAK82369        ----------------------------HTLPDLPYAYDALEPFISRQI-MELHHKK 32
SODM_GANMI      -----------------------------HVLPDLPYAYNALEPFISQQI-MELHHKK 30
Q9Y773          ----------------------------HTLPDLPYSYDALEPYVSQQI-MTLHHKK 32
SODM_AGABI      ---------------------------HVLPDLPYAYDALEPYISRQI-MELHHKK 30
SODM_ASPFU      ----------------------------YTLPPLPYPYDALQPYISQQI-MELHHKK 32
SODM_PENCH      ---------------------------HTLPPLPYAYDALEPVISKQI-MELHHQK 33
SODM_NEUCR      NATT--------------------MAGLATYSLPQLPYAYNALEPYISAQI-MELHHSK 61
SODF_TETPY      ------------------------------LNYEYSDLEPVLSAHL-LSFHHGK 23
O42919          GLRKGLALSPITHLLKRCSSVTDNVHRVNYCYNYHTVPNLSQRNLLPLFSPEALDIAWDQ 64
O74379          ----------------------MSFQQRFARALDVGYKNFLSK--DAVRNIFAYDN 32
AAK80516        ------------------------------MITEKKYNFNSLAGF--SNELLKAHYDI 26
```

```
SODF_BACFR      HLQTYVNNLNSLVPG-TE--------YEGKTVEAIVASAP---------DGAIFNNAGQV 72
SODF_PORGI      HLKTYVDNLNKLIIG-TE--------FENADLNTIVQKS----------EGGIFNNAGQT 71
AAK88724        HHKAYFDKLNTLAAG-TR--------YADMELAGIVKESAGSKAAA---DVKIFNNAAQA 109
SODG_PSEPU      HHNTYVVNLNNLVPG-TE--------FEGKTLEEIVKT-----SSG-----GIFNNAAQV 70
Q9AIX5          HHNTYVVNLNNLVPG-TE--------FEGKTLEEIVKT-----SSG-----GIFNNAAQV 71
Q9WWG8          HHNTYVVNLNNLVPG-TE--------FEGKTLEEIVKT-----SSG-----GIFNNAAQV 71
SODF_PSEAE      HHNTYVVNLNNLIPG-TE--------FEGKSLEEIVKS-----SSG-----GIFNNAAQV 70
Q9ZGN1          HHNTYVVNLNNLIPG-TE--------FEGKSLEDIVKT-----SSG-----GIFNNAAQV 71
BAB35788        HHQTYVTNLNNLIKG-TA--------FEGKSLEEIIRS-----SEG-----GVFNNAAQV 71
AAG56645        HHQTYVTNLNNLIKG-TA--------FEGKSLEEIIRS-----SEG-----GVFNNAAQV 71
SODF_ECOLI      HHQTYVTNLNNLIKG-TA--------FEGKSLEEIIRS-----SEG-----GVFNNAAQV 70
Q9RQQ7          HHNTYVVKLNGLIPG-TE--------FEGKTLEEIIKT-----STG-----GVFNNAAQI 76
Q9RCF0          HHNTYVVKLNGLVPG-TE--------FEGKTLEEIIKT-----STG-----GIFNNAAQV 71
Q9KQF3          HHNTYVVKLNGLIPG-TE--------FENKSLEEIIKT-----STG-----GIFNNAAQV 71
SODF_PHOLE      HHNTYVVKLNGLVEG-TE--------LAEKSLEEIIKT-----STG-----GVFNNAAQV 70
Q9JUW9          HHQTYITNLNNQIKG-TE--------FENLPLEEIVKK-----SSG-----GVFNNAAQT 71
Q9JZV6          HHQTYITNLNNQIKG-TE--------FENLPLEEIVKK-----SSG-----GVFNNAAQT 71
Q9F4F5          HHQTYITNLNNQIKG-TE--------FENLPLEEIVKK-----SSG-----GVFNNAAQT 71
SODF_BORPE      HHQTYVTNLNNLVPG-TE--------FENLSLEEIVKK-----SSG-----GVFNNAAQV 71
SODF_LEGPN      HHNTYVTNLNKLIPG-TE--------FESMTLEEIIMK-----AKG-----GIFNNAAQV 71
SODF_COXBU      HHRAYVNKLNKLIEG-TP--------FEKEPLEEIIRK-----SDG-----GIFNNAAQH 71
Q9AXR7          HHNAYVTNLNKLIEG-TD--------LANAPLEEIVKAAASNASKA-----GLFNNAAQV 100
O15904          HHAGYVNKLNSLIKG-TP--------MESCTIEELILG-----QTG-----AVFNNAAQI 71
SODF_BABBO      HHAGYVNKLNSLIKG-TP--------LESCTIEELILG-----QTG-----AVFNNAAQI 71
Q27740          HHAGYVNKLNTLIKD-TP--------FAEKSLLDIVKE-----SSG-----AIFNNAAQI 71
O77071          HHAGYVAKLNGFIEG-TA--------FAGKTLEEVIRT-----STG-----AIFNNAAQV 71
Q27791          HHQGYVTKLNAAAQTNSA--------LATKSIEEIIRT-----EKG-----PIFNLAAQI 73
O02616          HHQGYVTKLNAAAQTNSA--------LATKSIEEIIRT-----EKG-----PIFNLAAQI 73
AAK52814        HHMGYVTKLNAVAACNAA--------VAAKSVEEIIRT-----EKG-----PIFNLAAQI 73
O15640          HHKGYAVKLNAAAESNSG--------LASKSLVDIIKS-----EKG-----PAFNCAAQI 73
AAL03316        HHNTYVQNLNNLLKDKEE--------LQKKDLEEIIEW-----SSQN-SNAAIFNNAAQI 87
SODF_RICPR      HHNSYVQNLNNLIKDREE--------LQKKDLEEIIEW-----SSQN-AEVAILNNASQI 87
Q9RM31          HHQTYVNNLNNLIKD-TD--------FEKSSLFDILTK-----SSG-----GVFNNAAQI 70
SODF_HELPY      HHQTYVNNLNNLIKG-TD--------FEKSSLFDILTK-----SSG-----GVFNNAAQI 70
Q9R2E7          HHQTYLNNLNNLIKG-TD--------FEKSSLFAILTK-----SSG-----GVFNNAAQI 70
Q9S6R8          HHQTYVKNLNNLIKG-TD--------FEKSSLFDILTK-----SSG-----GVFNNAAQI 70
SODF_HELPJ      HHQTYVNNLNNLIKG-TD--------FEKSSLFAILTK-----SSG-----GVFNNAAQI 70
Q9R2E8          HHRTYVNNLNNLIKG-TD--------FEKSSLFAILTK-----SSG-----GVFNNAAQI 70
Q9R2E6          HHQTYVNNLNNLIKG-TE--------FEKSSLFDILTK-----SSG-----GVFKNAAQI 70
SODF_CAMCO      HHNTYVTNLNNLIKY-TE--------FASKDLVSIIKS-----SSG-----GVFNNAAQV 70
SODF_CAMJE      HHNTYVTNLNNLIKD-TE--------FAGKDLVSIIKT-----SNG-----GVFNNAAQV 70
SODF_ENTHI      HHATYVNKLNGLVKG-TE--------QEHKTLEELIKQK----PTQ-----AIYNNAAQA 72
Q9A2K4          HHAAYVTALNGLLNG-DD--------KG--SLEAVIKGA----GPG-----KVFNNAAQA 70
Q9M7R2          HHRTYVENLKKQVVGT-E--------LDGKSLEEIIVTA---YNKG--DILPAFNNAAQV 99
SODF_SOYBN      HHKTYVENLKKQVVGT-E--------LDGKSLEEIIVTS---YNKG--DILPAFNNAAQV 100
SODF_NICPL      HHRAYVDNLNKQIDGT-E--------LDGKTLEDIILVT---YNKG--APLPAFNNAAQA 75
AAK62615        HHRAYVDNLKKQVLGT-E--------LEGKPLEHIIHST---YNNG--DLLPAFNNAAQA 84
SODF_ARATH      HHRAYVDNLKKQVLGT-E--------LEGKPLEHIIHST---YNNG--DLLPAFNNAAQA 106
Q9FE21          HHRAYVDNLKKQVLGT-E--------LEGKPLEHIIHST---YNNG--DLLPAFNNAAQA 84
O65327          HHRAYVDNLKKQVLGS-E--------LEGKALEHIIQNT---YNNG--DLLPPFNNAAQA 84
Q9LU64          HHKTYVENLNKQILGT-D--------LDALSLEEVVLLS---YNKG--NMLPAFNNAAQA 126
O82583          HHRGYVDGLNRQILGT-E--------LAGLSLEEIVIKS---YNEG--DLLPTFNNAAQI 119
Q42683          HHRAYVDNMNKQVAGT-P--------LDGKSLEEIVLAS---WNNG--QPTPVFNNAAQV 108
Q9ZWM8          HQQDYVDSLNKQLATS-M--------FYGYTLEELIKEA---YNNG--NPLPEYNNAAQV 116
Q9SNQ0          HQQDYVDSLNKQLATS-M--------FYGYTLEELIKEA---YNNG--NPLPEYNNAAQV 116
Q9FMX0          HHRGYVDNLNKQLGKDDR--------LYGYTMEELIKAT---YNNG--NPLPEFNNAAQV 124
Q9LWS3          HQNIHVERLNGMIGGS-E--------WEGMSLGQMMLSS---FNEGREAPHPPFFHAAQI 99
SODF_SYNP7      HHAAYVNNYNNAVKDT-D--------LDGQPIEAVIKAI---AGDA--SKAGLFNNAAQA 76
SODF_SYNY3      HHAAYVNNFNNAVAGT-D--------LDNQSIEDVIKAV---AGDA--SKAGIFNNAAQA 75
```

```
O50257          HTAAYYANLNKAVAGT-P--------MATMRLEDVVKLV---AGDP--AKGVLFNNAAHL 72
Q9RFQ4          HHKAYVDNLNKLTEGT-E--------LANKSLEEVIKTS---FNDS--SKAGIFNNAAQV 77
Q9R6Y6          HHKAYVDNLNKLTDGT-E--------LADKSLEEVIQIA---FKDA--YKAGIFNNAAQV 77
SODF_PLEBO      HHAAYVTNLNKLVEGT-P--------MESLSLEDVIKQS---FGDS--SKVGVFNNAAQV 77
AAK86683        HHKAYVDNGNKLAAEA-G--------LSDLSLEEVVKKS---FGT----NAGLFNNAAQH 74
Q985K3          HHKAYVDNGNKLAAEA-G--------MGDLSVEEVVKQS---FGK----NAGLFNNAAQH 74
CAC45545        HHLAYVTNGNKLAEEA-G--------LSDLSLEDIVKKS---YGT----NQPLFNNAGQH 74
SODF_RHIME      HHLAYVTNGNKLAEEA-G--------LSDLSLEDIVKKS---YGT----NQPLFNNAGQH 73
AAK88862        HHKAYVDNLNKAITGT-E--------WEKSSLEDIVRGT---YEKGAVAQSGIFNNASQH 120
O30970          HHKAYVDNGNKLIAGT-E--------WEGKSVEEIVKGT---YCAGAVAQSGIFNNASQH 79
O15641          HHSAYVDKLN-TLG--KG--------CEGKTIEEIILATSGTTESK-----VMNNQAAQH 105
O02615          HHKAYVDKLN-ALAG-TT--------YDGKSIEEIILAVANDAEKK-----GLFNQAAQH 107
Q95051          HHQAYVDFANKNVPG-TE--------FEGKPIEEIIQKATG----------PLFNNVAQH 70
Q95052          HHQGYVDMANKIVPE-SE--------FKGKSVEEIIQNASG----------PVFNNVAQH 70
O43957          HHQSYIDTANKLIVG-SG--------FEGKPIEEIIQKAQG----------PLFNNVAQH 70
O84924          HHNTYVTNLNAAIEKHPE--------LGEKTVEELLADF---SSVPEDIQTAVRNNGGGH 79
SODM_LACLA      HHQTYVNNLNAAIEKHNE--------LDDLSLEELLTDL---SAIPEDIRTAVRNNGGGH 79
AAK99478        HHQTYVNNANAALEKHPE--------IGED-LEALLADV---ESIPADIRQALINNGGGH 78
SODM_STRPN      HHQTYVNNANAALEKHPE--------IGED-LEALLADV---ESIPADIRQALINNGGGH 77
AAK74904        HHQTYVNNANAALEKHPE--------IGED-LEALLADV---ESIPADIRQALINNGGGH 78
Q9AGW1          HHNTYVTNVNAALAKHLE--------IGED-LEKLLADV---ESIPADIRQAVINNGGGH 78
SODM_STRAG      HHATYVANANAALEKHPE--------IGED-LEALLADV---SQIPEDIRQAVINNGGGH 77
SODM_STRPY      HHATYVANTNAALEKHPE--------IGEN-LEELLADV---TKIPEDIRQALINNGGGH 77
SODM_STRMU      HHATYVANANAALEKHPE--------IGEN-LEVLLADV---EQIPADIRQSLINNGGGH 77
AAF64074        HHNTYVTNLNAALEGHPD--------LQNKSLEELLSNL---EALPESIRTAVRNNGGGH 79
SODM_BACCA      HHNTYVTNLNAALEGHPD--------LQNKSLEELLSNL---EALPESIRTAVRNNGGGH 78
Q9LBF6          HHNTYVTNLNAALEGHPD--------LQNKSLEELLSNL---EALPESIRTAVRNNGGGH 79
SODM_BACST      HHNTYVTNLNAALEGHPD--------LQNKSLEELLSNL---EALPESIRTAVRNNGGGH 78
SODM_BACSU      HHNTYVTNLNKAVEGNTA--------LANKSVEELVADL---DSVPENIRTAVRNNGGGH 78
O86168          HHNTYVTNLNKAVEGNTA--------LANKSVEELVADL---DSVPENIRTAVRNNGGGH 79
Q9ZF38          HHNTYVTKLNEAVAGKQD--------LESKSVEELVANL---DAVPENIRTAVRNNGGGH 79
Q9KD10          HHNTYVTKLNAALEGHSA--------LAEKSIEALVSDL---DAVPENIRTAVRNNGGGH 79
SODM_LISMO      HHNTYVTKLNEAVAGHPE--------LASKSAEELVTNL---DSVPEDIRGAVRNHGGGH 79
SODM_LISIV      HHNIYVTKLNEAVSGHAE--------LASKPGEELVANL---DSVPEEIRGAVRNHGGGH 79
BAB57715        HHNTYVTKLNAAVEG-TD--------LESKSIEEIVANL---DSVPANIQTAVRNNGGGH 78
Q9Z5W5          HHNTYVTKLNAAVEG-TD--------LESKSIEEIVANL---DSVPANIQTAVRNNGGGH 78
Q9K4V3          HHNTYVTKLNAAVEG-TD--------LESKSIEEIVANL---DSVPENIQTAVRNNGGGH 78
Q9F326          HHNTYVTKLNAAIEG-TD--------LENKSIEEIVANL---DSVPSDIQTAVRNNGGGH 78
BAB56295        HHNTYVTKLNATVEG-TE--------LEHQSLADMIANL---DKVPEAMRMSVRNNGGGH 78
Q99X82          HHNTYVTKLNATVEG-TE--------LEHQSLADMIANL---DKVPEAMRMSVRNNGGGH 78
Q9EZZ2          HHNTYVTKLNATVEG-TE--------LEHQSLADMIANL---DKVPEAMRMSVRNNGGGH 78
SOD2_PLEBO      HHAAYVNNLNAAIEKYSD--------LQSMSVEDLVTHL---DRVPEDVRTTVRNNAGGH 79
SOD3_PLEBO      HHQTYVDNLNTALKDQPN--------LQNLSIEAMLRDL---NAVPENIRNTIRNNAGGH 109
SOD1_PLEBO      HHATYVNNINETLKAYPD--------LQKQSVDSLIQNL---NQVPEAIRTKIRNNGGGH 120
SODM_PASHA      HHQAYINNANAALEAHP--------ELLEKCPGALIKDL---SQVPAEKRIAVRNNVGGH 79
Q59133          HHQTYVNNANAALEAHP--------ELLEKCPGALIKDL---SQVPAEKRTAVRNNVGGH 79
SODM_HAEDU      HHQAYVNNSNALLEKHP--------ELLEKCPGALLKDL---TQVPAEKRTAVRNNLGGH 78
SODM_PASMU      HHQAYVNNANAALENLP--------ELAQGCPGQLLTKL---AEVPADKLTAIRNNVGGH 79
SODM_HAEIN      HHQAYVNNANAALEGLPA-------ELVEMYPGHLISNL---DKIPAEKRGALRNNAGGH 79
BAB38257        HHQTYVNNANAALESLP--------EFANLPVEELITKL---DQLPADKKTVLRNNAGGH 79
AAG59102        HHQTYVNNANAALESLP--------EFANLPVEELITKL---DQLPADKKTVLRNNAGGH 79
SODM_ECOLI      HHQTYVNNANAALESLP--------EFANLPVEELITKL---DQLPADKKTVLRNNAGGH 78
SODM_SALTY      HHQTYVNNANAALENLP--------EFASLPVEELITKL---DQVPADKKTVLRNNAGGH 78
SODM_YEREN      HHQTYVNNANTVLESFP--------ELAKFSVEDLIKDL---DKVPAEKRTFMRNNAGGH 79
Q9RUV2          HHQTYVDNANKALEGTE---------FADLPVEQLIQQL---DRVPADKKGALRNNAGGH 78
SODM_BUCAI      HHQNYINNTNSILENTT---------FSSLPIEELISIL---NEIILEKKNALRNNAGGH 78
SODM_XANCP      HHQTYINNVNAALEGTEY---------ADLPIEELVSKL---KSLPENLQGPVRNNGGGH 78
SODF_METJ       HHQTYINNVNAALEGTSF---------ANEPVEALLQKL---DSLPENLRGPVRNNGGGH 77
Q9PAA4          HHATYINNLNAALEGTEY---------ADLPIEELLRNL---KSLPESLQGPVRNNGGGC 78
SODM_PSEPU      HHQTYVNGLNAAIEGTEW---------AEWPVEKLVGAV---KQLPESLRGAVTNHGGGH 78
Q9WWG7          HHQTYINNLNAAVEGTEF---------ADWSVEKLVASV---QQLPENLRPAVINQGGGH 78
```

