RICE UNIVERSITY

X-ray structural determination and biophysical characterization of HemAT, a chemotaxis receptor from *B. subtilis*

by
Wei Zhang

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

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

George N. Phillips, Jr., Professor, Advisor,
Department of Biochemistry & Cell Biology, Rice
University, Department of Biochemistry & Computer
Sciences, University of Wisconsin-Madison

John S. Olson, Professor, Department of
Biochemistry & Cell Biology, Rice University

Yousif Shamoo, Assistant Professor, Chair,
Department of Biochemistry & Cell Biology, Rice
University

Edward Nikonowicz, Associate Professor,
Department of Biochemistry & Cell Biology, Rice
University

Jordan Konisky, Professor, Department of
Biochemistry & Cell Biology, Rice University

Lon J. Wilson, Professor, Department of Chemistry,
Rice University

HOUSTON, TEXAS
AUGUST, 2003
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
ABSTRACT

X-ray structural determination and biophysical characterization of HemAT, a chemotaxis receptor from \textit{B. subtilis}

By

Wei Zhang

The heme-based aerotaxis transducer (HemAT) from \textit{B. subtilis} is a heme-containing protein and functions as an oxygen sensor. It can detect oxygen and transmit the signal generated from oxygen binding to regulatory proteins through its putative methyl-accepting chemotactic domain. Through other components, the signaling information is transferred to motor proteins, which control the direction of rotation of flagella and in turn lead to changes in the swimming behavior of bacteria. There is a great deal of information known about chemotaxis signaling transduction for \textit{Escherichia coli} and \textit{Salmonella typhimurium}. However, the detailed molecular mechanism of chemotaxis of \textit{Bacillus subtilis} is in a sense reversed, because attractant binding to chemotactic receptors strengthens the activity of the downstream histidine kinase, instead of inhibiting reaction in \textit{Escherichia coli} and \textit{Salmonella typhimurium}.

Multiple-wavelength anomalous dispersion (MAD) data were collected from crystals of HemAT using the intrinsic anomalous scatterer, iron, with synchrotron radiation. Three wavelength iron MAD data were collected to 2.8Å resolution. The native data set was collected to 2.15Å resolution. The crystallographic analysis reveals that the crystal belongs to \textit{P2}_{1}2_{1}2_{1} space group
with the cell dimension $a=50.00\AA$, $b=80.12\AA$, $c=85.95\AA$. There are two molecules in one asymmetric unit with 40% solvent content.

I have determined the crystal structures of the HemAT sensor domain in liganded and unliganded forms at resolutions of 2.15\AA{} and 2.7\AA{}. The structures show that the HemAT sensor domain is a dimeric protein with one heme group in each subunit. The structure of liganded form of HemAT sensor domain reveals a more symmetrical organization than that of the unliganded form. Tyrosine70 in one subunit shows distinct conformations in the liganded and unliganded structures. Our study suggests that disruption of HemAT symmetry plays an important role in initiating the chemotaxis signaling transduction pathway. Our kinetic and thermodynamic studies of ligand binding suggest that HemAT may employ negative cooperativity for detecting external ligand in the signal transduction. The sensor domain provides the structural evidence for such a molecular mechanism.
ACKNOWLEDGEMENTS

I would like to take this opportunity to express my appreciation to all the people who have helped me through my years at Rice and at Madison. Without their support, my research would not be completed.

I would like to express my gratitude to my advisor Dr. George Phillips, who opened a new scientific world for me and showed me the beauty of the reciprocal space and the elegance of Ewald sphere, and for sharing his wisdom and expertise and for his constant support. I would like to thank my advisor Dr. John S. Olson. His optimism always gives me more encouragement for real life and his keen scientific perception always helps me take to new directions. Many thanks are to Dr. Bog Stec, who is a professor at University of Texas at El Paso, inspires my interest in science and provides me many scientific guidance and inspirations. He and Dr. Jayashree Soman are always good teachers and friends for me.

I would also like to thank Drs. Yousif Shamoo, Edward Nikonowicz, Jordan Konisky, Lon Wilson as my committee members for their scientific guidance.

In particular, I am also very grateful to my friends at Rice University at Houston: John Bruning, Jeffrey Myers, Siyang Sun and Rafael Couñago. The good time we spent together and the countless bottles of beer always make the graduate student life wonderful and memorable. I would like to thank my friends at University of Wisconsin at Madison, Euiyoung Bae, Jason McCoy, Roman
Aranda, Byung Woo Han and Ryan Bannen in creating a friendly environment. I also like to thank people Drs. Wilson Radding, Craig Bingman, Gary Wessenberg and Ken Johnson for their help.

Finally, none of my work would have been possible without the support and understanding of my family.
Table of Contents

Abstract ii

Acknowledgements iv

Table of Contents vi

List of Figures x

List of Tables xiii

List of Equations xiv

1 Introduction 1

1.1 Background of chemotaxis 1

1.2 Two components system 6

1.3 Aerotaxis and heme-based biological sensors 9

1.3.1 Aerotaxis 9

1.3.2 Heme-based biological sensors 10

1.4 Structure and function of Other chemoreceptors 15

1.4.1 Ligand binding domain of Tar 15

1.4.2 Signaling domain of Tsr 16
1.4.3 Sequence information on chemoreceptor family 19

1.5 Comparisons between E. coli and B. subtilis chemotaxis 21

1.5.1 Protein components 21

1.5.2 Regulatory function and signal transduction 25

1.6 Research scope and goals 26

2 Fundamentals of Protein Crystallography 28

2.1 Introduction of X-ray Crystallography 28

2.2 Phase problem 29

2.3 Methodology development from DM, MIR to MR 32

2.4 MAD method 42

3 Protein Expression and Purification 46

3.1 Expression and purification of wild-type B. subtilis HemAT 46

3.2 Protein production of HemAT sensor domain and its mutants 47

3.2.1 Design of the enzymatic cleavage site 48

3.2.2 Expression and purification of HemAT sensor 49
domain and its mutants

4 Biophysical Characterization of HemAT protein 60
4.1 UV/Visible Absorption spectroscopic measurement 60
4.2 Circular dichroism spectroscopy 63
4.3 Analytical ultracentrifugation 65
4.4 FTIR spectroscopic measurements 68
4.5 Oxygen equilibrium determination 70
4.6 Kinetic study on ligand binding 72
4.6.1 Kinetic experiments 72
4.6.2 Oxygen binding reaction 73
4.6.3 Carbon monoxide binding reaction 75
4.6.4 Summary of kinetics of ligand binding 75

5 Preliminary Crystallographic Analysis 46
5.1 Protein crystallization and crystal manipulations 52
5.1.1 Protein crystal growth of intact HemAT 52
5.1.2 Protein crystal growth of HemAT sensor domain 53
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Crystallographic data collection</td>
<td>55</td>
</tr>
<tr>
<td>5.3</td>
<td>Structure determination and refinement</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>Structure analysis of sensor domain</td>
<td>78</td>
</tr>
<tr>
<td>6.1</td>
<td>Overall structure</td>
<td>78</td>
</tr>
<tr>
<td>6.2</td>
<td>Comparative analysis of HemAT in two ligation states</td>
<td>81</td>
</tr>
<tr>
<td>6.3</td>
<td>Ligand binding site</td>
<td>89</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparisons with other globins</td>
<td>91</td>
</tr>
<tr>
<td>6.5</td>
<td>Dimerization interface</td>
<td>96</td>
</tr>
<tr>
<td>6.5.1</td>
<td>HemAT dimer interface</td>
<td>96</td>
</tr>
<tr>
<td>6.5.2</td>
<td>nsHb and other globin dimer interface</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>Structural determination of intact HemAT protein</td>
<td>102</td>
</tr>
<tr>
<td>7.1</td>
<td>Model of intact HemAT protein</td>
<td>107</td>
</tr>
<tr>
<td>7.2</td>
<td>Diffraction data analysis</td>
<td>111</td>
</tr>
<tr>
<td>7.3</td>
<td>Patterson map of the diffraction data</td>
<td>115</td>
</tr>
<tr>
<td>7.4</td>
<td>MR solution and Recognition of pseudo symmetry</td>
<td>120</td>
</tr>
<tr>
<td>7.5</td>
<td>Summary</td>
<td></td>
</tr>
</tbody>
</table>
8  Structure and function relation 125
8.1  Symmetry breaking in signal transduction 125
8.2  O2, CO ligand binding reactions 130
8.3  Mutation investigation on the role of Tyr70 135
8.4  Summary 140

9  Biological implications 142
9.1  Molecular mechanism of chemotaxis in \textit{B. subtilis} 142
9.2  Biological implications on the strategy in chemotaxis 145

Appendix

References
## List of Figures

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1.1</td>
<td>Positions of the flagella on <em>E. coli</em> during swimming.</td>
<td>4</td>
</tr>
<tr>
<td>FIGURE 1.2</td>
<td>The tracks of a swimming bacterium.</td>
<td>4</td>
</tr>
<tr>
<td>FIGURE 1.3</td>
<td>Schematic representation of bacterial signal transduction network of <em>E. coli</em>.</td>
<td>8</td>
</tr>
<tr>
<td>FIGURE 1.4</td>
<td>Three heme-based sensing proteins and their structural domains</td>
<td>14</td>
</tr>
<tr>
<td>FIGURE 1.5</td>
<td>Sequence alignments of primary sequences of signaling domains of HemAT and <em>E. coli</em> chemoreceptors</td>
<td>20</td>
</tr>
<tr>
<td>FIGURE 1.6</td>
<td>The chemotaxis signaling transduction pathway model for <em>B. subtilis</em> (left) and <em>E. coli</em> (right).</td>
<td>24</td>
</tr>
<tr>
<td>FIGURE 1.7</td>
<td>Schematic representation of bacterial signal transduction pathway of <em>E. coli</em> and <em>B. subtilis</em>.</td>
<td>26</td>
</tr>
<tr>
<td>FIGURE 2.1</td>
<td>Argand diagram of the isomorphous replacement for centric case</td>
<td>38</td>
</tr>
<tr>
<td>FIGURE 2.2</td>
<td>Argand diagram of the isomorphous replacement for acentric case</td>
<td>39</td>
</tr>
<tr>
<td>FIGURE 2.3</td>
<td>Argand diagram of the isomorphous replacement for solving the phase ambiguity and the related equations</td>
<td>42</td>
</tr>
<tr>
<td>FIGURE 2.4</td>
<td>Argand diagram of decomposition of anomalous scattering factor</td>
<td>43</td>
</tr>
<tr>
<td>FIGURE 2.5</td>
<td>Argand diagram of the breakdown of Friedel’s pairs in the MAD</td>
<td>44</td>
</tr>
<tr>
<td>FIGURE 2.6</td>
<td>X-ray absorption plot of Iron using theoretic values</td>
<td>45</td>
</tr>
<tr>
<td>FIGURE 3.1</td>
<td>Coomassie stained SDS_PAGE of intact HemAT and Mass spectroscopic measurement of purified HemAT.</td>
<td>47</td>
</tr>
</tbody>
</table>
FIGURE 3.2  Coomassie stained SDS_PAGE of Factor Xa digestion of the HemAT sensor domain. 51
FIGURE 3.3  SDS_PAGE of purified HemAT sensor domain. 52
FIGURE 4.1  Absorption spectra of purified HemAT and sperm whale myoglobin. 55
FIGURE 4.2  Circular dichroism spectrum of intact HemAT 57
FIGURE 4.3  Determination of sedimentation coefficient 59
FIGURE 4.4  Analytical ultracentrifugation experiments. 62
FIGURE 4.5  FTIR spectra of intact HemAT and the sensor domain 65
FIGURE 4.6  Hill plot of oxygen equilibrium curve of intact HemAT 67
FIGURE 4.7  The correlation of observed rate constants of oxygen association reaction with ligand concentrations. 72
FIGURE 4.8  Oxygen association curve of HemAT 73
FIGURE 4.9  Times curves for O_2 dissociation from HemAT 74
FIGURE 4.10  CO association curve of HemAT 76
FIGURE 4.11  Times curves for CO dissociation from HemAT 77
FIGURE 5.1  A single crystal of intact HemAT. 82
FIGURE 5.2  Crystals of HemAT sensor domain. 84
FIGURE 5.3  Crystals of HemAT sensor domain in met form. 86
FIGURE 6.1  The molecular structure of HemAT sensor domain 91
FIGURE 6.2  DDM plots of the HemAT sensor domain 95
FIGURE 6.3  Ca trace of the G and H helices at dimerization interface. 97
FIGURE 6.4  Superposition of the ligand binding sites of the unliganded and liganded structures. 99
FIGURE 6.5  Structure based sequence alignment of Mb, HemAT 104
and Hb from *Vitreoscilla*

**FIGURE 6.6** Comparison of distal pockets of the liganded HemAT sensor domain and Mb.

**FIGURE 6.7** Ribbons diagrams of six dimeric hemoglobins

**FIGURE 7.1** Ribbon diagrams of intact HemAT model at two orientations.

**FIGURE 7.2** Patterson map calculated with experimental data with space group P4₁

**FIGURE 7.3** Patterson map calculated with one HemAT with space group P4₁2₁2

**FIGURE 7.4** Patterson map calculated with three HemAT molecules with space group P4₁

**FIGURE 7.5** Molecular solution of HemAT structure in space group P4₁ projected onto a-b plane.

**FIGURE 9.1** A model of the four known types of chemoreceptors found in *B. subtilis*.

**FIGURE 9.2** Revised representation of bacterial signal transduction pathway of *B. subtilis*. 
List of Tables

TABLE 3.1 Comparisons of absorption peaks of Mb and HemAT 54
TABLE 4.1 Rate and equilibrium constants for ligand binding to 80
HemAT and other hemoproteins
TABLE 5.1 Statistics of data collection of intact HemAT 82
TABLE 5.2 Statistics of MAD data collection 87
TABLE 5.3 Statistics of data collection and refinement of HemAT 90
sensor domain
TABLE 6.1 Summary of dimeric hemoglobins 113
TABLE 6.2 Detailed comparisons of the compositions of Hb dimer 114
TABLE 7.1 Example of diffraction spots on 0,0,L axis 118
TABLE 7.2 Analysis of Matthews Coefficient and Solvent Content 119
List of Equations

EQUATION 2.1 30
EQUATION 2.2 30
EQUATION 2.3 31
EQUATION 2.4 33
EQUATION 2.5 36
EQUATION 2.6 38
EQUATION 2.7 39
EQUATION 2.8 40
EQUATION 2.9 42
1 INTRODUCTION

1.1 Background on chemotaxis

Sensing the changes in the surrounding environments is essential for living organisms, both prokaryotes and eukaryotes, to survive and grow. Many microorganisms have the ability to migrate toward more favorable environments or away from unfavorable ones. This is called chemotaxis, one of the most important functions to help bacteria find better living conditions.

The investigation of this biological phenomenon will shed light on the understanding of the sensing mechanism in higher organisms, like how human sensory cells recognize smell and taste by following the route from simple to complex system, and from lower microorganisms to higher class of eukaryotes. However, prokaryotes like Escherichia coli are not as simple as people thought. Many findings reveal that an amazingly complex but well coordinated system in the prokaryotes is responsible for this behavior.

Peoples started to realize that simple microorganisms can exhibit random swimming movement almost three hundred years ago immediately after light microscopy was invented by Antony van Leeuwenhoek in the 1700’s (Berg 2000). In the 1900’s, researchers discovered that these motile bacteria not only can swim randomly, but also demonstrate the biased swimming behavior either towards certain chemicals, like sugars, amino acids, etc., or away from other
chemicals as documented in the literature (Adler 1969; Grishanin and Bibikov 1997). The chemical sources that attract organisms are called attractants and those make bacteria swim away from them are called repellents, respectively (Stock and Da Re 2000).

The chemotactic behavior of bacteria draws attention since the advancement of modern instrumentation and especially molecular biology have allowed investigators observe the events visually, and modify the genetic information in the living cells to test proposed hypotheses.

They observed that some motile bacteria have an ability to be drawn toward a place where attractants are present. But the systematic investigation of bacterial chemotaxis in a more quantitative way was pioneered by Julius Adler in the 1960s’ (Adler 1966; Adler 1969). He was trying to find answers for the question that how the bacteria detect the existence of the attractants, which was the first leading question concerning the chemotaxis. The concept of "chemoreceptor" was also first brought up by him (Adler 1966).

His experimental design is relatively simple but elegant, and can yield very quantitative estimations of the attractant concentration that can catch the attention of the bacteria. Briefly, a thin capillary filled with a variety of attractants was inserted into bacterial culture and incubated for a period of time. After that, the bacteria inside the capillary were spread onto a plate and grown overnight. The number of colonies was counted the next day to quantify the influences of an attractant on bacterial migration.
Using this method, Julius Adler examined thoroughly a great number of metabolizable chemical and their related products. With the combination of mutation screening, the results obtained strongly supported the ideas that it is the chemical itself alters the chemotactic behavior of the bacteria and there is a special type of protein that responsible for the sensing the presence of these chemicals. His investigation advanced our understanding of the bacterial chemotaxis to the molecular level.

In order to have an overall picture of bacterial chemotaxis, I also need to give a brief depiction of the locomotive organelle of bacteria. Flagella are long, filamentous organelle used by most of prokaryotes for swimming. A bacterium can either move in a straight line ("smooth swimming") or tumble in random direction ("tumbling") as flagella rotate counterclockwise or clockwise, respectively. A note needed to be added here is that this type swimming behavior and flagella rotation is a proper example for *E. coli*, not for *B. subtilis*. A more detailed explanation and comparison will be provided in the following chapters.

External attractive stimuli can cause bacteria to swim smoothly to a favorable environment, while repellent stimuli can make bacteria have the opposite reaction, resulting in more tumbling (Figure 1.1, 1.2).
**Figure 1.1** Positions of the flagella on *E. coli* during swimming. When the flagella rotate counterclockwise (A), they are drawn together into a single bundle, which acts as a propeller to produce smooth swimming. When the flagella rotate clockwise (B), they fly apart and produce tumbling. (Figures are adapted from (Alberts, Bray et al. 1994))

**Figure 1.2** The tracks of a swimming bacterium. In the absence of a chemotactic signal (A), periods of smooth swimming are interrupted by brief tumbles that randomly change the direction of swimming. Thus runs and tumbles occur in alternating sequence. In the presence of a chemotactic attractant (B), tumbling is partially suppressed whenever the bacterium happens to be swimming toward a higher concentration of the attractant, so that it gradually moves in the direction of the attractant – a biased random walk. (Figures are adapted from (Alberts, Bray et al. 1994))
Bacteria are small, and their overall size is only about a millionth of a meter in their longest dimension (about 1-2 microns) for both *Salmonella typhimurium* and *E. coli*. It seems impossible that the bacteria have such competence to measure concentration differences from one cell end to the other. The experimental evidence shown by Macnab et al. indisputably proved that bacteria can perceive the concentration changes of chemicals present in the surroundings in a temporal mode rather a spatial manner with their ingenious mixing apparatus (Macnab and Koshland 1972). This key observation on the time-dependent swimming behavior also leads to the initial establishment of excitation and adaptation mechanism of the chemotaxis. Following the excitatory response, bacteria can sense and compare temporal changes of external stimuli concentration. They will adapt to the new environment and restore to the prestimulus state if no detectable concentration changes are present (Grishanin and Bibikov 1997).

Later investigations in the research field of bacterial chemotaxis made tremendous progress on the understanding not only this signaling transduction pathway, but also on the detailed structural view of the both sensing complex and the motor machinery. These achievements were made possible by the development of molecular biology and modern techniques in structural determination, including cryo-EM, NMR and X-ray crystallography. The research interests also have shifted from gene sequence and function relationship to structure and function relationship. Despite the fact that these approaches are complementary to each other in terms of elucidating the molecular mechanism
from individual component and/or the whole pathway to the living systems, I am only able to tell a story on one of unique chemotactic protein from *B. subtilis* using different methods in this thesis. It is amazing to see the strategies microorganisms use in evolution to adjust to their enviroment in order to survive and proliferate in nature. By taking the lessons from the world of microorganisms, it would be enlightening to observe and be observed in the relatively large human environment and to apply our understanding of simple chemotaxis to the complex human behavior.

1.2 Two components system

The chemoreceptors do not influence the flagellar rotation directly. Rather they utilize a two-component regulatory system that is ubiquitously used in signal transduction pathway in prokaryotes and eukaryotes (Djordjevic and Stock 1998).

Two component systems are comprised of a histidine kinase protein and a response regulator (Stock, Robinson et al. 2000). A signaling ligand or ligand-bound receptor can interact with the histidine kinase protein to induce its autophosphorylation at a conserved histidine residue. The phosphate on the phosphorylated histidine kinase can then be transferred to an aspartate residue of its specific partner, a response regulator protein. Formation of the phosphoaspartate in the regulator protein will activate downstream events in the signaling pathway, resulting in initiating the transcription of specific genes or
changing bacteria rotational patterns by interacting with the flagellar motor protein FliM (Bunn and Poyton 1996; Stock, Robinson et al. 2000).

