EXPLORING NEW STRUCTURAL AND FUNCTIONAL SPACE IN THE GLUTATHIONE TRANSFERASE SUPERFAMILY FROM *Escherichia coli* K-12

By

Megan Christine Branch

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

August, 2011

Nashville, Tennessee

Approved:

Professor Richard N. Armstrong

Professor F. Peter Guengerich

Professor Walter J. Chazin

Professor Lawrence J. Marnett

Professor D. Borden Lacy

To my wonderful family

ACKNOWLEDGMENTS

I would like to express my gratitude to the following people:

My advisor, professor Dr. Richard Armstrong, for providing support, encouragement and scientific advice during my years of graduate research. I thank him for the opportunity to be part of this research group, helping me develop as a scientist and learn from his extensive scientific experience.

The outstanding members of my thesis committee, Dr. Fred Guengerich, Dr. Walter Chazin, Dr. Lawrence Marnett, and Dr. Borden Lacy. Thank you for your support and guidance throughout the years.

Dr. Joel Harp, the Director of Operations for Crystallography at the Center for Structural Biology, who provided invaluable scientific advice, encouragement, and support in my crystallographic endeavors. Thank you for taking the time to train a student with no previous experience and patiently answering my questions in the theoretical and practical aspects of protein crystallography.

Members of the Marnett, Guengerich, and Iverson research groups. In particular, Dr. Philip Kingsley for assistance with HPLC experiments and Dr. Elizabeth Loecken, Dr. Christal Sohl, and Dr. Charles Knutson for both scientific and academic support. Dr. Timothy Panosian for assistance with data collection and processing at the Advanced Photon Source.

I would like to gratefully thank all my current and former colleagues. In particular, Dr. Laura Bushenlener, Dr. Nina Stourman, and Dr. Paul Cook for invaluable

scientific discussions, patient support, and encouragement. Much of the work presented herein could not have been accomplished with the initial effort of Dr. Nina Stourman. She provided knockouts and clones, as well as scientific support and technical supervision that allowed me to accomplish this work. To my friend and fellow graduate student, Edward Prage, thank you for your support and humor; you made many long days in the laboratory more enjoyable.

I would like to thank the staff of the Biochemistry department Brenda Bilbrey, Peggy Fisher, Marlene Jayne, and Ellen Rochelle for their assistance during my graduate career.

The support of my friends and family has been paramount to my success over this long effort. To my mother and father, thank you for fostering my love of learning and always believing in me. Your love and support means so much. I would like to express my gratitude and love to my husband Kevin, your patience, encouragement and humor helped me through the most difficult moments in graduate school. Thank you so very much.

Finally, I would like to thank the Vanderbilt University Chemistry-Biology Interface Training Grant and the National Institutes of Health for financial support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATP	adenosine-5'-triphosphate
CDNB	1-chloro-2,4-dinitrobenzene
COG	clusters of orthologous groups
DCM	dichloromethane
DEAE	diethylaminoethyl
DiCHQ	dichlorohydroquinone
DNA	deoxyribonucleic acid
DTNB	5, 5'-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GSH	glutathione
GSP	glutathionylspermidine
GSSG	glutathione disulfide, oxidized glutathione
HA	hydroxyapatite
HCCA	2-hydroxychromene-2-carboxylic acid
HED	2-hydroxyethyldisulfide
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	luria broth
MES	2-(N-morpholino)ethanesulfonic acid

PDB	protein data bank
PEG	polyethyleneglycol
RMSD	root-mean-square deviation
SDS	sodium dodecyl sulfate
SP	sulphopropyl
TCHQ	tetrachlorohydroquinone
TriCHQ	trichlorohydroquinone
UDP	uridine diphosphate

CHAPTER I

INTRODUCTION

Detoxication Enzymes

All living organisms are increasingly exposed to a seemingly limitless variety of naturally occurring and synthetic chemicals. These chemicals range from metals and small inorganic compounds to large complex organic molecules. In agricultural settings these compounds include herbicides and pesticides, where as, in the chemical industry they include solvents and chemical manufacturing intermediates [1]. An organism's ability to detoxicate and excrete these chemicals is of central importance to itssurvival. If allowed to persist, these toxic chemicals have the ability to modify hydrogen-bonding interactions between DNA bases resulting in strand breakage and other genotoxic effects [2]. Inert hydrophobic molecules exhibit low solubility and associate with cellular lipids. Ultimately, both reactive and inert molecules must be transformed into hydrophilic species of low chemical potential for effective detoxication and elimination [3].

In response to possible cellular damage by electrophiles and reactive oxygen species arising from both foreign compounds and endogenous metabolic processes, organisms have developed effective biochemical mechanisms to protect themselves from chemical insult. Detoxication enzymes catalyze the oxidation, reduction, cleavage, rearrangement, and conjugation of an enormous variety of chemical compounds, thereby deactivating the reactive molecules and preparing them for further metabolism and rapid elimination [4]. Detoxication enzymes employ several strategies in order to avoid the

impossible task of manufacturing a new enzyme for each individual xenobiotic. They exhibit broad substrate specificity and can catalyze a range of reactions. Many organisms encode multiple copies of detoxication genes. The enzymes encoded by these genes have different and overlapping substrate specificities. Finally, enhanced expression of detoxication enzymes in response to xenobiotic exposure allows an organism to rapidly and specifically respond to a chemical threat [3].

Important families of detoxication enzymes include the cytochrome P450s, glutathione (GSH) transferases, sulfotransferases and UDP-glucuronosyltransferases. Cytochrome P450s are heme containing proteins that act through a general catalytic mechanism involving the abstraction of electrons or hydrogens followed by oxygen rebound [4]. By this mechanism, P450s catalyze diverse monooxygenation reactions at aliphatic and aromatic carbon, nitrogen, and sulfur atoms, including hydroxylations, epoxidations, heteroatom oxidations, dealkylations, deaminations and reductions [4]. The number of P450 substrates is easily in the thouands, and they are the primary enzyme targets in drug metabolism [4,5]. Often P450 enzymes transform xenobiotics into more hydrophilic compounds, preparing them for further metabolism. GSH transferases catalyze the addition of GSH to reactive electrophiles, which removes the reactive functional groups and tags the compound for removal from the cell by the ATP-dependent GSH-conjugate export pumps [6].

It is important to recognize that some enzyme-catalyzed detoxication reactions actually result in bioactivation of foreign compounds, converting them to more reactive metabolites. A classic example of metabolic activation is the metabolism of the polycyclic aromatic hydrocarbon benzo(*a*)pyrene (Figure I-1). This initially inert species

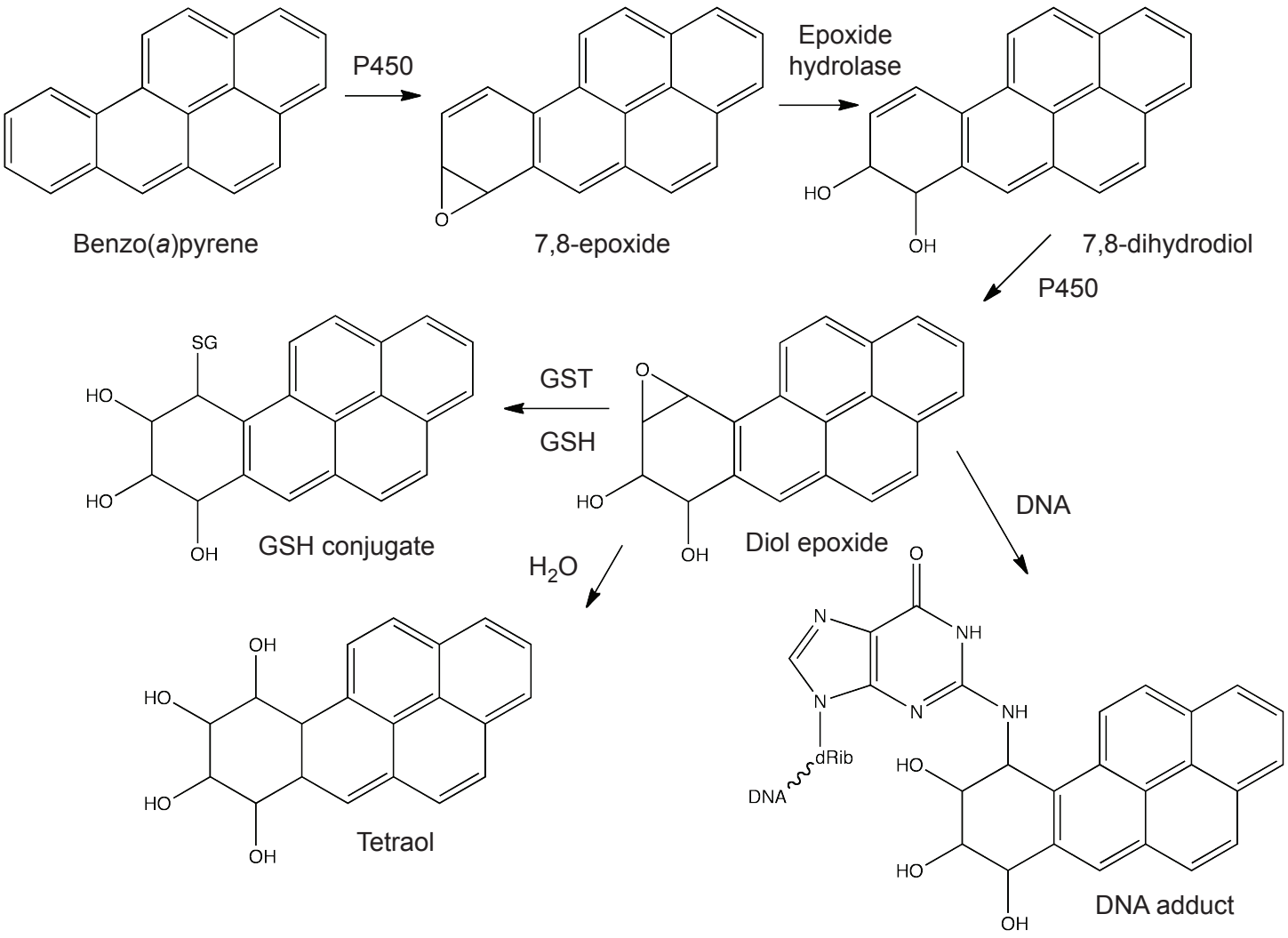


Figure I-1. Biotransformation and metabolic activation of the carcinogen benzo*(a)*pyrene. Conversion to a reactive diol epoxide, plus reactions of the diol epoxide with water, DNA and GSH. (adapted from 4)

is oxidized to an epoxide by cytochrome P450 monooxygenase and then hydrolyzed by epoxide hydrolase to form a diol. A second, rapid oxidation by P450 results in a highly reactive diol epoxide that can interact with water, or DNA or undergo GSH conjugation by a GSH transferase [4].

The Global Glutathione Transferase Superfamily

The discovery of an enzymatic activity for the addition of GSH to 1-chloro-2,4-dinitrobenzene (CDNB) in cytosolic liver extracts in 1961 marked the beginning of decades of research on the GSH transferase (E.C. 2.5.1.18) superfamily [7,8]. The GSH transferases are a multifunctional superfamily of proteins with fundamental roles in the cellular detoxication of a wide range of exogenous and endogenous compounds.

Metabolic Reactions

Generally, these enzymes catalyze the nucleophilic attack of the tripeptide glutathione [(γ-L-glutamyl-L-cysteinyl)glycine), GSH, Figure I-2] to nonpolar compounds containing an electrophilic carbon, nitrogen or sulfur atom (Scheme I-1) [9]. The substrates for GSH transferases include halonitrobenzenes, arene oxides, alkyl and aryl halides, quinones, and α,β-unsaturated carbonyls [9,10]. The resulting GSH-conjugates are often less reactive and more soluble than the parent xenobiotic compound. These conjugates are exported from cells by the transmembrane MRP (multi-drug resistance associated protein) family of transporters. Nine MRPs exist but MRP1 and MRP2 are able to export GSH-conjugates [10]. Once exported, the GSH-conjugates are metabolized by the mercapturic acid pathway. The γ-glutamyl and glycyl groups are removed and the amino

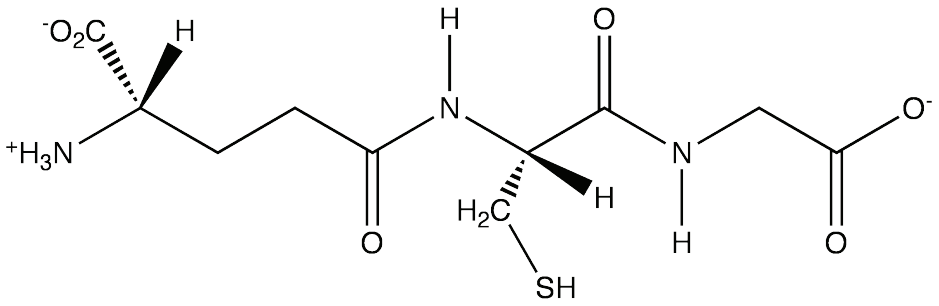
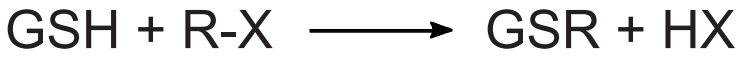


Figure I-2. The chemical structure of glutathione [γ-L-glutamyl-L-cysteinyl) glycine], GSH



Scheme I-1. General reaction catalyzed by a glutathione transferase.

group of the resulting cysteine conjugate is acetylated [6]. GSH transferases play key roles in the metabolism of a broad range of structurally diverse compounds, therefore, aiding in the cellular detoxication of alkylating agents. But they can also contribute to resistance to chemotherapeutic compounds [11]. GSH transferases are upregulated in certain tumor types and polymorphisms of human GSH transferase genes are implicated in responsiveness to cancer therapies [11].

In addition to GSH conjugation chemistry, members of this diverse superfamily also catalyze nucleophilic aromatic substitutions, epoxide ring openings, reversible Michael additions, and isomerization and peroxidase reactions [6,12]. Some GSH transferases do not participate in catalysis but instead bind substrate ligands for intracellular storage or participate in the transport of hormones, metabolites, drugs and hydrophobic compounds [10]. GSH transferases also have the ability to modulate signaling pathways and regulate the processes of transcription and translation [10].

Classification of Glutathione Transferases

There are at least four major groups of GSH transferases; (1) the cytosolic (or canonical) GSH transferases (2) the metallothiol bacterial fosfomycin resistance proteins (3) the mitochondrial kappa proteins and (4) the microsomal GSH transferases (also called Membrane Associated Proteins in Eicasanoid and Glutathione Metabolism, abbreviated MAPEG) [13].

The metallothiol fosfomycin proteins belong to the vicinal oxygen chelate superfamily, which includes glyoxylase I, methylmalonyl-CoA epimerase, extradiol dioxygenase and the bleomycin resistance proteins [14]. FosA, B and X proteins have been

identified in bacteria. These metal-dependent enzymes catalyze the addition of GSH, L-cysteine (or possibly another low molecular weight thiol) [15], or water to the antibiotic fosfomycin, rendering it inactive [16].

The mitochondrial or kappa class of GSH transferases was initially isolated from rat mitochondria. Mouse, rat and human have a single kappa GSH transferase [10]. Similar to cytosolic GSH transferases, the proteins possess a thioredoxin domain and an all-alpha-helical domain, but the topology is more similar to the *Escherichia coli* disulfide bond isomerase (DsbA) protein [17]. Microorganisms posses enzymes that may be related to the kappa class GSH transferase, one example is the 2-hydroxychromene-2-carboxylic acid (HCCA) isomerase protein from *Pseudomonas putida* which participates in naphthalene catabolism [18].

MAPEG proteins are integral membrane proteins with highly divergent functions in eicosanoid and GSH metabolism. These proteins have been divided into four subgroups with proteins within a subgroup sharing >20% sequence identity. Human MAPEG members participate in leukotriene biosynthesis and act as GSH transferases and peroxidases. They are typically trimers with four transmembrane α-helices per monomer [10,19].

The most extensively studied group of GSH transferases is the soluble or canonical enzymes. These proteins are found in all classes of eukarya and bacteria. They are classified into subgroups based roughly on protein sequence, function, and structure. Mammalian members of the same class share >40% amino acid identity and proteins between classes sharing <20% sequence identity [10]. Widely used class designations of the cytosolic GSH transferases include, -alpha, -mu, -omega, -phi, -pi, -sigma, -theta, and –

zeta [10,13]. In recent years with the discovery and characterization of new GSH transferases (particularly in microorganisms), new classes have been proposed. These include –beta, -delta, -lamda, -epsilon, -nu-1, -nu-2, -tau, -chi and -xi. Inferences of homology and the accurate classification of the cytosolic GSH transferases is a complex issue. For example, the discovery of class theta enzymes in plants and insects greatly increased the range of organisms with recognizable GSH transferase family members [13]. The theta class enzymes are much more phylogenetically diverse than the mammalian enzymes. Because of this great diversity, assignment to class theta has little meaning [13]. The same may be true for bacteria. Cytosolic GSH transferases from bacteria vary greatly in sequence, function, and structure [20]. There may be at least one specific bacterial class (beta) but this single class is insufficient to encapsulate the diversity of proteins in this domain. Careful sequence analysis, integrated with functional and structural information, will be necessary for the continued classification of cytosolic GSH transferases.

The Three Dimensional Structures of Cytosolic Glutathione Transferases

Considerable structural information is available for the cytosolic GSH- transferase superfamily. In general, despite low primary amino acid sequence identity, the overall fold and spatial features of the proteins are remarkably similar [21]. Most cytosolic GSH transferases function as dimeric proteins composed of two structurally distinct domains. Each subunit contains an N-terminal thioredoxin-like domain, and a C-terminal all α-helical domain. The thioredoxin-like domain comprises approximately one-third of the protein and consists of a β-α-β-α-β-β-α structural motif, which forms a four strand β sheet flanked by two α-helices on one side and one α-helix on the other [9]. The larger, C-

terminal, all α-helical domain is a unique protein fold with a bundle of four α-helices at the core [21]. The N-terminal domain contains the GSH binding site while the C-terminal domain contains the structural elements for xenobiotic recognition [9]. Representative structures of cytosolic GSH transferases are shown in Figure I-3.

The GSH molecule typically binds in an extended conformation with multiple hydrogen bonding interactions with the protein. Significant differences in the details of the hydrogen bonding patterns exist between the various enzymes [9]. The most conserved structural regions in the cytosolic GSH transferases is the core β-β-α motif responsible for recognition of the γ-glutamyl portion of GSH [9]. Amino acids at the turn between β-sheet 4 and α-helix 3 recognize the amino and carboxylate groups of the chiral α-carbon of the γ-glutamyl portion of GSH [9,21,22]. A global analysis of the sequence and structural similarities of the cytosolic GSH transferases allows classification into two major subgroups based on the amino acid that interacts directly with the nucleophilic sulfur of GSH [23]. Members of the first major subgroup (Y- GSTs), have a tyrosine positioned at the end of the first β-strand, while the second major subgroup has a serine or cysteine several positions later at the amino terminus of α-helix 1 (S/C-GSTs) [23]. When a tyrosine or serine is present the hydroxyl group of these amino acids acts as a hydrogen bond donor to the sulfur of GSH, lowering its pK_a to stabilize a nucleophilic thiolate [9,22]. In the omega class proteins, which contain an active site cysteine, thioltransferase activity is observed [24]. Zeta class enzymes have both a serine and cysteine residue, which is critical to their functions but also have a reactive active site cysteine that plays a role in binding GSH, although this cysteine is not required for catalysis [25,26].

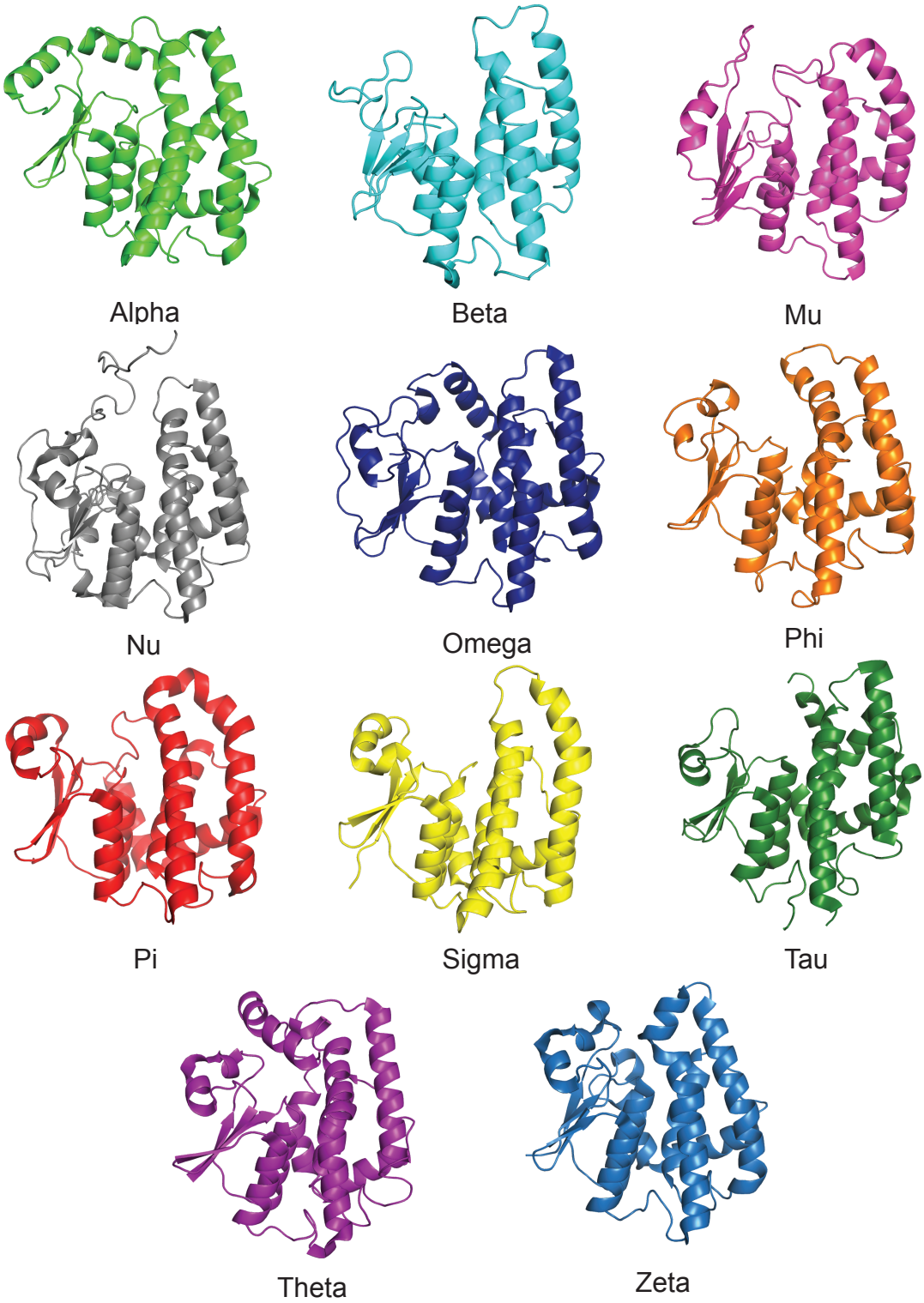


Figure I-3. Representative monomer folds for several classes of glutathione transferases. The structures are: Alpha, GSTA3_CHICK, 1VF1; Beta, GSTB_ECOLI, 1A0F; Mu, GSTM2_HUMAN, 3GTU; Nu, YGHU_ECOLI, 3C8E; Omega, GSTO_HUMAN, 1EEM; Phi, GSTF4_ARATH, 1GNW; Pi, GSTP_ONCV, 1TU7; Sigma, GST1_DROME, 1MOU; Tau, GSTU1_ORYSJ, 1OYJ; Theta, GSTT1_HUMAN, 2C3N; Zeta, MAAI_MOUSE, 2CZ2.