```
SODM_PSEAE      HHQTYVNNLNAALEGTPY---------AEQPVESLLRQL---AGLPEKLRTPVVNNGGGH 78
SODM_BORPE      HHQTYVNGLNAALEGAGLD--------SEEPVEQLLRRI---PALPPGIHGAVRNHGGGH 79
SODM_CHLRE      HHQTYVNNLNAALDKFPEL--------KDLGLVDLNKAVGT-DKLPKDVATVIRNNGGGH 81
SODM_ACICA      HHKAYVDNLNAQIKTYPE--------LDKTDLIQLQKQI--------SKYNTAVRNNGGGH 97
Q9PC60          HHKAYVDNLNAAIKDIPT--------LSGKTLEQLLAIA-------STLPPAVRNNAGGD 106
Q9A7E6          HHQAYVNALNGAVAVTPA--------LQGKSLDAVLSEV-------SRHPPVVRNNAGGH 109
Q9KW87          HHKTYFDKFVTAVQGSVL---------ETQSLDEIFAAI-------SQHSPAIRNNGGGY 74
Q9KW85          HHKTYYDKFLAAIQGSEL---------EHQTLTEIFSSI-------SQHSPAVRNNGGGY 74
AAK53166        HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 74
Q9KNN7          HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 91
AAK53165        HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 74
AAK53164        HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 74
AAK53167        HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 74
AAK53163        HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 74
AAK53547        HHRTYYDKFLSAIKGTEH---------EDRPLSEIFARV-------STLPAAVRNHGGGY 74
Q9KW86          HHRTYYDKFLSAIKGTDY---------QDQSLSEIFARV-------STLPAAVRNHGGGY 74
SODM_THETH      HHGAYVTNLNAALEKYPY--------LHGVEVEVLLRHL---AALPQDIQTAVRNNGGGH 80
SODM_THEAQ      HHGAYVTNLNAALEKYPY--------LQGAEVETLLRHL---TALPADIQAAVRNNGGGH 80
SODM_BORBU      HHNGFVMNLNSILEKMGK--------IHLTDVSNILKNI---HDFPEEFQTLIRNNAGGY 78
P96201          HQQAYVDKLNKALEKHPE--------LYGKSLYDILSNL---DDMPEDIMADLVNQGGGV 105
SODF_METTH      HHQAYVDGANALLRKLDE----ARESDT---DVDIKAALK-----------ELSFHVGGY 78
SODF_METTM      HHQAYVDGANGVLRKLDD----ARENGE---EVDIKAALK-----------ELSFHVGGY 75
Q9P9I6          HHQAYVTNANAAIEMMDK----ARKEGT---DFDYKATAK-----------AFSFNLAGH 75
SODF_PYRAE      HHQGYVNGANAALEKLEK----FRKGEA---QIDIRAVLR-----------DLSFHLNGH 76
SODF_AERPE      HHNTYVKGANAALEKIEK----HLKGEI---QIDVRAVMR-----------DFSFNYAGH 76
SODF_SULSO      HHKGYVNGANSLLERLEK----VVKGDLQTGQYDIQGIIR-----------GLTFNINGH 81
BAB67393        HHKGYVNGANSFLERLEK----IIRGEITSGQYDIQGLLR-----------GLVFNINGH 82
SODF_SULAC      HHKGYVNGANSLLDRLEK----LIKGDLPQGQYDLQGILR-----------GLTFNINGH 81
SODF_ACIAM      HHRGYVNGANSFVDRVNK----ILKGEISSGQYDIQGLLR-----------GLVFNINGH 82
Q9KCK8          HHQSYVEGLNKAEKEMER----ARRTNH---FQLIKHWER-----------EAAFNGAGH 161
SODF_BACSU      HHQSYVDGLNKAESELKK----ARATKN---YDLITHWER-----------ELAFHGAGH 149
BAB62412        HHATYVAGANAALEALEK----AREEGTN--PDQIRALSK-----------NLAFNLGGH 74
Q9APY3          HHATYVAGANAALEALEK----AREEGTN--PDQIRALSK-----------NLAFNLGGH 74
Q9F9R1          HHATYVKGVNDALAKLEE----ARAN-ED--HAAIFLNEK-----------NLAFHLGGH 73
Q9AM00          HHATYVKGVNDALAKLEE----ARAN-ED--HAAIFLNEK-----------NLAFHLGGH 73
SODM_MYCAV      HHATYVKGVNDALAKLEE----ARAN-ED--HAAIFLNEK-----------NLAFHLGGH 72
SODM_MYCLP      HHATYVKGVNDALAKLEE----ARAN-ED--HAAIFLNEK-----------NLAFHLGGH 72
SODM_MYCLE      HHAAYVKGVNDALAKLDE----ARAK-DD--HSAIFLNEK-----------NLAFHLGGH 72
SODM_MYCFO      HHAAYVKGVNDAVAKLDE----ARAN-GD--HAAIFLNEK-----------NLAFHLGGH 72
SODM_MYCSM      HHATYVKGVNDAIAKLEE----ARAN-GD--HAAIFLNEK-----------NLAFHLGGH 72
SODM_NOCAS      HHAAYVAGANTALEKLEA----AREA-GD--HSAIFLHEK-----------NLAFHLGGH 72
SODF_MYCTU      HHATYVKGANDAVAKLEE----ARAK-ED--HSAILLNEK-----------NLAFNLAGH 73
SODM_PROFR      HHKAYVDGANTALDKLAE----ARDK-AD--FGAINKLEK-----------DLAFNLAGH 72
Q9X469          HHAAYVKGANDTLEQLAE----AREK-ES--WGSINGLEK-----------NLAFHLSGH 73
SODF_STRCO      HHAAYVKGANDTLEQLAE----ARDK-ET--WGSINGLEK-----------NLAFHLSGH 72
Q9X6N3          HHAAYVKGANDTLEQLEE----ARDK-EA--WGAINGLQK-----------NLAFHLSGH 73
Q9HQF1          HHQGYVNGWNDAEETLAE-------NRETGDHASTAGALG-----------DVTHNGSGH 73
SOD1_HALCU      HHQGYVNGWNDAEETLAE-------NRETGDHASTAGALG-----------DVTHNGSGH 72
SOD1_HALSG      HHQGYVNGWNDAEETLAE-------NRETGDHASTAGALG-----------DVTHNGSGH 73
Q9HQ47          HHQSYVDGLNSAEETLAE-------NRETGDHASTAGALG-----------DVTHNGCGH 73
SODM_HALHA      HHQSYVDGLNSAEETLAE-------NRETGDHASTAGALG-----------DVTHNGCGH 73
SOD2_HALSG      HHQSYVDGLNSAEENLAG-------NRETGDHASTAGALG-----------DVTHNGCGH 73
SOD2_HALVO      HHQGYVNGWNADDETLAE-------NREAGEFGSSAGAVR-----------NVTHNGSGH 72
SOD1_HALVO      HHQGYVNGWNADDETLAE-------NREAGEFGSSAGAVR-----------NVTHNGSGH 73
SODM_HALMA      HHQGYVNGLESAEETLAE-------NRDAGDFGSSAAAWV-----------NVTHNGCGQ 76
SODF_AQUAE      HYKGYVAKYNEIQEKLADQ-NFADRSKANQNYSEYRELKV-----------EETFNYMGV 78
SODF_AQUPY      HYKGYVAKYNEIQEKLADQ-NFADRSKANQNYSEYRELKV-----------EETFNYMGV 78
BAB59201        HYKGYVNKLNEIWSRLPD----VDRSKANQNYSEFRALKL-----------EETFNYGGS 75
Q9HM56          HYKGYVNKLNEIWSRLPD----VDRSKANQNYSEFRALKL-----------EETFNYGGS 75
CAC42412        NYGGAVKRLNAITATLAE------LDMATAPVFTLNGLKR-----------EELIATNSM 72
SODM_ALCEU      NYGGAVKRLNAITATLAE------LDMATAPVFTLNGLKR-----------EELIATNSM 72
```

```
AAH10548        HHAAYVNNLNATEEK-------YHEALAKGDVTTQVALQP----------ALKFNGGGH 95
BAB22095        HHAAYVNNLNATEEK-------YHEALAKGDVTTQVALQP----------ALKFNGGGH 95
SODM_MOUSE      HHAAYVNNLNATEEK-------YHEALAKGDVTTQVALQP----------ALKFNGGGH 95
SODM_RABIT      HHAAYVNNLNATEEK-------YREALARGDVTAHVALQP----------ALKFKGGGH 76
AAH12423        HHAAYVNNLNVTEEK-------YQEALAKGDVTAQIALQP----------ALKFNGGGH 95
SODM_HUMAN      HHAAYVNNLNVTEEK-------YQEALAKGDVTAQIALQP----------ALKFNGGGH 95
SODM_CAVPO      HHAAYLNNLNIAEEK-------YQEALAKGDVTAQVALQP----------ALKFNGGGH 95
SODM_HORSE      HHAAYVNNLNVTEEK-------YQEALAKGDVTAQIALQP----------ALKFNGGGH 95
AAK97214        HHATYVNNLNVTEEK-------YKEALAKGDVTAQVSLQP----------ALKFNGGGH 97
Q9NB66          HHQTYVNNLNVAEEK-------LAEAKEKGDVSTIISLAP----------ALKFNGGGH 93
SODM_CHAFE      HHQTYVNNLNVAEEK-------LAEAKEKGDVSTIISLAP----------ALRFNGGGH 91
SODM_CAEEL      HHATYVNNLNQIEEK-------LHEAVSKGNVKEAIALQP----------ALKFNGGGH 95
SODN_CAEEL      HHATYVNNLNQIEEK-------LHEAVSKGNLKEAIALQP----------ALKFNGGGH 95
SODM_ONCVO      HHAAYVNALNQAEEK-------VKEALAKGDTQAAVAGTK----------LMNFNTGGH 95
SODM_DROME      HHQTYVNNLNAAEEQ-------LEEAKSKSDTTKLIQLAP----------ALRFNGGGH 88
SODM_CHLMU      HHQGYINNLNEALKS-------LDVASATQDLTGLIAINP----------ALRFNGGGH 75
SODM_CHLTR      HHQGYINNLNEALKS-------LDVANATQDLARLIAINP----------ALRFNGGSH 75
SODM_CHLPN      HHQIYINNLNAALKR-------LDAAETQQNLNELIALEP----------ALRFNGGGH 75
Q9M532          HHQTYITNYNKALEQ-------LHEATEKGDSSTVVKLQS----------AIKFNGGGH 105
SODM_HEVBR      HHQTYITNYNKALEQ-------LNDAIEKGDSAAVVKLQS----------AIKFNGGGH 100
SODM_NICPL      HHQTYVTNYNKALEQ-------LHDAISKGDAPTVAKLHS----------AIKFNGGGH 97
Q9SM64          HHQTYVTNYNKALEQ-------LHDAISKGDAPTVAKLHS----------AIKFNGGGH 97
SODM_CAPAN      HHQTYITNYNNALQQ-------LHDAINKGDSPTVAKLQG----------AIKFNGGGH 97
SODM_PEA        HHQTYITNYNKALEQ-------LHDAVAKADTSTTVKLQN----------AIKFNGGGH 109
O82584          HHQAYITNYNKALEQ-------LDEAIAKGDASKVVGLQS----------AIKFNGGGH 109
Q9SRK3          HHQAYVTNYNNALEQ-------LDQAVNKGDASTVVKLQS----------AIKFNGGGH 100
SODM_ARATH      HHQAYVTNYNNALEQ-------LDQAVNKGDASTVVKLQS----------AIKFNGGGH 100
O65324          HHQAYVTNYNNALEQ-------LDQAVNKGDASAVVKLQS----------AIKFNGGGH 100
Q42672          HHQAYVTNYNKALEQ-------LHEAISKGDASAAVKLQS----------AIKFNGGGH 96
Q96185          HHATYVANYNKALEQ-------LDAAVSKGDASAVVHLQS----------AIKFNGGGH 100
O82571          HHATYVAHYNKALEQ-------LDAAVSKGDASAVVHLQS----------AIKFNGGGH 94
P93606          HHATYVANYNKALEQ-------LDAAVSKGDASAVVHLQS----------AIKFNGGGH 100
Q43121          HHATYVANYNKALEQ-------LDAAVAKGDAPAIVHLQS----------AIKFNGGGH 100
SODM_ORYSA      HHATYVANYNKALEQ-------LDAAVAKGDAPAIVHLQS----------AIKFNGGGH 100
Q43803          HHATYVANYNKALEQ-------LDAAVAKGDAPAIVHLQS----------AIKFNGGGH 100
SODP_MAIZE      HHATYVGNYNKALEQ-------LDAAVAKGDASAVVQLQG----------AIKFNGGGH 102
SODN_MAIZE      HHATYVVNYNKALEQ-------LDAVVVKGDASAVVQLQG----------AIKFNGGGH 102
SODO_MAIZE      NHATYVVNYNKALEQ-------IDDVVVKGDDSAVVQLQG----------AIKFNGGGH 102
SODM_MAIZE      HHATYVANYNKALEQ-------LETAVSKGDASAVVQLQA----------AIKFNGGGH 104
Q43273          HHATYVANYNKALEQ-------LETAVSKGDASAVVQLQA----------AIKFNGGGH 104
Q9LYK8          HHQTYVTQYNKALNS-------LRSAMADGDHSSVVKLQS----------LIKFNGGGH 105
SODM_CANAL      HHQTYVNNLNASIEQ-------AVEAKSKGEVKKLVALEK----------AINFNGGGY 105
P79022          HHQTYVNGYNQAIEQ-------AARPR-QGEVKKTIELQK----------AINFHGGGY 98
SODM_YEAST      HHQTYVNGFNTAVDQFQELSDLLAKEPSPANARKMIAIQQ----------NIKFHGGGF 104
Q9P945          HHQTYVNSYNTAIEQ-------LQEAQASNNIAAQIALKP----------LINFHGGGH 99
Q9UQX0          HHQTYVNNLNAAQEK-------LADPN--LDLEGEVALQA----------AIKFNGGGH 93
O74200          HHRAYVTNLNKTIEK-----YYEGNESSLDSFINRLNLLT----------SIKFFAGGH 96
Q9P921          HHRAYVTNFNLALEK-----YNE-YDSSVD-LATRMNLLT----------SIKFHGGGH 59
AAK82369        HHQTYVNALNAAEQA-------YAKASTPKER---IALQA----------ALKFNGGGH 71
SODM_GANMI      HHQTYVNSLNAAEQA-------YAKASTPKER---IALQS----------ALKFNGGGH 69
Q9Y773          HHQTYVTALNAAEVS-------YAKTATPKER---IALQA----------ALRFNGGGH 71
SODM_AGABI      TSSDLCECAQHCRGC-------LRHSTAVVGG---FDLSL----------FFILTTLGH 69
SODM_ASPFU      HHQTYVNGLNAALEA-------QKKAAEATDVPKLVSVQQ----------AIKFNGGGH 74
SODM_PENCH      HHQTYINNLNAALSA-------QASATASNDVPTLISLQQ----------KLRFNGGGH 75
SODM_NEUCR      HHQTYVTNLNNALKV-------HVAAIASSDIPAQIAQQP----------AIKFNGGGH 103
SODF_TETPY      HHQAYVNNLNATYEQ-------IAAATKENDAHKIATLQS----------ALRFNLGGH 65
O42919          HQRQVVKELNDRVKG---------TELEDSSVFNIIFQTA-----ALPEHAATFQFASQA 110
O74379          HLRGLVQKECKIHQTPYR-------VPSDLMVQSASDPAR----------ANLFNYSSQL 75
AAK80516        YVN-YINDLNKLWKVSYTPG-----FRSDDTNHSLMRSLK----------SKEIYTLNGI 70
```

```
SODF_BACFR      LNHTLYFLQFAP---------KPAKNEPAGKLGEAIKRDFGSFENFKKEFNAASVGLFG- 122
SODF_PORGI      LNHNLYFTQFRP---------GKG-GAPKGKLGEAIDKQFGSFEKFKEEFNTAGTTLFG- 120
AAK88724        WNHVAYWDQFVP---------GGP-NRPTGDLAASLNETFGDYDGFIKRAVDVSDTVFG- 158
SODG_PSEPU      WNHTFYWNCLAP---------NAGGQPT-GALADAINAAFGSFDKFKEEFTKTSVGTFG- 119
Q9AIX5          WNHTFYWNCLEP---------NSGGQPT-GALADAINAAFGSFDKFKEEFTKTSVGTFG- 120
Q9WWG8          WNHTFYWNCLAP----------NAGGEPT-GALAEAINKAFGSFDKFKEEFSKTSIGTFG- 120
SODF_PSEAE      WNHTFYWNCLSP----------NGGGQPT-GALADAINAAFGSFDKFKEEFTKTSVGTFG- 119
Q9ZGN1          WNHTFYWNGLKP---------QGGGQPT-GPLADAINAAFGSFDKFKEEFTKVAIGTFG- 120
BAB35788        WNHTFYWNCLAP---------NAGGEPT-GKVAEAIAASFGSFADFKAQFTDAAIKNFG- 120
AAG56645        WNHTFYWNCLAP---------NAGGEPT-GKVAEAIAASFGSFADFKAQFTDAAIKNFG- 120
SODF_ECOLI      WNHTFYWNCLAP---------NAGGEPT-GKVAEAIAASFGSFADFKAQFTDAAIKNFG- 119
Q9RQQ7          WNHTFYWHCLAP---------NAGGEPT-GAVADAINAAFGSFEEFKAKFTDAAINNFG- 125
Q9RCF0          WNHTFYWHCLAP---------NAGGEPT-GAVADAINAAFGSFEEFKAKFTDSAINNFG- 120
Q9KQF3          WNHTFYWHCLSP---------NGGGEPT-GAVAEAINAAFGSFADFKAKFTDSAINNFG- 120
SODF_PHOLE      WNHTFYWNCLAP---------NAGGEPT-GEVAAAIEKAFGSFAEFKAKFTDSAINNFG- 119
Q9JUW9          WNHTFYWLGFTP---------KGQGKPS-GELAAAIDAKWGSFEKFQEAFSACAAGTFG- 120
Q9JZV6          WNHTFYWLGFTS---------KGQGKPA-GELAAAIDAKWGSFEKFQEAFNACAAGTFG- 120
Q9F4F5          WNHTFYWLGFTP---------KGQGKPA-GELAAAIDAKWGSFEKFQEAFNACAAGTFG- 120
SODF_BORPE      WNHTFYWNSLSP---------NGGGEPS-GALADAIKAKWGSVDAFKEAFNKSAAGNFG- 120
SODF_LEGPN      WNHTFYWHSMSP---------NGGGEPK-GRLAEAINKSFGSFAAFKEQFSQTAATTFG- 120
SODF_COXBU      WNHTFYWHCMSP---------DGGGDPS-GELASAIDKTFGSLEKFKALFTDSANNHFG- 120
Q9AXR7          WNHTFYWHSMKP---------GGGGAPH-GKVAELIERDFGSYEEFATQFKAAGGGQFG- 149
O15904          WNHTFYWNSMGP---------NCGGEPT-GPIRKKIEEKFGSFSAFKTDFSNLLAGHFG- 120
SODF_BABBO      WNHTFYWNSMGP---------NCGGEPT-GPIRKKIEEKFGSFSAFKTDFSNLLAGHFG- 120
Q27740          WNHTFYWDSMGP---------DCGGEPH-GEIKEKIQEDFGSFNNFKEQFSNILCGHFG- 120
O77071          WNHTFYFSSMKPPMS------GGGGEPT-GRLLDEIKKEFTSVENFKDEFSKVAAGHFG- 123
Q27791          FNHTFYWESMCP---------NGGGEPT-GKVADEINASFGSFAKFKEEFTNVAVGHFG- 122
O02616          FNHTFYWESMCP---------NGGGEPT-GKLADEINASFGSFAKFKEEFTNVAVGHFG- 122
AAK52814        FNHDFYWESISP---------NGGGEPS-GKLADAIADSFGSFARFKEEFTNAAVGHFG- 122
O15640          FNHDFFWRCLSR---------EAGGEPS-GPLASAIVDSFGTFASFKKEFTDAPNGHFG- 122
AAL03316        WNHTFFWHSIKPQ---------GGGKPS-GKILEQINKDFGSFEEFCEQFKQEAVGQFG- 136
SODF_RICPR      WNHTFFWYSIKPH---------GGGKPS-GKVFEQISKDFGSFEQFCAQFKQEAVGQFG- 136
Q9RM31          YNHDFYWDCLSP----------KATALS-DELKGALEKDFGSLEKFKEDFIKSATTLFG- 118
SODF_HELPY      YNHDFYWDCLSP----------KATALS-DELKGALEKDFGSLEKFKEDFIKSATTLFG- 118
Q9R2E7          YNHDFYWDCLSP----------KATALS-DELKGALEKDFGSLEKFKEDFIKSATTLFG- 118
Q9S6R8          YNHDFYWDCLSP----------KATALS-DELKGALEKDFGSLEKFKEDFIKSATTLFG- 118
SODF_HELPJ      YNHDFYWDCLSP----------KATALS-DELKEALEKDFGSLEKFKEDFIKSATTLFG- 118
Q9R2E8          YNHDFYWDCLSP----------KATAFS-DELKGALEKDFGSLEKFKEDFIKSATTLFG- 118
Q9R2E6          NNHDFYWDCLSP----------KATALS-DELKGALEKDFGSLEKFKEDFIKSATTLFG- 118
SODF_CAMCO      YNHDFYFDCIKPITGC--GCGGSCQSMD-ANLQAALEKEFGSLENFKAEFIKGATGVFG- 126
SODF_CAMJE      YNHDFYFDCIKPSTGC--GCGGSCQSID-ANLQAALEKEFGSLENFKAEFIKGATGVFG- 126
SODF_ENTHI      WNHAFYWKCMCG----------CGVKPS-EQLIAKLTAAFGGLEEFKKKFTEKAVGHFG- 120
Q9A2K4          WNHAFFWDGLSP----------TKTAPG-AELAAAIDATFGGMDALKEKFVAEGIGHFG- 118
Q9M7R2          WNHDFFWECMKPG---------GGGKPS-GELLELIERDFGSFEKFLDEFKAAAATQFG- 148
SODF_SOYBN      WNHDFFWECMKPG---------GGGKPS-GELLELIERDFGSFVKFLDEFKAAAATQFG- 149
SODF_NICPL      WNHQFFWESMKPN---------GGGEPS-GELLELINRDFGSYDAFVKEFKAAAATQFG- 124
AAK62615        WNHEFFWESMKPG---------GGGKPS-GELLALLERDFTSYEKFYEEFNAAAATQFG- 133
SODF_ARATH      WNHEFFWESMKPG---------GGGKPS-GELLALLERDFTSYEKFYEEFNAAAATQFG- 155
Q9FE21          WNHEFFWESMKPG---------GGGKPS-GELLALLERDFTSYEKFYEEFNAAAATQFG- 133
O65327          WNHEFFWESMKPG---------GGGKPS-GELLALLERDFTSYEKFYDEFNAAAATQFG- 133
Q9LU64          WNHEFFWESIQPG---------GGGKPT-GELLRLIERDFGSFEEFLERFKSAAASNFG- 175
O82583          WNHDFFWQSMKPD---------GGGKPF-GVLMELIERDFGSFEGMMAQFKNAALTQFG- 168
Q42683          WNHTFFWESMKPN---------GGGAPT-GALAEAITRDFGSLDKFKEEFKQAGMTQFG- 157
Q9ZWM8          WNHHFFWESMQPE---------GGGSPG-RGVLQQIEKDFGSFTNFREEFIRSALSLLG- 165
Q9SNQ0          WNHHFFWESMQPE---------GGGSPG-RGVLQQIEKDFGSFTNFREEFIRSALSLLG- 165
Q9FMX0          YNHDFFWESMQPG---------GGDTPQ-KGVLEQIDKDFGSFTNFREKFTNAALTQFG- 173
Q9LWS3          WNHDFYWRSMQPG---------GGGKPP-ERLLKFINRDFGSYDGMIRQFMDAASTQFG- 148
SODF_SYNP7      WNHSFYWNSIKPN---------GGGAPT-GALADKIAADFGSFENFVTEFKQAAATQFG- 125
SODF_SYNY3      WNHSFYWNCMKPG---------GGGQPS-GALADKINADFGSFDAFVEAFKQAGATQFG- 124
```

