COMMITTEE CERTIFICATION OF APPROVED VERSION

The committee for Yonghong Zhao certifies that this is the approved version of the following dissertation:

STRUCTURAL STUDIES OF THE YEDU STRESS PROTEIN

Committee:

Robert O. Fox, Ph.D., Supervisor

Vincent J. Hilser, Ph.D.

Hiram F. Gilbert, Ph.D.

Luis Reuss, M.D.

David W. Bolen, Ph.D.

Dean, Graduate School

STRUCTURAL STUDIES OF THE YEDU STRESS PROTEIN

By

Yonghong Zhao, M.S.

Dissertation
Presented to the Faculty of The University of Texas Graduate School of
Biomedical Sciences at Galveston
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Approved by the Supervisory Committee

Robert O. Fox, Ph.D.
Vincent J. Hilser, Ph.D.
Hiram F. Gilbert, Ph.D.
Luis Reuss, M.D.
David W. Bolen, Ph.D.

December 2004
Galveston, Texas

Key words: Protease, Metalloprotein, X-ray crystallography

In memory of my mother Fuqin Hao (1941-2001)
‘gone but not forgotten’

ACKNOWLEDGEMENTS

I am deeply grateful to many people who have helped me in my research and in my life of more than 5 years at UTMB.

First, I would like to thank my advisor Dr. Robert O. Fox. It has been a great opportunity to be a member of his research group and to carry out the project under his direction. I truly appreciate his constant teaching and support. I am also very grateful to Drs. Hiram F. Gilbert, Luis Reuss, David W. Bolen, and Vincent J. Hilser. As the supervisory committee members, they reviewed my proposal critically. Drs. Gilbert and Hilser made many very good suggestions. I leaned a great deal from Drs. Bolen and Reuss not only on the research, but also on courses that they taught during my first year at UTMB. I also would like to thank Dr. Nancy K. Wills. When I did my research rotations in Wills group and Bolen group, I learned many useful techniques that have been a great help for my research. I thank Drs. Edmund W. Czerwinski, Werner Braun, James C. Sacchettini, and John Caradonna for their help on bioinformatics and model building.

I thank Dr. White A. White for maintaining X-ray crystallography facility at UTMB. He also has been very helpful on data collection and processing, model building, refinement, and visualization software. I thank Dr. Henry D. Bellamy for beamline support at CAMD.

I thank members of Fox group for their help and kindness. Dr. Xiuzhen Fan has been very helpful on molecular biology and mass spectrometry. Deqian Liu showed me how to set up crystallization trials and handle crystals. Warna D. Kaluarachchi did the peptide digestion experiment. Dr. Munia Mukherjee and Ye Peng also helped me a lot on many experiments.

I also thank the current and former directors of the graduate program in Cellular Physiology and Molecular Biophysics, Drs. Malcolm S. Brodwick, Philip T. Palade, Simon A. Lewis, Aileen K. Ritchie, and Burgess N. Christensen. I appreciate their guidance and support.

Last but not least, there are my family members. I thank my wife Xin Guo, my parents, and my sisters for their endless support. I thank my five-month son, Jason Xin Zhao, for the joy he has brought to the family since he was born on May 31, 2004. I also appreciate my parents-in-law for their additional support after Jason was born.

STRUCTURAL STUDIES OF THE YEDU STRESS PROTEIN

Publication No._____

Yonghong Zhao, Ph.D.
The University of Texas Graduate School of Biomedical Sciences at Galveston, 2004

Supervisor: Robert O. Fox

The knowledge of stress proteins is important for understanding stress response, pathology of a broad set of diseases, and the development of therapeutics. The *Escherichia coli* YedU stress protein, also known as Hsp31, is highly induced upon heat shock. To obtain a better understanding for the possible molecular function of the YedU stress protein, it was expressed, purified, and crystallized. The crystal structure of YedU was determined at 2.2 Å resolution in a multiple isomorphous replacement (MIR) experiment.

YedU monomer has an alpha/beta/alpha sandwich domain and a second smaller domain. Between the sandwich domain and the second smaller domain, there is a putative catalytic triad composed of Cys184, His185, and Asp213. A metal-binding site was identified, where a zinc(II) ion is coordinated by a 2-His-1-carboxylate motif composed of His85, Glu90, and His122. The possible functions of the metal-binding site and the Cys184-His185-Asp213 triad are discussed.

It was reported that YedU has chaperone activity *in vitro*. YedU forms dimers in solution and the dimer interface was identified. The molecular surface of the YedU homodimer exhibits a number of solvent-exposed hydrophobic patches that are reminiscent of substrate-binding sites of molecular chaperones. To investigate the role of a central hydrophobic patch in substrate binding, four mutants were made, each replaced a hydrophobic residue with a charged residue. Compared with the wild type YedU protein, the chaperone activity of these mutants was only slightly reduced, suggesting that these residues alone do not play a dominant role in substrate binding at high temperatures.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAA	ATPase Associated with a variety of cellular Activities
ADH	Alcohol dehydrogenase
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BME	2-mercaptoethanol
BSA	Bovine serum albumin
CS	Citrate synthase
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FOM	Figure of merit
GABA	Gamma-aminobutyric acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid
H-NS	Histone-like nucleoid structuring protein
HSP	Heat shock protein
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kDa	Kilodalton
MAD	Multiple wavelength anomalous dispersion
MIR	Multiple isomorphous replacement
MR	Molecular replacement
NAD	Nicotinamide adenine dinucleotide
NCS	Non-crystallographic symmetry

PCMB	p-chloromercuribenzoate
PCR	Polymerase chain reaction
PDB	Protein data bank
PMSF	Phenyl-methylsulfonyl fluoride
RMSD	Squared root of mean square deviations
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis

CHAPTER 1: INTRODUCTION

This chapter provides a basic introduction to stress response and stress proteins, functions and structures of chaperones and proteases, protein quality control, and primary sequence analysis of the YedU stress protein.

1.1 STRESS RESPONSE AND STRESS PROTEINS

Cellular stress can be defined as the threat of damage to macromolecules (Kultz, 2003). Many types of environmental stress (e.g., thermal stress, oxidative stress, osmotic stress, and heavy metal stress) cause deleterious changes in protein conformation. Stress response represents an attempt by the cell to counteract the deleterious effects of the stressor. Stress response is critical for the cell to survive in different environmental conditions (Hengge-Aronis, 2002). The exposure of cells to stress triggers or up-regulates the expression of a set of proteins that are collectively called stress proteins or, for historical reasons, heat shock proteins (Macario, 1995). As a consequence of stress, house-keeping proteins are down-regulated or shut off, while stress proteins are up-regulated, or called into action if they were not constitutively expressed. Stress proteins enable the cell to react in an appropriate manner in order to survive in stress conditions. They perform a diverse array of functions including transcription regulation, chaperone, protease, as well as other more specific functions. The knowledge of how stress proteins protect cells is important for understanding the pathology of a broad set of diseases and the development of therapeutics (Macario and Conway de Macario, 2000).

In many cases, molecular mechanisms that underlie biological phenomena are better-understood in prokaryotes, particularly in *Escherichia coli*, one of the most widely studied organisms. In the following sections, function, structure, and mechanism of

action of major *E. coli* stress proteins will be described briefly. Stress proteins are highly conserved in all three life kingdoms, bacteria, archaea, and eukaryotes. These conservations between stress proteins from different organisms go beyond the primary sequence; they also pertain to structure, function, and mechanism of action. Thus, information on function, structure, and mechanism of action of a stress protein from the bacterium *E. coli* is partially applicable to eukaryotes (Macario and Conway de Macario, 2000).

1.1.1 Regulation of Stress Response

In *E. coli* the expression of stress response genes is mediated mainly by modulation of specificity of RNA polymerase by replacement of the promoter recognition sigma (σ) subunit (Ishihama, 2000). Seven different species of sigma subunit, RpoD (σ^{70}), RopN (σ^{54} or σ^{N}), RpoS (σ^{38} or σ^{S}), RpoH (σ^{32} or σ^{H}), RpoF (σ^{28} or σ^{F}), RpoE (σ^{24} or σ^{E}), and FceI (σ^{FecI}), have been identified in *E. coli*, each involved in transcription of a specific set of genes. The bacteria constitutively express a set of housekeeping genes that are involved in basic functions needed for the sustenance of the cell. Most of the housekeeping genes are transcribed by the holoenzyme containing the sigma subunit RpoD, while the holoenzyme containing the sigma subunit RpoS is essential for transcription of general stress response genes that are activated by a variety of stress conditions. Specific stress response genes are transcribed by RNA polymerase holoenzymes containing alternative minor sigma subunits. The holoenzymes containing the sigma subunit RpoN transcribe those genes that are activated by deficiency of nitrogen; the holoenzymes containing the sigma subunit RpoH transcribe heat shock genes; holoenzyme containing the sigma subunit RpoF is responsible for the expression of the third wave of flagella and chemotaxis genes; the holoenzyme containing the sigma subunit RpoE is needed for transcription of genes whose products deal with protein defects such as misfolded proteins in the periplasm; and the sigma subunit FceI is involved in transcriptional activation of fec operon expressing ferric citrate transporters.

The promoter recognition specificity of holoenzymes containing different sigma subunits is not entirely understood. Sigma subunit specificity at promoters is affected by the interplay of RNA polymerase holoenzymes with other regulatory factors such as H-NS (histone-like nucleoid structuring protein), LRP (leucine-responsive regulatory protein), CRP (cAMP receptor protein), IHF (integration host factor), and FIS (factor for inversion stimulation). For example, it has been shown that some promoters, normally preferentially recognized by the holoenzyme containing the sigma subunit RpoS, can be transcribed by the holoenzyme containing the sigma subunit RpoD in the absence of H-NS (Hengge-Aronis, 1996).

In order to properly initiate and shut off stress responses, the cellular concentration of each sigma subunit is tightly controlled at multiple levels. For example, many factors regulate the cellular concentration of the sigma subunit RpoS at the levels of translation and protein stability (Hengge-Aronis, 1996). The histone-like protein H-NS and the LysR-like regulator LeuO repress RpoS translation, while host factor HF-1 and histone-like protein HU stimulate RpoS translation. Posttranslational RpoS degradation is mediated by the chaperone DnaK, the response regulator RssB, and the protease ClpPX. The net result of these regulations is low levels of RpoS in exponential-growth phase and increased levels during stationary phase or under stress conditions.

Heat shock response is under the control of the alternative sigma subunit RpoH whose intracellular level is very low during growth at 30°C and increases transiently upon heat shock (Yura et al., 2000). Heat induction of the sigma subunit RpoH mainly occurs at the posttranscriptional level, although regulations exist at four different levels: transcription and translation of the *rpoH* gene and activity and stability of the RpoH protein. An extended secondary structure in the rpoH transcript blocks translation at low temperatures. At high temperatures, melting of that structure permits ribosome entry followed by translation initiation. Posttranslationally, the fate of RpoH protein is determined by its interaction with a number of other proteins. Normally, RpoH is inactivated through interactions with DnaK and DnaJ under nonstress conditions. Sequestration of RpoH by the chaperones also protects it from FtsH-mediated proteolysis.

Upon heat shock, accumulation of unfolded proteins strips away the DnaK/DnaJ chaperones, freeing RpoH to initiate transcription of heat shock genes (Yura et al., 2000). At high temperatures, a conformational change of RpoH abolishes its interaction with DnaK/DnaJ. The structurally altered RpoH is rapidly turned over by cellular proteases, thus turning off the heat shock response.

1.1.2 Molecular Chaperones

Molecular chaperones are ubiquitous and highly conserved proteins that bind to and stabilize an otherwise unstable conformer of another protein (Hartl and Hayer-Hartl, 2002). At the molecular level, chaperones protect against protein aggregation, solubilize protein aggregates, assist in protein folding, target irreversibly damaged proteins to degradation, and sequester damaged proteins to larger aggregates. Molecular chaperones are essential for all cells in all states, and they are needed even more under stress conditions that induce protein damage. In *E. coli* major chaperones include AAA+ chaperones (ClpA, ClpB, ClpX, and ClpY), DnaK (Hsp70) and its co-chaperone DnaJ (Hsp40), GroEL (Hsp60) and its co-chaperone GroES (Dougan et al., 2002a).

The *E. coli* ClpA, ClpB, ClpX, and ClpY are members of the AAA+ superfamily (Neuwald et al., 1999), which is proposed as an expansion of the AAA family of ATPases (ATPase Associated with a variety of cellular Activities). Based on their sequence homology, these AAA+ chaperones can be classified into two groups, group I consists of ClpA, ClpB, ClpC, and ClpE, while group II consists of ClpX and ClpY. Among group I AAA+ chaperones, ClpC and ClpE are from gram-positive bacteria. Group I AAA+ chaperones have an N-terminal variable region and two highly conserved nucleotide binding domains AAA1 and AAA2. Group II AAA+ chaperones contain only a single nucleotide binding domain that is a homologue of AAA2. ClpX has an N-terminal zinc binding domain, which is not present in ClpY. ClpY contains an extra domain (I domain) that inserts into the AAA2 domain. Since ClpA and ClpX are

associated with ClpP protease and ClpY is associated with ClpQ protease, more details about ClpA, ClpX, and ClpY will be described in section 1.1.3.

ClpB (Hsp104) collaborates with the DnaK system to disassemble and refold large protein aggregates in an ATP-dependent manner (Mogk et al., 1999). At very high temperatures, strong protein aggregation occurs because the large increase in misfolded proteins overwhelms the limited holder chaperone capacity of the cell. Following heat stress, aggregated proteins are first resolubilized by ClpB and DnaK. The subsequent refolding of the disaggregated proteins is assisted by the KJE chaperones (DnaK, its cochaperone DnaJ, and the nucleotide exchange factor GrpE). In addition to an N-terminal (N) domain and two nucleotide binding domains, AAA1 and AAA2, ClpB also has a middle (M) domain. The crystal structure has been reported for ClpB of *Thermus thermophilus* (Lee et al., 2003a) and it reveals the inherent flexibility of the chaperone. Six ClpB monomers form a hexamer that has a two-tiered ring structure, with a diameter of 140 - 150 Å and a height of 90 Å. The N and M domains are highly mobile and move independently as rigid bodies. The role of the mobile N domain in ClpB disaggregase function is unknown. Immobilization of the M domain by disulfide bridges inactivates ClpB disaggregase without affecting ClpB oligomerization or ATPase activity. Addition of dithiothreitol (DTT), that breaks disulfide bridges, can restore ClpB disaggregase activity, indicating a direct correlation between M domain mobility and ClpB disaggregase activity (Lee et al., 2003a). Thus, the M domain is proposed to act as a molecular crowbar in converting larger protein aggregates into smaller ones.

DnaK belongs to Hsp70 family and it is the most abundant *E. coli* chaperone (Slepenkov et al., 2002). DnaK binds to and stabilizes nascent polypeptide and stress-denatured protein substrates, preventing them from forming aggregates. DnaK also regulates other chaperone proteins through its stress-dependent association with the RpoH sigma subunit. DnaK has two domains, an N-terminal ATP-binding domain and a C-terminal substrate binding domain. Binding and hydrolysis of ATP regulate the affinity of substrate to the C-terminal domain. It has higher substrate affinity when ADP is bound and the substrate is released when ATP is bound. Substrate binding and release

is regulated through its interactions with the co-chaperone DnaJ and the nucleotide exchange factor GrpE. The J domain of DnaJ interacts with the ATPase domain of DnaK and stimulates substrate binding. The release of ADP from the ATPase domain of DnaK is triggered by GrpE. Then ATP binds to DnaK and the substrate is released. The released substrate either folds into native structure, or fails in its folding attempt and is recaptured by other chaperones.

GroEL belongs to the Hsp60 family and it is the most studied of all molecular chaperones (Xu et al., 1998). Homologues of GroEL have been found in all three life kingdoms as well as in organelles. GroEL facilitates protein folding not only by preventing aggregation but also by providing an isolated environment that favors the native state conformer. Deletion of GroEL is lethal, indicating that it is critical for the folding of one or more essential cell proteins. Each GroEL monomer has three domains, an equatorial domain, an apical domain, and a hinge-like intermediate domain. Fourteen monomers of GroEL form double-ring shaped protein complexes. The equatorial domain provides most of the intra-ring interactions and all of the inter-ring interactions. The apical domain forms the opening to the central channel. The intermediate domain links the equatorial domain to the apical domain. GroEL has ATPase activity and the ATP-binding pocket resides in the equatorial domain. GroES, the cochaperone of GroEL, folds into a sevenfold rotationally symmetric, dome-shaped structure and binds the *cis* GroEL ring. The GroEL-GroES-ADP complex structure revealed that all of the nucleotide-binding sites on the *cis* GroEL ring are fully occupied by ADP, while sites on the *trans* GroEL ring are empty (Xu et al., 1997). The inner surface of the *trans* ring is largely hydrophobic. The *trans* conformation of GroEL is responsible for binding its substrate and hydrophobic interactions are important in substrate recognition. The presentation of up to seven apical hydrophobic surfaces provides for cooperative binding of substrates and a potential unfolding mechanism to reset polypeptides for another attempt in folding. ATP and GroES bind to the ring containing substrate, creating a stable *cis* ring and stripping the substrate into the central chamber "Anfinsen cage" (Xu et al., 1998). Now, the central chamber is lined with mostly polar residues, creating an isolated environment that favors the polypeptide to fold into its native structure. When

all ATP is hydrolyzed, the weakened *cis* ring is "primed" for release of GroES. Another substrate, ATP and GroES bind to the *trans* ring on the other side and the weakened *cis* ring releases ADP, GroES, and the product.

HtpG (Hsp90), Hsp33, and IbpA/IbpB (sHsp) are abundant *E. coli* cytosolic proteins that have chaperone activity *in vitro*, but their physiological roles are poorly understood. HtpG is dispensable in *E. coli*, while its eukaryotic homologue Hsp90 is essential in yeast (Bardwell et al., 1988). Although HtpG exhibits chaperone activity *in vitro*, little is known about its *in vivo* functions.

Hsp33 is a redox-regulated molecular chaperone (Jakob et al., 1999). At the transcriptional level, the Hsp33 is under heat shock control; at the posttranslational level, Hsp33 is under oxidative stress control. Hsp33's redox sensor is located in the C-terminus of the protein. Under reducing conditions, four cysteines in the C-terminus of the Hsp33 protein coordinate zinc in a low-affinity substrate-binding conformation. Upon oxidation, these cysteines form two intramolecular disulfide bonds in a high-affinity substrate-binding conformation.

IbpA and IbpB are two sequence-related proteins of 14 kDa and 16 kDa, respectively (Veinger et al., 1998). Both IbpA and IbpB can form multimers that consist of about 100-150 subunits. They belong to the small heat shock protein (sHsp) family that all have a core region of about 85 residues (the α-crystallin-like domain). In addition to the α-crystallin-like domain, both IbpA and IbpB have an N-terminal domain that is essential for oligomerization. *In vitro*, it was shown that IbpB oligomers can bind denatured proteins and deliver them to the KJE chaperones (DnaK/DnaJ/GrpE) for subsequent refolding (Veinger et al., 1998). *In vivo*, IbpA and IbpB are found associated with endogenous protein aggregates during heat shock and with recombinant proteins in inclusion bodies.

In addition to these chaperones described above, the *E. coli* trigger factor is an abundant cytosolic protein and has peptidyl-prolyl-cis/trans-isomerase (PPIase) activity *in vitro* (Hesterkamp et al., 1996). *In vivo*, it binds to nascent polypeptides to prevent

their aggregation. Trigger factors also cooperates with other chaperones in promoting proteolysis of abnormal proteins.

1.1.3 Proteases

Under stress conditions, damaged proteins that are beyond repair are degraded by cytoplasmic proteolysis. Prokaryotes do not have a labeling molecule such as eukaryotic ubiquitin for tagging protease substrates, thus bacterial proteases must have the ability to recognize those proteins that shall be degraded. Major proteases of the *E. coli* cytosol include Lon, ClpP, and ClpQ (HslV) (Dougan et al., 2002a). All three are ATP-dependent proteases.

Lon is the primary protease degrading misfolded proteins in *E. coli* (Gottesman, 1996). Degradation of these abnormal proteins usually requires the participation of chaperones. In addition to abnormal proteins, Lon protease also recognizes and degrades SulA (a cell division inhibitor) and RcsA (a positive regulator of capsule transcription). The Lon protease is made up of four identical subunits of 87 kDa. Lon is an ATP-dependent serine protease and each subunit carries an active-site serine and an ATP-binding site. In eukaryotes, Lon is present in mitochondria and it is essential for mitochondrial function.

The mature ClpP protein is 21.5 kDa with a serine protease-like active site (Porankiewica et al., 1999). ClpP alone is unable to degrade peptides longer than six residues. To gain full proteolytic activity, ClpP must associate with one of two related subunits, ClpA or ClpX. Both ClpA and ClpX have ATPase activity and belong to the Clp/Hsp100 family of molecular chaperones. The active protease is composed of two central heptameric rings of ClpP flanked by two hexameric rings of ClpA or ClpX. An essential ClpP docking site was identified as a tripeptide [LIV]-G-[FL] in the AAA2 domain of ClpA and ClpX. The ClpAP protease degrades large proteins down to small peptides and has no apparent sequence specificity. Recently, ClpS was identified as a ClpA-specific adaptor protein (Dougan et al., 2002b). ClpS enhances ClpA recognition

and consequently proteolysis of aggregated proteins. ClpXP protease has substrate specificity and has been linked to different cellular activities.

The *E. coli* ClpQ (HslV) protease is a bacterial homologue of the eukaryotic proteasome (Gottesman, 1996). ClpQ degrades specific substrate protein with the assistance of the ClpY chaperone in an ATP-dependent manner. Substrates of ClpYQ overlap with those of the Lon protease and include SulA (a cell division inhibitor), RcsA (a positive regulator of capsule transcription), and RpoH (the heat-shock sigma subunit). ClpQ has an active-site threonine and shares ~20% sequence similarity and a conserved fold with eukaryotic 20S proteasome β-subunit. ClpQ itself has no protease activity. The protease is activated upon the association with the ClpY chaperone. Twelve subunits of ClpQ form two central hexameric rings that are flanked by two hexameric rings of ClpY. The crystal structure of ClpYQ suggests that the unique I domain of ClpY is responsible for substrate recognition and initial steps in the transfer of substrate to the active site located in the cavity of the ClpQ rings (Bochtler et al., 2000; Sousa et al., 2000).

1.1.4 Protein Quality Control

Chaperones and proteases form an interconnected network playing a pivotal role in protein quality control that is important to maintain cellular homeostasis under normal conditions, and is critical for the cell to survive under stress conditions (Dougan et al., 2002a). Diverse human disorders, including several devastating neurodegenerative diseases such as Alzheimer's and Parkinson's diseases, are thought to arise from misfolded proteins that overwhelm or escape from the protein quality control machinery (Soto, 2001).

Chaperones prevent protein aggregation and assist in protein *de novo* folding and in refolding of misfolded proteins. Proteases remove irreversibly damaged proteins from the cellular pool. In some cases, chaperones and proteases also coordinate the cellular response to stress by modifying the abundance or activity of transcription regulators. As the major components of the cellular protein quality control machinery, many chaperones

and proteases, though constitutively expressed, are synthesized at greatly increased levels in response to stress.

Although the native fold of a protein is encoded in its amino acid sequence, protein folding inside the cell is usually not a spontaneous process (Hartl and Hayer-Hartl, 2002). About one third of newly synthesized *E. coli* proteins require the assistance of molecular chaperones and the input of metabolic energy to reach their native structure efficiently. Generally, nascent chains are held by trigger factor (TF) while they are being synthesized by the ribosome. After whole polypeptide chains are released from the ribosome, most proteins smaller than 20 to 30 kDa (~65-80% of total cellular proteins) fold rapidly without further assistance. Some larger proteins (~10-20% of total cellular proteins) are first stabilized by DnaK. They fold upon one or several cycles of ATP-dependent binding to and release from DnaK. Another 10 to 15% of all polypeptide chains can fold correctly only inside the central chamber of the GroEL/GroES chaperones. Deletion of GroEL is lethal, indicating that some of the *E. coli* essential proteins require the assistance of GroEL/GroES chaperones for folding (Fayet et al., 1989).

Misfolded and partially folded proteins arise constitutively due to slow rates of folding, intrinsic structural instability, and biosynthetic errors. Under stress conditions, the protein misfolding problem becomes severe, as many stressors denature proteins. The cellular protein quality control machinery counteracts the aggregation of nonnative proteins, assists in refolding of misfolded proteins, and degrades damaged proteins that are beyond repair (Dougan et al., 2002a). In the first line of defense, the misfolded proteins may be refolded by the KJE (DnaK/DnaJ/GrpE) or the ELS (GroEL/GroES) chaperones, or removed by Lon protease. Under severe conditions of stress, the cellular chaperones become overloaded and misfolded proteins tend to aggregate. When stress is alleviated or withdrawn, the aggregated proteins may be rescued by the ClpB and KJE chaperones, or degraded by the ClpAPS protease, minimizing the toxic effect of protein aggregation and enhancing cell survival.

1.2 THE YEDU STRESS PROTEIN

Although stress responses and some well-known stress proteins have been studied for decades, the inventory of stress proteins is still expanding. After the *E. coli* genome sequencing was completed (Blattner et al., 1997), DNA microarray technique was used to study the *E. coli* gene expression profiling under different stress conditions (Richmond et al., 1999; Pomposiello et al., 2001; Zheng et al., 2001; Weber et al., 2002; Tucker et al., 2002; Quillardet et al., 2003). Richmond and co-workers identified 77 genes that were up-regulated upon heat shock (Richmond et al., 1999). It was not surprising that the proteins encoded by these genes include well-known stress proteins such as ClpB, HtpG, DnaK, GroEL, ClpQ, and ClpP. It also includes many genes that encode proteins with no known function. It is very likely that these stress-inducible proteins perform novel functions in counteracting the effects of stressors. Bardwell and co-workers studied two of these stress-inducible proteins, YrfH and YrfI. It was shown that the YrfH protein, renamed as Hsp15, is a ribosome-associated protein involved in the recycling of free 50S subunits (Korber et al., 2000). The YrfI protein, renamed as Hsp33, was shown to be a redox-regulated chaperone (Jakob et al., 1999). The target of this study, the YedU protein, is also one of these stress-inducible proteins with no known function (Richmond et al., 1999).