Most cytosolic GSH transferases function as homo- or hetero- dimers. Heterodimers exist between isoenzymes within a gene subfamily but do not occur between intergene subfamilies [21]. The principal subunit interactions occur between the thioredoxin-like domain of one subunit and the α-helical domain of its partner [9]. The subunit interface of the alpha, mu, and pi class is characterized by a ball-and-socket hydrophobic interaction. The side chain of a phenylalanine residue (the ball) located on the loop between α-helix 2 and β-sheet 3, protrudes into a hydrophobic socket formed from residues on α-helices in the all α-helical domain (Figure I-4) [9]. The ball-and-socket hydrophobic interaction is missing in the sigma, theta and omega class proteins. Proteins in these three classes all have a crevice at the dimer interface, making it more open and decreasing the buried surface area (Figure I-5). The omega interface is dominated by non-polar interactions and the buried surface area at the dimer interface is only 1,960 $Å^2$, compared to 2,700-3,400 $Å^2$ in other classes [12].

Monomeric forms of cytosolic GSH transferases have been described. Of particular interest are the intracellular chloride ion channels (CLIC). The CLIC1 protein is a structural homologue to GSH transferases [27]. The protein possesses an intact GSH binding site and has a redox-active cysteine, capable of forming a covalent mixed disulfide with GSH [27]. The protein undergoes a reversible transition from a monomeric to a dimeric state due to the formation of an intramolecular disulfide bond [28]. It has been proposed that this large conformational rearrangement facilitates membrane insertion [28].

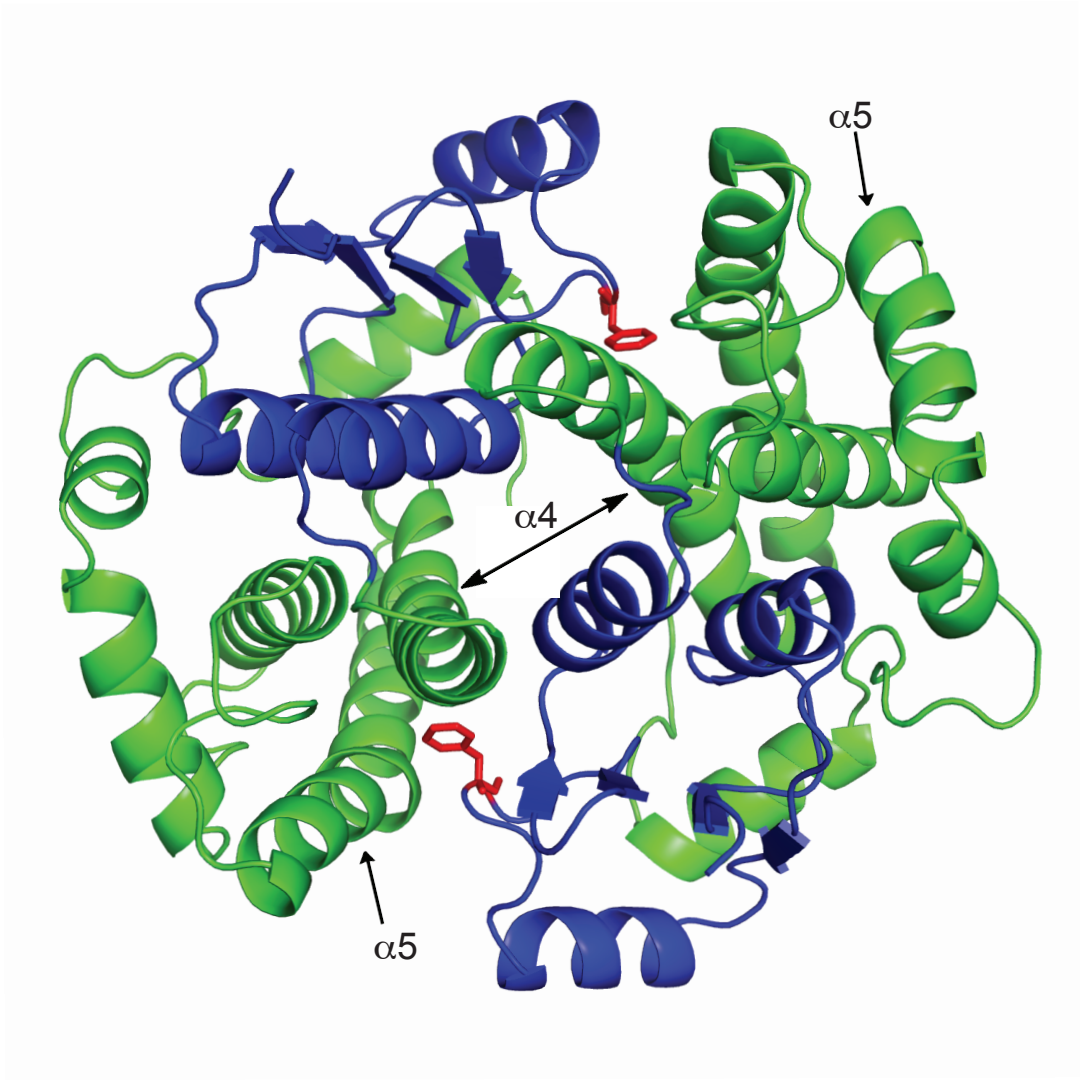


Figure I-4. Illustration of the hydrophobic ball-and-socket dimer interface typical in alpha, mu and pi glutathione transferases. A ribbon diagram of the alpha class glutathionetransferase from chicken is shown (GSTA3_CHICK, PBD ID: 1VF1). The thioredoxin-like domains are colored in blue and phenylalanine 52 is highlighted in stick representation and colored in red.

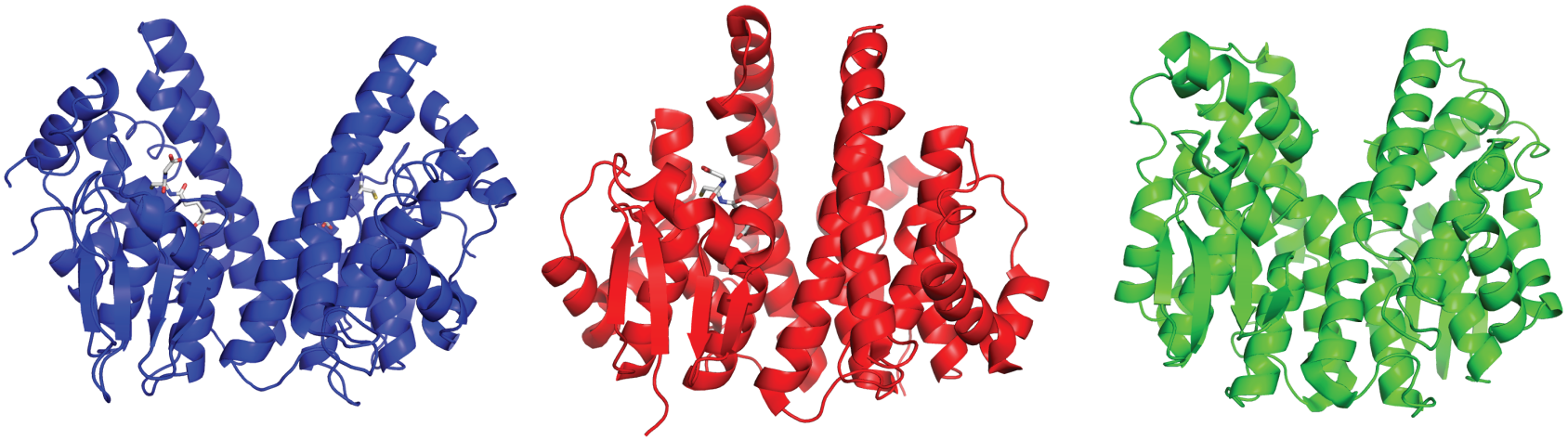


Figure I-5. Protein structures of omega (blue), sigma (red) and theta (green) class glutathione transferases highlight the crevice found between the dimer interface in these classes (omega, GSTO_HUMAN, 1EEM; sigma, GST1_DROME, 1MOU; theta, GSTT1_HUMAN, 2C3N).

Glutathione Transferases In Bacteria

T. Shishido reported the first evidence of a GSH transferase in bacteria thirty years ago in a strain of *E. coli* [29]. Today, we appreciate the large number of GSH transferases that are widely distributed throughout prokaryotes and bacteria in particular. Many bacteria have multiple GSH transferase genes (Figure I-6), but most remain only as annotations in sequenced genomes with little functional or structural information ascribed to them. Until recently relatively few bacterial GSH transferases have been characterized in detail and those that have been studied show modest catalytic activity and a lack of substrate specificity [30]

The first and most well studied bacterial GSH transferases are proteins from *Proteus mirabilis* (PmGST) and *E. coli* (EGST). These proteins, combined with two GSTs from soil bacteria (OaGST from *Ochrobacturm anthropi* and BphK/BxGST from *Burkholderia xenovorans*), form the hallmark members of the beta class in bacteria [20]. These enzymes exhibit modest catalytic activity with (CDNB) when compared to their eukaryotic counterparts [31,32]. Despite their low level of sequence identity (<35%), the structures of these four proteins superimpose remarkably well with an RMSD of only 0.983 Å over 166 structurally equivalent C_α atom pairs. All have an active site cysteine residue. In the OaGST and BxGST structures, the average distance between the sulfurs is ~3.3 Å, indicating a hydrogen bond between the GSH-thiolate and the hydrogen of the cysteine on the protein [33,34]. A mixed cysteine-GSH disulfide was observed in the PmGST structure and the distance between the two sulfurs is 2.2 Å [35]. Interestingly, this active site cysteine is not required for GSH conjugation activity with CDNB in the PmGST or EGST enzymes, but does significantly affect the activity of OaGST [30,36-39]. The *in vivo*

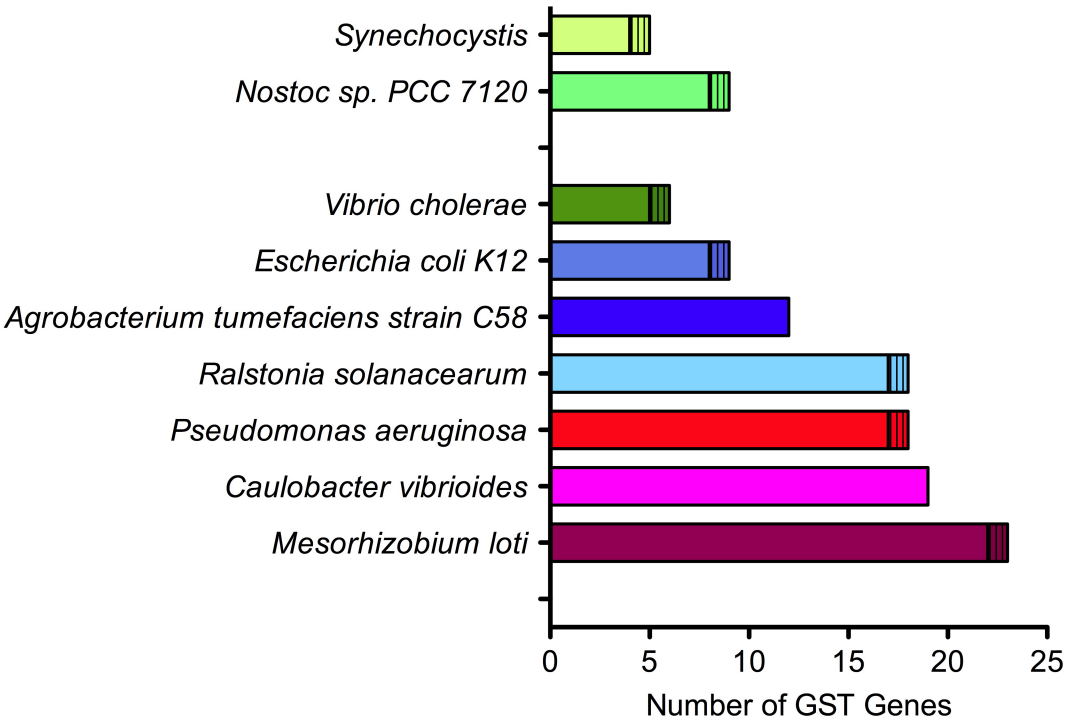


Figure I-6. Multiple GSH-transferase genes are present in various bacterial genomes. Data gathered from the NCBI database, Clusters of Orthologous Groups of Proteins (COGs) (http://www.ncbi.nlm.nih.gov/COG/). Number of glutathione transferase sequences from COG0625 (glutathione transferase) and COG0435 (predicted glutathione transferase). Sequences from COG0625 are denoted as a solid bar while COG0435 is denoted with a horizontal slash.

physiological functions of PmGST and EGST have not been established. However, the gene that encodes the BphK enzyme in *B. xenovorans* is located within a gene cluster responsible for the degradation of biphenyl and chlorobiphenyl compounds. Consequently, the BphK enzyme has been found to catalyze the dehalogenation of compounds in the metabolism of polychlorinated biphenyls [33]. *O. anthropi* is a soil bacterium that is able to grow on atrazine as a carbon source. OaGST is thought to be involved in atrazine conjugation with GSH, although the exact mechanism is not known [40].

Cyanobacterial genomes may encode multiple GSH transferase genes because of the high levels of cytosolic GSH in these organisms [41,42]. Indeed, in the two sequenced cyanobacterial genomes in the COG database (http://www.ncbi.nlm.nih.gov/COG/), *Synechocystis sp.* contains five GSH transferases and one predicted GSH transferase. The *Nostoc sp.* PCC 7120 genome encodes eight GSH transferases and one predicted GSH transferase. The TeGST from *Thermosynechoccus elongates* BP-1 and the SeGST from *Synechoccus elongatus* PCC 6301 have been characterized. These proteins completely lack cysteine residues and are shorter in sequence (~186 amino acids) compared to other bacterial GSH transferases (~200 amino acids) [43]. They exhibit modest catalytic activity towards naturally occurring isothiocyanates (k_{cat}/K_m $^{\text{phenethyl-ITC}}$ = 300-600 $mM^{-1}s^{-1}$), but their physiological roles in the bacterium have been not established [43]. They have been assigned to the new 'chi' class of cytosolic GSH transferases.

Bacterial GSH transferases also play important roles in the biodegradation of chlorinated xenobiotics. Halogenated compounds are widely used in industry and are an important group of environmental pollutants. Dichloromethane (DCM) dehalogenase is a

GSH-dependent enzyme that catalyzes the hydrolytic dechlorination of dichloromethane in a number of facultative methyltrophic bacteria. DCM dehalogenase allows the bacteria to grow on DCM as a sole carbon source [44]. The DM11 enzyme from *Methylophilus leisingeri* and DM4 enzyme from *Methylophilus dichloromethanicum* are well characterized [45,46]. They are similar in sequence to theta class GSH transferases and share a conserved serine in the N-terminal region that is essential for catalysis [47]. The tetrachlorohydroquinone (TCHQ) dehalogenase enzyme from *Sphingobium chlorophenolicum* utilizes two molecules of GSH to reductively dechlorinate TCHQ to trichlorohydroquinone (TriCHQ) and then dichlorohydroquinone (DiCHQ) [48]. The gene encoding this enzyme is found in a gene cassette that allows the soil bacteria to grow on pentachlorophenol as a sole carbon source [49].

Two GSH transferases are located in the isoprene degradation gene cluster in the gram-positive bacterium *Rhodococcus sp.* (strain AD45) [50]. Isoprene is an atmospheric reactive hydrocarbon and the *Rhodococcus* genus of bacterium is well characterized for its ability to biodegrade organic pollutants [51]. The *isoI* gene encodes a GSH transferase that catalyzes the GSH-dependent ring opening of isoprene monoxide, the primary oxidation product of isoprene. The second GSH transferase, encoded by the *isoJ* gene, is thought to remove a molecule of GSH from a downstream degradation product of isoprene [52]. However, the full pathway of isoprene degradation has not been described.

In the gram-negative soil bacteria, *Sphingomonas paucimobilis*, three GSH transferases are involved in the process of lignin degradation [53]. *S. paucimobilis* is able to degrade a variety of lignin-related biphenyl chemical compounds. The *ligE*, *ligF* and *ligG* genes are located next to each other [53]. LigE and LigF are enantioselective GSH

transferases that consume GSH while cleaving the β–aryl ether linkage of lignin. LigG catalyses the removal of GSH from the conjugates produced by LigE and LigF [54].

The enzyme HCCA isomerase is a GSH-dependent enzyme responsible for the interconversion of HCCA and *trans*-*o*-hydroxybenzylidene pyruvic acid in the naphthalene catabolic pathway in *Pseudomonas putida* [18]. HCCA is structurally related to the kappa class GSH transferases in that the α-helical domain is inserted into the thioredoxin fold [18]. At this time the distribution of HCCA isomerase proteins in bacteria is not understood.

The membrane associated proteins in eicosanoid and glutathione metabolism (MAPEG) family of GSH transferases are also distributed in bacteria. MAPEG members have been identified in *E. coli*, *Vibrio cholerae*, and *Synechocystis sp.* among others [19]. The sequences of these proteins are only ~132 amino acids in length and are believed to form homotrimers with membrane spanning domains. These proteins share varying degrees of sequence similarity. The Y1147 MAPEG member from *Synechocystis sp.* only shares 7% sequence identity with the Q9KNE3 protein from *V. cholerae*, but the *E. coli* and the *V. cholerae* proteins share 32% sequence identity. The bacterial MAPEG family members show distinct evolutionary relationships from eukaryotic mammalian and non-mammalian MAPEG members [55]. Membrane fractions from bacteria over expressing the *E. coli* and *Synechocystis sp.* MAPEG proteins have GSH transferase activity with CDNB, but the physiological role of the proteins in bacteria is still unknown [55].

The Glutathione Transferase Superfamily in *Escherichia coli* K-12

The genome of *E. coli* K-12 encodes multiple GSH transferase homologues (Figure I-7). Nine of these genes are believed to encode cytosolic proteins belonging to the classical or canonical family of GSH transferases. These proteins belong to two clusters of orthologous groups; COG0625, representing the glutathione transferases and COG0435, the predicted glutathione transferases. Initially, only two of the nine gene products were assigned a defined function.

The *gst* gene encodes a genuine GSH transferase and demonstrates modest GSH conjugation activity with CDNB [57]. However, its *in vivo* physiological substrates and biological role are not known. The protein adopts a classical GSH transferases fold and was crystallized in the presence of glutathione sulfonate [37,58]. In contrast to mammalian GSH transferases, the active site residues of EGST, including Cys-10 and His-106, are not strictly required for GSH-conjugation activity [36,39]. Analysis of the sequences of the GSH transferases homologues from *E. coli* in the context of the EGST crystal structures reveal that the most highly conserved amino acids are involved in stabilization of the core of the structure [58].

The *sspA* (stringent starvation protein A) gene in *E. coli* encodes a protein that may be structurally related to GSH transferases. There are now four structures of sspA proteins in various bacteria. The proteins crystallize as dimers and adopt a classical GSH transferases fold with an N-terminal thioredoxin-like domain and a C-terminal all α-helical domain [59]. However, the SspA proteins do not bind GSH or use the molecule in any catalytic reaction. Instead, the *E. coli* sspA protein was identified almost 32 years ago as a RNA polymerase-associated protein required for the transcriptional activation of

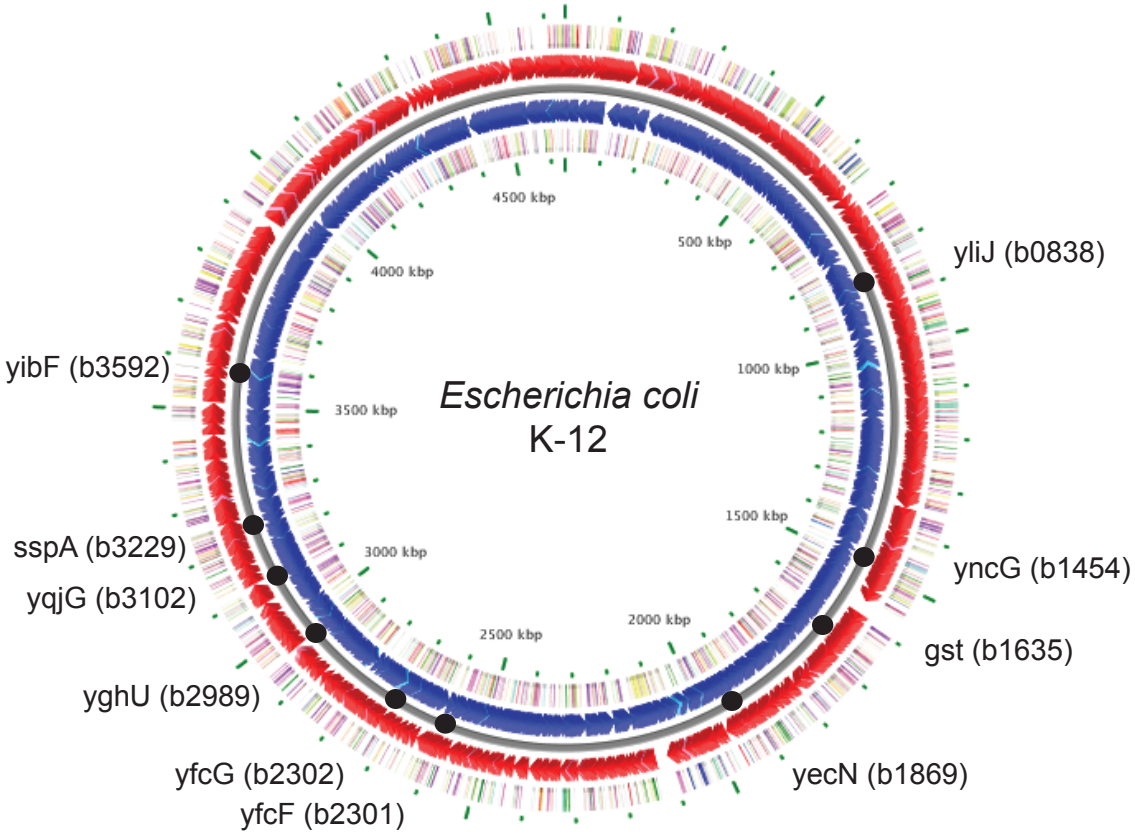


Figure I-7. The glutathione transferase superfamily in *Escherichia coli* K-12. Eight members belong to COG0625 (glutathione transferase, yibF, sspA, yghU, yfcG, yfcF, gst, yncG, yliJ), one to COG0435 (predicted glutathione transferase, yqjG) and one is believed to encode a MAPEG family member (yecN).

phage P1 late genes [60]. While the exact role of SspA in transcription is still unknown, it is implicated in survival during nutrient starvation and prolonged stationary phase [61]. SspA has been shown to play an important role in acid resistance during stationary phase by down regulating the DNA binding protein H-NS [62]. The H-NS protein binds curved double stranded DNA sequences and is thought to act as a global regulator of transcription.

A recent report has described the efficient dehalogenation of bromoacetate by the *E. coli* GSH transferase, YliJ [64]. The Keio collection of > 4,000 *E. coli* single gene knockouts was screened for hypersensitivity to bromoacetate. In total, nine proteins were confirmed as displaying both hypersensitivity when the gene was knocked out, and conferring resistance to bromoacetaete when the protein was overexpressed. YliJ was found to catalyze the dehalogenation of bromoacetate with a k_{cat} 24 ± 1 s^{-1} and a k_{cat}/K_m = 5.4 x 10^3 $M^{-1}s^{-1}$, ($K_m^{bromoacetate}$ = 5 ± 0.3 mM). The substrate specificity of the dehalogenation reaction with YliJ was investigated with chloroacetate, iodoacetate, bromoacetamide, 2-bromopropionate, and 3-bromopropionate. Only iodoacetate (in addition to bromoacetate), was reactive with YliJ. The physiological substrate of YliJ was hypothesized to be a small-molecule containing a carboxylate moiety [64].

The *E. coli* proteins YfcF and YfcG have been reported to display very low GSH-conjugation activity with CDNB and tepid GSH-dependent peroxidase activity towards cumene hydroperoxide [63]. Chapter II describes the functional and structural properties of YfcG as an efficient and unique disulfide-bond oxidoreductase.

Thiol Substrates in *Escherichia coli*

Of central importance to the assignment of enzyme function in the GSH transferases family from *E. coli* is whether they utilize or interact with a thiol substrate. Two such thiol substrates occur in *E. coli*, GSH and glutathionylspermidine (GSP). GSH is synthesized by the consecutive action of two enzymes, γ-glutamyl-cysteine synthetase and GSH synthetase [65]. GSH synthesis is affected by the availability of cysteine and the activity of the γ-glutamyl-cysteine synthetase enzyme, the rate-limiting step [66]. The intracellular GSH concentration in *E. coli* is ~ 17 mM [67].