```
O50257          WNHTFYWAGMKPG---------GGGTPP-AKVADALKASFGSVEACMEQLSEAAKTQFA- 121
Q9RFQ4          WNHTFFWNSLKPA---------GGGAPT-GDLAARIDKDFGSFDKFKEEFSNAAATQFG- 126
Q9R6Y6          WNHTFFWNSLKPA---------GGGAPT-GEFAAKINQDFGSFDKLKEEFSNAAATQFG- 126
SODF_PLEBO      WNHTFFWNCLKAG---------GGGAPT-GELAAKIDAAFGSLDKFKEEFSNAAATQFG- 126
AAK86683        YNHVHFWKWMKKD---------GGGNKLPGKLEQAFASDLGGYDKFKADFIAAGTTQFG- 124
Q985K3          YNHIHFWKWMKKG---------GGGNKLPAALQKAFDSDLGGYDKFKADFVAAGTTQFG- 124
CAC45545        YNHVHFWKWMKKG---------GGGTSLPGKLDAAIKSDLGGYDKFRADFSAAGAGQFG- 124
SODF_RHIME      YNHVHFWKWMKKG---------GGGTSLPGKLDAAIKSDLGGYDKFRADFSAAGAGQFG- 123
AAK88862        WNHNLFWEIMGP----------QNHAIPSPLEDALTQSFGTVALFKQNFALAGASQFG- 168
O30970          WNHAQFWEMMGPG---------EDKKMPGALEKALVESFGSVAKFKEDFAAAGAGQFG- 128
O15641          FNHSFFWKCLSPG---------GKKIPKTLENAIANEFGSVDDFTVSFQQAGVNNFG- 153
O02615          FNHTFYFRCITPN----------GKAMPKSLESAVTAQFGSVEQFKDAFVQAGVNNFG- 155
Q95051          FNHAFFWNCLTAK----------KQEVPAGVASFLAKHFESVDNFKAQFVQKASTVFG- 118
Q95052          FNHSFFWKCLTAT----------TQEVPAAVASFLSKHFESVDNFKAQFVQKASTVFG- 118
O43957          FNHSFFWKSLSAE----------KVAVPAHVAELLKKNFGSVEKFQETFTAKASTVFG- 118
O84924          ANHTFFWEILGPNAGG---------EPTGAIKEAIEETFGSFEDFKEEFKTAATGRFG- 128
SODM_LACLA      LNHSQFWLWLRPNTDG--SE--N---HADGEIGDAIAKEFGSFETFKTEFKAAATGRFG- 131
AAK99478        LNHALFWELMTPEKTA----------PSAELAAAIDATFGSFEEFQAAFTAAATTRFG- 126
SODM_STRPN      LNHALFWELMTPEKTA----------PSAELAAAIDATFGSFEEFQAAFTAAATTRFG- 125
AAK74904        LNHALFWELMTPEKTA----------PSAELAAAIDATFGSFEEFQAAFTAAATTRFG- 126
Q9AGW1          LNHALFWELMTAEETA----------PSAELAADIDATFGSFEDFQAAFTTAATTRFG- 126
SODM_STRAG      LNHALFWELMSPEETQ----------ISQELSEDINATFGSFEDFKAAFTAAATGRFG- 125
SODM_STRPY      LNHALFWELLSPEKQD----------VTPDVAQAIDDAFGSFDAFKEQFTAAATGRFG- 125
SODM_STRMU      LNHALFWELLSPEKTK----------VTAEVAAAINEAFGSFDDFKAAFTAAATTRFG- 125
AAF64074        ANHSLFWTILSPNGGG--E--------PTGELAEAINKKFGSFTAFKDEFSKAAAGRFG- 128
SODM_BACCA      ANHSLFWTILSPNGGG--E--------PTGELAEAINKKFGSFTAFKDEFSKAAAGRFG- 127
Q9LBF6          ANHSLFWTILSPNGGG--E--------PTGELAEAINKKFGSFTAFKDEFSKAAAGRFS- 128
SODM_BACST      ANHSLFWTILSPNGGG--E--------PTGELADAINKKFGSFTAFKDEFSKAAAGRFG- 127
SODM_BACSU      ANHKLFWTLLSPNGGG--E--------PTGALAEEINSVFGSFDKFKEQFAAAAAGRFG- 127
O86168          ANHKLFWTLLSPNGGG--E--------PTGALAEEINSVFGSFDKFKEQFAAAAAGRFG- 128
Q9ZF38          ANHSLFWKLLSPNGGG--A--------PTGELAEAINSKFGSFDQFKEDFAAAAAARFG- 128
Q9KD10          ANHTLFWQILSPNGGG--A--------PTGELADAINAEFGSFDQFKEKFADAAANRFG- 128
SODM_LISMO      ANHTLFWSILSPNGGG--A--------PTGNLKAAIESEFGTFDEFKEKFNAAAAARFG- 128
SODM_LISIV      ANHTLFWSSLSPNGGG--A--------PTGNLKAAIESEFGTFDEFKEKFNAAAAARFG- 128
BAB57715        LNHSLFWELLSPN--S--E--------EKGTVVEKIKEQWGSLEEFKKEFADKAAARFG- 125
Q9Z5W5          LNHSLFWELLSPN--S--E--------EKGTVVEKIKEQWGSLEEFKKEFADKAAARFG- 125
Q9K4V3          LNHSLFWELLTPN--S--E--------EKGTVVDKIKEQWGSLDAFKEEFADKAAARFG- 125
Q9F326          LNHSLFWQLLTPN--S--E--------EKGTVIDKIKEEWGSLDKFKDEFAKKAAGQFG- 125
BAB56295        FNHSLFWEILSPN--S--E--------EKGGVIDDIKAQWGTLDEFKNEFANKATTLFG- 125
Q99X82          FNHSLFWEILSPN--S--E--------EKGGVIDDIKAQWGTLDEFKNEFANKATTLFG- 125
Q9EZZ2          FNHSLFWEILSPN--S--E--------EKGGVIDDIKAQWGTLDEFKNEFANKATTLFG- 125
SOD2_PLEBO      VNHTMFWEIMGANGSG--A--------PTGAISEAINNSFGSFDAFKQQFNDAGTKRFG- 128
SOD3_PLEBO      LNHTIFWQIMSPDGGG--Q--------PTGAIAQAINQTFGNFESFRKQFNEAGGDRFG- 158
SOD1_PLEBO      VNHTMFWQIMAPKAGG--T--------PTGAVAKAIDQTFGSFDAFKQQFNKAGADRFG- 169
SODM_PASHA      VNHTLFWKGLK--TG----------TTLQGALKEAIERDFGSVEAFQSEFEKAATTRVG- 126
Q59133          VNHTLFWKGLK--TG----------TTLQGALKDAIVRDFGSVEAFQAEFEKAAATRFG- 126
SODM_HAEDU      VNHTLFWKGLK--TG----------TTLQGALKDAIIRDFGSVKAFQAEFEQAAATRFG- 125
SODM_PASMU      LNHSLFWKSLK--KG----------TTLQGALKDAIVRDFGSVEAFQAEFEKAAATRFG- 126
SODM_HAEIN      TNHSLFWKSLK--KG----------TTLQGALKDAIERDFGSVDAFKAEFEKAAATRFG- 126
BAB38257        ANHSLFWKGLK--KG----------TTLQGDLKAAIERDFGSVDNFKAEFEKAAASRFG- 126
AAG59102        ANHSLFWKGLK--KG----------TTLQGDLKAAIERDFGSVDNFKAEFEKAAASRFG- 126
SODM_ECOLI      ANHSLFWKGLK--KG----------TTLQGDLKAAIERDFGSVDNFKAEFEKAAASRFG- 125
SODM_SALTY      ANHSLFWKGLK--TG----------TTLQGDLKAAIERDFGSVDNFKAEFEKAAATRFG- 125
SODM_YEREN      ANHSLFWKGLK--LG----------TTLTGDLKAAIERDFGSVDSFKEKFEAAAATRFG- 126
Q9RUV2          ANHSMFWQIMG--QGQ--GQ--NGANQPSGELLDAINSAFGSFDAFKQKFEDAAKTRFG- 131
SODM_BUCAI      INHSFFWKSLK--SG----------TVLTNDLKIEIEKQFGTIDEFKEKFESVALNHFG- 125
SODM_XANCP      ANHSLFWTVLS-PNGG--G-------EPKGEVAKAIDKDIGGFEKFKEAFTKAAVSRFG- 127
SODF_METJ       ANHSLFWKVLT-PNGG--G-------EPKGALADAIKSDIGGLDTFKEAFTKAALTRFG- 126
Q9PAA4          VNHSLFWTVMA-QGGG--G-------KPKGEVAKAIDQQLGGFETFKDAFTKAALSRFG- 127
SODM_PSEPU      ANHSLFWTVMS-PQGG--G-------EPHGQLAQAIASQLGGFDAFKEAFTKAALTRFG- 127
Q9WWG7          ANHSLFWEVMV-PGGG--G-------LPTGAVASAINEQLGGYESFKEAFTKAALTRFG- 127
```

```
SODM_PSEAE      ANHSLFWTVMS-PQGG--G-------RPDGDLGRAIDEQLGGFEAFKDAFTKAALTRFG- 127
SODM_BORPE      ANHSLLWTVMS-PSGG--G-------RPDGRLAADIQAQLGGHDAFQAAFTQAALGRFG- 128
SODM_CHLRE      YNHSFFWKVMTNPSNT--N-------GPNGDVKAAIEASFGSVDEMKAKFNAAAAGRFG- 131
SODM_ACICA      FNHTFFWESLAATNKT--G-------QPSPALVKQITQDFGSLDAFKQKFNEAASGRFG- 147
Q9PC60          WNHSEFWKMMAPVGKG--G-------KPSAALETQIKKDFGSLDAFKERFNKAATGRFG- 156
Q9A7E6          YNHTLFWSLMAPSGAG--G-------APSSKLLAQIDKDFGSLDAFKKAFETAGLQRFG- 159
Q9KW87          YNHILYWNCMSPEGGG--E--------PQGALAQAINDKFGHFEKFKDEFAQAAVNTFG- 123
Q9KW85          YNHILYWQCMSPNGGG--Q--------PQGALAHAIQEKFGTFAAFKEAFSQAAVNTFG- 123
AAK53166        YNHIVYWNCMKPNAGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 123
Q9KNN7          YNHIVYWNCMKPNAGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 140
AAK53165        YNHIVYWNCMKPNAGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 123
AAK53164        YNHIVYWNCMKPNAGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 123
AAK53167        YNHIVYWNCMKPNAGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 123
AAK53163        YNHIVYWNCMKPNGGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 123
AAK53547        YNHIVYWNCMKPNSGG--E--------PQGELAAEIERQFGSFAQFKEAFSQAAVNTFG- 123
Q9KW86          YNHILYWNCMKPNGGG--E--------PQGELAAAIERHFGSFAQFKEAFSQAAVNTFG- 123
SODM_THETH      LNHSLFWRLLTPGGAK--E--------PVGELKKAIDEQFGGFQALKEKLTQAAMGRFG- 129
SODM_THEAQ      LNHSLFWRLLTPGGAK--E--------PVGELKKAIDEQFGGFAALKEKLTQAAMGRFG- 129
SODM_BORBU      SNHTLYFRTLRPGNKD--N--------LFEKFKDDINAAFGSLDVLKANLKDTAMKIFG- 127
P96201          YNHEFYWSILGKG-CN--R--------PVAEIADAIDRDFGSFEEFKEKFKQCGISTFG- 153
SODF_METTH      VLHLFFWGNMGPADE---CGGEPS-----GKLAEYIEKDFGSFERFRKEFSQAAISAEG- 129
SODF_METTM      VLHLFFWGNMGPADE---CGGEPD-----GRLAEYIEKDFGSFQRFKKEFSQAAVSAEG- 126
Q9P9I6          VLHDYFWWEMTPASN---SSKEPV-----GELSEVIKEDFGSFDRFKKEFSQVASSVEG- 126
SODF_PYRAE      ILHSIFWPNMAPPGK---GGGKPG-----GKIADLINKFFGSFEKFKEEFSQAAKNVEG- 127
SODF_AERPE      IMHTIFWPNMAPPGK---GGGTPG-----GRVADLIEKQFGGFEKFKALFSAAAKTVEG- 127
SODF_SULSO      KLHALYWENMAPSGK---GGGKPG-----GALADLINKQYGSFDRFKQVFTETANSLPG- 132
BAB67393        KLHALYWQNMAPAGK---GGGKPG-----GALADLIDKQYGSFDKFKQLFTEAANSLPG- 133
SODF_SULAC      KLHAIYWNNMAPAGK---GGGKPG-----GALADLINKQYGSFDRFKQVFSESANSLPG- 132
SODF_ACIAM      KLHSLYWQNMAPAGK---GGGKPG-----GVIGDLIEKQYGSFEKFKALFTEAANSLPG- 133
Q9KCK8          YLHSIFWEVMSP--H---GGGEPK-----GELRRQIERDFGSFARFKNHFSQAAEKVEG- 210
SODF_BACSU      YLHSIFWFSMHP--N---GKRRPT-----GALFQMIDLSFGSYSAFKEHFTQASKKVEG- 198
BAB62412        TNHSVFWKNLSPNG-----GGEPT-----GELAEAINRDFGSFAKFQDHFNSAALGLQG- 123
Q9APY3          TNHSVFWKNLSPNG-----GGEPT-----GELAEAINRDFGSFAKFQDHFNSAALGLQG- 123
Q9F9R1          VNHSIWWKNLSPDG-----GDKPT-----GELAAAIDDAFGSFDKFRAQFSAAANGLQG- 122
Q9AM00          VNHSIWWKNLSPDG-----GDKPT-----GELAAAIDDAFGSFDKFRAQFSAAANGLQG- 122
SODM_MYCAV      VNHSIWWKNLSPDG-----GDKPT-----GELAAAIDDAFGSFDKFRAQFSAAANGLQG- 121
SODM_MYCLP      VNHSLWWKNLSPDG-----GDKPT-----GELAAAIDDAFGSFDKFRAQFSAAANGLQG- 121
SODM_MYCLE      VNHSIWWKNLSPNG-----GDKPT-----GGLATDIDETFGSFDKFRAQFSAAANGLQG- 121
SODM_MYCFO      VNHSIWWKNLSPNG-----GDKPT-----GDLAAAIDDQFGSFDKFQAQFTAAANGLQG- 121
SODM_MYCSM      INHSIWWKNLSPNG-----GDKPT-----GELAAAIDDQFGSFDKFQAQFTAATNGLQG- 121
SODM_NOCAS      VNHSIWWKNLSPNG-----GDKPV-----GELAAAIDDQFGSFDKFRAQFTAAANGLQG- 121
SODF_MYCTU      VNHTIWWKNLSPNG-----GDKPT-----GELAAAIADAFGSFDKFRAQFHAAATTVQG- 122
SODM_PROFR      VNHSVFWKNMAPKGS---APERPT-----DELGAAIDEFFGSFDNMKAQFTAAATGIQG- 123
Q9X469          ILHSIYWQNMTGPKD---GGGEPLAQDGVGDLADAITESFGSFAGFKAQLTKAAATTQG- 129
SODF_STRCO      ILHSIYWHNMTG--D---GGGEPLDKDGVGELADAIAESFGSFAGFRAQLTKAAATTQG- 126
Q9X6N3          ILHSIYWHNMTG--D---GGGEPLAADGVGDLADAITESFGSYAGFKSQLTKAAATTQG- 127
Q9HQF1          ILHTLFWQSMSPAG-----GDEPS-----GALADRIAADFGSYENWRAEFEAAASAASG- 122
SOD1_HALCU      ILHTLFWQSMSPAG-----GDEPS-----GALADRIAADFGSYENWRAEFEAAASAASG- 121
SOD1_HALSG      ILHTLFWQSMSPAG-----GDEPS-----GALADRIVADFGSYENWRAEFEAAASAASG- 122
Q9HQ47          YLHTMFWEHMSPDG-----GGEPS-----GALADRIAADFGSYENWRAEFEVAAGAASG- 122
SODM_HALHA      YLHTMFWEHMSPDG-----GGEPS-----GALADRIAADFGSYENWRAEFEVAAGAASG- 122
SOD2_HALSG      YLHTMFWEHMSPDG-----GGEPS-----GALADRIAADFGSYENWRAEFEVAAGAASG- 122
SOD2_HALVO      ILHDLFWQNMSPEG-----GDEPE-----GLLAERIAEDFGSYEAWKGEFEAAAGAAAG- 121
SOD1_HALVO      ILHDLFWQNMSPEG-----GDEPE-----GLLAERIAEDFGSYEAWKGEFEAAAGAAAG- 122
SODM_HALMA      DLHTLFWENMDPNG-----GGEPE-----GELLDRIEEDFGSYEGWKGEFEAAASAAGG- 125
SODF_AQUAE      VLHELYFGMLAPGG-----KGEPS-----EALKKKIEEDLGGLDACTNELKAAAIAFRG- 127
SODF_AQUPY      VLHELYFGMLTPGG-----KGEPS-----EALKKKIEEDIGGLDACTNELKAAAMAFRG- 127
BAB59201        FLHELYFEGLTP-S-----HSEVP-----DEFKDAVAKDFGSYEKWLEDFKATGTAFRG- 123
Q9HM56          LLHELYFEGLTP-K-----HSEVP-----KEFKDAVAKDFGSYEKWLEDFKATGTAFRG- 123
CAC42412        ILHEVYFDSLGD-------GGSLD-----GALKTAIERDFGSVERWQAEFTAMGKALGGG 120
SODM_ALCEU      ILHEVYFDSLGD-------GGSLD-----GALKTAIERDFGSVERWQAEFTAMGKALGGG 120
```