The *E. coli* YedU protein was first described as a 30-kDa protein whose expression is markedly enhanced in an hns-deletion strain (Yoshida et al., 1993). H-NS is a histone-like nucleoid structuring protein that plays a role in bacterial nucleoid organization and in the expression of genes involved in stress responses (Tendeng and Bertin, 2003; Atlung and Ingmer, 1997). Although the precise role of H-NS in bacterial physiology is still not clear, mutations of H-NS have been associated with numerous phenotypes including de-repression of β-glucoside metabolism (Defez and de Felice, 1981), increased resistance to acidic conditions and to high osmolarity (Hommais et al.,

2001), and loss of motility (Soutourina and Bertin, 2003). Beside the YedU protein, the expression of *E. coli* GadA, GadB, YciE, YciF, HdeA, and HdeB proteins is also enhanced in the hns-deletion strain (Yoshida et al., 1993). Interestingly, HdeA, GadA and GadB are all involved in *E. coli* acid resistance. HdeA is found in the periplasmic space, where it supports an acid resistance phenotype (Gajiwala and Burley, 2000). Functional studies demonstrate that HdeA is activated by a dimer-to-monomer transition at low pH, leading to suppression of protein aggregation caused by acid denaturation. GadA and GadB catalyze the decarboxylation of glutamate to gamma-aminobutyric acid (GABA), which is then transported out of the cell in exchange for new substrate (de Biase et al., 1999). It was proposed that GadC functions as an antiporter (Hersh et al., 1996). GadA, GadB, and the putative antiporter GadC constitute the gad acid-resistance system that confers the ability for bacterial survival for at least 2 hours in a strongly acidic environment (pH<2.5). The function of HdeB, YciE and YciF is still unknown, although HdeB is thought to form a heterocomplex with HdeA and may also play a role in acid resistance (Waterman and Small, 1996).

Recently, it was shown that transcription of the *yedU* gene is also moderately induced (1.68-fold) by irradiation with ultraviolet light (Quillardet et al., 2003). These expression profiles suggest that the *yedU* gene encodes a stress protein that may play a role in the heat shock response as well as other stress responses.

1.2.1 Primary Sequence Analysis of the Yedu Stress Protein

The predicted product of the *yedU* gene is a 283-residue protein. The sequence of the N-terminus of YedU protein purified from the hns-deletion strain is TVQTSKNPQ, indicating that the first Met is removed by methionine aminopeptidase *in vivo* (Yoshida et al., 1993). As the result of this posttranslational excision, the YedU stress protein is a 282-residue protein (swissprot accession P31658). A BLAST search against the GenBank non-redundant (nr) databases revealed that the YedU protein contains an

evolutionarily conserved DJ1/ThiJ/PfpI domain, which has been linked with distinct activities including RNA-protein interaction, biosynthesis of thiamine, and proteolysis (Bandyopadhyay and Cookson, 2004).

Genes that encode YedU-homologous proteins have been found in the genomes of several human pathogens including *Shigella flexneri* (GenBank accession gi:24113342, 98% identity to the YedU protein), *Pseudomonas fluorescens* (gi:23059816, 60%), *Vibrio cholerae* (gi:9658312, 57%), *Pseudomonas aeruginosa* (gi:15596332, 54%), and *Staphylococcus aureus* (gi:15923541, 54%). The recent sequencing of *Gibberella zeae* indicated that the YedU homologue (gi:42549093, 50%) is also present in eukaryotes. These proteins will be referred to as class I YedU orthologs. All have a conserved DJ1/ThiJ/PfpI domain, a large N-terminal extension, and two insertions (Figure 1-1).

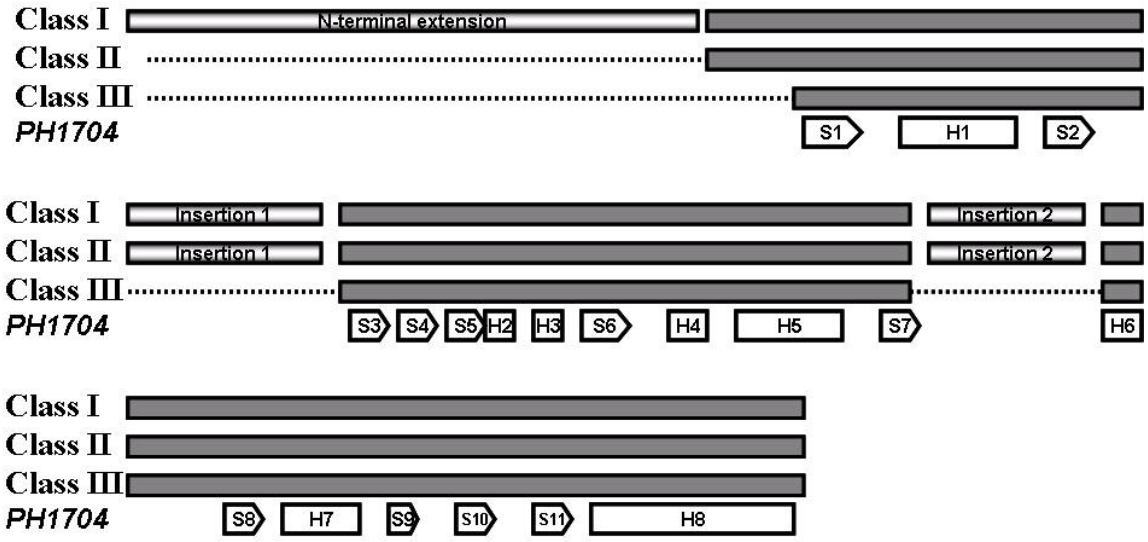

Figure 1-1 Class I, II, and III YedU orthologs

Class I orthologs have a large N-terminal extension of about 60 residues. Both Class I and Class II orthologs have two insertions. Class III orthologs contain a single DJ1/ThiJ/PfpI domain. The secondary structural elements of the PH1704 protease (class III) are shown. "S" and "H" refer to strand and helix, respectively.

YedU also shares 20-30% sequence identity with a group of proteins that have about 230 residues, forming a core DJ1/ThiJ/PfpI domain and a second smaller domain (Wilson et al., 2004). These proteins are present in a variety of prokaryotic and eukaryotic species and will be referred to as class II YedU orthologs (Figure 1-1). Yeast YDR533Cp, a class II ortholog, is induced by heat shock and oxidative stress. However, the molecular function of YDR533Cp and other class II orthologs is still unknown.

Proteins with a single DJ1/ThiJ/PfpI core domain will be referred to as class III YedU orthologs (Figure 1-1). They can be further classified into three groups, the PfpI cysteine proteases (class IIIA orthologs), the DJ-1 proteins (class IIIB orthologs), and the ThiJ kinases (class IIIC orthologs). Among them, the PfpI proteases will be referred to as class IIIA orthologs since they are more closely related to YedU than the DJ-1 proteins and the ThiJ kinases (Bandyopadhyay and Cookson, 2004). The PfpI cysteine proteases are a family of intracellular proteases found in most bacteria and archaea (Halio et al., 1996). The PH1704 protease from *Pyrococcus horikoshii* belongs to class IIIA orthologs. It forms hexamers in which a putative catalytic triad is composed of Cys100, His101, and a Glu474' contributed by an adjacent subunit (Du et al., 2000). YedU shares less than 20% sequence identity with the PH1704 protease. Two of the catalytic triad residues, Cys100 and His101, are conserved in YedU, but the Glu474' is not. The DJ-1 proteins will be referred to as class IIIB orthologs. Human DJ-1 protein is a medically important protein because of its involvement in early-onset Parkinson's disease (Bandmann, 2004). It was first identified as an oncogene that transforms NIH3T3 cells in cooperation with ras oncogene. It is also involved in the fertilization process. It has been shown that the human DJ-1 protein plays a role in the oxidative stress response that may underlie its association with early-onset Parkinson's disease (Bonifati et al., 2003). However, the exact molecular function of the DJ-1 protein is still unknown. The ThiJ kinases will be referred to as class IIIC orthologs. The *E. coli* ThiJ protein is related to the biosynthesis of thiamine, catalyzing the phosphorylation of hydroxymethylpyrimidine (HMP) to HMP monophosphate (Mizote et al., 1999).

1.2.2 Structure – A Bridge Between Protein Sequence and Function

About 220 prokaryotic and eukaryotic genomes have been completely sequenced and there are more than 900 genome sequencing projects are in progress (Genomes OnLine Database; http://www.genomesonline.org/). The goal of genome sequencing efforts is to discover the molecular and cellular functions of all the gene products. In newly sequenced genomes, genes are annotated mainly on the basis of sequence homology to proteins of known functions. However, in any genome, 30% to 50% of the predicted gene products either have no known homologues or show too little sequence homology (less than 30% identity) to proteins of known functions. Beside sequence-based methods, many other approaches are used to infer gene function, including gathering information about (i) the phenotype of the gene knockout; (ii) the temporal, spatial, and physiological regulation of proteins; (iii) the protein-interaction partners; and (iv) the protein structure determination. One rationale for the latter approach is that protein structures are more conserved than sequences, thus the structure can predict functional features that may not be obvious at the sequence level (Eisenstein et al., 2000). In the past several years, protein structure determination has provided valuable information for functional annotation as evidenced by many works in which the function is deduced from the structure (Kim et al., 2003). Structural studies can prompt functional annotation in several ways.

First, the nature of the ligand, the ligand-binding site, and the catalytic residues can be readily identified when a protein-bound ligand or cofactor is co-crystallized with the protein. Such information is the most useful for functional annotation. For example, Kim and co-workers showed that the MJ0577 orphan protein co-purified and co-crystallized with ATP (Zarembinski et al., 1998). This suggested that the protein might be an ATPase, which was subsequently confirmed with biochemical experiments.

Second, many structural studies with proteins of unknown function suggest a role for the protein in interaction with nucleic acids or other proteins based on the presence of

some well-characterized domains that bind nucleic acids or other proteins. For example, there are well-characterized RNA-binding domains including RNA recognition motifs, KH motif, dsRBD, S1, zinc finger and basic-rich motifs (Dreyfuss et al., 2002). The crystal structure of the *E. coli* YhbY protein indicated that it resembles the C-terminal domain of the translation initiation factor IF3, which binds to 16S rRNA (Ostheimer et al., 2002). The β–sheet surface of the YhbY protein contains numerous basic residues that may contribute to recognition and binding to RNA. Sedimentation experiments demonstrated that *E. coli* YhbY protein co-sediments with free 50S ribosomal subunits.

More generally, protein structures can reveal structural similarity that is not evident from sequence analysis. The *Haemophilus influenzae* YrbI protein has an alpha/beta hydrolase fold, which is most closely related to the fold of the haloacid dehalogenase (HAD) family (Parsons et al., 2002). This structural information prompted a more detailed sequence analysis that revealed remote homology to the phosphatase subfamily of HAD hydrolases. Subsequently, the phosphatase activity of the YrbI protein was demonstrated in biochemical studies.

However, a structure alone will provide conclusive functional information only in very rare cases. Very often, structural similarity can suggest a general class of biochemical function of a protein, but can not make specific suggestions. In these instances, further experimental and informatic approaches are necessary for functional annotation.

No structure of homologous proteins that share 30% or more sequence identity to the YedU stress protein has been determined previously. To obtain a better understanding for its possible molecular function, the YedU stress protein was expressed, purified, and crystallized. Its crystal structure was determined at 2.2 Å resolution in a multiple isomorphous replacement (MIR) experiment. Recently, it was reported that the YedU protein has chaperone activity *in vitro* (Sastry et al., 2002; Malki et al., 2003). The YedU crystal structure indicates that the surface of the YedU protein has a number of solvent-exposed hydrophobic patches that are potential substrate-binding sites for its chaperone activity. The YedU crystal structure also revealed a metal-binding site and a

potential nucleophile, providing novel information on its possible functions. The following chapters describe the work in detail. Chapter 2 summarizes the preparation and preliminary characterization of YedU crystals. Chapter 3 describes data collection and structure determination. Chapter 4 describes the YedU crystal structure, including its tertiary structure, quaternary structure, structural comparisons with orthologs, the metal-binding site, and the potential nucleophile. Chapter 5 describes the molecular surface of the YedU homodimer.

CHAPTER 2: PREPARATION OF YEDU CRYSTALS

This chapter describes the construction of p192 plasmid for overexpression of the YedU protein, purification procedures, crystallization screening and optimization, and preliminary crystallographic analysis of YedU crystals.

2.1 CONSTRUCTION OF THE P192 PLASMID

The genomic DNA of the *E. coli* MG1655 strain (ATCC) was extracted using a genomic DNA purification kit (Promega). The open reading frame (ORF) of the yedU gene was amplified from the genomic DNA by PCR using primers ECYEDU-S2 (5'-GGTATTGAGGGTCGCACTGTTCAAACAAGTAA-3') and ECYEDU-AS2 (5'-AGAGGAGAGTTAGAGCCACGTAACAGGGATTAACC-3'). The PCR product was purified using the QIAquick PCR purification kit (Qiagen) and then treated with T4 DNA polymerase (Novagen) in the presence of only dGTP. Since T4 DNA polymerase has exonuclease activity, it generates vector compatible overhangs at the end of the PCR product. The T4 DNA polymerase-treated PCR product was annealed with linear pET-30 Xa/LIC vector by incubation at room temperature for 5 minutes. The annealing reaction was then transformed into NovaBlue Singles (NBS) competent cells. The transformation was plated on LB (Luria-Bertani, 1.0% tryptone, 0.5% yeast extract, 1.0% NaCl) agar plates containing 30 μg/ml kanamycin. Plasmids were purified from several colonies and sequenced at the UTMB recombinant DNA laboratory. After its sequence identity to GenBank accession number gi:2367124 (Blattner et al., 1997) was confirmed, the pET-30 Xa/LIC based expression plasmid containing the *yedU* open reading frame was named p192.

2.2 EXPRESSION AND PURIFICATION OF THE YEDU PROTEIN

The p192 plasmid was transformed into the *E. coli* BL21(DE3) competent cells (Novagen). A single colony was picked up and grown overnight at 37 °C with shaking in 20 ml LB broth containing 30 μg/ml kanamycin. The overnight cell culture was diluted into 2 liters of 2xYT broth (1.6% Tryptone, 1.0% yeast extract, 0.5% NaCl) containing 30 μg/ml kanamycin. The cell culture was grown at 37 °C with shaking. After two hours, A_{600} (absorbance at 600 nm) of the cell culture was measure every 15 minutes to monitor the growth. Isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.8 mM when the cell culture was in was mid-log phase (A_{600} of 0.4-0.6). After induction, the cell culture was continued to grow at 37 °C with shaking for 3 hours before cells were harvested by centrifugation at 2,000 × g for 45 minutes. Cell pellets were frozen and stored at -80 °C if they were not processed immediately.

Recombinant YedU protein has a hexahistidine tag (His-tag) and was purified from the cell lysate by metal-affinity chromatography. The fusion tag was then removed by the Factor Xa protease. The YedU protein was further purified by size-exclusion chromatography.

2.2.1 Nickel-Nitrilotriacetic Acid (Ni-NTA) Affinity Column

The cell pellets were suspended in buffer 1A (10 mM imidazole, 300 mM NaCl, 50 mM Tris pH 8.2). Phenyl-methylsulfonyl fluoride (PMSF) and 2-mercaptoethanol (BME) were also added to final concentrations of 1 mM and 10 mM, respectively. Cells were lysed by repeated sonications on ice (6 bursts, 30 seconds / burst with 30 seconds interval between bursts). Cell lysates were clarified by centrifugation at 38,000 × g for 30 minutes. The supernatant was loaded onto a column containing 20 ml Ni-NTA

superflow resin (Qiagen) pre-equilibrated with buffer 1A. The column was washed with 100 ml buffer 1A containing 10 mM BME before it was developed at 4.0 ml/min with a 0-600 mM imidazole gradient in 30 minutes. YedU-containing fractions were eluted between 200 and 400 mM imidazole and they were pooled together and dialyzed against 10 mM BME, 150 mM NaCl, 25 mM Tris pH 8.2 (Figure 2-1A). At this point, A_{280} (absorbance at 280 nm) was measured and the protein concentration was estimated using a molar extinction coefficient of 39,060 $M^{-1}cm^{-1}$ (extinction coefficient calculation will be described in section 2.2.3).

2.2.2 Removal of Fusion Tag and Size-Exclusion Chromatography

There is a Factor Xa protease-cleavage site between the His-tag and the YedU protein. To remove the fusion tag, Factor Xa protease (Novagen) was added (5 U / mg of recombinant YedU protein) and the reaction was carried overnight at room temperature. After the completeness of cleavage was confirmed on a 15% SDS-PAGE (Figure 2-1A), the protein was concentrated and loaded onto a HiLoad™ 26/60 Superdex75 size-exclusion chromatographic column (Pharmacia). The column was developed with 0.5 mM dithiothreitol (DTT), 150 mM NaCl, 25 mM Tris pH 8.2 at a flow rate of 2.0 ml/min (Figure 2-1B). YedU-containing fractions were pooled together and concentrated to about 20 mg/ml by ultrafiltration through an YM-10 membrane (Amicon). DTT was added to a final concentration of 5 mM. Glycerol were also added to 20% (v/v) final concentration before the protein was frozen in liquid nitrogen and stored at -80 °C.

The Superdex75 size-exclusion chromatographic column was calibrated with vitamin B_{12} (1.35 kDa) and globular proteins of known molecular weight. Elution volumes of bovine thyroglobulin (670 kDa), bovine serum albumin (67 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa) and vitamin B_{12} (1.35 kDa) were 113 ml, 138 ml, 153 ml, 197 ml, and 287 ml, respectively. For size exclusion chromatography, a linear correlation between the logarithms of protein molecular weight and the distribution coefficients has been observed for globular proteins (Locascio et al., 1969), where the

distribution coefficient (Kd) is defined as (Ve-Vo)/Vi (Ve, Vo, and Vi are elution volume, void volume, and internal volume, respectively). YedU (31059 Da/monomer) migrated through the Superdex75 size-exclusion chromatographic column as a homodimer whose apparent molecular weight is about 60 kDa (Figure 2-1B).

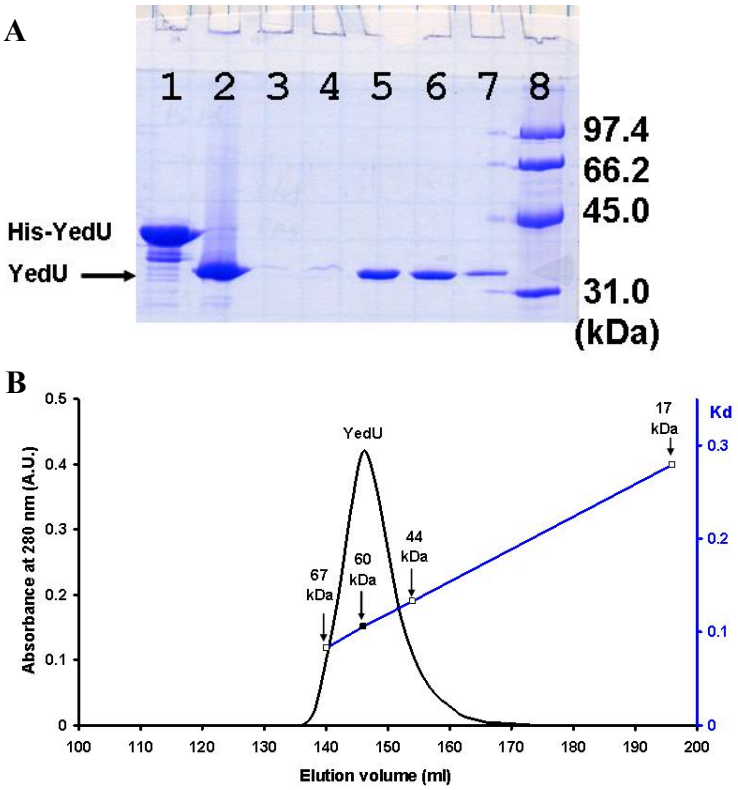

Figure 2-1 Purification of the YedU protein

(*A*) 15% SDS-PAGE with Coomassie blue G-250 staining. Lane 1, the recombinant YedU protein with an N-terminal His-tag; Lane 2, cleavage of the fusion tag by the Factor Xa protease; Lanes 3-7, Superdex75 sizing column fractions; and Lane 8, standard protein marker. (*B*) Size-exclusion chromatographic profile of the YedU protein. The elution volumes of protein standards, bovine serum albumin (67 kDa), chicken ovalbumin (44 kDa), and horse myoglobin (17 kDa), are shown by arrows. The distribution coefficients (Kd) are plotted with the logarithms of molecular weight. A.U., arbitrary units.

2.2.3 Determination of Protein Concentrations

The concentrations of YedU protein were determined by measuring its absorbance at 280 nm (A_{280}). The calculated molar extinction coefficient (ProtParam sever http://au.expasy.org/tools/protparam.html; Gill and von Hippel, 1989) of the YedU protein is 32,550 $M^{-1}cm^{-1}$ in 6.0 M guanidium hydrochloride, 0.02 M phosphate buffer pH 6.5. Concentrated YedU was diluted 100-fold into 6.0 M guanidium hydrochloride, 0.02 M phosphate buffer pH 6.5 and 150 mM NaCl, 50 mM Tris pH8.2, respectively. A_{280} of YedU in native condition (150 mM NaCl, 50 mM Tris pH8.2) was 1.20 times greater than that in the denatured condition. Thus, a correction factor of 1.20 was applied to yield the molar extinction coefficient of 39,060 $M^{-1}cm^{-1}$ at 280 nm, which was used to calculate YedU concentration under native conditions.

2.2.4 Expression and Purification of Selenomethionyl Protein

The introduction of the tunable synchrotron radiation sources has made it feasible for determining the phase angles of protein crystals using multiple wavelength anomalous dispersion (MAD) method (Hendrickson et al., 1988). In general, selenium (Se) can be introduced into a protein by growing microorganisms in media with methionine substituted by selenomethionine. If diffraction data of crystals containing selenomethionyl protein are collected at three different wavelengths ("edge", "white line", and "remote"), respectively, anomalous scattering can be exploited for phase angle determination.

In order to apply MAD method to determine the phase angles of the YedU crystals, selenomethionyl protein was produced and purified. The *E. coli* BL21(DE3)pLysS cells (Novagen) bearing plasmid p192 were grown overnight at 37 °C with shaking in 10 ml LB broth containing 30 μg/ml kanamycin and 34 μg/ml chloramphenicol. The overnight cell culture was spun down and resuspended in 1 ml M9 minimum broth (2.0% glucose, 1 mM $MgSO_4$, 42 mM Na_2HPO_4, 24 mM KH_2PO4, 9

mM NaCl, 19 mM NH_4Cl). The resuspended cells were diluted into 1 liter of M9 medium containing 30 μg/ml kanamycin and 34 μg/ml chloramphenicol. The cell culture was grown at 37 °C with shaking to mid-log phase (A_{600} of 0.4-0.6). Selenomethionine (50 mg), isoleucine (50 mg), leucine (50 mg), valine (50 mg), lysine (100 mg), phenylalanine (100 mg), and threonine (100 mg) were added to the cell culture to shut down endogenous methionine synthesis. After an additional 15 minutes of growth at 37°C with shaking, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.8 mM to induce the expression of YedU. After induction, the cell culture was grown at 37 °C with shaking for 13 hours before cells were harvested by centrifugation. The selenomethionyl protein was purified and crystallized following the same procedures described in this chapter. Incorporation of selenium into YedU protein was confirmed by mass spectrometry. However, these crystals of selenomethionyl protein did not diffract well (resolution ~10 Å).

2.3 CRYSTALLIZATION AND CRYSTAL HANDLING

The crystallization screening was performed with a sparse-matrix screening kit (Hampton Research). In a hanging-drop vapor-diffusion setting at room temperature, needles and plates were obtained with reagent #6 (30% PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris HCl pH 8.5), #9 (30% PEG 4000, 0.2 M Ammonium Acetate, 0.1 M Na Citrate pH 5.6), and #39 (2.0 M $(NH_4)_2SO_4$, 2% PEG 400, 0.1 M Na HEPES pH 7.5) within two weeks. Subsequently, grid screenings were performed at different pH values (8.5, 8.0, 7.5, 7.0, 6.5, 6.0, and 5.5) and different precipitant concentrations (20%, 24%, 28%, 32% PEG 4000 or 1.4 M, 1.6 M, 1.8 M, 2.0 M $(NH_4)_2SO_4$). Single crystals were produced in droplets containing 3 μl of YedU protein (15 mg/ml) and 3 μl of reservoir solution containing 1.8 M $(NH_4)_2SO_4$, 5 mM DTT, 2% PEG 600, 0.1 M HEPES pH7.5. Typical size of the YedU crystals was ~ 0.10mm × 0.10mm × 0.15 mm (Figure 2-2). For cryoprotection, the crystals were soaked in cryo-buffers containing 2.0 M $(NH_4)_2SO_4$, 2%

PEG 600, 0.1 M HEPES pH 7.5 with increasing glycerol concentrations (v/v, 0.08%, 0.2%, 0.8%, 4%, 15%, and 30%). After the crystals were soaked briefly in cryo-buffers with 30% glycerol, they were frozen in liquid nitrogen.

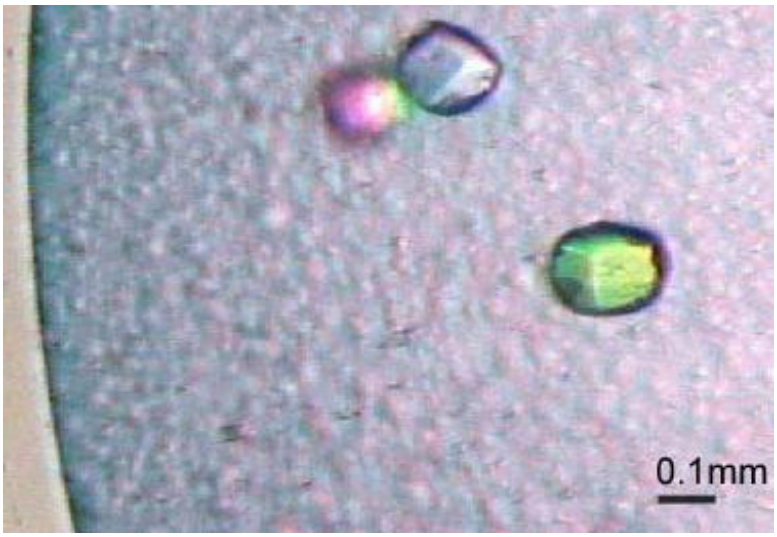

Figure 2-2 The YedU crystals

These crystals were growing in a hanging-drop setting with reservoir solution containing 1.8 M $(NH_4)_2SO_4$, 5 mM DTT, 2% PEG 600, 0.1 M HEPES pH 7.5.