GSP was first isolated and characterized by Tabor and Tabor in 1975 [68]. Its synthesis is controlled by the bifunctional enzyme glutathionylspermidine synthetase/amidase (GspSA) (Figure I.8). The N-terminal domain of the GspSA protein catalyzes the ATP-dependent condensation of the glycyl carboxylate of GSH with the N1 position of spermidine. The C-terminal amidase domain catalyzes the hydrolysis of the GSP product [69,70]. GSP is produced in late stationary phase under anaerobic conditions and constitutes the primary thiol under those conditions [71]. This leads to a 'thiol switch' in which GSH levels decrease and GSP levels increase during anaerobic growth. A recent report describes a mechanism of regulation of the GspSA protein under conditions of oxidative stress [72]. Through chemical modification and mass spectrometry studies, an active site cysteine in the amidase domain of the GspSA protein was identified as being oxidized to sulfenic acid when *E. coli* is exposed to oxidative stress. Modification of this amino acids allows accumulation of GSP by inactivation of the amidase activity of the protein. Upon elimination of the oxidative stress, the amidase activity can be restored by reaction of the sulfenic acid with a reducing agent (GSH or GSP). The resulting mixed

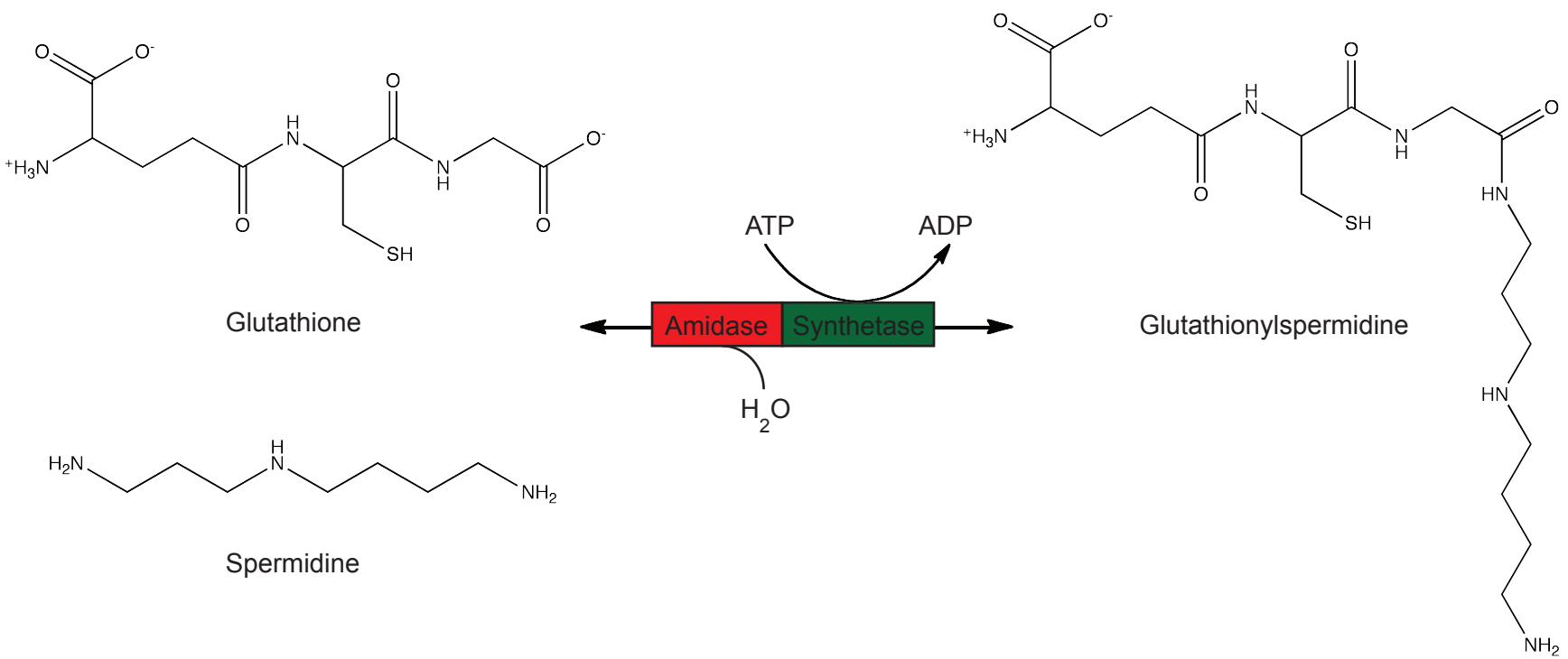


Figure I-8. Reactions catalyzed by the bifunctional enzyme glutathionylspermidine synthetase/amidase (GspSA).

disulfide at Cys-59 would then need to be reduced through another thiol exchange reaction to fully regenerate the amidase activity of the protein. While the exact biochemical role of GSP is still not understood, proteins with mixed disulfides of GSP and protein thiols have been observed under oxidizing conditions in *E. coli* [72]. GSP-modified proteins may represent a protective mechanism for protein thiols under oxidizing conditions or a novel post-translational modification.

Several years ago we noted that the *yghU* gene (an *E. coli* GSH transferases) is located adjacent to the *gss* gene [58,73]. We hypothesized that YghU might regulate GspSA or in some way interact with GSP. However, a number of experiments using purified GspSA and YghU have failed to demonstrate a role for YghU in the utilization, synthesis or degradation of GSP. Chapter IV describes the functional and structural properties of YghU, as a disulfide-bond oxidoreductase related to YfcG.

About The Dissertation

Genome sequencing projects have revealed that GSH transferases are widely distributed in bacteria but most remain only as annotations in sequenced genomes. In a 1997 review, Stéphane Vuilleumier posed the question, "Bacterial glutathione transferases: what are they good for?" [56]. Now, 14 years later, we are just beginning to appreciate the breath of functions in bacterial GSH transferases. The goal of this project was to assign function to members of the GSH transferases homologs from *E. coli* K-12. Several strategies are required for the accurate assignment of function to proteins of unknown function. These approaches include; 1) analysis using informatics and sequence similarity, 2) co-localization of genes providing operon/metabolic context, 3)

transcriptional analysis, 4) phenotypic response to gene knockouts, 5) structural biology, 6) and functional assays of the protein [73,105]. The work described in this report has focused on functional and structural studies of three GSH transferase homologs from *E. coli*, YfcG, YghU and YqjG.

CHAPTER II

ANALYSIS OF THE STRUCTURE AND FUNCTION OF YfcG FROM *Escherichia coli* REVEALS AN EFFICIENT AND UNIQUE DISULFIDE-BOND REDUCTASE

Abstract

YfcG is one of nine GSH transferases homologues encoded in the *E. coli* genome. The protein exhibits robust disulfide-bond oxidoreductase activity toward 2-hydroxyethyl disulfide but little to no GSH transferases activity with various electrophilic substrates. The crystal structure of YfcG was determined to 2.3 Å, and a molecule of glutathione disulfide (GSSG) was observed in the active site of the enzyme even though the crystals were grown in the presence of GSH. The protein is unlike other reductases in that it lacks cysteine residues in the active site. Based on these biochemical and structural results we suggest that the reductase activity of YfcG is unique in that no sulfhydryl groups in the YfcG protein are covalently involved in redox chemistry.

Introduction

The well-studied bacterium, *E. coli* K-12 possesses nine cytosolic GSH-transferases. GSH transferases typically catalyze the addition of GSH to molecules with electrophilic functional groups, thereby playing important roles in cellular detoxication of endogenously produced and exogenous xenobiotic compounds[9]. However, members of this diverse superfamily can also function as peroxidases, isomerases or have non-catalytic roles in ion transport, and the regulation of transcription.[27,59,61,62,74-76]. Most of the *E. coli* GSH transferases proteins have not been characterized biochemically and

structurally. Thus, the biological functions of the proteins have not been established. The YfcG protein has been reported to possess GSH transferases and GSH-dependent peroxidase activities and be involved in bacterial resistance to hydrogen peroxide [63]. However, it remains unclear if these weak catalytic activities can truly be related to resistance to oxidative stress in the bacteria. In this report, the three dimensional structure of YfcG, from *E. coli*, bound to GSSG is described. Compelling biochemical evidence suggests that this structure represents an oxidized version of an efficient disulfide-bond oxidoreductase.

Materials And Methods

Materials

E. coli genomic DNA was from ATCC (Manassas, VA). Primers were custom ordered from Invitrogen (Carlsbad, CA). Restriction enzymes were from New England Biolabs (Ipswich, MA). The pET20b(+) vector and monobromobiamine was from EMB Chemicals, Inc (Gibbstown, NJ). BL21-Gold (DE3) cells were from Stratagene (La Jolla, CA). Luria broth, ampicillin, IPTG, HEPES, DTT, streptomycin sulfate, sodium chloride, MES, glycerol, ammonium sulfate, and GSH were all from Research Products International (Mt. Prospect, IL). EDTA, ammonium acetate, DTNB, GSSG, CDNB, ethacrynic acid, trans-4-phenyl-3-butene-2-one, 1,2-epoxy-3-(4-nitrophenoxy) propane, 1,4-dichloro-2-nitrobenzene, *tert*-Butyl hydroperoxide, cumene hydroperoxide, benzoyl peroxide, lauroyl peroxide and 2-hydroxyethyl disulfide were ordered from Sigma Aldrich (St. Louis, MO). DEAE-Sepharose and SP-Sepharose resins were obtained from GE Healthcare (Piscataway, NJ). The hydroxyapatite resin and 15% w/v SDS-PAGE gels

were obtained from Biorad (Hercules, CA). Potassium phosphate (monobasic), magnesium chloride, hydrogen peroxide and acteonitrile were all from Fisher Scientific (Pittsburgh, PA). The spermidine and *Saccharomyces cerevisiae* GSH reductase were obtained from VWR International (West Chester, PA).

Methods

Cloning, Site-directed Mutagenesis, Protein Expression and Purification of YfcG

The *yfcG* gene was amplified from *Escherichia coli* K-12 genomic DNA (ATCC, Manassas, VA) using forward primer AAAAAAACATATGATCGATCTCTATTTCGCCCCGAC containing the restriction site *Nde*1 and reverse primer TTTTAAAGCTTAACTATCCGAACGCTCATCACCGAG containing the restriction site *Hin*dIII. The amplified and digested *yfcG* gene was subcloned into the pET20b(+) expression vector. The pET20b(+) plasmid containing the *yfcG* gene was transformed into BL21(DE3) cells and grown overnight at 37°C on LB agar containing ampicillin. A single colony was inoculated in liquid LB media supplemented with 0.1 mg/mL ampicillin and grown overnight at 37°C with shaking at 225 rpm. Cells were diluted 100 times in fresh LB media containing 0.1 mg/mL ampicillin and grown under the same conditions until an OD_{600} of 0.8 was reached. The cultures were induced with 0.5 mM IPTG, and the temperature was reduced to 18°C and grown an additional 18 hours. Cells were harvested by centrifugation at 4,000 x *g* for 10 minutes at 4°C. The liquid media was decanted and the pellets frozen at -20°C. The frozen cell pellet was resuspended in 50 mM HEPES buffer containing 1 mM EDTA and 1 mM DTT at pH 7.0 and lysed by

sonication (4 cycles at 70% duty cycle output for 3 minutes with 7 minutes of gentle stirring between each cycle). The cellular debris was removed by centrifugation at 31,000 x *g* for 30 minutes at 4°C. The supernatant was removed and treated with 10 mg/mL streptomycin sulfate, then centrifuged again at 31,000 x *g* for 30 minutes at 4°C. The supernatant was dialyzed overnight in 20 mM HEPES buffer containing 1 mM EDTA and 1 mM DTT at pH 7.0. The protein was loaded on a DEAE-Sepharose column equilibrated in the dialysis buffer and eluted from the column with a linear gradient of 0 to 0.4 M NaCl in the same buffer. Protein containing fractions, determined using A_{280} and SDS-PAGE, were combined and dialyzed overnight against 20 mM MES buffer containing 1 mM EDTA and 1 mM DTT, pH 6.0. The protein was loaded on a SP-Sepharose column (equilibrated with the dialysis buffer) and eluted from the column with a linear gradient of 0 to 0.4 M NaCl in the same buffer. Protein containing fractions were determined by the same procedure and combined and dialyzed overnight against 20 mM KH_2PO_4 buffer containing 1 mM EDTA and 1 mM DTT, pH 7.0. For final purification, the protein was loaded onto a HA column preequilibrated in the dialysis buffer and fractions were collected immediately. The YfcG protein eluted in the flow-through fractions and those fractions containing protein were pooled, and dialyzed in 20 mM KH_2PO_4 buffer containing 1 mM EDTA and 1 mM DTT, pH 7.0. For protein used in the HED assay, the final dialysis buffer did not contain DTT. To prevent protein precipitation in the absence of DTT, the final dialysis buffer contained 300 mM NaCl and 10% w/v glycerol in addition to KH_2PO_4 and EDTA. The purified protein (estimated to be ≥99% pure by SDS-PAGE) was concentrated to ~ 5 mL (70 mg/mL) under nitrogen using an Amicon stirred cell and a YM10 membrane.

Preparation of the C59A Mutant of Glutathionylspermidine Synthetase/Amidase

GSP was synthesized and purified using a procedure similar to that described by Bollinger J.M.[70]. The *gss* gene was subcloned from genomic *E. coli* DNA and into the pET20b(+) expression vector as described for *yfcG*. The following primers and restriction sites were used: forward primer, AAAAAAGAATTCATGAGCAAAGGAACGACCAGC (*Eco*RI) and reverse primer, TTTAAGCTTACTTTTTCACCACAATTAACGGTT (*Hin*dIII). To prevent hydrolysis of the conjugate by the enzyme, Cys-59 was mutated to alanine with the QuikChange Site-Directed Mutagenesis Kit from Stratagene according to the manufacturer's protocol. The following primers were used, GGC CAC AAG TGG CAA **TCC** GTT GAA TTT GCT CGC CG (forward) and CGG CGA GCA AAT TCA AC**G GA**T TGC CAC TTG TGG CC (reverse). The pET20b(+) plasmid containing the *gss* gene with the cysteine mutation was transformed in BL21(DE3) cells and expressed per the procedure described for YfcG. After harvesting, the cells were sonicated and a streptomycin sulfate precipitation was performed. An ammonium sulfate precipitation at 65% saturation was also performed. The pellet containing the precipitated proteins was collected by centrifugation at 22,100 x *g* for 40 minutes at 4°C. The pellet was dissolved in 20 mM HEPES buffer containing 1 mM EDTA and 1 mM DTT at pH 7.0 and dialyzed against the same buffer overnight. The protein was partially purified with a DEAE-Sepharose column equilibrated in the dialysis buffer. The GspSA (C59A) protein was eluted from the column with a linear gradient of 0 to 0.4 M NaCl in the same buffer. Fractions containing the partially purified protein were combined and concentrated.

Enzymatic Synthesis of Glutathionylspermidine

For the GSP synthesis, 5 mM GSH, 10 mM spermidine, 15 mM $MgCl_2$ and 30 μM GspSA (C59A) in 100 mM KH_2PO_4 buffer at pH 6.5 were combined in a 30 mL reaction. The reaction was initiated by the addition of 15 mM ATP. The reaction ran overnight and was terminated by heating the mixture to 95°C for 3 min. The precipitate was removed by centrifugation and the supernatant was loaded onto a SP-Sepharose column equilibrated with 50 mM KH_2PO_4, pH 7.0. The column was washed with the equilibration buffer and the protein was eluted by a linear gradient (0-300 mM) of ammonium acetate, pH 5.2. A small volume of each fraction was removed and mixed with 5 mM DTNB to check for the presence of a thiol (as indicated by a yellow color change). A 5 μL aliquot of the fractions containing thiols was derivatized with monobromobimane and analyzed by reverse phase HPLC with a Beckman Ultrasphere C18 column (4.6 mm x 25 cm). The mobile phase was (A) acetonitrile and (B) 140 mM ammonium acetate (pH 5.0) with a 15-25% gradient over 20 min. Peaks were identified by the retention times of GSP and GSH standards at 3.9 min and 14.5 min respectively. Fractions containing only GSP were pooled, lyophilized and stored at -20°C.

Preparation of Glutathionylspermidine-disulfide

A solution of 20 μmol GSP [8.7 mg in 10 mL of water (pH 7.0)] was mixed with 20 μmol of diamide (3.4 mg in 10 mL of water) in the manner described by Kosower and Kosower [77]. The reaction was allowed to proceed for 3 min at room temperature. The reagent was removed by several extractions with dichloromethane (1 volume of CH_2Cl_2/3

volume of reaction solution). After reaction with diamide no remaining GSP could be detected upon reaction with 5,5'-dithiobis-(2-nitrobenzoic acid) DTNB. The complete conversion of GSP to GSP-disulfide was confirmed by reduction of GSP-disulfide back to GSP in the presence of sodium borohydride. A small volume of GSP-disulfide was incubated with a 1.5 molar excess of sodium borohydride at pH 9.0. The reaction was allowed to proceed for 3 min before the pH was adjusted to 6.0, to decompose the excess $NaBH_4$. The concentration of GSP in the final stock solution was measured by titration with DTNB. GSP concentration for the reduced GSP-disulfide agreed with the concentration of GSP used in the initial diamide oxidation within 5%.

Crystallization and Determination of the Structure of YfcG

Crystallization of YfcG

Crystals were grown by hanging-drop vapor diffusion at room temperature (approximately 20°C). The crystallization buffer in the well contained 20% w/v PEG 3000 and 100 mM acetate at pH 4.5. The 10 μL drop contained a 50:50 mixture of the crystallization buffer and a solution of 33 mg/mL YfcG and 10 mM GSH (the final drop concentration of YfcG was 16.5 mg/mL and 5 mM of GSH). Crystals grew in about 2 weeks.

X-Ray Data Collection and Processing

Diffraction data were collected at the Vanderbilt University Center for Structural Biology Biomolecular Crystallography Facility. The cryoprotectant used was the crystallization buffer supplemented with 30% glycerol. Crystals were screened initially

using an Xcalibur PX2 Ultra sealed tube X-ray generator with an ONYX CCD area detector (Oxford Diffraction) and Oxford Cryojet (Oxford Instruments) cryostat to maintain crystals at 100K. Complete data sets were acquired using a Bruker Microstar rotating-anode X-ray generator and a Bruker Proteum PT135 CCD area detector. Crystals were maintained at 100K using a Bruker Kryo-Flex cryostat. Data collection sweeps were optimized using Cosmo (Bruker AXS, 2008) software and data integrated and scaled using SADABS (Bruker AXS, 2008) and XPREP (Bruker AXS, 2008) in the PROTEUM2 (Bruker AXS, 2008) package. X-ray data collection statistics are presented in Table II-1.

Structure Determination

The structure was determined using the molecular replacement method. The Fold and Function Assignment System web server [78] was used to search for known structures to use as the molecular replacement model. The structure (PDB ID 1PN9) gave the best score and was chosen as the model for the molecular replacement. The Matthews coefficient for one polypeptide chain in the asymmetric unit is 2.6 $Å^3$/Da. The program PHASER [79] was used to place the model in the asymmetric unit and calculate the initial phases. Using the iterative script, (www.solve.lanl.gov) which includes cycles of density modification and automated model building with RESOLVE [80,81] and molecular refinement with REFMAC5 [82], 160 of the 215 residues were built and 119 side chains were placed. The model was completed by cycles of examining and building in the graphics display program COOT [83], and refining in REFMAC5. The final model included 205 residues, one GSSG ligand, and 93 water molecules. Ten residues at the C-terminus

Parameters	**Values**
PDB file name	3GX0
Space group	$P4_12_12$
Unit cell dimensions (Å)	68.49, 68.49, 111.99
Resolution limits (Å)	18.1-2.3 (2.4-2.3)
Number of independent reflections	12,571 (1,462)
Completeness (%)	99.8 (98.5)
Redundancy	10.9 (6.8)
Average I/average σ(I)	3.35
R_{sym} (%)	13.0 (55.1)
R-factor (overall) (%) / Number of reflections	19.4 (25.1) / 11,342
R-factor (working) (%) / Number of reflections	19.1 (24.9) / 10,758
R-factor (free) (%) / Number of reflections	25.0 (35.7) / 584
Number of protein atoms	1,658
Number of hetero-atoms[a]	133
Average B Values	
Protein atoms (Å^2) 1658	27.9
Glutathione disulfide (Å^2) 40	30.7
Solvent (Å^2) 93	26.9
Weighted root-mean-square deviation of ideality	
Bond lengths (Å)	0.017
Bond angles (deg)	1.663
General planes (Å)	0.0070

Table II-1. X-Ray Data Collection and Least-Squares Refinement Statistics for the Structure of YfcG. Values for the highest resolution bin are given in parentheses.
[a]These include 1 glutathione disulfide molecule and 92 water molecules.

are not seen in the electron density map. The model was checked with the validation tools inside COOT and the Molprobity program [84]. The Ramachandran plot highlighted only one outlier, E71. The final refinement statistics are shown in Table II-1.

Steady State Fluorescence Titration

Equilibrium dissociation constants were determined for the complex of YfcG with GSH, glutathione sulfonate (GSO_3^-), GSP, GSH-disulfide and GSP-disulfide using fluorescence titration. Measurements were made on a SPEX Fluorolog-3 spectrofluorimeter in the constant wavelength mode. The samples were excited at 295 nm and the emission was collected at 340 nm over 30 seconds. The change in fluorescence intensity of 1 μM YfcG in 100 mM KH_2PO_4 buffer, at pH 7.0, was measured after 5 min preincubation with each addition of thiol. Fluorescence data for each concentration of thiol were collected three times and averaged. Three independent titrations were completed for each of the five thiol substrates and the results were averaged. Each data set was corrected for dilution caused by the addition of ligand. Inner filter effects were corrected for by calculating a corrected fluorescence, F_c.

$$F_c=F*\text{antilog}[(A_{295}+A_{340})/2] \quad (1)$$

The data were fit to a one-site binding model according to Equation 2, where Y = change in fluorescence and X = concentration of thiol in millimolar.

$$Y = Y_{max}/[1+(K_D/X)] \quad (2)$$

Steady State Enzyme Kinetics

Glutathione Transferase Activity Assays

The GSH transferase activity of YfcG was measured with a variety of electrophilic substrates including CDNB, ethacrynic acid, trans-4-phenyl-3-butene-2-one, 1,2-epoxy-3-(4-nitrophenoxy) propane and 1,4-dichloro-2-nitrobenzene. Measurable activity was only detected for the CDNB substrate. Reactions were followed at 340 nm using a Perkin-Elmer Lambda 45 Spectrophotometer. In each reaction 1 μM YfcG was pre-incubated with varying concentrations of GSH (0.5-25 mM) for 5 minutes in 100 mM KH_2PO_4 buffer, pH 6.5. The reactions were started by addition of CDNB in acetonitrile (1 mM final concentration). All measurements were made at 25°C. Initial rate measurements for the production of the CDNB-GSH conjugate were calculated using the slope of the increase in absorbance at 340 nm ($\varepsilon_{GS\text{-}DNB}$= 9,600 $M^{-1}cm^{-1}$). The data were corrected for the spontaneous GSH-CDNB reaction. All measurements were made in triplicate and averaged. The data were fit to the Michaelis-Menten equation (Equation 3):

$$Y = \frac{(V_{max} \cdot X)}{(K_m + X)} \qquad (3)$$

where X is the substrate concentration and Y is the enzyme velocity. Values of V_{max} and K_m were determined using equation 3 and GraphPad Prism software (La Jolla, CA). The k_{cat} value was determined using equation 4.

$$V_{max} = k_{cat}\,[E] \qquad (4)$$

Glutathione-dependent Peroxidase Activity Assays

The GSH-dependent peroxidase activity of YfcG was measured using a coupled assay with GSH reductase similar to the method described in Flohe and Gunzler[85]. A variety of peroxides including hydrogen peroxide, benzoyl peroxide, lauroyl peroxide, cumene hydroperoxide, and *tert*-butyl hydroperoxide were tested. Activity was only detected with the cumene hydroperoxide substrate; however, the background reaction was significant (73-90% of enzyme catalyzed reaction). Reactions were followed at 340 nm using a Perkin-Elmer Lambda 45 Spectrophotometer. Each reaction contained 1 μM YfcG, varying concentrations of GSH (0.1-3.0 mM), 1 mM EDTA, 0.15 mM NADPH, 1 unit (6 μg/mL) GSH reductase in 100 mM KH_2PO_4 buffer, pH 7.0. The reactions were started by addition of 1.2 mM cumene hydroperoxide. All measurements were made at 37°C. The data were corrected for the spontaneous reduction of cumene hydroperoxide by GSH. The decrease in absorbance at 340 nm was monitored over 3 minutes (ε_{NADPH} = 6,220 $M^{-1}cm^{-1}$) and the slope of this line was used to calculate the initial rate. The data were fit to the Michaelis-Menten equation 3 using GraphPad Prism and values of V_{max} and K_m were determined. The k_{cat} value was determined using equation 4.