```
AAH10548        INHTIFWTNLSP------KGGGEPKGE----LLEAIKRDFGSFEKFKEKLTAVSVGVQG- 144
BAB22095        INHTIFWTNLSP------KGGGEPKGE----LLEAIKRDFGSFEKFKEKLTAVSVGVQG- 144
SODM_MOUSE      INHTIFWTNLSP------KGGGEPKGE----LLEAIKRDFGSFEKFKEKLTAVSVGVQG- 144
SODM_RABIT      INHTIFWTNLSP------NGGGEPKGE----LLEAIKRDFGSFDKFKERLTAVSVGVQG- 125
AAH12423        INHSIFWTNLSP------NGGGEPKGE----LLEAIKRDFGSFDKFKEKLTAASVGVQG- 144
SODM_HUMAN      INHSIFWTNLSP------NGGGEPKGE----LLEAIKRDFGSFDKFKEKLTAASVGVQG- 144
SODM_CAVPO      INHSIFWTNLSP------NGGGEPKGE----LLEAIKRDFGSFDKFKEKLTAVSVGVQG- 144
SODM_HORSE      INHTIFWTNLSP------NGGGEPKGK----LLDAIKRDFGSFDKFKEKLTAVSAGVQG- 144
AAK97214        INHTILWTNLSP------SGGGEPKGE----LMEAIKRDFGSFANFKEKLTAVSVGVQG- 146
Q9NB66          INHSIFWQNLSA------DGG-EPEGE----LLAAINRDFGSVENMKNQLSAQTVAVQG- 141
SODM_CHAFE      INHSIFWQNLSA------DGG-EPEGE----LLAAINRDFGSVENMKNQLSAQTVAVQG- 139
SODM_CAEEL      INHSIFWTNLAK------DGG-EPSAE----LLTAIKSDFGSLDNLQKQLSASTVAVQG- 143
SODN_CAEEL      INHSIFWTNLAK------DGG-EPSKE----LMDTIKRDFGSLDNLQKRLSDITIAVQG- 143
SODM_ONCVO      INHTLFWEGLTAV-----KNSGEPNSE----LMTAIKKDFGSLETMIDKLNAKTIAIQG- 145
SODM_DROME      INHTIFWQNLSP-------NKTQPSDD----LKKAIESQWKSLEEFKKELTTLTVAVQG- 136
SODM_CHLMU      INHSLFWEMLAPQN---KGGGTPPRHE----LLKLIEKFWGSFDNFLKNFISSSAAVQG- 127
SODM_CHLTR      INHSLFWEMLAPQG---KGGGVPPRHE----LLKLIEKFWGSFDNFLKNFITSSAAVQG- 127
SODM_CHLPN      INHSLFWETLAPID---QGGGQPPKHE----LLSLIERFWGTMDNFLKKLIEVAAGVQG- 127
Q9M532          INHSIFWKNLAPVG---EGGGELPHGS----LGWAIDKDFGSLEKLIQKMNTQGAAVQG- 157
SODM_HEVBR      VNHSIFWKNLAPVR---EGGGELPHGS----LGWAIDADFGSLEKLIQLMNAEGAALQG- 152
SODM_NICPL      INHSIFWKNLAPVR---EGGGEPPKGS----LGWAIDTNFGSLEALVQKMNAEGAALQG- 149
Q9SM64          INHSIFWKNLAPVR---EGGGEPPKGS----LGWAIDTNFGSLEALVQKMNAEGAALQG- 149
SODM_CAPAN      INHSVFWKNLAPTR---EGGGEPPKGS----LGSAIDTNFGSLEAVIQKMNAEGAALQG- 149
SODM_PEA        INHSIFWKNLAPVS---EGGGEPPKES----LGWAIDTNFGSLEALIQKINAEGAALQ-- 160
O82584          VNHSIFWKNLAPTK---EGGGEPPKGA----LGWAIDTDLGSFEALIQKISAEGAALQG- 161
Q9SRK3          VNHSIFWKNLAPSS---EGGGEPPKGS----LGSAIDAHFGSLEGLVKKMSAEGAAVQG- 152
SODM_ARATH      VNHSIFWKNLAPSS---EGGGEPPKGS----LGSAIDAHFGSLEGLVKKMSAEGAAVQG- 152
O65324          VNHSIFWKNLAPVK---EGGGEPPKGS----LGGAIDTSFGSLEGLVKKMSAEGAAVQG- 152
Q42672          INHSIFWKNLAPTR---EGGGEPPKSS----LGWEIDNHFGSLDALIQKMSAEGAAVQG- 148
Q96185          VNHSIFWKNLKPIS---EGGGEAPHGK----LGWAIDEDFGSIEKLIKKMNAEGAALQG- 152
O82571          VNHSIFWKNLKPIS---EGGGEAPHGK----LGWAIDEDFGSIEKLIKKMNAEG-ALQG- 145
P93606          VNHSIFWKNLKPIS---EGGGEPPHGK----LGWAIDEDFGSIEKLIKKMNAEGAALQG- 152
Q43121          VNHSIFWNNLKPIS---EGGGDPPHAK----LGWAIDEDFGSFEALVKKMSAEGAALQG- 152
SODM_ORYSA      VNHSIFWNNLKPIS---EGGGDPPHAK----LGWAIDEDFGSFEALVKKMSAEGAALQG- 152
Q43803          VNHSIFWNNLKPIS---EGGGEPPHAK----LGWAIDEDFGSFEALVKKMSAEGAALQG- 152
SODP_MAIZE      VNHSIFWKNLKPIS---EGGGEPPHGK----LGWAIDEDFGSFEALVKRMNAEGAALQG- 154
SODN_MAIZE      FNHSIFWENLKPIS---EGG-EPPHGK----LGWAIDEDFGSFEALVKRMNAEGAALQG- 153
SODO_MAIZE      VNHSIFWKNLKPIS---EGGGEPPHGK----LGWAIDEDFGSFEALVKRMNAEGAALQG- 154
SODM_MAIZE      VNHSIFWKNLKPIS---EGGGEPPHGK----LGWAIDEDFGSFEALVKKMNAEGAALQG- 156
Q43273          VNHSIFWKNLKPIS---EGGGEPPHGK----LGWAIDEDFGSFEALVKKMNAEGAAFQG- 156
Q9LYK8          VNHAIFWKNLAPVH---EGGGKPPHDP----LASAIDAHFGSLEGLIQKMNAEGAAVQG- 157
SODM_CANAL      LNHCLWWKNLAPVS---QGGGQPPSEDS--KLGKQIVKQFGSLDKLIEITNGKLAGIQG- 159
P79022          TNHCLFWKNLAPEK---QGGGEPPAEDS--EFAKRIVEQYGSLDNLKAITNGKLAGIQG- 152
SODM_YEAST      TNHCLFWENLAPES---QGGGEPPTG----ALAKAIDEQFGSLDELIKLTNTKLAGVQG- 156
Q9P945          INHTLFWENLAPKN---AGGGEPPSG----ALSKAINESFGSLENFQGQMNTALAAIQG- 151
Q9UQX0          INHSLFWKILAPQK---EGGGKPVTSG---SLHKAITSKWGSLEDFQKEMNAALASIQG- 146
O74200          INHSLYWENLLPNK---QGGGEIING----PLVEAIKKEWKSVDEFIRIFNMQLAGIQG- 148
Q9P921          INHSLYWESLLPPK---EGGGQVIDG----PLVDAIKKEWGSVDQFIRTFNTHLSGIQG- 111
AAK82369        INHSLFWKNLAPAASEGKGNGGVLHDGP---LKKAILESFGSFENFKKEFNTTTAAIQG- 127
SODM_GANMI      INHSLFWKNLAPAKSEGKGNGGALADGP---LKSAIEQNWGSVDNFIKEFNATTAAIQG- 125
Q9Y773          INHSLFWKNLAPAASEGKGKGGELKPGP---LKDAIDSNFGSLDSLKKEFNAATAAIQG- 127
SODM_AGABI      INHSLFWQNLAPAA----GAGGQLKPGP---LKDAIDQTFGGLDNLKKEFNTTTAGIQG- 121
SODM_ASPFU      INHSLFWKNLAPEK---SGG-GKIDQAP--VLKAAIEQRWGSFDKFKDAFNTTLLGIQG- 127
SODM_PENCH      INHSLFWKNLTPPG---TPA-NDIAGAP--ALREAIVSRWGSHEAFVKAFGAELLGLQG- 128
SODM_NEUCR      INHSLFWKNLAPAE---TPETNYSKAAP--SLAAEIEKTWGSFDEFKKAFSAALLGIQG- 157
SODF_TETPY      VNHWIYWDNLAPVK---SGGGVLPDEHS--PLTKAIKEKWGSYENFITLFNTRTAAIQG- 119
O42919          YNNHFFFQSLIGKRAA--DAKKNSKYEANAAINKAVNENFGSKENLLSKIHELASNSFG- 167
O74379          VNHDFFFSGLISPERP-----SADADLGAINLKPGIDASFGSFGELKSQMVDVGNSVFG- 129
AAK80516        KLHELYFENMTGKN-----------TKPYGPILKEIGKQFSSYNNFISYLTEVAISMKG- 118
                  :   :  :                      .   .
```

```
SODF_BACFR      SGWAWLSVDKDGK---------------LHITKEPNG SNPVRAG----------LKPL 155
SODF_PORGI      SGWVWLASDANGK---------------LSIEKEPNA GNPVRKG----------LNPL 153
AAK88724        TGWVWLTRDDDNK---------------LALIGYEDG NNPVAVG----------RPAY 191
SODG_PSEPU      SGWGWLVKKADGS---------------LALASTIGA GCPLTSG----------DTPL 152
Q9AIX5          SDWGCLVKKADGS---------------LALASTIGA GCPLTIG----------DTPL 153
Q9WWG8          SGWGWLVKKADGS---------------LALASTIGA GNPLTSG----------DTPL 153
SODF_PSEAE      SGWGWLVKKADGS---------------LALASTIGA GNPLTSG----------DTPL 152
Q9ZGN1          SGWAWLVKKADGS---------------LALASTIGA GNPLTSG----------DTPL 153
BAB35788        SGWTWLVKNSDGK---------------LAIVSTSNA GTPLTTD----------ATPL 153
AAG56645        SGWTWLVKNSDGK---------------LAIVSTSNA GTPLTTD----------ATPL 153
SODF_ECOLI      SGWTWLVKNSDGK---------------LAIVSTSNA GTPLTTD----------ATPL 152
Q9RQQ7          SSWTWLVKKADGS---------------LEIVNTSNA ATPLTEEG---------TTPL 159
Q9RCF0          SSWTWLVKKADGS---------------LAIVNTSNA ATPITEEG---------VTPL 154
Q9KQF3          SSWTWLVKKADGT---------------LAITNTSNA ATPLTEEG----------VTPL 154
SODF_PHOLE      SSWTWLVKNANGS---------------LAIVNTSNA GCPITEEG---------VTPL 153
Q9JUW9          SGWAWLVKTPAGG---------------LDLVSTSNA ATPLTTE----------NTPL 153
Q9JZV6          SGWAWLVKTPAGG---------------LDLVSTSNA ATPLTTE----------NTPL 153
Q9F4F5          SGWAWLVKTPVGG---------------LDLISTSNA ATPLTTE----------NTPL 153
SODF_BORPE      SGWTWLVKKADGT---------------LDIVNTSNA ATPLTTA----------DKAL 153
SODF_LEGPN      SGWAWLVQDQSGA---------------LKIINTSNA GTPMTEG----------LNAL 153
SODF_COXBU      SGWAWLVKDNNGK---------------LEVLSTVNA RNPMTEG----------KKPL 153
Q9AXR7          SGWAWLVADKEGK---------------LKVFQTHDA GCPLTEG----------LTPL 182
O15904          SGWGWLVLKDDGT---------------ADIVQTHDA GSPLKENL---------GRPL 154
SODF_BABBO      SGWGWLVLKDDGT---------------ADIVQTHDA GSPLKENL---------GRPL 154
Q27740          SGWGWLALNNNNK---------------LVILQTHDA GNPIKDNT----------GIPI 154
O77071          SGWAWLVWDKQGKK--------------VGIEQTHDA GTPITEPM----------KVPL 158
Q27791          SGLAWLVKDTNSGK--------------LKVYQTHDA GCPLTEPN----------LKPL 157
O02616          CGLAWPVKDTNSGK--------------LKVYQTHDA GCPLTEPN----------LKPL 157
AAK52814        SGWAWLVQDTTTKK--------------LKVFQTHDA GCPLTEAD----------LKPI 157
O15640          SGWAWLVKDKSSGK--------------LKVLQTHDA GCPLTEPN----------LVPM 157
AAL03316        SGWAWLIYHDN--R--------------LQIIKTANA GTPIAHG----------MKPL 168
SODF_RICPR      SGWTWVVYHDN--K--------------LQIIKTSNA GTPIVNF----------MKPI 168
Q9RM31          SGWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
SODF_HELPY      SGWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
Q9R2E7          SGWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
Q9S6R8          SGWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
SODF_HELPJ      SGWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
Q9R2E8          SAWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
Q9R2E6          SGWNWAAYNLDTQK--------------IEIIQTSNA QTPVTDK----------KVPL 152
SODF_CAMCO      SGWFWLVYNTKNQK--------------LEFVGTSNA ATPITED----------KVPL 160
SODF_CAMJE      SGWFWLVYNTKNQK--------------LEFVGTSNA ATPITED----------KVPL 160
SODF_ENTHI      SGWCWLV-EHDGK---------------LEIIDTHDA VNPMTNG----------MKPL 152
Q9A2K4          SGWVWLVSDASGA---------------LKIISTHDA ESPVTQD----------LTAL 151
Q9M7R2          SGWAWLAYKASKLDGENAANPPSADEDNKLVVIKSPNA VNPLVWGG---------YYPL 198
SODF_SOYBN      SGWAWLAYRARKFDGENVANPPSPDEDNKLVVLKSPNA VNPLVWGG---------YYPL 199
SODF_NICPL      SGWAWLAYKP-----------------EEKKLALVKTPNA ENPLVLG----------YTPL 158
AAK62615        AGWAWLAYSN-----------------EKLKVVKTPNA VNPLVLG----------SFPL 165
SODF_ARATH      AGWAWLAYSN-----------------EKLKVVKTPNA VNPLVLG----------SFPL 187
Q9FE21          AGWAWLAYSN-----------------EKLKVVKTPNA VNPLVLG----------SFPL 165
O65327          AGWAWLAYAD-----------------NKLKVVKTPNA VNPLVLG----------SFPL 165
Q9LU64          SGWTWLAYKANRLDVANAVNPLPKEEDKKLVIVKTPNA VNPLVWD----------YSPL 224
O82583          SGWAWLVYKANRLDVGNAVNPCPTEKDYKLIIEKTPNA VNPLIWD----------YNPI 217
Q42683          SGWAWLNADKT----------------GKLSISKSPNA VNPVVEG----------KTPI 190
Q9ZWM8          SGWVLLVLKRK----------------ERKLSVVHTQNA ISPLALG----------DIPL 199
Q9SNQ0          SGWVWLVLKRK----------------ERKFSVVHTQNA ISPLALG----------DIIM 199
Q9FMX0          SGWVWLVLKRE----------------ERRLEVVKTSNA INPLVWD----------DIPI 207
Q9LWS3          SGWVWLCYKTSKLPHVKSRSPIPSDNYGRLVISKSPNA INPLVWG---------------- 193
SODF_SYNP7      SGWAWLVL--DNG----------------TLKITKTGNA DTPIAHG----------QTPL 157
SODF_SYNY3      SGWAWLVL--DNG---------------TLKVTKTGNA ENPMTAG----------QTPL 156
```