The histidine kinase protein and response regulator protein in E. coli and S. typhimurium (CheA and CheY) have been identified and characterized (Stock and Mowbray 1995; Surette, Levit et al. 1996; Levit, Liu et al. 1999). Receptor binding requires assistance from another protein, CheW (Gegner, Graham et al. 1992). When CheY receives the phosphor group from the phosphorylated CheA, it can bind to the flagellar motor-switch protein to maintain the basic clockwise rotation ("tumble"). The attractant bound receptors can lower the binding affinity of the complexes, which subsequently decrease the amount of phosphorylated CheA (Stock, Robinson et al. 2000). The flagella will restore the counterclockwise rotation ("smooth swimming") due to the inability of CheY to bind to it, so bacteria can move in the direction of attractant.

The structure of truncated CheA from T. maritima, which lacks both the amino terminal conserved histidine region and the response regulator CheY binding site, and the structure of the CheY-binding domain of histidine kinase CheA in complex with CheY have recently been determined by crystallographic methods (Welch, Chinardet et al. 1998; Bilwes, Alex et al. 1999). CheA contains separate domains for interacting with receptors, CheW, ATP and the other CheA subunit. There are also other proteins involved in the signal transmission between receptor and flagella. CheW can help CheA and the receptors form a large complex whose affinity is mediated by the ligation state and methylation of the receptor (Bunn and Poyton 1996). CheY is the response regulator protein in the
chemotactic signaling pathway responsible for receptor modification (Gegner, Graham et al. 1992). The site of phosphorylation in CheY (Asp) is a conserved residue. Unphosphorylated CheY has 6 times higher affinity than phosphorylated CheY for CheA, while phosphorylated CheY has higher affinity than unphosphorylated CheY for the flagellar FliM protein that is responsible for driving flagella rotation (Falke, Bass et al. 1997; Welch, Chinardet et al. 1998).

Figure 1.3 Schematic representation of bacterial signal transduction network of E. coli. Nearly all of the protein components have been identified. The histidine kinase, CheA; aspartate kinase, CheY; methyl-transferase, CheR; and the methyl-esterase, CheB. This picture is taken from Stock (Stock 1999).
1.3 Aerotaxis and heme-based biological sensors

1.3.1 Aerotaxis

One of the most widely studied chemotaxis systems is aerotaxis, which is how bacteria move in search of a favorable gas environment. Like many other chemotactic responses, the aerotactic process requires the coordinated function of the similar protein components and passes the signal through a shared two-component pathway to control cellular behavior. The aerotactic receptor can receive external stimuli from the binding of nitrogen, oxygen carbon monoxide or oxygen. In the context of this thesis, aerotaxis is designated as the chemotactic response to oxygen and suggests how bacteria search for a favorable environment with proper oxygen concentration in order to survive and grow.

The importance of seeking optimal concentration of oxygen relies on the fact that oxygen is the terminal electron acceptor of electron transport chain during energy generation (Stock 1997).

The first aerotaxis receptor found in *E. coli* is Aer, which has molecular weight about 56kD with 506 residues with one prosthetic group flavin adenine dinucleotide (FAD). Analysis of the Aer primary sequence showed that Aer has two distinct structural domains, the N terminus PAS like domain and the C terminus MCP domain which has great similarity to the cytoplasmic domain of other chemoreceptors (Bibikov, Biran et al. 1997; Bibikov, Barnes et al. 2000). The function of Aer protein had been investigated and has been shown to act indirectly as an oxygen sensor in *E. coli* by measuring redox potential in the cell
with its prosthetic group FAD (Rebbapragada, Johnson et al. 1997; Bibikov, Barnes et al. 2000).

The second aerotaxis receptor found in \textit{E. coli} is the Dos protein which stands for direct oxygen sensor and was identified from sequence comparisons (Delgado-Nixon, Gonzalez et al. 2000). Dos is also a multiple domain protein with 60\% sequence homologous to FixL heme domain. Both of \textit{E. coli} oxygen sensing proteins belong to the same structural category, though with different prosthetic groups. The detailed studies on these proteins are being carried out by several labs (Repik, Rebbapragada et al. 2000; Park, Suquet et al. 2002).

Unlike oxygen sensing proteins in \textit{E. coli}, HemAT is found in \textit{B. subtilis} and is a 432 amino acids protein with two structural domains. The C terminus has very high sequence similarity to the cytoplasmic domain of chemoreceptors, but the N terminus has myoglobin like fold (Hou, Larsen et al. 2000). Its homologous protein in \textit{Halobacterium salinarium} is the first myoglobin-like protein found in Archaea kingdom, which may be the evolutionary origin of myoglobin, a well-known oxygen transport and storage protein (Hou, Larsen et al. 2000; Hou, Freitas et al. 2001).

In terms of the primary structure, the different sensing domains of the aerotactic proteins from two species determine their distinct molecular functions at the receptor level for detecting the presence of oxygen (Garrity and Ordal 1995; Grishanin and Bibikov 1997).
1.3.2 Heme-based biological sensors

The number of heme-based sensing proteins has been dramatically increased recently, especially the sensing protein for gaseous molecules, like O2, CO, and NO, which can coordinate to the heme group to play a regulatory role (Rodgers 1999; Chan 2001).

The functions of these heme-based sensing proteins are to monitor the availability of corresponding gaseous molecules, through a hemoglobin-like allosteric mechanism to regulate cellular behavior, like transcriptional initiation, nitrogen fixation, and cellulose synthesis (Gilles-Gonzalez, Ditta et al. 1991; Shelver, Kerby et al. 1997; Chang, Tuckerman et al. 2001). Nearly all of these sensing proteins have multiple domains, based on the amino acids sequence comparisons. The heme-containing domain is presumed to be the sensing domain and can coordinate the gaseous ligands on the prosthetic group heme, which in turn will trigger conformational changes. Structural information generated from the ligand binding will transfer to the signaling domain, which can either initiate transcription or regulate the enzymatic activity.

Three heme-based sensing proteins have been identified, and the schematic representation of their domain organization is shown in Figure 1.4.

The FixL protein and its transcriptional activator FixJ from *Bradyrhizobium japonicum* and *Rhizobium meliloti* play important roles in the regulations of nitrogen fixation (Gilles-Gonzalez, Ditta et al. 1991; Gilles-Gonzalez and Gonzalez 1993; Gilles-Gonzalez, Gonzalez et al. 1994). Under conditions of low concentration of oxygen like anaerobic root nodules, FixL
protein can autophosphorylate itself and is able to transfer a phosphoryl group to its partner FixJ to initiate the transcription of other nitrogen fixation genes. When in the high concentration of oxygen, the FixL protein is in a ligand bound inactive form (Gilles-Gonzalez, Gonzalez et al. 1994). The crystal structure of the sensing domain of FixL has been determined with different ligand bound and revealed that it has the PAS fold (Gong, Hao et al. 1998; Gong, Hao et al. 2000; Hao, Isaza et al. 2002). A molecular mechanism for oxygen sensing has been proposed: The formation and loss of salt bridge between Arg220 and propionate 7 plays important role during the transition from active to inactive states (Chan 2001).

The CooA from photosynthetic bacterium *Rhodospirillum rubrum* itself is a transcription factor and has two structural domains, the N terminus heme-containing CO sensing domain and the C terminus DNA binding domain (Shelver, Kerby et al. 1997). Associated with the uniqueness of *Rhodospirillum rubrum* bacterium, which can take CO as its energy source by oxidation to CO₂, it presents an even more interesting aspect of the heme-based sensing protein (Kerby, Ludden et al. 1995). The kinetic and spectroscopic data showed that DNA binding to CooA protein can have direct influence on the ligand binding behavior and support the functional role of CooA (Uchida, Ishikawa et al. 1998).

The crystal structure of CooA without its DNA and effectors has been determined to 2.6 Å (Lanzilotta, Schuller et al. 2000). With comparison to its homologous protein in active form a plausible model has been proposed to elucidate the allosteric mechanism of the CO dependent activation of CooA. In
the unliganded states, His77 and Pro2 are the coordinating residues. When CO is bound to CooA protein, it will take the position of Pro2. However, in the ferric states, the His77 is displaced with Cys75 residues (Uchida, Ishikawa et al. 2000). Detailed studies are still under way to fully comprehend the complex behavior of CooA activation.

The soluble guanylate cyclase (sGC) is a heterodimeric protein with NO as its ligand (Zhao, Brandish et al. 1999). NO is a signaling molecule produced by NO synthase (NOS) and plays critical roles in the regulation of vascular relaxation and neurotransmission, etc (Marletta 1994). As shown in Figure 1.4, soluble guanylate cyclase is composed of one α subunit with 691 residues and one β subunit with 619 residues which has the heme containing domain using His105 as coordinating residue (Zhao and Marletta 1997). The disruption of His105 and heme group by NO ligand binding is believed to be responsible for the activation of this protein (Dierks, Hu et al. 1997).

In summary, the discovery of more and more heme-based sensing proteins will lead us to have more knowledge on how hemes play roles in the regulation of the protein behavior and the cellular activity. The investigations of the conformational changes induced by these ligands at the structural level will reveal new insights into the molecular mechanism of the heme proteins.
**Heme O₂ sensors**

- **Bj FixL**
  - PAS
  - Histidine Kinase
  - 155 to 258 to 505

- **Ec Dos**
  - PAS
  - Phosphodiesterase
  - 138 to 336 to 807

- **Bs HemAT**
  - Myoglobin fold
  - Tsr structure
  - 178 to 432

**Heme CO sensors**

- **Rr CooA**
  - CooA heme
  - DNA BD
  - 135 to 222

**Heme NO sensors**

- **sGCα**
  - Guanylate cyclase
  - 468 to 691

- **sGCβ**
  - sGC heme
  - Guanylate cyclase
  - 385 to 619

**Figure 1.4** Three heme-based sensing proteins and their structural domains based on their amino acid sequences and homologous structures. Adopted from (Chan 2001) with modifications.
1.4 Structure and function of Other chemoreceptors

1.4.1 Ligand binding domain of Tar

The Aspartate receptor (Tar) of *E. coli* is one of the most characterized chemoreceptors. There are two types of chemoreceptors present in the *E. coli* cell, according to the number of copies of each protein, high abundance receptors, including Tar and serine receptor (Tsr), having about 10 times more than the low abundance receptors including Trg and Tap receptor. The low abundance receptors require the presence of the high abundance receptors to perform proper taxis functions due to the lack of the important protein interactions sites to methyltransferase, which are only found in high-abundance receptors (Feng, Baumgartner et al. 1997; Barnakov, Barnakova et al. 1998 ; Feng, Lilly et al. 1999).

Tar receptor is a homodimeric transmembrane protein with the periplasmic sensing domain, cytoplasmic signaling domain and transmembrane regions. Tar protein not only can detect the aspartate molecule by direct binding but also can mediates response to maltose molecule by indirectly help of maltose binding protein (Gardina, Conway et al. 1992).

The aspartate receptor from *E. coli* and *S. typhimurium* exhibits different ligand binding behavior. The aspartate ligand binding to *E. coli* protein shows a stoichiochemistry of one molecule per dimer. However the binding to the *S. typhimurium* protein has two binding affinities sites (Borkovich, Alex et al. 1992; Milligan and Koshland 1993).
Also structural investigations of the ligand binding domain of aspartate receptor revealed that the unliganded structure displays perfect two fold symmetry with two ligand binding sites. However, with only one aspartate molecule bound in the protein-ligand complex structure the second binding site is disrupted (Milburn, Prive et al. 1991). Kinetic and thermodynamic investigations have shown that when *S. typhimurium* protein binds the first aspartate molecule, the affinity for the second ligand binding decreases. The Hill coefficient is in the range 0.6 ~ 0.8, which is called negative cooperativity. As for *E. coli* protein, these studies shows that only one aspartate molecule can bind to the receptor, which is an extreme case of negative cooperativity, “half-of-the-sites reactivity”. It is found that serine receptor (Tsr) also exhibits negative cooperativity for serine ligand binding (Lin, Li et al. 1994).

So it firmly established based on the structural, biochemical data that the negative cooperativity is an intrinsic feature of these chemoreceptors. The advantage of this structural mechanism is that it can allow the bacteria to be very robust when detecting a wide range of ligand concentrations.

### 1.4.2 Signaling domain of Tsr

The serine receptor (Tsr) was first identified through mutation screening (Hedblom and Adler 1980). Later on the nucleotide sequence information was made available and its amino acid sequence comparison reveals a great deal of
similarity to both the sensing domain and signaling domain of Tar (Boyd, Kendall et al. 1983).

The bacterial chemotaxis process is made of two processes, excitation and adaptation. After the ligand binding to the chemoreceptor, the receptor will presumably experience conformational changes, which are responsible for relaying the structural information to the signaling domain, then to the downstream signaling cascade. The structural changes can be reversed or enhanced by the modification of certain residues on the signaling domain. Eventually, the receptor will restore its sensitivity to the pre-stimulation state, even though there is higher concentration of ligands present in the surroundings. This adaptation is largely due to the methylation and demethylation of glutamate or glutamine residues by methyl-transferase or methyl-esterase. The modifications can attenuate the electrostatic interactions between the helices to control the extent of the ligand-binding caused conformational changes (Surette and Stock 1996). The identification of these residues important for adaptation on Tsr not only provides the exact positions of these residues but also gives great help on the structure determination of the cytoplasmic domain of Tsr (see below) (Rice and Dahlquist 1991).

The structure of cytoplasmic domain of Tsr (residue 286 to 526) has been determined to 2.6Å resolution with the modifications of the methylation sites (from wild-type QEQE to QQQQ (Q-mutant)). The structure revealed features unique to the protein (Kim, Yokota et al. 1999). It is a homodimeric protein with a very elongated shape of about 200Å long, mainly made of a four helical bundle
of coiled coils. These two anti-parallel helices are linked by a 180° U-turn. The methylation sites are very close to each other and exposed to solvent. Notably, residues in this region are not well resolved due to high temperature factors. This result suggests that the methylation regions are more mobile than the rest of the protein. The highly conserved domain is adjacent to the loop, and residues within this region have very low temperature factors.

Another interesting feature of the structure is that the molecule forms a trimer of dimers around the three-fold crystallographic symmetry axis. There are extensive molecular interactions between each helix at the lower end of the structure, which is presumably the protein interaction site for histidine kinase and/or the helper protein.

Since there are greater sequence similarities among the cytoplasmic domains of chemoreceptor, as well as the characteristic seven residue repeat which represent the coiled coil secondary structure, the structure of cytoplasmic domain of Tsr can be regarded as the representative of the structure family (Le Moual and Koshland 1996).

The intact *E. coli* Tsr receptor is a dimer made of three regions, the periplasmic sensing domain, transmembrane regions and signaling domain. The length of the whole molecule is about 380Å. The length of the cytoplasmic domain extending from the inner membrane is about 260Å (Kim, Yokota et al. 1999).
1.4.3 Sequence information on chemoreceptor family

The cytoplasmic domain of HemAT (residues 198 to 432) shares 30% sequence identity to E. coli chemoreceptor Tsr (Serine receptor), Tar (Aspartate receptor) and Aer (E. coli aerotactic receptor) (Figure 1.5). Alignment of C-terminal domain sequences identifies the methylation regions, K1 and R1, as well as the highly conserved domain. From sequence analysis, B. subtilis HemAT and H. salinarum HemAT have fewer methylation residues in comparison to other bacterial chemoreceptors.

The highly conserved domain of B. subtilis HemAT shows greater sequence identity to E. coli chemoreceptors than other regions in the protein. This sequence evidence indicates that HemAT and E. coli chemoreceptors may adapt to the similar three-dimensional architecture with protein kinase CheA and helper protein CheW. Mutational studies in Tar confirmed that some residues in this domain are crucial for interaction with the kinase CheA and/or the helper protein CheW, and also suggested that it provides a surface for docking CheA and/or CheW (Simon Shimizu, Le Novere et al. 2000). Nearly 60% of the residues at this region are hydrophobic residues, suggesting that this domain may be involved in the protein-protein interactions. Therefore, this region is probably the domain responsible for transmitting external signals detected by the chemoreceptor to the two-component system involved in chemotaxis.
| HemAT.Bs  | 198  | LHQE  | TSGSIA  | N  | SETT | R | S | VQELVDEG | S | GSKA | CT | VTS | T  | S | EK  |
| HemAT.Hs  | 222  | LEAS  | DVAERTD | R | M  | T  | D | GVRMAD  | R | SKIE | S | SASS | R | R | W  |
| Tsr       | 248  | LRTD  | GDVNGA  | N | GSA  | A | T | GNNDLSSRE  | E | QAAS | A | ETA | A  |
| Aer       | 253  | CRMY  | DVSQQV  | S | RNGE | ET | A | EGTDA  | E | NHTCG  | D | VNQ  | Q | TVAT  |
| Tar       | 261  | LDTD  | THVQESDA | N | ATGTRAG | AAGNTDL  | S  | ERTTQ  | A | GSA | A | ETA | A  |

**Figure 1.5** Sequence alignments of primary sequences of signaling domains of HemAT and *E. coli* chemoreceptors, Tsr, Tar and Aer. Putative methylation sites are boxed at K1 and R1 regions. HCD represents the highly conserved domain that interacts with CheA and CheW.
1.5 Comparisons between *E. coli* and *B. subtilis* chemotaxis

1.5.1 Protein components

Most chemoreceptors are methyl accepting chemotaxis proteins (MCP) that are methylated or demethylated during excitation and adaptation. This system is referred to as a two-component regulatory system in the signal transduction pathway of prokaryotes (Djordjevic and Stock 1998).

Two component systems are comprised of a histidine kinase protein and a response regulator (Stock, Robinson et al. 2000). The histidine kinase protein can be induced to autophosphorylate itself. The phosphate then can be transferred to response regulator protein. Formation of the phosphoaspartate in the regulator protein will regulate downstream events in the signaling pathway, resulting in changing flagellar rotational behaviors (Bunn and Poyton 1996; Stock, Robinson et al. 2000).

For *E. coli* and *S. typhimurium*, the histidine kinase protein and response regulator proteins, CheA and CheY, have been identified and characterized (Stock and Mowbray 1995; Surette, Levit et al. 1996; Levit, Liu et al. 1999). Receptor binding requires assistance of CheW protein (Gegner, Graham et al. 1992). Much biochemical evidence shows that CheA and CheW can form 2:2:2 complexes with all known MCPs (Falke, Bass et al. 1997; Mowbray and Sandgren 1998). Bacteria maintain a basal level of such large complexes in the cytoplasm, since it is essential for the autophosphorylation activity of CheA. When CheY receives the phosphoryl group from the phosphorylated CheA, it can
bind to the flagellar motor-switch protein to maintain the basic clockwise rotation ("tumble") (Stock, Robinson et al. 2000).

CheW can help CheA and the receptors form a large complex whose affinity is mediated by the ligation state and methylation of the receptor (Bunn and Poyton 1996). CheY is the response regulator protein in the chemotactic signaling pathway responsible for receptor modification (Gegner, Graham et al. 1992). Unphosphorylated CheY has a 6 times higher affinity than phosphorylated CheY for CheA, while phosphorylated CheY has higher affinity than unphosphorylated CheY for the flagellar FliM protein (Falke, Bass et al. 1997; Welch, Chinardet et al. 1998).

Receptor methylation also plays a crucial role in the regulation of signal transmission. CheR is a receptor methyltransferase that can catalyze the methylation reaction of conserved glutamate residues in receptors with S-adenosylmethionine (Bischoff and Ordal 1992; West, Martinez-Hackert et al. 1995). There is a conserved peptide sequence in the C-terminus of most of receptors (Asn-Trp-Glu-Thr-Phe), which is necessary for CheR to associate with the receptor (Le Mouyal and Koshland 1996; Fabret, Feher et al. 1999). After the histidine kinase protein CheA is autophosphorylated during ligand binding, phosphorylated CheA transfers the phosphoryl group not only to its regulatory partner CheY, but also to the methylesterase CheB (Bilwes, Alex et al. 1999). The methylation state at the specific site in the receptor is balanced by CheR and phosphorylated CheB and can also alter receptor conformation to affect the activity of the histidine kinase protein CheA (Stock 1997).
As for *B. subtilis*, a more complicated system has been described. Recent studies suggest that *B. subtilis* still retains the chemotaxis framework of the enteric bacteria, although some aspects are unique. The homolog of CheA, CheW, CheY, CheR, CheB have been identified in *B. subtilis* (Bischoff and Ordal 1992; Fabret, Feher et al. 1999). Experimental evidence demonstrates that these proteins function in the chemotaxis pathway of *B. subtilis* (Fuhrer and Ordal 1991; Bischoff, Bourret et al. 1993; Garrity and Ordal 1995). The genes from *B. subtilis* can complement the behavior of null mutants in *E. coli*, and vice versa. Further, methylesterase CheB and methyltransferase CheR from one bacterium have catalytic activity in both organisms (Garrity and Ordal 1995).

However, there are two chemotaxis-related proteins, CheC and CheD, in *B. subtilis* that do not have orthologs in *E. coli* (Rosario, Kirby et al. 1995). Researchers propose that CheC and CheD regulate formation of the CheW:CheA:receptor complex in *B. subtilis*, however, their roles are still under investigation.