Disulfide-bond Oxidoreductase Activity Assays

The disulfide-bond oxidoreductase activity of YfcG was measured using a coupled assay with GSH reductase similar to the method utilized by Vlamis-Gardikas[86]. Briefly, reactions containing 2 mM EDTA, 0.2 mM NADPH, 6 μg/mL (1 unit) GSH reductase, 0.1 μM YfcG and varying concentrations of GSH (0.1 to 4 mM) were prepared in 100 mM Tris-Cl, pH 8.0. The reaction was started by addition of 1 mM 2-hydroxyethyl

disulfide. All measurements were made at 25°C. The decrease in absorbance at 340 nm was monitored over 4 minutes (ε_{NADPH} = 6,220 $M^{-1}cm^{-1}$) and the slope of this line was used to calculate the initial rate. All measurements were done in duplicate and averaged. The data was fit to the Michaelis-Menten equation 3 using GraphPad Prism and values of V_{max} and K_m were determined. The k_{cat} value was determined using equation 4.

Results And Discussion

Substrate and Ligand Binding and Characterization of the Enzyme Activity of YfcG

The dissociation constants for various thiol and disulfide ligands and substrates were determined using fluorescence titration, by monitoring the change in intrinsic fluorescence of the protein upon ligand binding (Table II-2). The *yfcG* gene is expressed in late stationary phase[73], where a significant fraction of GSH is found as the GSP conjugate[71]. GSP was produced and purified as described in the materials and methods section and used as a ligand in binding titrations and a substrate in enzyme assays. Interestingly, GSP binds YfcG 10x more tightly than GSH (K_d^{GSH} = 330 ± 10 μM, K_d^{GSP} = 29 ± 7 μM). However, no thiol transferase or peroxidase activity was observed when GSP was used as a nucleophile. Despite its initial characterization by Tabor and Tabor over 36 years ago, the physiological role of GSP in *E. coli* remains largely unknown[68]. A recent report, has described the formation of GSP mixed disulfides with protein thiols *in vivo* and a mechanism of regulation of glutathionylspermidine synthetase/amidase under oxidative stress[72]. In our studies, we have not seen catalytic activity with YfcG and GSP as a thiol nucleophile. YfcG was crystallized in the presence of GSP and the structure refines (at a resolution of 2.3 Å), with respectable R and R_{free} values (R = 0.205 and R_{free}

Compound	K_d (μM)
GSH	330 ± 10
GSO_3^-	23 ± 5
GSP	29 ± 7
GSP-disulfide	180 ± 40
GSSG	2.4 ± 0.5

Table II-2. Dissociation constants of thiol and disulfide ligands and substrates as determined by fluorescence titration at 25°C. (GSH, glutathione; GSO_3^-, glutathione sulfonate; GSP, glutathionylspermidine; GSP-disulfide, glutathionylspermidine disulfide; GSSG, glutathione disulfide)

= 0.233). However, the protein structure is less well ordered than the YfcG·GSSG structure, particularly in the active site. No electron density is observed for residues 32-50 and 110-133 or the GSP molecule. The reason for the disorder is not known. Because of a lack of functional and structural data, we cannot currently explain the tighter binding of GSP (vs. GSH) to YfcG. The role of YfcG in the biochemistry of GSP is under active investigation.

The most interesting result of the binding titration studies is that GSSG binds YfcG 100x more tightly than GSH (K_d^{GSSG} = 2.4 ± 0.5 μM). GSSG is the tightest binding substrate of those tested which may indicate a functional role for GSSG in the active site of the enzyme. This is especially significant because a molecule of GSSG was observed in the active site of the enzyme in the crystal structure, even though the crystals were grown in the presence of GSH (see next section).

The thiol transferase activity of YfcG was tested against a panel of electrophilic substrates using both GSH and GSP as thiol nucleophiles. Of the electrophilic substrates surveyed (CNDB, ethacrynic acid, trans-4-phenyl-3-butene-2-one, 1,2-epoxy-3-(4-nitrophenoxy)propane and 1,4-dichloro-2-nitrobenzene), only CDNB was a substrate for a GSH transferase reaction (k_{cat} = 0.1 ± 0.01 s^{-1} and k_{cat}/K_m^{GSH} = 29 ± 5 $M^{-1}s^{-1}$). This very low activity towards a typical GSH transferase substrate, CDNB, may indicate that YfcG does not play a role in cellular detoxication reactions in *E. coli*.

The YfcG protein has been reported by others to have low (but previously not quantified) GSH-dependent peroxidase activity [63]. We investigated this activity with a panel of substrates including H_2O_2, t-butyl hydroperoxide, cumeme hydroperoxide, benzyl peroxide and lauryl peroxide. YfcG exhibited weak GSH-dependent peroxidase

activity with cumene hydroperoxide (k_{cat} = 0.27 ± 0.03 s^{-1} and k_{cat}/K_m^{GSH} = 520 ± 200 $M^{-1}s^{-1}$). No activity was observed toward any other peroxide substrate. The potential role of YfcG in resistance to peroxides and defense against oxidative stress requires further investigation but our experiments do not indicate that YfcG has a significant GSH-dependent peroxidase activity.

Observation of GSSG in the active site of the enzyme in the crystal structure suggests that YfcG might be involved in disulfide-bond oxidoreductase chemistry. The disulfide-bond reductase activity of the protein was measured with 2-hydroxyethyl disulfide using a coupled assay. YfcG is a very efficient disulfide reductase toward this model substrate, with activity on par with glutaredoxins (Table II-3).

Structural Characterization of YfcG

The structure of YfcG was determined to 2.3 Å. YfcG was crystallized in the presence of GSH, 20% w/v PEG 3000, and 0.1 M sodium acetate (pH 4.5). A ribbon representation of the structure of the dimer of YfcG is shown in Figure II-1. The structure has a typical GSH transferase fold with an N-terminal thioredoxin-like domain and a C-terminal all α-helical domain. The functional homodimer is formed via primarily hydrophobic interactions along two helices, one from the thioredoxin-like domain and one from the all α-helical domain.

Remarkably, a molecule of GSSG is found in the active site even though the protein was crystallized in the presence of GSH. The electron density for GSSG in the structure is shown in Figure II-2. The disulfide bond of GSSG is located on the surface of

(A)

Enzyme	k_{cat} (s^{-1})	k_{cat}/K_M^{GSH} ($M^{-1}s^{-1}$)	K_m^{GSH} (mM)
YfcG	29 ± 2	(1.8 ± 0.3) x 10^4	1.6 ± 0.3
YfcG-C166A	30 ± 2	(3.0 ± 0.7) x 10^4	1.0 ± 0.2

(B)

Enzyme	k_{cat} (s^{-1})	k_{cat}/K_M^{GSH} ($M^{-1}s^{-1}$)	K_m^{GSH} (mM)
Grx1	115 ± 16	3.8 x 10^4	3.0 ± 0.7
Grx2	554 ± 28	3.1 x 10^5	1.8 ± 0.2
Grx3	182 ± 20	7.3 x 10^4	2.5 ± 0.5
hGrx	320 ± 13	1.7 x 10^5	1.9 ± 0.2

Table II-3. (A) Steady state kinetic constants for the reduction of 2-hydroxyethyldisulfide (1.0 mM) by YfcG and its C166A mutant, compared to **(B)** glutaredoxin 1, 2, 3 from *Escherichia coli* and human glutaredoxin. (Values in B taken from reference 86).

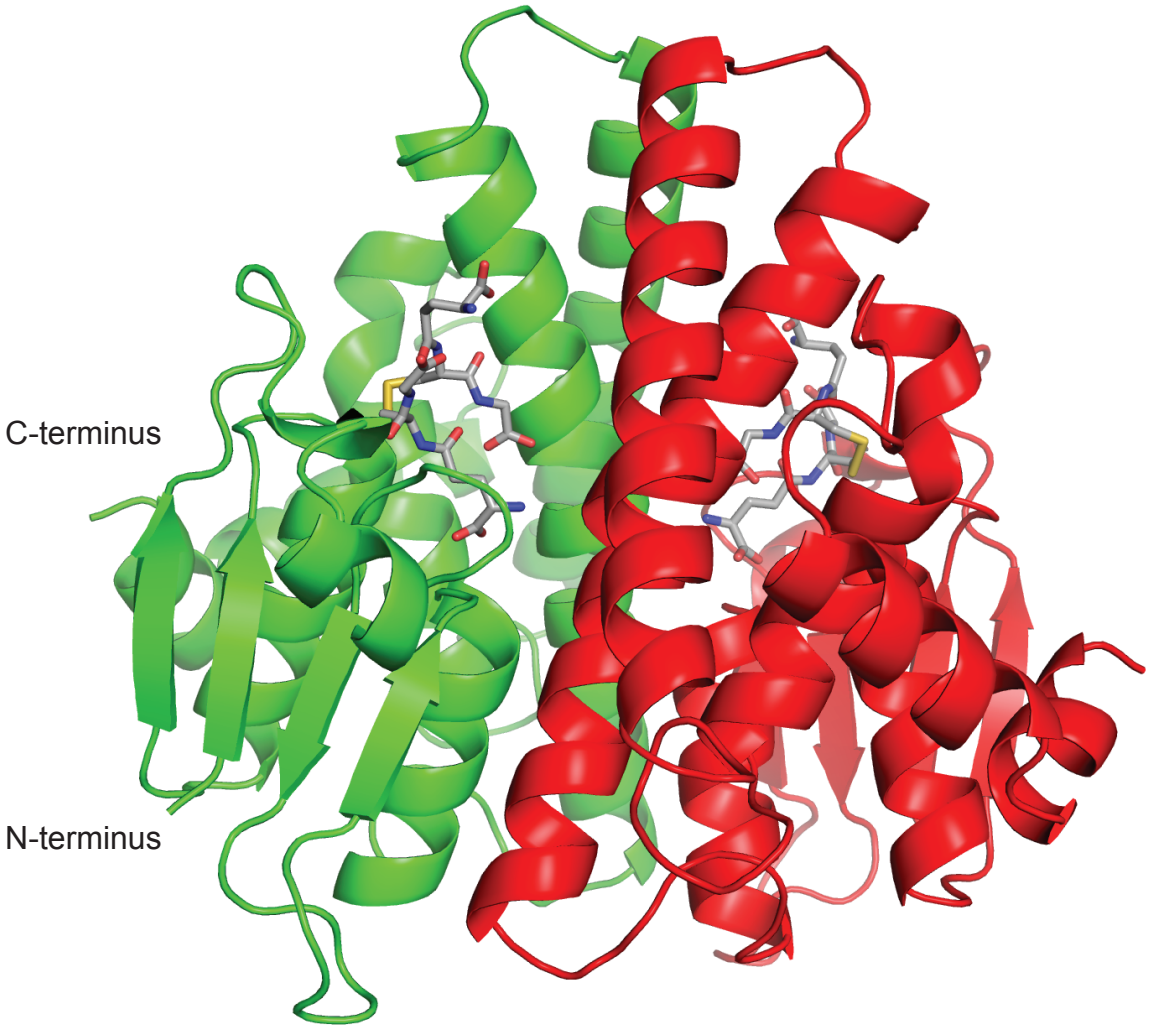


Figure II-1. Ribbon diagram of the structure of the dimer of YfcG with GSSG bound in the active sites. The two subunits are shown in red and green, respectively, and the two molecules of GSSG are shown in stick representation.

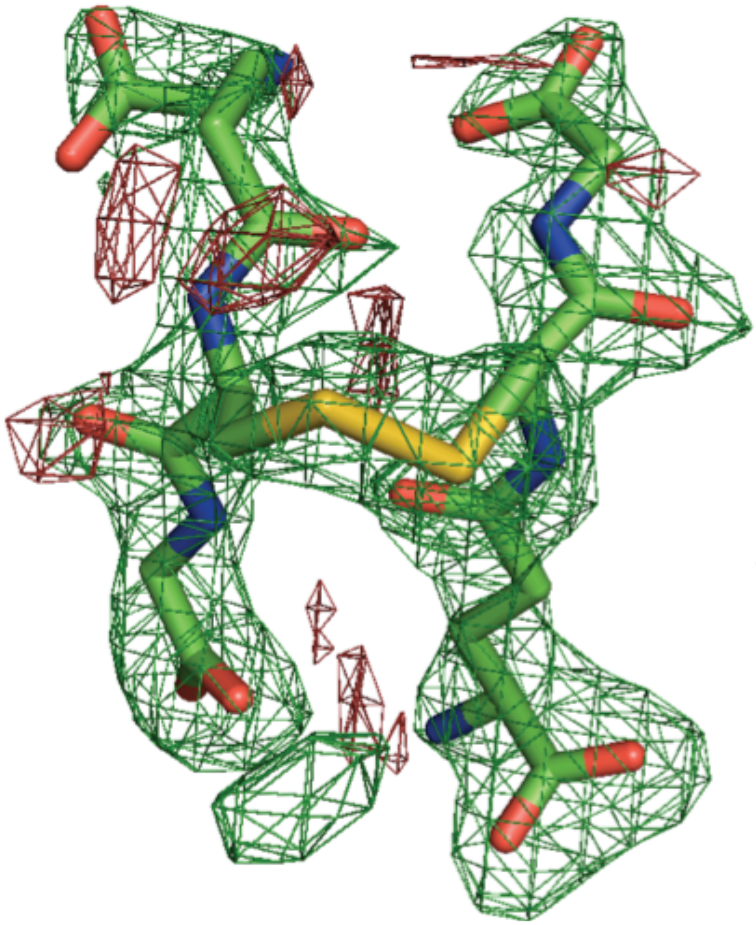


Figure II-2. Omit map of the electron density for the molecule of GSSG observed in the active site of YfcG. The map, contoured at 3σ, was calculated with coefficients of the form $F_O - F_C$ in which the observed and calculated structure factor amplitudes were calculated from the model lacking the coordinates of the GSSG molecule. The final model is shown in stick representation. The disulfide bond is shown in yellow. The average B factor for the GSSG atoms in the final refined structure is 30.7 $Å^2$.

the protein (Figure II-3). This crystallographic observation corresponds to the robust disulfide-bond oxidoreductase activity measured in functional assays.

As displayed in Figure II-4, one half of the GSSG molecule (GS1) makes several hydrogen bonding interactions to the primary subunit. These include; the glycine carboxylate hydrogen bonding to the side chain nitrogen of Gln-38, the cysteine carbonyl oxygen forming hydrogen bonding interactions to the main chain nitrogen of Ile-52, and the γ-glutamyl carbonyl oxygen hydrogen bonding to the side chain nitrogen of Asn-11. Finally, the γ-glutamyl nitrogen of GS1 hydrogen bonds to Glu-71, while Ser-72 forms hydrogen bonding interactions with the γ-glutamyl carboxylate. Glu-71 and Ser-72 are located at the turn between β-strand 4 and α-helix 3 in the thioredoxin domain. This ββα core of the thioredoxin domain is a conserved structural element responsible for the recognition of the γ-glutamyl portion of GSH [9]. The other half of the GSSG molecule (GS2), has few hydrogen bonding interactions with the protein. However, the γ-glutamyl-α-carboxylate of GS2 forms a critical ionic interaction with Arg-132 on α-helix 5 of the opposite subunit.

Typical disulfide-bond oxidoreductases, eg. thioredoxins and glutaredoxins, have one or two catalytic cysteine residues, which form mixed disulfide intermediates with the substrate disulfide. These intermediates are reduced by two molecules of GSH, leading to formation of GSSG. The YfcG protein only has a single cysteine residue that is located >15 Å away from the bound GSSG in the active site. Mutation of this cysteine to alanine did not significantly change the kinetic constants for the disulfide-bond reductase reaction (Table II-3) suggesting the cysteine does not participate in this activity of the protein.

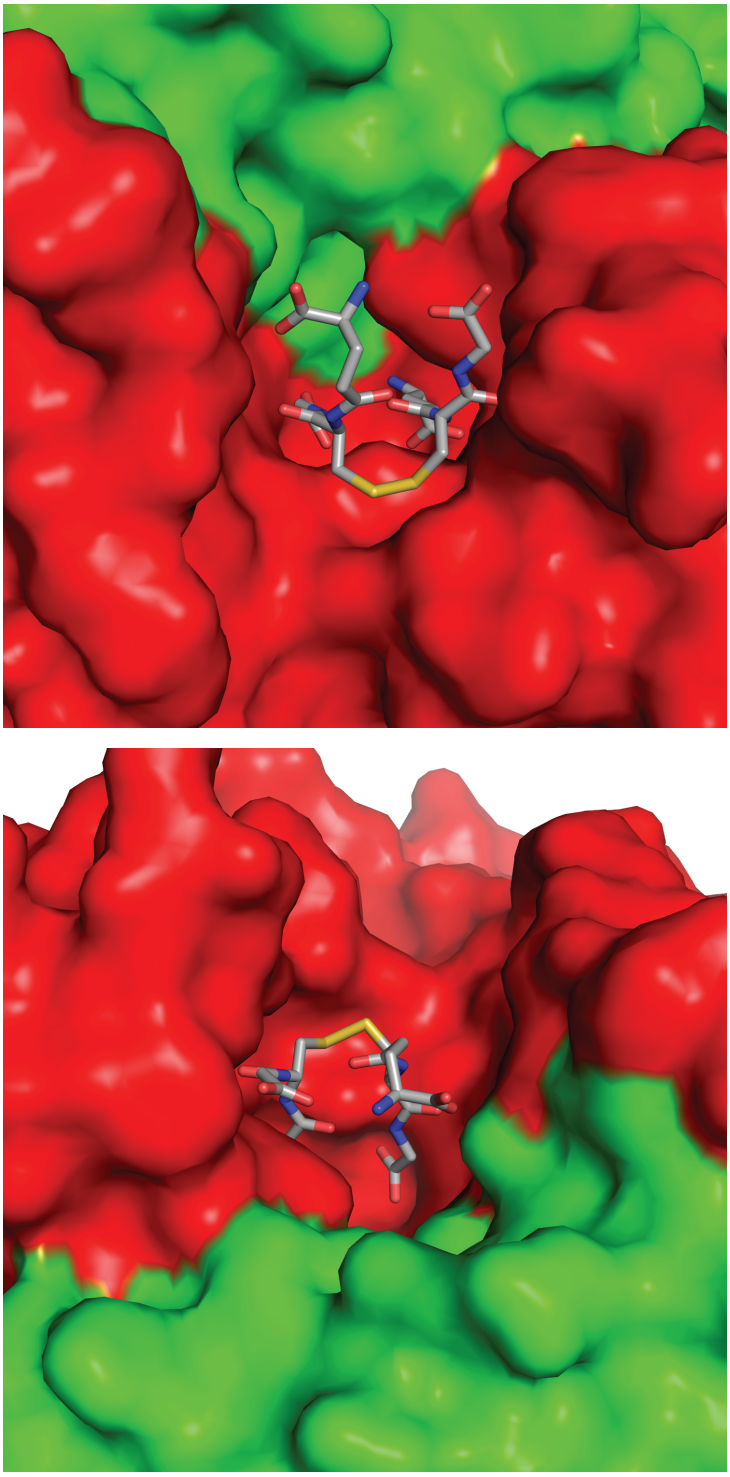


Figure II-3. Space filling representations of the YfcG-GSSG complex illustrating the exposure of the disulfide bond of GSSG on the surface of the protein.

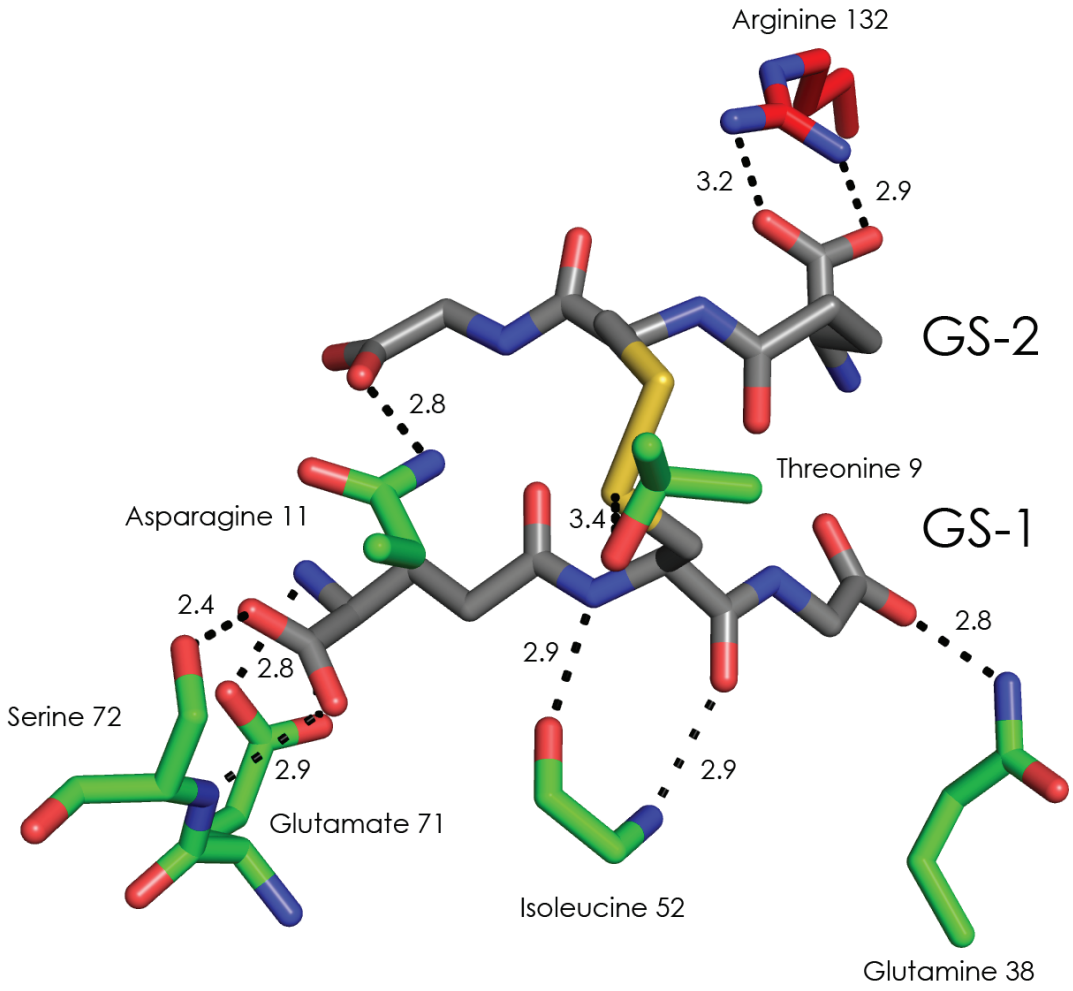


Figure II-4. Details of the GSSG binding site. Subunits one and two are colored in red and green, respectively. Potential hydrogen bonding interactions are shown as dashed lines with distances shown in angstroms (Å).

Many GSH transferases possess an active site tyrosine, serine, or cysteine residue that interacts with the thiol group of GSH. This residue functions to lower the pK_a of GSH from approximately 9 to approximately 6-7 and is thought to be a key component of catalysis. In the YfcG structure, the hydroxyl group of Thr-9 makes hydrogen bonding interactions with the sulfur of GS1. This threonine is located in a similar position to the tyrosine/serine/cysteine found in other GSH transferases, and is located at the amino terminus of α-helix 1 in the thioredoxin domain (Figure II-5). This is the first structurally confirmed case of a threonine used to stabilize the GSH thiolate.