```
O50257          SGWAWLAKGTENGK------------PVLKVMKTGNA DNPMTQG----------LTPV 157
Q9RFQ4          SGWSWLID--DGG--------------TLKVIKTPNA ENPLAHG----------KKAL 158
Q9R6Y6          SGWAWLID--DGG--------------TLKVTKTPNA ENPLAHG----------QKAL 158
SODF_PLEBO      SGWAWLVD--DGG--------------TLKVTKTPNA ENPLVHG----------QKPL 158
AAK86683        SGWAWLSVK----------------NGKLEISKTPNG ENPLVHG----------ADPI 156
Q985K3          SGWAWVSVK----------------DGKLAISKTPNG ENPLVHG----------ASPI 156
CAC45545        SGWAWLSVK----------------NGKLEISKTPNG ENPLVHG----------ATPI 156
SODF_RHIME      SGWAWLSVK----------------NGKLEISKTPNG ENPLVHG----------ATPI 155
AAK88862        SGWAWLVVDT---------------DGKLKITRTENG VNPLCFG----------QKAL 201
O30970          SGWAWLVKDS---------------DGALKITKTENG VNPLCFG----------QTAL 161
O15641          SGWTWLCVDPRTK-------------ELRIDNTSNA GCPLTSG----------LRPI 187
O02615          SGWTWLCVDPSNKN------------QLVIDNTSNA GCPLTK---------------- 185
Q95051          SGWCYLAQNKDKT--------------ISINQYSNA LNPVKDG----------GVPL 151
Q95052          SGWCYLAVNKDGS--------------VSINQYSNA LNPVKDG----------GVPL 151
O43957          SGWAYLYKTKDGK--------------LEIGQYSNA ANPVKDG----------LTPI 151
O84924          SGWAWLVVKD-GK--------------LAITSTANQ DSPLMDG----------QTPV 160
SODM_LACLA      SGWAWLVVDEAGK--------------LKVVSTANQ DNPISEG----------LTPV 164
AAK99478        SGWAWLVVNKEGK--------------LEVTSTANQ DTPISEG----------KKPI 159
SODM_STRPN      SGWAWLVVNKEGK--------------LEVTSTANQ DTPISEG----------KKPI 158
AAK74904        SGWAWLVVNKEGK--------------LEVTSTANQ DTPISEG----------KKPI 159
Q9AGW1          SGWAWLVVNKEGK-------------LEVVSTANQ DTPISEG----------KTPI 159
SODM_STRAG      SGWAWLVVNAEGK--------------LEVLSTANQ DTPIMEG----------KKPI 158
SODM_STRPY      SGWAWLVVNKEGQ--------------LEITSTANQ DTPISEG----------KKPI 158
SODM_STRMU      SGWAWLVVDKEGK--------------LEVTSTANQ DTPISQG----------LKPI 158
AAF64074        SGWAWLVVN-NGE--------------LEITSTPNQ DSPIMEG----------KTPI 160
SODM_BACCA      SGWAWLVVN-NGE--------------LEITSTPNQ DSPIMEG----------KTPI 159
Q9LBF6          PGWAWLVVN-NGE--------------LEITSTPNQ DSPIMEG----------KTPI 160
SODM_BACST      SGWAWLVVN-NGE--------------LEITSTPNQ DSPIMEG----------KTPI 159
SODM_BACSU      SGWAWLVVN-NGK--------------LEITSTPNQ DSPLSEG----------KTPI 159
O86168          SGWAWLVVN-NGK-------------LEITSTPNQ DSPLSEG----------KTPI 160
Q9ZF38          SGWAWLVVN-NGE--------------LEITSTPNQ DSPLSEG----------KTPI 160
Q9KD10          SGWAWLVVN-DGK--------------LEITSTPNQ DTPLMEG----------KTPI 160
SODM_LISMO      SGWAWLVVN-DGK--------------LEIVSTANQ DSPLSDG----------KTPV 160
SODM_LISIV      SGWAWLVVN-NGK--------------LEIVSTANQ DSPLSEG----------KTPV 160
BAB57715        SGWAWLVVN-NGQ-------------LEIVTTPNQ DNPLTEG----------KTPI 157
Q9Z5W5          SGWAWLVVN-NGQ--------------LEIVTTPNQ DNPLTEG----------KTPI 157
Q9K4V3          SGWAWLVVN-NGN--------------LEIVTTPNQ DNPITEG----------KTPI 157
Q9F326          SGWAWLVVDKDGK-------------LEIVSTPNQ DNPITEG----------KTPI 158
BAB56295        SGWTWLVVN-DGK--------------LEIVTTPNQ DNPLTEG----------KTPI 157
Q99X82          SGWTWLVVN-DGK--------------LEIVTTPNQ DNPLTEG----------KTPI 157
Q9EZZ2          SGWTWLVVN-DGK--------------LEIVTTPNQ DNPLTEG----------KTPI 157
SOD2_PLEBO      SGWVWLVRSQQGD--------------LQILSTPNQ DSPLIEG----------HTPI 161
SOD3_PLEBO      SGWVWLVRNAQGQ--------------LQIVSTANQ DNPIMEG----------NYPI 191
SOD1_PLEBO      SGWAWLVSDRQGK--------------LSITSTANQ DNPLMSNP---------NAYPI 204
SODM_PASHA      SGWAWLVLEE-GK--------------LAVVSTANQ DNPLMGKEVAGV-----SGYPI 164
Q59133          SGWAWLVVEG-GK--------------LAVVSTANQ DNPIMGKEVAGV-----SGFPI 164
SODM_HAEDU      SGWAWLVLEE-GK--------------LAVVSTANQ DSPIMGKDVAGV-----SGYPI 163
SODM_PASMU      SGWAWLVLQENGK--------------LAVVSTANQ DSPVMGKAIAGC-----EGYPL 165
SODM_HAEIN      SGWAWLVLTAEGK--------------LAVVSTANQ DNPLMGKEVAGC-----EGFPL 165
BAB38257        SGWAWLVLKGD-K--------------LAVVSTANQ DSPLMGEAISGA-----SGFPI 164
AAG59102        SGWAWLVLKGD-K--------------LAVVSTANQ DSPLMGEAISGA-----SGFPI 164
SODM_ECOLI      SGWAWLVLKGD-K--------------LAVVSTANQ DSPLMGEAISGA-----SGFPI 163
SODM_SALTY      SGWAWLVLKGD-K--------------LAVVSTANQ DSPLMGEAISGA-----SGFPI 163
SODM_YEREN      SGWAWLVLKDDGK--------------LAVVSTANQ DSPLMGEAVSGA-----SGFPI 165
Q9RUV2          SGWAWLVVKD-GK--------------LDVVSTANQ DNPLMGEAIAGV-----SGTPI 169
SODM_BUCAI      SGWVWLVNQN-GV--------------LSIVSTVNQ DSPLMGKLISNT-----YGYPI 163
SODM_XANCP      SGWAWLSVTPDKK--------------LVVESTANQ DSPLFE----------GNTPI 160
SODF_METJ       SGWAWLSVTPEKK--------------LVVESTGNQ DSPLST----------GNTPI 159
Q9PAA4          SGWSLLNVTPSKT--------------LVVESSANQ DSPLCD----------GKTPI 160
SODM_PSEPU      SGWAWLSVTPQKT--------------LVVESSGNQ DSPLMF----------GNTPI 160
Q9WWG7          SGWAWLSVTPDKK--------------LVVESSGNQ DSPLMN----------GNTPI 160
```

```
SODM_PSEAE      SGWAWLSVTPQGS----------------LLVESSGNQ DSPLMN----------GNTPI 160
SODM_BORPE      SGWAWLTVTPAGR----------------LRVDSSANQ DSPLME----------GNTPI 161
SODM_CHLRE      SGWAWLSVKPDGS----------------LSIDSTPNQ DNPLMTALPDVA-----GGIPL 170
SODM_ACICA      SGWAWLIVTPNGK----------------LAVTSTPNQ DNPLMDLSETK-------GTPL 184
Q9PC60          SGWAWMILTSSG-----------------LQITSTPNQ DNPLMDVAEVR-------GQPL 192
Q9A7E6          SGWAWLIWSEGK-----------------LRVVSTPNQ DNPLMDDAPVK-------GAPI 195
Q9KW87          SGFAWLVVK-DGD----------------IHITSTSNQ DNPLMDVADVR-------GEPI 159
Q9KW85          SGFAWLVVK-EGD----------------IHITSTSNQ DNPLMDIASVR-------GEPI 159
AAK53166        SGFVWLIVQ-QGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 159
Q9KNN7          SGFVWLIVQ-QGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 176
AAK53165        SGFVWLIVQ-QGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 159
AAK53164        SGFVWLIVQ-QGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 159
AAK53167        SGFVWLIVQ-QGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 159
AAK53163        SGFVWLIVQ-QGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 159
AAK53547        SGFVWLIVQ-HGQ-----------------LSITSTSNQ DNPLMDVVAVR-------GEPI 159
Q9KW86          SGFVWLIVQ-EGQ-----------------LSISSTSNQ DNPLMDVVAVR-------GEPI 159
SODM_THETH      SGWAWLVKDPFGK-----------------LHVLSTPNQ DNPVMEG----------FTPI 162
SODM_THEAQ      SGWAWLVKDPFGK-----------------LHVISTANQ DNPVMGG----------FAPI 162
SODM_BORBU      SGWAWLVLCPESG-----------------LKVISMPNQ DSPLMNS----------YKPI 160
P96201          SGWAWLVSDKDGK-----------------LEIMSTKDQ SSPISLG----------LIPI 186
SODF_METTH      SGWAVLTYCQR-TDR---------------LFIMQVEKH NVNVIPH  ---------FRIL 163
SODF_METTM      SGWAVLTYCQR-TDR---------------LFIMQVEKH NVNVIPH  ---------FRIL 160
Q9P9I6          SGWAVLTYCKD-TER---------------LMIMQIEKH NVNLVPD  ---------YPII 160
SODF_PYRAE      VGWAILVYEPL-EEQ---------------LLILQIEKH NLMHAAD  ---------AQVL 161
SODF_AERPE      VGWGVLAFDPL-TEE---------------LRILQVEKH NVLMTAG  ---------LVPI 161
SODF_SULSO      TGWAVLYYDTE-SGN---------------LQIMTFENH FQNHIAE  ---------IPII 166
BAB67393        TGWTVLYYDTE-SGN---------------LEIMTFENH FQNHIAE  ---------LPII 167
SODF_SULAC      SGWTVLYYDNE-SGN----------------LQIMTVENH FMNHIAE  ---------LPVI 166
SODF_ACIAM      TGWTVLYYEVE-NGN---------------LQIMTFENH FQNHIAE  ---------LPIL 167
Q9KCK8          VGWAILVWAPR-AHR---------------LEILQAEFH QNLSQQD  ---------AIPL 244
SODF_BACSU      VGWAILVWAPR-SGR---------------LEILTAEKH QLFSQWD  ---------VIPL 232
BAB62412        SGWAVLGYDHI-SGR---------------LVIEQLTDQ QGNISVD  ---------ITPV 157
Q9APY3          SGWAVLGYDHI-SGR---------------LVIEQLTDQ QGNISVD  ---------ITPV 157
Q9F9R1          SGWAVLGYDTV-GSR---------------LLTFQLYDQ QANVPLG  ---------IIPL 156
Q9AM00          SGWAVLGYDTV-GSRL------------LTLLTFQLYDQ QANVPLG  ---------IIPL 159
SODM_MYCAV      SGWAVLGYDTL-GSR---------------LLTFQLYDQ QANVPLG  ---------IIPL 155
SODM_MYCLP      SGWAVLGYDTL-GSR----------------LLTFQLYDQ QANVPLG  ---------IIPL 155
SODM_MYCLE      SGWAVLGYDTL-GNK---------------LLTFQLYDQ QANVSLG  ---------IIPL 155
SODM_MYCFO      SGWAVLGYDSL-GDR---------------LLTFQLYDQ QANVPLG  ---------IIPL 155
SODM_MYCSM      SGWAVLGYDSL-GGR---------------LLTFQLYDQ QANVPLG  ---------IIPL 155
SODM_NOCAS      SGWAVLGYDTL-GQK---------------LLTFQLYDQ QANVPLG  ---------IIPL 155
SODF_MYCTU      SGWAALGWDTL-GNK---------------LLIFQVYDH QTNFPLG  ---------IVPL 156
SODM_PROFR      SGWASLVWDPL-GKR----------------INTLQFYDH QNNLPAG  ---------SIPL 157
Q9X469          SGWGVLAYEPL-SGR---------------LIVEQVYDH QGNVGQG  ---------ATPI 163
SODF_STRCO      SGWGVLAYEPL-SGR---------------LIVEQIYDH QGNVGQG  ---------STPI 160
Q9X6N3          SGWGVLAYEPV-SGK---------------LIVEQVYDH QGNVGQG  ---------SVPI 161
Q9HQF1          --WALLVYDSH-SNT---------------LRNVAVDNH DEGALWG  ---------SHPI 154
SOD1_HALCU      --WALLVYDSH-SNT---------------LRNVAVDNH DEGALWG  ---------SHPI 153
SOD1_HALSG      --WALLVYDSH-SNT---------------LRNVAVDNH DEGALWG  ---------SHPI 154
Q9HQ47          --WALLVYDPV-AKQ---------------LRNVAVDNH DEGALWG  ---------SHPI 154
SODM_HALHA      --WALLVYDPV-AKQ----------------LRNVAVDNH DEGALWG  ---------SHPI 154
SOD2_HALSG      --WALLVYDPV-AKQ---------------LRNVAVDNH DEGALWG --------- SHPI 154
SOD2_HALVO      --WALLVYDSF-SNQ---------------LRNVVVDKH DQGALWG --------- SHPI 153
SOD1_HALVO      --WALLVYDSF-SNQ---------------LRNVVVDKH DQGALWG --------- SHPI 154
SODM_HALMA      --WALLVYDPC-AKQ---------------LRNVPVDKH DQGALWG --------- SHPI 157
SODF_AQUAE      --WAILGLDIF-SGR---------------LVVNGLDAH NVYNLTG --------- LIPL 159
SODF_AQUPY      --WAILGLDIF-SGR---------------LVVNGLDAH NVYNLTG --------- LIPL 159
BAB59201        --WAILVFDLN-YGK---------------LRNIGSDAH NVGLIWN --------- SIAI 155
Q9HM56          --WAILVFDLN-YGK---------------LRNIGSDAH NVGLIWN --------- SIAI 155
CAC42412        SGWVLLTYSPR-DGR---------------LVNQWASDH- AHTLAG --------- GTPV 153
SODM_ALCEU      SGWVLLTYSPR-DGR---------------LVNQWASDH- AHTLAG --------- GTPV 153
```

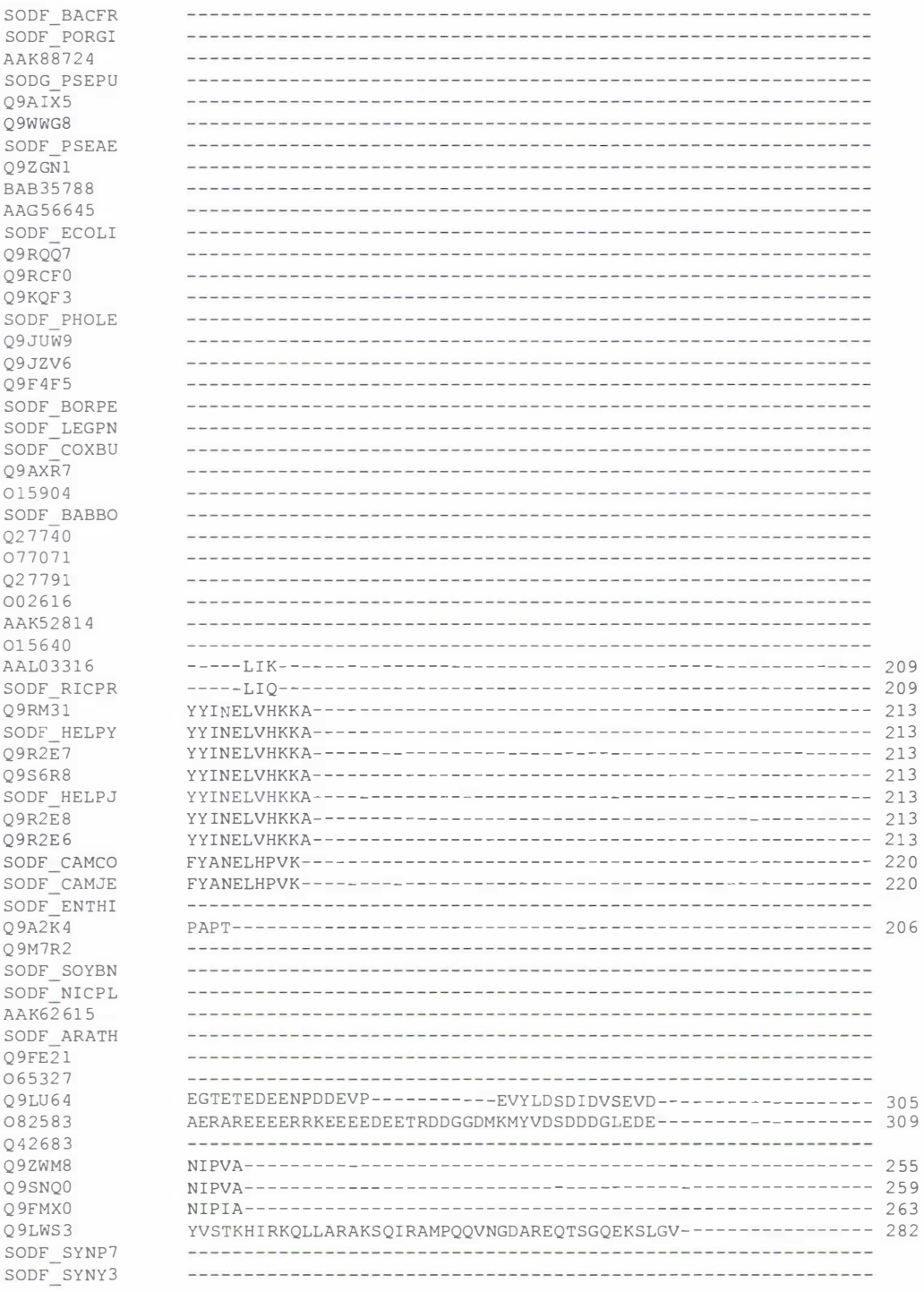
```
AAH10548        SGWGWLGFNKE-QGR ----------LQIAACSNQDPLQGTTG ----------LIPL 179
BAB22095        SGWGWLGFNKE-QGR ----------LQIAACSNQDPLQGTTG ----------LIPL 179
SODM_MOUSE      SGWGWLGFNKE-QGR ----------LQIAACSNQDPLQGTTG ----------LIPL 179
SODM_RABIT      SGWGWLGFNKE-QGH ----------LQIAACANQDPLQGTTG ----------LIPL 160
AAH12423        SGWGWLGFNKE-RGH ----------LQIAACPNQDPLQGTTG ----------LIPL 179
SODM_HUMAN      SGWGWLGFNKE-RGH ----------LQIAACPNQDPLQGTTG ----------LIPL 179
SODM_CAVPO      SGWGWLGFNKE-RGC ----------LQIAACSNQDPLQGTTG ----------LIPL 179
SODM_HORSE      SGWGWLGFNKD-QGR ----------LQIVACPNQDPLQGTTG ----------LIPL 179
AAK97214        SGWGWLGYNKE-QGR ----------LQIAACANQDPLQGTTG ----------LIPL 181
Q9NB66          SGWGWLGYNKQ-KGT ----------LQIATCPNQDPLEATTG ----------LVPL 176
SODM_CHAFE      SGWGWLGY-IA-EGA ----------LQIATCPNQDPLEATTG ----------LVPL 173
SODM_CAEEL      SGWGWLGYCPK-GKI ----------LKVATCANQDPLEATTG ----------LVPL 178
SODN_CAEEL      SGWGWLGYCKK-DKI ----------LKIATCANQDPLEG--- ----------MVPL 175
SODM_ONCVO      SGWGWLAYDKE-MKR ----------LQLACCPNQDLLEPTTG ----------LIPL 180
SODM_DROME      SGWGWLGFNKK-SGK ----------LQLAALPNQDPLEASTG ----------LIPL 171
SODM_CHLMU      SGWGWLAFCPK-KQE ----------LMIQTTANQDPLEATTG ----------MIPL 162
SODM_CHLTR      SGWGWLAFCPQ-KQE ----------LMVQTTANQDPLEATTG ----------MIPL 162
SODM_CHLPN      SGWAWLGFCPA-KQE ----------LVLQATANQDPLEPLTG ----------KLPL 162
Q9M532          SGWVWLGLEKE-SKR ----------LVVETTSNQDPLVTKGP ----------LVPL 192
SODM_HEVBR      SGWVWLALDKE-LKK ----------LVVETTANQDPLVTKGPT ---------LVPL 188
SODM_NICPL      SGWVWLGVDKE-LKR ----------LVIETTANQDPLVSKGAN ---------LVPL 185
Q9SM64          SGWVWLALDKE-LKK ----------LVVETTANQDPLVTKGPT ---------LVPL 185
SODM_CAPAN      SGWVWLGLDKE-LKR ----------LVIETTANQDPLVIKGPN ---------LVPL 185
SODM_PEA        ----WLGLDKD-LKR ----------LVVETT- QDPLVTKGAS ---------LVPL 190
O82584          SGWVWLALDKE-LKK ----------VTVETTANQDPLVTKGLH ---------LVPL 197
Q9SRK3          SGWVWLGLDKE-LKK ----------LVVDTTANQDPLVTKGGS ---------LVPL 188
SODM_ARATH      SGWVWLGLDKE-LKK ----------LVFDTTANQDPLVTKGGS ---------LVPL 188
O65324          SGWVWLGLDKE-LKK ----------LVVDTTANQDPLVTKGGS ---------LVPL 188
Q42672          SGWVWLGLDKE-LKH ----------LVVETTPNQDPLVTKNPN ---------KVPL 184
Q96185          SGWVWLALDKE-AKR ----------LSVETTPNQDPLVTKGSN ----------LHPL 188
O82571          SGWVWLALDKE-AKG ----------LSVETTPNQDPLVTKGSN ---------LHPL 181
P93606          SGWVWLALDKE-AKK ----------LSVETTPNQDPLVTKGSN ---------LYPL 188
Q43121          SGWVWLALDKE-AKK ----------LSVETTANQDPLVTKGAN ---------LVPL 188
SODM_ORYSA      SGWVWLALDKE-AKK ----------LSVETTANQDPLVTKGAN ---------LVPL 188
Q43803          SGWVWLALDKE-AKK ----------LSVETTANQDPLVTKGAN ---------LVPL 188
SODP_MAIZE      SGWVWLALDKE-PKK ----------LSVETTANQDPLVTKGAS ---------LVPL 190
SODN_MAIZE      SGWVWLALDKE-PKK ----------LSVETTANQDPLVTKGAS ---------LVPL 189
SODO_MAIZE      SGWVWLALDKE-AKK ----------VSVETTANQDPLVTKGAS ---------LVPL 190
SODM_MAIZE      SGWVWLALDKE-AKK ----------VSVETTANQDPLVTKGAS ---------LVPL 192
Q43273          SGWVWLALDKE-AKK ----------VSVETTANQDPLVTKGAS ---------LVPL 192
Q9LYK8          SGWVWFGLDRE-LKR ----------LVVETTANQDPLVTKGSH ---------LVPL 193
SODM_CANAL      SGWAFIVKNKA-NGDT----------IDVITTANQDTVTDPN ----------LVPL 194
P79022          SGWAFIVKNKE-NGGA----------LDVITTANQDTVLGP ----------FVPL 186
SODM_YEAST      SGWAFIVKNLS-NGGK----------LDVVQTYNQDTVTGP ----------LVPL 190
Q9P945          SGWAWLVQDKQ-TG-S----------IAIKTYANQDPVVGQ ----------FKPL 184
Q9UQX0          SGWAWLIVD---KDGS----------LRITTTANQDTIVKS ----------KPI 177
O74200          SGWAWLVKSPF-SQR ----------LSIQITMNQDVVTQG ----------KVI 180
Q9P921          SGWCWLVKIPS-SRQ ----------LFIQTTMNQDLVTQG ----------KVI 143
AAK82369        SGWGWLGYNPS-SKR ----------VELATTANQDPLLT ----------HVPL 159
SODM_GANMI      SGWGWLGLNPA-TKR ----------LEITTTANQDPLLS ----------HVPI 157
Q9Y773          SGWGWLGLNPS-TKR ----------LEIATTPNQDPLLT ----------HIPI 159
SODM_AGABI      SGWGWLGVNPS-NKR ----------LEISTTPNQDPLLN ----------LVPI 153
SODM_ASPFU      SGWGWLVTDG--KGK ----------LDITTTHDQDPVTG ----------AAPV 159
SODM_PENCH      SGWGWLVSKGGAKGR ----------LEIVTTKDQDPVNAP ----------DVPV 162
SODM_NEUCR      SGWGWLVKESTAEKGR----------LRIITTKDQDPVVGG ----------EVPV 192
SODF_TETPY      SGWGWLGYDTV--SKS----------LRLFELGNQDMPEWSS ----------IVPL 153
O42919          ACWLWIVIDDYNRLNLLRTFQAG-SPYLWTRWQSNDPHLISSVPDYSARPRKY   AHVPI 224
O74379          DGWLWLVYSPEKS---------- LFSLLCTYN  ASNAFLWGTGFPKFRTN- AIVPL 173
AAK80516        --WVILSIDSLDGR----------FHLFGND  FHDTSAILAS ----------HPL 150
```