There are great similarities between *B. subtilis* and *H. halobium*. They seem to have comparable mechanisms for chemotaxis (Olsen, Woese et al. 1994; Fabret, Feher et al. 1999). Later in evolution these complex signaling pathways may have become simplified and optimized in *E. coli*. HemAT is only present in *B. subtilis* and *H. halobium* (Zhang, Brooun et al. 1996; Hou, Larsen et al. 2000), with no homology in *E. coli*. The aerotactic transducer in *E. coli* (Aer) was recently discovered as a FAD containing protein (Bibikov, Biran et al. 1997; Rebbapragada, Johnson et al. 1997; Bibikov, Barnes et al. 2000). Elucidation of
the function of both this FAD protein and HemAT may provide more insight on aerotaxis in bacteria.

\[ \text{Figure 1.6} \] The chemotaxis signaling transduction pathway model for *B. subtilis* (left) and *E. coli* (right). A, CheA; B, CheB; C, CheC; D, CheD; R, CheR; V, CheV; W, CheW; Y, CheY; K, FliM; G, FilG. FilM, FilG, FilN/FilY are flagellar proteins. CH3 is the methyl group attached to glutamates of receptors. CH3OH represents methanol released by methylesterase CheB. CheZ is found only in *E. coli* so far, CheV and CheD are in *B. subtilis* only. Modified from (Fabret, Feher et al. 1999).
1.5.2 Regulatory function and signal transduction

The most prominent difference between the two organisms is their movement in the absence of attractants. The default behavior of *E. coli* is smooth swimming, which is altered by phosphorylated CheY. The default swimming behavior of *B. subtilis* is tumbling, which is also altered by phosphorylated CheY (Bischoff, Bourret et al. 1993; Ottemann and Koshland 1997). Thus, addition of attractants increases phosphorylated CheY levels in *B. subtilis* and lowers phosphorylated CheY levels in *E. coli*.

From the evolutionary point of view, the ancestors of *E. coli* might have adjusted their chemotactic behavior at some point in the pathway reversing the final function of CheY. It appears to be favorable for bacteria to move smoothly most of the time, since bacteria need to change the direction quickly in the presence of repellents, especially under unfavorable conditions, rather than tumble until death (Bischoff, Bourret et al. 1993; Garrity and Ordal 1995).
Figure 1.7 Schematic representation of bacterial signal transduction pathway of *E. coli* and *B. subtilis*. The default behavior of *E. coli* is smooth swimming, which is regulated by phosphorylated CheY. The default swimming-behavior of *B. subtilis* is tumbling, which is also regulated by the level of phosphorylated CheY.

1.6 Research scope and goals

HemAT (heme-based aerotactic transducer) is a novel soluble heme containing protein that has been recently identified from *B. subtilis*. It functions as an aerotactic transducer that can sense oxygen and transmit the signal generated from ligand binding to regulatory proteins through its putative methyl-accepting chemotactic domain. Since to-date most of the chemoreceptors identified in bacteria are trans-membrane proteins and due to the difficulty in crystallizing intact membrane proteins, the available structures of chemotactic
receptors are limited only to the periplasmic domain. Thus it would be desirable to obtain the complete structure of an aerotactic transducer at the atomic level.

Structural characterization of HemAT would provide explicit structural information on the ligand-induced conformational changes of intact receptor, and their relationships to other regulatory signaling components. These structural features make HemAT an ideal model to study signal sensing and transmission in the chemotactic pathway of bacteria.

The goal of my thesis is to describe the HemAT structure using crystallographic methods, obtain different ligand-bound structures, and analyze their structural differences. Thus, I hope to get a detailed understanding of ligand induced conformational changes in the chemoreceptor and determine its properties and interactions with other regulatory proteins. Such structural differences will shed light on the mechanism of ligand-induced signal transduction in the chemotactic signaling pathway. Moreover, such structural knowledge can also be applied to this trans-membrane chemoreceptor family whose crystal structures are needed but not readily obtained.
2 FUNDAMENTALS OF PROTEIN CRYSTALLOGRAPHY

2.1 Introduction of X-ray Crystallography

Protein X-ray crystallography can be viewed as a mature science at present, with modern progress in this field including automation of crystal manipulation, data collection using synchrotron radiation, diffraction data processing and the automatic map generation and model building. All of these events lead to the industrial-scale systemic protein structure determination, structural genomics. The physics of X-ray diffraction have been known at beginning of 20th century. The development of this field comes along with the progress of the human understanding route from simple to complex, from two atom organic structure to the mega Dalton ribosome complex. In this chapter, I will only provide some fundamentals of crystallography and the motivating highlights during this scientific exploration based on my understanding of this subject.

The foundation of the modern X-ray crystallography can be traced back to the time when Röntgen discovered X-ray radiation, which leads to many subsequent scientific discoveries. The milestone event of the practical exploitation of X-ray began when people found that organic salt crystals can diffract X-rays. The successful interpretation of the X-ray diffraction pattern of table salt, sodium chloride, by Lawrence Bragg marked the new era of crystallography. Though more and more small organic compounds have been
solved this way to show their three dimensional atomic arrangement, the methodology was largely trial and error instead of systemic knowledge. Like many other scientific advancements, progress in crystallography has taken a long period of time to reach maturity. Along with brilliant technical innovations and theoretical progress, the application of Fourier analysis or the Patterson synthesis have made dramatic improvements that allowed large molecules to be determined, the number of atoms in the solved structures has increased with time.

A landmark of research in protein crystallography was performed in 1934, when Dorothy Hodgkin obtained the first diffraction image of pepsin crystals. The excitement of observing the diffraction spots from a protein crystal and the vanishing of anxiety after rechecking the realness these spots the next morning still leaves remarkable and vivid impression in my mind. Her achievements announced that protein molecules have unique and regular three-dimensional atomic structures and similar approaches could be used to investigate other protein structures. The story of the successful structure determinations of myoglobin and hemoglobin by Max Perutz and John Kendrew generated even greater excitement in the protein crystallography field and opened a new window to see the atomic world, after overcoming tremendous difficulties. Their innovative methods are still in use at the moment.

2.2 Phase problem

X-ray diffraction is a physics phenomenon, and it happens when X-ray beams impinge on a crystal having regular and repeated identical units in three-
dimensional space. X-rays are diffracted by the crystal and will produce special arrangement of spots due to an interference effect, called a diffraction pattern. These diffraction spots vary in their intensities and positions for different crystal systems. They contain information about the atomic arrangement of the crystals and can be used to trace back to its structure.

\[ F_{(h,k,l)} = \sum_{i=1}^{N} f_{(i)} \exp[2\pi \cdot i \cdot (hx_{(i)} + ky_{(i)} + lz_{(i)})] \quad \text{(equation 2.1)} \]

Equation 2.1 is the mathematic formula to compute the diffraction structure factor. The left side of the equation \((F_{(h,k,l)})\) represents the amplitude of the diffraction spots. The \(f_{(i)}\) term is the atomic scattering factor which is dependent of the number of electrons within each atom, and the diffracted beam angle, etc. This equation has two implications, first is that the position of each individual atom determines the final magnitude of the diffraction spots and second, is that each diffraction spot is the summation of all of the contributions to structure factors from the atoms within the whole unit cell \((N)\). However, the \(F_{(h,k,l)}\) term is complex number, and we only can obtain its indirect form, intensity, from real experimental observation. This relationship can be manifested in the following form, in which the \(K\) is the scaling coefficient.

\[ I_{(h,k,l)} = |F_{(h,k,l)}|^2 \cdot K \quad \text{(equation 2.2)} \]
Based on the above facts, the next question will be can we get the molecule's structure from the diffraction pattern or the structure factors? Equation 2.1 can be rewritten as the following formula due to the nature of Fourier transform:

\[ \rho_{(x,y,z)} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{(h,k,l)}| \exp[-2\pi i \cdot (hx + ky + lz) + i\alpha_{(h,k,l)}] \]  (equation 2.3)

After the Fourier transform, the left-hand side represents the electron density function at each position in the unit cell. The amplitude of the structure factor can be obtained from the experimental intensity measurement, but the angle \( \alpha_{(h,k,l)} \) in the structure factor is lost. This is the source of the so-called "phase problem" in crystallography. In reality, the phase angle is proved to have more importance than the amplitudes in the structure determination.

In order to get the atomic structure of the molecules in the unit cell, many methods have been established to recover the lost phase information, from the beginning of the trial and error method, which can solve only structures with only a few atoms to direct method (DM), which can solve up to one thousand atoms structure. The rest methods are particular valuable for protein crystallography, like Multiple Isomorphic Replacement (MIR), Molecular Replacement (MR) and Multi-wavelength Anomalous Diffraction (MAD). At present time, the combination of these methods seems give us tools which can solve the phasing problem on any protein molecule with imaginable size.
2.3 Methodology development from DM, MIR to MR

Since the days when the structure of table salt, sodium chloride was solved by trial and error methods, crystallographers have expended much effort to investigate other methods to solve the phase problem due to the increasingly difficulty encountered when the number of atoms of the structure gets larger. By solely using the intensity data to solve the phase problem with the probability theory, Karle and Hauptman first provided a direct method to solve a structure with up to a thousand atoms (Karle and Hauptman 1956). These direct methods rely on the fact that the intrinsic atomicity property of the protein crystal structure and the larger number of measured amplitude data compared to the number of atoms in the structure (Hauptman 2000). The correlation between the measured intensity and the structure factor amplitudes is the focus of the direct method. Assumptions also need to be made to fully exploit this method, like the atoms in the crystal are discrete with idealized shape and the electron density for each atom’s position is positive. The correlation can be treated as probabilistic distribution. One example is the structure invariants which reveals that the some structure factor amplitudes are independent of the origin (Karle 1989). The formation of tangent formula has the essential role in the development of direct method (Sayre 1952; Karle and Hauptman 1956).

But limitations become evident when the number of atoms increases within the crystal structure, because the statistical relationships become weaker.

As long as the crystal diffracts to atomic resolution (1.2Å) with up to one thousand atoms in the unit cell, the structure determination can take the
advantage of direct methods (Hauptman 1997). This method and Patterson function also play important roles in locating the substructure of isomorphous replacement and multi-wavelength anomalous diffraction datasets (Woolfson 1987).

Patterson function analysis is another way to solve the structure of small molecules without any direct knowledge of the phase angle. The function was first introduced by Patterson in 1934 (Patterson 1934). The mathematic formula represents the Fourier transform of the measured intensity data without the phase information of the diffraction spots.

\[
P_{(U,V,W)} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{(h,k,l)}|^2 \exp[-2\pi i \cdot (hU + kV + lW)]
\]  

(equation 2.4)

It provides a map with all interatomic vectors, thus a Patterson map will have total N*N peaks in the unit cell with N*(N-1) off origin peaks. The peak at origins of unit cells is predominant since it is the summation of electrons of all atoms in the unit cell. Another feature of the Patterson function is that it is centrosymmetric because the interatomic vectors can have two directions.

Patterson functions provide very convenient approach to determine a structure with a few atoms. The potential of this function already was fully exploited in conjunction of isomorphous replacement, molecular replacement and multiple wavelength anomalous dispersion methods.

The molecular replacement approach has been routinely used in the protein structure determination, in the case of wild type protein and mutant
proteins, apo-protein and its protein ligand complex, between very high sequence homology proteins, etc. A known structure is used as the model to search against the new diffraction data set. The solution can be found only if the assumption for molecular replacement still remains valid that the expected intermolecular phase variations between the model and experimental data are within a certain threshold. For wild type and mutant protein structure or protein and protein-ligand complex, this phase relationship does not have large variation. But the situation for using homologous proteins as models will depend on the extent of resemblance. Commonly used criteria is at least 40~60% sequence similarities between the two proteins. HemAT sensor domain shares only ~20% sequence similarities to other heme-containing proteins, in particular myoglobin and rice hemoglobin. The molecular replacement results were not satisfactory in an initial attempt to solve its structure using the crystal structures of myoglobin and rice hemoglobin. After the structure was solved with iron MAD analysis, the structural comparisons reveal a great deal of structural differences, such as the rearrangement of helices and extra structural compositions. Empirically, the rms deviations need to be less than 1Å between model and data for molecular replacement to work (Baker, Anderson et al. 1995). The failure of both trial models provides vivid examples to emphasize the sequence similarity criteria.

The molecular replacement in principle is a 6 dimensional search problem, which has been investigated and seen remarkable progress along with present increasingly powerful computers. The 6 dimensional search has been formed as a global optimization problem with many local minima in the target function. The
development of a new strategy to search the global minima starting with low resolution data already can solve some problems which are difficult for traditional approaches (Jamrog, Zhang et al. 2003). Stochastic approaches also have developed to tackle the molecular replacement in 6 dimensions, like reverse Monte Carlo minimization and evolutionary algorithms (Glykos and Kokkinidis 2000; Kissinger, Gehlhaar et al. 2001). The evolutionary algorithm has been implemented in the EPMR program and is able to solve a Cytochrome c₆ structure from *Arthrobacter maxima* with 6 molecules in one asymmetric unit and 24 molecules in the unit cell, when proper inter-molecule constraints are applied. It is a really challenging task for traditional approaches (Kerfeld, Sawaya et al. 2002).

There are still other methods for 6 dimensional searches, more or less with connection of the two sequential 3 dimensional searches or combination for exhaustive searches (Tong 1996; Sheriff, Klei et al. 1999). The disadvantage of the higher dimensional search is that it is still computational intensive, especially for the reverse Monte Carlo algorithm (Glykos and Kokkinidis 2000).

The most commonly used algorithm for molecular replacement is to separate the 6 dimensional searches to two sets of 3 dimensional searches, rotational and translational. This division greatly decreases the computational cost almost exponentially (n⁶ versus 2*n³), where n represents the number of parameters in the crystals. The target function of molecular replacement can be viewed as following formula:
\[ X' = [C]X + D \]  

(equation 2.5)

In this equation, the \( X' \) represents the target function of the unknown structure. \([C]\) is the rotational matrix which needs to apply to the search model in order to get the same orientation (Rossmann 2001). This rotational matrix can be obtained using Patterson functions, whose centrosymmetric properties eliminate the complications of the translational vector. When the appropriate rotational matrix is identified, the Patterson functions of the search model and unknown will have the maximal superposition, which can be formulated in a more defined mathematic form, recognized by Rossmann in 1962 (Rossmann and Blow 1962). The development of the fast rotation function also greatly lowers the computational cost as compared to previous one (Navaza 1994). The method of overlapping two Patterson functions is also described as a real space search. Direct search is another method by rotating the molecule in the unit cell instead of the Patterson space, calculating the corresponding structure factors and comparing this calculated amplitudes with the experimental amplitudes (Brunger and DeLano 1995). The direct search can presumably produce more accurate results because it prevents potential problems associated with the Patterson vectors, but with the added expanse of longer computation times.

The Patterson function search works best when the number of close distances within molecules is greater than between molecules, namely the cross-vectors in the Patterson map arise mainly from the molecule itself. This situation
has been clearly illustrated in a review article (Grosse-Kunstleve and Adams 2001).

But the solution may not be easy to achieve if the above assumption is violated. This usually happens when the protein molecule has an elongated shape. It is very difficult to differentiate the self-vectors from cross vectors, which is just the case for intact HemAT protein.

In equation 2.4, D represents the translational vector to bring the search model to the exact position of the unknown structure. It can be tackled either with a traditional Patterson search for cross-vector matching using the correlation coefficient as criteria, or by using an R factor search, both cases need to take account of the symmetry operations and the results from the rotational search (Crowther and Blow 1967). The development of Patterson correlation refinement also increases the accuracy of the molecular replacement (Brunger 1990).

Due the rapid growth in the Protein Data Bank and the structure genomics initiative, more and more protein structures will be solved with fewer new folds being discovered. Molecular replacement will be an increasingly and frequently used method to deal with the phase problem. The cases for this approach also can be enhanced when theoretically predicted models become more accurate (Marti-Renom, Stuart et al. 2000).

Multiple isomorphous replacement was the first successful technique to tackle the phase problem of a protein and was introduced by Perutz. (Perutz 1956). By diffusing heavy atoms, like mercury, platinum or gold etc. into the protein crystal, these atoms can bind to several special amino acids, like cysteine
methionine, lysine, etc (Islam, Carvin et al. 1998) (see complete review). Incorporating the heavy atoms into the protein crystal, they have significant impact on the diffraction. The changes in the intensity by incorporating the heavy atoms had been estimated very early (Crick and Magdoff 1956).

For centric reflections, their structure factors project onto the real axis. For example, the P2₁ space group has centric reflections on h0l. The phase angle is either 0 degree or 180 degree for these reflections. The influences of heavy atoms on centric reflection are shown in Figure 2.1.

![Image of Argand diagram](diagram.png)

**Figure 2.1** Argand diagram of the isomorphous replacement for centric case

The intensity change for centric reflections is given as following equation:

\[ \Phi \Delta I = 2 \cdot \sqrt{\frac{N_H}{N_p} \cdot \frac{f_H}{f_p}} \]  \hspace{1cm} \text{(equation 2.6)}

The left side is called “the fractional change of intensity”, meaning the average changes in the structure factors of protein due to the heavy atoms (Crick
and Magdoff 1956). The $N_H$ term is the number of heavy atoms with atomic scattering factor $f_H$. The $N_P$ and $f_P$ parts stand for protein with the same meaning. $f_p$ can be estimated as 6.7 for protein molecule if we take the average molecular weight for amino acids, that is about 110 Dalton with 7.7 non-hydrogen atoms for one residue (Hendrickson and Ogata 1997).

In reality, it is more meaningful to calculate approximately the intensity changes for acentric reflections.

Figure 2.2 Argand diagram of the isomorphous replacement for acentric case

The intensity change for acentric reflections is given as following equation:

$$\Phi \Delta I = \sqrt{2} \cdot \sqrt{\frac{N_H}{N_P}} \cdot \frac{f_H}{f_P}$$

(equation 2.7)

We can estimate the fractional change of intensity computationally for HemAT sensor intact protein in the case of two mercury atoms had been attached
to one protein molecule (two heavy atoms per dimer). There are approximately 7000 atoms in the dimeric protein. The signal of the intensity change is about 27% with full occupancy, and about 13% with half occupancy. The error associated with experimental measurement of intensity is about the same value to that with half occupancy (Drenth 1999).

Based on above explanations, the isomorphous replacement seems to be a reasonable way to solve the phase problem with just one heavy atom derivative. The Figure 2.2 can help to form an expression to calculate the structure factor of one heavy atom derivative.

$$|F_{PH}| = \sqrt{|F_P|^2 + |F_H|^2 + 2|F_P||F_H|\cos(\alpha_H - \alpha_P)}$$

(equation 2.8)

Due to the symmetric property of the cos function, two sets of protein angles can be obtained to comply with equation 2.8. So this is the mathematic origin for the phase ambiguity of single isomorphous replacement method. At least a second heavy atom derivative is needed to resolve this ambiguity in order to reconstruct the three-dimensional electron density map if the crystal isomorphism is maintained well, since the two set solutions from the second data will have one identical to one of the previous solutions.

The positions of heavy atoms can be located based on the difference Patterson map employing $|F_{PH}| - |F_P|$ as coefficient in the Fourier synthesis with the help of Harker sections which are special set plane in the Patterson map derived from the symmetry operations (Harker 1956). Or these positions can be
found using direct method. After the heavy atoms are identified, the phase angle for structure factors of protein itself can be properly determined to provide a reasonable map.

However, the phase ambiguity is not absolutely unsolvable with just single set of heavy atom derivative. The approach based on that is called single isomorphous replacement (SIR). This method can be clearly demonstrated with following Argard diagram, see the original reference for more complete description (Wang 1985).

\[
\begin{align*}
2F_{\text{SIR}} &= F_P + F_{\text{False}} \\
2mF_P \exp(i\alpha_{\text{SIR}}) &= F_P \exp(i\alpha_P) + F_P \exp(i2\alpha_H) \exp(-i\alpha_P) \\
2\rho_{\text{SIR}} &= \rho_{\text{Protein}} + \rho_{\text{Noise}}
\end{align*}
\]

**Figure 2.3** Argand diagram of single isomorphous replacement for solving the phase ambiguity and the related equations (Wang 1985)

This diagram and equations reveal that there is only one correct answer in the two sets of angles for the protein structure factor containing single heavy atom derivative data. The other one as indicated \(\alpha_{\text{False}}\) only contributes to the
noise of the electron density in the last equation in the Figure 2.3, which requires it to be filtered out in order to see the correct protein density. With taking into account the fact that the electron density of solvent has lower value than protein has, the double space procedure with solvent flattening was a powerful tool for solving the phase problem in the 1980's. But this procedure was not implemented into a software package available to public yet, except the solvent flattening becomes standard method to improve electron density after obtaining initial phase information by other techniques.

Gradually, Multi-wavelength anomalous diffraction (MAD) emerged in 1991 and was drawn much attention to this new technique for solving the phase problem.