While most disulfide-bond oxidoreductases have one or two protein cysteine residues that participate covalently in the redox chemistry, it is possible for a non-covalently bound thiol disulfide couple, like 2GSH/GSSG, to function in the same manner. YfcG may function as an oxidant or reductant depending on the redox potential of the cellular compartment and the redox potential of the protein-thiol-disulfide couple. YfcG most likely functions in the cytoplasm of the cell, where the GSH/GSSG ratio is typically 200 to 1. Because the protein binds GSSG ≥100 fold more tightly than GSH (Table II-2), and prefers GSSG in the presence of at least 10-fold excess of GSH (crystal structure), the redox potential of GSH on the surface of YfcG may be significantly perturbed. The redox potential of GSH in aqueous solution is estimated to be ~ -250 mV. If the difference in binding energy of GSSG and GSH is expressed as the redox potential of bound GSH, then the redox potential of YfcG·2GSH complex could be as much as -50 mV more reducing (~ -300 mV) than GSH in aqueous solution. These remarkable physical chemical properties of YfcG are consistent with it being a disulfide-bond oxidoreductant in the cytosol of *E. coli*. A possible mechanism for this reaction is

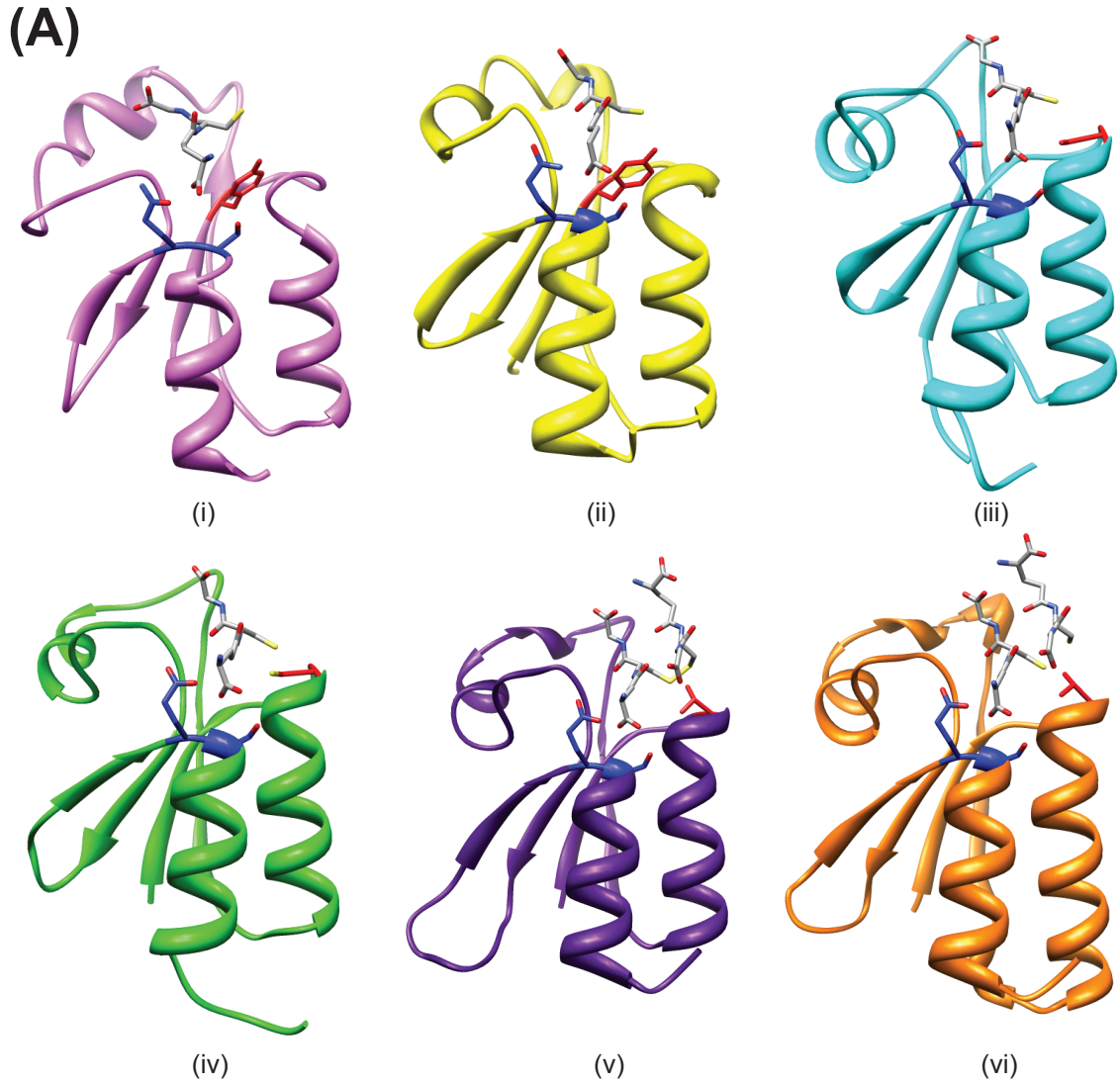


Figure II-5. (A) Thioredoxin domain in six representative GSTs highlighting the conserved QS or ES motif responsible for recognition of the γ-glutamyl portion of GSH (in blue) and the active site residue (in red). **(B)** Structure based sequence alignment of the structures shown in (A). Residues in α-helices are highlighted in yellow and those in β-strands are highlighted in green. [(i) 1FHE, mu class; (ii) 1TU7, pi class; (iii) 1OYJ, tau class; (iv) 1EEM, omega class; (v) 3GX0, nu class; (vi) 3C8E, nu class]

outlined in the next chapter (reference Figure III-6.). However, the actual physiological substrates that are reduced by the enzyme have not been identified.

Conclusions

The structural and functional properties of YfcG, a GSH transferase homolog, from *Escherichia coli* have been investigated in this report. Data from fluorescence titrations show that YfcG binds GSSG 100x more tightly than GSH. In addition, the enzyme is an efficient disulfide-bond oxidoreductase with the substrate 2-hydroxyethyl disulfide. A molecule of GSSG was observed in the active site of the enzyme in the crystal structure even though the crystals were grown in the presence of GSH. The protein does not contain an active site cysteine, which enables catalysis. Instead, an active site threonine residue is within hydrogen bonding distance to GSSG. The functional and structural evidence suggest that YfcG is a unique disulfide-bond oxidoreductase in *E. coli*.

Performing a similarity search against the protein databank using BLAST or the protein database structural server DALI, reveals that the structure of YfcG is very similar to another *E. coli* GSH transferases protein, YghU. In the next chapter we will discuss the functional and structural properties of the YghU protein and its relation to YfcG.

CHAPTER III

STRUCTURE AND FUNCTION OF YghU, A NU-CLASS GLUTATHIONE TRANSFERASE RELATED TO YfcG FROM *Escherichia coli*

Abstract

The structural and biochemical properties of YghU, a GSH transferases homolog from *Escherichia coli* K-12, are investigated in this report. The crystal structure of YghU was determined at 1.5 Å resolution and two molecules of GSH were observed in each active site of the enzyme. This crystallographic observation is consistent with biphasic equilibrium binding data that indicate one tight and one weak binding site for GSH. The protein exhibits little to no GSH transferases activity with typical electrophilic substrates but does possess modest GSH-dependent peroxidase activity with several organic hydroperoxides. Most importantly, the protein exhibits robust disulfide-bond oxidoreductase activity with 2-hydroxyethyl disulfide, on par with that reported for YfcG. In fact, structural superposition of the YfcG•GSSG and YghU•2GSH complexes reveals remarkable similarity, especially with regard to the active site and overlayed GSSG and 2 GSH from each structure. From these observations we conclude that the two structures represent oxidized and reduced forms of GSH-dependent disulfide-bond oxidoreductases. The structural and biochemical properties of YghU and YfcG along with global analysis of the GSH transferases superfamily indicate that they are members of the same, but previously unidentified, subfamily of GSH transferases homologues, which we suggest be called the *nu*-class GSH transferases.

Introduction

GSH transferases typically catalyze the conjugation of the tri-peptide GSH to a variety of electrophilic substrates. In this manner the proteins play important roles in cellular detoxication of endogenous and xenobiotic electrophiles [9]. However, the GSH transferases superfamily is a functionally diverse one, and there are several examples of proteins in this superfamily with distinct functions in redox chemistry, ion transport, and the regulation of transcription[27,59,61,62,74-76]. With the completion of dozens of microbial genome sequencing projects we now understand that bacterial genomes contain multiple GSH transferases genes of widely divergent sequence and unknown functions. The genome of *Escherichia coli* K-12 contains nine cytosolic GSH transferases. It is likely that several of these *E. coli* GSH transferases homologs will have functional roles outside that of typical GSHconjugation chemistry.

Several years ago we noted that the *gss* gene was located next to the *yghU* gene, a GSH transferases homolog, in the *E. coli* genome [58]. The *gss* gene encodes the glutathionylspermidine synthetase/amidase protein (GspSA). This bifunctional enzyme controls the synthesis of GSP by catalyzing the ATP-dependent condensation of the glycl carboxylate of GSH with the N1 position of spermidine at the C-terminal domain, as well as the hydrolysis of the GSP product in the N-terminal amidase domain [69,70]. We hypothesized that the gene context of *gss* and *yghU* could be significant and that the YghU protein might regulate the GspSA enzyme or act as a GSP-transferase [58]. However, further studies have failed to demonstrate a regulatory role for YghU in the synthesis or hydrolysis of GSP under a variety of conditions [73]. In addition, a gene knockout of YghU had no significant effects on GSP levels under anaerobic conditions in *E. coli*. While

YghU does bind GSP with modest affinity, there is currently no evidence of its roles in the utilization, synthesis or degradation of GSP. Recently, a mechanism of regulation of the GspSA protein has been proposed [72]. Through chemical modification and mass spectrometry studies, an active site cysteine in the amidase domain of the GspSA protein was identified as being oxidized to sulfenic acid when *E. coli* is exposed to oxidative stress[72]. Modification of this amino acid allows accumulation of GSP by inactivation of the amidase activity of the protein. Upon elimination of the oxidative stress, the amidase activity can be restored by reaction of the sulfenic acid with a reducing agent (GSH or GSP). The resulting mixed disulfide at Cys-59 would then need to be reduced through another thiol exchange reaction to fully regenerate the amidase activity of the protein.

In this report, we describe the structural and biochemical properties of YghU, which differs from our initial hypothesis. The data presented here indicate that it is a disulfide-bond oxidoreductase closely related to YfcG from *E. coli.*

Materials And Methods

Materials

E. coli K-12 genomic DNA was from ATCC (Manassas, VA). Primers were custom ordered from Invitrogen (Carlsbad, CA). Restriction enzymes were from New England Biolabs (Ipswich, MA). The pET20b(+) vector and B834(DE3) cells were from EMB Chemicals, Inc (Gibbstown, NJ). BL21-Gold (DE3) cells were from Stratagene (La Jolla, CA). Luria broth, ampicillin, NADPH, IPTG, HEPES, DTT, streptomycin sulfate, and MES were all from Research Products International (Mt. Prospect, IL). EDTA, ammonium acetate, GSH, GSSG, CDNB, ethacrynic acid, trans-4-phenyl-3-butene-2-

one, 1,2-epoxy-3-(4-nitrophenoxy) propane, 1,4-dichloro-2-nitrobenzene, tert-Butyl hydroperoxide, cumene hydroperoxide, *S. cerevisiae* GSH reductase, and all amino acids were ordered from Sigma Aldrich (St. Louis, MO). DEAE-Cellulose and SP-Sepharose resins were obtained from GE Healthcare (Piscataway, NJ). Hydrogen peroxide and linoleic acid was from Fisher Scientific (Pittsburgh, PA). The soybean lipoxidase was from VWR (West Chester, PA). The (5Z,8Z,11Z,13E,15S)-15-hydroperoxyicosa-5,8,11,13-tetraenoic acid (15(S)-HpETE) substrate was a generous gift from Dr. Alan Brash (Vanderbilt University).

Methods

Cloning, Expression and Purification of YghU

[The work described in this section was initially performed by Dr. Nina Stourman.]

The *yghU* gene was amplified from *E. coli* K-12 genomic DNA (ATCC, Manassas, VA) using primers containing restriction sites for *NdeI* and *EcoRI*. The forward primer was: AAA AAA TCT AGA CAT **ATG** ACA GAC AAT ACT TAT CAG CCC and the reverse primer: TTT GGA TCC G AA TTC **TTA** CCC CTG ACG CTT ATC TTC CG. The *yghU* gene was digested and ligated into the *pET20b* vector. *E. coli* BL21(DE3) cells were transformed with the *pET20b*(*yghU*) plasmid, a liquid culture of the transformed bacteria was grown overnight from a single colony in LB media containing 100 μg/mL ampicillin at 37˚ C. The cells were diluted 100 times in a fresh media and grown to OD_{600} of about 0.8. Protein production was induced by addition of 0.4 mM IPTG. The cells were grown for 4 hrs at 30˚ C to avoid the formation of inclusion bodies. The cells were harvested by centrifugation at 6,000 x *g* for 10 min and

the pellet was frozen at -20˚ C for storage. The frozen pellet was resuspended in 20 mM HEPES, 1 mM EDTA (pH 7.0), lysed by sonication and centrifuged at 17,000 x *g* for 30 min. The supernatant was treated with streptomycin sulfate (10 mg/mL), centrifuged for 30 min at 17,000 x *g* and dialyzed overnight against 20 mM HEPES, 1 mM EDTA (pH 7.5). The dialyzed protein was loaded on to a DEAE-cellulose column equilibrated with 20 mM HEPES, 1 mM EDTA (pH 7.5) and eluted from the column with a linear gradient of NaCl (0 – 0.4 M) in the same buffer. Fractions containing the protein (determined by A_{280} and SDS-PAGE) were pooled together, dialyzed against 20 mM MES, 1 mM EDTA (pH 5.8) loaded on an SP-Sepharose column equilibrated with 20 mM MES, 1 mM EDTA (pH 5.8) and eluted with the linear gradient of NaCl (0 – 0.4 M) in the same buffer. Fractions containing the protein (determined by A_{280} and SDS-PAGE) were pooled together, dialyzed against 20 mM KH_2PO_4, 1 mM EDTA (pH 7.0), concentrated, and flash frozen.

Expression and purification of the selenomethionine derivative of YghU

[The work described in this section was performed by Dr. Matthew Schaab.]

The selenomethionine derivative was prepared from *E. coli* B834 DE3 cells transformed with the expression vector containing the *yghU* insert. A single colony was used to inoculate a 100 mL culture of M9 media containing 100 µg/mL ampicillin. The culture was grown overnight at 37 °C. A 10 mL aliquot of the overnight culture was used to inoculate a 1 L culture of M9 media supplemented with 30 mg of all common amino acids. The cultures were grown at 37 °C with shaking at 250 rpm. When the OD_{600} reached 0.6, the cells were harvested by centrifugation at 8000 x *g* for 10 minutes. The

cells were resuspended in 1 L of M9 media supplemented with 30 mg of every common amino acid and 50 mg of selenomethionine substituted for methionine. The cells were induced by the addition of 100 mg of IPTG. The cells were grown at 15 °C and 225 rpm for 16 hours. The cells were harvested by centrifugation at 8000 x *g* for 10 minutes. The cell paste was frozen and stored at -80 °C. The selenomethionyl-enzyme was purified as described previously for the native protein.

Steady-State Enzyme Kinetics.

Glutathione Transferase Activity Assays

[The work described in this section was performed by Dr. Nina Stournman.]

The GSH transferase activity of YghU was assayed with several common substrates including CDNB, ethacrynic acid, trans-4-phenyl-3-butene-2-one, 1,2-epoxy-3-(4-nitrophenoxy) propane and 1,4-dichloro-2-nitrobenzene. With these potential substrates, YghU exhibited measurable activity only with CDNB. Reactions with CDNB were followed with a Perkin-Elmer Lambda 45 Spectrophotometer at 25˚ C in 1 mL cuvettes. For each reaction, 0.7 μM YghU was pre-incubated with various concentrations of GSH (0.5 μM to 20 mM) or GSP (8 μM to 1 mM) in 100 mM KH_2PO_4 buffer (pH 7.0) for 4 min. The reaction was initiated by adding a solution of CDNB in CH_3CN to a final concentration 2 mM. The rate of production of the GSH or GspSH conjugate with CDNB was calculated using a linear regression to fit the initial rate measured by increase of absorbance at 340 nm ($\Delta\varepsilon$ = 9,600 $M^{-1}cm^{-1}$). When necessary, corrections were made for the uncatalyzed reaction between the thiols and CDNB. All measurements were done in

triplicate and averaged. Data were analyzed with a non-linear least squares fit to the Michaelis-Menten equation 3. The k_{cat} value was determined using equation 4.

Glutathione Disulfide Inhibition Assays

The assays to estimate the inhibition of the addition of GSH to CDNB by GSSG were carried out at 25 ˚C in 100 mM KH_2PO_4 (pH 6.5) containing 3.0 μM YghU and 250 μM CDNB. The GSH concentration was varied between 30 and 400 μM. The inhibited reaction contained 15 μM GSSG.

Disulfide-bond Oxidoreductase Activity Assays

The disulfide-bond oxidoreductase activity of YghU was measured with the model substrate, 2-hydroxyethyldisulfide at pH 8.0 and 25 ˚C by the coupled assay previously described [86,87].

Glutathione-dependent Peroxidase Activity Assays

The peroxidase activity of YghU was evaluated with several peroxides including hydrogen peroxide, *tert*-butyl hydroperoxide, cumene hydroperoxide, linoleic acid 13(*S*)-hydroperoxide and 15(*S*)-HpETE. Little or no activity was detected with hydrogen peroxide and *tert*-Butyl hydroperoxide. The linoleic acid hydroperoxide was synthesized using a method similar to that described Iacazio, et al[88]. Briefly, 1 mM linoleic acid was combined with 0.1 mg/mL soybean lipoxidase in 100 mM Tris-HCl buffer, pH 7.4. The product was extracted with chloroform, dried under argon and resuspended in methanol. The concentration of linoleic acid 13(*S*)-hydroperoxide was measured by absorbance at

234 nm using a ε_{234} = 25,000 $M^{-1} \cdot cm^{-1}$. The organic hydroperoxidase activity of YghU with cumene hydroperoxide and linoleic acid hydroperoxide was measured with the coupled assay with GSH reductase from yeast according to the method described by Flohe and Gunzler[85]. Reactions were followed with a Perkin-Elmer Lambda 45 Spectrophotometer at 25° C in 1 mL cuvettes. Reaction mixtures containing, 1 μM YghU, 2 mM GSH, and 0.8 U of GSH reductase from yeast in 100 mM KH_2PO_4 (pH 7.0), were pre-incubated at 25˚ C for 5 min, followed by addition of NADPH to a final concentration of 0.15 mM and further incubation for 3 min. The reactions were initiated by addition of cumene hydroperoxide at concentrations between 5-150 μM or linoleic acid at concentrations between 10-400 μM. The decrease of absorbance at 340 nm ($\Delta\varepsilon$ = 6,220 $M^{-1}cm^{-1}$) was monitored over 3-5 min and used to calculate the initial rate.

An HPLC-based method similar to that described by Jakobsson P.J., was used to determine lipid peroxidase activity of YghU with 15(*S*)-HpETE [89]. Reactions containing 0.15 μM protein, 3 mM GSH and various concentrations of 15(*S*)-HpETE were incubated in 50 μL of 100 mM KH_2PO_4 buffer (pH 7.0) containing 0.005% BSA (w/v). After 3 minutes the reaction was terminated with 100 μL of acetonitrile containing 0.2% acetic acid. Before analysis 50 μL of water was added and the protein was removed by centrifugation (10 min at 14,000 rpm). Samples were analyzed using reverse-phase HPLC with a Beckman Ultrasphere C18 column (4.6 mm x 25 cm). The mobile phase was acetonitrile:H_2O:acetic acid at 60:40:0.01 v/v with a flow rate of 1.0 mL/min. A photodiode array detector (Waters) set to 236 nm was used to measure absorbance values and Empower software (Waters Corporation) was used to integrate peak areas. Amounts of 15(*S*)-HETE were calculated based on peak area from known amounts of standard

15(*S*)-HETE (retention time of 10.9 minutes) and 15(*S*)-HpETE (retention time of 11.9 minutes). All measurements of peroxidase activity were done in triplicate and averaged. Data were fit to the Michaelis-Menten equation 3 by non-linear regression analysis. The k_{cat} value was determined using equation 4.

Steady-State Fluorescence Titrations

[The work described in this section was performed by Dr. Nina Stourman.]

Dissociation constants for the complex of YghU with GSH and GSP were determined by titration of the intrinsic fluorescence of the protein. Fluorescence measurements were conducted at 25 ˚C on a SPEX Fluorolog-3 spectrofluorimeter (Jobin Yvon Horiba, Edison, NJ) in the constant-wavelength mode. The samples were excited at 295 nm, and the emission was monitored at 340 nm and integrated over a 30 sec period. All measurements were done in triplicate and averaged. Inner filter effects were taken into consideration using equation 1 to calculate a corrected fluorescence, F_c.

The change of fluorescence intensity of 1 μM protein in 20 mM KH_2PO_4 buffer (pH 7.0) was measured after 5 min pre-incubation with GSH to final concentrations 0.025 to 10 mM or GSP to a final concentrations of 0.025 to 8 mM. The GSH binding data were fit to a two-site binding model using equation 5, while the binding data for GSP were fit to a single-site binding model (equation 6). In these equations S = concentration of the thiol substrate and F = absolute value of the fluorescence.

$$F = ([S] \bullet \Delta F_1)/([S] + K_{d1}) + ([S] \bullet \Delta F_2)/([S] + K_{d2}) \qquad (5)$$

$$F = ([S] \bullet \Delta F)/([S] + K_d) \quad (6)$$

Cytoscape Cluster Analysis of the Glutathione Transferase Superfamily.

The data used to construct the network in Figure III-8 was complied and filtered by Dr. Shoshana Brown at The California Institute for Quantitative Biomedical Research at USCF (Prof. Patricia Babbitt, Principle Investigator). The cytoscape software for visualization of these networks is available as a free download from the Cytoscape Consortium (http://www.cytoscape.org). All full-length sequences in the Pfam GST_N and GST_C families (as of July 2010) were downloaded. Sequences of fewer than 150 amino acids were considered fragments and removed from the data set. The sequence set was filtered to remove sequences with greater than 60% identity. An all-by-all BLAST of the remaining sequences was then performed with an E-value cut off of $1e^{-14}$. The resulting networks were visualized in Cytoscape version 2.7.0 using the organic layout. Nodes represent sequences and edges represent BLAST e-values of $1e^{-18}$ or more significance.

Crystallization and Determination of the Structure of YghU

Crystallization of YghU

[The work described in this section was performed by Dr. Mathew Schaab.]

YghU was crystallized by hanging-drop vapor diffusion at 25 ˚C with a reservoir solution of 0.1 M Bis-Tris, 0.2 M NaCl (pH 5.5) containing 25% w/v PEG 3350. The drop consisted of a 1:1 mixture of reservoir solution and a solution of YghU (20 mg/mL) in 30 mM NaH_2PO_4 (pH 7.0) containing 1 mM DTT and 20 mM GSH. Crystals grew in

one week as long rods and belonged to the $P2_12_12_1$ space group with unit cell dimensions of a = 58.197, b = 73.455, c = 130.810. The crystals were transferred to mother liquor supplemented with 35% ethylene glycol as a cryo-protectant before flash-freezing.

X-ray Data Collection and Processing

[The work described in this section was performed by Dr. Joel Harp and Dr. Jane Ladner]

Crystals were flash-cooled and maintained at 100 K for cryogenic data collection. The quality of the X-ray diffraction by the crystals was determined in-house using a Bruker Microstar microfocus rotating anode X-ray generator with Montel confocal multilayer X-ray optics. In-house data were collected using a Bruker-Nonius X8 kappa goniometer and Proteum PT135 CCD area detector. Crystals were maintained at 100 K using Bruker KryoFlex cryostats.

A native data set was collected using the mail-in crystallography service at beamline 22-BM of the Southeastern Regional Collaborative Access Team, SER-CAT, at the Advanced Photon Source, Argonne National Laboratory. A single selenomethionyl derivative crystal was used to collect MAD data at three wavelengths at the SERCAT 22-BM beam line. The wavelengths of data collection were as follows, 0.9793 Å (peak), 0.9792 Å (inflection) and 0.9717 Å (remote). The crystal was maintained at 100 K. All data were reduced using HKL2000. The data sets were closely comparable with respect to their quality and completeness. The average statistics for the three MAD data sets were resolution range = 50 - 1.89 Å, completeness = 97.7 %, redundancy = 6.9, I/s(I) = 19.3,

linear R factor = 0.087. Data quality statistics for the native data reflections collected are presented in Table III-1.