2.4 PRELIMINARY CRYSTALLOGRAPHIC ANALYSIS

The goal of preliminary crystallographic analysis is to determine the diffraction quality of the YedU crystal, its lattice type and space group, and to estimate the number

of molecules in the unit cell. According to Bragg's law ($\lambda = 2d\sin\theta$; λ, wavelength; d, resolution; θ, reflection angle), the resolution of diffraction data is determined by radiation wavelength and crystal-to-detector distance for a given detector when the crystal gives measurable diffractions at the edge of the detector. The DENZO and SCALEPACK programs of the HKL package were used to index, integrate, and scale diffraction image/data (Otwinowski and Minor, 1997).

A few diffraction images were first collected from a YedU crystal at -273 °C on a MacScience rotating anode X-ray generator with a DIP2030 imaging plate. When the crystal-to-detector distance was set at 250 mm, the crystal diffracted to 2.9 Å at the edge of the image plate with an X-ray wavelength of 1.5418 Å. However, many diffraction spots overlapped each other, indicating that the YedU crystal has a long axis. In order to characterize the YedU crystal, the crystal-to-detector distance was set at 350 mm so that the diffraction spots were separated (Figure 2-3). Data were collected to 3.8 Å resolution for 90.0º (0.5º/image) and indexed with the DENZO program of the HKL package (Otwinowski and Minor, 1997). The lattice and unit-cell distortion table of the DENZO outputs indicated that the primitive hexagonal lattice was the highest symmetry with an acceptable distortion. Thus, the YedU crystal has a primitive hexagonal lattice with unit cell parameters of $a = b \approx 53.29$Å, $c \approx 348.01$Å, $\alpha = \beta = 90º$, and $\gamma = 120º$ (Table 2-1). The DENZO output files were scaled by the SCALEPACK program of the HKL package. Since there are 5 possible Laue groups ($P3_n$, $P3_n12$, $P3_n21$, $P6_n$, and $P6_n22$) for the primitive hexagonal lattice, each Laue group was tested and the SCALEPACK outputs indicated that the highest symmetry Laue group $P6_n22$ gave an R_{sym} value of 0.063, comparable with that of the primary Laue group $P3_n$. Systematic absences suggested that the YedU crystal belongs to the space groups $P6_122$ or $P6_522$ since only reflections (0, 0, 6n) were present along the screw axis c*. The handedness of the screw axis can not be determined at this stage without phase information.

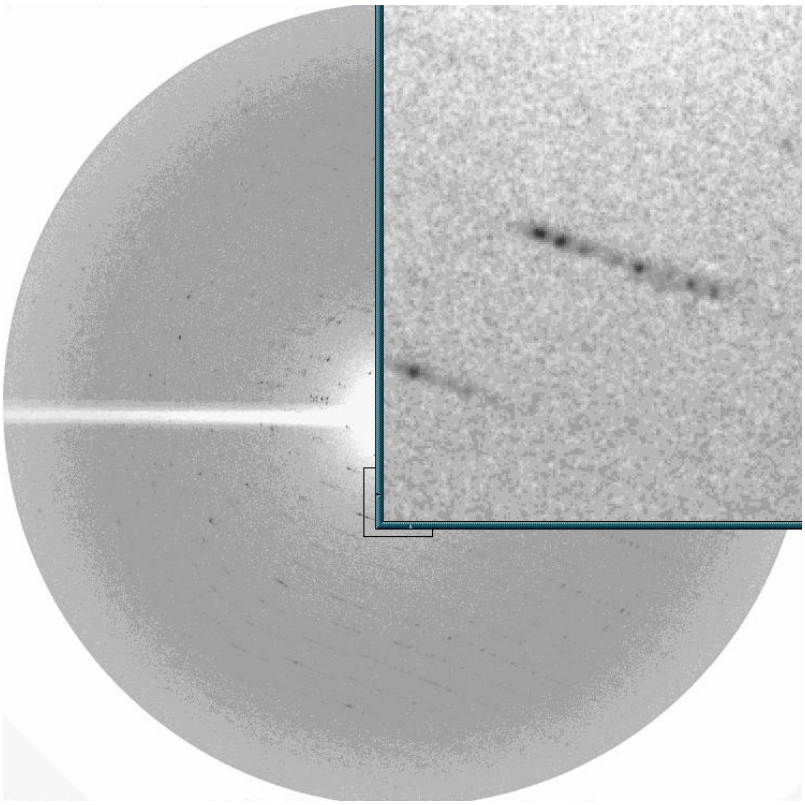

Figure 2-3 Diffraction pattern of the YedU crystal

The image was collected from a YedU crystal at -273 °C on a MacScience rotating anode X-ray generator with a DIP2030 imaging plate. the crystal-to-detector distance was set at 350 mm. The resolution limit at the edge of the imaging plate was 3.8 Å.

The Matthews coefficient has been widely used to estimate the number of molecules in the unit cell and in one asymmetric unit (Matthews, 1968). The Matthews coefficient (V_M) is the ratio of the unit cell volume and molecular weight. For most protein crystals, V_M is between 1.7 and 3.5 $Å^3$/Da with most values around 2.15 $Å^3$/Da. For space groups $P6_122$ or $P6_522$, there are 12 asymmetric units in a unit cell. The molecular weight of the YedU monomer is 31059 Da. The calculated V_M was 2.26

$Å^3$/Da (or 1.13 $Å^3$/Da) if there was one (or two) YedU monomers in an asymmetric unit. Because 1.13 $Å^3$/Da was too low, it was reasonable to estimate that there was one YedU monomer in an asymmetric unit, corresponding to 6 YedU dimers in a unit cell. The volume fraction of protein ($V_{protein}$) was approximately $1.23/V_M \times 100\% = 54.4\%$. The solvent content of the YedU crystals was $1 - V_{protein} = 45.6\%$.

Table 2-1 YedU crystal information

Space group	$P6_122$ or $P6_522$
Unit cell (Å)	
a = b	52.8
c	346.9
V_M ($Å^3$/Da)	2.26
Solvent content (%)	45.6

CHAPTER 3: DETERMINATION OF THE YEDU CRYSTAL STRUCTURE

This chapter describes procedures for the structure determination of the YedU stress protein, including a brief introduction to macromolecular crystallography and the multiple isomorphous replacement (MIR) method, preparation of derivatives, data collection and reduction, the MIR solution of phase angles, model building, and refinement. Identification of a metal-binding site is also described in this chapter.

3.1 INTRODUCTION

Macromolecular crystallography relies on the basic theory of X-ray diffraction by a single crystal, which was well understood before the development of modern molecular biology. A crystal is a three-dimensional periodic arrangement of molecules. The unit cell of a crystal is defined as its smallest unit, which, if repeated, could generate the whole crystal (Drenth, 1999). When a protein crystal is irradiated with a monochromatic X-ray beam, a three-dimensional array of atoms will diffract X-rays in a pattern that is determined by the spatial arrangement of the atoms (lattice) and with intensities that are properties of the atomic positions. Mathematically, the structure factor F(h k l) can be expressed as the summation over all atoms j in the unit cell,

$$F(h\ k\ l) = \sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \qquad (3.1)$$

where f_j is the atomic scattering factor of atom j; (x_j, y_j, z_j) is the spatial position of atom j; and (h k l) are the reflection indices that meet the Laue conditions (Drenth, 1999). The intensity of an individual (h k l) is proportional to the square of the amplitude of the structure factor F(h k l). Instead of summing over all separate atoms, we

can integrate over all electrons in the unit cell and express the structure factor as a function of the electron density $\rho(xyz)$ in the unit cell,

$$\mathrm{F(h\ k\ l)} = V\iiint_{xyz} \rho(xyz)\exp[2\pi i(hx_j + ky_j + lz_j)]dxdydz \qquad (3.2)$$

As a vector, F(h k l) can also be written as F(h k l)exp[iα(h k l)], where F(h k l) is its amplitude and α(h k l) is its phase angle. In macromolecular crystallography, the diffraction pattern is known experimentally and the goal is to calculate the electron density $\rho(xyz)$ at every position (x y z) in the unit cell. In Eq. 3.2, F(h k l) is the Fourier transform of $\rho(xyz)$. Since the Fourier transform is reversible, $\rho(xyz)$ can be calculated as the Fourier transform of all F(h k l),

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l |\,\mathrm{F}(hkl)\,|\ \exp[-2\pi i(hx_j + ky_j + lz_j) + i\alpha(hkl)] \quad (3.3)$$

Both the amplitude and the phase angle of all F(h k l) are required to calculate an electron density map. The amplitude of the structure factor is the square root of its intensity that is measured in a diffraction experiment. However, the phase angle of the structure factor can not be determined directly in a diffraction experiment. The determination of phase angles is well known as the "phase problem" in crystallography.

Three widely used methods have been developed to determine the phase angles, multiple isomorphous replacement (MIR) method, multiple wavelength anomalous dispersion (MAD) method, and molecular replacement (MR) method.

The MIR method has been widely used for solving the phase problem. An electron rich atom (usually heavy metals such as Hg, Au, Pb, and Pt) is attached to the protein molecule in a derivative crystal. If the heavy atom is located at specific positions in the unit cell, it will contribute to the diffraction of the derivative crystal. In the ideal situation, the conformation, position, and orientation of the protein molecule as well as the unit cell parameters are almost the same in the native crystal and in the derivative crystal. The diffraction intensity differences between the native and derivative pattern are mainly due to the attached heavy atoms. From these differences the positions of the

heavy atoms can be determined. The information of these heavy atom positions can then lead to an estimation of the phase angles of the structure factors in a Harker diagram (Harker, 1956).

With the advent of tunable synchrotron radiation and the *in vivo* selenomethionine substitution technique, the MAD method has been routinely used for phase determination in protein crystallography (Hendrickson, 2000). The MAD method utilizes the anomalous scattering of certain atoms at the wavelength close to their absorption edge. From dispersive and anomalous differences the anomalous scatterers are localized and preliminary phase angles are estimated. Beside selenium, other heavy atoms such as Fe, Zn, Mo, Pt, and Hg are also appropriate anomalous scatters that are frequently used in MAD phasing.

The MR method is based on the observation that proteins homologous in their sequence have a very similar fold. If the structure of such a homologous protein is known, the phase angles calculated from the known structure may be used as an initial solution that can subsequently be refined.

After the initial phase angles are obtained with MIR, MAD, or MR method, they can be further refined with density modification methods such as solvent flattening, molecular averaging (if noncrystallographic symmetry is present in an asymmetric unit), and histogram matching (Wang, 1985; Kleywegt, 1996; and Lunin, 1988). The refined phase angles are used to calculate an electron density map. If the quality of the electron density map is good, structural elements such as alpha-helices and beta-sheets are readily recognizable and a model corresponding to the amino-acid sequence of the protein can be built into the electron density.

Since no structure of homologous proteins that share 30% or more sequence identity to the YedU protein has been previously determined, the MR method is not feasible for YedU structure determination, thus, the phase problem must be solved experimentally either by MIR method or by MAD method. The attempt of utilizing the selenomethionine-based MAD method was stopped since the selenomethionine substituted YedU crystals diffracted poorly, as described in section 2.2.4. The crystal

structure of YedU was determined by obtaining phase angles using the multiple isomorphous replacement (MIR) method.

3.2 PREPARATION OF DERIVATIVES

Heavy-atom derivatives are usually prepared by soaking native crystals in solutions of heavy-atom ions. In many cases, heavy-atom ions bind to one or a few specific sites on the protein without perturbing its conformation or crystal packing. For example, Hg(II) is known to react readily with surface cysteine residues and Pt can from stable adducts with cysteine, histidine, and methionine. The search for isomorphous derivatives is empirical and requires screening different ionic compounds at various concentrations and pH values. Several criteria define a good heavy-atom derivative. First, the derivative crystals should be isomorphous with the native crystals, i.e., the heavy atom does not disturb the conformation of the protein or crystal packing. Unit cell dimensions are quite sensitive to such disturbances, so an isomorphous derivative crystal must have the same unit-cell dimensions as those of the native crystal. Practically, a change in the unit-cell dimensions of one fourth of resolution limit is tolerable. Sometimes, nonisomorphism can occur without seeing a change in unit-cell dimensions. Poor refinement of heavy-atom parameters can be indicative of this kind of nonisomorphism. Second, diffraction changes must be measurable between the derivative data and native data in at least a modest number of reflections. The phase angles are derived from these diffraction changes, so they must be clearly detectable and large enough to measure accurately. Practically, when a few frames of a potential derivative are scaled with a native data set, χ^2 (goodness of fit) will indicate the power of a derivative. If the derivative crystal is isomorphic with native crystals, a χ^2 value of ~10 indicates that it is a promising derivative whose diffraction intensities are significantly different from those of the native crystal. Finally, derivative crystals usually do not diffract as well as native crystals, but they must diffract to a reasonably high resolution in

order to obtain a set of initial phase angles that can lead to an interpretable electron density map.

The YedU protein has two cysteine residues. Several mercurial compounds were tested in order to obtain mercurial derivative crystals. The first derivative was a crystal soaked overnight in mother liquors containing 1 mM $HgCl_2$, then soaked in cryo-buffers containing increased concentrations of glycerol (v/v, 0.08%, 0.2%, 0.8%, 4%, 15%, and 30%) without $HgCl_2$. Each soaking step took about 30 minutes. After the crystals were soaked in cryo-buffers containing 30% glycerol, they were frozen in liquid nitrogen. The $HgCl_2$ derivative was isomorphous with native crystals. However, the occupancy of Hg turned out to be too low and the B-factor was high, indicating that the Hg was not well ordered. To increase the occupancy, the native crystals were washed with mother liquors several times to remove residual dithiothreitol (DTT) before adding an organic mercurial compound. After washing, crystals were soaked overnight in mother liquors containing of 1 mM p-chloromercuribenzoate (PCMB) and then soaked in cryo-buffers containing increased concentration of glycerol (v/v, 0.08%, 0.2%, 0.8%, 4%, 15%, 30%; 30 minutes for each step) without PCMB. The PCMB soaked crystals were frozen in liquid nitrogen and they yielded a derivative model where the Hg had a lower B-factor, indicating that it is well ordered. Compared with $HgCl_2$, it is likely that the aromatic ring of PCMB restricts the mobility of the Hg, thus keeping it well ordered in the crystals.

Recently, solvent halide ions have been used to obtain derivative crystals (Dauter and Dauter, 2001). Traditional methods of derivative preparation usually require soaking of crystals in low concentrations of heavy-atom solutions (~1 mM) for a period of time ranging from a few hours to a few weeks. Low heavy-atom concentrations are imperative because crystals are usually unstable in the presence of high concentrations of heavy atoms. As a result, a long soaking time is required to reach equilibrium in low concentrations of heavy atoms. The ability to freeze crystals for data collection makes it possible to prepare derivatives by soaking in higher concentrations for shorter time since any degradation of the crystal can be halted by freezing in liquid nitrogen.

To prepare iodide derivatives, YedU crystals were soaked in cryo-buffers containing of increased concentration of glycerol (v/v, 0.08%, 0.2%, 0.8%, 4%, 15%; 30 minutes for each step). The crystals were then transferred into cryo-buffers containing 160 mM KI and 30% glycerol. The soaking time was relatively short (<5 minutes) before the crystals were frozen in liquid nitrogen. In order to keep the crystal isomorphous, the $(NH_4)_2SO_4$ concentration in the cryo-buffers was adjusted for the KI addition to keep the ionic strength unchanged.

3.3 DATA COLLECTION AND REDUCTION

Data collection is an important step in macromolecular crystallography since all the structure information is derived from these diffraction data. The more completely and accurately the reflections are measured, the more reliable the determined structure will be. A native dataset was collected at 2.65 Å resolution in-house at the UTMB X-ray crystallography center. A mercurial derivative dataset and an iodide derivative dataset were collected at 2.8 Å resolution. A mercury site and several iodine sites were located by the SOLVE program (Terwilliger and Berendzen, 1999) and an interpretable electron density map was calculated. To further improve the quality of the electron density map, diffraction data of higher resolution were collected at the PX-1 protein crystallography beamline at The Center for Advanced Microstructures and Devices (CAMD, Baton Rouge, LA). The resolutions of the native dataset, the mercurial derivative dataset, and the iodide derivative dataset were extended to 2.2 Å, 2.5 Å, and 2.3 Å, respectively, at the PX-1 beamline (Figure 3-1). Since the YedU crystals have a long unit-cell (c = 347.40 Å), the crystals must be aligned properly in order to avoid reflection overlaps. Before each data collection, several test frames were collected to determine the orientation of crystal c-axis. The crystal was then aligned so that the angle between the crystal c-axis and the goniometer rotation axis was in the range of 15°-30°. The PX-1 protein crystallography beamline is equipped with a 165 mm CCD detector (MAR Research).

The crystal-to-detector distance was set at 200 mm with a 20° 2-theta offset, corresponding to 2.0-Å resolution at the edge of the detector. Although all unique reflections can be collected in 30° for space group $P6_122$ or $P6_522$, diffraction images were recorded every 0.5° for more than 100° to increase redundancy. This also compensated for the redundancy lost upon the 20° 2-theta positioning of the detector. The diffraction data were first indexed by the DENZO program and scaled by the SCALEPACK program of the HKL package (Otwinowski and Minor, 1997). An $I/\sigma(I)$ of 2.0 in the outer resolution shell was used as the cutoff to determine the resolution limits of each dataset. The statistics and relevant information are listed in Table 3-1.

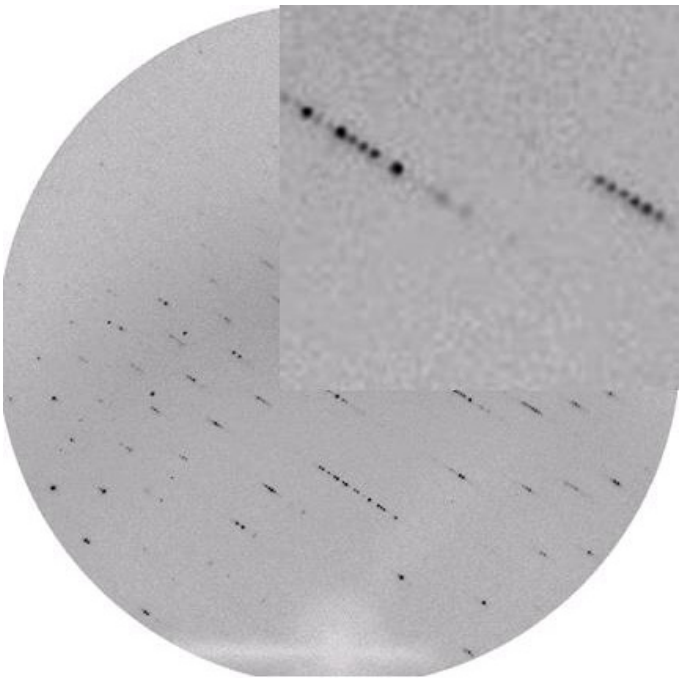

Figure 3-1 A diffraction image of the YedU crystal

The image was collected from a native YedU crystal at the PX-1 protein crystallography beamline (Baton Rouge, LA). The crystal-to-detector distance was set at 200 mm with a 20° 2-theta offset. The resolution limit at the edge of the detector was 2.0 Å.

Table 3-1 Data collection and statistics

Dataset	Native	PCMB (Hg)	KI
Space group	$P6_122/P6_522$	$P6_122/P6_522$	$P6_122/P6_522$
Unit cell (Å)			
a = b	52.98	52.85	53.02
c	347.40	347.40	347.22
Resolution range (Å)	90-2.2	90-2.5	90-2.3
(Outer shell)	(2.25-2.2)	(2.56-2.5)	(2.35-2.3)
Wavelength (Å)	1.5418	1.008	1.5418
Measurements	290299	228337	202376
Unique reflections[1]	15151(779)	9885(496)	11811(421)
Completeness[1]	94.5(77.5)	90.1(73.5)	83.4(48.3)
$I/\sigma(I)$ [1]	14.7(3.9)	12.7(2.3)	13.6(2.5)
R_{sym}[1,2]	0.075(0.319)	0.107(0.664)	0.105(0.581)

[1]Values for the outer resolution shell of data are given in parentheses.

[2]$R_{sym} = \sum\sum | I_i - I_m | / \sum\sum I_i$, where I_i is the intensity of the measured reflection and I_m is the mean intensity of all symmetry-related reflections.

3.4 THE MIR SOLUTION OF PHASE ANGLES

After the native data and derivative data were collected, indexed and scaled, the first step in solving the phase problem by the MIR method is to locate the heavy atoms inside the derivative crystals. For an isomorphous derivative, a difference Patterson map is calculated using the diffraction difference between the native and derivative dataset (Drenth, 1999).

$$P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}(F_{\mathrm{PH}} - F_{\mathrm{P}})^2 \exp[-2\pi i(hu + kv + lw)] \qquad (3.4)$$

The difference Patterson map will have peaks corresponding to vectors between symmetry-related heavy-atom positions on the Harker sections. In the case of multiple heavy-atom binding sites, the difference Patterson map will also have peaks corresponding to vectors between different heavy-atom positions. Heavy-atom coordinates can be calculated from the positions of peaks in the difference Patterson map either manually or by computer programs such as SOLVE (Terwilliger and Berendzen, 1999). Finding heavy-atom sites for YedU derivative crystals was straightforward since the SOLVE program located a single mercury binding site. Once the mercury-binding site was known, the iodine-binding sites in the iodide derivatives could be located by using the difference Fourier technique. The SOLVE program located 7 iodine-binding sites. The MIR phases were calculated by the SOLVE program using both mercurial and iodide derivative data sets. The overall Figure of Merit (FOM) at the resolution of 2.4 Å was 0.40. The FOM is a statistical factor indicating the quality of the phase angles. For a given reflection (h, k, l), FOM is defined as:

$$\mathrm{FOM} = |F(h, k, l)_{\mathrm{best}}| / |F(h, k, l)| \qquad (3.5)$$

Where $F(h, k, l)_{\mathrm{best}}$ is the centroid of the probability distribution of F(h, k, l) and F(h, k, l) is the observed amplitude of the structure factor. The overall FOM is the average of FOM for all individual reflections. The phase angles were further refined by

the RESOLVE program using the techniques of density modification (Terwilliger, 2000). At this point, the overall FOM to 2.4 Å resolution had increased to 0.61. Using the refined phases angles, the RESOLVE program calculated an electron density map that was used to build a 3-dimension (3-D) atomic model of the YedU stress protein (Figure 3-2).

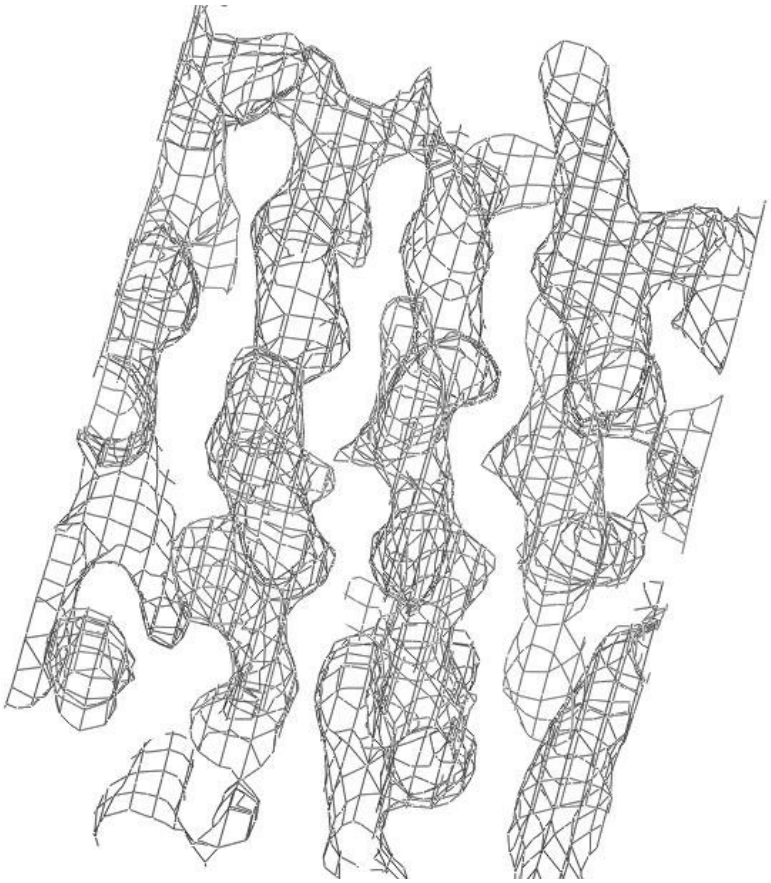

Figure 3-2 The electron density map of the YedU stress protein

The 2.4 Å resolution electron density map was calculated from MIR phase angles following solvent flattening. This map is contoured at 1.0 σ.