2.4 MAD method

Multi-wavelength anomalous dispersion (MAD) method is the most powerful technique available to solve the phase problem and is based on a physical phenomenon, anomalous scattering. Briefly, when the wavelength of the incident X-ray is close to the natural oscillation frequency of an electron in an atom, the electron within the atom will absorb and re-emit the energy of the X-ray, causing the phase of the diffracted X-ray to be altered. Diffraction from such an event will create a complex component in the atomic scattering factor, which results from the decomposition of the anomalous signal in two directions, as shown in equation 2.9 and Figure 2.4.
\[ f = f_0 + f' + i \cdot f'' \]  

(equation 2.9)

**Figure 2.4** Argand diagram of decomposition of anomalous scattering factor

For this method, the crystal is required to have heavy atoms, which have a distinct absorption edge. They can be included during protein production or utilizing the heavier atoms originally present in the protein itself (i.e. Fe atom) or added by soaking. The amount of anomalous scattering can be examined by taking an X-ray absorption spectrum to find out the absorption edge of the specific heavy atom (Hendrickson 1991).

The imaginary part of the anomalous scattering factor is always 90 degrees ahead of the normal scattering factor, leading to the breakdown of Friedel’s law, namely \( F(h,k,l) \neq F(-h,-k,-l) \), see Figure 2.5.
Figure 2.5 Argand diagram of the breakdown of Friedel's pairs in the MAD.

Access to a tunable synchrotron radiation facility is also necessary for successful collection data (Hendrickson 1991). Diffraction data can be collected at several wavelengths to take advantage of the anomalous dispersion, such as near, at, and far from the absorption edges to be used. The heavy atoms can be located from a difference Patterson map. When the positions of heavy atoms in the unit cell are available, one can use the phase of these heavy atoms to calculate structure factor phases and hence the electron density map of the whole protein (Krishna Murthy 1996).

The difference between Friedel's pairs can be used to estimate the size of the anomalous signals. Exactly like the equation 2.7 which use to get the isomorphous replacement signal, the anomalous signals can be calculated with the similar manner with substitution $f_H$ with $f''_A$.

As a comparison, for the HemAT sensor domain, the iron was used as intrinsic scatterer in the protein molecule during the MAD data collection, $(f''_A$ is
about 4 electrons, see Figure 2.6). The anomalous signal from iron for this protein can be estimated to be 4.5% at the absorption edge. On the other hand, MAD data collection could also be made possible by incorporating selenomethionine into the protein or utilizing the other heavy atoms in the protein.

In this case, for MAD phasing with seleno-methionine of the HemAT sensor domain, with the assumption the natural abundance of methionine residue about 1 over 60 and about 6 electrons for the $f'_{\alpha}$, the anomalous signal from selenium atoms will be 11.6%. This value can be accurately determined at present experimental conditions, which has approximately 10% error.

Based above calculation, it clearly shows the difficulty we encountered during structure determination of HemAT sensor domain using the Iron MAD method.

Figure 2.6 X-ray absorption plot of Iron using theoretical values
3 PROTEIN EXPRESSION AND PURIFICATION

3.1 Expression and purification of wild-type *B. subtilis* HemAT

The HemAT gene from *B. subtilis* was commercially cloned into a pET29b expression plasmid by PCR methods (ATG, Inc.). The recombinant protein of *B. subtilis* HemAT is expressed in *E. coli* BL21(DE3) cells which carry a pET29b-HemAT plasmid. Protein expression is under the control of T7 promoter. When the cell growth temperature was 37°C both before and after IPTG induction, it was found that nearly all of the HemAT protein was in insoluble form at these temperatures. Lowering the growth temperature to 28°C and prolonging induction were helpful in obtaining soluble HemAT and yielded 5-7 mg of protein per liter cell culture. (Later, it was also found that the insoluble protein in the cell pellet can be solublized by 6 M Urea or 6 M Guanidine HCl. When this is done the yield from 1 liter cell culture would be about 40–60 mg. This procedure was not used due to the problem of developing a refolding protocol.)

Protein purification was carried out on the soluble form using perfusion chromatography on a BioCad instrument. After completing ammonium sulfate fractionation, passing through hydrophobic interaction column (HIC), the SOURCE 15Q weak anion exchange column appeared to be the best for binding HemAT. The fractions containing HemAT were monitored by absorption at 280 nm and 410 nm, pooled and concentrated to a final protein concentration of
10mg/ml. The gel filtration columns Superose12 give well-resolved peaks. The ratio of absorbance at 280nm to 410nm is 0.5 to 0.6. Using this protocol, I was able to purify enough amount of HemAT for biophysical and structural study.

Figure 3.1 Left: Coomassie stained SDS-PAGE of intact HemAT. Lane 1: protein standard; lane 2: After HIC column; lane 3: 5µl of eluant from Source 15Q column; lane 4: 10µl of eluant from Source 15Q column; lane 5: 5µl of eluant from Superose 12 column; lane 6: 10µl of eluant from Superose 12 column; lane 7: 20µl of eluant from Superose 12 column; lane 8: 20µl of eluant from S400 column from the other batch. Right: Mass spectroscopic measurement of purified HemAT from eluant from Superose 12 column. The highest peak has apparent molecular weight 48757 Da, which is extremely close to the expected monomeric polypeptide chain of HemAT (48767.4 Da).
3.2 Protein production of *B. subtilis* HemAT sensor domain and its mutants

3.2.1 Design of the enzymatic cleavage site and histidine affinity tag

It is a common to attach six histidines at the N- or C- terminus for metal affinity purification during liquid chromatography, sometimes with a varied length of consecutive histidines for this purpose, ranging from 4 to 10 (Bridges, Gruenke et al. 1998). One example is that the 6 histidines can facilitate the purification of beta-Galactosidase, a MW of 60 kDa protein (Daabrowski, Sobiewska et al. 2000). We assumed that the 6 histidines tag can also succeed for HemAT protein which is about 48 kDa for intact monomer and 22 kDa for sensor domain monomer.

In many previous works, researchers unambiguously found that the sequence upstream of Factor Xa cleavage site (IEGR) can interfere with Factor Xa activity (Gardella, Rubin et al. 1990; Markmeyer, Ruhlmann et al. 1990; Morganti, Huyer et al. 1996; Davis, Elisee et al. 1999). They postulated it is due to steric hindrance effect. This problem can be solved by placing a linker which is composed of small side chain residues, like Ser, Gly, Ile, etc. Some sequences used in previous work include:
There are cases that do not use a linker between His(6) and IEGR. However, one is only partially successful attempting to remove the plasmid encoded residues. The sequence is Ala-Ser-His(6)-IEGR-Met----- (Morganti, Huyer et al. 1996). In another case, the sequence (Met-Lys-His(6)-IEGR-Met-Lys--) was added at N-terminus, and yielded purified protein. But they failed to show the evidence to support that His(6) can be cleaved by Factor Xa (Reece, Rickles et al. 1993).

From the sequences shown above, the immediate downstream residue of IEGR is not Met. This is because the Factor Xa cleavage site is mainly used for eukaryotic gene expression and it is not necessary to have a Met at the first position. However, the instructions provided by Novagen, Inc. support a construct that has a Met residue downstream of the IEGR sequence.

Based on these facts, a linker between His(6) and IEGR is necessary for enzymatic digestion, like -Gly-Gly-Ser-. Obtaining the full length protein of HemAT and the sensor domain was constructed at the initial stage of this project. A full length HemAT with a His(6) tag and a Factor Xa cleavage site at N-terminus and the sensor domain with a His(6) tag and a Factor Xa cleavage site
at N-terminus were all made for this work. A number of constructs have been made according this way, see details in summary.

3.2.2 Expression and purification of HemAT sensor domain and its mutants

The HemAT gene from *B. subtilis* was commercially cloned into a pET29b expression plasmid by PCR methods (ATG, Inc.). In addition to the six histidine residues on the N-terminus of the HemAT sequence, one Factor Xa cleavage site (Ile-Glu-Gly-Arg) sequence was inserted before the first residue, methionine, of the sensor domain HemAT (1-178 residues). *E. coli* (BL21(DE3)) cells transformed with pET29b-HemAT plasmid were grown overnight at 37°C on 2xYT (Tryptone16.0g, Yeast extract10.0g Sodium chloride 5.0g per liter) agar plates with addition of 30μg/ml kanamycin. A single colony was picked to inoculate 1 liter 2xYT media with antibiotics. Protein induction was carried out by addition of 1mM IPTG at final concentration. The growth temperature was adjusted to 28°C after the induction to continue growth 6 ~ 10 hours.

*B. subtilis* HemAT sensor domains were heterologously expressed in *E. coli*. Cell growth temperature was initially at 37°C for this strain both before and after IPTG induction. However, it was found that large portion of HemAT protein was in insoluble form under these conditions. Lowering the growth temperature to 30°C and prolonging the induction time were very helpful methods in obtaining soluble HemAT and yielded 20 ~30mg of protein per liter
of cell culture. No further effort was taken to solubilize the inclusion bodies, since large amounts of protein were in soluble form.

The Ni affinity column facilitated the purification by providing relatively high purity protein in one step. After the enzymatic cleavage of the His-tag, the purity of HemAT protein was further improved with anion exchange and hydroxyapatite chromatography. The sample purity at each purification step was monitored by SDS-PAGE. The purity of the protein sample used for crystallization was examined by electro-spray mass spectrometry and the measured mass was within 10 daltons of the expected value (21,812 Da, without heme group). Protein used for crystallization was oxidized to the more stable cyano- form, because initial trials on the unliganded form were not successful.

---

Figure 3.2 Coomassie blue stained SDS_PAGE of Factor Xa digestion of the HemAT sensor domain. Lane A, after Ni-column purification; lane B-E, digestion at 2h, 4h, 8h, 16h using commercial enzyme; lane F-I, digestion at 2h, 4h, 8h, 16h using enzyme provided by Dr. J.S. Olson.
Figure 3.3  SDS-PAGE of purified HemAT sensor domain. Lane 1: Protein standard with label on the left. Lane 2: eluant from Ni-chelating column. After breaking the cells, centrifugation, supernatant was loaded to Ni-chelating affinity column. Lane 3: after Factor Xa enzymatic digestion. Lane 4: eluant from a hydroxyapatite column.
4 BIOPHYISCAL CHARACTERIZATION OF HEMAT PROTEIN

4.1 UV/Visible Absorption Spectroscopic measurement

UV/Visible absorbance spectra of purified intact HemAT were collected on a Shimadzu UV-2401 PC spectrophotometer by using quartz cuvettes with 1cm path length in 100mM sodium phosphate buffer, pH 7.0 at 22oC. Deoxygenated HemAT was obtained by adding a few grain of dithionite solid to the cuvette containing purified HemAT protein. Fully oxygenated HemAT was prepared by passing the dithionite reduced HemAT through a small Sephadex G-25 column, and then flushing the sample with pure oxygen gas in a stoppered cuvette to obtain 100% saturation. The carbon monoxide derivative was obtained by bubbling pure carbon monoxide into the cuvette instead of oxygen.

The spectrum of HemAT suggests that the purified protein is largely in the oxygen bound form in the solution (100mM sodium phosphate buffer, pH 7.0). This finding also suggests that the function of HemAT is to ensure that bacteria stay in an aerobic environment by binding oxygen. Three major peaks are seen in the spectrum of oxygen-bound HemAT, 412 nm (Soret), 544nm (β-band), and 578nm (α-band), compared with peaks of 418nm, 543nm and 581nm for sperm whale myoglobin. This evidence clearly shows that HemAT is a heme protein and it contains bound O₂. However, our studies show slight differences from the Hou's spectra (Table 4.1). The spectrum of deoxy-HemAT is taken by adding
sodium dithionite to remove the bound oxygen ligand. In Hou's result the Soret peak and multiple small peaks at visible regions indicated that the sample used in their experiment was contaminated by other forms of heme.

<table>
<thead>
<tr>
<th></th>
<th>Soret</th>
<th>β-band</th>
<th>α-band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxy-</td>
<td>Mb</td>
<td>418</td>
<td>543</td>
</tr>
<tr>
<td></td>
<td>HemAT</td>
<td>412</td>
<td>544</td>
</tr>
<tr>
<td></td>
<td>Hou's</td>
<td>406</td>
<td>538</td>
</tr>
<tr>
<td>Deoxy-</td>
<td>Mb</td>
<td>434</td>
<td>556</td>
</tr>
<tr>
<td></td>
<td>HemAT</td>
<td>431</td>
<td>560</td>
</tr>
<tr>
<td></td>
<td>Hou's</td>
<td>425</td>
<td></td>
</tr>
<tr>
<td>Co-</td>
<td>Mb</td>
<td>423</td>
<td>542</td>
</tr>
<tr>
<td></td>
<td>HemAT</td>
<td>421</td>
<td>541</td>
</tr>
<tr>
<td></td>
<td>Hou's</td>
<td>415</td>
<td>536</td>
</tr>
</tbody>
</table>

Table 3.1 Comparisons of absorption peaks of sperm whale myoglobin, our partially purified HemAT and HemAT from Hou's studies.
Figure 4.1 Absorption spectra of purified HemAT and sperm whale myoglobin. Upper figure shows the spectra of sperm whale myoglobin, lower for HemAT. The met-HemAT is not measured. N₂ represents the spectrum of HemAT in the solution during purification with no chemical manipulations.
4.2 Circular Dichroism Spectroscopy

Circular dichroism spectra of purified intact HemAT were collected on an AVIV model 62A DS spectrophotometer at 22°C. A quartz cuvette with 1cm path length was used in the experiment. Protein was diluted to a concentration of 50μg/ml in 100mM sodium phosphate buffer, pH 7.0. Circular dichroism wavelength scans were taken from 190 to 250nm with bandwidth 1nm and 1 minute averaging time. The final spectrum was reported as extinction coefficient (Δε) versus wavelength after subtracting the baseline spectrum from the protein sample spectrum and converting the measured CD signal to extinction coefficient. The secondary structure analysis was examined using the software provided by AVIV.

Protein sequence alignment reveals that the N-terminus of HemAT (residue 1~170) exhibits limited homology to sperm whale myoglobin, and the C-terminus of HemAT (residues 198 to 432) shares 30% sequence identity to E. coli serine receptor Tsr and other methyl-accepting chemotactic proteins. Modeling of cytoplasmic domain of 20 chemotactic receptors indicates that this region is mainly helical structure with characteristic seven-residue repeats. Moreover, the crystal structure of the cytoplasmic domain of E. coli Tsr was determined and revealed that Tsr was a dimeric protein with an elongated coiled-coil of anti-parallel helices. In order to estimate the content of α-helical and β-sheet structure, the secondary structure of intact HemAT was examined using circular dichroism spectroscopy. CD spectra of HemAT indicate that it is typical helical protein based on the characteristic double minima at 218nm and 208nm.
The content of secondary structure composition was estimated based on other protein structures. The results reveal that HemAT has 67% α-helix, 6.9% β-sheet, 12.5% turn and 15.9% random coil. The helical composition is slightly less than that of sperm whale myoglobin (78.0% α-helix content). The results from circular dichroism spectroscopy are consistent with the expected structure of HemAT based on homology modeling with other chemoreceptors and myoglobin-like proteins. Our circular dichroism experiment provided spectroscopic evidence that the secondary structure of HemAT is primarily α-helical.

Figure 4.2 Circular dichroism spectrum of intact HemAT. Protein sample was at concentration of 50μg/ml in 100mM NaH₂PO₄, pH 7.0. Circular dichroism wavelength scans were taken from 190 to 250nm.
4.3 Analytical Ultracentrifugation

I observed that purified intact HemAT did not behave like a globular protein during our protein purification process, and were unable to determine its accurate molecular weight using size exclusion chromatography. The column elution profiles suggested that HemAT was a single dimeric species with an elongated shape. This contradicts the result reported by Aono et al. (Aono, Kato et al. 2002). Their result showed that HemAT is a homo-tetrameric protein on the basis of size exclusion chromatography experiment. Moreover, the cytoplasmic domain of E. coli Tar receptor also displayed very dynamic dissociation behavior at different pH conditions of a solution (Seeley, Weis et al. 1996). In order to resolve this discrepancy, analytical ultracentrifugation was employed to determine its accurate molecular weight and oligomeric state.

Sedimentation velocity and sedimentation equilibrium experiments were performed on a Beckman model XL-A analytical ultracentrifuge at 10°C. The purified intact HemAT was prepared in 50mM Tris.HCl buffer, pH 7.0, 5mM NaCl, and 4% glycerol. The sedimentation velocity experiment was carried out at rotor speed of 60,000 rpm and monitored by the absorbance at 415nm. The second moment method was used for sedimentation velocity data analysis.

The average sedimentation coefficients and diffusion coefficients were calculated from the slope of the plot ln(r) versus $o^2t$ and the plot of the boundary spreading, $z$, versus $t$ (correlation coefficient 0.99 for both values, Figure 4.3). Sedimentation equilibrium experiment was carried out at 8,000, 11,000, and 14,000 rpm using six-sector cells at three different protein concentrations.
**Figure 4.3** Determination of Sedimentation coefficient based on Transport Method. The best fit to equation \( \ln(r) = \omega^2 st \) gives sedimentation coefficient \( s \) 2.63465E-13 with Correlation coefficient 0.99484.

Sedimentation equilibrium experiments were carried out to determine the molecular weight independently of its frictional coefficients. A molecular weight of 98,292Da was obtained from these analyses and is remarkably close to the theoretical value of the HemAT dimer based on sequence information (98,766Da including the heme moiety). The experiments also revealed that HemAT is in a
homodimeric form in solution without dissociating or associating to other oligomeric forms.

Results from sedimentation velocity experiments suggest that the protein is homogenous in solution under our current experiment conditions. The sedimentation coefficient, however, is small, 2.63s, in comparison to that expected value for globular protein of similar molecular weight (7~8s). This low value also suggests that HemAT has an extremely elongated shape.

Information on the overall shape can also be obtained from our analytical ultracentrifugation experiments. The molecular weight was used in conjunction with sedimentation coefficient value to calculate the frictional coefficient of HemAT. The ratio of the experimental frictional coefficient (f) to theoretical frictional coefficient for a spherical particle (fo) is 2.5, which demonstrates that wild type HemAT has a highly asymmetric shape.

Thus, we conclude that HemAT is an elongated homodimer with axial ratio greater than 10, which is consistent with the structure of other chemoreceptors.
Figure 4.4 Analytical ultracentrifugation experiments. A.) Sedimentation equilibrium curve of purified intact HemAT. Data shown were recorded at 415 nm at rotor speed 11,000 rpm. The residual distributions are shown at the upper panel. B.) Sedimentation velocity experiment of purified intact HemAT. Experiments were carried out at rotor speed of 60,000 rpm and the absorbance was monitored at 415 nm. C.) Sedimentation coefficient distribution g(s*) plot. Peak of the curve corresponds to sedimentation coefficient 2.63s.
4.4 FTIR Spectroscopic Measurements

The stretching frequency of hemeprotein-CO complex has been investigated using FTIR spectroscopy and computational calculations with a Posisson-Boltzman method (Li, Quillin et al. 1994; Phillips, Teodoro et al. 1999). These results demonstrated that the determination of ν(C-O) can yield valuable structural information on residue composition and serve as a sensitive probe to measure the electrostatic potentials in the ligand binding sites because of the inverse relationship between the electrostatic fields and ν(C-O). So this method have been applied to many other hemeproteins on the investigations of the ligand binding properties (Chu, Katakura et al. 2000; Kaposi, Vanderkooi et al. 2001; Kaposi, Wright et al. 2001).

The infrared spectral data collections of both intact protein and sensor domain were carried out on a Nicolet Nexus 470 FTIR spectrometer with the assistance of OMNIC software. The samples were concentrated to about 2mM, reduced with dithionite and equilibrated with 1 atm CO gas in a small tube in 10mM sodium phosphate buffer, pH7.0 before taking the data. The FTIR cuvette was filled with 20μL protein sample each time. Baseline spectrum was first collected for background correction. The absorbance change was recorded as a function of wavenumber with the resolution of 2cm\(^{-1}\) at 25°C. Total 64 scans were collected each time and averaged to give final infrared spectra after subtraction of the background spectrum.

The FTIR spectra of both wild type intact HemAT and sensor domain were determined in 10mM sodium phosphate buffer, pH 7.0 at 25°C. Their
vibrational spectra have only single peak with frequencies $\nu$(C-O) centered at 1967.3 cm$^{-1}$ for intact HemAT-CO complex and 1966.7 cm$^{-1}$ for sensor domain-CO complex in the range from 1900 cm$^{-1}$ to 2000 cm$^{-1}$. Due to the resolution used in the data collection (2 cm$^{-1}$), the C-O stretching frequencies for intact HemAT, wild type sensor domain and its mutants are nearly identical. This observation suggested that the electrostatic field around CO ligand and possibly the conformations of residues in ligand binding pocket of sensor domain are not perturbed by the truncation of the signaling domain. It also suggested that the mutations taken place at Tyr70 position do not affect greatly the bonding between CO and the heme group. The doublet feature of the Y70W mutant may reflect the fact of multiple conformations of the bulky side chain.