```
SODF_BACFR      LGFDV----WEHAYYLDYQNRR-----ADDVNKLWE-IIDWDVVEKRL------------ 193
SODF_PORGI      LGFDV----WEHAYYLTYQNRR-----ADHLKDLWS-IVDWDIVESRY------------ 191
AAK88724        LGIDI----WEHAYYLDYENRK-----PDHIRAVLDKLVNWRVVEERLKA---------- 232
SODG_PSEPU      LTCDV----WEHAYYIDYRNLR-----PKYVEAFWNL-VNWAFVAEQFEGKTFKA----- 197
Q9AIX5          LTCDV----WEHAYYIDYRNLR-----PKYVEAFWNL-VNWAFVAEQFEGKTYKV----- 198
Q9WWG8          LTCDV----WEHAYYIDYRNLR-----PKYVEAFWNL-VNWDFVAKQFAS---------- 193
SODF_PSEAE      LTCDV----WEHAYYIDYRNLR-----PKYVEAFWNL-VNWDFVAKNFAA---------- 192
Q9ZGN1          LTCDV----WEHAYYIDYRNLR-----PKYVEAFWNL-VNWDFVAANYAA---------- 193
BAB35788        LTVDV----WEHAYYIDYRNAR-----PGYLEHFWAL-VNWEFVAKNLAA---------- 193
AAG56645        LTVDV----WEHAYYIDYRNAR-----PGYLEHFWAL-VNWEFVAKNLAA---------- 193
SODF_ECOLI      LTVDV----WEHAYYIDYRNAR-----PGYLEHFWAL-VNWEFVAKNLAA---------- 192
Q9RQQ7          LTVDL----WEHAYYIDYRNVR-----PDYMNGFWAL-VNWDFVAEDRQLS--------- 200
Q9RCF0          LTVDL----WEHAYYIDYRNLR-----PDYMNGFWAL-VNWEFVAREPSAF--------- 195
Q9KQF3          LTVDL----WEHAYYIDYRNVR-----PDYMNGFWAL-VNWDFVAQNLAK---------- 194
SODF_PHOLE      LTVDL----WEHAYYIDYRNLR-----PSYMDGFWAL-VNWDFVSKNLAA---------- 193
Q9JUW9          LTCDV----WEHAYYIDYRNSR-----PNYLKGFWEI-VNWDEVAKRFAALS-------- 195
Q9JZV6          LTCDV----WEHAYYIDYRNSR-----PNYLKGFWEI-VNWDEVAKRFAALS-------- 195
Q9F4F5          LTCDV----WEHAYYIDYRNSR-----PNYLKGFWEI-VNWDEVAKRFAA---------- 193
SODF_BORPE      LTCDV----WEHAYYIDYRNAR-----PKYLENFWAL-VNWEFAAKNFA----------- 192
SODF_LEGPN      LTCDV----WEHAYYIDYRNRR-----PDYIEAFWSL-VNWDFASSNLK----------- 192
SODF_COXBU      MTCDV----WEHAYYIDTRNDR-----PKYVNNFWQV-VNWDFVMKNFKS---------- 193
Q9AXR7          LTCDV----WEHAFYIDYKNDK-----AKYVQSFWGA-VNWDFVASNFPQGELRASKL-- 230
O15904          LCCDV----WEHAYYIDYKNDR-----LSYINSWWNL-VNWDFANKNLEAPFKWS----- 199
SODF_BABBO      LCCDV----WEHAYYIDYKNDR-----LSYINSWWNL-VNWDFANKNLEAPFKWS----- 199
Q27740          LTCDI----WEHAYYIDYRNDR-----ASYVKAWWNL-VNWNFANENLKKAMKK------ 198
O77071          LCCDV----WEHAYYLDRKNDR-----PAYIKAWWNV-VNWDFASKNLENALK------- 201
Q27791          LTCDV----WEHAYYVDYKNDR-----AAYVQTFWNV-VNWKNVERQL------------ 195
O02616          LTCDV----WEHAYYVDYKNDL-----AGYVQAFWNV-VNWKNVERQL------------ 195
AAK52814        LACDV----WEHAYYIDYKNDR-----PAYVQTFWNE-VNWAHAEEQLLKS--------- 198
O15640          LTCDI----WEHAYYIDYRNDR-----ASYVNAFWNM-VDWDFASSQL------------ 195
AAL03316        LACDV----WEHAYYIDYRNKR-----PDYVDIFIKHMINWKFVEDN------------- 206
SODF_RICPR      LACDV----WEHAYYIDYRNKR-----SDYIDIFIRHMINWKFVEDN------------- 206
Q9RM31          LVVDV----WEHAYYIDHKNAR-----PVYLEKFYGH-INWHFVSQCYEWAKKEGLGSVD 202
SODF_HELPY      LVVDV----WEHAYYIDHKNAR-----PVYLEKFYGH-INWHFVSQCYEWAKKEGLGSVD 202
Q9R2E7          LVVDV----WEHAYYIDHKNAR-----PVYLEKFYGH-INWHFVSQCYEWAKKEGLGSVD 202
Q9S6R8          LVVDV----WEHAYYIDHKNAR-----PVYLEKFYGH-INWHFVSQCYEWAKKEGLGSVD 202
SODF_HELPJ      LVVDV----WEHAYYIDHKNAR-----PVYLEKFYGH-VNWHFVSQCYEWAKKEGLGSVD 202
Q9R2E8          LVVDV----WEHAYYIDHKNVR-----PVYLEKFYGH-INWHFVSQCYEWAKKEGLGSVD 202
Q9R2E6          LVVDV----WEHAYYIDHKNVR-----PVYLEKFYGH-INWHFVSQCYEWAKKEGLGSVD 202
SODF_CAMCO      LVVDV----WEHAYYVDHRNAR-----PAYLEKFYAH-INWEFVAKAYEWALKEGMGSVS 210
SODF_CAMJE      LVVDV----WEHAYYVDHRNAR-----PAYLEKFYAH-INWEFVAKAYEWALKEGMGSVS 210
SODF_ENTHI      LTCDV----WEHAYYIDTRNNR-----AAYLEHWWN-VVNWKFVEEQL------------ 190
Q9A2K4          VVADV----WEHAYYLDYQNLR----KGFLEAVFDNLINWTLADSQYAAAKSGAQGWAY 202
Q9M7R2          LTIDV----WEHAYYLDFQNRR-----PDYISVFMDKLVSWDAVSSRLEQAKALSA---- 245
SODF_SOYBN      LTIDV----WEHAYYLDFQNRR-----PDYISVFMDKLVSWDAVSSRLEQAKALITSA-- 248
SODF_NICPL      LTIDV----WEHAYYLDFQNRR-----PDYISIFMEKLVSWEAVSSRLKAATA------- 202
AAK62615        LTIDV----WEHAYYLDFQNRR-----PDYIKTFMTNLVSWEAVSARLEAAKAASA---- 212
SODF_ARATH      LTIDV----WEHAYYLDFQNRR-----PDYIKTFMTNLVSWEAVSARLEAAKAASA---- 234
Q9FE21          LTIDV----WEHAYYLDFQNRR-----PDYIKTFMTNLVSWEAVSARLEAAKAASA---- 212
O65327          LTIDV----WEHAYYLDFQNRR-----PDYIKTFMNNLVSWEAVSSRLEAAKAASA---- 212
Q9LU64          LTIDT----WEHAYYLDFENRR-----AEYINTFMEKLVSWETVSTRLESAIARAVQREQ 275
O82583          LVVDV----WEHAYYLDYQNRR-----PDFVSTFMDKLISWEAASARLEAAMAQAAEAQA 268
Q42683          LTVDV----WEHAYYIDVQNRR-----PDYITTFMEKLINWDAVAQRYARATK------- 234
Q9ZWM8          INLDL----WEHAYYLDYKDDR-----RMYVTNFIDHLVSWDTVTLRMMRAEAFVNLGEP 250
Q9SNQ0          LNIDISTFVFQHAYYLDYKDDR-----RMYVTNFIDHLVSWDTVTLRMMRAEAFVNLGEP 254
Q9FMX0          ICVDV----WEHSYYLDYKNDR-----AKYINTFLNHLVSWNAAMSRMARAEAFVNLGEP 258
Q9LWS3          ---------HSHAYYLDYEDRR-----SDYVSTFLEKLVSWETVESRLKKAVQRAVERDE 239
SODF_SYNP7      LTIDV----WEHAYYLDYQNRR-----PDYISTFVEKLANWDFASANYAAAIA------- 201
SODF_SYNY3      LTMDV----WEHAYYLDYQNRR-----PDYIADFLGKLVNWDFVAANLAAA--------- 198
```

```
O50257          LTIDV----WEHAYYLDYQNKR-----PDYVKAFFEKLVNWDEVATRL---------- 196
Q9RFQ4          LTLDV----WEHAYYIDFRNAR-----PAFIKNFLDNLVNWDFAAENLAKA--------- 200
Q9R6Y6          LTLDV----WEHAYYIDFRNAR-----PAFIKNYLDNLVNWDFAAANYAKA--------- 200
SODF_PLEBO      LTLDV----WEHAYYLDFQNAR-----PAFIKNFLDNLVNWDFVAQNLAA---------- 199
AAK86683        LGVDV----WEHSYYIDYRNLR-----PKYLEAFVDNLINWDYVLERYEAAAK------- 200
Q985K3          LGVDV----WEHSYYIDYRNAR-----PKYLEAFVDSLINWDHVLELYEKAK-------- 199
CAC45545        LGVDV----WEHSYYIDYRNAR-----PKYLEAFVDNLINWDYVLELYEAAAK------- 200
SODF_RHIME      LGVDV----WEHSYYIDYRNAR-----PKYLEAFVDNLINWDYVLELYEAAAK------- 199
AAK88862        LGCDV----WEHSYYIDFRTKR-----PAYVENFLDNLVNWQAVAARL------------ 240
O30970          LGCDV----WEHSYYIDFRNKR-----PAYLTNFLDKLVNWENVASRM------------ 200
O15641          FTADV----GEHAYYKDFENRP-----RDYLKELWQI-VDWEFVCQMYEKATK------- 230
O02615          ---DCAP----SSQWMCGSMRT-----TRTLKTAGRI-T---------------------- 211
Q95051          LCVDT    WEHAWYIDYENRK-----AEYFNKFWDA-CNWEFLEKNLKAAGLI------ 195
Q95052          LCVDT    WEHAWYIDYENRK-----AEYFNKFWGV-VNWNFVEENLKKAKLI------ 195
O43957          LTVDT    WEHAWYIDYENRK-----AEYFKNYWNH-VNWNFVEQNLKAAGL------- 194
O84924          LGLDV----WEHAYYLKYKNVR-----PDYINAFWSV-VNWDKVNEYFAKA--------- 201
SODM_LACLA      LGLDV----WEHAYYLKYHNVR-----PDYIEAFFNL-VNWDKVNELYAKAK-------- 206
AAK99478        LGLDV----WEHAYYVKYRNVR-----PDYIKAFFSV-INWNKVDELYAAAK-------- 201
SODM_STRPN      LGLDV----WEHAYYVKYRNVR-----PDYIKAFFSV-INWNKVDELYAAAK-------- 200
AAK74904        LGLDV----WEHAYYVKYRNVR-----PDYIKAFFSV-INWNKVDELYAAAK-------- 201
Q9AGW1          LGLDV----WEHAYYVKYRNVR-----PDYIKAFFSV-INWNKVNELYAAAK-------- 201
SODM_STRAG      LGLDV----WEHAYYLNYRNVR-----PNYIKAFFEI-INWNKVNELYQAAKA------- 201
SODM_STRPY      LALDV----WEHAYYLNYRNVR-----PNYIKAFFEI-INWKKVSELYQAAK-------- 200
SODM_STRMU      LALDV----WEHAYYLNYRNVR-----PNYIKAFFEV-INWNTVARLYAEALTK------ 202
AAF64074        LGLDV----WEHAYYLKYQNRR-----PEYIAAFWNI-VNWDEVAKRYSEAKAK------ 204
SODM_BACCA      LGLDV----WEHAYYLKYQNRR-----PEYIAAFWNI-VNWDEVAKRYSEAKAK------ 203
Q9LBF6          LGLDV----WEHAYYLKYQNRR-----PEYIAAFWNI-VNWDEVAKRYSEAKAK------ 204
SODM_BACST      LGLDV----WEHAYYLKYQNRR-----PEYIAAFWNV-VNWDEVAKRYSEAKAK------ 203
SODM_BACSU      LGLDV----WEHAYYLNYQNRR-----PDYISAFWNV-VNWDEVARLYSDENNGTNKVLI 209
O86168          LGLDV----WEHAYYLNYQNRR-----PDYISAFWNV-VNWDEVARLYSEAK-------- 202
Q9ZF38          LGLDV----WEHAYYLNYQNRR-----PDYIKAFWNV-VNWDEVAPLYSEAK-------- 202
Q9KD10          LGLDV----WEHAYYLNYQNRR-----PDYISAFWNV-VNWDEVAKRYNEAK-------- 202
SODM_LISMO      LGLDV----WEHAYYLKFQNRR-----PEYIETFWNV-INWDEANKRFDAAK-------- 202
SODM_LISIV      LGLDV----WEHAYYLKFQNRR-----PEYIDTFWNV-INWDERNKRFDAAK-------- 202
BAB57715        LGLDV----WEHAYYLKYQNKR-----PDYIGAFWNV-VNWEKVDELYNATK-------- 199
Q9Z5W5          LGLDV----WEHAYYLKYQNKR-----PDYIGAFWNV-VNWEKVDELYNATK-------- 199
Q9K4V3          LGLDV----WEHAYYLKYQNKR-----PDYISAFWNV-VNWEKVDELYNAAK-------- 199
Q9F326          LGLDV----WEHAYYLKYQNKR-----PDYIDAFWNV-VNWNKVDELYEAATK------- 201
BAB56295        LLFDV----WEHAYYLKYQNKR-----PDYMTAFWNI-VNWKKVDELYQAAK--------- 199
Q99X82          LLFDV----WEHAYYLKYQNKR-----PDYMTAFWNI-VNWKKVDELYQAAK--------- 199
Q9EZZ2          LLFDV----WEHAYYLKYQNKR-----PDYMTAFWNI-VN---------------------- 187
SOD2_PLEBO      MGNDV----WEHAYYLKYQNRR-----PEYLNAWWNV-LNWEEINRRFDAAMSGH----- 206
SOD3_PLEBO      MGNDV----WEHAYYLRYQNRR-----ADYLNNWWNV-VNWNEINRRFQAASQQQARS-- 239
SOD1_PLEBO      LGNDV----WEHAYYLKYQNRR-----AEYLTNWWNV-VNWQAVNQRYAQAQRK------ 248
SODM_PASHA      LVLDV----WEHAYYLNYQNRR-----PDFIKAFWNV-VNWDEAARRFEEKVATCGCAK- 213
Q59133          LALDV----WEHAYYLHYQNRR-----PDYIKAFWNV-VNWDEAARRFEEHASHCGCAK- 213
SODM_HAEDU      FTLDV----WEHAYYLHYQNRR-----PDYIKAFWNV-VNWDEASRRFEEKQAGCGCTK- 212
SODM_PASMU      LGLDV----WEHAYYLKFQNRR-----PDYIKEFWHV-VNWDFVAERFEKKVEHCNCTK- 214
SODM_HAEIN      LGLDV----WEHAYYLKFQNRR-----PDYIKEFWNV-VNWDFVAERFEQKTAHSNCAK- 214
BAB38257        LGLDV----WEHAYYLKFQNRR-----PDYIKEFWNV-VNWDEAAARFAAKK-------- 206
AAG59102        LGLDV----WEHAYYLKFQNRR-----PDYIKEFWNV-VNWDEAAARFAAKK-------- 206
SODM_ECOLI      MGLDV----WEHAYYLKFQNRR-----PDYIKEFWNV-VNWDEAAARFAAKK-------- 205
SODM_SALTY      LGLDV----WEHAYYLKFQNRR-----PDYIKEFWNV-VNWDEAAARFALK---------- 204
SODM_YEREN      VGLDV----WEHAYYLKFQNRR-----PDYIKAFWNV-VNWDEAAARFAQAK-------- 207
Q9RUV2          LGVDV----WEHAYYLNYQNRR-----PDYLAAFWNV-VNWDEVSKRYAAAK-------- 211
SODM_BUCAI      IGLDI----WEHAYYLKYQNRR-----LDYIKSFWNV-VNWEEASNRLQK---------- 203
SODM_XANCP      LGLDV----WEHAYYLKYQNRR-----PEYIGAFYNA-VNWEEVERRYHAAIA------- 203
SODF_METJ       LGLDV----WEHAYYLKYQNRR-----PEYIGAFFNV-VNWDEVSRRYQEALA------- 202
Q9PAA4          LCLDV----WEHAYYLHYQNRR-----ADYVNAFYNV-INWDEVERRYVAAMS------- 203
SODM_PSEPU      FGLDV----WEHAYYLKYQNRR-----PEYIGAFYNV-IDWAEVERRYLEALK------- 203
Q9WWG7          LGLDV----WEHAYYLRYQNRR-----PEYINAFYNV-INWDEVSRRYQAALA------- 203
```