3.5 MODEL BUILDING AND REFINEMENT

The final step and the ultimate goal of macromolecular crystallography are to get an accurate atomic model that corresponds to the 3-D structure of the macromolecule of interest. The RESOLVE (Terwilliger, 2000) and TEXTAL (Holton et al. 2000) programs built two fragmented models of modest quality, each containing backbone atoms for less than half of the residues. The model building of YedU was done using the graphic program XtalView (McRee, 1999), starting from the partial models generated by the RESOLVE and TEXTAL programs. The core region of the protein was well ordered and some secondary structural elements such as helices and strands could readily be identified in the electron density map calculated by the RESOLVE program (Figure 3-2). It was obvious that YedU contains a core domain of an alpha/beta/alpha sandwich fold. Thus, the crystal structure of the PH1704 protease was used to guide the model building of this region (Du et al., 2000; PDB accession 1G2I). However, the electron density was poorly connected between some secondary structural elements. Initially, fragments of polypeptide were constructed into the electron density of the well-ordered region. It is very important to avoid model bias in the process of model building. If the electron density for an individual residue was not convincing enough to determine its side chain rotamer, an alanine or glycine residue was built into its position instead. The initial electron density map from the RESOLVE program was always referred to as an unbiased "experimental map" throughout the refinement. The starting fragmented model was refined by the CNS program using maximum-likelihood and simulated-annealing methods (Brunger et al., 1998). Before refinement, the CNS program randomly selects 5% of the observed reflections and mark them as test set (T). The remaining 95% of observed reflections are marked as working set (W) and refinement was carried out with the working set only. During refinement, the quality of the model was monitored by R-free (the free R-factor), which is defined as:

$$\text{R-free} = \frac{\sum_{hkl \subset T} || F_{obs} | - k | F_{calc} ||}{\sum_{hkl \subset T} | F_{obs} |} \qquad (3.6)$$

where F_{obs} and F_{calc} are the observed and calculated amplitude of a given reflection (h, k, l) in the test set and k is a scale factor. After the R-free converged in each refinement round, a set of combined phases were calculated from two sets of phases, one from the refined partial structure model and one from the experimental electron density map. Usually the poor electron density regions were improved in a new electron density map calculated from the combined phases and more residues could be built into the partial structure model. The above procedure was repeated many times until the R-free dropped to less than 30% when the model was refined against the native dataset in the resolution range of 30.0-2.2 Å. At this stage, the structure model included almost all the residues of the YedU protein and water molecules were gradually added to the model.

3.6 IDENTIFICATION OF A METAL-BINDING SITE

After most ordered water molecules were included in the model, a difference Fourier map (Fo - Fc) showed that there is a strong density between the side chains of His85, Glu90, His122, and their symmetry equivalents on a crystallographic two-fold axis (Figure 3-3). The density was too strong to be explained by water molecules. Metal ions such as Mn(II), Fe(II), Zn(II), Co(II), Ni(II), and Cu(II) were tested. Two symmetry-related zinc(II) ions fit the density best.

X-ray absorption at the zinc K-edge was later measured at the CAMD PX-1 protein crystallography beamline. The absorbance of X-rays by YedU crystals was scanned from 9630 to 9700 eV in steps of 1 eV. The spectra, although noisy, exhibited a clear K-edge absorption over a 7 eV range centered at 9666 eV (Figure 3-3), which was very close to the K-edge of zinc (9660 eV). The cryo-buffers alone did not exhibit such a zinc K-edge absorbance, suggesting that the zinc is associated with the protein.

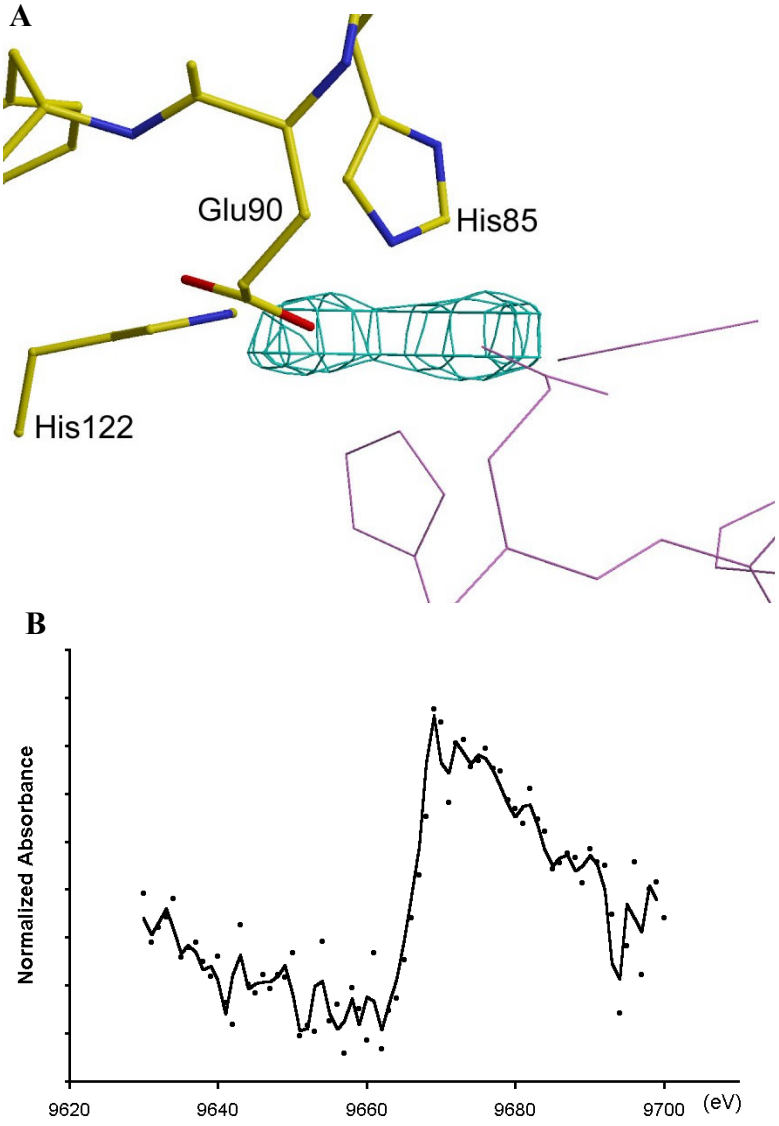

Figure 3-3 Identification of a metal-binding site

(*A*) A view of the Fo-Fc difference electron density. The map is contoured at 6.0 σ. Residues His85, Glu90, and His122 are shown as sticks. Their symmetry-related equivalents are also shown (thin sticks). (*B*) The K-edge X-ray absorption of the YedU crystal.

3.7 FINAL REFINEMENT STATISTICS

The final YedU structure model contains 278 residues, a zinc(II) ion, and 120 ordered water molecules. The R-free is 24.30% (Table 3-2). PROCHECK (Laskowski et al., 1993) indicates that the majority of the residues (99.2%) are in the favorable and allowed regions of the Ramachandran plot. Two residues, Cys184 and Arg254, are in the disallowed regions and their phi-psi (ϕ/ψ) angles are 57°/-115° and 65°/-117°, respectively. Cys184 is located at the turn of a nucleophile elbow-like motif. An unusual ψ angle between -100° and -150° has been one of the main characteristics of the active-site nucleophile of alpha/beta hydrolases (Hakansson, 2002), which will be discussed in section 4.5. Arg254 is located on a tight turn between two beta-strands. Both Cys184 and Arg254 were examined carefully, and their conformations were consistent with the experimental electron density map.

Table 3-2 Refinement statistics

Resolution (Å)*	30.0-2.2 (2.25-2.2)
Total reflections (working / test)*	14142 (750) / 714 (40)
Number of molecules	
Protein residues	278
Water	120
Zinc(II) ion	1
R-factor / R-free (%)	20.30 / 24.30
RMSD from ideality	
Bond lengths (Å)	0.006
Bond angles (°)	1.3
Ramachandran analysis (%)	
Most favored regions	87.3
Additional allowed regions	11.4
Generously allowed regions	0.4
Disallowed regions	0.8

*Values for the outer resolution shell of data are given in parentheses.

CHAPTER 4: THE YEDU CRYSTAL STRUCTURE

This chapter describes the overall structure of the YedU protein and the potential active sites revealed by the structure. Section 4.1 describes the tertiary structure. Section 4.2 describes the quaternary structure. Structural comparison with class II and class III orthologs is in section 4.3. Section 4.4 describes a metal-binding site. Section 4.5 describes a putative catalytic triad.

4.1 THE TERTIARY STRUCTURE OF THE YEDU MONOMER

Except for the first 3 residues and the last residue, the backbones of all other residues are traceable in the electron density map and included in our model. The overall secondary structure of the YedU protein is a mixture of helices and strands (Figure 4-1). The globular structure (~ 50Å × 45Å × 30Å) can be subdivided into two domains, domain 1 (D1) consisting residues 34-56, 71-209, and 230-282, and a smaller second domain (D2) consisting residues 1-29, 57-70, and 210-229. Domain 1 is an alpha/beta/alpha sandwich fold that is characteristic of DJ-1/ThiJ/PfpI domains. The central beta-sheet consists of seven strands (S7, S6, S3, S8, S9, S13, and S12), with S12 as the only anti-parallel strand (Figure 4-1). The central β-sheet is flanked by helices H2, H3, H4, and H11 on one side and by helices H5, H6, H7, H8, and H10 on the other side. Domain 2 is composed of an anti-parallel beta-sheet (S1↑-S2↓-S5↓-S4↑) and two helices, H1 and H9.

Interactions between domains 1 and 2 are extensive and bury ~2,200 $Å^2$ of solvent-accessible surface area. The inter-domain interactions are mainly hydrophobic: Pro20 and Leu25 of the D1 domain make van der Waals interactions with Trp106 of the D2 domain; Leu60 makes contact with Leu157 and Ile158; Trp228 contacts Pro161 and

Leu191. The side chain of Asp210 makes hydrogen bonds with the backbone amide of Gly231 and Glu232, and the side chain of Asp213 makes hydrogen bonds with the imidazole ring of His154 and His185. Between the D1 and D2 domains, there is a narrow tunnel that connects the bulk solvent with a potential nucleophile Cys184, which will be described in section 4.5.

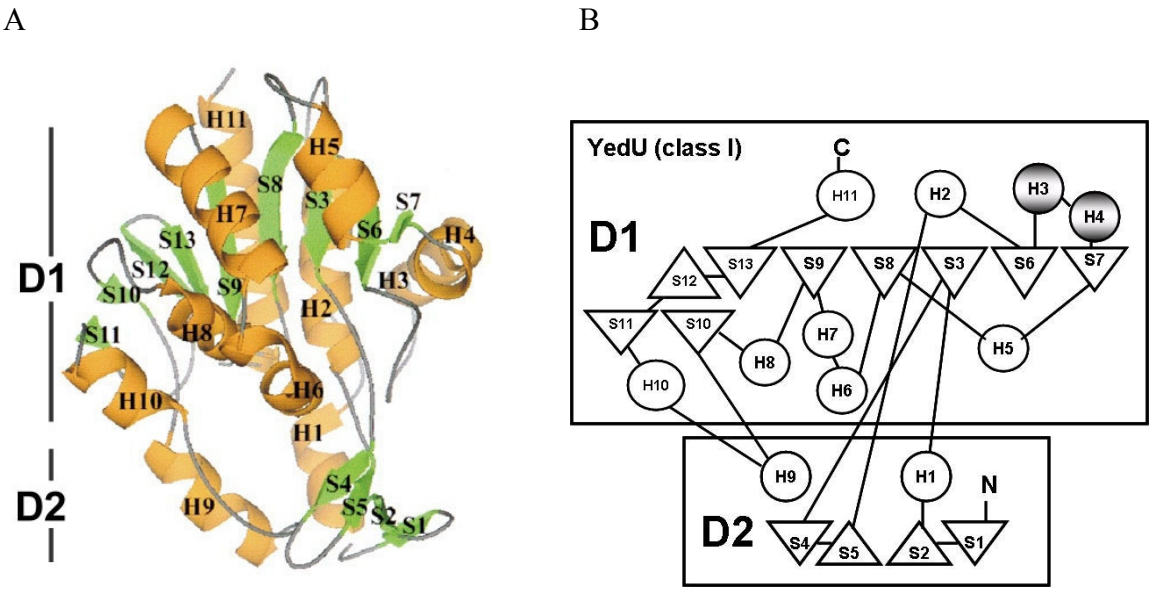

Figure 4-1 The tertiary structure of the YedU monomer

(*A*) Ribbon representation of the YedU monomer. "H" and "S" refer to helix and strand that are colored green and orange, respectively. (*B*) Schematic topology diagram of the YedU monomer. Triangles represent beta-strands. Circles represent alpha-helices. The upper part of the structure is a sandwich domain (D1) composed of a central 7-strand β-sheet flanked by helices. Below the sandwich domain, there is a smaller domain (D2) composed of an anti-parallel β-sheet (S1↑-S2↓-S5↓-S4↑) and two helices (H1 and H9).

4.2 QUATERNARY STRUCTURE OF THE YEDU PROTEIN

Two crystal contacts (Figure 4-2), each at a crystallographic two-fold axis, bury a significant amount of surface area calculated by the GETAREA program (Fraczkiewicz and Braun, 1998).

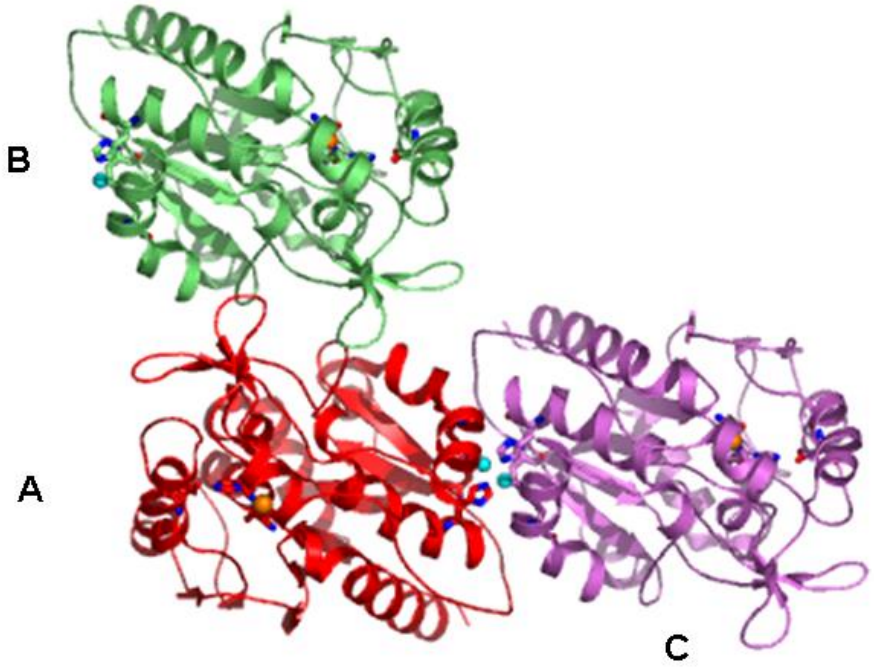

Figure 4-2 The YedU crystal contacts

Monomers A, B, and C are colored red, lime, and violet, respectively. The zinc(II) ions are shown as cyan spheres.

The zinc-binding site is presented in the contact between monomers A and C, which buries about 2000 $Å^2$ of solvent-accessible surface area (1000 $Å^2$ from each monomer). The side chains of His85, Glu90, His122, and their symmetry equivalents coordinate two zinc(II) ions at the center of the contact. Together with the zinc coordinating residues, the van der Waals interactions between residues Pro41, Pro43, Tyr82, Ala86, Pro117, Gln121, Ser124, Asn128, and their symmetry equivalents form the core of the AC contact with about 1300 $Å^2$ of buried surface areas. Side chains of additional residues forming the peripheral part of this contact do not pack well and some are indeed disordered. These disordered side chains include those of Lys42, Arg45, Lys47, Lys49, Glu92, Lys130, and Glu144 (Figure 4-3). Most residues involved in this contact are not conserved among class I orthologs (Figure 4-4).

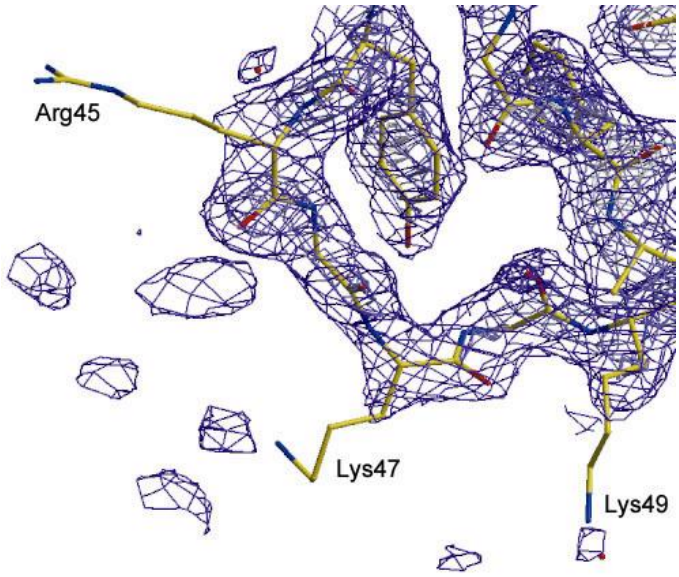

Figure 4-3 A view of the electron density around residues 45-49

The sigma-A weighted 2Fo-Fc electron density map is contoured at 1.0 σ. The side chains of Arg45, Lys47, and Lys49 are shown as sticks.

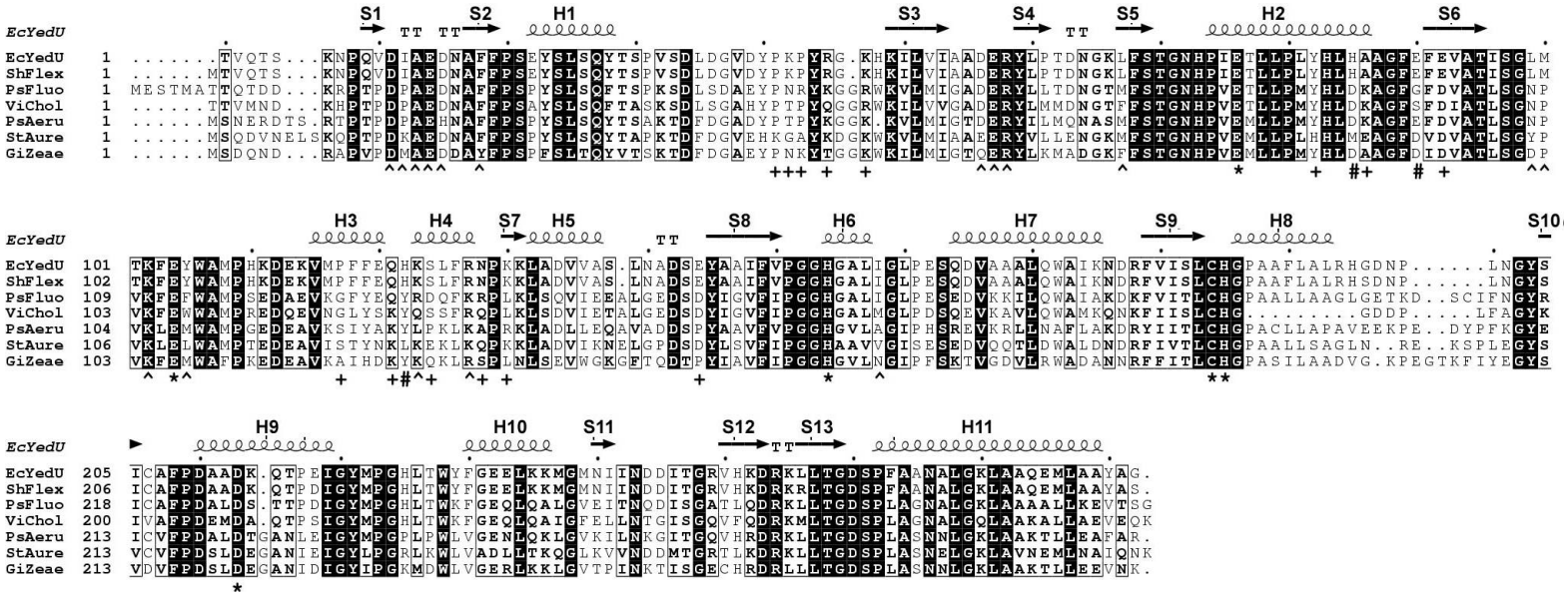

Figure 4-4 Sequence alignment of class I YedU orthologs

Abbreviations are: EcYedU, *Escherichia coli* YedU protein (gi:2506683); ShFlex, *Shigella flexneri* (gi:24113342, percentage identityto the YedU protein 98%); PsFluo, *Pseudomonas fluorescens* (gi:23059816, 62%); ViChol, *Vibrio cholerae* (gi:9658312, 58%); PsAeru, *Pseudomonas aeruginosa* (gi:11347751, 56%); StAure, *Staphylococcus aureus* (gi:15923541, 56%); GiZeae, *Gibberella zeae* (gi: 42549093, 50%). The secondary structures of YedU are shown above the sequences. Residues involved in the AB contact and the AC contact are indicated by "^" and "+", respectively, below the sequences. Zinc-coordinating residues are marked with "#". Residues that are hydrogen-bonded with the putative catalytic triad residues are marked with "*".

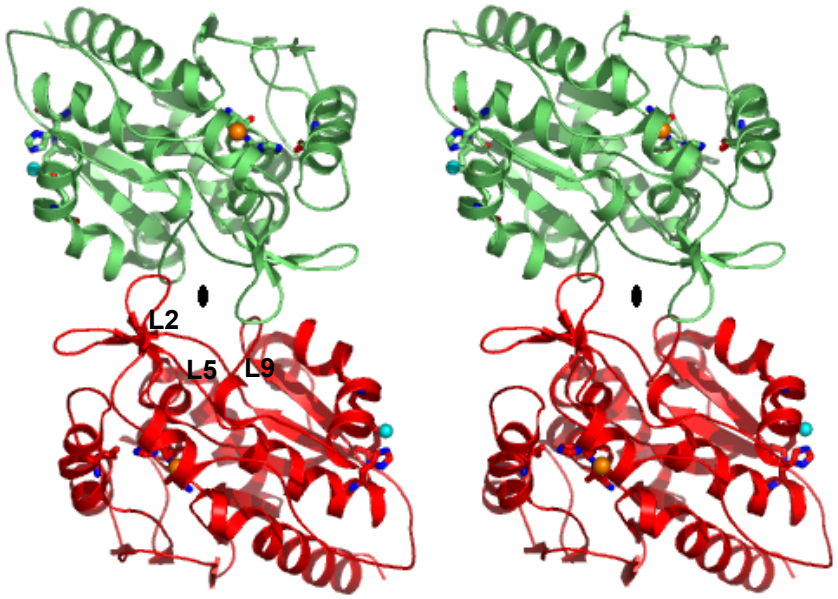

Figure 4-5 A stereo view of the YedU homodimer

Subunits A and A' are colored red and lime, respectively. The putative catalytic triad residues and the zinc-coordinating residues are shown as sticks. The Sg atom of Cys184 and the zinc(II) ion are shown as orange and cyan spheres, respectively.

The contact between monomers A and B also occurs on a two-fold axis and buries about 2080 $Å^2$ of solvent-accessible surface areas. Most of the residues involved in this contact are located on three secondary structural elements, β-turn L2 between strands S1 and S2, loop L5 between strands S3 and S4, and loop L9 between strand S6 and helix H3 (Figure 4-5). The β-turn L2 inserts into the other subunit by loop "handshaking," making intensive contact with the other subunit. L5 is located at the center of the contact and interact with L2, L5, and L9 of the other subunit. Seven residues (Leu99 - Tyr105) of L9 are buried or partially buried; they make contacts with L2, L5, and strand S5 of the other

subunit. There are three pairs of salt bridges between Glu14 (of L2) and Arg58' (L5), Asp15 (L2) and Lys123' (H4), Glu57 (L5) and Lys102' (L9), and their symmetry-related equivalents. Leu67 of strand S5 is also buried. Other residues that are partially buried include Phe19 and Ile158. Most of these residues making the contact are conserved or conservatively substituted throughout class I orthologs (Figure 4-4).

Size exclusion chromatography indicates that YedU forms dimers in solution. Although the buried surface areas are comparable in two crystal contacts, the AC contact is less intensive because many side chains that form the peripheral part of it are indeed disordered (Figure 4-3). Thus, the more conserved and better packed AB contact most likely occurs in the solution dimer (Figure 4-5). The D2 domain of subunit A contacts both the D1' and D2' domains of the other subunit. No D1-D1' interactions are present at dimer interface.

The crystal structure of YedU described here was published and deposited in the Protein Data Bank (Zhao et al., 2003; PDB accession 1ONS). The crystal structure was also determined independently by two other research groups (Quigley et al, 2003; PDB accession 1N57; Lee et al., 2003b; PDB accession 1IZY). Although the crystal forms are different (Table 4-1), all structures are essentially the same in terms of the monomer structure. Our structure 1ONS is consistent with structure 1N57 in identifying the preferred dimer interface. In addition, we identified the zinc binding site, which was not reported in structure 1N57. The absence of the zinc(II) ion in structure 1N57 is likely due to the presence of ethylenediaminetetraacetic acid (EDTA) in their crystallization experiments (Table 4-1). The structure 1IZY also reported the metal binding site, although it did not include Glu90 as one of the coordinating residues (Lee et al., 2003b). In contrast to structures 1ONS and 1N57, structure 1IZY reported the AC contact as their preferred dimer interface, which appears to be incorrect. Recently, a new crystal form of YedU was reported (Quigley et al, 2004; PDB accession 1PV2). In this crystal form, a triclinic cell (space group P1) contained four dimers of YedU. The AB contact was observed in the new crystal form and the AC contact was not, confirming that the AB contact should be the preferred dimer interface.