Li and Phillips reported that the vibrational frequency of the wild type sperm whale myoglobin increased from 1945 to 1961 $\sim$ 1967 cm$^{-1}$, when Val68 was replaced with Thr or His64 was changed to non-polar residues or polar residues like Thr, Tyr or Trp (Li, Quillin et al. 1994; Phillips, Teodoro et al. 1999). The $\nu$(C-O) of HemAT protein has about the comparable value (1967 cm$^{-1}$) and thus strongly suggests that the CO group has characteristic of a triple bond. The distal ligand binding pockets seems very unlikely to have a positively charged group close to the ligand as revealed in the sequence alignment (Hou, Larsen et al. 2000).
Figure 4.5 FTIR spectra of intact HemAT protein and the sensor domain. These data were measured at protein concentration of 2mM in 10mM sodium phosphate buffer, pH 7. The absorbance change was recorded as a function of wavenumber with the resolution of 2cm-1 at 25°C.
4.5 Oxygen Equilibrium Determination

Oxygen equilibrium curves were obtained using Shimadzu spectrophotometer equipped with an Imai cell and O₂ electrode (Imai 1981). The system was calibrated with 100% nitrogen, 20% oxygen, and 100% oxygen at each time. 50µM of purified intact HemAT in 100mM sodium phosphate buffer, pH 7.0 was injected into the cuvette and the formation of deoxygenated protein was monitored by absorbance change at 560nm in an isothermal environment, 25°C. Absorbance data were collected at preset changes of P₀₂. Spectra from 400-700nm were collected at 100% and 0% oxygen concentrations to monitor the whole process. Reoxygenation experiments were also carried out to confirm the deoxygenation experiments.

The oxygen equilibrium curve of intact HemAT is displayed as Hill plot with an n-value (Hill coefficient) ~0.45. The low n-value indicates that the oxygen binding reaction of HemAT has either having heterogeneous binding components or negative cooperativity. Such multi-phasic binding is not unique for HemAT, and it was also observed for the ligand binding to the other chemoreceptors, Tar and Tsr (Biemann and Koshland 1994; Lin, Li et al. 1994). Thus two independent binding sites were used to model the oxygen equilibrium curve. The Kₐ for the two binding sites were fitted to be 1.2µM and 16.2µM, the concentration of O₂ at half saturation, P₅₀, is ~8µM. These data suggest that one binding site has more than 10 fold higher affinity than the other one. The kinetic data predicted Kₐ values of 5 to 100µM whereas the direct equilibrium measurements gave a range of 1-20µM. This discrepancy is probably due to the
heterogeneity observed in both the kinetic and equilibrium experiments where analysis in terms of only 2 components is probably an oversimplification. Nevertheless, the ranges do overlap.

**Figure 4.6** Hill plot of oxygen equilibrium curve of intact HemAT. Isotherm oxygen dissociation process of HemAT was recorded at 25°C with protein concentration of ~50 M in 10mM sodium phosphate buffer, pH 7.0. Y is the fractional saturation.
4.6 Kinetic studies of ligand binding

4.6.1 Kinetic Experiments

The association rates \((k')\) for oxygen and carbon monoxide binding to intact HemAT and sensor domain were measured by flash photolysis techniques as described previously (Rohlfs, Mathews et al. 1990). Deoxygenated sample was diluted to final concentration of 50\(\mu\)M in 100mM sodium phosphate buffer, pH 7.0. The sample was injected into a stoppered 1-mm path length cuvette and converted to fully liganded form by flushing with pure oxygen or carbon monoxide. The absorbance changes were monitor at 436nm or 422nm.

Rapid mixing techniques were carried on a Gibson-Dionex stopped-flow apparatus equipped with an On-line-Instruments-System (OLIS) model 3820 data collection system at 20\(^\circ\)C. The dissociation rates \((k)\) for oxygen and carbon monoxide were determined by using ligand replacement techniques as described previously (Olson 1981). The fully oxygenated HemAT was rapidly mixed with an anaerobic stock solution of 1mM (100%) carbon monoxide in 100mM sodium phosphate buffer, pH 7.0. The replacement rate constants were obtained by fitting the experimental time courses to a double exponential. The \(O_2\) dissociation rate constants were calculated from replacement rates, \(r_{O_2}\), by compensating for the competition between \(O_2\) and CO rebinding \((k_{O_2}=r_{O_2}(1+k'_{O_2}[O_2]/k'_{CO}[CO]))\). The rate constant for CO dissociation from
HemAT was determined by mixing the CO derivative at ~100μM free carbon monoxide with a 2mM (100%) solution of NO (Olson, Foley et al. 2003).

4.6.2 Oxygen Binding Reaction

Kinetics parameters for O₂ and CO binding to HemAT proteins are listed in Table 4.1. Oxygen rebinding to HemAT after laser photolysis showed two phases for both intact HemAT and the sensor domain (Figure 4.7). The observed association rate constants \( k'_{\text{obs}} \) of the fast phase reaction are in a linear relationship with the ligand concentration (Figure 4.6), which provides evidence to support the validity of the experiment. The correlation coefficient of the best linear fitting is 99% between the observed data for both cases. The association rate constant \( k'_{\text{O2}} \) for both proteins are nearly the identical, 18μM⁻¹s⁻¹ at pH7.0 20°C, which is comparable to that of sperm whale myoglobin (Springer, Sligar et al. 1994). The rate constants of the mutants are about 3–4 fold faster than both wild type proteins.

On the other hand, the slow phases of O₂ rebinding reactions are in an inversely linear relation with the log scaled O₂ concentration. The correlation coefficient of the best linear fitting is 97% between the observed data for the slow phase of both proteins. This implies that the O₂ ligand affinity to the protein decreases in the slow phase reaction when ligand concentration gets higher.
Despite that it is not clear exact cause, we hypothesize that the two ligand-binding sites may not adopt a similar conformation at the low ligand concentration. The disappearance of the slow phase at very high O₂ concentration may reflect the structural homogeneity, so the O₂ rebinding would only one rate constant.

Dissociation of O₂ from HemAT shows biphasic time courses (Figure 4.8). The dissociation rate constants (k₀₂) for the fast phase of wild type intact protein are 2,000s⁻¹ and 85s⁻¹ for slow phase. The k₀₂ for the fast phase is ~100 times faster than that observed for sperm whale myoglobin (15s⁻¹) (Springer, Sligar et al. 1994). High oxygen dissociation rate constants are also observed in *Glycera dibranchiata* HbII (Leu at E7 position, 1,800s⁻¹) and sperm whale myoglobin mutants in which His64 is replaced with apolar amino acid (2,000 ~ 10,000s⁻¹), possibly due to the lack of a distal pocket water molecule that would require exchange of the water for ligand (Rohlfs, Mathews et al. 1990). The dissociation rate constants for wild type sensor protein are very similar, only slightly smaller for slow phase (50s⁻¹).

The k₀₂ for the slow phase of HemAT is about ~4 times faster than that for sperm whale myoglobin. The oxygen affinity constant (K₀₂ = k'₀₂/k₀₂) for the fast and slow phase are estimated to be 0.01 and 0.2~0.3μM⁻¹, respectively, both of which are less than that from sperm whale myoglobin as previously mentioned
(Thorsteinsson, Bevan et al. 1999; Hou, Freitas et al. 2001). The estimated $O_2$ affinity of slow phase component is similar to that for sperm whale myoglobin containing the H64Q mutation, and close to the $O_2$ affinity of HemAT determined previously by other researchers (Thorsteinsson, Bevan et al. 1999; Hou, Freitas et al. 2001).

The dissociation rate constants for both fast phase and slow phase for mutants are significantly faster, about 10~30 folds. The enormously fast off rates are the critical factor for limiting the $O_2$ affinity to the mutants.

From the liganded structure HemAT sensor domain, the cyano ligands are stabilized by two hydrogen bonds from both side-chain of Tyr70 and a water molecule in the distal pocket. The mutations made on sensor domain (Y70F, Y70L and Y70W) in fact remove the hydrogen-bonding donor structurally. The lack of stabilization force may play a critical role on the ligand binding, especially having great impact on the off rates. It is reasonable to assume that ligand binding geometry of $O_2$ may be similar to that of cyano-liganded structure.
Figure 4.7 The correlation of observed rate constants of oxygen association reaction with ligand concentrations. Top panel shows the fast phase of the ligand rebinding, and lower panel shows the slow phase for both wild type intact and sensor domain.
**Figure 4.8** Oxygen association curve of HemAT determined by flash photolysis at 436nm. Top panel shows the fast phase of the ligand rebinding at 100% $O_2$. Lower panel shows the time courses of wt sensor of HemAT at different ligand concentrations. Protein sample was in 100mM sodium phosphate buffer, pH 7.0 at 20°C. The insert represents the slow phase of ligand rebinding at concentration of 20% $O_2$. The y-axis is the normalized fraction of HemAT protein.
Figure 4.9 Times curves for $O_2$ dissociation from HemAT measured by ligand replacement technique (100% $O_2$ and 10% CO). The fully oxygenated HemAT was mixed with 0.1mM CO solution rapidly, and the absorbance change was measured at 422nm. With two different time scales, the plot shows clearly the two phases during the ligand dissociation reaction. Protein sample was in 100mM sodium phosphate buffer, pH 7.0 at 20°C. Two-exponential fitting method was used to model the $O_2$ dissociation curve.
4.6.3 Carbon Monoxide Binding Reaction

The kinetic parameters for the reaction of HemAT proteins with carbon monoxide are also listed in Table 4.1. Very interestingly, the time courses for CO rebinding after laser photolysis are monophasic (Figure 4.10), and the association rate constant \( (k'_\infty) \) is 0.34\( \mu \text{M}^{-1}\text{s}^{-1} \), which is only 2-fold smaller than that of myoglobin. The dissociation time courses for CO are also monophasic (Figure 4.11), unlike those for \( \text{O}_2 \). The calculated CO affinity is 5.07\( \mu \text{M}^{-1} \), about ~5 times smaller than that for sperm whale myoglobin.

The ligand binding reactions of three mutants also show the comparable behavior, monophasic. All three association rate constants \( (k'_\infty) \) are in the same range of magnitude. The dissociation rate constants are 2 fold smaller than that of wild type proteins. Only mutant Y70F has 2~3 fold higher affinity than wild type proteins. The mutant Y70L and Y70W have about the same CO affinity as wild type.

It is surprising to observe that the mutations do not affect ligand binding of CO by either association or dissociation. The crucial structural elements in the \( \text{O}_2 \) binding are not present here, which may lead to the conclusion that the structural composition of CO binding is significantly different from what we see in the cyano-liganded structure.
Figure 4.10 CO association curve of HemAT determined by flash photolysis at 436nm. The plots shows the single phase of the ligand rebinding at concentration of 100% CO. Protein sample was in 100mM sodium phosphate buffer, pH 7.0 at 20°C. The y-axis is the normalized fraction of HemAT protein.
Figure 4.11 Time courses for CO dissociation from HemAT measured by ligand replacement technique at 422nm (10% CO and 100% NO). The plot shows clearly the single phase during the ligand dissociation reaction. Protein sample was in 100mM sodium phosphate buffer, pH 7.0 at 20°C. Two-exponential fitting method was used to model the O₂ dissociation curve.
4.6.4 Summary of Kinetics of Ligand Binding

HemAT shows two affinities for O₂, one with a Kₐ of 1-5μM and another with a Kₐ of 20-100μM. The high affinity component shows a Kₐ that is similar to that obtained previously (Hou, Freitas et al. 2001). Even though there are structural evidence available for asymmetric conformations (and/or nonidentical ligand binding sites) in the HemAT homodimer, we are still unable to unambiguously differentiate the structural heterogeneity from negative cooperativity. However, negative cooperativity has been observed for aspartate receptor from *E. coli* and *S. typhimurium* and for the serine receptor from *E. coli*, and by analogy, it also need to be taken into account for the low n-value seen for O₂ binding to HemAT (Biemann and Koshland 1994; Lin, Li et al. 1994).

If negative cooperativity occurs, the biphasic O₂ dissociation time courses suggest that HemAT may experience structural rearrangement in the ligand-binding site of one subunit when O₂ molecule dissociates from the second subunit. The changes caused by ligand dissociation would potentially stabilize bound O₂ molecule in the first site. Thus, in the reverse process, during O₂ binding to deoxygenated HemAT, the first ligand would bind tightly and the second would bind more weakly. However, currently there is no indication of a two-phase process for HemAT sensing in the literature, and the two O₂
dissociation rates could represent conformational heterogeneity with slow interconversion with respect to ligand binding.

If structural heterogeneity is responsible for the O₂ ligand binding behaviour, the two nonidentical binding sites will have different binding affinity as revealed in our kinetics experiments. The conformational changes on one site can also be seen by the other site.

Taking together with the FTIR data, the kinetics of CO binding to HemAT protein suggest that CO ligand may adopt dramatically different arrangements in the complex. The distal pocket residue Tyr70 may not have the identical conformation to that in either cyano- or oxy- liganded structures. The hydrogen bond seemingly does not appear to be a major factor. It is tantalizing to postulate that the positions of Tyr70 side chains of both subunits may be out of the pocket.

Regardless of the exact explanation, HemAT is always partially oxygenated when it is exposed to air. This is very likely due to rapid loss of oxygen from the site showing the large $k_{O₂}$ value. This observation needs to be taken into account in any discussion of the biological roles of HemAT in O₂ sensing.
<table>
<thead>
<tr>
<th>Protein</th>
<th>(k'_O_2)</th>
<th>(k_{O_2})</th>
<th>(K_{O_2})</th>
<th>(k'_C_O)</th>
<th>(k_{C_O})</th>
<th>(K_{C_O})</th>
<th>(K_{C_O}/K_{O_2})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\mu M^{-1}s^{-1})</td>
<td>(s^{-1})</td>
<td>(\mu M^{-1})</td>
<td>(\mu M^{-1}s^{-1})</td>
<td>(s^{-1})</td>
<td>(\mu M^{-1})</td>
<td></td>
</tr>
<tr>
<td>wt SwMb (^a)</td>
<td>17</td>
<td>15</td>
<td>1.1</td>
<td>0.51</td>
<td>0.019</td>
<td>27</td>
<td>25</td>
</tr>
<tr>
<td>SwMb H64Q (^b)</td>
<td>24</td>
<td>110</td>
<td>0.19</td>
<td>0.94</td>
<td>0.012</td>
<td>78</td>
<td>411</td>
</tr>
<tr>
<td>SwMb H64L (^c)</td>
<td>98</td>
<td>4110</td>
<td>0.023</td>
<td>26</td>
<td>0.024</td>
<td>1100</td>
<td>48000</td>
</tr>
<tr>
<td>Glycera HbII (^c)</td>
<td>190</td>
<td>1800</td>
<td>0.11</td>
<td>27</td>
<td>0.042</td>
<td>640</td>
<td>6000</td>
</tr>
<tr>
<td>Aplysia Mb (^c)</td>
<td>15</td>
<td>70</td>
<td>0.21</td>
<td>0.5</td>
<td>0.02</td>
<td>25</td>
<td>120</td>
</tr>
<tr>
<td>B.s. HemAT</td>
<td>18.94</td>
<td>1880</td>
<td>0.01</td>
<td>0.34</td>
<td>0.067</td>
<td>5.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>87</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt Sensor</td>
<td>18.61</td>
<td>1810</td>
<td>0.01</td>
<td>0.43</td>
<td>0.070</td>
<td>6.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensor Y70F</td>
<td>52.52</td>
<td>18962</td>
<td>0.003</td>
<td>0.47</td>
<td>0.030</td>
<td>15.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1778</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensor Y70L</td>
<td>89.45</td>
<td>69776</td>
<td>0.001</td>
<td>0.16</td>
<td>0.029</td>
<td>5.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3463</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensor Y70W</td>
<td>57.01</td>
<td>22099</td>
<td>0.003</td>
<td>0.28</td>
<td>0.043</td>
<td>6.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3067</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Data for sperm whale myoglobin and human hemoglobin \(\alpha\) subunit taken from ref. (Springer, Sligar et al. 1994).

\(^b\) Sperm whale myoglobin mutant H64Q taken from ref (Thorsteinsson, Bevan et al. 1999).

\(^c\) Sperm whale myoglobin mutant H64L, Glycera HbII and Aplysia Mb taken from ref. (Rohlfs, Mathews et al. 1990).

Table 4.1 Rate and equilibrium constants for ligand binding to HemAT
5 PRELIMINARY CRYSTALLOGRAPHIC ANALYSES

5.1 Protein crystallization and crystal manipulations

5.1.1 Protein crystal growth of intact HemAT protein

The protein used in crystallization setup was prepared by oxidizing wild type HemAT with excess potassium cyanide. The sample was passed through a 5ml Sephadex G-25 column, equilibrated with 50mM Tris.HCl, pH 7.0, 5mM NaCl and 1mM EDTA. The protein concentration was 10~15mg/ml. Crystallization conditions were screened using hanging-drop vapor-diffusion method at different temperatures (Hampton Crystal Screens I, II, Emerald Wizard Screens I, II and home made reverse screening, temperature 4°C, 10°C and 25°C). The diffraction data were collected at cryo-temperature by using fiber loop at APS beamline at wavelength 0.9Å with oscillation angle 1 degree and exposure time 90 seconds. The distance between crystal and detector was 350mm for all datasets.

Crystals of wild type HemAT were grown from 10~18% PEG6000, 100mM KH$_2$PO$_4$, pH 7.0, and 200mM (NH$_4$)H$_2$PO$_4$ at room temperature. Crystals of HemAT showed red color with dimensions around 0.2x0.05x0.05 mm (Figure 5.1). The crystals diffract to a resolution of about 7Å. The space group was determined to be P4$_1$2$_1$2 or P4$_2$2$_1$2 with unit cell dimensions of a=b=247Å, c=185Å.
**Figure 5.1** A single crystal of intact HemAT. Crystal dimensions are about 0.2×0.05 × 0.05 mm

<table>
<thead>
<tr>
<th></th>
<th>Intact HemAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>$P4_12_12$ or $P4_32_12$</td>
</tr>
<tr>
<td><strong>Unit cell parameters</strong></td>
<td>$a=b=247\text{Å}$, $c=185\text{Å}$</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>150 ~ 7.1</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>38,323</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>8,752</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>96.7 (91.4)</td>
</tr>
<tr>
<td><strong>$R_{sym}$ (%)</strong></td>
<td>11.5 (40.9)</td>
</tr>
</tbody>
</table>

**Table 5.1** Statistics of Data Collection of intact HemAT
5.1.2 Protein crystal growth of HemAT sensor domain

5.1.2.1 HemAT sensor domain in cyano form

The protein used in crystallization setups was prepared by converting purified HemAT to cyano- liganded form with excessive amount of potassium cyanide. The sample was then passed through a small Sephadex G-25 column, equilibrated with 10mM phosphate buffer, pH 7.0, 5mM NaCl and 4% glycerol. The final protein concentration was 10 ~15mg/ml. The sample was aliquoted and frozen at -80°C immediately. Each time when doing crystallization experiments, the protein sample was thawed and centrifuged using a bench top microcentrifuge at 14,000 rpm for 15 minutes in a cold room. Crystallization conditions were initially screened using the hanging-drop vapor-diffusion method using a variety of precipitants at different temperature.

Crystals of HemAT sensor domain were grown from 10~18% sodium citrate, 100mM KH$_2$PO$_4$, pH 7.0 at room temperature. However, these crystals formed needle like clusters and were unsuitable for X-ray diffraction studies. After testing crystallization conditions with detergent additives, better crystals were obtained. Large and reproducible crystals were obtained at 0.1% concentration (w/v) of n-octyl-β-D-glucoside (Sigma). The morphology was dramatically improved, but clusters were still formed. By breaking down the crystal clusters, single plate-like crystals were obtained. Crystals of HemAT
showed red color with the largest one with dimensions of about 0.2x0.05x0.05 mm for the single plate crystal.

Figure 5.2 Cluster of reddish plate-like crystals of *B. subtilis* HemAT sensor domain. Typical dimensions of a single crystal are approximately 0.2x0.05x0.05mm. Sodium citrate was used as precipitant in the crystallization condition.

5.1.2.2 HemAT sensor domain in unliganded form

The unliganded HemAT crystal was obtained by reducing the cyano liganded crystal with excessive amount of sodium dithionite. The time courses for the reduction reaction were tested to get a fully reduced form. The crystal color change indicated the transition from the liganded form to the unliganded form, as has been used for other hemeproteins (Quilllin, Arduini et al. 1993; Brucker, Olson et al. 1996; Hao, Isaza et al. 2002).
5.1.2.3 **HemAT sensor domain in met form**

The met form of purified HemAT was made by treating the protein sample with potassium ferric cyanide. To remove the excess amounts of the oxidizing chemical, the sample was then passed through a small Sephadex G-25 column, equilibrated with 10mM phosphate buffer, pH 7.0, 5mM NaCl and 4% glycerol. The crystallization conditions used for cyano form did not produce protein crystals in the met form. Crystallization conditions were rescreened and obtained after screening a variety of precipitants (Hampton Crystal Screening I, II and UW-144) and additives at room temperature using the hanging-drop vapor-diffusion method. The same crystallization condition can also yield similar sized reddish crystals using small batch method.