```
SODM_PSEAE      LGLDV----WEHAYYLKYQNRR-----PEYIGAFYNV-IDWREVARRYAQALA------- 203
SODM_BORPE      LGLDV----WEHAYYLQYQNRR-----PEYIEAFYRV-VDWAEVARRYEIALAELGREAA 211
SODM_CHLRE      LGLDV----WEHAYYLKYQNRR-----PEYIAAWWNV-VNWEQVAENYKAAQAGTVPL-- 218
SODM_ACICA      LGLDV----WEHAYYLKYQNRR-----ADYIKAFWSV-VNWNKVNERYNEAIKK------ 228
Q9PC60          LALDV----WEHAYYLKYKYKR-----ADYLNAWWTV-VNWNEVNHLFEVAKKEQHNLNH 242
Q9A7E6          LANDV----WEHAYYLKHQNNR-----GAYLGGWWQV-VNWNKANALFDAAAR------- 238
Q9KW87          LALDV----WEHAYYISYQNRR-----PEYIDAWWNV-VNWRAVEENYARALV------- 202
Q9KW85          LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VDWQVVEQNYAQAIG------- 202
AAK53166        LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 205
Q9KNN7          LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 222
AAK53165        LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 205
AAK53164        LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 205
AAK53167        LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 205
AAK53163        LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 205
AAK53547        LALDV----WEHAYYIRYQNRR-----PEYIDAWWNV-VNWEAVSENYAIALTQAA---- 205
Q9KW86          LALDV----WEHAYYIRYQNRR-----PEYIDAWWSV-VNWEAVSENYAIALTQAA---- 205
SODM_THETH      VGIDV----WEHAYYLKYQNRR-----ADYLQAIWNV-LNWDVAEEFFKKA--------- 203
SODM_THEAQ      VGIDV----WEHAYYLKYQNRR-----ADYLQAIWNV-LNWDVAEEIYKGA--------- 203
SODM_BORBU      LGIDV----WEHAYYLKYQNRR-----IEYVDAFLKA-LNWEEVSKVYNEVIN------- 203
P96201          LTMDV----WEHAYYLKYQNRR-----PEYIDYFFDI-INWKKCEEYYNNR--------- 227
SODF_METTH      LVLDV----WEHAYYIDYRNVR-----PDYVEAFWNI-VNWKEVEKRFED---------- 203
SODF_METTM      MVLDV----WEHAYYIDYRNVR-----PDYVEAFWNI-VNWKEVEKRFDD---------- 200
Q9P9I6          MDLDT----WEHAYYIDYRNDR-----AKFIEAFWNI-VDWEEIDKYFRK---------- 200
SODF_PYRAE      LALDV----WEHAYYLQYKNDR-----GSYVDNWWNV-VNWDDVERRLQK---------- 201
SODF_AERPE      LVIDV----WEHAYYLQYKNDR-----GSYVENWWNV-VNWDDVEKRLEQ---------- 201
SODF_SULSO      LILDE----FEHAYYLQYKNKR-----ADYVNAWWNV-VNWDAAEKKLQK---------- 206
BAB67393        LILDE----FEHAYYLQYKNKR-----ADYVNAWWNL-VNWDEADKKLQK---------- 207
SODF_SULAC      LIVDE----FEHAYYLQYKNKR-----GDYLNAWWNV-VNWDDAEKRLQK---------- 206
SODF_ACIAM      LILDE----FEHAYYLQYKNKR-----ADYVNNWWNL-VNWDFADKKLQQ---------- 207
Q9KCK8          LPLDV----WEHAYYLQYKNER-----KDYIENWWNV-VNWDAVEKRYNA---------- 284
SODF_BACSU      LPLDV----WEHAYYLQYKNDR-----ASYVDHWWNV-VDWREAEKRFEQ---------- 272
BAB62412        LMLDM----WEHAFYLQYKNVK-----ADYVKAVWNV-FNWDDAAARFAA---------- 197
Q9APY3          LMLDM----WEHAFYLQYKNVK-----ADYVKAVWNV-FNWDDAAARFAA---------- 197
Q9F9R1          LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWADVQKRYAA---------- 196
Q9AM00          LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWADVQKRYAA---------- 199
SODM_MYCAV      LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWADVQKRYAA---------- 195
SODM_MYCLP      LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWADVQKRYTA---------- 195
SODM_MYCLE      LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWADVQSRYMA---------- 195
SODM_MYCFO      LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWEDVQNRYAA---------- 195
SODM_MYCSM      LQVDM----WEHAFYLQYKNVK-----ADYVKAFWNV-VNWDDVQNRFAA---------- 195
SODM_NOCAS      LQVDM----WEHAFYLQYKNVK-----ADYVTAFWNV-VNWADVQDRFGK---------- 195
SODF_MYCTU      LLLDM----WEHAFYLQYKNVK-----VDFAKAFWNV-VNWADVQSRYAA---------- 196
SODM_PROFR      LQLDM----WEHAFYLQYKNVK-----GDYVKSWWNV-VNWDDVALRFSE---------- 197
Q9X469          LVFDA----WEHAFYLQYKNQK-----VDFIEAMWAV-VNWQDVAKRHAA---------- 203
SODF_STRCO      LVFDA----WEHAFYLQYKNQK-----VDFIDAMWAV-VNWQDVARRYEA---------- 200
Q9X6N3          LVFDA----WEHAFYLQYKNQK-----VDFIEAMWQV-VNWQDVAKRYAA---------- 201
Q9HQF1          LALDV----WEHSYYYDYGPDR-----GSFVDAFFEV-VDWDEPTERFEQ---------- 194
SOD1_HALCU      LALDV----WEHSYYYDYGPDR-----GSFVDAFFEV-VDWDEPTERFEQ---------- 193
SOD1_HALSG      LALDV----WEHSYYYDYGPDR-----GSFVDAFFEV-VDWDEPTERFEQ---------- 194
Q9HQ47          LALDV----WEHSYYYDYGPDR-----GSFVDAFFEV-IDWDPIAANYDD---------- 194
SODM_HALHA      LALDV----WEHSYYYDYGPDR-----GSFVDAFFEV-IDWDPIAANYDD---------- 194
SOD2_HALSG      LALDV----WEHSYYYDYGPDR-----GSFVDAFFEV-IDWDPIAANYDD---------- 194
SOD2_HALVO      LALDV----WEHSYYHDYGPAR-----GDFVSAFFEV-VDWDEPAARYEQ---------- 193
SOD1_HALVO      LALDV----WEHSYYHDYGPAR-----GDFVSAFFEV-VDWDEPAARYEQ---------- 194
SODM_HALMA      LALDV----WEHSYYYDYGPAR-----GDFIDAFFEV-VDWDKAAEEYEK---------- 197
SODF_AQUAE      IVIDT----YEHAYYVDYKNKR-----PPYIDAFFKN-INWDVVNERFEK---------- 199
SODF_AQUPY      IVIDT----YEHAYYVDYKNKR-----PPYIDAFFKN-INWDVVNERFEK---------- 199
BAB59201        LTMDV----YEHAYYVDYGAKR-----APYLDAFMKN-VNWPVVLKRLEK---------- 195
Q9HM56          LTMDV----YEHAYYVDYGAKR-----APYLDAFLKN-VNWPVVLDRLNR---------- 195
CAC42412        LALDM----YEHSYHMDYGAKA-----AAYVDAFMQN-IHWQRAA--TR---------- 190
SODM_ALCEU      LALDM----YEHSYHMDYGAKA-----AAYVDAFMQN-IHWQRAANLARGCRAGLTEAGT 203
```

```
AAH10548        LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENVTERYTA---------- 219
BAB22095        LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENVTERYTA---------- 219
SODM_MOUSE      LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENVTERYTA---------- 219
SODM_RABIT      LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-ITWENVTERYMA---------- 200
AAH12423        LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENVTERYMA---------- 219
SODM_HUMAN      LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENVTERYMA---------- 219
SODM_CAVPO      LGIDV----WEHAYYLQLKNVR-----PDYLKAIWKV-IKNS----------------- 211
SODM_HORSE      LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENVSERYMA---------- 219
AAK97214        LGIDV----WEHAYYLQYKNVR-----PDYLKAIWNV-INWENASSRYES---------- 221
Q9NB66          FGIDV----WEHAYYLQYKNVR-----PDYVKAIWNV-ANWKDITARFNA---------- 216
SODM_CHAFE      FGIDV----WEHAYYLQYKNVRRIMLKPSGISLIGRI-SLQGSMQQSKNA---------- 218
SODM_CAEEL      FGIDV----WEHAYYLQYKNVR-----PDYVNAIWKI-ANWKNVSERFAK---------- 218
SODN_CAEEL      FGIDV----WEHAYYLQYKNVR-----PDYVHAIWKI-ANWKNISERFAN---------- 215
SODM_ONCVO      FCIDV----WEHAYYLQYKNLR-----PDFVKAIWKI-ANWKIISDRYIK---------- 220
SODM_DROME      FGIDV----WEHAYYLQYKNVR-----PSYVEAIWDI-ANWDDISCRFQE---------- 211
SODM_CHLMU      LGVDV----WEHAYYLQYKNAR-----LDYLKNFPSI-INWDYIESRFVE---------- 202
SODM_CHLTR      LGVDV----WEHAYYLQYKNAR-----MDYLKSFPSI-INWDYIENRFVE---------- 202
SODM_CHLPN      LGVDV----WEHAYYLQYKNVR-----MDYLKAFPQI-INWGHIENRFSE---------- 202
Q9M532          VGIDV----WEHAYYLQYKNVR-----PDYLKNIWKV-VNWKYASDIYAN---------- 232
SODM_HEVBR      LGIDV----WEHAYYLQYKNVR-----PDYLKNIWKV-MNWKYASEVYAK---------- 228
SODM_NICPL      LGIDV----WEHAYYLQYKNVR-----PDYLKNIWKV-MNWKYANEVYEK---------- 225
Q9SM64          LGIDV----WEHAYYLQYKNVR-----PDYLKNIWKV-INWKYASEVYEK---------- 225
SODM_CAPAN      LGIDV----WEHAYYLQYKNVK-----PDYLKNIWKV-INWKYAAEVYEK---------- 225
SODM_PEA        LGIDV----WEHAYYLQYKNVR-----PDYLKNIWKV-INWKHASEVYEK---------- 230
O82584          LGIDV----WEHAYYLQYKNVR-----PDYLKNIWGV-INWKYASEVYEK---------- 237
Q9SRK3          VGIDV----WEHAYYLQYKNVR-----PEYLKNVWKV-INWKYASEVYEK---------- 228
SODM_ARATH      VGIDV----WEHAYYLQYKNVR-----PEYLKNVWKV-INWKYASEVYEK---------- 228
O65324          VGIDV----WEHAYYLQYKNVR-----PDYLKNVWKV-INWKYASEVYEK---------- 228
Q42672          LGIDV----WEHAYYLQYKNVR-----PDYLKNIWKV-VNWKYACEVYQQ---------- 224
Q96185          LGIDV----WEHAYYLQYKNVR-----PDYLTNIWKV-VNWKYAGEEYEK---------- 228
O82571          LGIDV----WEHAYYLQYKNVR-----PDYLTNIWKV-VNWKYAGEEYEK---------- 221
P93606          LGIDV----WEHAYYLQYKNVR-----PDYLTNIWKV-VNWKYAGEEYEK---------- 228
Q43121          LGIDV----WEHAYYLQYKNVR-----PDYLSNIWKV-MNWKYAGEVYEN---------- 228
SODM_ORYSA      LGIDV----WEHAYYLQYKNVR-----PDYLSNIWKV-MNWKYAGEVIEN---------- 228
Q43803          LGIDV----WEHAYYLQYKNVR-----PDYLSNIWKV-MNWKYAGEVYEN---------- 228
SODP_MAIZE      LGIDV----WEHAYYLQYKNVR-----PDYLNNIWKV-MNWKYAGEVYEN---------- 230
SODN_MAIZE      LGIDV----WEHAYYLQYKNVR-----PDYLNNIWKV-MNWKYAGEVYEN---------- 229
SODO_MAIZE      LGIDV----WEHAYYLQYKNVR-----PDYLNNIWKV-MNWKYAGEVYEN---------- 230
SODM_MAIZE      LGIDV----WEHAYYLQYKNVR-----PDYLNNIWKV-MNWKYAGEVYEN---------- 232
Q43273          LGIDV----WEHAYYLQYKNVR-----PDYLNNIWKV-MNWKYAGEVYEN---------- 232
Q9LYK8          IGIDV----WEHAYYPQYKNAR-----AEYLKNIWTV-INWKYAADVFEK---------- 233
SODM_CANAL      IAIDA----WEHAYYLQYQNVK-----ADYFKNLWHV-INWKEAERRFEF---------- 234
P79022          VAIDA----WEHAYYLQYQNVK-----ADYFKAIWNV-INWKEAEKRYLI---------- 226
SODM_YEAST      VAIDA----WEHAYYLQYQNKK-----ADYFKAIWNV-VNWKEASRRFDA---------- 230
Q9P945          LGIDA----WEHAYYLQYQNRK-----AEYFKAIWEV-INWKAVEKRFA----------- 223
Q9UQX0          IGIDA----WEHAYYPQYENRK-----AEYFKAIWNV-INWKEAESRYSN---------- 217
O74200          LGIDA----WEHSYYVQYLNNK-----TKYFENIWNV-INWKVMNQRFEQ---------- 220
Q9P921          LGIDA----WEHAYYIQYFNNK-----VKYFENIWNV-I--------------------- 172
AAK82369        IGVDI----WEHAFYLQYLNVK-----ADYLNAIWNV-INFEEAEKRYLE---------- 199
SODM_GANMI      IGVDI----WEHAFYLQYLNVK-----ADYLAAIWIV-INFKEAERRLIE---------- 197
Q9Y773          IGVDI----WEHAFYLQYLNVK-----VDYLNAIWSV-INFKEAEKRFTE---------- 199
SODM_AGABI      IGVDI----WEHAFYLQYLNVK-----ADYLNAIWSV-INFDEAQRRYVE---------- 193
SODM_ASPFU      FGVDM----WEHAYYLQYLNDK-----ASYAKGIWNV-INWAEAENRYIA---------- 199
SODM_PENCH      FGVDM----WEHAYYLQYLNNK-----AGYVEGIWKI-IHWAEAEKRYTA---------- 202
SODM_NEUCR      FGVDM----WEHAYYLQYLNGK-----AAYVENIWKV-INWKTAEERFQG---------- 232
SODF_TETPY      LTIDV----WEHAYYLDYQNLR-----PKYLTEVWKI-VNWREVEKRYLQ---------- 193
O42919          LNLCL----WNHAYYKDYGLLNR----SRYIDTWFDC-IDWSVIEERLTNSLANSEQSSH 275
O74379          LCVNL----WQYAYLDDYGLNGK----KMYITKWWDM-INWTVVNNRFQATRIESL---- 220
AAK80516        MVLDV----QEHSYFKDFINDK-----GKYVETFINN-LDWKILNDRFKAYIFQNKLRTL 200
                          :
```


SODF_BACFR
SODF_PORGI
AAK88724
SODG_PSEPU
Q9AIX5
Q9WWG8
SODF_PSEAE
Q9ZGN1
BAB35788
AAG56645
SODF_ECOLI
Q9RQQ7
Q9RCF0
Q9KQF3
SODF_PHOLE
Q9JUW9
Q9JZV6
Q9F4F5
SODF_BORPE
SODF_LEGPN
SODF_COXBU
Q9AXR7
O15904
SODF_BABBO
Q27740
O77071
Q27791
O02616
AAK52814
O15640
AAL03316
SODF_RICPR
Q9RM31
SODF_HELPY
Q9R2E7
Q9S6R8
SODF_HELPJ
Q9R2E8
Q9R2E6
SODF_CAMCO
SODF_CAMJE
SODF_ENTHI
Q9A2K4
Q9M7R2
SODF_SOYBN
SODF_NICPL
AAK62615
SODF_ARATH
Q9FE21
O65327
Q9LU64
O82583
Q42683
Q9ZWM8
Q9SNQ0
Q9FMX0
Q9LWS3
SODF_SYNP7
SODF_SYNY3
-----LIK-
-----LIQ-
YYINELVHKKA
YYINELVHKKA
YYINELVHKKA
YYINELVHKKA
YYINELVHKKA
YYINELVHKKA
YYINELVHKKA
FYANELHPVK
FYANELHPVK
PAPT
EGTETEDEENPDDEVP----------EVYLDSDIDVSEVD--
AERAREEEERRKEEEEDEETRDDGGDMKMYVDSDDDGLEDE
NIPVA
NIPVA
NIPIA
YVSTKHIRKQLLARAKSQIRAMPQQVNGDAREQTSGQEKSLGV-
209
209
213
213
213
213
213
213
213
220
220
206
305
309
255
259
263
282