Table 4-1 Different crystal forms of the YedU protein

PDB ID	1ONS	1N57	1IZY	1PV2
Space group	$P6_122$	C2	$P6_5$	P1
Unit cell (Å)				
a, b	52.98, 52.98	52.15, 82.00	52.97, 52.97	54.5, 99.0
c	347.40	64.48	314.40	116.8
Crystallization condition	pH 7.5 1.8 M $(NH_4)_2SO_4$ 2% PEG 600 5 mM DTT	pH 6.5 25% PEG 3350 1 mM EDTA 4 mM DTT 300 mM $MgCl_2$	pH 8.5 2.0 M $(NH_4)_2SO_4$	pH 7.5 26% PEG 6000 1 mM EDTA 50 mM NaCl

4.3 STRUCTURAL COMPARISON WITH ORTHOLOGS

A DALI search (Holm and Sander, 1995) revealed that YedU was most structurally similar to the yeast YDR533Cp protein (Z score = 23.0, C-alpha RMSD = 2.6 Å, 22% sequence identity), the *Pyrococcus horikoshii* PH1704 protease (Z score = 17.3, RMSD = 2.0 Å, 18% sequence identity), and the human DJ-1 protein (Z score = 18.0, RMSD = 2.4 Å, 13% sequence identity). YDR533Cp, PH1704, and DJ-1 belong to class II, class IIIA, and class IIIB YedU orthologs, respectively. Superposition of YedU, YDR533Cp, PH1704, and DJ-1 monomers shows the high degree of structural similarity found in the DJ-1/ThiJ/PfpI domains of these proteins (Figure 4-6). The overall alpha/beta/alpha sandwich fold is well preserved in these proteins. However, at the edge of the alpha/beta/alpha sandwich fold, the Val115-Arg127 region of YedU and the corresponding region of YDR533Cp fold as helices (H3 and H4), whereas a beta hairpin is present in the corresponding region of PH1704 and DJ-1 (Figure 4-7). In addition, PH1704 and DJ-1 lack the smaller second domain (D2) that is formed by insertion

residues in YedU and YDR533Cp. Although the smaller second domain is also present in YDR533Cp, its fold is different from that of YedU. The human DJ-1 protein has an extra C-terminal helix (Figure 4-7).

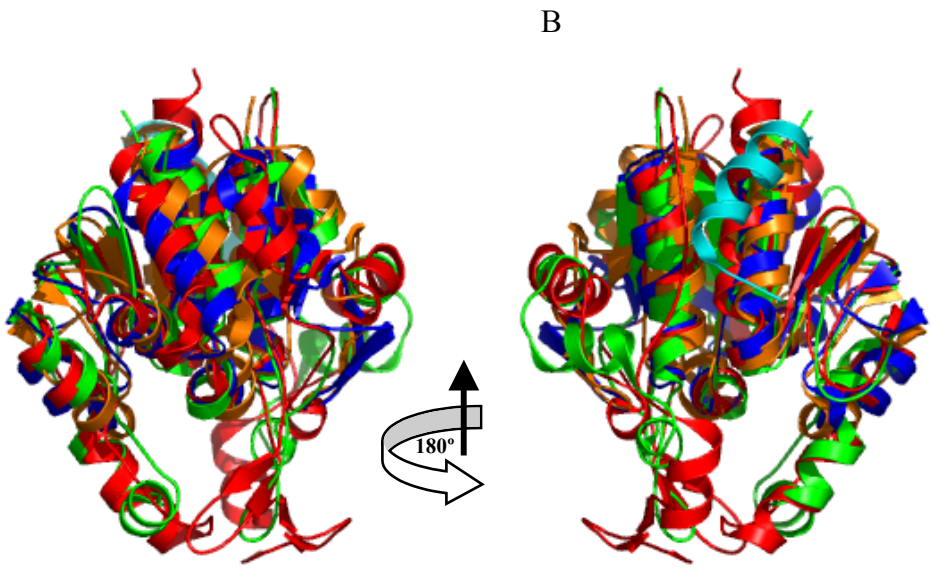

Figure 4-6 Overlay of YedU, YDR533Cp, PH1704, and DJ-1 monomers

(*A*) Superposition of YedU (red), YDR533Cp (green), PH1704 (blue), and DJ-1 (orange) monomers. (*B*) The superposition in a different view after 180º rotation. The extra C-terminal helix of DJ-1 is colored cyan.

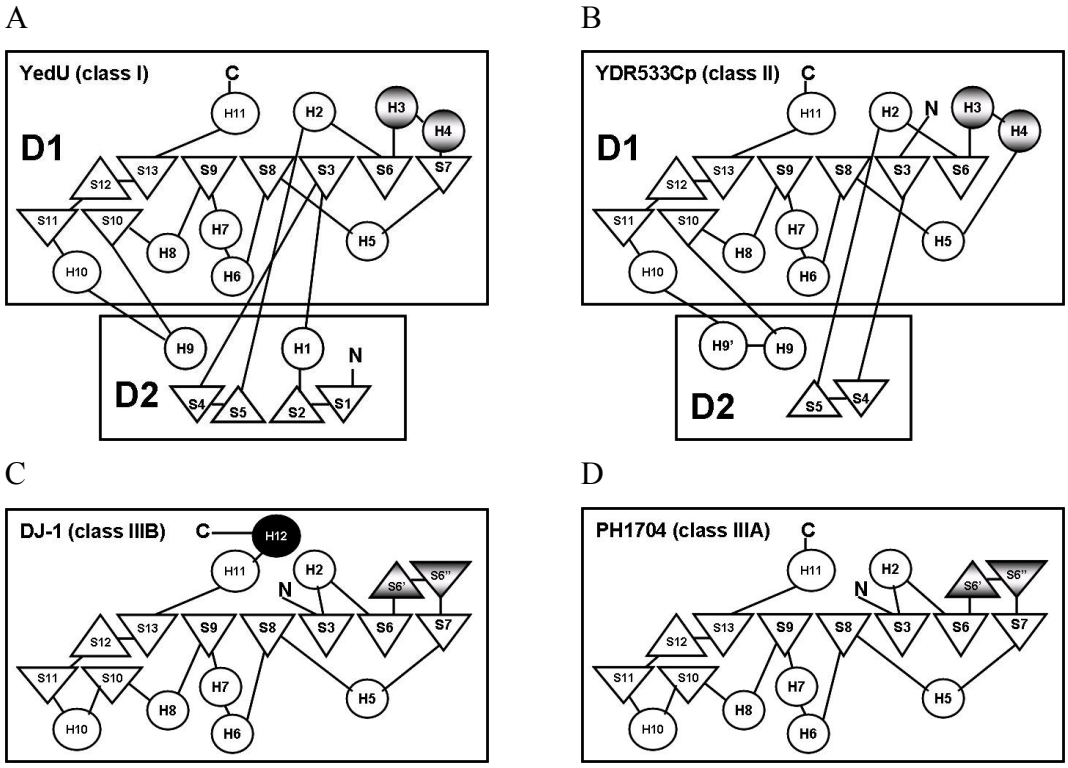

Figure 4-7 Topology diagrams of YedU, YDR533Cp, PH1704, and DJ-1

Triangles and circles represent beta-strands and alpha-helices, respectively. Secondary structural elements of YDR533Cp, PH1704, and DJ-1 are named as their equivalents in YedU.

The quaternary structures of YedU, YDR533Cp, PH1704, and DJ-1 are different. The dimer interface of YedU is not seen in YDR533Cp, PH1704, and DJ-1 because they all lack the N-terminal extension that contains most of the residues involved in the dimerization of YedU (Figure 4-8A).

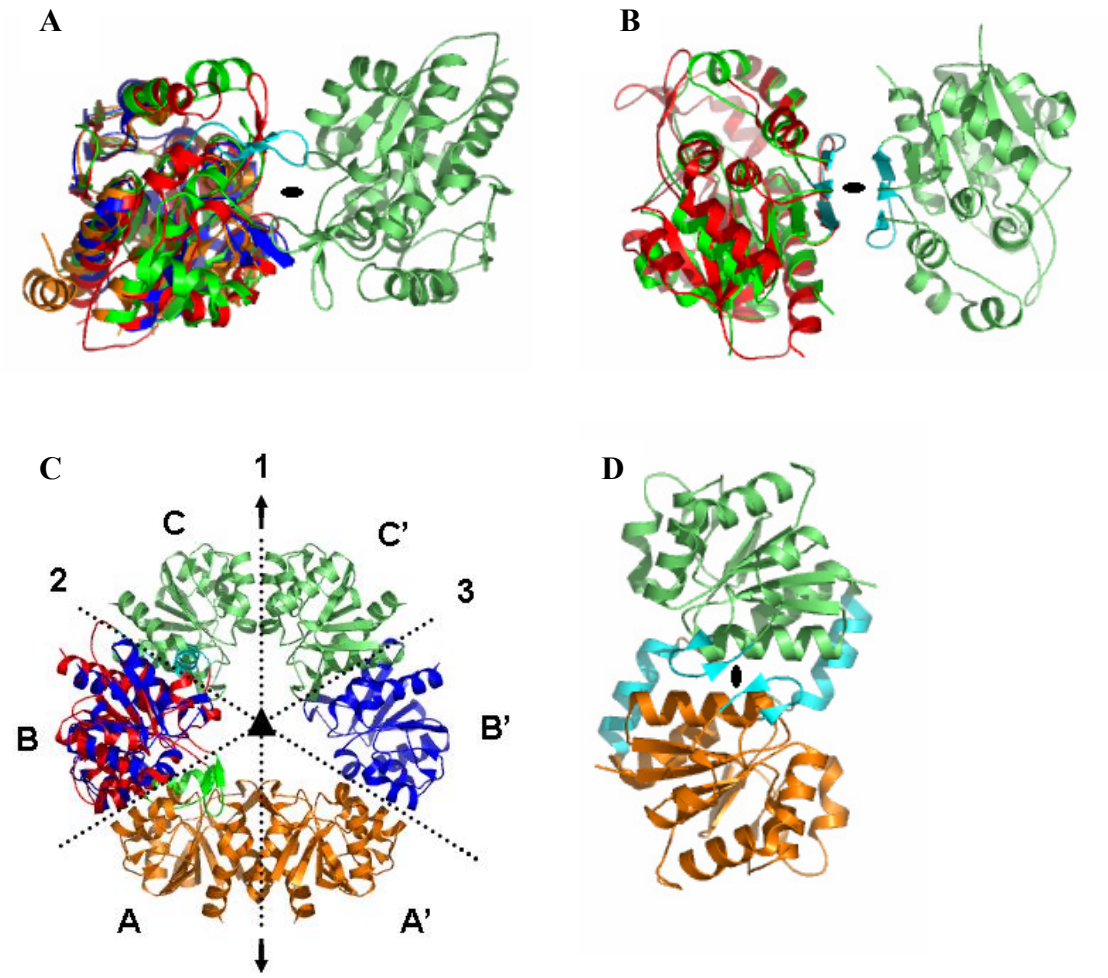

Figure 4-8 The quaternary structures of YedU orthologs

(*A*) The YedU homodimer. Monomers A and A' are colored red and lime, respectively. The beta hairpin S1-S2 is colored cyan in monomer A. The monomers of YDR533Cp (green), PH1704 (blue), and DJ-1 (orange) are superposed onto monomer A. (*B*) The YDR533Cp homodimer. Monomer A and A' are colored green and lime, respectively. Strands S10, S11, and S12 at the dimer interface are colored cyan in both monomers. The YedU monomer (red) is superposed onto monomer A. (*C*) The PH1704 hexamer. It is a trimer of dimers. The YedU monomer (red) is superposed onto monomer B. The secondary structures of YedU that prevent the AB and BC contacts are colored green and cyan, respectively. (*D*) The DJ-1 homodimer. Monomers A and A' are colored orange and lime, respectively. The beta hairpin S6'-S6'' and the H12 helix are colored cyan.

YDR533Cp also forms dimers in solution. Dimerization of YDR533Cp involves contacts between strands S10, S11, and S12 on each monomer (Wilson et al., 2004). These strands from both monomers form a quasi-six-stranded beta-barrel at the dimer interface (Figure 4-8B). Although strands S10, S11, and S12 are also present in YedU, the beta-barrel dimer interface is not seen in YedU. The dimer interface of YDR533Cp is modest and buries ~1,080 $Å^2$ surface areas, comparing with ~2,080 $Å^2$ surface areas buried at the dimer interface of YedU. There are a pair of salt bridges between the side chains of Asp212 and Lys197' of the other monomer. Otherwise, the interactions are mainly hydrophobic. Residues Leu199, Val202, Tyr208, and Ile210 are buried and form the hydrophobic core of the interface. Structure-based sequence alignment indicated that these hydrophobic residues are not preserved in YedU, eliminating this potential interface (Figure 4-9).

```
EcYedU: 231|GEELKKMGMNIIN--DDITGRVHKDRKLLTGDSPFAANALGK|270
YDR533: 188|EDVAKKYGAKYLAPVGPWDDYSITDGRLVTGVNPASAHSTAV|229
```

Figure 4-9 Alignment of YedU and YDR533Cp

Residues 231-270 of YedU are aligned with their structurally equivalent residues in YDR533Cp by the ProSup program (Lackner et al., 2000). The residues Leu199, Val202, Tyr208, and Ile210 that form the hydrophobic core of the YDR533Cp dimer interface are underlined.

The PH1704 protease forms hexamers that can be considered as a trimer of dimers, resulting in two types of subunit interactions (Du et al., 2000). Contact between monomers B and C makes the dimer, and contact between monomer A and B makes the

trimer of dimers by a threefold non-crystallographic symmetry (NCS) operation (Figure 4-8C). Both contacts are absent from YedU because of the helix H3, which prevents the BC interaction, and the beta-hairpin S4-S5 and the helix H9, which prevents the AB interaction.

The human DJ-1 protein exists as a dimer that is very robust as dimerization buries 2,615 $Å^2$ of surface areas (Honbou et al., 2003; Huai et al., 2003; Lee et al., 2003b; Tao and Tong, 2003; and Wilson et al., 2003). The crystal structure indicates that helices H2 of both monomers form the core of the dimer. At one edge of the interface, the beta hairpin S6'-S6'' and its symmetry related equivalent form an intersubunit four-stranded anti-parallel beta-sheet (Figure 4-8D). At the other edge, the interactions are formed between the C-terminal helix H12 that does not have a structural equivalent in YedU.

4.4 A METAL-BINDING SITE

4.4.1 Description of the Zinc Coordination

A zinc(II) ion is coordinated by the side chains of His85, Glu90, and His122 (Figure 4-10). His85 is located at the end of the H2 helix and Glu90 is contributed by the loop following the H2 helix. His122 is from a short loop between the H3 and H4 helices. In the crystal, the zinc(II) ion is tetrahedrally coordinated by His85 NE2, His122 NE2, both Glu90 OE1, and NE2 of a symmetry related His85'. His122 has the potential to disperse charge through hydrogen-bonding of its ND1 with the carbonyl oxygen of Phe118. The involvement of a symmetry related His85' in coordination of the zinc(II) ion is due to crystal packing, while the metal-binding site should be fully exposed on the surface of the YedU homodimer in solution. It is likely that a water molecule replaces this symmetry related His85' in coordinating the zinc(II) ion in solution.

The metal ion present in the YedU crystal has been documented to be a zinc(II) ion using X-ray fluorescence. No zinc ions were added during purification and

crystallization, indicating that the zinc(II) ion could be the biologically relevant metal ion. The zinc-binding site is on the protein surface, suggesting that it is not a structural zinc ion. In the presence of 1 mM of the chelating agent EDTA, the YedU protein still migrates as a dimer on the Superdex75 size exclusion chromatographic column. This further proved that the zinc ion is not critical for protein folding or dimerization. Most likely, it is a catalytic zinc ion underlying a yet-to-be-discovered function of the YedU protein. It is also possible that iron is the biologically relevant metal ion for the YedU protein but zinc was incorporated upon iron oxidation and depletion during overexpression and purification.

Sequence alignment indicates that zinc-binding residues His85, His122, and Glu90 are not conserved (Figure 4-4), indicating evolutionary divergence among class I YedU orthologs.

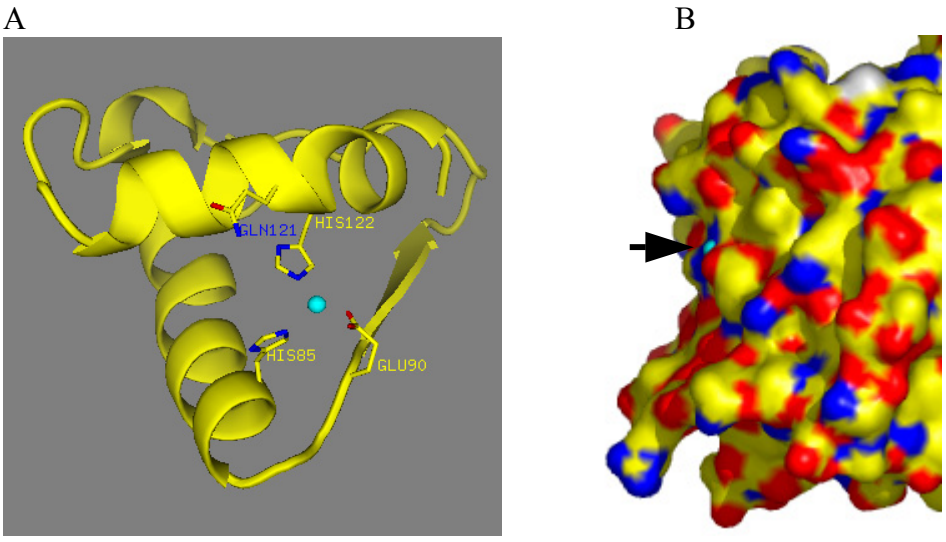

Figure 4-10 The metal-binding site of YedU

(*A*) A zinc(II) ion is coordinated by the side chains of His85, Glu90, and His122. (*B*) The metal binding site is on the surface of the YedU protein. The arrow points to the zinc ion (cyan).

4.4.2 Enzymes with a Catalytic Zinc(II) Ion

Zinc(II) ion has been found in the catalytic sites of many enzymes including, but not limited to, the following families of alcohol dehydrogenase, carboxypeptidase, astacin family of zinc proteases, beta-lactamase, and carbonic anhydrase (Auld, 2001). In these enzymes, a catalytic zinc(II) ion generally forms complexes with three nitrogen, oxygen and sulfur donors of histidine, glutamate/aspartate, and cysteine residues. In addition, water is always a ligand to the catalytic zinc.

The dimeric alcohol dehydrogenases (ADH) are nicotinamide adenine dinucleotide (NAD) dependent and catalyze the reversible oxidation of alcohols to aldehydes. Each subunit of the dimeric enzyme has a coenzyme binding domain and a catalytic domain that are separated by a cleft. ADHs contain both a catalytic and a structural zinc site that reside in the catalytic domain. The catalytic zinc is coordinated by two cysteine residues, a histidine residue, and a water molecule in a tetrahedral geometry (Xie et al., 1997).

The carboxypeptidase family of exopeptidases catalyze the degradation of short peptides. These enzymes complement the actions of endopeptidases and lead to the formation of amino acids. The catalytic zinc of carboxypeptidases A, B, D, and M is coordinated by ND1 of two histidine residues, both OE1 and OE2 of a glutamate residue, and a water molecule in a pentahedral geometry.

The astacin family of metalloendopeptidases has a signature sequence motif of **H**Exx**H**xxGxx**H** and they are involved in activation of growth factors, degradation of polypeptides, and processing of extracellular proteins (Bond and Beynon, 1995). Astacin has a compact bilobar structure with a deep active-site cleft that divides it into two parts. The N-terminal 100 amino acids are folded as a five-stranded beta-sheet and two long alpha-helices. One of the helices contains the **H**Exx**H**xxGxx**H** signature motif that includes three histidines as zinc ligands. An additional tyrosine and a water molecule

that also interacts with a glutamate are the fourth and fifth ligands, finishing the pentahedral, or trigonal bipyramid, coordination of the catalytic zinc(II) ion.

Although the majority of beta-lactamases utilize a serine in the hydrolysis of the beta-lactam ring, there are pathogenic bacteria that have a metallo-beta-lactamase that contains a catalytic zinc (Daiyasu et al., 2001). The beta-lactamase from *Bacillus cereus* is folded into a beta/beta sandwich with helices on each external face (Carfi et al., 1995). The active site is located at one edge of the beta/beta sandwich and near the N-terminal end of a helix. The catalytic zinc(II) ion is coordinated by three histidine residues and a water molecule in a tetrahedral arrangement.

Carbonic anhydrases catalyze the reversible hydration of CO_2. There are three distinct classes (designated alpha, beta and gamma) of carbonic anhydrases that have no significant sequence identity and have evolved independently. The catalytic zinc of alpha-class carbonic anhydrases is tetrahedrally coordinated to the NE2 of two histidines from a beta-strand, ND1 of a histidine from another beta-strand, and a water molecule. The beta-class carbonic anhydrases has a different catalytic zinc coordination site that includes two cysteines, a histidine, and a water molecule. The gamma-class carbonic anhydrases retains three histidine ligands as in the alpha-class carbonic anhydrases.

Compared with these zinc enzymes, YedU has a unique fold that is dissimilar to the folds described above. Although the zinc coordination of YedU is most similar to that of carboxypeptidases by two histidines and a glutamate, there are substantial differences. Instead of ND1, YedU utilizes its NE2 of histidine in coordinating the zinc(II) ion (Figure 4-10). In addition, the zinc site of YedU is on the surface of the protein, while that of carboxypeptidases is located at the bottom of a deep cleft and surrounded by a second coordination shell.

4.4.3 Iron(II) Enzymes with a 2-His-1-Carboxylate Triad

It is also possible that iron is the biologically relevant metal ion for the YedU protein but zinc was incorporated upon iron depletion and oxidation during overexpression and purification. Similar substitutions have been found for several iron enzymes including rubredoxin, desulforedoxin, and peptide deformylase (Eidsness et al., 1992; Czaja et al., 1995; Rajagopalan et al., 1997). Furthermore, the metal coordination found in YedU is very similar to that of many non-heme iron(II) enzymes containing 2-His-1-carboxylate motif (Que, 2000). The 2-His-1-carboxylate motif serves as an excellent monoanionic three-pronged platform for binding divalent ions. The three remaining sites on the opposite face of the octahedron are made available for exogenous molecular oxygen, cofactor, and substrate. The coupling between the reduction of molecular oxygen and oxidation of substrate is strongly promoted by poising the metal site to bind molecular oxygen only when substrate and cofactor are present (Que, 2000). The 2-His-1-carboxylate motif has been found in four classes of non-heme iron(II) enzymes including pterin-dependent hydroxylases, extradiol-cleaving catechol dioxygenases, rieske dioxygenases, and α-keto acid-dependent oxygenases.

The pterin-dependent hydroxylases are a small family of closely related enzymes that catalyze aromatic hydroxylation of phenylalanine, tyrosine, and tryptophan by using tetrahydrobiopterin (BH_4) as a two-electron-donating cofactor (Erlandsen and Stevens, 1999). Their signature sequences are $\mathbf{H}X_4\mathbf{H}X_{40}\mathbf{E}$. The pterin-dependent hydroxylases exist in solution as homotetramers and each subunit has an N-terminal regulatory domain, a catalytic domain, and a C-terminal tetramerization domain. Within the catalytic domain, 14 alpha-helices and 8 beta-strands fold in a basket-like arrangement. The cofactor and substrate sites are close to the active-site iron, which resides at the bottom of a deep cleft (Figure 4-11A).

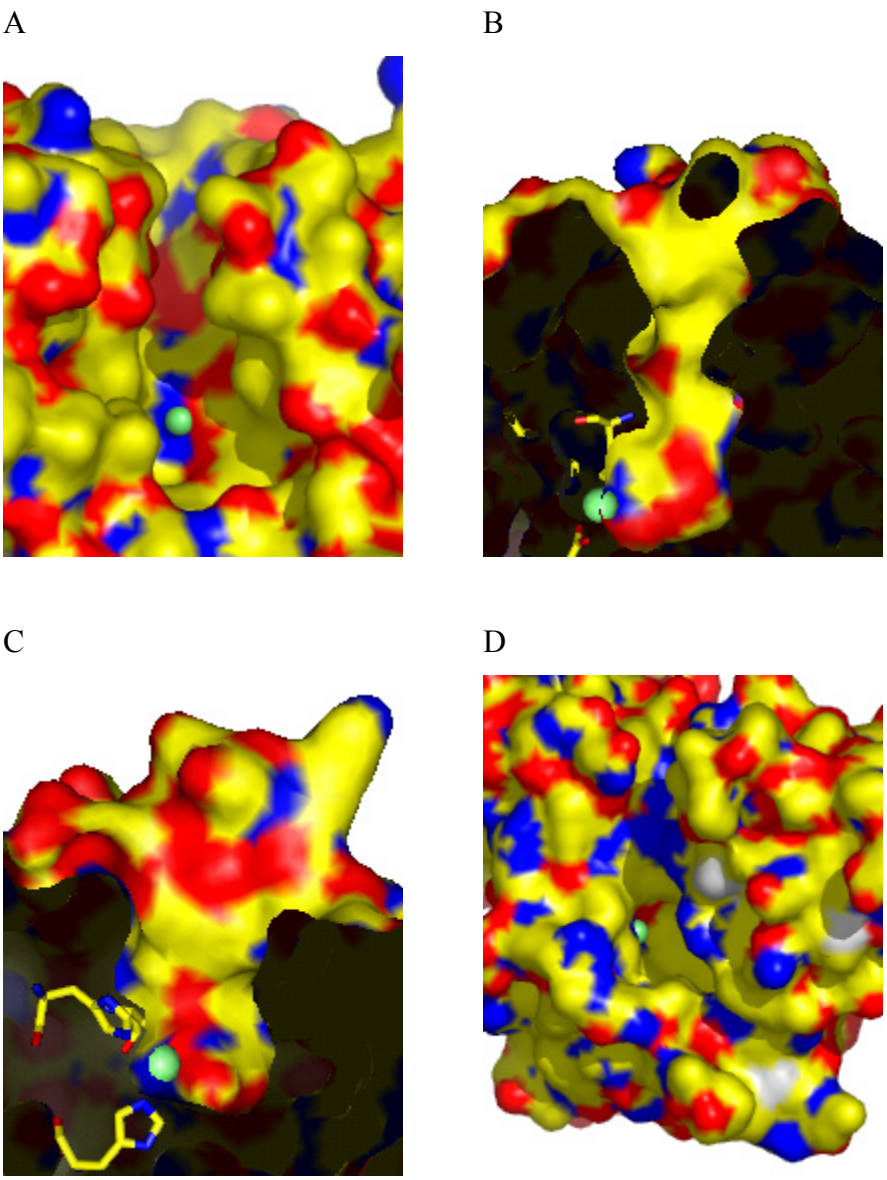

Figure 4-11 The metal-binding sites of non-heme iron(II) enzymes

The iron is shown as a lime sphere in (*A*) Phenylalanine hydroxylase (PDB accession 1PAH), (*B*) Napthalene 1,2-dioxygenase (PDB accession 1NDO), (*C*) 2,3-Dihydroxybiphenyl 1,2-dioxygenase (PDB accession 1KW3), and (*D*) Deacetoxycephalosporin C synthase (PDB accession 1RXF). Carbon, oxygen, and nitrogen are colored yellow, red and blue, respectively.