Crystals of HemAT sensor domain in met form were grown from 25% PEG1500, 10mM Hepes, pH 7.4 with 100mM BaCl₂ at room temperature. When adding BaCl₂ as additive in the solution, it produces white precipitants largely due to the interactions of phosphate buffer and Ba ion. Both of them are necessary to produce HemAT sensor domain crystals in met form.

The morphology of these crystals is a very thin square plate, dramatically different from the cyano form. The single plate crystal has size about 0.1x0.1x0.008mm. The crystal has been examined using synchrotron radiation and diffracted to a resolution of about 3.5 Å. The space group is believed to be a new one with unit cell dimensions of a=46Å, b=52Å, c=140Å. Attempts to optimize the size of the crystals have not been successful.
Figure 5.3 Crystals of the HemAT sensor domain in met form. Typical dimensions of a single crystal are approximately 0.1x0.1x0.008mm. PEG1500 was used as precipitant in the crystallization condition.

5.2 Crystallographic data collection

Multiple-wavelength anomalous dispersion (MAD) data were collected at the iron edge using flash cooling techniques at the Advance Photo Source (APS) BioCARS beamline using ADSC Quantum-4 detector. The cryo-protectant was made of 2M sodium citrate, 20% glycerol and 1M lithium sulfate. The X-ray fluorescence scan was taken near iron edge (7.12496 kev). Though the absorption signal was small as we expect from calculation in Chapter 2, the X-ray absorption scan readily identified the anomalous scatterer as iron atom. Three data sets were collected at peak, inflection point and remote wavelength. Additionally, another native data set was collected after the MAD data collection from the same crystal. The native data was collected to 2.15Å resolution. The collected data were processed on site using Denzo and Scalepack (Otwinowski and Minor 1997). The summary of the data collection statistics is given in Table 3.1.
<table>
<thead>
<tr>
<th>Wavelength (Å)</th>
<th>Resolution (Å)</th>
<th>Reflection</th>
<th>Completeness (%)</th>
<th>R_{sym} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7416</td>
<td>2.85</td>
<td>8218 (649)</td>
<td>96.5(78.8)</td>
<td>0.103 (0.223)</td>
</tr>
<tr>
<td>1.7400</td>
<td>2.90</td>
<td>7902(682)</td>
<td>97.4(87.1)</td>
<td>0.112 (0.333)</td>
</tr>
<tr>
<td>1.6712</td>
<td>3.10</td>
<td>6467(552)</td>
<td>97.0(85.8)</td>
<td>0.113 (0.217)</td>
</tr>
<tr>
<td>1.0000</td>
<td>2.15</td>
<td>18610(1588)</td>
<td>96.1(84.1)</td>
<td>0.067(0.220)</td>
</tr>
</tbody>
</table>

**Table 5.2** Statistics of MAD data collection. The wavelengths used in the data collection and their corresponding resolution are listed. Numbers in parentheses are for the highest resolution shell.

The solvent content analysis of HemAT sensor protein crystals showed that Matthews coefficient would be 4.12 and the solvent content 69.9% if there were one subunit in one asymmetric unit. The solvent content would be 39.8% and Matthews coefficient were 2.06 for two subunits in one asymmetric unit. It is therefore reasonable to assume that two subunits existed in one asymmetric unit. The space group was determined to be P2_1_2_1_2_1 based on systematic absences with cell dimensions a=50.00Å, b=80.12Å, c=85.95Å.

### 5.3 Structure determination and refinement

There is limited sequence similarity between the HemAT sensor domain and other members of the myoglobin family. However, all attempts to solve the
sensor domain structure using molecular replacement failed. The iron in the heme group was used as a candidate anomalous scatterer to collect a MAD data set. The X-ray absorption scan readily showed the iron atom, and then MAD data were collected at peak, inflection point and remote wavelengths. The data were collected to 2.8Å resolution (Table 1) near the iron edge. A native data set was also collected to a resolution of 2.15Å from the same crystal after MAD data collection (wavelength at 1.00 Å).

The initial heavy atom positions were found by SOLVE (Terwilliger and Berendzen 1999) from the three-wavelength MAD data (peak, inflection point and remote). However, the map produced from SOLVE was noisy and did not reveal any recognizable protein features. The model built by RESOLVE based on this map only gave a few candidate helices. Another approach was taken to find the heavy atoms sites using programs from the CCP4 package (Collaborative Computational Project Number 4 1994). Four data sets including native data were processed and scaled with program TRUNCATE and SCALEIT. Direct methods were used to find the heavy atoms using RANTAN. The identified heavy atom sites were refined with MLPHARE. Two heavy atoms were found in the asymmetric unit. It was found that the sites found from CCP4 and from SOLVE were the same, being symmetry related. There were two heavy atom sites corresponding to the iron positions from two subunits in one asymmetric unit. I then found that several helices produced by RESOLVE bore non-crystallographic two-fold symmetry. Along with homology model building based
on other heme proteins, the map was improved after density modification taking
the initial phases from Mlphare. The protein chain was then traced from the map.

Refinement was carried out with SHELXL for the liganded structure and
CNS for the unliganded structure with the liganded structure as model (Sheldrick,
Schneider et al. 1997; Brunger, Adams et al. 1998).

For the unliganded structure, the model from the liganded structure was
adjusted to fit a simulated annealing omit map prepared with phases from the
liganded structure. It was refined initially with simulated annealing protocols
implemented in program CNS. The final several cycles of refinement of the
unliganded structure was also completed using SHELXL. The refinement
methods yielded nearly identical structures. There were no restraints applied to
between ligand binding pocket, ligand and heme group during the refinement in
order to remove the bias. Model building was performed using program
XtalView (McRee 1999). (It should be noted that the distal pocket of subunit A
of the unliganded structure is completely devoid of density at the ligand binding
site. Subunit B, however, shows a small amount of residual density that lies too
close to the relaxed position of Leu92 to represent significant ligand occupancy.)
The summary of the data collection and refinement statistics is listed in Table 3.2.
<table>
<thead>
<tr>
<th></th>
<th>Liganded</th>
<th>Unliganded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>( P2_12_12_1 )</td>
<td>( P2_12_12_1 )</td>
</tr>
<tr>
<td>Unit cell parameters Å</td>
<td>( a=49.99, \ b=80.12, ) c=85.94</td>
<td>( a=49.45, \ b=79.72, ) c=85.14</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>27 ( \sim ) 2.15</td>
<td>27 ( \sim ) 2.71</td>
</tr>
<tr>
<td>Observations</td>
<td>83,301</td>
<td>55,718</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>18,722</td>
<td>9,226</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>96.0 (83.7)</td>
<td>95.1 (74.0)</td>
</tr>
<tr>
<td>( R_{\text{sym}} ) (%)*</td>
<td>6.3 (21.8)</td>
<td>9.8 (21.6)</td>
</tr>
<tr>
<td>( I/\sigma ) *</td>
<td>16.58 (5.35)</td>
<td>18.38 (6.80)</td>
</tr>
<tr>
<td>Refinement statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R_{\text{crystal}} ) (%)</td>
<td>19.70</td>
<td>20.80</td>
</tr>
<tr>
<td>( R_{\text{free}} ) (%)</td>
<td>25.78</td>
<td>32.90</td>
</tr>
<tr>
<td>Model statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>2,666</td>
<td>2,621</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>204</td>
<td>185</td>
</tr>
<tr>
<td>rmsd bond lengths (Å)</td>
<td>0.007</td>
<td>0.009</td>
</tr>
<tr>
<td>rmsd bond angles (°)</td>
<td>1.324</td>
<td>1.100</td>
</tr>
</tbody>
</table>

**Table 5.3** Statistics of Data Collection and Refinement of HemAT sensor domain.
6 STRUCTURE ANALYSIS OF SENSOR DOMAIN

6.1 Overall structure

The crystal structure of the sensor domain of HemAT (amino acids 1-178) has been determined at 2.15Å for a cyano-liganded form and at 2.7Å for an unliganded form. There is two-fold non-crystallographic symmetry for the two subunits in one asymmetric unit. Two-fold symmetry constraints were applied to the subunits during the initial refinement, but removed when the R-factor plateaued at 30%. The sensor domain was revealed to be a homodimeric protein in which the dimerization interface forms a four-helical bundle as a core and this core is closely packed with the remaining helices.

Like many isolated dimeric sensor domains in dilute solutions HemAT is monomeric, but it is dimeric in crystals. The weaker dimerization is presumably due to the loss of dimer-forming interactions from the methyl-accepting domain. For example, the sporulation response regulatory protein Spo0B, the cytoplasmic domain of osmosensor protein EnvZ and *E. coli* PhoQ sensor domain show similar behavior (Hidaka, Park et al. 1997; Zhou, Madhusudan et al. 1997; Varughese, Madhusudan et al. 1998; Tomomori, Tanaka et al. 1999; Lesley and Waldburger 2001) (and C. Bingman personal communication). The dimerization interface of the sensor domain of HemAT has a buried surface area comparable to "standard-size" interfaces found in other protein complexes (Lo Conte, Chothia et al. 1999). Gel filtration chromatography of the sensor domain of
HemAT demonstrated that the higher molecular weight dimeric form of HemAT sensor domain is present, but not dominant in dilute solution.

Despite these changes, the sensor domain subunit maintains a classic globin fold. There are eight α-helices with one extended chain at the N-terminus of each subunit. In order to assist the comparison with other proteins in the globin family, helical segments in the sensor domain structure have been named according to the nomenclature of globin classic fold. It has one helix before A helix, which we name the Z helix. Keeping using such convention, the sensor domain lacks a D-helix, so the helix arrangement in HemAT sensor domain is Z, A, B, C, E, F, G and H. Partial electron density of the N-terminal structure is observed in A subunit of our dimer, but not in the B subunit, due to disorder. The final liganded structure contains 169 residues for subunit A (residue 10 to 178), 158 residues for subunit B (residue 21 to 178), one heme group and one cyanide ligand for each subunit, and 204 water molecules. The unliganded structure has 168 residues for subunit A (residue 11-178), 156 residues for subunit B (residue 23 to 178) with 186 water molecules, and one heme group for each subunit.
Figure 6.1 The molecular structure of the HemAT sensor domain represented with ribbon diagrams. (A) Stereo view of the structure. The signaling domain would be located further down on the page. (B) Top view showing the flanking of the core helices, G and H, by the rest of the molecule. The helices are labeled
to corresponding to nomenclature of globin-fold. Subunit A is colored cyan and subunit B is colored yellow.

6.2 Comparative analysis of HemAT in two ligation states

The overall structure of the unliganded form is very similar to the liganded structure. The structural superposition was carried out using the ESCET program (Schneider 2002). The overall root mean square (rms) difference in the Cα atom position of the liganded and unliganded forms is 0.42Å, when the two structures are superimposed using residues 25-175 in subunit A and residues 25-175 in subunit B. The magnitude of this difference is greater than the individual coordinate errors estimated from a Luzzati plot (mean RMSD 0.22Å for the liganded structure and 0.33Å for the unliganded structure). The rms difference for the A subunits between the liganded and unliganded forms is smaller than that for the B subunits. The rms difference between Cα atom positions in the A subunits of the liganded and unliganded structures is 0.37Å versus 0.48Å for the B subunits. This suggests that the two subunits of the HemAT sensor domain undergo different rearrangements in the transition from the liganded to the unliganded states. We also see the heme group in the A subunit shift by 0.19Å, whereas the B subunit shifts by 0.49Å.
Figure 6.2 Error-scaled difference distance matrix (DDM) plots of the HemAT sensor domain. (A) DDM plot of subunit A versus subunit B within the liganded structure. (B) DDM plot of subunit A versus subunit B within the unliganded
structure. The result shows that there is less dimeric symmetry in the unliganded structure. (C) DDM plot of the liganded versus unliganded structures. The changes in distances smaller than 0.7Å are omitted. The color gradient indicates differences between 0.7Å and 1.4Å, where red represents expansion and blue contraction. The results show more dramatic changes in the subunit B, representing asymmetrical effects of the ligand departure.

In the second analysis, a least squares method was used to superpose subunits A and B within the dimer for each of the liganded and unliganded HemAT structures (residues 25-175) in order to reveal any intra-subunit differences. The rms difference for Cα atom positions for the liganded form is 0.39Å, and rms difference for all atoms is 1.05Å. The rms differences between the two subunits in the unliganded form shows larger differences than that in the liganded structure, 0.47Å difference for Cα atoms and 1.22Å for all atoms in the structure. This evidence indicates that the HemAT sensor undergoes small but measurable conformational changes within the dimer when ligand is depleted from the liganded form.

To further characterize the details of the differences, an objective comparison of the liganded and unliganded structures was carried out using an
error-scaled difference distance matrix method (Schneider 2002). Comparison was made not only between liganded and unliganded dimeric structures, but also carried out between the two individual subunits of each dimeric structure.

These plots produced by this method are advantageous because they do not depend on the superposition of one structure upon another, but compare internal distances between atoms. From the difference distance matrix plots comparing A and B subunits, the liganded structure has smaller differences between its two subunits than that of the unliganded form (Figure 6A and B). The plot of the liganded structure is basically featureless and suggests that its subunits have nearly identical conformation. However, the plot comparing subunits within the unliganded structure shows significant features. For instance, the residues from Leu73 to Ser77, which connecting helices B and C, experience greater than 1Å differences between subunit A and subunit B. These results show that the liganded structure of HemAT sensor has a more symmetrical organization than the unliganded structure.

Moreover, significant differences are observed in the plot generated based on the difference distance matrix of Cα atoms of the liganded and unliganded dimers (Figure 6.2C). The plot shows that differences between the B subunits of each structure are larger than that in the A subunits. The largest displacements are observed in the BC corner and the F and G helices. All of these variations
occur near the heme pocket, which might be expected since this is where the ligand is bound.

Only modest domain motions are observed in going from the liganded to the unliganded states. The largest change is that the F helix (residue 112-129) on the subunit B shows a 14 degree rotation with a 0.186 Å translation, which is detected with the help of program DynDom (Hayward, Kitao et al. 1997). This motion is likely responsible for the dramatic orientation change of proximal His123 residues.

A more detailed examination has been carried out on the dimerization interface with a superposition of two structures. The G and H helices of the two subunits form an anti-parallel four helical bundle. The H helices are likely continuous with the extended helical structure of the signaling domain of HemAT. As stated above, the rms differences for A subunits between two structures are smaller than that for B subunits, but both G helices show larger displacement than H helices in going from the unliganded to liganded structures. In Figure 6.3, the directions of displacements of the Cα atoms larger than 0.25 Å are shown with arrows. Their directions point from the unliganded to the liganded form, and their amplitudes are exaggerated by a factor of 25, because of the small size of the shifts.
Figure 6.3 Ca trace of the G and H helices at the dimerization interface. The unliganded structure is colored in green and the liganded structure in yellow. These difference vectors for G and H helices are shown as red arrows with a threshold of 0.25 Å. (A) Stereograph of side view of the helices. (B) Stereo view
of looking down into the cell. (C) Schematic representation of helical motions. Left is the averaged difference vectors were projected onto a line parallel to the H helices. Right is the same vectors projected onto a plane perpendicular to these helices are shown. Helices of subunit A (GA, HA) are colored in yellow, and these of subunit B (GB, HB) are in green. The numerical values for each helix are shown with the estimated standard deviation (0.04Å) for these helices based on error propagation of individual coordinate errors.

These difference vectors for each G (133-151) and H helices (157-175) were summed on main chain atoms (N, C, O, Cα) and averaged per atom. The averaged difference vectors of each helix then were projected onto a plane parallel to the H helices and onto a plane perpendicular to these helices. The estimated standard deviations of the average atomic motions is 0.04Å, which were calculated using the individual error estimates of each coordinate propagated assuming independent and random uncertainties according to Taylor (Taylor 1982).

In the unliganded structure, both G helices exhibit downward movement relative to the H helices (away from the presumed location of the membrane). The magnitude for the averaged difference vectors of 76 atoms of 19 residues is 0.17Å ± 0.04Å for helix G of subunit B, and 0.12Å ± 0.04Å for that of subunit A
on this projection. Helix H shows small upward movement in subunit B (0.04Å ± 0.04Å) and nearly no upward or downward movement in subunit A (Figure 6.3 C). Overall, the displacement between G and H helices in B subunit is more significant than that occurs in A subunit, even though the average displacement of individual atoms is small. This is consistent with the observation of a larger displacement of the heme in the B subunit and the increase in asymmetry going from liganded to unliganded forms. Not only do these helices show a small translational movement, but also a small rotational movement, which is visualized with difference vectors projected on a plane perpendicular to these helices. Subunit B has a tendency to have small rotations relative to the position of subunit A. These rotations are only 1 to 3 estimated standard deviations in magnitude, but still measurable.

The helical displacements of the HemAT sensor domain are small, but not much smaller than the results obtained for Tar ligand binding domain, which shows an average displacements of 0.5Å (Milburn, Prive et al. 1991; Yeh, Biemann et al. 1996). Biochemical studies of the cytoplasmic domain of the Tar receptor also suggest that the conformational changes involved are smaller than the flexibility allowed by an artificially introduced disulfide bond (Bass and Falke 1999). These results indicate that displacements in the signaling process are small and varied, but they still produce a downstream signal.
6.3 Ligand binding site

The C and E helices form one side of the portion of ligand binding site with the B helix completing the distal pocket (Figure 2). The F helix runs nearly parallel to the heme plane and therefore is close to the heme over large span. These features plus these close contacts of G and H helices make the HemAT sensor domain more compact than other 8-helix globin proteins.

The ligand-binding site of the sensor domain comprises Phe69, Tyr70, Ile83, Leu92, Thr95 and Leu96 on the distal side and His123 as the covalent attachment on the proximal side. The hydrophobic composition of this pocket is similar to the binding site in dimeric hemoglobin from *V. stercoraria*, but the cavity of HemAT is larger. The heme plane is buried deeply in the cleft so that there are more contacts with the protein matrix. The cyanide ligand is bound to heme iron and stabilized by the hydroxyl group of Tyr70 with a hydrogen bond. The cyanide ligand shows similar geometry in both binding sites of the liganded structure. Moreover, there is one distal pocket water within each binding sites of the liganded structure. The distal pocket water is stabilized by hydrogen bond to gamma oxygen atom on the Thr95 side chain. A hydrogen bond is also formed between a water molecule and the cyanide ligand.
Figure 6.4 Superposition of the ligand binding sites of the unliganded and liganded structures. The unliganded structure is colored in green and the liganded structure in yellow. (A) Left: side view of superposition of A sites; right: its top view of A sites. (B) Left: side view of superposition of B sites; right: its top view
of B sites. The results show the dramatic motion of Tyr70 and the heme in the B subunit. (C) Left: Electron density map of the B subunit of the liganded structure contoured at a level of 1 sigma and with dashed lines representing hydrogen bonds. Right: electron density map of B subunit of the unliganded structure contoured at level of 0.7sigma.

The most dramatic change in going from the liganded to unliganded form is that the side chain of Tyr70 of subunit B moves by 100° around its Cα-Cβ bond and shows some signs of disorder. In contrast both Tyr70 residues in the liganded structure point toward the heme group and are stabilized by hydrogen bonding to the ligand (Figure 6.4 C). Space filling models of the binding site in the unliganded structure indicate that there is no space available for either a ligand or water molecule in the pocket. Movements of side chains of Leu92, Leu96 and Phe69 into the cavity center contribute to the collapse of the pocket as the protein matrix relaxes to the unliganded state.

6.4 Comparisons with other globins

The structure-based sequence alignment of the HemAT sensor domain, hemoglobin from V. stercoraria and sperm whale myoglobin displays limited homology (Figure 2). The HemAT sensor domain has 25% sequence similarity to
hemoglobin from *V. stercoraria*, and only 15% similarity to sperm whale myoglobin. The proximal residue (His) on the F helix of the three proteins is absolutely conserved. The distal pocket residues of HemAT and hemoglobin from *V. stercoraria* are similar and are shifted more toward the B helix and away from the E helix than in myoglobins and vertebrate hemoglobins.

The structure of the HemAT sensor domain has a somewhat more compact conformation than other globin proteins, such as sperm whale myoglobin. The H and G helices run nearly parallel to each other making up the dimer interface. Furthermore, due to the lack of a D helix, the C and E helices are connected closely. Such structural rearrangements make the angle formed between E helix and heme plane smaller, and compress the crevice for holding the heme group.

However, the dimerization of HemAT is unique among these dimeric hemoglobins. Unlike any other, the dimerization interface of HemAT sensor domain is composed of G and H helices with part of Z helix and B-C corner. These dimeric contacts in hemoglobins have been reviewed in a previous report and the HemAT is again distinctive (Hargrove, Brucker et al. 2000). The buried surface area in HemAT dimers is among of the largest contact surface of homodimeric hemoglobins, similar in magnitude to hemoglobins from *Scapharca inequivalvis* and from *Caudina arenicola*, which are about 2000Å².
Figure 6.5 The structure-based sequence alignment of sperm whale myoglobin (Mb_Sw), HemAT sensor domain (HemAT Bs) and hemoglobin from Vitreoscilla stercorea (Hb Vs). (A) Ribbon diagrams of the three proteins in
their monomeric forms and the nomenclature for helical segments in each structure. (B) Alignment of three amino acid sequences based on their structures. Secondary structures are shown at the bottom. Identical residues among all three sequences are colored in green, and in yellow between two sequences. Similar residues are colored in cyan. The critical distal pocket Tyr and His and proximal pocket His residues are shown in red letters.