Structure Determination and Refinement

[The work described in this section was performed by Dr. Joel Harp and Dr. Jane Ladner.]

Phasing of the data was accomplished using SHELXC, SHELXD, and SHELXE using the graphical user interface, HKL2MAP[90-93]. All four data sets were input to HKL2MAP with the resolution limit set to 2.5 Å. SHELXD found all eight selenium positions with a maximum correlation coefficient of 67.8%. Refinement of positions in SHELXE resulted in a reported contrast of 0.425 and connectivity of 0.905 and pseudo-free correlation coefficient of 74.0%. Visualization of the map clearly showed density for tyrosine side chains with holes in the ring density.

Phases from SHELXE were imported into the CCP4 suite and a preliminary structure was built using ARP/warp [94]. ARP/wARP was able to build the model to 93% completion. The model from ARP/wARP was then refined using the native data to 1.50 Å combined with rounds of manual inspection and fitting using Xfit [95].

The final model was refined with REFMAC [82] and consisted of 4649 non-hydrogen protein atoms, four GSH molecules (80 atoms) and 471 solvent atoms. The A chain is missing residues 1-4 and the electron density for residue 5 is poor. Chain B is missing residues 1-5 and residues 15-18 have very poor electron density. Alternative conformations are included for 11 residues in chain A and 12 residues in chain B. Refinement statistics are given in Table III-1.

Parameters	**Values**
PDB file name	3C8E
Space group	$P2_12_12_1$
Unit cell dimensions (Å)	58.197, 73.455, 130.810
Resolution limits (Å)	50-1.41 (1.46-1.41)
Number of independent reflections	95609 (6087)
Completeness (%)	87.7 (56.5)
Redundancy	8.9 (3.0)
Average I/average σ(I)	-
R_{sym} (%)	10.9 (65.3)
R-factor (overall) (%) / Number of reflections	16.5 / 80150
R-factor (working) (%) / Number of reflections	16.2 (21.5) / 72113
R-factor (free) (%) / Number of reflections	19.7 (27.2) / 8037
Number of protein atoms	14523
Number of hetero-atoms[a]	551
Average B Values	
Protein atoms (Å^2)	13.8
Glutathione (Å^2)	12.0
Solvent (Å^2)	22.9
Weighted root-mean-square deviation of ideality	
Bond lengths (Å)	0.019
Bond angles (deg)	1.834
General planes (Å)	0.01

Table III-1 X-ray data collection and least-squares refinement statistics for the structure of YghU. Values for the highest resolution bin are given in parentheses.
[a]These include 4 glutathione molecules and 470 water molecules.

Results and Discussion

Characterization of the Enzyme Activity and Substrate Binding of YghU

The GSH transferases activity of YghU was measured with a variety of electrophilic substrates. YghU has modest GSH transferases activity with 1 mM CDNB (k_{cat} = 0.109 ± 0.005 sec^{-1}; k_{cat}/K_m^{GSH} = 1400 ± 300 $M^{-1}sec^{-1}$; K_M^{GSH} = 80 ± 20 μM). No activity was observed with the other electrophilic substrates tested including, ethacrynic acid, trans-4-phenyl-3-buten-2-one, and 1,2-epoxy-3-(4-nitrophenoxy)propane. No detectable activity was observed with CDNB when GSP was used as a thiol nucleophile.

The intrinsic fluorescence of YghU was monitored upon titration with GSH and GSP. In the case of GSH, a biphasic binding equilibrium was observed as an increase in fluorescence followed by a decrease in fluorescence (Figure III-1). The observation of a higher binding affinity binding site (K_{d1}^{GSH} = 0.07 ± 0.03 mM) and a lower affinity binding site (K_{d2}^{GSH} = 1.3 ± 0.2 mM) is consistent with either negative cooperativity in the binding of two molecules of GSH to the dimer or the binding of two molecules of GSH, with different binding affinities, to each active site in the dimer. The same fluorescence titration with GSP yielded a monophasic binding equilibrium with K_d^{GSP} = 1.0 ± 0.1 mM.

Attempts to titrate the intrinsic fluorescence of YghU with GSSG did not produce sufficient signal (ΔF) for a reliable titration. However, when GSSG was used as an inhibitor in the CDNB assay the observed k_{cat} and k_{cat}/K_M^{GSH} values decreased ten-fold with the additional of only 15 μM GSSG ([YghU] = 3 μM). This suggests that the K_i^{GSSG} is much, much less than 10 μM and that GSSG actually binds YghU more tightly that

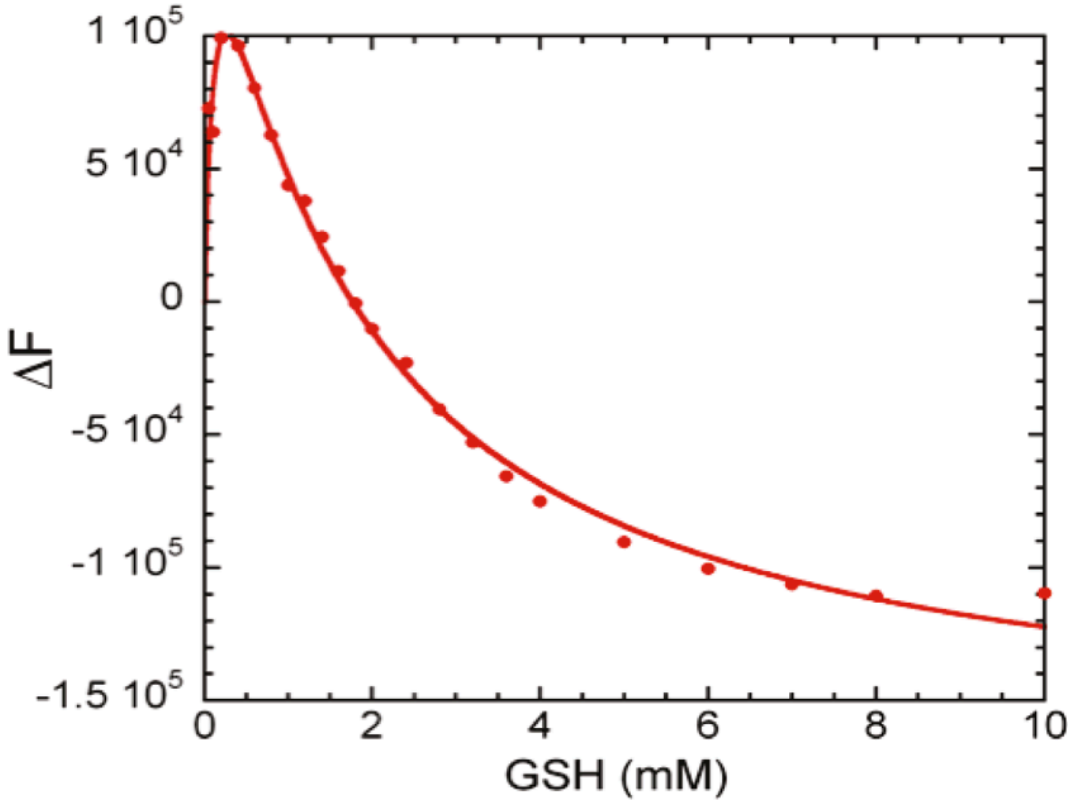


Figure III-1 Titration of the intrinsic protein fluorescence of YghU with GSH. The experimental data were fit to a two-site binding model (equation 5) with the following values: K_{d1} = 0.07 ± 0.03 mM, and ΔF_1 = (2.1 ± 0.6) x 10^5; K_{d2} = 1.3 ± 0.2 mM, and ΔF_2 = (-8.7 ± 0.5) x 10^5 (r = 0.999).

GSH. However, the low catalytic activity of YghU in the CDNB assay and the high background reaction of GSH and CDNB prevented determination of an exact mechanism of inhibition of YghU by GSSG and a reliable measurement of K_i.

Structural Characterization of YghU

The 1.5 Å resolution structure of YghU in the presence of GSH was determined by MAD phasing. A ribbon representation of the structure of the dimer of YghU is shown in Figure III-2. The protein adopts a typical GSH transferases fold with an N-terminal thioredoxin-like domain and a C-terminal all α-helical domain. The sequence of YghU is approximately 50 amino acids longer than other GSH transferases. This additional sequence can be seen as an N-terminal extension that wraps up over the active site and then crosses over to the opposite subunit. The N-terminus limits access to the active site as shown in Figure III.3. There is also additional sequence at the C-terminus, which consists of a short α-helix and random coil structures in the crystal structure.

The most striking and functionally significant detail of the structure is the two molecules of GSH found in each active site. The two GSH molecules are bound in an anti-parallel arrangement with a distance of 3.4 Å between the sulfur atoms. This distance could reasonably correspond to a hydrogen bond between a thiolate anion and a thiol, and is too long for a typical disulfide bond (usually 2.05 Å). An omit map of the electron density for the two molecules of GSH is displayed in Figure III-4.

The details of the potential hydrogen-bonding interactions are shown in Figure III-5. One molecule of GSH is bound in the typical position with the γ-glutamyl portion forming hydrogen bonding interactions with the 'recognition motif' of Glu-117 and Ser-

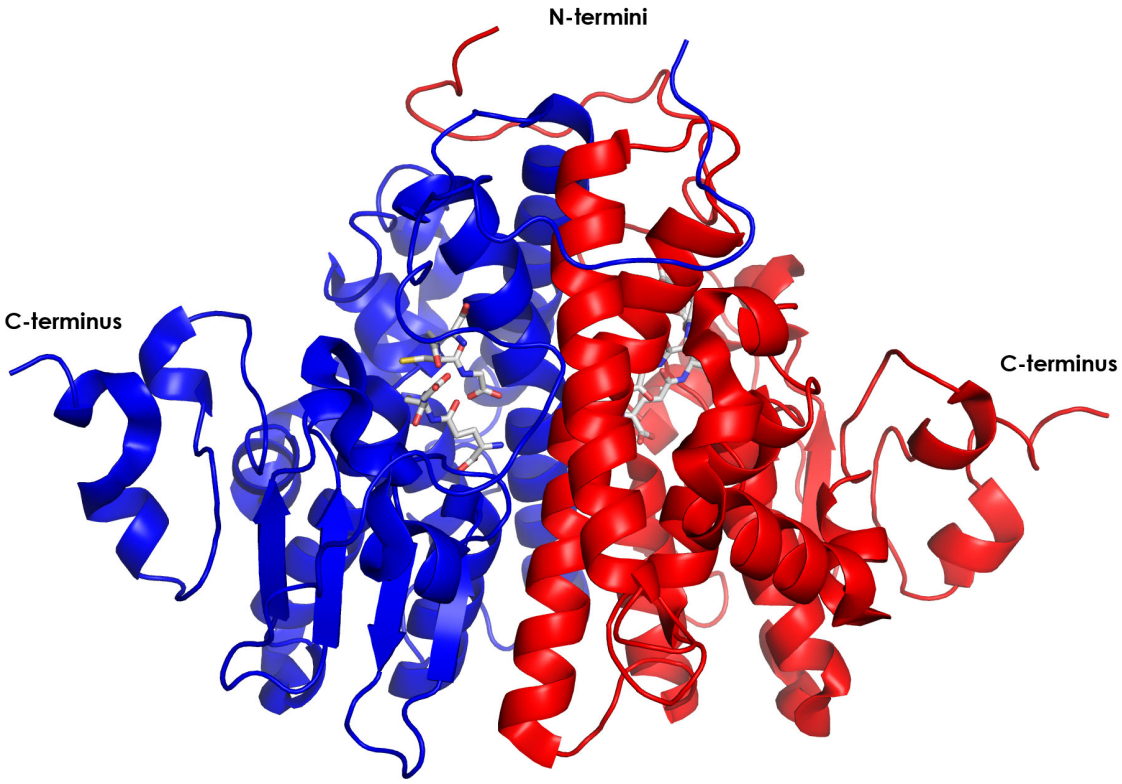


Figure III-2 Ribbon diagram of the structure of the dimer of YghU with two GSH molecules bound in the active sites. The two subunits are shown in red and blue respectively and the four molecules of GSH are shown in stick representation.

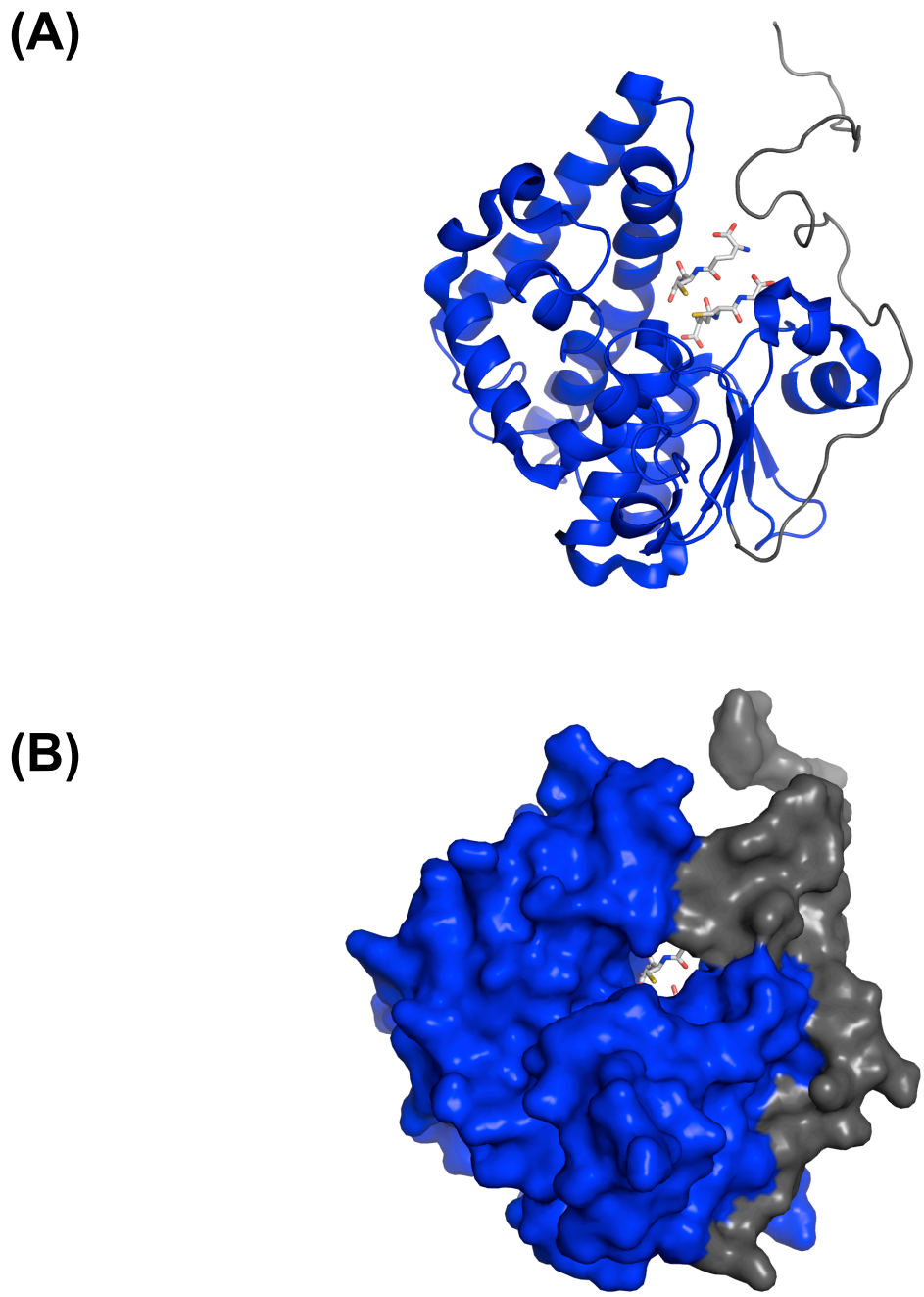


Figure III-3 (A) Structure of the monomer of YghU. The 50 amino acid N-terminal extension is colored in grey while the rest of protein is colored in blue. (B) The occlusion of the active site by the N-terminus can be seen in the space-filling model. The opening is approximately 11 Å x 11 Å.

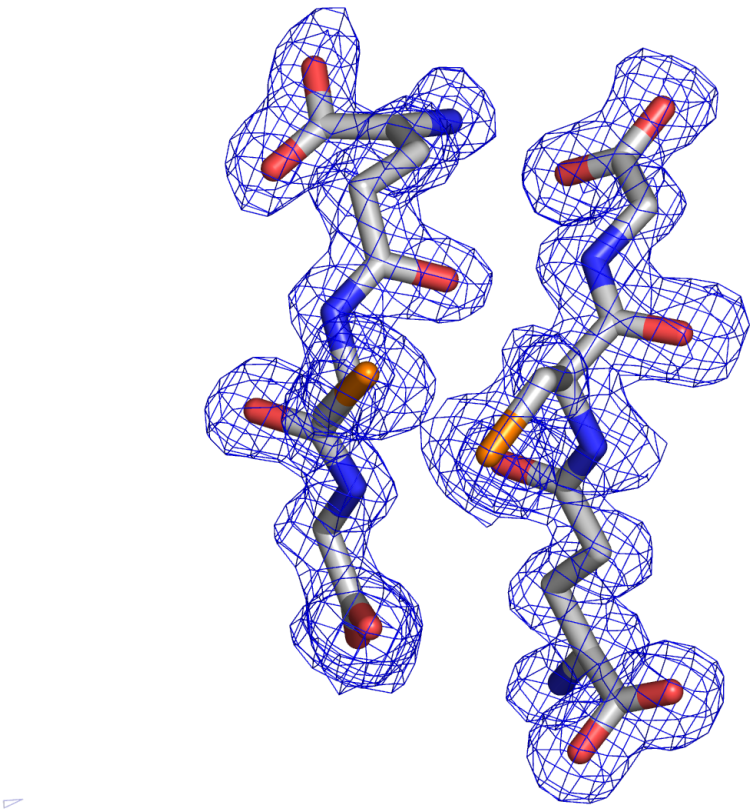


Figure III-4 Omit map of the electron density for the two molecules of GSH observed in the active site of YghU. The map, contoured at 2.5σ, was calculated with coefficients of the form $F_o - F_c$ in which the observed and calculated structure factor amplitudes were calculated from the model lacking the coordinates of the two molecules of GSH. The final model of the two molecules is shown in stick representation. The two sulfur atoms in the middle are colored in orange.

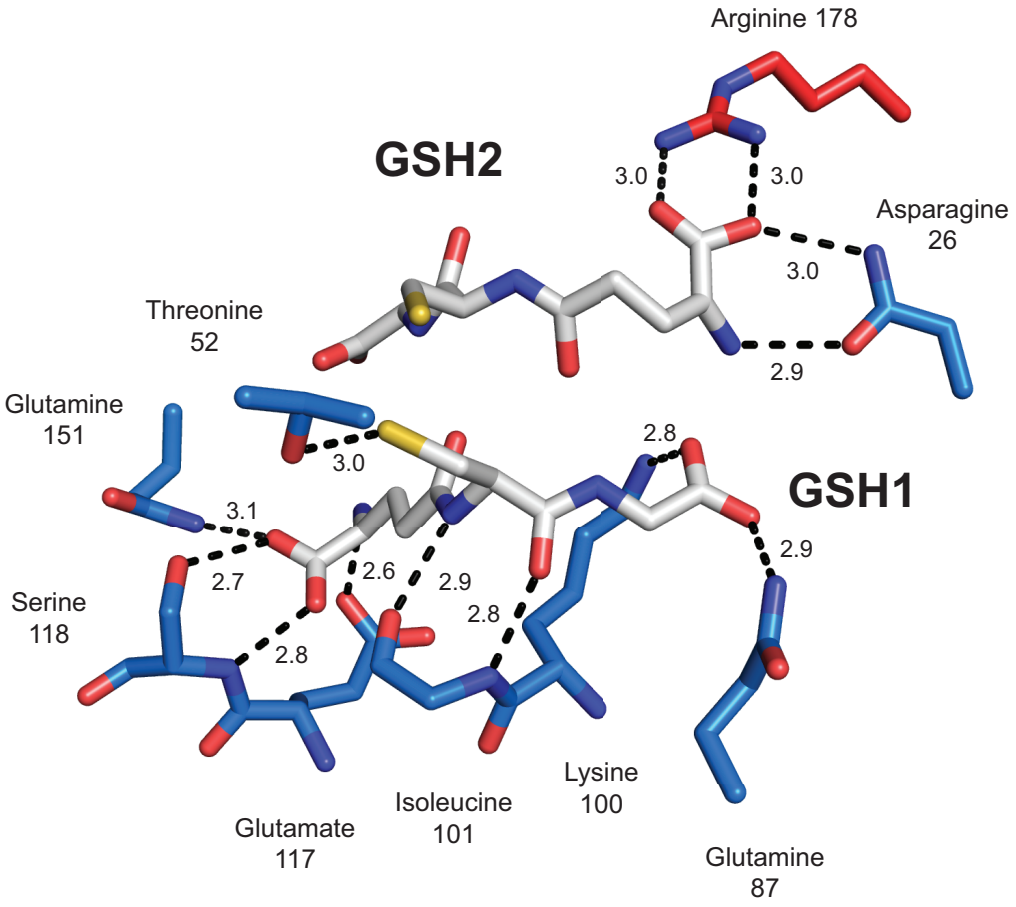


Figure III-5 Details of the GSH binding site of YghU. Subunits one and two are colored in blue and red, respectively. Potential hydrogen bonding interactions are shown as dashed lines with distances shown in angstroms (Å).

118 (reference Figure II-5). As seen in the YfcG structure, a threonine residue is located at the amino terminal end of α-helix 1 and hydrogen bonds with the sulfhydryl of the bound thiol, in this case the cysteine sulfhydryl of GSH1. The GSH1 molecule also makes three additional hydrogen bonding interactions with the protein; the cysteine nitrogen of GSH1 hydrogen bonds to the backbone carbonyl oxygen of Ile-101, the cysteine carbonyl oxygen of GSH1 hydrogen bonds to the amide nitrogen of Ile-101, and both the side chain of Lys-100 and Gln-87 hydrogen bond to the glycine carboxylate of GSH. The second GSH molecule forms fewer hydrogen bonding interactions with the protein. Similar to the YfcG structure, the γ-glutamyl-α-carboxylate of GSH2 in the YghU structure forms a critical ionic interaction with Arg-178 from the opposite subunit. The γ-glutamyl-α-carboxylate of GSH2 also forms hydrogen bonding interactions with Asn-26, an interaction that is unique to the YghU structure. Asn-26 is located on the N-terminal extension that wraps around the active site of the enzyme. It is not known if this interaction is an artifact of the crystal packing of the N-terminal region or of biological signifance. The observation that GSH1 forms more hydrogen bonding interactions with the protein than GSH2 suggests, but does not prove, that GSH1 is the more tightly bound molecule in the titration experiments.

The Structure Reveals Function

The observation of two molecules of GSH in the active site of the crystal structure of YghU suggests that the protein might be involved in redox reactions. Some GSH-trasnferases are known to act as GSH-dependent peroxidases and reduce organic hyroperoxides using two molecules of GSH. It would appear that YghU is set up for the

efficient reduction of peroxides. The GSH-dependent peroxidase activity of YghU was assayed with a variety of peroxide and hydroperoxide substrates. While the enzyme does not reduce H_2O_2 or tert-butyl hydroperoxide at significant rates, it does have modest activity with cumene hydroperoxide, 13(*S*)-linoleic acid hydroperoxide and 15(*S*)-HpETE hydroperoxide as summarized in Table III-2. The turnover number for these substrates is low but the values of k_{cat}/K_m (on the order of 10^3), suggest that these types of peroxides may be natural substrates for YghU. However, the identities of specific hydroperoxide substrates for YghU in *E. coli* is still unknown. This point is especially important considering *E. coli* does not possess arachidonic acid and thus any hydroperoxides derived from it.