Rieske dioxygenases catalyze the NADH-mediated *cis*-dihydroxylation of an arene double bond, the first step in the biodegradation of aromatic molecules by soil bacteria. Rieske dioxygenases have a sequence motif of **H**X_4**H**X_{148}**D** that contains iron ligands. Naphthalene dioxygenase (NDO) is a $\alpha_3\beta_3$ hexamer (Kauppi et al., 1998). The catalytic domain resides in the alpha-subunit. The catalytic domain is dominated by a nine-stranded antiparallel beta-sheet. Three of the insertions between the strands contain the iron ligands His208, His213, and Asp262. A second coordination shell contains Asp205 and Asp361 that are hydrogen-bonded to His208 and His213, respectively. Electrons from the NADH cofactor that does not bind directly to the iron are transferred to NDO via an iron-sulfur flavoprotein and a Rieske ferredoxin. The catalytic iron(II) ion and the substrate binding site are at the bottom of a deep cavity (Figure 4-11B).

Following the formation of catechols, they are degraded by intradiol-cleaving and extradiol-cleaving catechol dioxygenases for the biodegradation of aromatic molecules in the soil (Han et al., 1995). While the intradiol-cleaving enzymes utilize an iron(III) ion, the extradiol-cleaving enzymes typically use an iron(II) ion to activate molecular oxygen. At the sequence level, the extradiol-cleaving catechol dioxygenases have an **H**X_{n1}**H**X_{n2}**E** motif that contains the iron ligands separated by two long linkers X_{n1} and X_{n2}. In its tertiary structure, 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD) from *Pseudomonas cepacia* has four superimposable copies of a $\beta\alpha\beta\beta\beta$ module. The spatial order and polarity of the strands are $\beta_1^+\beta_4^+\beta_3^-\beta_2^+$. The metal site is positioned between strands β_1 and β_4 of two $\beta\alpha\beta\beta\beta$ modules in edge-to-edge contact. The substrate binding site is within a deep cavity in which the iron resides (Figure 4-11C).

Oxidative reactions catalyzed by alpha-keto acid-dependent oxygenases are steps in the biosynthesis of many metabolites such as plant hormones and antibiotics (Schofield and Zhang, 1999). The oxidation of substrate is coupled to the conversion of the cofactor keto acid into the carboxylate derivative of the keto acid and CO_2. In the case of 2-oxoglutarate (2OG), the most frequently used cofactor, it is converted to succinate and CO_2. During the reaction, one of the oxygen of the dioxygen molecule is incorporated

into the substrate and the other goes to succinate. So alpha-keto acid-dependent oxygenases can be classified as dioxygenases. A **H**X(**D/E**)X_n**H** motif is found in all sequences of these enzymes. The alpha-keto acid-dependent oxygenases have a beta-strand core folded into a distorted jelly-roll structure. Iron-binding residues are located on strands forming beta-barrel core, at the bottom of a long cleft (Figure 4-11D).

Although conserved sequence motifs can be found within each class of enzymes, there is no sequence homology among the various classes. Each class of enzymes adopts a particular fold that is maintained within the class but distinct from those of other classes. It appears that nature has evolved a common structural motif to bind iron(II) out of a number of divergent protein sequences and folds. The YedU protein shows no homology with any of the families described above. The fold of YedU is also different from these known non-heme iron(II) enzymes. The 2-His-1-carboxylate motif of YedU is located on a relatively flat surface (Figure 4-10B), compared with the deep clefts (or cavities) seen in these non-heme iron(II) enzymes (Figure 4-11).

4.5 A PUTATIVE CATALYTIC TRIAD

4.5.1 Alpha/Beta Hydrolases

The YedU protein has a structural resemblance with the alpha/beta hydrolase superfamily characterized by a strand-turn-helix nucleophile elbow and an alpha/beta/alpha sandwich fold (Nardini and Dijkstra, 1999; Holmquist, 2000). The sandwich fold generally consists of a central five to eight stranded beta-sheet core flanked on both sides by alpha-helices (Figure 4-12). The central beta-sheet displays a left-handed superhelical twist, with the first and last strands crossing each other at an angle of about 90°. The degree of twisting of the central beta-sheet can show significant differences among members of the superfamily. The twisted sheet forms a half-barrel.

The first and last helices pack onto the convex surface of the half barrel, whereas other helices pack onto the concave surface. Although the numbers and spatial positions of the alpha-helices packed on the concave surface of the half-barrel show significant difference, the helix of the nucleophile elbow is well conserved and always positioned in the center of the concave surface.

The alpha/beta hydrolase fold provides a stable scaffold for the active sites of a wide variety of enzymes. The active-site residues always form a catalytic triad: a Ser, Cys, or Asp as the nucleophile, a His, and an acidic residue that are hydrogen-bonded with each other. The nucleophile residue forms hydrogen bond with the His through it ND1 atom, while the NE2 of His is hydrogen-bonded with the acidic residue Glu or Asp. The nucleophile Ser, Cys, or Asp is always located at the sharp turn of the strand-turn-helix nucleophile elbow, where it can be easily approached by the substrate. The nucleophile elbow is also characterized by the consensus sequence small-x-nucleophile-x-small-small. The nucleophile residue is forced to adopt energetically unfavorable main chain torsion angles by the tightness of this strand-turn-helix motif (Hakansson, 2002).

The geometry of the nucleophile elbow also contributes to the formation of an "oxyanion hole" that is required to stabilize the negatively charged transition-state intermediate during hydrolysis. The "oxyanion hole" is usually formed by two backbone nitrogen atoms: the first is always from the residue immediately following the nucleophile, whereas the second is generally located between an adjacent strand and helix. Although the His is the only residue of the catalytic triad that is absolutely conserved, the position, shape, and length of the His-containing loop can differ significantly. The position of the acidic residue also shows considerable difference among members of the superfamily.

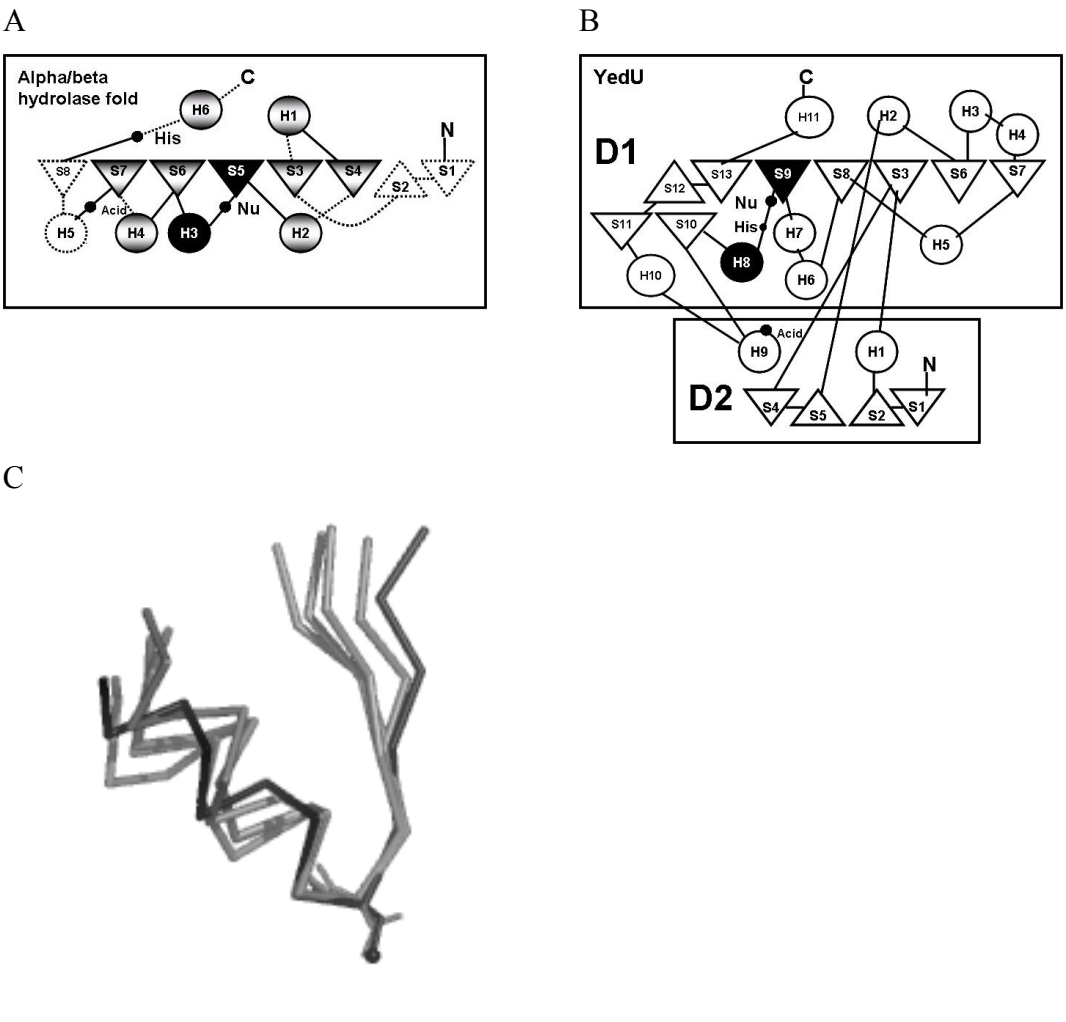

Figure 4-12 Comparison of the alpha/beta hydrolase fold and YedU

(*A*) The alpha/beta hydrolase fold. Triangles represent beta-strands. Circles represent alpha-helices. The solid and shadow triangles and circles represent minimal alpha/beta hydrolase fold. Dashed lines, triangles, and circles are possible extensions and insertions. The location of catalytic triad is indicated by black dots. (*B*) The YedU protein. The nucleophile elbow-like motif is represented by the solid triangle and circle. (*C*) Superposition of the nucleophile elbow motifs of YedU (black) and alpha/beta hydrolases (gray). The alpha/beta hydrolases are Cephalosporin C deacetylase (PDB accession 1ODS), bacterial lipase (PDB accession 1JI3), bacterial cocaine esterase (PDB accession 1JU3), and human dipeptidyl peptidase IV (DPP-IV) (PDB accession 1NU6).

While maintaining a conserved catalytic machinery, the alpha/beta hydrolase fold tolerates insertions, ranging from a few amino acid residues to an extra domain. This gives members of the alpha/beta hydrolase superfamily a striking ability for adaptation and evolution. After the alpha/beta hydrolase fold was first described a decade ago (Ollis et al., 1992), the number of identified members of the superfamily has grown rapidly. The members diverged from a common ancestor into a number of enzymes with a wide range of hydrolytic functions, together with other proteins with no recognized catalytic activity. The alpha/beta hydrolase superfamily is a typical example of evolutionary differentiation within a common fold. Coverage of the active site is the major structural variation has been seen within the alpha/beta hydrolase superfamily. In order to accommodate different substrates, loops are inserted in the center of the molecule, usually at the C-terminal of the central strands. These insertions shape the substrate-binding site that is near the catalytic triad. They may be a few residues or large enough to form an extra domain that either seals off the catalytic triad, except for a narrow tunnel, or acts as a movable lid that regulates the entry of substrate and release of product.

4.5.2 A Putative Cys184-His185-Asp213 Catalytic Triad

Residues 179-193 (S9 and H8) of YedU form a nucleophile elbow-like motif that is very similar to those of alpha/beta hydrolases (Figure 4-12). A potential nucleophile Cys184 is located at the sharp turn between the S9 strand and the H8 helix. The phi/psi (ϕ/ψ) angles (57°/-115°) of Cys184 fall in the disallowed region of the Ramachandran plot, this is characteristic of the nucleophile in a nucleophile elbow (Hakansson, 2002). Residues Cys184, His185, and Asp213 of YedU form a putative catalytic triad that is reminiscent of those of papain-type cysteine proteases. The Cys184 is hydrogen-bonded with the ND1 atom of His185. The NE2 atom of His185 is hydrogen-bonded with the Asp213. The triad is surrounded by polar residues that form an extensive hydrogen-bond network (Figure 4-13). These residues are conserved among all class I YedU orthologs.

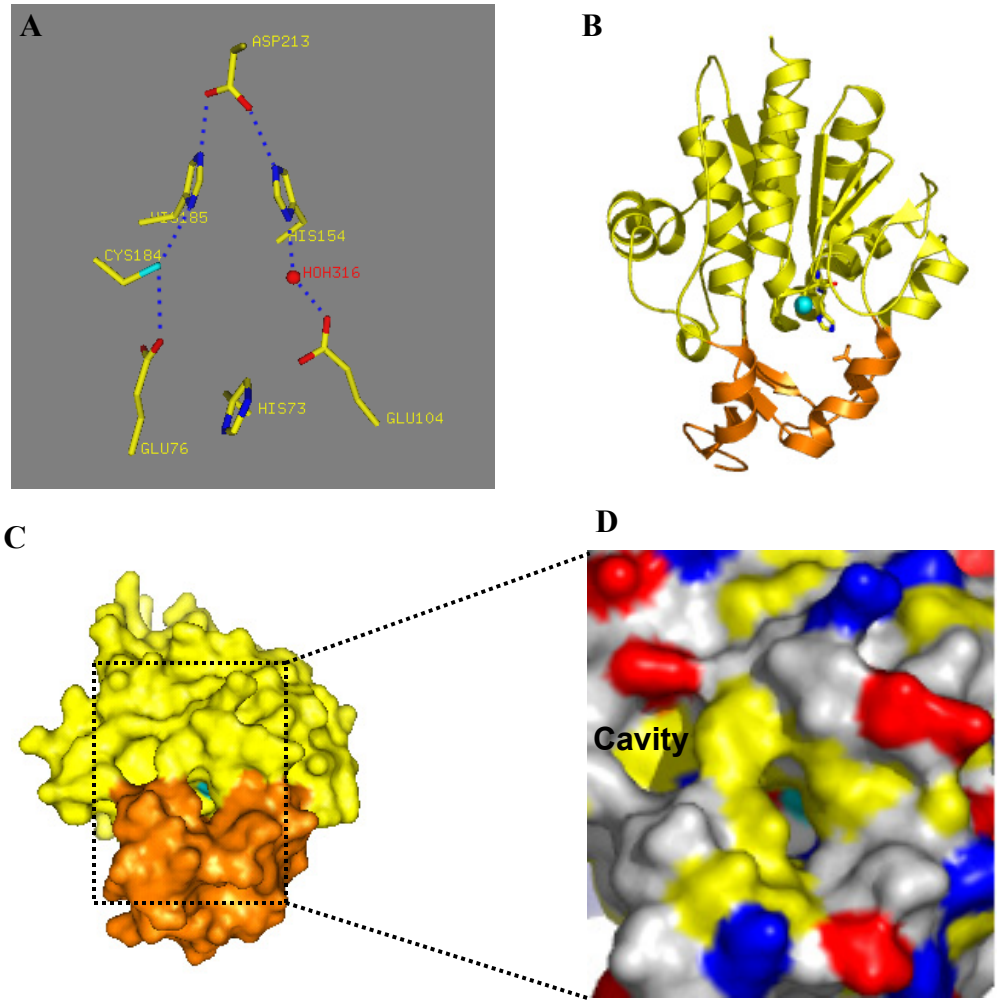

Figure 4-13 The putative Cys184-His185-Asp213 catalytic triad

(*A*) Cys184, His185, and Asp213 form a putative catalytic triad. The Sg atom of Cys184 is colored cyan. The triad is surrounded by polar residues Glu76, Glu104, His154, and a water molecule that form an extensive hydrogen-bond network. The hydrogen-bonds are shown as dotted lines. (*B*) The triad is located at the interface of the D1 (yellow) and D2 (orange) domains. (*C*) A narrow tunnel is formed at the interface of the D1 (yellow) and D2 (orange) domains. The Sg atom of Cys184 can be seen at the bottom of the tunnel. (*D*) A zoom in view of the tunnel. There is a cavity near the tunnel. The side chains of hydrophobic residues (Pro, Ala, Val, Leu, Ile, Met, Phe, Tyr, and Trp) are colored yellow; the side chains of positively charged residues (His, Lys, and Arg) are blue; the side chains of negatively charged residues (Asp and Glu) are red; other residues and all backbone atoms are white.

The backbone nitrogen atoms from Gly153 and His185 are located 4.2 Å and 3.3 Å, respectively, from the Sg atom of Cys184, ideally positioned to form an "oxyanion hole", which is present in many alpha/beta hydrolases for stabilization of tetrahedral intermediates during catalysis (Nardini and Dijkstra, 1999; Holmquist, 2000). Gly153 is located between the S8 strand and the H6 helix adjacent to the nucleophile elbow. It is conserved among all class I YedU orthologs.

The His185 immediately follows the potential nucleophile Cys184, while Aps213 is contributed by the H9 helix of the D2 domain. The triad of YedU is located at the interface of the D1 and D2 domains (Figure 4-13). A narrow tunnel between the D1 and D2 domains connects the triad with the bulk solvent region. The tunnel is lined with hydrophobic residues Phe208, Ile246, Pro262, and Phe263 of the D1 domain on one side, and Tyr28, Ile219, Tyr221, and Thr216 of the D2 domain on the other side. The width and depth of the tunnel are about 6 Å and 11 Å, respectively, restricting the accessibility of bulky compounds or proteins to the triad. Residues Asp213, Thr216, Tyr221, and Met222 form a hydrophobic pocket that could be a substrate binding site at the bottom of the tunnel.

At the bottom of the tunnel, a narrow passageway connects the tunnel to a neighboring cavity which is about 8 Å wide and 7 Å deep (Figure 4-13). The tunnel and the cavity are separated by residues His73, Pro262, Phe263, and Tyr28, where the narrow connecting passageway is less than 4 Å wide between His73 and Pro262. Although the interior of the cavity is predominately hydrophobic, its opening is amphipathic and lined by the backbones of Thr29 - Ser33 and side chains of Ser30 and Asp34.

4.5.3 The Catalytic Triads of YDR533Cp and Protease PH1704

Putative catalytic triads have also been found in the yeast YDR533Cp protein (Cys138-His139-Glu170) and the *Pyrococcus horikoshii* PH1704 protease (Cys100-His101-Glu474') that belong to class II orthologs and class IIIA YedU orthologs,

respectively. Furthermore, all three proteins have a conserved sequence motif CHG located at the turn of the nucleophile elbow. The strand-turn-helix nucleophile elbow of YedU aligns very well with those of YDR533Cp and PH1704. When the C-alpha atoms of five residues (the conserved CHG motif with a flanking residue at both ends) are superposed, the rms deviation values are 0.28 Å (between YedU and YDR533Cp) and 0.12 Å (between YedU and PH1704). In fact, YedU, YDR533Cp, and PH1704 monomers can be aligned by superposing these five residues.

The putative catalytic triad of YedU is most similar to that of YDR533Cp. Overlap of two triads yields rms deviation values on the three C-alpha atoms of 0.24 Å (Figure 4-14). The acidic residue Glu170 of YDR533Cp is contributed by an alpha-helix in the D2 domain, very similar to the sequential and spatial arrangement seen in YedU. The hydrogen-bonding network is also present in YDR533Cp, formed between Glu170, His108, Tyr18 atom, and Asp57 (Figure 4-14). Furthermore, YDR533Cp's triad is also located at the bottom of a cleft that restricts the accessibility of the triad (Wilson et al., 2004). The function of the YDR533Cp's putative catalytic triad is still unclear.

PH1704 has been known as a cysteine endopeptidase that can hydrolyze bulky substrates such as azocasein and gelatin (Du et al., 2000). The geometry of YedU's triad is similar to that of PH1704 with a rms deviation value on the three C-alpha atoms of 0.64 Å. Both have an additional Glu that makes a hydrogen bond with the Sg atom of the nucleophile Cys (Figure 4-14). However, there are substantial differences. Unlike YedU, PH1704 relies on oligomerization to complete the catalytic triad since Glu474' is contributed by an alpha-helix in an adjacent subunit. The hydrogen-bond network of PH1704 is less extensive than that of YedU (Figure 4-14). Most importantly, PH1704's catalytic triad is more accessible (Figure 4-14), allowing it to act on bulky substrates such as azocasein and gelatin (Du et al., 2000).

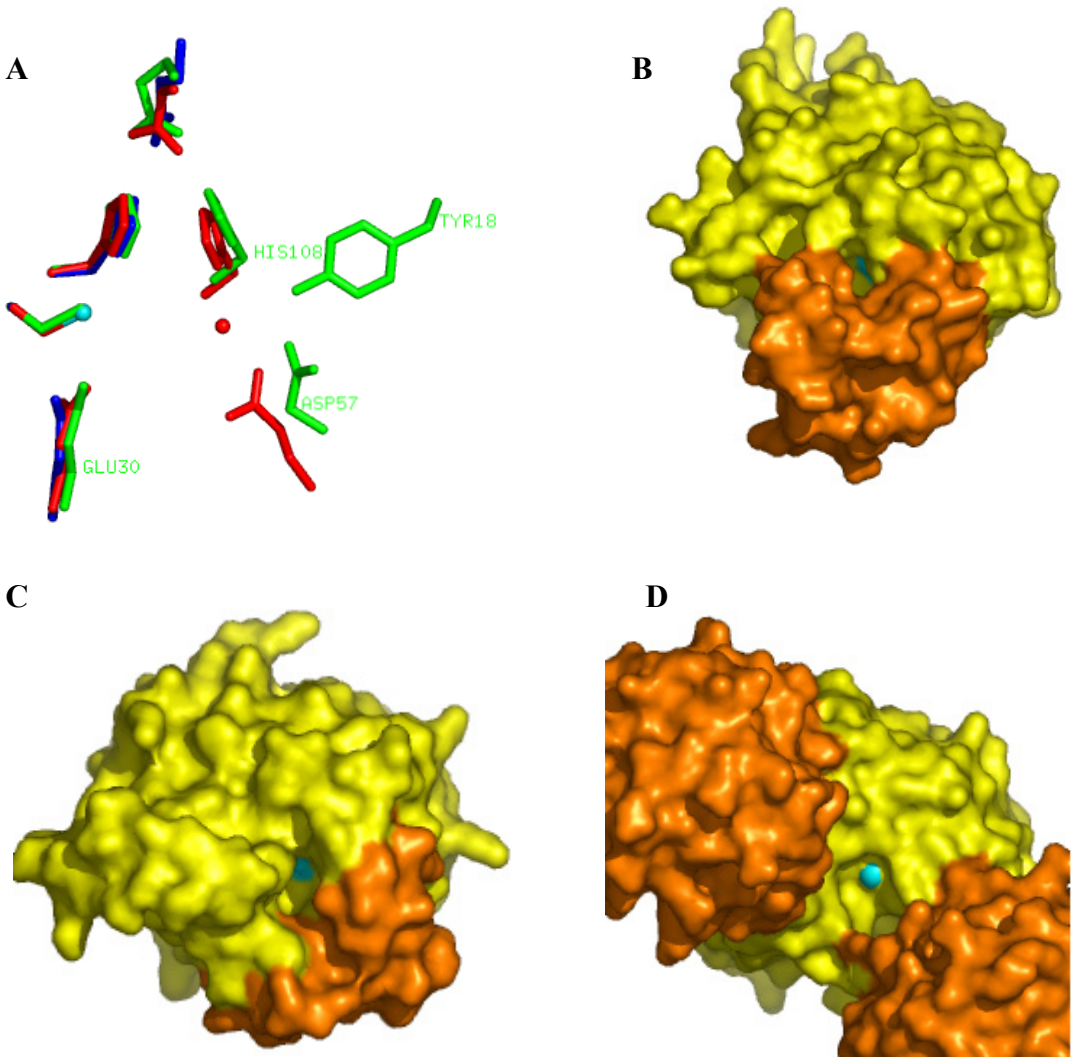

Figure 4-14 The catalytic triads of YedU, YDR533Cp, and PH1704

(*A*) Superposition of putative catalytic triads of YedU, YDR533Cp, and PH1704. YedU is colored red and the Sg atom of Cys184 is shown as a cyan sphere. YDR533Cp and PH1704 are colored green and blue, respectively. (*B*) The potential nucleophile Cys184 (cyan) of YedU is located at the bottom of the tunnel between the D1 (yellow) and D2 (orange) domains. (*C*) The potential nucleophile Cys138 (cyan) of YDR533Cp is located at the bottom of a cleft. The D1 and D2 domains are colored yellow and orange, respectively. (*D*) The putative nucleophile Cys100 (cyan) of the PH1704 protease is located between two monomers that are colored yellow and orange, respectively.

4.5.4 Functional Investigation of the Cys-His-Asp Triad of YedU

We have performed biochemical experiment to test the proteolytic activity of YedU. It did not display any detectable protease activity on the bulky substrate azocasein (Zhao et al., 2003). Another group also reported similar results (Malki et al., 2003).

The dimension of the tunnel connecting YedU's triad with the solvent suggests that it would be difficult for a bulky substrate to reach the potential nucleophile Cys184. It explains why YedU does not have detectable protease activity. However, the end of a peptide may be able to access Cys184 through the tunnel, whose width is about 6 Å. YedU was measured for exopeptidase activity toward peptides Dps (STALKC, amino acid sequence in single letter abbreviations) and Pep302 (GQGQQYGGY). These two peptides were used because of their immediate availability. YedU did not display any detectable exopeptidase activity in a range of temperatures, with or without ATP and $MgCl_2$ (Zhao et al., 2003). Negative results were also reported when exopeptidase activity of YedU was assayed using the synthetic substrates Suc-LLVY-amc, Boc-LAR-amc, Ac-YVAD-amc, AAF-amc, ALK-amc, Hippuryl-Phe, Hippuryl-Arg, Suc-AAA-*p*-nitroanilide and Pyr-HF(NO_2)FAL-NH_2 (Malki et al., 2003).