However, it is interesting that the protein is monomeric at low concentrations during its isolation. This may not be surprising since much of the coiled coil domain of the extended signaling region has been deleted.

The proximal pockets of the HemAT sensor domain, myoglobins and hemoglobins including that from *V. stercoraria* are very similar. Histidine on the F helix acts as an anchor to covalently bind the prosthetic heme group in the pocket, and hydrophobic side chains interact with the porphyrin ring around its perimeter. However, the heme group of the HemAT sensor domain is buried deeper than the heme groups in sperm whale myoglobin and hemoglobin from *V. stercoraria* (Figure 6.6). The calculated cavity has a volume of 995Å³ for the HemAT sensor domain, 772Å³ for Mb and 847Å³ for the hemoglobin from *V. stercoraria*, respectively. The HemAT protein still has no trouble binding small
gaseous ligands, however, the result is that the heme group in HemAT has less solvent exposed area than Mb and hemoglobins (Hou, Larsen et al. 2000).

Figure 6.6 Comparison of the distal pockets of the liganded HemAT sensor domain and sperm whale myoglobin. (A) The distal pocket of the liganded HemAT sensor domain. (B) The distal pocket of the liganded sperm whale myoglobin. The result shows that the deeper encapsulation of the heme in the sensor domain of HemAT.
6.5 Dimerization interface

6.5.1 HemAT dimer interface

The dimerization interface of the HemAT sensor domain comprises two long helices (G and H), part of the Z helix and the B-C corner from each subunit (Figure 1). The buried intrasubunit contact surface is about $1800 \text{Å}^2$ between these substructures, more than 20% of the whole surface area of a monomer. Two hydrophobic stretches of amino acids are observed on the G and H helices, from Leu140 to Ile145 on G and from Leu158 to Ile162 on H. The two hydrophobic patches make close contacts with each other and with the same regions on the partner subunit. These patches of hydrophobic interactions make up roughly one third of area of the homodimeric interface. There is also a large water cavity present at the interface. The cavity is about $293 \text{Å}^3$ and contains six water molecules, which form a hydrogen network with Thr166, Lys167 and Asn170. Thus, hydrogen bonding appears to be another energetic force that stabilizes the dimer.

One can compare the results with the dimerization interaction of Spo0B. Both the HemAT sensor domain and Spo0B form a similar four-helix hairpin structure at the dimer interface. But the interface of Spo0B is made primarily of hydrophobic residues, which contribute to the stability of the dimer in solution and crystal (Varughese, Madhusudan et al. 1998).
6.5.2 nsHb and other globin dimer interface

The biological function performed by a protein molecule is determined generally by its primary structure, and its uniquely predefined tertiary structure. However, the studies of structure and function relationship of protein oligomers give emphasis to the allosteric regulating role of the quaternary structure. For example, the hemoglobin proteins in different organisms exist monomer, dimer or oligomer in their native state (Bolognesi, Bordo et al. 1997; D'Alessio 1999).

Clearly, the dimerization interface of HemAT presents a unique type of composition and arrangement in the quaternary structure among the family of dimeric hemoglobins. The investigation of the structure and function properties of these dimeric hemoglobin has suggested that the interactions between each subunit is very crucial for the oxygen ligand binding (Shionyu, Takahashi et al. 2001). The factors include amino acid composition, hydrophobic v.s. hydrophilic, the position involved for the subunit interactions, the size of the surface area and the distance between two irons in the separate porphyrin rings (Kitto, Thomas et al. 1998; Shionyu, Takahashi et al. 2001). Unlike the monomeric form of hemoglobin, myoglobin, its kinetic behavior is largely influenced by the local environment of the ligand binding pocket, either electrostatically or sterically (Springer, Egeberg et al. 1989; Springer, Sligar et al. 1994; Olson and Phillips 1997). As the case for dimeric hemoglobin, the consideration of their ligand binding kinetics will also need to include the role of the quaternary arrangement played.
As already suggested in the nonsymbiotic plant hemoglobin, the inter subunit contacts may play a functionally important role in regulating the ligand binding kinetics for the specific type hemoglobin (Hargrove, Brucker et al. 2000). Larger surface area has also been observed for these hemoglobins displaying cooperative oxygen binding.

The structural organization involved in the dimerization interface varied from protein to protein. The subunit contact of nonsymbiotic plant hemoglobin (*Oryza sativa*) is made between CD region and partial G helix with surface area 554Å² (Hargrove, Brucker et al. 2000). The sea worm dimeric hemoglobin (*Urechis caupo*) has interface with both AB region and E helices from both subunits, its surface area about 520Å² (Kolatkar, Hackert et al. 1994). For the dimeric hemoglobins in clam (*Scapharca inaequivalvis*) and sea cucumber (*Caudina arenicola*), they have very similar quaternary and dimerization structural arrangement with E and F helices from two subunits and 980Å² surface area (Mitchell, Kitto et al. 1995; Pardanani, Gambacurta et al. 1998). For the α₁β₁ human hemoglobin, the interface is between Partial B, G, H helices from two subunits with 825Å² surface area.

The quaternary structure of hemoglobin from bacteria (*Vitreoscilla stercoraria*) shows an interesting structural feature in the interface, which is based on partial F and H helices from two subunits with about 434Å² of surface area (Tarricone, Galizzi et al. 1997). Several lines of experimental data are not consistent with each other on cooperative ligand binding, and are inconclusive at
present time. Nevertheless, the two ligand binding sites have different ligand binding properties (Tarricone, Galizzi et al. 1997).

For HemAT (*Bacillus subtilis*), both whole G and H helices of the two subunits involved in the formation of the dimerization interface in a form of an anti-parallel four helical bundle. Since the H helices are likely continuous with the extended helical structure of the signaling domain of HemAT, the interface may be even larger than we see on the truncated structure.

The buried area in HemAT dimerization interface has the largest contact surface among other homodimeric hemoglobins, two times larger than hemoglobins from *Scapharca inequivalvis* and from *Caudina arenicola*, which are about 900Å². However, it is interesting that the protein is monomeric at low concentrations during its isolation. This may not be surprising since much of the coiled coil domain of the extended signaling region has been deleted.

Based on the calculation of the contact surface area, it is not difficult to recognize that these cooperative hemoglobins have larger subunit interaction area. The relationship of the contact area to cooperativity in ligand binding is still not clear, so the question of how dimeric hemoglobins like HemAT with the largest contact (~2000Å²) has negative cooperativity, these with the middle size contact area (~1000Å²) has positive cooperativity and these with the small size contact area (~500Å²) has no cooperativity still remains unanswered.
<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Species</th>
<th>Surface area (Å²) *</th>
<th>Composition</th>
<th>Cooperativity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1VHB^e</td>
<td><em>Vitreoscilla stercoraria</em></td>
<td>445</td>
<td>Partial F, H helix</td>
<td>Yes/No</td>
</tr>
<tr>
<td>1ITH^d</td>
<td><em>Urechis caupo</em></td>
<td>510</td>
<td>AB region, E helix</td>
<td>No</td>
</tr>
<tr>
<td>1D8U^a</td>
<td><em>Oryza sativa</em></td>
<td>480</td>
<td>CD region, partial G helix</td>
<td>No</td>
</tr>
<tr>
<td>3HBB^b</td>
<td><em>Homo sapiens</em></td>
<td>825</td>
<td>Partial B, G, H helices</td>
<td>Yes</td>
</tr>
<tr>
<td>1HLB^f</td>
<td><em>Caudina arenicola</em></td>
<td>950</td>
<td>E, F helix</td>
<td>Yes</td>
</tr>
<tr>
<td>4HBI^c</td>
<td><em>Scapharca inaequalvis</em></td>
<td>995</td>
<td>E, F helix</td>
<td>Yes</td>
</tr>
<tr>
<td>1OR4</td>
<td><em>Bacillus subtilis</em></td>
<td>1800</td>
<td>G, H helices partial Z helix and BC region</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* [http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html](http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html)  
1D8U^a (Hargrove, Brucker et al. 2000)  
3HBB^b (Fermi, Perutz et al. 1984)  
4HBI^c (Pardanani, Gambacurta et al. 1998)  
1ITH^d (Kolatkar, Hackert et al. 1994)  
1VHB^e (Tarricone, Galizzi et al. 1997)  
1HLB^f (Mitchell, Kitto et al. 1995)  

**Table 6.1** Summary of dimeric hemoglobins
<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Residue Composition in the interface</th>
</tr>
</thead>
</table>
| 1VHBc  | A: Leu2, Gln4, Ile7, Glu69, Asn70, Pro72, Ala73, Leu75, Pro76, Asp131, Val132, Gln135, Val136, Asp139, Gln143  
|        | B: Leu2, Gln4, Ile7, Asn70, Pro72, Ala73, Leu75, Pro76, Asp131, Val132, Gln135, Val136, Asp139, Gln143 |
| 1ITHd  | A: Phe15, Leu16, Lys19, Gly20, Cys21, Leu22, Gln23, Ala63, Leu66, Thr67, Asn70, Asp73, Lys74, Asp77  
|        | B: Phe15, Leu16, Lys19, Gly20, Cys21, Leu22, Gln23, Ala63, Leu66, Thr67, Asn70, Asp73, Lys74, Asp77 |
| 1D8Ua  | A: Lys42, Glu45, Val46, Ala47, Pro48, Ser49, Asp115, Ala116, His117, Glu119, Val120, Phe123, Lys143  
|        | B: Lys42, Glu45, Val46, Ala47, Pro48, Ser49, Asp115, Ala116, His117, Glu119, Val120, Phe123, Lys143 |
| 3HHBb  | A: Glu27, Glu30, Arg31, Leu34, Ser35, Phe36, His103, Cys104, Leu106, Val107, Ala110, Ala111, His112, Pro114, Phe117, Pro119, Ala120, His122, Ala123, Asp126  
|        | B: Arg30, Val33, Val34, Tyr35, Pro51, Met55, Asn108, Val109, Val111, Cys112, Ala115, His116, Gly119, Lys120, Phe122, Thr123, Pro124, Pro125, Gln127, Ala128, Gln131, Lys132 |
| 1HLBf  | A: Tyr4, Lys29, Lys30, Arg53, Leu54, Asp64, Lys65, Arg67, Gly68, His69, Ile71, Ile72, Tyr75, Gln78, Asn79, Asp82, Gln83, Asn86, Asp88, Asp89, Cys92, Val93, Glu95, Lys96, Phe97, Val99, Asn100, Arg104  
| 4HBlg  | B: Tyr4, Lys29, Lys30, Arg53, Asp64, Lys65, Arg67, Gly68, His69, Ile71, Ile72, Tyr75, Gln78, Asn79, Asp82, Gln83, Asn86, Asp88, Asp89, Cys92, Val93, Glu95, Lys96, Phe97, Val99, Asn100, Arg104 |
|        | B: Asp33, Lys36, Gln37, Lys39, Met40, Val41, Arg42, Asn72, Leu73, His75, Glu76, Leu79, Pro130, Lys131, Trp132, Met134, Gly135, Ala136, Gln138, Glu139, Leu141, Leu142, Ile145, Asp146, Glu149, Thr153, Gln155, Gln156, Leu158, Leu159, Ile162, Lys163, Thr166, Lys167, Asn170, Gln173, Gln174, Leu177, Glu178 |

**Table 6.2** Detailed comparisons of the compositions of Hb dimer
Figure 6.7 Ribbons diagrams of six dimeric hemoglobins.
7 STRUCTURAL DETERMINATION OF INTACT HEMAT PROTEIN

7.1 Model of intact HemAT protein

Based on the sequence analysis shown previously, the signaling domain of HemAT shared 30% sequence identity to those of *E. coli* chemoreceptors. This information indicates that HemAT and *E. coli* chemoreceptors may adapt to the similar three-dimensional structure with the characteristic hepated repeat.

The crystal structure of the cytoplasmic domain of *E. coli* Tsr was determined to 2.6\AA{} resolution by Kim et al. in 1999 and is a dimeric protein with an elongated coiled-coil of anti-parallel helices (Kim, Yokota et al. 1999). The highly conserved domain is next to the loop, and residues within this region have very low temperature factors.

Based on the sequence analysis, previous crystallographic studies and the solved HemAT sensor domain, I propose a model that for *B. subtilis* HemAT.

The dimeric protein has two major domains about 260\AA{} in overall length. This sensor domain can detect oxygen by binding to the iron in the heme plane, causing a conformational change, which is transmitted to signaling domain of the receptor. The signaling domain of HemAT contains a slightly shorter coiled-coil of two anti-parallel helices than those found in the cytoplasmic domain of Tsr with 20 more residues in the C-terminal domain.
Figure 7.1 Ribbon diagrams of intact HemAT model at two orientations, based on the HemAT sensor domain and cytoplasmic domain of Tsr.
7.2 Diffraction data analysis

The intact HemAT protein was crystallized and diffraction data were obtained using synchrotron radiation (wavelength 0.9Å), and many efforts have been made to determine the structure at the present resolution.

The initial crystallographic analysis reveals that its space group belongs to P4₁2₁2 or P4₃2₁2 with unit cell dimensions of a=b=247Å, c=185Å.

<table>
<thead>
<tr>
<th>H</th>
<th>K</th>
<th>L</th>
<th>I</th>
<th>Sigma</th>
<th>I/Sigma</th>
<th>√</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1046.4</td>
<td>29.5</td>
<td>35.47</td>
<td>√</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>5</td>
<td>62.5</td>
<td>20.8</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>6</td>
<td>-34.7</td>
<td>25.3</td>
<td>-1.37</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>7</td>
<td>-25.8</td>
<td>29.1</td>
<td>-0.89</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1843.9</td>
<td>57.9</td>
<td>31.85</td>
<td>√</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>9</td>
<td>6.7</td>
<td>42.3</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>1426.8</td>
<td>80.2</td>
<td>17.79</td>
<td>?</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>11</td>
<td>-55.2</td>
<td>52.1</td>
<td>-1.06</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>13</td>
<td>116.5</td>
<td>65</td>
<td>1.79</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>14</td>
<td>3.4</td>
<td>76.1</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>15</td>
<td>1306.7</td>
<td>104.5</td>
<td>12.50</td>
<td>?</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>16</td>
<td>23283.2</td>
<td>533</td>
<td>43.68</td>
<td>√</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>17</td>
<td>1351.8</td>
<td>112.7</td>
<td>11.99</td>
<td>?</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>18</td>
<td>537.1</td>
<td>118.1</td>
<td>4.55</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>19</td>
<td>-103.1</td>
<td>164.6</td>
<td>-0.63</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>21</td>
<td>-212.3</td>
<td>121.6</td>
<td>-1.75</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>22</td>
<td>-171</td>
<td>140.6</td>
<td>-1.22</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>23</td>
<td>333.4</td>
<td>137.5</td>
<td>2.42</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>24</td>
<td>248.6</td>
<td>147</td>
<td>1.69</td>
<td>√</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>25</td>
<td>3764.1</td>
<td>186.8</td>
<td>20.15</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 7.1 Example of diffraction spots on 0,0,L axis.

The deduced space group was based on the examination of the systematic absences in the diffraction data. The following table was processed and scaled
with space group P422 in order to clearly demonstrate the absence of spots along the "l" axis. It is very possible to have a four-fold symmetry along the l axis, but these spots with high ratio of I/ SigmaI are still not clear.

The solvent content analysis with the crystallographic parameters in Table 6.2 reveals that for space group $P4_12_12$ or $P4_22_2$, it is reasonable to assume that there are at least three molecules in the asymmetric unit. Considering the crystallization condition of growing the crystal from less than 5% PEG6000 and 10% glycerol, it is not difficult to anticipate high solvent content for the crystal.

Despite the recent revisit of Matthews coefficient (Vm) and solvent content on 10471 protein crystal forms shows that the distribution for the Vm is about average value of $2.69 \text{Å}^3$/Dalton with a solvent content of ~47%. The initial estimated partial specific volume (psv) for proteins is 0.74 cm$^3$/g, which is in good agreement with original values (Matthews 1968; Kantardjieff and Rupp 2003).

<table>
<thead>
<tr>
<th>Number of HemAT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mw (kDa)</td>
<td>96</td>
<td>192</td>
<td>288</td>
<td>384</td>
</tr>
<tr>
<td>Vm($\text{Å}^3$/Da)</td>
<td>14.70</td>
<td>7.35</td>
<td>4.90</td>
<td>3.67</td>
</tr>
<tr>
<td>Solvent content</td>
<td>91.63 %</td>
<td>83.26 %</td>
<td>74.89 %</td>
<td>66.52 %</td>
</tr>
</tbody>
</table>

**Table 7.2** Analyses of Matthews Coefficient and Solvent Content
7.3 Patterson map of the diffraction data

As I described in Chapter 2, Patterson synthesis can yield valuable information on the positions of the atoms of a molecule. The mathematic formula represents the Fourier transform of the measured intensity data without the phase angle. It provides a map with all interatomic vectors, thus Patterson map will have total $N^2$ peaks in the unit cell with $N^2(N-1)$ off origin peaks.

The experimental Patterson map is computed using the diffraction data processed and scaled as space group of $P4_1$. (Different space groups have been tested to produce the experimental Patterson maps due to the ambiguity. The choice of this particular space is apparently first to examine)

The Patterson map shows several significant features that are very constructive to further investigation. Firstly, it reveals that three elongated density regions dominate in the Patterson spaces on the xy-section ($z=0$). Three distinct layers of density exhibit on the xz-section ($y=0$). By combining the self-rotation real space search which gives one extraordinary peak at ($\psi=14.257$, $\phi=0.000$, $\kappa=180.000$, larger than 5sigma) besides the symmetry related peak, one molecule should run parallel to z-plane and with 14.257° tilted along x- and y- axes (the number of molecule is dependent of the space group). The other useful information can also be extracted from the experimental Patterson map is the position of third (or second) molecule, which should run along the diagonal axis of unit cell.
Figure 7.2 Patterson map calculated with experimental data with space group \( P4_1 \).
7.4 MR solution and recognition of pseudo symmetry

After the preliminary analysis mentioned above (space group \(P4_12_12\) or \(P4_32_12\) with unit cell dimensions of \(a=b=247\,\text{Å}, \ c=185\,\text{Å}\)), the phase problem seems likely to be solved by molecular replacement. The traditional rotation function requires rotated Patterson maps to superpose, and the result is determined based on Patterson correlation. It is easily realized that the considerable elongated shape of intact HemAT (260 Å long), which traverse the single unit cell on two directions, makes this approach unrealistic, because the accuracy of potential solutions depends on the assumption that the cross vectors and self vectors are well separated within certain radius from the origin. The length of the molecule makes it impossible to provide a reasonable an integration sphere. The application programs, which rely on the rotational search first, for example CNS real space search, AMORE and Molrep, are unable to yield correct solutions.

Due to the low-resolution diffraction of the crystal and difficulty in recovering the solution from the initial molecular replacement, the brute force approaches have also been put to the test to tackle the problem even though no satisfactory solutions were produced. \textit{Ab initio} structure determination start with taking the diffraction data alone, for example a method proposed by Subbiah by treating all atoms as hard spheres as to position them in the related space group with the criteria of R factor (Subbiah 1991), and the improved approach with connectivity constraint as guidance to lead the movement of these atoms in the
real space developed in Phillips’ laboratory (Zou and Phillips 1997). Neither of these methods produced convincing results.

Following the clues from the experimental Patterson map, positioning the first molecule along a or b axis in real space can be achieved with direct rotation function search implemented in CNS.

Figure 7.3 Patterson map calculated with one HemAT with space group P4₁2₁2₂.
The method is to rotate the molecule and then calculate the structure factor for each new orientation. This is no approximation on inter atomic vector length and the Patterson correlation coefficient is estimated for the whole unit cell. It is superior to the traditional method and can enhance the likelihood of finding the correct solutions, even though at expenss of long computation time. The subsequent PC refinement of can also improve the overall orientation of the search model.

At this stage, the correlation coefficient between the experimental data and one HemAT molecule in an asymmetric unit of space group P4₁2₁2 was 0.668. The correctness of the position can also be noticed from the Patterson map computed solely with above findings.

The difficulty in finding the second HemAT molecule in P4₁2₁2 space group has prompted us to reexamine the diffraction data and the experimental Patterson map. The positioning second HemAT molecule in P4₁2₁2 along the diagonal axis inevitably produces penetrated molecules that are completely impossible and indicates an error somewhere. The recognition of the possible pseudo-symmetry created by the homodimeric structure of HemAT may account for the systematic absence on the “h” axis was a breakthrough step in deciding to degenerate the high symmetry space group to lower symmetry ones (P4₁ or P4₃) and leads to the possible localization of the third molecule in space group P4₁.

After positioning the third molecule, the R-factor is about 55%. The further refinement to improve the agreement by adjusting carefully selected rigid body segment refinement has reduced the R-factor to 47%. The quality of the
data and search model maybe the contributing factors in limiting the convergence of refinement.