Another possible function of YghU is that of a disulfide-bond oxidoreductase. In this regard, YghU was found to have robust activity toward 2-hydroxyethyl disulfide (k_{cat} = 74 ± 6 s^{-1}, k_{cat}/K_M^{GSH} = (6.14 ± 1.3) x 10^4 $M^{-1}s^{-1}$). The kinetic constants measured for YghU in this assay are very similar to those measured for YfcG (k_{cat} = 29 ± 2 s^{-1}, k_{cat}/K_m^{GSH} = (1.8 ± 0.2) x 10^4 $M^{-1}s^{-1}$) as well as those measured for *E. coli* glutaredoxins (reference Table II-3). From these results we can conclude that YghU appears to be a fairly efficient disulfide bond reductase, an activity that is due to its ability to bind two molecules of GSH in the active site.

A New Class of Glutathione Transferases

The structural and functional properties of YghU are remarkably similar to those of YfcG. Both enzymes have moderate activity towards hydroperoxides, with YghU being the more efficient of the two. Most importantly, is their shared robust disulfide

Substrate	K_m^{ROOH} **(mM)**	k_{cat} **(s^{-1})**	k_{cat}/K_M^{ROOH} **($M^{-1}s^{-1}$)**
Cumene Hydroperoxide	16 ± 2	0.050 ± 0.001	3500 ± 400
13(S)-Linoleic Acid Hydroperoxide	130 ± 30	0.19 ± 0.02	1500 ± 400
15(S)-HpETE	28 ± 4	0.096 ± 0.004	3400 ± 1500

Table III-2 Kinetic constants for reactions catalyzed by YghU with various organic hydroperoxides at 25ºC.

bond reductase activity towards 2-hydroxyethyl disulfide. YfcG and YghU are comparable in this activity to *E. coli* glutaredoxins. However, in contrast to glutaredoxins, which use an active site cysteine or cysteine pair to reduce disulfide substrates, YfcG contains no active site cysteines and YghU has no cysteine residues at all. A possible mechanism for the reduction and oxidation of a substrate by YghU and YfcG is outlined in Figure III-6. It is important to point out that disulfide bond oxidoreductase activity is not a common one for GSH transferases. Only two other *E. coli* GSH transferases proteins have described oxidoreductase activity. EGST and YqjG have very low and modest disulfide bond oxidoreductase activity, but both proteins have active site cysteine residues that are necessary for this activity [96,97]

A structural superposition of the YghU·2GSH and the YfcG·GSSG complexes makes an even more compelling argument for their relationship (Figure III-7). The RMSD between 177 C_α atom pairs is 0.817 Å in the overlaid structures. The cores of the proteins are remarkably similar, especially with regard to the thioredoxin domain. The main differences in the two structures are the N-terminal and C-terminal extensions observed in the YghU structure but not in the YfcG structure. While the close superposition of the polypeptide chain is not surprising given their sequence similarity (38% identical) (Figure III-7), the almost identical superposition of the bound thiol(s), GSSG (YfcG) and 2 GSH (YghU), cannot be ignored. Of particular interest are the hydrogen bonding interactions between the hydroxyl group of a threonine residue located at the amino terminus of α-helix 1 and the sulfur of the bound thiol(s). Most GSH transferases have an electrophilic side chain located in the loop regions between the end of β-strand 1 and the start of α-helix 1 in the thioredoxin domain (reference Figure II-5).

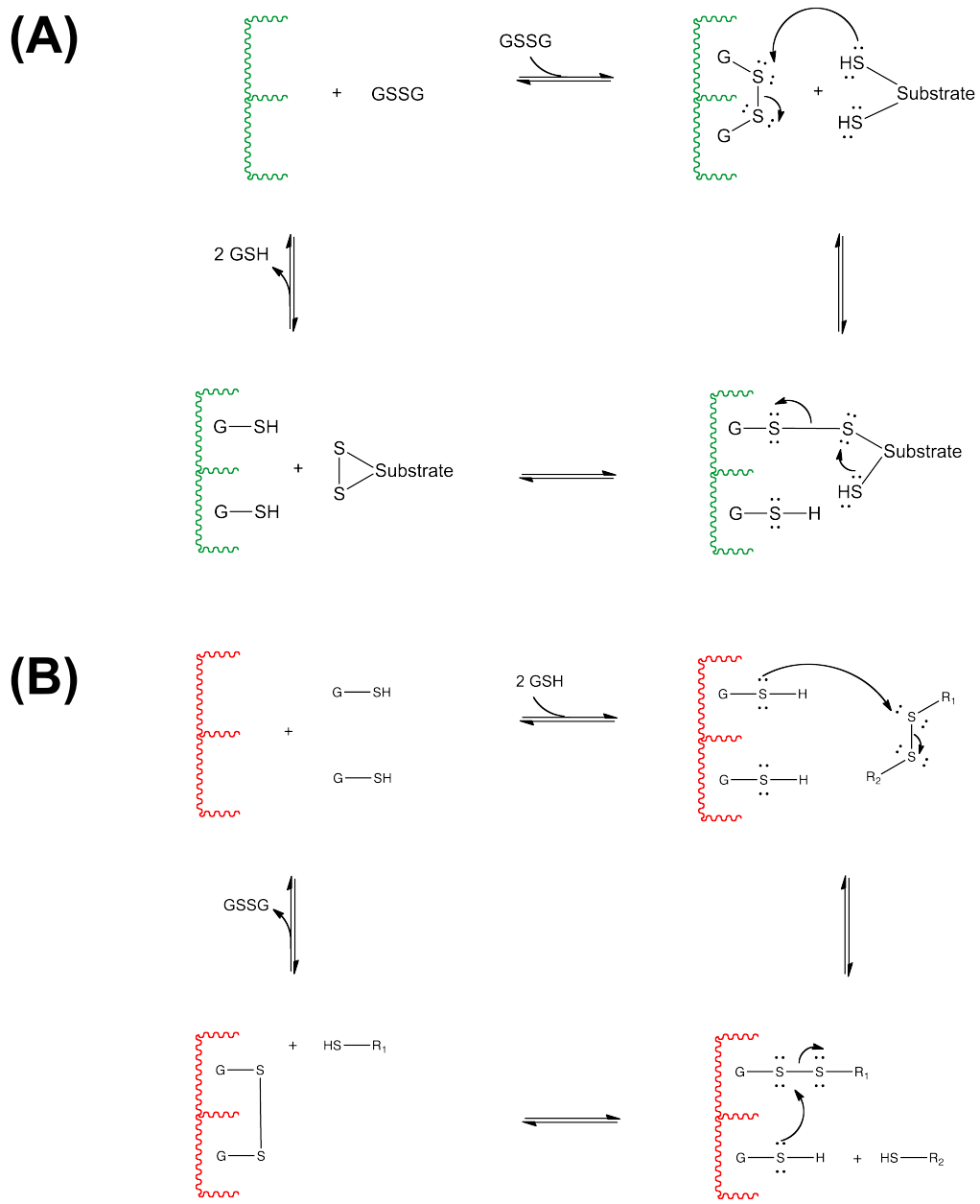
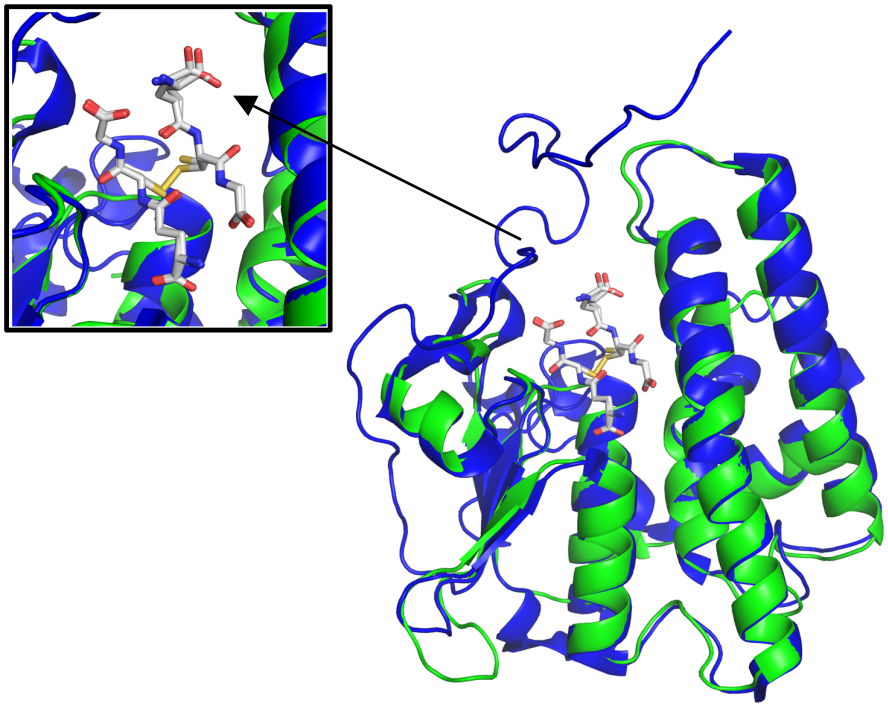


Figure III-6 Possible mechanism for the reduction and oxidation of a substrate by YfcG (A) and YghU (B).

(A)



(B)

```
CLUSTAL 2.1 multiple sequence alignment

Q46845|YGHU_ECOLI      MTDNTYQPAKVWTWDKSAGGAFANINRPVSGPTHEKTLPVGKHPLQLYSLGTPNGQKVTI 60
P77526|YFCG_ECOLI      -------------------------------------------MIDLYFAPTPNGHKITL 17
                                                                 ::**   ****:*:*:

Q46845|YGHU_ECOLI      MLEELLALGVTGAEYDAWLIRIGDGDQFSSGFVEVNPNSKIPALRDHTHN---PPIRVFE 117
P77526|YFCG_ECOLI      FLEE------AELDYRLIKVDLGKGGQFRPEFLRISPNNKIPAIVDHSPADGGEPLSLFE 71
                       :***      :  :*    : :*.*.** . *:.:.**.****: **:      *: :**

Q46845|YGHU_ECOLI      SGSILLYLAEKFGYFLPQDLAKRTETMNWLFWLQG-AAPFLGGGFGHFYHYAPVKIEYAI 176
P77526|YFCG_ECOLI      SGAILLYLAEKTGLFLSHETRERAATLQWLFWQVGGLGPMLGQNH-HFNHAAPQTIPYAI 130
                       **:******** * **.::  :*: *::**** *  .*:** .. ** * ** .* ***

Q46845|YGHU_ECOLI      NRFTMEAKRLLDVLDKQLAQHKFVAGDEYTIADMAIWPWFGNVVLGGVYDAAEFLDAGSY 236
P77526|YFCG_ECOLI      ERYQVETQRLYHVLNKRLENSPWLGGENYSIADIACWPWVNAWTR-------QRIDLAMY 183
                       :*: :*::** .**:*:* :  ::.*::*:***:* ***..  .        : :* . *

Q46845|YGHU_ECOLI      KHVQRWAKEVGERPAVKRGRIVNRTNGPLNEQLHERHDASDFETNTEDKRQG 288
P77526|YFCG_ECOLI      PAVKNWHERIRSRPATG--------QALLKAQLGDERSDS------------ 215
                         *:.* :.: .***.         :. *: ** :.:. *
```

Figure III-7 (A) Structural superposition of the YghU•2GSH and YfcG•GSSG complexes. The inset shows the remarkable similarity of the bound 2GSH and GSSG molecules in the active site. **(B)** Multiple sequence alignment of YghU and YfcG created using the CLUSTALW2 program. Amino acids involved in thiol(s) binding are shown in red.

Typically, these residues are tyrosine, serine, or cysteine, and serve to lower the pK_a of the sulfhydryl group on GSH. YghU and YfcG are the first structurally confirmed cases of GSH transferases that utilize a threonine hydroxyl in this capacity. In both protein structures an arginine from the opposite subunit forms a critical ionic interaction with the second molecule of GSH (YghU) or the second half of the bound GSSG (YfcG). The use of (1) a threonine in the active site of the enzyme and (2) an opposite subunit arginine for the stabilization of the bound thiol(s) may be a structural signature for a new class of GSTs that appear to function as disulfide-bond oxido-reductases.

Nu-class Glutathione Transferases

In order to analyze the YfcG and YghU enzymes in the wider context of the whole GSH transferases superfamily, all full-length GSH transferases sequences were downloaded from the pfam database (http://pfam.sanger.ac.uk/) and visualized using Cytoscape (as detailed in the methods section). The resulting network (Figure III-8) clearly suggests that the two *E. coli* proteins, YghU and YfcG, and a number of other sequences cluster closely together. In the tradition of naming various subfamilies of GSH transferases with a Greek letter, we suggest that this particular subfamily be designated as nu. This nu-class cluster, is distinct from other well-characterized GSH transferase proteins in the alpha, beta, mu, omega, phi, pi, sigma, tau, theta and zeta families.

Sequences relating to YghU and YfcG in this cluster were further analyzed by network analysis, multiple sequence alignment and for taxonomic diversity. These sequences are primarily from bacterial and fungal sources (Figure III-9). Just under half (46%) of the sequences are bacterial with the largest population being from

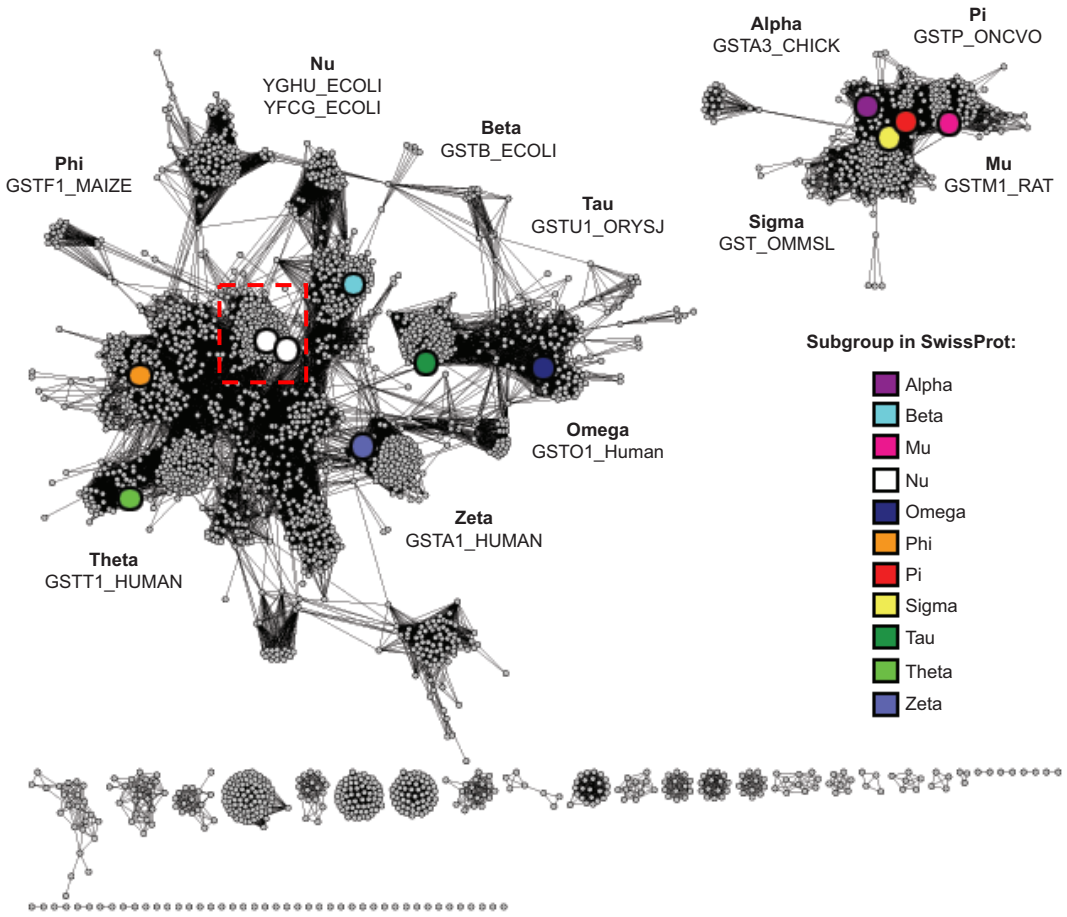


Figure III-8 YghU and YfcG form a nu class of glutathione transferases. The overall sequence similarity network contains 2,851 sequences and 97,144 edges. Edges represent BLAST *E* values of 10^{-18} or more stringent. Large nodes are colored by the classification of the amino acid sequences in SWISS-PROT and indicate a representative member of each subgroup as follows: Alpha, GSTA3_CHICK (UniProt Accession P26697); Beta, GSTB_ECOLI (P0ACA7); Mu, GSTM1_RAT (P04905); Nu, YFCG_ECOLI (P77526) and YGHU_ECOLI (Q46845); Omega, GSTO_HUMAN (P78417); Phi, GSTF1_MAIZE (P12653); Pi, GSTP_ONCVO (P46427); Sigma, GST_OMMSL (P46088); Tau, GSTU1_ORYSJ (Q10CE7); Theta, GSTT1_HUMAN (P30711); Zeta, GSTZ1_HUMAN (O43708). Sequences in the red box were analyzed further for taxonomic diversity and by multiple sequence alignment.

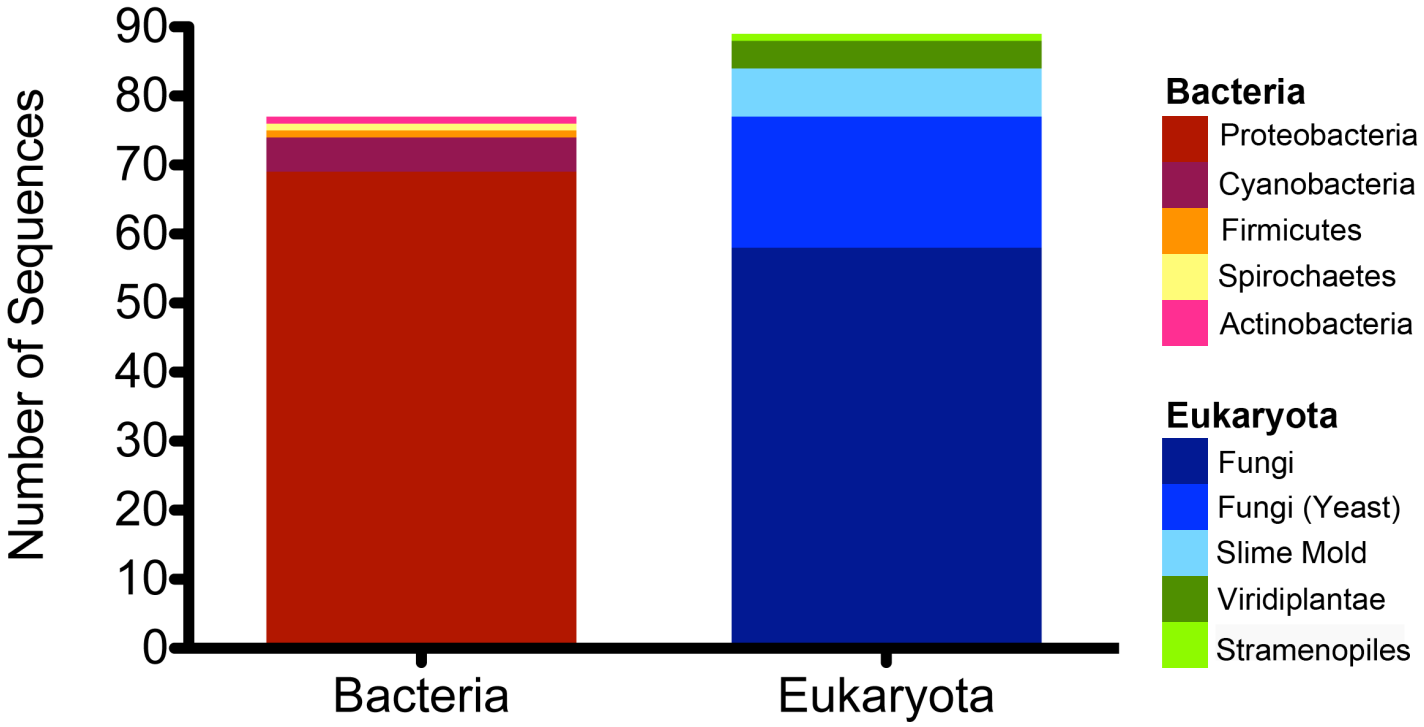


Figure III-9 Taxonomic distribution of the nu-class cluster. Approximately, 46% of the sequences are bacterial with the largest population being from the phylum proteobacteria. Of the 89 eukaryotic sequences, 87% are from fungal organisms.

proteobacteria (90% bacterial, 42% total). There are five cyanobacteria sequences. Of the 89 eukaryotic sequences 77, (87%) are from fungal sources. Interestingly, 19 sequences are from yeast, including several from *Saccharomyces* and *Candida* strains. Many of these sequences are annotated as Ure2 proteins.

Ure2 is the protein determinant of the non-chromosomal genetic element [*URE3*] found in many yeast strains [98]. The protein is involved, through an unknown mechanism, in a signal transduction cascade responsible for the regulation of nitrogen catabolism [99]. It is also associated with resistance to heavy metal toxicity (especially cadmium) *in vivo*, and shows GSH-dependent peroxidase activity with hydrogen peroxide, t-butyl hydroperoxide, and cumene hydroperoxide *in vitro* [100]. Structurally, the Ure2 protein contains a flexible N-terminal region rich in asparagine and glutamine residues and a C-terminal thioredoxin and α-helical domains which are < 20% identical to GSH transferases [99]. The N-terminal region regulates the assembly of Ure2 protein to an aggregate form which gives rise to amyloid-like fibrils *in vitro* [98]. Ure2 proteins bind GSH fairly tightly ($K_d \cong 20$ μM). In the crystal structure of Ure2p from *S. cerevisae*, a conserved asparagine residue forms hydrogen-bonding interactions with the thiol of GSH [99]. Recently, robust disulfide-bond oxidoreductase activity was described for Ure2, $k_{cat}^{app(HEDS)} = 1.2\ s^{-1}$, in the 2-hydroxyethyl disulfide assay [100]. This disulfide bond reductase activity is distinct from the classical glutaredoxin motif, as the Ure2 protein contains no cysteine residues. YghU is similar in this regard in that it also contains no cysteine residues. YfcG has one cysteine residue, located > 15Å away from the active site and is not involved in the disulfide reductase activity (reference Table II-3).

Overall the cysteine content of the nu-class cluster of sequences is greatly reduced when compared to other GSH transferases sequences and to the cysteine content of proteins in commonly studied organisms. In fact, almost 40% of the nu-class sequences contain no cysteine residues at all (Figure III-10A). The percentage of sequences with no cysteine residues in the nu-class cluster is much greater than other GSH transferases sequences, in which only 8.4% have no cysteine residues. When comparing the cysteine content of randomly selected GSH transferases sequences to sequences of the nu-class cluster, the average percentage of cysteine residues per sequence drops from 1.1% to 0.5% (Figure III-10B). The occurrence of only 0.5% cysteine residues in protein sequences in the Nu-class cluster is also well below the average occurrence of cysteine in various well-studied organisms and the mathematical probability for the occurrence of cysteine (Figure III-10C)[101,102].

The YghU, YfcG and the Ure2 proteins share functional properties including GSH-dependent peroxidase activity and disufide-bond oxidoreductase activity. The proteins are also characterized by a lack of cysteine residues, a property true for many of the nu-class sequences. Therefore, proteins in the nu-class cluster could represent a new category of disulfide-bond oxidoreductases that operate through a glutaredoxin-independent mechanism and are thus unaffected by increases in reactive oxygen species, which would otherwise irreversibly inactive cysteine residues.