YedU showed exopeptidase activity against two synthetic substrates (Ala-amc and Gly-amc) with a small amino acid at N terminus (Lee et al., 2003b). Exopeptidase activity was not observed on the synthetic substrates Asp-amc, Arg-amc, Pro-amc, Leu-amc, Ser-amc, Val-amc, Tyr-amc, Thr-amc, Ile-amc, and Glu-amc. Although proteolytic activity of YedU toward bovine serum albumin (BSA) was demonstrated, it was extremely weak (Lee et al., 2003b). At 37°C, it required 12 hours to show some level of degradation when BSA was incubated with YedU at 3:1 molar ratio.

The hydrolytic activity of YedU was also measured by using 4-nitrophenyl phosphate as a substrate, a standard substrate used for phosphatase assays (Tabatabai and Bremner, 1969). The assay mixture contained 100 μg/ml of 4-nitrophenyl phosphate in 1 mM $MgCl_2$, 10 μM $ZnCl_2$, and 10 mM Tris pH 7.0. Reactions were carried out with

varying concentrations of YedU (1-5 μg/ml) in 1.0 ml of reaction volume. After 5-minutes of incubation, the reaction was stopped with 100 μl of 0.1 M NaOH and the absorbance was measured at 405nm. No phosphatase activity was detected for YedU at 25°C, 37°C or 45°C (Kaluarachchi and Fox, unpublished data).

CHAPTER 5: CHAPERONE ACTIVITY OF YEDU AND MUTANTS

This chapter describes the reported chaperone activity of the YedU protein, the molecular surface of the YedU homodimer, and the chaperone activity of subsite 1 mutants.

5.1 INTRODUCTION

Recently, it was reported that YedU has chaperone activity *in vitro* (Sastry et al., 2002; Malki et al., 2003). YedU can suppress the thermal aggregation of alcohol dehydrogenase (ADH) and citrate synthase (CS) at 45 °C, and promote their renaturation when the temperature drops to 23°C. YedU is also able to promote the renaturation of chemically denatured citrate synthase (CS) and α-glucosidase. The YedU concentrations (around 1 μM) required for half-maximal protein renaturation are similar to that of other molecular chaperones (Malki et al., 2003). Interestingly, ATP (2 mM concentration) inhibits the chaperone activity of YedU, although no ATPase activity is detected (Sastry et al., 2002).

In general, molecular chaperones recognize hydrophobic residues and unstructured backbone regions in their substrates, i.e., structural features exposed by nonnative proteins but normally buried upon protein folding. The substrate-binding sites of several well-studied chaperones have been identified through structural and mutagenesis studies. They can be grouped broadly into four classes: clamps, chambers, specialized surfaces, and structural complements (Stirling et al., 2003).

The substrate-binding site of DnaK resembles a clamp, which is formed by a beta-sheet cradle and a flexible alpha-helical lid. There is a deep cleft between loops L1,2 and L3,4 of the substrate-binding domain (Pellecchia et al., 2000). Structural studies revealed that a substrate peptide (sequence NRLLLTG) binds in this cleft. The middle Leu of the substrate peptide is completely buried in a deep hydrophobic pocket lined by residues Ile401, Thr403, Phe426, Val436, and Ile438. The backbone of substrate peptide forms hydrogen bonds with the chaperone. The long hydrophobic side chain of Met404 at the tip of loop L1,2 wraps over the substrate peptide to make contact with Ala429 from loop L3,4. This generates a hydrophobic arch that holds the bound peptide like a clamp.

Fourteen subunits of GroEL form two stacked rings that create chambers used to sequester non-native proteins. The apical domain of GroEL lines the opening of the chamber and provide multivalent binding sites for substrates (Farr et al., 2000). There is a groove formed by a pair of parallel helices H and I of the apical domain (Buckle et al., 1997). The groove is lined by the side chains of Ile230, Leu234, Leu237, Glu238, Ala241, Glu257, Thr261, Val264, Asn265, Ile270, and Val271, forming several hydrophobic pockets surrounded by a surface rich in polar and charged moieties. These residues were identified as essential residues in binding polypeptide substrates in several mutagenesis and structural studies (Fenton et al., 1994; Buckle et al., 1997; Chen and Sigler, 1999). The substrates bury their bulky hydrophobic side chains in the hydrophobic pockets and form a network of hydrogen bonds between their backbones and surrounding polar moieties of the apical domain.

Relatively flat or corrugated surfaces are also frequently used as substrate-binding sites. For example, SecB forms a dimer of dimers that has two hydrophobic grooves on opposite faces of the molecule (Xu et al., 2000; Dekker et al., 2003). Each groove can be roughly divided into two subsites. Subsite 1 is located at the middle of the groove and lined with conserved aromatic residues Phe39, Trp43, Phe132, and Pro133. Subsite 2 is located at the shallow end of the groove and lined with hydrophobic residues Leu47, Phe49, Leu51, Leu135, Leu137, Pro139, and Val140. The wheat HSP16.9, a small heat shock protein, forms a dodecamer consisting of two disks, each comprising six alpha-

crystallin-like domains organized in a trimer of dimers (van Montfort et al., 2001). Other studies suggested that temperature-dependent dissociation and association of sub-assembled species (dimers) regulate their exposed hydrophobic residues between substrate binding and higher-order assembly (Bova et al, 2000). Combining structural studies and biochemistry studies, two substrate-binding sites have been mapped on the surface of the dimer: a shallow groove and a flat patch. The shallow groove is located between strands β4 and β8 and lined by hydrophobic residues Val69, Val71, Val73, Val118, Val121, Ala123, Leu125, Leu130, and Val132. The flat patch is made up by aromatic residues Trp48 and Phe110. The shallow groove and the flat patch are on opposite sides of the α-crystallin-like domain, facing outside and inside, respectively, upon assembly of a dodecamer.

In exceptional cases, specialized chaperones provide a specific structural site to stabilize their substrate by structural complements. As an example, the *E. coli* periplasmic chaperone PapD specifically assists the folding of the Pap pilus subunits (Sauer et al., 2002). Each pilus subunit contains an immunoglobulin domain that lacks a beta strand, which is provided by an adjacent subunit in the assembled pilus structure. Before assembly, each pilus subunit is stabilized by a complementary beta-strand from PapD.

5.2 MOLECULAR SURFACE OF THE YEDU HOMODIMER

The molecular surface of the YedU homodimer exhibits a number of solvent-exposed hydrophobic patches. The largest one is an elongated, S-shaped hydrophobic surface that runs across the molecular two-fold axis (Figure 5-1). When the YedU homodimer is viewed down the molecular two-fold axis, four symmetry related protrusions, two from each monomer, can be seen. In each monomer, the large side chains of His110 and Lys111 extend out and form a positively-charged protrusion. The

side chain of Glu22 is located at the tip of a negatively-charged protrusion formed by N-terminal residues Thr4 and Val5 whose backbones are packed at the joint point of the H1 and H9 helices. The distances between the negatively-charged protrusion and the positively-charged protrusions are 17 Å and 29 Å, within each monomer and between monomers, respectively. Thus, a parallelogram can be defined by these protrusions. The lengths of the parallelogram's two sides are 17 Å and 29 Å, respectively, with an angle of 120° or 60° between each side. Between these protrusions, there is a groove whose depth is about 6 Å at the center on the molecular two-fold axis. The bottom of the groove is lined mainly with hydrophobic residues Val10, Ile12, Phe18, Phe19, Leu25, and Trp106. The hydrophobic surface formed by these residues and their symmetry equivalents will be referred as subsite 1. The groove is quite open, allowing the hydrophobic surface to extend to subsite 2 located at the base of the positively charged protrusion. Subsite 2 is lined with residues Tyr105, Met108, Met116, and Phe119. In each subunit, the hydrophobic surface continues to wrap around the protrusion on Pro117 and reaches subsite 3 that is lined with residues Leu35, Val38, Tyr82, Ala86, and Phe118. Overall, this creates an elongated, S-shaped hydrophobic surface that starts from subsite 3 of one subunit, wraps around the positively-charged protrusion and passes subsite 2, runs across the molecular two-fold axis as subsite 1, passes subsite 2', wraps round the positively-charged protrusion and reaches subsite 3' of the other subunit. The overall length of this hydrophobic surface is more than 70 Å and may provide binding sites for polypeptide of more than 20 amino-acid residues in extended conformation. The hydrophobic nature of subsites 1, 2, and 3 makes them ideal to bind hydrophobic side chains of a non-native polypeptide substrate, while these charged and polar residues of protrusion may interact with the backbone of the substrate polypeptide.

The opening of the triad tunnel is also a potential substrate-binding site (Figure 4-13D). The H1 and H9 helices, and a loop between the S11 and S12 strands form a basin that is located outside the central parallelogram, near the negatively charged protrusion on each subunit. The tunnel's opening is located at the bottom of the basin and is mainly hydrophobic because of residues Tyr28, Phe263, Ile246, Phe208, Ala212, Ile213, and Thr216. These hydrophobic residues are surrounded by Asp245, Asp244, Gln215,

Glu218, Lys6, Tyr23, and Gln27 that are mostly charged or polar. The H1 and H9 helices, and the loop between the S11 and S12 strands do not form a fully closed circle. Thus, the basin has a breach that faces the neighboring cavity, which is located outside the parallelogram, between the negatively charged protrusion and the positively charged protrusions on each subunit.

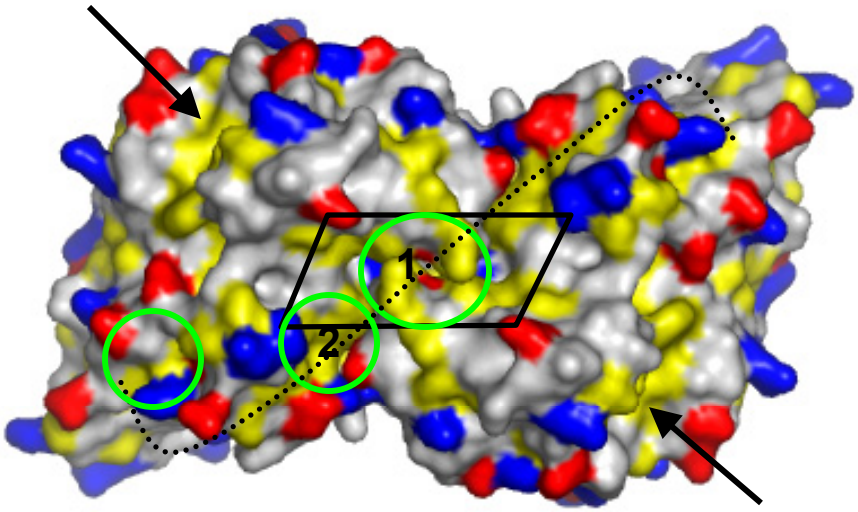

Figure 5-1 The molecular surface of the YedU homodimer

The view is looking down the molecular two-fold axis. The side chains of hydrophobic residues (Pro, Ala, Val, Leu, Ile, Met, Phe, Tyr, and Trp) are colored yellow; the side chains of positively charged residues (His, Lys, and Arg) are colored blue; the side chains of negatively charged residues (Asp and Glu) are colored red; other residues and all backbone atoms are colored white. The parallelogram defined by four protrusions is shown. Subsites 1, 2, and 3 are highlighted by green circles. A dotted line shows the S-shaped hydrophobic surface. Arrows point to the potential catalytic triad on both monomers.

5.3 CHAPERONE ACTIVITY OF SUBSITE 1 YEDU MUTANTS

As shown in the previous section, analysis of YedU's molecular surface highlighted several hydrophobic patches as candidates of potential substrate-binding site for its chaperone activity (Figure 5-1). Mutagenesis studies were performed to investigate the role of subsite 1 of the S-shaped hydrophobic surface in mediating substrate binding.

Four YedU mutants (V10D, I12D, F18E, and F19E) were made; each replaced a hydrophobic residue with a charged residue in subsite 1. Since these mutations are located on the surface of the protein, it was expected that they would not change the folding and dimerization of the YedU protein. Starting with the p192 plasmid (YedU-pET30), the point mutations were generated with QuikChange site-directed mutagenesis kit (Stratagene) by using primers that contain the desired mutations (Table 5-1). The mutants were expressed and purified following the same procedures used for wild type YedU protein (wtYedU). The mutants eluted at the same volume as wtYedU on the Superdex75 size-exclusion column, indicating that they also exist as dimers in solution.

Table 5-1 Primers for subsite 1 YedU mutants

Name	Sequence
YedU-V10D-5	5'-CAAGTAAAAATCCGCAGGACGATATTGCTGAAGATAATGC-3'
YedU-V10D-3	5'-GCATTATCTTCAGCAATATCGTCCTGCGGATTTTTACTTG-3'
YedU-I12D-5	5'-GTAAAAATCCGCAGGTCGATGATGCTGAAGATAATGCATTC-3'
YedU-I12D-3	5'-GAATGCATTATCTTCAGCATCATCGACCTGCGGATTTTTAC-3'
YedU-F18E-5	5'-GCTGAAGATAATGCAGAATTCCCTTCAGAATATTCG-3'
YedU-F18E-3	5'-CGAATATTCTGAAGGGAATTCTGCATTATCTTCAGC-3'
YedU-F19E-5	5'-GCTGAAGATAATGCATTCGAACCTTCAGAATATTCG-3'
YedU-F19E-3	5'-CGAATATTCTGAAGGTTCGAATGCATTATCTTCAGC-3'

Suppression of substrate thermal aggregation by wtYedU and subsite 1 mutants was assayed in right-angle light scattering experiments (Sastry et al., 2002).

Alcohol dehydrogenase (ADH) at 2.4 μM final concentration and YedU proteins at a series of concentrations were added to 150 mM Tris pH 7.4, prewarmed to 43°C. Samples (2.0 ml) were transferred to the thermostatic cell of a FluoroMax-2 spectrofluorometer (Spex Industries, Inc.) and held at 43 °C. Right-angle light scattering was recorded every 30 seconds with excitation and emission wavelengths set at 350 nm, while the excitation and emission slit widths were set at 2.5 nm and 5.0 nm, respectively. Consistent with the reported chaperone activity of YedU (Sastry et al., 2002), the wild type protein at 3-fold, 6-fold, and 9-fold molar excess (all calculated based on monomer concentrations) over ADH can suppress its thermal aggregation. The suppression of ADH thermal aggregation by YedU is modest at 3-fold molar excess, while it saturates at 6-fold molar excess. The experiment was repeated with subsite 1 mutants. At 3-fold molar excess over ADH, these single amino acid substitutions display only a modest change in the ability of YedU to inhibit ADH aggregation at 43°C (Figure 5-2A).

The suppression of thermal aggregation experiment was repeated with another substrate, citrate synthase (CS), at 0.11 μM final concentration in 40 mM HEPES pH 7.5. Similarly, none of these subsite 1 mutants exert significant effect on the aggregation-suppression activity of the YedU protein at 43°C (Figure 5-2B). These experiments suggest that single amino acid substitutions can not disrupt subsite 1, or the disruption of subsite 1 itself has no significant effect on YedU's chaperone activity at high temperatures.

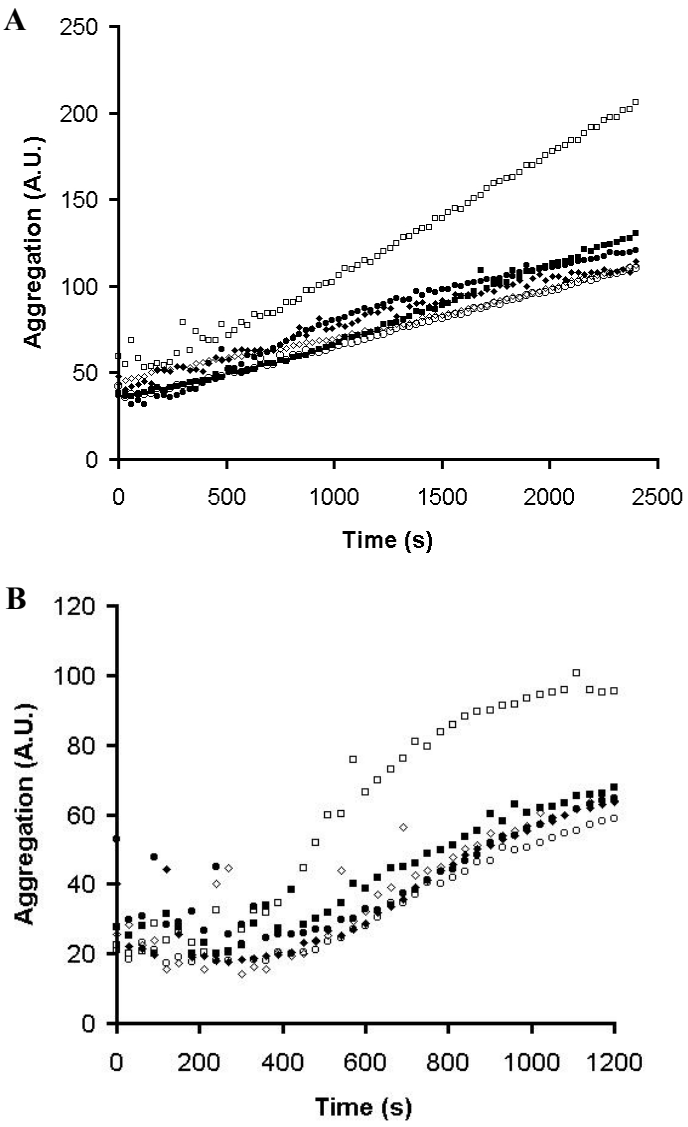

Figure 5-2 Chaperone activity of YedU mutants

Single amino-acid substitutions in subsite 1 do not have significant influence on the ability of YedU to suppress the heat-induced aggregation of (*A*) alcohol dehydrogenase (ADH) and (*B*) citrate synthase (CS). The aggregation of 2.4 μM ADH or 0.11 μM CS at 43°C was monitored without additive (open squares), or in the presence of wild type YedU protein (open circles), V10D (filled squares), I12D (filled circles), F18E (filled diamonds), or F19E (open diamonds) at 3:1 molar excess over the substrates. Results are reported in arbitrary units (A.U.).

5.4 THE LINKER-LOOP SHIELDED HYDROPHOBIC PATCHES

Bis-1-anilino-8-naphthalenesulfonate (bis-ANS) binding experiments suggest that high temperatures promote the exposure of additional structured hydrophobic surfaces in YedU (Sastry et al., 2002). Residues 29-47 of YedU exist in an extended conformation and form a long linker-loop that runs approximately in parallel with the H2 and H11 helices. In the triclinic (P1) crystal form reported recently, the linker-loop exhibits increased flexibility, and so does the positively-charged protrusion (Quigley et al., 2004). It was proposed that the mobility of the linker-loop and the positively- charged protrusion would lead to exposure of an additional hydrophobic patch on both subunits near the dimer interface (Figure 5-3). The exposure of this additional hydrophobic patch would link subsite 2 and 3 together, forming a large hydrophobic surface that may serve as a high-affinity substrate-binding site at high temperatures.

Consistent with our result shown in the previous section, a recent investigation also showed that subsite 1 residues did not play a dominant role in substrate binding at high temperatures (Sastry et al., 2004). They found that linker-loop-shielded hydrophobic residues are important for substrate interactions at high temperatures. Furthermore, the chaperone activity of YedU was abolished upon immobilization of the linker-loop by a disulfide bridge (Sastry et al., 2004). These experiments suggest that exposure of the proposed high-affinity substrate-binding site is critical for its chaperone activity at high temperatures.

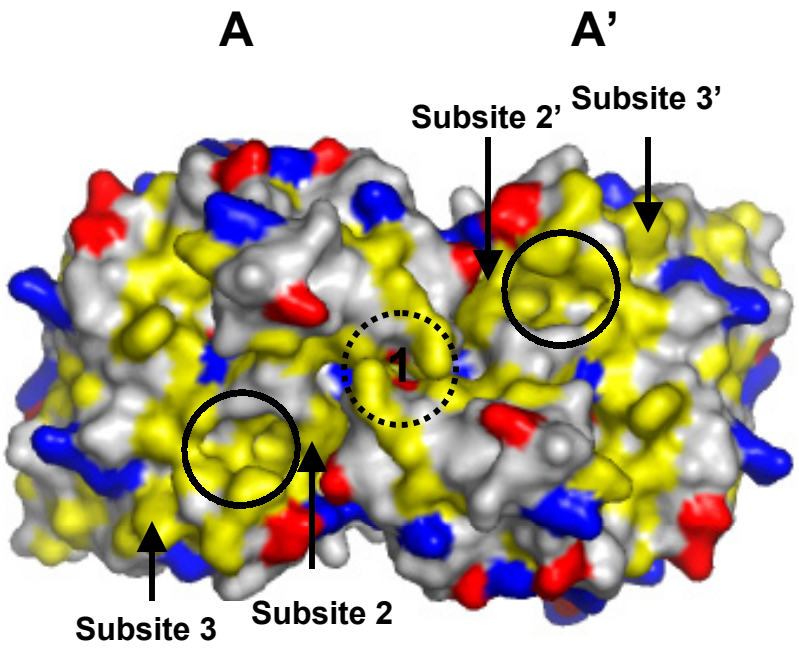

Figure 5-3 The linker-loop shielded hydrophobic patches

Removal of the linker-loop and the positively-charged protrusion results in the exposure of an additional hydrophobic patch (solid circle) on each YedU monomer. Subsite 1 is highlighted by a dotted circle. Arrows point to subsite 2 and 3 in monomer A, and subsite 2' and 3' in monomer A'. The side chains of hydrophobic residues (Pro, Ala, Val, Leu, Ile, Met, Phe, Tyr, and Trp) are colored yellow; the side chains of positively charged residues (His, Lys, and Arg) are colored blue; the side chains of negatively charged residues (Asp and Glu) are colored red; other residues and all backbone atoms are colored white.

CHAPTER 6: SUMMARY

The work reported in this dissertation describes the determination of the structure of the *E. coli* YedU stress protein in a multiple isomorphous replacement (MIR) experiment. The 2.2 Å crystal structure of YedU provides an accurate molecular model for further analysis of the mechanism underlying its reported chaperone activity (Sastry et al., 2002; Malki et al., 2003). In addition, the YedU crystal structure revealed two potential active sites, thus providing novel information on its possible functions.

The YedU monomer has an alpha/beta/alpha sandwich domain (D1) and a second smaller domain (D2) (Figure 6-1A). YedU is a dimer in solution and the dimer interface was identified. The molecular surface of the YedU homodimer exhibits a number of solvent-exposed hydrophobic patches (Figure 6-1B). To investigate the role of the subsite 1 in mediating substrate binding, single amino-acid substitutions were made, each replaced a hydrophobic residue with a charged residue. Compared with the wild type protein, the chaperone activity of these mutants was only slightly reduced. The data suggests that these residues alone do not play a dominant role in substrate binding at high temperatures.

A metal binding site was identified, where a zinc(II) ion is coordinated by a 2-His-1-carboxylate motif composed of His85, Glu90, and His122 (Figure 6-1C). The presence of zinc in the YedU crystals was confirmed by the X-ray fluorescence experiment. Each YedU monomer has a metal-binding site on the surface. There was no evidence that the zinc(II) ion plays a structural role in the folding or dimerization of the YedU protein, implying that it might be a catalytic (or regulatory) zinc ion. Since no zinc were added during purification and crystallization, the zinc(II) ion could be the biologically relevant metal ion. It is also possible that iron(II) ion is the biologically relevant metal ion for the YedU protein but zinc was incorporated upon iron depletion and oxidation during overexpression and purification. Further investigation of this

potential active site should consider both zinc(II) ion and iron(II) ion as potential relevant metal ions.

Each YedU monomer has a putative Cys184-His185-Asp213 catalytic triad surrounded by conserved residues that form an extensive hydrogen-bond network (Figure 6-1D). It can not be said with certainty whether the Cys184-His185-Asp213 triad is indeed catalytic, although its presence strongly suggests a possible hydrolase activity for the YedU protein. The triad of YedU is located at the bottom of a deep narrow tunnel between the D1 and D2 domains. The tunnel is lined with conserved hydrophobic residues, which might determine its substrate specificity.

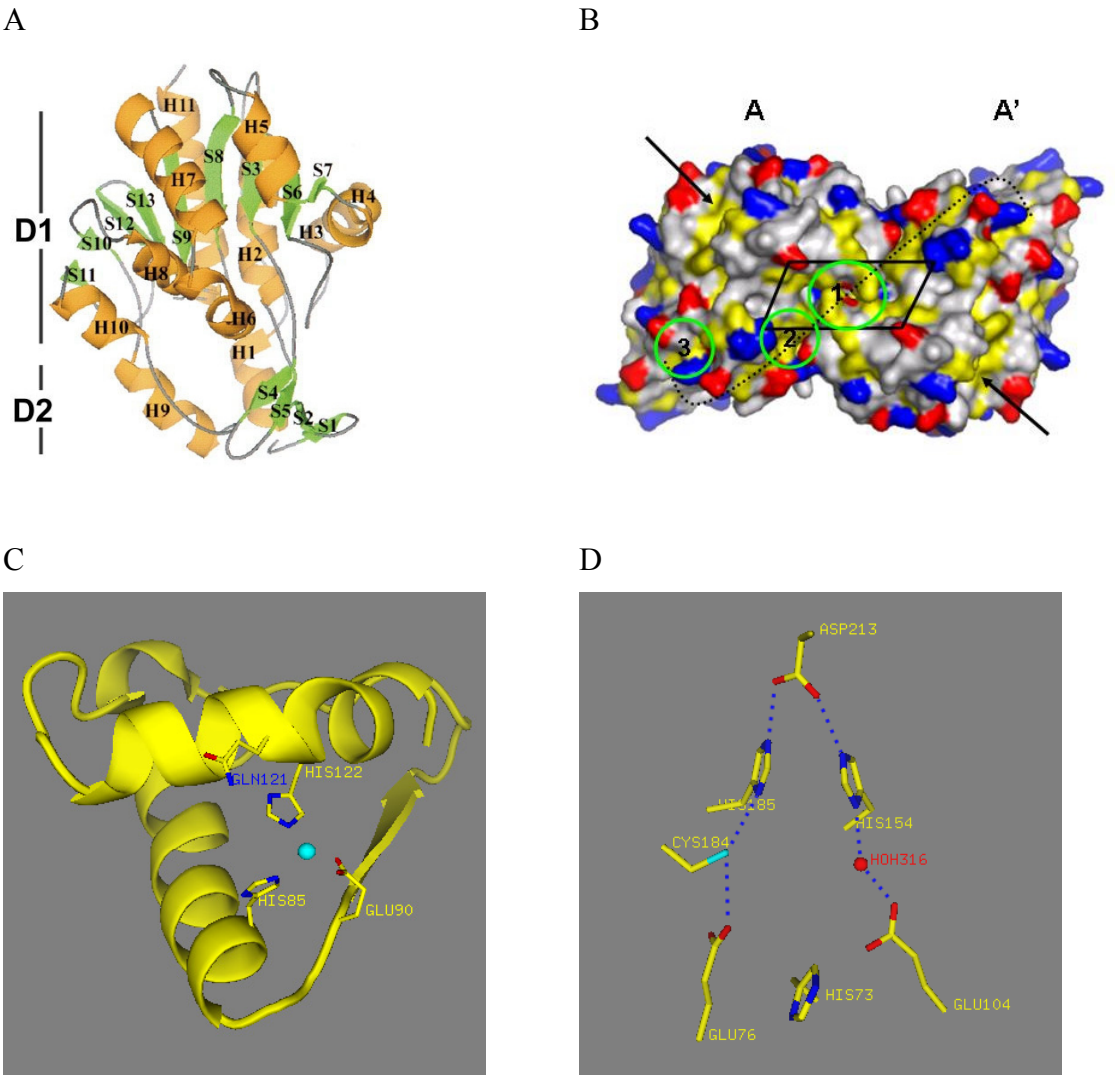

Figure 6-1 The YedU crystal structure

(*A*) The tertiary structure of the YedU monomer. (*B*) The molecular surface of the YedU homodimer. (*C*) The metal-binding site. (*D*) The putative catalytic triad.