```
O50257          ------------------------------------------------------------
Q9RFQ4          ------------------------------------------------------------
Q9R6Y6          ------------------------------------------------------------
SODF_PLEBO      ------------------------------------------------------------
AAK86683        ------------------------------------------------------------
Q985K3          ------------------------------------------------------------
CAC45545        ------------------------------------------------------------
SODF_RHIME      ------------------------------------------------------------
AAK88862        ------------------------------------------------------------
O30970          ------------------------------------------------------------
O15641          ------------------------------------------------------------
O02615          ------------------------------------------------------------
Q95051          ------------------------------------------------------------
Q95052          ------------------------------------------------------------
O43957          ------------------------------------------------------------
O84924          ------------------------------------------------------------
SODM_LACLA      ------------------------------------------------------------
AAK99478        ------------------------------------------------------------
SODM_STRPN      ------------------------------------------------------------
AAK74904        ------------------------------------------------------------
Q9AGW1          ------------------------------------------------------------
SODM_STRAG      ------------------------------------------------------------
SODM_STRPY      ------------------------------------------------------------
SODM_STRMU      ------------------------------------------------------------
AAF64074        ------------------------------------------------------------
SODM_BACCA      ------------------------------------------------------------
Q9LBF6          ------------------------------------------------------------
SODM_BACST      ------------------------------------------------------------
SODM_BACSU      MGPCFFMSSFYISRFE-----------------------------------------------  225
O86168          ------------------------------------------------------------
Q9ZF38          ------------------------------------------------------------
Q9KD10          ------------------------------------------------------------
SODM_LISMO      ------------------------------------------------------------
SODM_LISIV      ------------------------------------------------------------
BAB57715        ------------------------------------------------------------
Q9Z5W5          ------------------------------------------------------------
Q9K4V3          ------------------------------------------------------------
Q9F326          ------------------------------------------------------------
BAB56295        ------------------------------------------------------------
Q99X82          ------------------------------------------------------------
Q9EZZ2          ------------------------------------------------------------
SOD2_PLEBO      ------------------------------------------------------------
SOD3_PLEBO      ------------------------------------------------------------
SOD1_PLEBO      ------------------------------------------------------------
SODM_PASHA      ------------------------------------------------------------
Q59133          ------------------------------------------------------------
SODM_HAEDU      ------------------------------------------------------------
SODM_PASMU      ------------------------------------------------------------
SODM_HAEIN      ------------------------------------------------------------
BAB38257        ------------------------------------------------------------
AAG59102        ------------------------------------------------------------
SODM_ECOLI      ------------------------------------------------------------
SODM_SALTY      ------------------------------------------------------------
SODM_YEREN      ------------------------------------------------------------
Q9RUV2          ------------------------------------------------------------
SODM_BUCAI      ------------------------------------------------------------
SODM_XANCP      ------------------------------------------------------------
SODF_METJ       ------------------------------------------------------------
Q9PAA4          ------------------------------------------------------------
SODM_PSEPU      ------------------------------------------------------------
Q9WWG7          ------------------------------------------------------------
```

```
SODM_PSEAE      ------------------------------------------------------------
SODM_BORPE      ------------------------------------------------------------
SODM_CHLRE      ------------------------------------------------------------
SODM_ACICA      ------------------------------------------------------------
Q9PC60          ------------------------------------------------------------
Q9A7E6          ------------------------------------------------------------
Q9KW87          ------------------------------------------------------------
Q9KW85          ------------------------------------------------------------
AAK53166        ------------------------------------------------------------
Q9KNN7          ------------------------------------------------------------
AAK53165        ------------------------------------------------------------
AAK53164        ------------------------------------------------------------
AAK53167        ------------------------------------------------------------
AAK53163        ------------------------------------------------------------
AAK53547        ------------------------------------------------------------
Q9KW86          ------------------------------------------------------------
SODM_THETH      ------------------------------------------------------------
SODM_THEAQ      ------------------------------------------------------------
SODM_BORBU      ------------------------------------------------------------
P96201          ------------------------------------------------------------
SODF_METTH      ----------IL---------------------------------------------------- 205
SODF_METTM      ----------LF---------------------------------------------------- 202
Q9P9I6          ----------MRK--------------------------------------------------- 203
SODF_PYRAE      ----------ALNGQIALKL------------------------------------------- 211
SODF_AERPE      ----------ALNNAKPLYLLPQ---------------------------------------- 214
SODF_SULSO      ----------YLTK-------------------------------------------------- 210
BAB67393        ----------YLNK-------------------------------------------------- 211
SODF_SULAC      ----------YLNK-------------------------------------------------- 210
SODF_ACIAM      ----------YMKK-------------------------------------------------- 211
Q9KCK8          ----------AKQLRWTPY--------------------------------------------- 293
SODF_BACSU      ----------AKEVVWKLY--------------------------------------------- 281
BAB62412        ----------ASK--------------------------------------------------- 200
Q9APY3          ----------ASK--------------------------------------------------- 200
Q9F9R1          ----------ATSKAQGLIFG------------------------------------------- 207
Q9AM00          ----------ATSKAQGLIFG------------------------------------------- 210
SODM_MYCAV      ----------ATSKAQGLIFG------------------------------------------- 206
SODM_MYCLP      ----------ATSKTQGLIFG------------------------------------------- 206
SODM_MYCLE      ----------ATSKTQGLIFD------------------------------------------- 206
SODM_MYCFO      ----------ATSKTNGLIFG------------------------------------------- 206
SODM_MYCSM      ----------ATSKTNGLIFG------------------------------------------- 206
SODM_NOCAS      ----------AVNQGKGLIFG------------------------------------------- 206
SODF_MYCTU      ----------ATSQTKGLIFG------------------------------------------- 207
SODM_PROFR      ----------ARVA-------------------------------------------------- 201
Q9X469          ----------AKERGDSLLLKP------------------------------------------ 215
SODF_STRCO      ----------AKSRTNTLLLAP------------------------------------------ 212
Q9X6N3          ----------AKERVNVLLLAP------------------------------------------ 213
Q9HQF1          ----------AAERFE------------------------------------------------ 200
SOD1_HALCU      ----------AAERFE------------------------------------------------ 199
SOD1_HALSG      ----------AAERFE------------------------------------------------ 200
Q9HQ47          ----------VVSLFE------------------------------------------------ 200
SODM_HALHA      ----------VVSLFE------------------------------------------------ 200
SOD2_HALSG      ----------VVSLFE------------------------------------------------ 200
SOD2_HALVO      ----------AVELFE------------------------------------------------ 199
SOD1_HALVO      ----------AVELFE------------------------------------------------ 200
SODM_HALMA      ----------SVSHFE------------------------------------------------ 203
SODF_AQUAE      -------AMKAYEALKDFIK-------------------------------------------- 212
SODF_AQUPY      -------AMKAYEALKDFIK-------------------------------------------- 212
BAB59201        ------AKKAYEAFKS------------------------------------------------ 205
Q9HM56          -------AKKAYEAFKS----------------------------------------------- 205
CAC42412        ----------FAAAVRD----------------------------------------------- 197
SODM_ALCEU      PMPNATSSLELAAALHGLNAPQLIDVRRKPAFDASEQMIAGAAWRNPDELGNWIATLDAN 263
```

```
AAH10548        ----------CKK---------------------------------------------- 222
BAB22095        ----------CKK---------------------------------------------- 222
SODM_MOUSE      ----------CKK---------------------------------------------- 222
SODM_RABIT      ----------CK----------------------------------------------- 202
AAH12423        ----------CKK---------------------------------------------- 222
SODM_HUMAN      ----------CKK---------------------------------------------- 222
SODM_CAVPO      ------------------------------------------------------------
SODM_HORSE      ----------CKK---------------------------------------------- 222
AAK97214        ----------CRK---------------------------------------------- 224
Q9NB66          ----------AK----------------------------------------------- 218
SODM_CHAFE      ----------TSCSGA------------------------------------------- 224
SODM_CAEEL      ----------AQQ---------------------------------------------- 221
SODN_CAEEL      ----------ARQ---------------------------------------------- 218
SODM_ONCVO      ----------ARG---------------------------------------------- 223
SODM_DROME      ----------AKKLGC------------------------------------------- 217
SODM_CHLMU      ----------MSK---------------------------------------------- 205
SODM_CHLTR      ----------MSKQ--------------------------------------------- 206
SODM_CHLPN      ----------IISSK-------------------------------------------- 207
Q9M532          ----------ECPSA-------------------------------------------- 237
SODM_HEVBR      ----------ECPSS-------------------------------------------- 233
SODM_NICPL      ----------ECP---------------------------------------------- 228
Q9SM64          ----------ESP---------------------------------------------- 228
SODM_CAPAN      ----------ECP---------------------------------------------- 228
SODM_PEA        ----------ESS---------------------------------------------- 233
O82584          ----------ESA---------------------------------------------- 240
Q9SRK3          ----------ENN---------------------------------------------- 231
SODM_ARATH      ----------ENN---------------------------------------------- 231
O65324          ----------ECK---------------------------------------------- 231
Q42672          ----------FTPLPASRD---------------------------------------- 233
Q96185          ----------VLA---------------------------------------------- 231
O82571          ----------VLA---------------------------------------------- 224
P93606          ----------VLA---------------------------------------------- 231
Q43121          ----------ATA---------------------------------------------- 231
SODM_ORYSA      ----------ATA---------------------------------------------- 231
Q43803          ----------ATA---------------------------------------------- 231
SODP_MAIZE      ----------VLA---------------------------------------------- 233
SODN_MAIZE      ----------VLA---------------------------------------------- 232
SODO_MAIZE      ----------VLA---------------------------------------------- 233
SODM_MAIZE      ----------VLA---------------------------------------------- 235
Q43273          ----------VLA---------------------------------------------- 235
Q9LYK8          ----------HTRDLDIN----------------------------------------- 241
SODM_CANAL      ------------------------------------------------------------
P79022          ----------N------------------------------------------------ 227
SODM_YEAST      ----------GKI---------------------------------------------- 233
Q9P945          ------------------------------------------------------------
Q9UQX0          ----------R------------------------------------------------ 218
O74200          ------------------------------------------------------------
Q9P921          ------------------------------------------------------------
AAK82369        ----------AVGGA----KL-------------------------------------- 206
SODM_GANMI      ----------ATK---------------------------------------------- 200
Q9Y773          ----------GLSGS----KL-------------------------------------- 206
SODM_AGABI      ----------ATQGS----KL-------------------------------------- 200
SODM_ASPFU      ----------GDKGGHPFMKL-------------------------------------- 210
SODM_PENCH      ----------GVEN---PLKL-------------------------------------- 210
SODM_NEUCR      ----------SREDAFADLKALL------------------------------------ 245
SODF_TETPY      ----------AIE---------------------------------------------- 196
O42919          IS--------------------------------------------------------- 277
O74379          ------------------------------------------------------------
AAK80516        SRSDGDFGVCPFMSMQDTFY--------------------------------------- 220
```

```
SODF_BACFR      ------------------------------------
SODF_PORGI      ------------------------------------
AAK88724        ------------------------------------
SODG_PSEPU      ------------------------------------
Q9AIX5          ------------------------------------
Q9WWG8          ------------------------------------
SODF_PSEAE      ------------------------------------
Q9ZGN1          ------------------------------------
BAB35788        ------------------------------------
AAG56645        ------------------------------------
SODF_ECOLI      ------------------------------------
Q9RQQ7          ------------------------------------
Q9RCF0          ------------------------------------
Q9KQF3          ------------------------------------
SODF_PHOLE      ------------------------------------
Q9JUW9          ------------------------------------
Q9JZV6          ------------------------------------
Q9F4F5          ------------------------------------
SODF_BORPE      ------------------------------------
SODF_LEGPN      ------------------------------------
SODF_COXBU      ------------------------------------
Q9AXR7          ------------------------------------
O15904          ------------------------------------
SODF_BABBO      ------------------------------------
Q27740          ------------------------------------
O77071          ------------------------------------
Q27791          ------------------------------------
O02616          ------------------------------------
AAK52814        ------------------------------------
O15640          ------------------------------------
AAL03316        ------------------------------------
SODF_RICPR      ------------------------------------
Q9RM31          ------------------------------------
SODF_HELPY      ------------------------------------
Q9R2E7          ------------------------------------
Q9S6R8          ------------------------------------
SODF_HELPJ      ------------------------------------
Q9R2E8          ------------------------------------
Q9R2E6          ------------------------------------
SODF_CAMCO      ------------------------------------
SODF_CAMJE      ------------------------------------
SODF_ENTHI      ------------------------------------
Q9A2K4          ------------------------------------
Q9M7R2          ------------------------------------
SODF_SOYBN      ------------------------------------
SODF_NICPL      ------------------------------------
AAK62615        ------------------------------------
SODF_ARATH      ------------------------------------
Q9FE21          ------------------------------------
O65327          ------------------------------------
Q9LU64          ------------------------------------
O82583          ------------------------------------
Q42683          ------------------------------------
Q9ZWM8          ------------------------------------
Q9SNQ0          ------------------------------------
Q9FMX0          ------------------------------------
Q9LWS3          ------------------------------------
SODF_SYNP7      ------------------------------------
SODF_SYNY3      ------------------------------------
```

```
O50257          ------------------------------------
Q9RFQ4          ------------------------------------
Q9R6Y6          ------------------------------------
SODF_PLEBO      ------------------------------------
AAK86683        ------------------------------------
Q985K3          ------------------------------------
CAC45545        ------------------------------------
SODF_RHIME      ------------------------------------
AAK88862        ------------------------------------
O30970          ------------------------------------
O15641          ------------------------------------
O02615          ------------------------------------
Q95051          ------------------------------------
Q95052          ------------------------------------
O43957          ------------------------------------
O84924          ------------------------------------
SODM_LACLA      ------------------------------------
AAK99478        ------------------------------------
SODM_STRPN      ------------------------------------
AAK74904        ------------------------------------
Q9AGW1          ------------------------------------
SODM_STRAG      ------------------------------------
SODM_STRPY      ------------------------------------
SODM_STRMU      ------------------------------------
AAF64074        ------------------------------------
SODM_BACCA      ------------------------------------
Q9LBF6          ------------------------------------
SODM_BACST      ------------------------------------
SODM_BACSU      ------------------------------------
O86168          ------------------------------------
Q9ZF38          ------------------------------------
Q9KD10          ------------------------------------
SODM_LISMO      ------------------------------------
SODM_LISIV      ------------------------------------
BAB57715        ------------------------------------
Q9Z5W5          ------------------------------------
Q9K4V3          ------------------------------------
Q9F326          ------------------------------------
BAB56295        ------------------------------------
Q99X82          ------------------------------------
Q9EZZ2          ------------------------------------
SOD2_PLEBO      ------------------------------------
SOD3_PLEBO      ------------------------------------
SOD1_PLEBO      ------------------------------------
SODM_PASHA      ------------------------------------
Q59133          ------------------------------------
SODM_HAEDU      ------------------------------------
SODM_PASMU      ------------------------------------
SODM_HAEIN      ------------------------------------
BAB38257        ------------------------------------
AAG59102        ------------------------------------
SODM_ECOLI      ------------------------------------
SODM_SALTY      ------------------------------------
SODM_YEREN      ------------------------------------
Q9RUV2          ------------------------------------
SODM_BUCAI      ------------------------------------
SODM_XANCP      ------------------------------------
SODF_METJ       ------------------------------------
Q9PAA4          ------------------------------------
SODM_PSEPU      ------------------------------------
Q9WWG7          ------------------------------------
```

```
SODM_PSEAE      ------------------------------------
SODM_BORPE      ------------------------------------
SODM_CHLRE      ------------------------------------
SODM_ACICA      ------------------------------------
Q9PC60          ------------------------------------
Q9A7E6          ------------------------------------
Q9KW87          ------------------------------------
Q9KW85          ------------------------------------
AAK53166        ------------------------------------
Q9KNN7          ------------------------------------
AAK53165        ------------------------------------
AAK53164        ------------------------------------
AAK53167        ------------------------------------
AAK53163        ------------------------------------
AAK53547        ------------------------------------
Q9KW86          ------------------------------------
SODM_THETH      ------------------------------------
SODM_THEAQ      ------------------------------------
SODM_BORBU      ------------------------------------
P96201          ------------------------------------
SODF_METTH      ------------------------------------
SODF_METTM      ------------------------------------
Q9P9I6          ------------------------------------
SODF_PYRAE      ------------------------------------
SODF_AERPE      ------------------------------------
SODF_SULSO      ------------------------------------
BAB67393        ------------------------------------
SODF_SULAC      ------------------------------------
SODF_ACIAM      ------------------------------------
Q9KCK8          ------------------------------------
SODF_BACSU      ------------------------------------
BAB62412        ------------------------------------
Q9APY3          ------------------------------------
Q9F9R1          ------------------------------------
Q9AM00          ------------------------------------
SODM_MYCAV      ------------------------------------
SODM_MYCLP      ------------------------------------
SODM_MYCLE      ------------------------------------
SODM_MYCFO      ------------------------------------
SODM_MYCSM      ------------------------------------
SODM_NOCAS      ------------------------------------
SODF_MYCTU      ------------------------------------
SODM_PROFR      ------------------------------------
Q9X469          ------------------------------------
SODF_STRCO      ------------------------------------
Q9X6N3          ------------------------------------
Q9HQF1          ------------------------------------
SOD1_HALCU      ------------------------------------
SOD1_HALSG      ------------------------------------
Q9HQ47          ------------------------------------
SODM_HALHA      ------------------------------------
SOD2_HALSG      ------------------------------------
SOD2_HALVO      ------------------------------------
SOD1_HALVO      ------------------------------------
SODM_HALMA      ------------------------------------
SODF_AQUAE      ------------------------------------
SODF_AQUPY      ------------------------------------
BAB59201        ------------------------------------
Q9HM56          ------------------------------------
CAC42412        ------------------------------------
SODM_ALCEU      RPVTVYCVHGHQVSQDCAALLEHLLLALARRFSYGNT 300
```

```
AAH10548        ------------------------------------
BAB22095        ------------------------------------
SODM_MOUSE      ------------------------------------
SODM_RABIT      ------------------------------------
AAH12423        ------------------------------------
SODM_HUMAN      ------------------------------------
SODM_CAVPO      ------------------------------------
SODM_HORSE      ------------------------------------
AAK97214        ------------------------------------
Q9NB66          ------------------------------------
SODM_CHAFE      ------------------------------------
SODM_CAEEL      ------------------------------------
SODN_CAEEL      ------------------------------------
SODM_ONCVO      ------------------------------------
SODM_DROME      ------------------------------------
SODM_CHLMU      ------------------------------------
SODM_CHLTR      ------------------------------------
SODM_CHLPN      ------------------------------------
Q9M532          ------------------------------------
SODM_HEVBR      ------------------------------------
SODM_NICPL      ------------------------------------
Q9SM64          ------------------------------------
SODM_CAPAN      ------------------------------------
SODM_PEA        ------------------------------------
O82584          ------------------------------------
Q9SRK3          ------------------------------------
SODM_ARATH      ------------------------------------
O65324          ------------------------------------
Q42672          ------------------------------------
Q96185          ------------------------------------
O82571          ------------------------------------
P93606          ------------------------------------
Q43121          ------------------------------------
SODM_ORYSA      ------------------------------------
Q43803          ------------------------------------
SODP_MAIZE      ------------------------------------
SODN_MAIZE      ------------------------------------
SODO_MAIZE      ------------------------------------
SODM_MAIZE      ------------------------------------
Q43273          ------------------------------------
Q9LYK8          ------------------------------------
SODM_CANAL      ------------------------------------
P79022          ------------------------------------
SODM_YEAST      ------------------------------------
Q9P945          ------------------------------------
Q9UQX0          ------------------------------------
O74200          ------------------------------------
Q9P921          ------------------------------------
AAK82369        ------------------------------------
SODM_GANMI      ------------------------------------
Q9Y773          ------------------------------------
SODM_AGABI      ------------------------------------
SODM_ASPFU      ------------------------------------
SODM_PENCH      ------------------------------------
SODM_NEUCR      ------------------------------------
SODF_TETPY      ------------------------------------
O42919          ------------------------------------
O74379          ------------------------------------
AAK80516        ------------------------------------
```