7.5 Summary

The intact HemAT crystal only diffracts to 7Å, and the structure determination of intact HemAT is a difficult task. A correct solution has been found after successfully positioning three molecules with a combination of efforts, even though structure refinement is another challenging mission.
Figure 7.4 Patterson map calculated with three HemAT molecules with space group P4₁.
Figure 7.5 Molecular solution of HemAT structure in space group P4₁ projected onto a-b plane.
8  STRUCTURE AND FUNCTION RELATIONS

8.1  Symmetry Breaking in Signal Transduction

From our analysis of the liganded sensor domain structure above, the two subunits of the HemAT sensor domain have similar structures based on the distance difference matrix plot. In contrast, large variations are observed between the two subunits of the unliganded structure. This indicates that the liganded structure has a more symmetrical organization than the unliganded structure. Besides the small helical shifts and rotations of the four-helical bundle in the dimerization interface, the most significant changes between liganded and unliganded structures occur at the ligand-binding pocket. Subunit A maintains a similar arrangement of side chain packing, but subunit B undergoes dramatic changes when HemAT sensor domain relaxes to an unliganded structure from the liganded state.

In the ligand-binding pocket of subunit A of the unliganded structure, the side chains of Phe69, Leu92, Thr95 and Leu96 move slightly into the distal pocket to fill the empty space left by the ligand. But for subunit B of the unliganded structure, these side chains not only move into the ligand-binding pocket to fill the space left by ligand, but also the heme plane tilts by an angle of 9 degrees towards the F helix. These structural rearrangements lead to a flipping of the side chain of Tyr70 up and out of the pocket. The side chain flipping only
occurs in the subunit B of the unliganded structure, which results in a dramatic
disruption of the symmetry between the two subunits. Other increases in asymmetry of the side chain packing are also seen.

The dimerization interface reveals small but perceptible conformational changes that may be related to chemotaxis signal transduction. Both G helices of the unliganded structure have a downward movement with respect to the equivalent segments of the liganded structure. The movement has a greater extent for the G helix of subunit B than that for subunit A (Figure 4A, C). For the H helix neither shows a significant movement. Therefore, the resultant displacements of the G helices of the unliganded structure show downward movement with respect to H helices on each subunit, with the G helix on subunit B showing somewhat more displacements.

When the averaged difference distance vectors are projected onto the plane perpendicular to middle line of dimerization interface, we can see the relative movement of individual helices. All motions are greater than the estimate standard deviation of 0.04Å, although sometimes not by much. The helical core of the unliganded structure has a tendency to make counterclockwise rotational movements if we view down the plane perpendicular to middle line of dimerization interface.

So when the HemAT sensor domain undergoes the transition from unliganded to liganded state, the conformational changes at the dimerization
interface will cause a downward shift of the G helices. If one takes the G helices as the reference, H helices will have an upward movement. (Since it is the H helices that continue with the elongated signaling domain, it is not obvious what to consider the reference point.) The rotational movements of the four helices have different magnitudes, which could cause some "unwinding" of the coiled-coil signaling domain.

Our structures contrast with the observations that conformational changes in the ligand binding domain of aspartate receptor occur only on one helix (Milburn, Prive et al. 1991; Yeh, Biemann et al. 1996). However, based on studies of the ribose receptor (Trg), some mutational substitutions in the periplasmic domain (R71H, S72L, I78T, or Q79L) can induce the transmembrane signaling stronger than does wild type (Yaghmai and Hazelbauer 1992; Beel and Hazelbauer 2001; Beel and Hazelbauer 2001). These mutations mimic the presence of two ligands bound to the receptor. This finding suggests that Trg and perhaps other chemoreceptors are able, under extreme conditions, to change the conformations of two helices (Beel and Hazelbauer 2001; Beel and Hazelbauer 2001). Thus, it is very possible that the unliganded structure of the HemAT sensor domain represents a similar state as the occupancy-mimicked Trg receptor described above.

Based on our and others' structural investigations, two general hypotheses can be advanced for relating conformational changes in the sensor domains to
changes in the signaling domains of these receptors, and hence kinase activity regulation. In the first model, the asymmetry of sensor domain subunit is propagated to the signaling domains. When HemAT is unliganded or the aspartate receptor liganded, the asymmetrical structures of both sensor and signal domains are maintained by asymmetrical helix propagations, inhibiting kinase activity, perhaps by limiting kinase binding. The second model assumes that the asymmetry is present only in the sensing domains, and the signaling domains have symmetrical organization in all states. Any effects in the sensor domain would be transmitted through vectorial or rotational movements of symmetrical coiled coils to signaling domains. The structures of the isolated sensor or signaling domains cannot differentiate these two models.

8.2 O2 and CO ligand binding reactions

We can gain some understanding on the biological function of HemAT from the spectral and kinetics data in our study. Based on the homology modeling, N-terminus of HemAT has a myoglobin-like fold which distinguishes it from the other two heme-based oxygen sensors, FixL and Dos, whose heme-containing domains have a PAS fold (Delgado-Nixon, Gonzalez et al. 2000; Gong, Hao et al. 2000). These findings suggest that nature may evolve different molecular mechanisms for detecting oxygen. One type sensor, just like HemAT, adopts the myoglobin-like fold, which can bind the specific gaseous molecule,
whose presence can cause some conformational change and such structural information can be passed onto C-terminal signaling domain. The homology of HemAT was also discovered from the Archaea, which suggest that oxygen sensing is likely the ancient function of myoglobin and hemoglobin. The primary function of these ancient myoglobins may have been to detect environmental diatomic oxygen, and later this function evolved to remove oxygen by reaction with NO to store it and transport it for aerobic metabolism.

Negative cooperativity and heterogeneity has been observed in the ligand binding reactions of Tar and Tsr (Biemann and Koshland 1994; Lin, Li et al. 1994). The appearance of half-the-site reactivity has often been seen in enzymatic catalysis. Various molecular mechanisms have been proposed with some structural investigation (Koshland 1996; Anderson, O'Neil et al. 1999; Peterson and Smith 1999). Site-directed mutagenesis and structure studies have been carried out to investigate possible negative allosteric mechanisms of the aspartate receptor of *S. typhimurium* and *E. coil* (Kolodziej, Tan et al. 1996; Yeh, Biemann et al. 1996; Yu and Koshland 2001). These studies revealed that a key residue Ser-68 plays a crucial role in the regulation of cooperativity in the aspartate receptor. Ligand binding at one site triggers downward shift of one α helix that breaks the symmetry of whole structure and destroys the second ligand binding sites (Biemann and Koshland 1994; Yu and Koshland 2001). Thus, negative cooperativity might be an intrinsic structural property of this class of
chemoreceptors. However, positive cooperativity has been reported for both Tar and Tsr receptors when the receptor–kinase complexes were examined in kinase inhibition assay *in vitro* (Bornhorst and Falke 2000; Li and Weis 2000). These results could readily explain why bacteria are sensitive to both low and high ligand concentrations. Negative cooperativity increases the responsiveness of these chemoreceptors over a wide range of ligand concentrations. The decreased sensitivity in chemoreceptor could be compensated by the positive cooperativity of its histidine kinase. The broad range of sensitivity of these receptors has been demonstrated on an investigation of the robustness of chemotaxis network (Barkai and Leibler 1997; Yi, Huang et al. 2000). Combined with positive cooperativity of receptor–kinase complexes to ligand binding, we believe that such molecular machinery may reveal new mechanism on signal transduction of chemotaxis.
9 BIOLOGICAL IMPLICATIONS

9.1 Molecular mechanism of chemotaxis in *B. subtilis*

The chemoreceptors found in *E. coli* and *B. subtilis* have very high amino sequence similarity within their cytoplasmic signaling domains, especially within the highly conserved domain (Le Moual and Koshland 1996; Hou, Larsen et al. 2000). In addition, they have the characteristic seven-residue repeat of coiled coils. This evidence suggests that these chemoreceptors probably adopt a similar three-dimensional architecture.

However, there are some significant differences at the receptor level between two microorganisms, *B. subtilis* and *E. coli*. Six chemoreceptors found in *E. coli* are transmembrane proteins. Based on their sequence analysis and modeling studies, most of them very likely carry 260Å long cytoplasmic domains and form about 245Å long four helical bundles (Kim, Yokota et al. 1999). Unlike *E. coli*, there are four types of chemoreceptors that have been discovered in *B. subtilis*, type 1 chemoreceptors have two transmembrane domains and large periplasmic domain (about 230 residues), including mcpA, tlpA, mcpB, tlpB and mcpC. Type 2 chemoreceptors also have two transmembrane domains, but with a small periplasmic domain (about 170 residues). TlpC and yvaQ are in this category. Type 3 chemoreceptors have only one transmembrane domain but with
about 150 residues in periplasmic space, which includes yoaH. All of proteins in these three categories have nearly the same length for the cytoplasmic domain, about 300\AA, which is 40\AA longer than that of E. coli chemoreceptors (Kim, Yokota et al. 1999). Type 4 chemoreceptors are cytoplasmic proteins, including HemAT and yfmS. Taking together with recent findings that chemoreceptors tend to form a planar network near the inside surface of the cell membranes to enhance the signaling process, a model has been constructed for the relative positions of chemoreceptors of Bacillus subtilis (Figure 9.1) (Maddock and Shapiro 1993; Bray, Levin et al. 1998; Ames, Studdert et al. 2002). The implication is that the heme sensor domain of HemAT would sit about 40\AA inside the membrane sensing oxygen after it has diffused through the membrane and over a small distance.
Figure 9.1 A model of the four known types of chemoreceptors found in *B. subtilis*. Type 1 is colored in purple; type 2 in green; type 3 in cyan and type 4 in yellow. The lipid bilayer is colored in blue and gray. The cytoplasmic domain of *B. subtilis* proteins is about 300Å long, which is 40Å longer than the cytoplasmic domain of *E. coli* chemoreceptors. (A) Side view of these chemoreceptors. (B) Bottom-up view from cytoplasm to periplasmic space showing how the type 4 chemoreceptors are likely to be arranged in the planar network of other receptors.
9.2 Biological implications on the strategy in chemotaxis

The crystal structure of the HemAT sensor domain has the characteristic globin fold, though it only has very limited amino acids sequence homology to other hemoglobin-like proteins. It is unlike the heme-based oxygen sensors, FixL and Dos, whose heme-containing domain adopts a PAS fold (Stock 1997; Fabret, Feher et al. 1999).

The cavity holding heme group is larger and deeper than other globins and nearly all of heme group is buried within the protein scaffold. The homodimeric interface is composed of the entire G and H helices with partial structures from Z helix and BC corner. Thus the dimerization interface is among the larger ones seen for homodimeric hemoglobins. This also further demonstrates that hemoglobin-like proteins have evolved in such a way that the dimerization interface can be located in a variety of regions on the protein surface (Hargrove, Brucker et al. 2000).

Based on the results of the kinetics and equilibrium binding experiments that the ligand binding is biphasic for HemAT, it suggests that HemAT might utilize heterogeneity or negative cooperativity to expand the dynamic range for detecting the diatomic oxygen and transferring the structural information to the downstream histidine kinase. This speculation includes the evidence obtained for other chemoreceptors, like Tar and Tsr, where it has been suggested that negative cooperativity and heterogeneity is an important part of the molecular mechanism
for the ligand binding reactions (Biemann and Koshland 1994; Lin, Li et al. 1994; Koshland 1996).

Along with the observation that the liganded form of HemAT is more symmetrical than the unliganded one, we propose the following model for the signaling transduction by HemAT. The unliganded form of HemAT is in an energetically unfavorable form, not unlike the T state in hemoglobin. When environmental oxygen molecules diffuse into cytoplasm of *B. subtilis*, one subunit of the unliganded HemAT will bind one $O_2$ molecule. But, unlike hemoglobin, the ligand binding affinity of the second subunit is decreased due to structural arrangement of first binding event. Thus the partially liganded HemAT will exist often. As organisms move along the oxygen gradient, they would meet more and more oxygen molecules. Due to negative cooperativity or heterogeneity, it could take several orders of magnitude of increasing $O_2$ concentrations to saturate the second subunit, thus allowing a wide dynamic range of responses. Other regulatory mechanisms may also play an important role in desensitizing partially liganded HemAT, like the methylation and demethylation processes. The "design" of the HemAT structure seems elegantly to satisfy both sensitivity and robustness requirements for chemotaxis.
<table>
<thead>
<tr>
<th><strong>B. subtilis</strong></th>
<th><strong>E. coli</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Repellent binding</td>
<td>Attractant binding</td>
</tr>
<tr>
<td>inhibit</td>
<td>inhibit</td>
</tr>
<tr>
<td>CheA kinase activity lower</td>
<td>CheA kinase activity lower</td>
</tr>
<tr>
<td>phosphorylation of CheY strengthen</td>
<td>phosphorylation of CheY strengthen</td>
</tr>
<tr>
<td>the default CW to CCW</td>
<td>the default CCW rotation</td>
</tr>
<tr>
<td>Tumbling</td>
<td>Smooth swimming</td>
</tr>
</tbody>
</table>

**Figure 9.2** Revised representation of bacterial signal transduction pathway of *B. subtilis*. The default behavior of *E. coli* is smooth swimming, which is regulated by phosphorylated CheY. The default swimming-behavior of *B. subtilis* is tumbling, which is also regulated by the level of phosphorylated CheY. When changing the attractant in *B. subtilis* to repellent, the whole events along the signal transduction cascade are exactly similar to that on the *E. coli* pathway.

Our structural evidence also readily explains the differences between the signaling transduction pathways between *E. coli* and *B. subtilis*. When changing the attractant in *B. subtilis* to repellent, the whole events along the signal transduction cascade are exactly similar to that on the *E. coli* pathway. So due to the opposite default swimming behavior, the information flow must be reversed.
at receptor level. Our study on HemAT just gives exactly structural evidence to explain the structure and function relationship on the chemoreceptors.

The *E. coli* sensor goes from a symmetrical to an asymmetrical structure, inhibiting the downstream histidine kinase upon ligand binding (Milburn, Prive et al. 1991; Yeh, Biemann et al. 1993; Biemann and Koshland 1994; Yu and Koshland 2001). In contrast, the *B. subtilis* sensor goes from an asymmetrical to a more symmetrical form on ligand binding, activating the downstream histidine kinase. Thus although the connection between structural symmetry of receptor and kinase activation is conserved, the information switch is reversed at the level of the receptor, consistent with previous null mutation rescue studies.
APPENDIX

Protein Information:
Protein Sequence:
432 Amino Acids
Molecular Weight 48766.49Da

MLFKKDRQKE TAYFDSNGQ QKNRIQLTNK HADVKQLKM VRLGDAELVV
LEQLQPLIQE NIVIVDAFY KNLDHESLM DIINDHSSVD RLKQTLPKH
QEMFAGVIDD EFIEKRNRIA SIHLRIGLLP KWMGAFQEL LLSMIDIYEA
SITNQFELLK AIKATTKILN LEQQLVLEAF QSEYNQTRDE QEEKKNLLHQ
KIQETSQGIA NLFSETSRSV QELVDKSEGQ SQASKAGTVT SSTVEEKSIQ
GKELEVQQK QMNKIDTSLV QIEKEMVLDV BIAQQIEKIF GIVTGIAEQT
NLLSLNASIE SARAGEHKG PAVVANEVRK LSEDYKKTVS TVSELVNTNTN
TQINVSKHI KDVNELVSES KEKMTQINRL FDEIVHSMKI SKEQSGKIDV
DLQAPLGGLQ BVGRAWHVA ASVDSLVILT BE

Notes: Glycerol Stock #1226
Protein Information:
Protein Sequence:
438 Amino Acids
Molecular Weight 49607.35 Da

MLFKKDRKQE TAYFSDSNGQ QKNRQLTNK HADVKKQLKM VRLGDAELYV
LEQLQPLIQE NIVNIVDAFY KNLDDSLLM DIINHSSVD RLIKQTLKRHI
QEMFAGVIDD BFIEKRNRIA SIHLRIGLLP KWYMGAQEL LLSSMIDIYEA
SITNQQCLLLK AIKATTKILN LEQQLVLQAF QSEYNQTRDE QEEKNNLHG
KIQTSGSIA NLFSERSRSV QELVDSKSEG I SQASKATVT SSTVEESIG
GKELEIVQQK QMNKDTSVLV QIEKEMVKLK EIAOQIEKIF GIVTGIAEQT
NLSSILNASIE SARASEHGKG FAVVANEVRK LSEDHKTVS TVSELVNTNN
TQINVSKHI KDYVELOVES KEKMTQINRL FDEIVHSMKI SKEQSGKIDV
DLQAFGLGLQ EVSRAVSHVA ASVDSLVILT EHHHHHH

Notes: Glycerol Stock #1501
Protein Information:
Protein Sequence:
438 Amino Acids
Molecular Weight 49607.35Da

HHHHHMLFK KDRQETAYF SDSNGQQKNR IQLTNKHADV KKQLKMVRG
DAELVLEQL QPLQENIVN IVDAFYKNLD HESSLMDIIN DKSSVDRLKQ
TLKRIQEMF AGVIDDEFTI KRNRIASILH RIGLLPKWYM GAFQELLLSM
IDIYEAISITN QSELLKAKEA TTKILNLEQEQ LVLEAFQSEY NQTRDEQEGEEK
KNLLHQKIQE TSGSIANLFS ETSRSVQELV DKSSEGISQAQ KAGTVTSSTV
EEKSIQGKKKE LEVQQKQMNK IDTSLVQIEK EMVKLDERTAQ QIEKIPGITV
GIAEQTNLLS LNASIESARA GEGHKGFAVV ANEVRKLED TKKTSTVSE
LVNNTNTQIN IVSHKIDVNY ELVESSEKEM TQINRDFDEI VHSMKISKEQ
SGKIDVDLQA FLGGLQEVSR AVSHVAASV SLVILTEE

Notes: Glycerol Stock #1506
Protein Information:
Protein Sequence:
193 Amino Acids
Molecular Weight 22492.92 Da

MAHHHHHRGS GIEGRMLFKK DRKQETAYFS DSNGQQKNRI QLTNKHADVK
KQLKMRLGD AELYVELQLO PLIQENIVNI VDAFKNLHD ESSLMDIIND
HSSVDRVKQH LKRHIQEMFA GVIDDEFIEK RNRIASIHLR ILGHPKWMG
AFQELLLSMI DIYEASITTNQ QELLKAIKAT TKILNLEQQL VLE

Notes: Glycerol Stock #1719
Protein Information:
Protein Sequence:
193 Amino Acids
Molecular Weight 22476.92 Da

MAHHHHHHS GIEGRMLFKK DRKQETAYFS DSNQQQKRI QLTNKHAVDK
KQLKMVRLLD AELYVLEQLQ PLIQENIVNI VDAFKNLDH ESSLMDIIND
HSSVDRLQQT LKRHIQEMPA GVIDDEFIEK RNRIASIHLR IGLLFKWMG
AFQELLLSMT DIYEASITNQ QELLKAIKAT TKILNLEQQL VLE

Notes: Glycerol Stock #2520
Protein Information:
Protein Sequence:
193 Amino Acids
Molecular Weight 22442.90Da

MAHHHHHGS GIEGRMLFKK DRKQETAYFS DSNGQQKNRI QLTNKHADVK
KQLMVRQKLD AELVLEQLQ PLIQENIVNI VDAFQKNLDH ESSLMIND
HSSYDLKQT LKRIQEMPA GVIDDEFIEK RNRIASIHKL IGLLPKWYM
AFQELLLSMI DIYEASITNQ QELLKAIKAT TKILNELQQL VLE

Notes: Glycerol Stock #2521
Protein Information:
Protein Sequence:
193 Amino Acids
Molecular Weight 22515.96 Da

MAHHHHHGSGIEGRMLFKK DRKQETAYFS DSNQQQRNRI QLTNKHADVK
KQLKMVRLGD AELYVLEQLQ PLIQENIVNI VDAFPKNLDH ESSLMDIIND
HSSVDRKQT LKRHIQEMFA GVIDDEFIEK RNRJASTHML IGLLPKWYMG
AFQELLSSNI DIYEASITNQ QELLKAIKAT TKILNLEQQL VLE

Notes: Glycerol Stock #2522
Protein Information:
Protein Sequence:
230 Amino Acids
Molecular Weight 26816.57Da

Notes: Glycerol Stock #1974
Protein Information:
Protein Sequence:
230 Amino Acids
Molecular Weight 26828.63Da

MAHHHHHGSGIEGRMLFKKDRKQETAYFSDSNGQQKNRIQLTNKHADVKKQLKVMRVLGDAELYVLEQLQPLIQENIVNIVDAPYKNLDESSLMDDINDHSSVDRLIKQTLRHIIQEMPNGVIDDEFIEKRNRIASIHLRIGLLPKWMGAFQELLSMMIDYEASITNQQELLKAIKATTKILNLLEQQLVLEAFQSEYNQTRDEQEEKKNLHQQIQLSNGSSIANLFSE

Notes: Glycerol Stock #1976
REFERENCES