REFERENCES

Atlung, T. & Ingmer, H. (1997). H-NS: A modulator of environmentally regulated gene expression. Mol. Microbiol. 24, 7-17.

Auld, D.S. (2001). Zinc coordination sphere in biochemical zinc sites. Biometals 14, 271-313.

Bandmann, O. (2004). DJ-1: the second gene for early onset Parkinson disease. Neurology 62, 357-358.

Bandyopadhyay, S. & Cookson, M.R. (2004). Evolutionary and functional relationships within the DJ1 superfamily. BMC Evol. Biol. 4, 6.

Bardwell, J.C. & Craig, E.A. (1988). Ancient heat shock gene is dispensable. J. Bacteriol. 170, 2977-2983.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. Science 277, 1453-1474.

Bochtler, M., Hartmann, C., Song, H.K., Bourenkov, G.P., Bartunik, H.D. & Huber, R. (2000). The structures of HsIU and the ATP-dependent protease HsIU-HsIV. Nature 403, 800-805.

Bond, J.S. & Beynon, R.J. (1995). The astacin family of metalloendopeptidases. Protein Sci. 4, 1247-1261.

Bonifati, V., Rizzu, P., Squitieri, F., Krieger, E., Vanacore, N., van Swieten, J. C., Brice, A., van Duijn, C. M., Oostra, B., Meco, G. & Heutink, P. (2003). DJ-1(PARK7), a novel gene for autosomal recessive, early onset parkinsonism. Neurol. Sci. 24, 159-160.

Bova, M. P., McHaourab, H. S., Han, Y. & Fung, B. K. (2000). Subunit exchange of small heat shock proteins. Analysis of oligomer formation of alphaA-crystallin and Hsp27 by fluorescence resonance energy transfer and site-directed truncations. J. Biol. Chem. 275, 1035-1042.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998). Crystallography & NMR system: A new

software suite for macromolecular structure determination. Acta Crystallogr. D 54, 905-921.

Buckle, A.M., Zahn, R. & Fersht, A.R. (1997). A structural model for GroEL-polypeptide recognition. Proc. Natl. Acad. Sci. 94, 3571-3575.

Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J.M. & Dideberg, O. (1995). The 3-D structure of a zinc metallo-beta-lactamase from Bacillus cereus reveals a new type of protein fold. EMBO J. 14, 4914-4921.

Chen, L. & Sigler, P.B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. Cell 99, 757-768.

Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. 31, 3320-3323.

Czaja, C., Litwiller, R., Tomlinson, A. J., Naylor, S., Tavares, P., LeGall, J., Moura, J. J. G., Moura, I. & Rusnak, F. (1995). Expression of Desulfovibrio gigas desulforedoxin in Escherichia coli. Purification and characterization of mixed metal isoforms. J. Biol. Chem. 270, 20273-20277.

Daniels, D.S. & Tainer, J.A. (2000). Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O(6)-alkylguanine-DNA alkyltransferase. Mutat. Res. 460, 151-163.

Daiyasu, H., Osaka, K., Ishino, Y. & Toh, H. (2001). Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. FEBS Lett. 503, 1-6.

Dauter, Z. & Dauter, M. (2001). Entering a new phase: using solvent halide ions in protein structure determination. Structure 9, R21-26.

De Biase, D., Tramonti, A., Bossa, F. & Visca, P. (1999). The response to stationary-phase stress conditions in Escherichia coli: role and regulation of the glutamic acid decarboxylase system. Mol. Microbiol. 32, 1198-1211.

Defez, R. & De Felice, M. (1981). Cryptic operon for beta-glucoside metabolism in Escherichia coli K12: genetics evidence for a regulatory protein. Genetics 97, 11-25.

Dekker, C., de Kruijff, B. & Gros, P. (2003). Crystal structure of SecB from Escherichia coli. J. Struct. Biol. 144, 313-319.

Dougan, D.A., Mogk, A. & Bukau, B. (2002a). Protein folding and degradation in bacteria: to degrade or not to degrade? That is the question. Cell Mol. Life Sci. 59, 1607-1616.

Dougan, D.A., Reid, B.G., Horwich, A.L. & Bukau, B. (2002b). ClpS, a substrate modulator of the ClpAP machine. Mol. Cell 9, 673-83.

Drenth, J. (1999). Principles of protein X-ray crystallography (2nd ed.). Springer-Verlag, New York, NY.

Dreyfuss, G., Kim, V.N. & Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat. Rev. Mol. Cell. Biol. 3, 195-205.

Du, X., Choi, I.G., Kim, R., Wang, W., Jancarik, J., Yokota, H. & Kim, S.H. (2000). Crystal structure of an intracellular protease from Pyrococcus horikoshii at 2- Å resolution. Proc. Natl. Acad. Sci. 97, 14079-14084.

Eidsness, M.K., O'Dell, S.E., Kurtz, D.M. Jr., Robson, R.L. & Scott, R.A. (1992). Expression of a synthetic gene coding for the amino acid sequence of Clostridium pasteurianum rubredoxin. Protein Eng. 5, 367-371.

Eisenstein, E., Gilliland, G.L., Herzberg, O., Moult, J., Orban, J., Poljak, R.J., Banerjei, L., Richardson, D. & Howard, A.J. (2000). Biological function made crystal clear - annotation of hypothetical proteins via structural genomics. Curr. Opin. Biotechnol. 11, 25-30.

Erlandsen, H. & Stevens, R.C. (1999). The structural basis of phenylketonuria. Mol. Genet. Metab. 68, 103-125.

Farr, G.W., Furtak, K., Rowland, M.B., Ranson, N.A., Saibil, H.R., Kirchhausen, T. & Horwich, A.L. (2000). Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. Cell 100, 561-573.

Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. (1989). The GroES and GroEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. J. Bacteriol. 171, 1379–1385.

Fenton, W.A., Kashi, Y., Furtak, K. & Horwich, A.L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371, 614-619.

Fraczkiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. J. Comp. Chem. 19, 319-333.

Gajiwala, K.S. & Burley, S.K. (2000). HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. J. Mol. Biol. 295, 605-612.

Gill, S.C. & von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182, 319-326.

Gottesman, S. (1996). Proteases and their targets in Escherichia coli. Annu. Rev. Genet. 30, 465-506.

Hakansson, K. (2002). The strand-helix motif is a recurring theme in biological hydrolysis. Does the conformation of the Ramachandran outlier enhance its electrophilicity? Int. J. Biol. Macromol. 30, 273-277.

Halio, S.B., Blumentals, I.I., Short, S.A., Merrill, B.M. & Kelly, R.M. (1996). Sequence, expression in Escherichia coli, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon Pyrococcus furiosus. J. Bacteriol. 178, 2605-2612.

Han, S., Eltis, L.D., Timmis, K.N., Muchmore, S.W. & Bolin, J.T. (1995). Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading pseudomonad. Science 270, 976-980.

Harker, D. (1956). X-ray diffraction applied to crystalline proteins. Adv. Biol. Med. Phys. 4, 1-22.

Hartl, F.U. & Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295, 1852-1858.

Hendrickson, W.A. (2000). Synchrotron crystallography. Trends Biochem. Sci. 25, 637-643.

Hendrickson, W.A., Smith, J.L., Phizackerley, R.P. & Merritt, E.A. (1988). Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. Proteins 4, 77-88.

Hengge-Aronis, R (1996). Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in Escherichia coli. Mol. Microbiol. 21, 887-893.

Hengge-Aronis, R. (2002). Recent insights into the general stress response regulatory network in Escherichia coli. J. Mol. Microbiol. Biotechnol. 4, 341-346.

Hersh, B.M., Farooq, F.T., Barstad, D.N., Blankenhorn, D.L. & Slonczewski, J.L. (1996). A glutamate-dependent acid resistance gene in Escherichia coli. J. Bacteriol. 178, 3978-3981.

Hesterkamp, T. & Bukau, B. (1996). The *Escherichia coli* trigger factor. FEBS Lett. 389, 32-34.

Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. 20, 478-480.

Holmquist, M. (2000). Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. Curr. Protein Pept. Sci. 1, 209-235.

Holton T.R., Ioerger, T.R., Christopher, J.A. & Sacchettini, J.C. (2000). Determining Protein Structure from Electron-density Maps using Pattern Matching, Acta Crystallogr. D 56, 722-734.

Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J.-P., Danchin, A. & Bertin, P. (2001). Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. Mol. Microbiol. 40, 20-36.

Honbou, K., Suzuki, N.N., Horiuchi, M., Niki, T., Taira, T., Ariga, H. & Inagaki, F. (2003). The crystal structure of DJ-1, a protein related to male fertility and Parkinson's disease. J. Biol. Chem. 278, 31380-31384.

Hotelier, T., Renault, L., Cousin, X., Negre, V., Marchot, P. & Chatonnet, A. (2004). ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. Nucleic Acids Res. 32, D145-D147.

Huai, Q., Sun, Y., Wang, H., Chin, L.S., Li, L., Robinson, H. & Ke, H. (2003). Crystal structure of DJ-1/RS and implication on familial Parkinson's disease. FEBS Lett. 549, 171-175.

Hulo, N., Sigrist, C.J.A., Le Saux, V., Langendijk-Genevaux, P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P. & Bairoch A. (2004). Recent improvements to the PROSITE database. Nucleic Acids. Res. 32, D134-D137.

Humm, A., Fritsche, E. & Steinbacher, S. (1997). Structure and reaction mechanism of L-arginine, glycine amidinotransferase. Biol. Chem. 378, 193-197.

Ishihama, A. (2000). Functional modulation of Escherichia coli RNA polymerase. Annu. Rev. Microbiol. 54, 499-518.

Jakob, U., Muse, W., Eser, M. & Bardwell, J.C. (1999). Chaperone activity with a redox switch. Cell 96, 341-352.

Kamphuis, I.G., Kalk, K.H., Swarte, M.B. & Drenth, J. (1983). Structure of papain refined at 1.65 A resolution. J. Mol. Biol. 179, 233-256.

Kauppi, B., Lee, K., Carredano, E., Parales, R.E., Gibson, D.T., Eklund, H. & Ramaswamy, S. (1998). Structure of an aromatic ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. Structure 6, 571–586.

Kim, S.H., Shin, D.H., Choi, I.G., Schulze-Gahmen, U., Chen, S. & Kim R. (2003). Structure-based functional inference in structural genomics. J. Struct. Funct. Genomics 4, 129-135.

Kim, S.-J., Jeong, D.-G., Chi, S.-W., Lee, J.-S. & Ryu, S.-E. (2001). Crystal structure of proteolytic fragments of the redox-sensitive Hsp33 with constitutive chaperone activity. Nat. Struct. Biol. 8, 459-466.

Kleywegt, G.J. (1996). Use of Non-crystallographic Symmetry in Protein Structure Refinement. Acta Crystallogr. D 52, 842-857.

Korber, P., Stahl, J.M., Nierhaus, K.H. & Bardwell, J.C. (2000). Hsp15, a ribosome-associated heat shock protein. EMBO J. 19, 741-748.

Kultz, D. (2003). Evolution of the cellular stress proteome: from monophyletic origin to ubiquitous function. J. Exp. Biol. 206, 3119-3124.

Lackner, P., Koppensteiner, W.A., Sippl, M.J. & Domingues, F.S. (2000). ProSup: a refined tool for protein structure alignment. Protein Eng. 13, 745-752.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283-291.

Lee, S., Sowa, M.E., Watanabe, Y.H., Sigler, P.B., Chiu, W., Yoshida, M. & Tsai, F.T. (2003a). The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell 115, 229-40.

Lee, S.J., Kim, S.J., Kim, I.K., Ko, J., Jeong, C.S., Kim, G.H., Park, C., Kang, S.O., Suh, P.G., Lee, H.S. & Cha, S.S. (2003b). Crystal structures of human DJ-1 and Escherichia coli Hsp31, which share an evolutionarily conserved domain. J. Biol. Chem. 278, 44552-44559.

Li, J. & Sha, B. (2003). Crystal structure of the E. coli Hsp100 ClpB N-terminal domain. Structure 11, 323-328.

Locascio, G.A., Tigier, H.A. & Battle, A.M. del C. (1969). Estimation of molecular weights of proteins by agarose gel filtration. J. Chromatogr. 40, 453-457.

Lunin, V.Y. (1993). Electron-denisty histograms and the phase problem. Acta Crystallogr. D 49, 90-99.

Macario, A.J. (1995). Heat-shock proteins and molecular chaperones: implications for pathogenesis, diagnostics, and therapeutics. Int. J. Clin. Lab. Res. 25, 59-70.

Macario, A.J. & Conway de Macario, E. (2000). Stress and molecular chaperones in disease. Int. J. Clin. Lab. Res. 30, 49-66.

Malki, A., Kern, R., Abdallah, J. & Richarme, G. (2003). Characterization of the Escherichia coli YedU protein as a molecular chaperone. Biochem. Biophys. Res. Commun. 301, 430-436.

Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. 33, 491-497.

McRee, D.E. (1999). XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Density. J. Structural Biology 125, 156-165.

Mizote, T., Tsuda, M., Smith, D.D., Nakayama, H. & Nakazawa, T. (1999). Cloning and characterization of the thiD/J gene of Escherichia coli encoding a thiamin-synthesizing bifunctional enzyme, hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase. Microbiology 145, 495-501.

Mogk, A. & Bukau, B. (2004). Molecular chaperones, structure of a protein disaggregase. Curr. Biol. 14, R78-80.

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H. & Bukau, B. (1999). Identification of thermolabile Escherichia coli proteins, prevention and reversion of aggregation by DnaK and ClpB. EMBO J. 18, 6934-6949.

van Montfort, R.L.M., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. Nat. Struct. Biol. 8, 1025-1030.

Nardini, M. & Dijkstra, B. W. (1999). Alpha/beta hydrolase fold enzymes: the family keeps growing. Curr. Opin. Struct. Biol. 9, 732-737.

Neuwald, A.F., Aravind, L., Spouge, J.L. & Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. 9, 27-43.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., et al. (1992). The alpha/beta hydrolase fold. Protein Eng. 5, 197-211.

Ostheimer, G.J., Barkan, A. & Matthews, B.W. (2002). Crystal structure of E. coli YhbY: a representative of a novel class of RNA binding proteins. Structure 10, 1593-1601.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol. 276, 307-326 (Academic Press, New York, NY).

Pandya, D.P. (2001). Oxidant injury in coronary heart disease (Part-I). Compr. Ther. 27, 284-292.

Parsons, J.F., Lim, K., Tempczyk, A., Krajewski, W., Eisenstein, E. & Herzberg, O. (2002). From structure to function: YrbI from Haemophilus influenzae (HI1679) is a phosphatase. Proteins 46, 393-404.

Pellecchia, M., Montgomery, D. L., Stevens, S. Y., Van Der Kooi, C. W., Feng, H., Gierasch, L. M. & Zuiderweg, E. R. P. (2000). Structural Insights Into Substrate Binding by the Molecular Chaperone DnaK. Nat. Struct. Biol. 7, 298-303.

Pompeo, F., Brooke, E., Kawamura, A., Mushtaq, A. & Sim, E. (2002). The pharmacogenetics of NAT, structural aspects. Pharmacogenomics 3, 19-30.

Pomposiello, P.J., Bennik, M.H. & Demple, B. (2001). Genome-wide transcriptional profiling of the Escherichia coli responses to superoxide stress and sodium salicylate. J. Bacteriol. 183, 3890-3902.

Porankiewicz, J., Wang, J. & Clarke, A.K. (1999). New insights into the ATP-dependent Clp protease: Escherichia coli and beyond. Mol. Microbiol. 32, 449-458.

Que, L., Jr. (2000). One motif--many different reactions. Nat. Struct. Biol. 7, 182-184.

Quigley, P.M., Korotkov, K., Baneyx, F. & Hol, W.G. (2003). The 1.6-A crystal structure of the class of chaperones represented by Escherichia coli Hsp31 reveals a putative catalytic triad. Proc. Natl. Acad. Sci. 100, 3137-3142.

Quigley, P.M., Korotkov, K., Baneyx, F. & Hol, W.G. (2004). A new native EcHsp31 structure suggests a key role of structural flexibility for chaperone function. Protein Sci. 13, 269-277.

Quillardet, P., Rouffaud, M.A. & Bouige, P. (2003). DNA array analysis of gene expression in response to UV irradiation in Escherichia coli. Res. Microbiol. 154, 559-572.

Radford, N.B., Fina, M., Benjamin, I.J., Moreadith, R.W., Graves, K.H., Zhao, P., Gavva, S., Wiethoff, A., Sherry, A.D., Malloy, C.R. & Williams, R.S. (1996). Cardioprotective effects of 70-kDa heat shock protein in transgenic mice. Proc. Natl. Acad. Sci. 93, 2339-2342.

Rajagopalan, P. T. R., Yu, X. C. & Pei, D. (1997). Peptide Deformylase: A New Type of Mononuclear Iron Protein. J. Am. Chem. Soc. 119, 12418-12419.

Raushel, F.M., Thoden, J.B. & Holden, H.M. (1999). The amidotransferase family of enzymes: molecular machines for the production and delivery of ammonia. Biochemistry 38, 7891-7899.

Richmond, C.S., Glasner, J.D., Mau, R., Jin, H. & Blattner, F.R. (1999). Genome-wide expression profiling in Escherichia coli K-12. Nucleic Acids Res. 27, 3821-3835.

Sastry, M.S., Korotkov, K., Brodsky, Y. & Baneyx, F. (2002). Hsp31, the Escherichia coli yedU gene product, is a molecular chaperone whose activity is inhibited by ATP at high temperatures. J. Biol. Chem. 277, 46026-46034.

Sastry, M.S., Quigley, P.M., Hol, W.G. & Baneyx F. (2004). The linker-loop region of Escherichia coli chaperone Hsp31 functions as a gate that modulates high-affinity substrate binding at elevated temperatures. Proc. Natl. Acad. Sci. 101, 8587-8592.

Sauer, F.G., Pinkner, J.S., Waksman, G. and Hultgren, S.J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. Cell, 111, 543-551.

Schofield, C.J. & Zhang, Z. (1999). Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. Curr. Opin. Struct. Biol. 9, 722-731.

Slepenkov, S.V. & Witt, S.N. (2002). The unfolding story of the Escherichia coli Hsp70 DnaK: is DnaK a holdase or an unfoldase? Mol. Microbiol. 45, 1197-1206.

Soto, C. (2001). Protein misfolding and disease; protein refolding and therapy. FEBS Lett. 498, 204-207.

Sousa, M.C., Trame, C.B., Tsuruta, H., Wilbanks, S.M., Reddy, V.S. & McKay, D.B. (2000). Crystal and solution structures of an HslUV protease-chaperone complex. Cell 103, 633-643.

Soutourina, O.A. & Bertin, P.N. (2003). Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol. Rev. 27, 505-523.

Stirling, P.C., Lundin, V.F. & Leroux, M.R. (2003). Getting a grip on non-native proteins. EMBO Reports 4, 565-570.

Tabatabai, M. A. & Bremner, J.M. (1969). Use of p-nitrophenyl phosphate for assay of soil phosphatase activity. Soil Biol. Biochem. 1, 301-307.

Tao, X. & Tong, L. (2003). Crystal structure of human DJ-1, a protein associated with early onset Parkinson's disease. J. Biol. Chem. 278, 31372-31379.

Tendeng, C. & Bertin, P.N. (2003). H-NS in Gram-negative bacteria, a family of multifaceted proteins. Trends Microbiol. 11, 511-518.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. D 56, 965-972.

Terwilliger, T.C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D 55, 849-861.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Tucker, D.L., Tucker, N. & Conway, T. (2002). Gene expression profiling of the pH response in Escherichia coli. J. Bacteriol. 184, 6551-6558.

Veinger, L., Diamant, S., Buchner, J. & Goloubinoff, P. (1998). The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. J. Biol. Chem. 273, 11032-11037.

Vijayalakshmi, J., Mukhergee, M.K., Graumann, J., Jakob, U. & Saper, M. (2001). The 2.2 Å crystal structure of Hsp33: a heat shock protein with redox-regulated chaperone activity. Structure 9, 367-375.

Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol. 115, 90-112.

Waterman, S.R. & Small, P.L. (1996). Identification of sigma S-dependent genes associated with the stationary-phase acid-resistance phenotype of Shigella flexneri. Mol. Microbiol. 21, 925-940.

Wilson, M.A., Collins, J.L., Hod, Y., Ringe, D. & Petsko, G.A. (2003). The 1.1-A resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease. Proc. Natl. Acad. Sci. 100, 9256-9261.

Wilson, M.A., St Amour, C.V., Collins, J.L., Ringe, D. & Petsko, G.A. (2004). The 1.8-A resolution crystal structure of YDR533Cp from Saccharomyces cerevisiae: a member of the DJ-1/ThiJ/PfpI superfamily. Proc. Natl. Acad. Sci. 101, 1531-1536.

Weber, A. & Jung, K. (2002). Profiling early osmostress-dependent gene expression in Escherichia coli using DNA macroarrays. J. Bacteriol. 184, 5502-5507.

Xie, P., Parsons, S. H., Speckhard, D. C., Bosron, W. F. & Hurley, T. D. (1997). X-ray structure of human class IV sigmasigma alcohol dehydrogenase. Structural basis for substrate specificity. J. Biol. Chem. 272, 18558-18563.

Xu, Z., Horwich, A.L. & Sigler, P.B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperone complex. Nature 388, 741-750.

Xu, Z., Knafels, J. D. & Yoshino, K. (2000). Crystal structure of the bacterial protein export chaperone secB. Nat. Struct. Biol. 7, 1172-1177.

Xu, Z. & Sigler, P.B. (1998). GroEL/GroES, structure and function of a two-stroke folding machine. J. Struct. Biol. 124, 129-141.

Yoshida, T., Ueguchi, C., Yamada, H. & Mizuno, T. (1993). Function of the Escherichia coli nucleoid protein, H-NS: molecular analysis of a subset of proteins whose expression is enhanced in a hns deletion mutant. Mol. Gen. Genet. 237, 113-122.

Yura, T., M. Kanemori & M. T. Morita. (2000). The heat shock response: regulation and function, p. 3-18. In G. Storz and R. Hengge-Aronis (ed.), Bacterial stress responses. ASM Press, Washington, D.C.

Zarembinski, T.I., Hung, L.W., Mueller-Dieckmann, H.J., Kim, K.K., Yokota, H., Kim, R. & Kim, S.-H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. Proc. Natl. Acad. Sci. 95, 15189-15193.

Zhao, Y., Liu, D., Kaluarachchi, W. D., Bellamy, H. D., White, M. A. & Fox, R. O. (2003). The crystal structure of Escherichia coli heat shock protein YedU reveals three potential catalytic active sites. Protein Sci. 12, 2303-2311.

Zheng, M., Wang, X., Templeton, L.J., Smulski, D.R., LaRossa, R.A. & Storz, G. (2001). DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. J. Bacteriol. 183, 4562-4570.

VITA

Yonghong Zhao was born in Henan province, China on July 20, 1972, son of Wensheng Zhao and Fuqin Hao. Yonghong married Xin Guo in 2002 and currently the couple has one boy, Jason Xin Zhao. Yonghong was enrolled in the Cellular Physiology and Molecular Biophysics (CPMB) Graduate Program at The University of Texas Medical Branch (UTMB) in 1999. Yonghong gained significant research experience and published two peer-reviewed papers while at UTMB. He also served as the co-chair of the organizing committee for the 6th Life Science Symposium held at UTMB in 2002.

Education

B.S. Biophysics, 1993, Nankai University, Tianjin, China

M.S. Biophysics, 1996, Institute of Biophysics, Beijing, China

Publications

Zhao, Y. & Kim, L.-S. (1997). Analysis of dynamics in countercurrent chromatography. Progress in Biochemistry & Biophysics (in Chinese) 24, 144-147.

Zhao, Y. & Kim, L.-S. (1998). Centrifugal countercurrent partition chromatography with a chain-winding column. J. of Liquid Chromatogr. & Related Tech. 21, 2387-2396.

Zhao, Y., Liu, D., Kaluarachchi, W.D., Bellamy, H.D., White, M.A. & Fox, R.O. (2003). The crystal structure of Escherichia coli heat shock protein YedU reveals three potential catalytic active sites. Protein Sci. 12 , 2303-2311.

Liu, D., Zhao, Y., Fan, X., Sun, Y. & Fox R.O. (2004). Expression, crystallization and preliminary crystallographic analysis of YciE, a stress protein from Escherichia coli. Acta Crystallogr. D 60, 1888-1889.