However, it is clear at this point that the structural features and mechanisms of thiol binding and utilization may differ in this class of proteins. If interrogated at a more stringent cut off value, the sequences in the nu-class cluster clearly subdivide into three groups (Figure III-11). While GSH-dependent peroxidase and disulfide bond

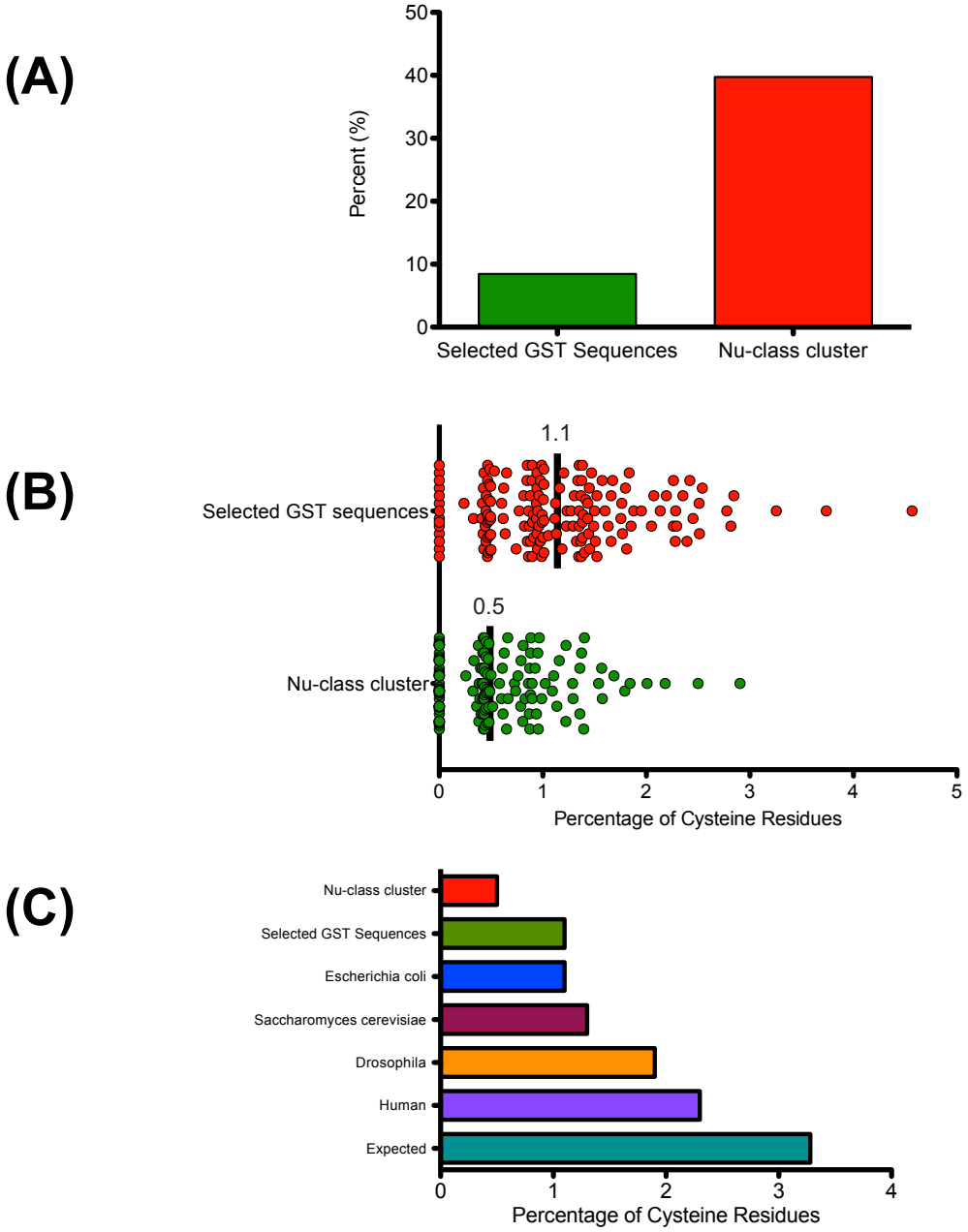
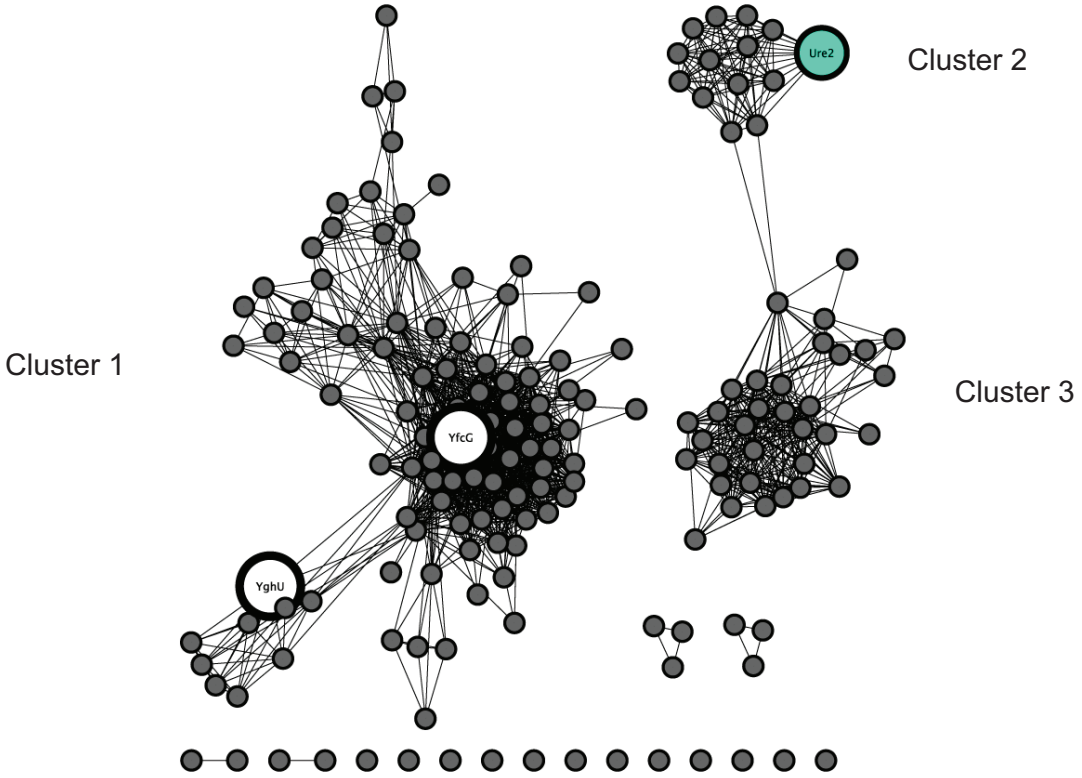


Figure III-10 Analysis of the cysteine content of the sequences in the nu-class cluster compared to other GSH transferase sequences and protein sequences in commonly studied organisms. **(A)** Percentage of sequences with no cysteine residues in the nu-class cluster compared to an equal number of randomly selected GSH transferase sequences. **(B)** Percentage of cysteine residues in nu-class sequences compared to randomly selected GSH transferase sequences. The vertical black bar indicates the average. **(C)** The occurrence of cysteine in proteins in the nu-class cluster compared to other GSH transferase sequences and the cysteine content of commonly studied organisms. The cysteine content is expressed as a percentage. Because cysteine has 2 codons out of 61 codons the mathematical probability for the occurrence of cysteine is 3.28% (expected, bottom row). Data for *E. coli*, *S. cerevisiae*, *Drosophila*, and *Human* are taken from reference 102).

(A)



(B)

	111	Beta 1	121	Alpha 1	131	1
Cluster1	- - - - - M I D L Y T - - - - - W P T P N G R K V S I M L - - -					
sp|Q46845|YGHU_ECOLI	V G K H P L Q L Y S - - - - - L G T P N G Q K V T I M L - - -					
sp|P77526|YFCG_ECOLI	- - - - - M I D L Y F - - - - - A P T P N G H K I T L F L - - -					
Cluster2	P M E G Y T L F S H - - - - - R S A P N G F K V A I V L - - -					
sp|P23202|URE2_YEAST	P L E G Y T L F S H - - - - - R S A P N G F K V A I V L - - -					
Cluster3	- - - - - - I K L Y G - - - - - P G G P N P W K V A I I L - - -					

Figure III-11 (A) Overall sequence similarity network of 166 nu-class glutathione transferase sequences created in Cytoscape v.2.0.7. Edges represent BLAST *E* values of 10^{-45} or more stringent. **(B)** Multiple sequence alignment of the consensus sequences from clusters 1, 2,and 3 compared to the sequences of YghU and YfcG from *E. coli* and Ure2 from S. *cerevisiae*. The amino acids corresponding to beta sheet 1 (highlighted in green) and alpha helix 1 (highlighted in yellow) of the thioredoxin domain are shown. Amino acids contacting the sulfhydryl of the bound thiol in the crystal structure are boxed in red. (Ure2 PDB ID: 1JZR, reference 99). The multiple sequence alignments were created using the ClustalW2 program and viewed using Chimera.

oxidoreductase activities have been described for enzymes in all three groups there are important differences in amino acid sequence between the three groups. The largest cluster contains the YfcG and YghU proteins. More than 86% of these sequences contain a conserved threonine residue in the equivalent position as the threonine found contacting the bound 2 GSH or GSSG in the YghU and YghU crystal structures (Figure III.11.B). The second set of sequences is exclusively Ure2 sequences from yeast. They are characterized by an N-terminal region rich in asparagine and glutamine residues, which are important for prion formation in these proteins. The Ure2 proteins contain an alanine residue in place of the threonine seen in the YfcG/YghU cluster. However, a conserved asparagine was found to hydrogen bond with the thiol group of GSH in the crystal structure of Ure2p from *S. cerevisiae*. While there is no structural information for proteins in the third cluster, multiple sequence alignments indicate that a conserved glycine residue is in place of the threonine residue.

Conclusions

The YghU and YfcG proteins encoded in the genome of *E. coli* represent a previously unrecognized class of GSH transferases that have unique structural and catalytic properties. These two proteins have a distinct active site and the ability to bind GSSG with high affinity or two molecules of GSH simultaneously. The proteins also share disulfide-bond oxidoreductase and GSH-dependent peroxidase activity. Despite the similarities, YghU and YfcG have significant chemical and structural differences, including their preferred oxidation states and the N-terminal and C-terminal extensions of YghU.

YghU and YfcG belong to a new class of GSH transferases, which include the Ure2 proteins from yeast. Proteins in the nu-class share GSH-dependent peroxidase and disulfide-bond oxidoreductase activities and are characterized by a marked reduction in cysteine content. Proteins in the cluster may represent a new category of disulfide-bond oxidoreductases that operate through a cysteine-independent mechanism and thereby circumvent oxidative damage to protein thiols which would normally occur under conditions of oxidative stress. However, the mechanism of thiol(s) binding may differ in this class of proteins and the exact substrates and the *in vivo* activity of the proteins remains to be elucidated.

CHAPTER IV

STRUCTURAL AND FUNCTIONAL STUDIES OF YqjG, A GLUTATHIONE TRANSFERASE HOMOLOG FROM *Escherichia coli* K-12, REVEALS IT BELONGS TO A NOVEL STRUCTURAL CLASS

Abstract

The 2.2 Å resolution crystal structure of the GSH transferase homolog, YqjG from *E. coli* reveals that is belongs to a new structural class of GSH transferases. YqjG adopts a classical GSH transferase fold but a unique splayed dimer. The protein exhibits modest disulfide bond reductase activity and may possess *S*-glutathionyl-reductase activity with other substrates. The GSH binding properties and characteristics of the dimer interface of YqjG are compared to two other proteins with a similar splayed dimer. A global analysis of the GSH transferase superfamily has identified 500 other sequences that belong to this novel class. The majority of these sequences are bacterial. Protein in this class exhibit *S*-glutathionyl-(chloro) hydroquinone reductase and *S*-(phenylacyl) hydroquinone reductase activity [97,103]. Possible hydroquinone substrates for YqjG in *Escherichia coli* are presented.

Introduction

Recent advances in sequencing technologies have lead to exponential increases in gene and protein sequences in databases like GenBank and SWISS-Prot (Figure IV-1). This sequence information has paved the way for increased understanding of the composition and organization of genes in many organisms [104]. However, the accurate

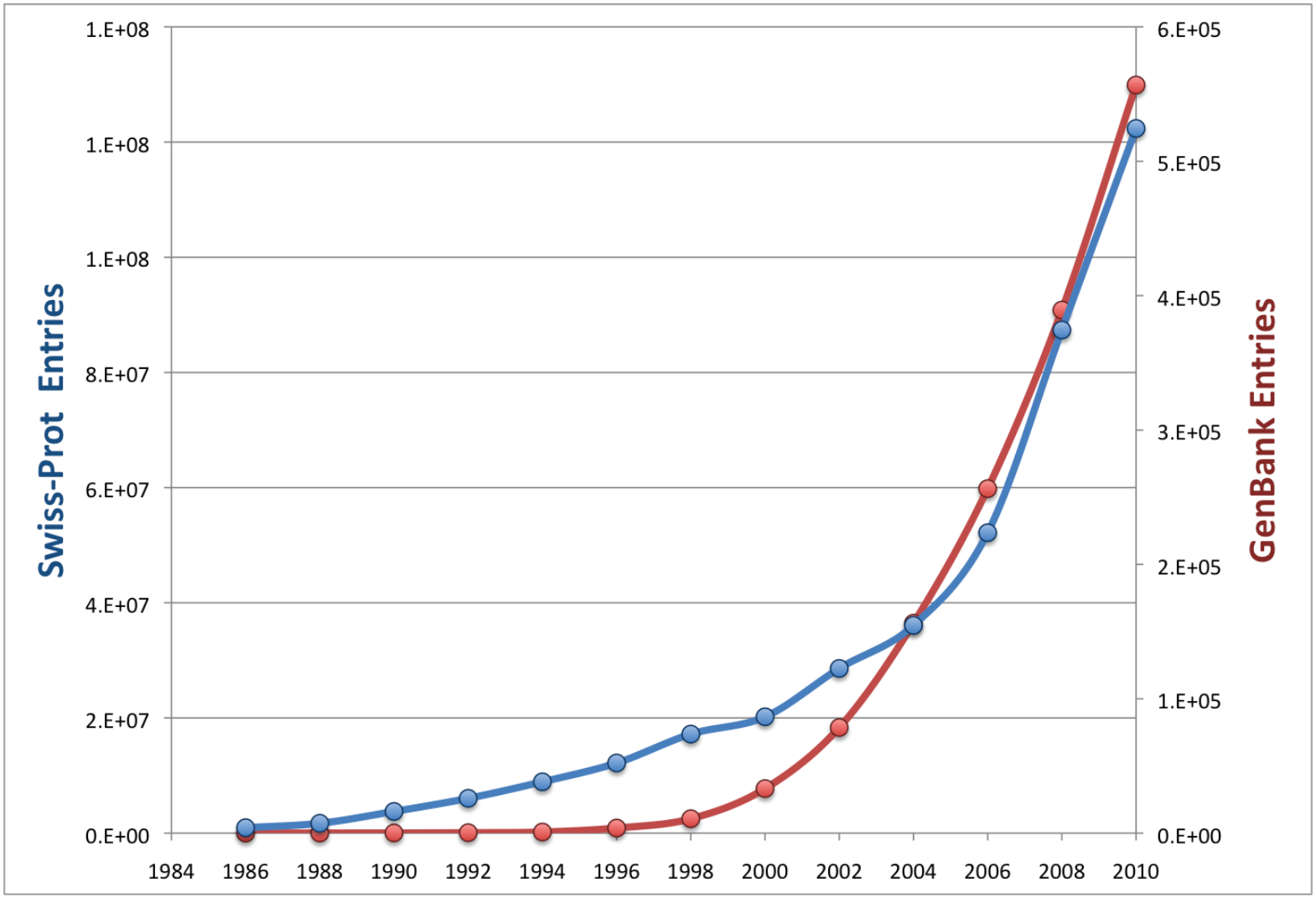


Figure IV-1. Recent exponential increases in gene and protein sequences. As of April 15, 2011 GenBank contained 135,440,924 enteries and as of May 3, 2011 Swiss-Prot contained 528,048 enteries.

determination of *in vitro* enzyme function and an *in vivo* physiological function has remained a limitation of genomic biology [105].

GSH transferases are a multifunctional superfamily of proteins with fundamental roles in cellular detoxication of exogenous and endogenous compounds [9]. While evidence of the first GSH transferase in bacteria was reported 29 years ago, only in recent years has the number of GSH transferases genes in bacteria been appreciated [29]. In the latest release of the Pfam database (March 2011), the GST_N family contained over 10,000 sequences from approximately 1,500 species [106]. Over half of these sequences (6,036) are from bacteria. Bacterial genomes contain multiple GSH transferase genes (reference figure I-6) of divergent sequence and unknown function [20]. For example, the genome of *E. coli* K-12 encodes nine cytosolic GSH transferase homologues while *Mesorhizobium loti* encodes 22 cytosolic GSH transferase homologues. Bacterial GSH transferases have described roles in the degradation of xenobiotics, cellular protection against chemical and oxidative stress, and resistance to antibiotics [9]. They are also involved in metabolic processes such as the biotransformation of dichloromethane, the degradation of lignin and atrazine and the reductive dechlorination of pentachlorophenol [20]. Bacterial GSH transferases can also take on non-enzymatic roles, act as transcription factors and regulate the processes of transcription [27,59,61,62,74-76]. Clearly, an annotation as a GSH transferase is not sufficient to describe the physiological function of a protein in an organism.

Due to an overwhelming amount of sequence information, enzymologists need to integrate bioinformatics with functional and structural information in the context of the whole enzyme superfamily. The integration of bioinformatics is a powerful tool for detecting functional and structural relationships between genes and the understanding of

evolutionary relationships in gene families [104]. Analysis of an entire protein superfamily also allows us to identify regions of sequence and structure space that are unexplored.

This report focuses on identifying and characterizing the functional and structural properties of an outlier in the GSH transferase superfamily. Outliers are members of the GSH transferase superfamily that do not fall into the traditional classes. YqjG is a GSH transferase homolog encoded in the *E. coli* genome. YqjG is annotated in UniProt as a member of the omega class of GSH transferase proteins. Omega class enzymes have previously been described in eukaryotes, and possess dehydroascorbate and thiol transferase activity and catalyze the reduction of monomethylarsonate [24]. In this report we describe the functional properties of YqjG as well as the GSH bound crystal structure of the enzyme at 2.2 Å resolution. Using bioinformatics analysis we determined that the YqjG protein 'clusters' with a subset of other proteins from eukaryotes and bacteria that are distinct from the omega class. Proteins in this new class may function as *S*-glutathionyl-chloro hydroquinone reductases [97,103] and fold in a unique way as to facilitate a novel dimer interface.

Materials and Methods

Materials

E. coli K-12 genomic DNA was from ATCC (Manassas, VA). Primers were custom ordered from Sigma (St. Louis, MO). Restriction enzymes were from New England Biolabs (Ipswich, MA). The pET20b(+) vector was from EMB Chemicals, Inc (Gibbstown, NJ), and the BL21(DE3) cells were from Agilent (Wilmington, DE). Luria broth, ampicillin, NADPH, IPTG, HEPES, DTT, EDTA, GSH and streptomycin sulfate

were from Research Products International (Mt. Prospect, IL). The DEAE resin was from GE Healthcare (Piscataway, NJ) and the HA resin was from Biorad (Hercules, CA). The sodium chloride and potassium phosphate were from Fisher (Pittsburgh, PA). The GSH reductase from *S. cerevisiae* was from Roche (Indianapolis, IN). Finally, all the amino acids, cumene hydroperoxide, 1-chloro-2,4-dinitrobenzene, hydroxyethyl disulfide, cystine, fufuryl disulfide, D-penicillamine and D-pantethine disulfide were all from Sigma Aldrich (St. Louis, MO).

Methods

Cloning, Expression and Purification of YqjG

The *yqjG* gene was PCR amplified from the *E. coli* K-12 genomic DNA using primers containing a 5' restriction site for *Nde*I and a 3' restriction site for *Xho*I. (Forward primer 5'- GGG GGG GGG CAT ATG GGT CAA CATG ATT GAC GGC GTC TG, Reverse primer 5'- CAT GGA CGA GAT GTT CGC TTC GGT TAA CTC GAG TTT TTT) The purified PCR product was digested and ligated into a pET20b(+) vector (Novagen). The correct *yqjG* gene sequence in the pET20b vector was confirmed by sequencing, using the T7 primer. The *E. coli* yqjG-pET20b plasmid was used to transform BL21DE3 cells. A single colony was used to start a 100 mL overnight culture of LB media supplemented with 0.1 mg/mL ampicillin, which was grown at 37°C with shaking at 225 rpm. The overnight culture was diluted 100x into 1 L of LB media and grown under the same conditions until the optical density at 600 nm reached 0.8. The cultures were then cooled to 18°C, induced with 0.5 mM IPTG and allowed to express for 18 hours. The cells were harvested by centrifugation and the cell pellet was frozen at -

20°C. The frozen cell pellet was resuspended in buffer containing 20 mM HEPES pH 7.0, 1 mM DTT and 1 mM EDTA. The cells were lysed by sonication (4x, 3 min) and then centrifuged at 31,000 x *g* for 30 minutes. The supernatant was treated with streptomycin sulfate (10 mg/mL) and then centrifuged as before. The resulting supernatant was dialyzed in buffer containing 20 mM HEPES (pH 7.0), 1 mM DTT and 1 mM EDTA. The YqjG protein was purified using weak anion exchange with a DEAE sepharose column equilibrated in the dialysis buffer. The protein was eluted from the column with a linear gradient of 0-400 mM NaCl. Fractions containing the protein were confirmed by absorbance at 280 nm as well as SDS-PAGE. These fractions were pooled and dialyzed in 20 mM KH_2PO_4 (pH 7.0) buffer containing 1 mM DTT and 1 mM EDTA. The protein was then loaded onto an HA column equilibrated in the dialysis buffer and fractions were collected immediately as the protein eluted in the flow through. Fractions containing the purified protein were determined by SDS-PAGE, concentrated, and frozen at -80°C.

Expression and Purification of the Selenomethionine Derivative of YqjG

The selenomethionine derivative of YqjG was expressed from the yqjG-pET20b plasmid in BL21DE3 cells. An overnight culture of M9 media containing 0.1 mg/mL ampicillin and 5 mg/L thiamine was grown at 37°C. The overnight culture was diluted 100x into 1L of M9 media and grew at 37°C to an optical density of 0.8. The media was cooled to 18°C and then supplemented with 100 mg each of lysine, threonine and phenylalanine and 50 mg each of leucine, isoleucine, valine and selenomethionine. After 20 minutes, 0.5 mM IPTG was also added and the selenomethioneine protein was

expressed for 16 hours. The purification of selenomethine-YqjG was carried out using the same procedure as for the native enzyme.

Steady-State Enzyme Kinetics

Glutathione Transferase Activity Assay

The GSH transferase activity of YqjG was assayed with the electrophilic substrate, CDNB. Reactions containing 1 μM YqjG, 1 mM CDNB and 10 mM GSH were followed with a Perkin-Elmer Lambda 45 Spectrophotometer at 25°C in 1 mL cuvettes. Very little activity was observed over the spontaneous reaction, preventing determination of steady-state kinetic parameters.

Glutathione-dependent Peroxidase Activity Assays

The GSH-dependent peroxidase activity of YqjG was evaluated with cumene hydroperoxide using a coupled assay with GSH reductase from yeast as described [107]. Reactions contained 1 mM EDTA, 6 μg/mL GSH reductase, 2 mM GSH, 0.15 mM NADPH, and 1 μM YqjG and were initiated by 1.2 mM CuOOH. No activity over background was observed for YqjG in this assay.

Disulfide-bond Reductase Activity Assays

The disulfide-bond reductase activity of YghU was measured with a panel of disulfide substrates including hydroxyethyl disulfide, cystine (cysteine disulfide), fufuryl disulfide, and D-penicillamine and D-pantethine disulfide using the coupled assay previously described [86,87]. The data were fit to the Michaelis-Menten equation 3 using

GraphPad Prism and values of V_{max} and K_m were determined. The k_{cat} value was determined using equation 4.

Steady State Fluorescence Titration

Dissociation constants for the complex of YqjG with GSH were determined by titration of the intrinsic fluorescence of the protein. Measurements were made on a SPEC Fluorolog-3 spectrofluorimeter in constant wavelength mode. The samples were excited at 295 nm and the emission was collected at 340 nm over 30 seconds. The change in fluorescence intensity of 1 μM YqjG in 100 mM KH_2PO_4 buffer at pH 7.0 was measured after a 5 minute preincubation with GSH to a final concentration of 3.0 mM. Data were corrected for dilution caused by the addition of the ligand and for inner filter effect using equation 1. The data were fit to a one-site binding model according to equation 2.

Crystallization and Determination of the Structure of YqjG

Crystallization of YqjG

Crystallization conditions were surveyed using the hanging drop method of vapor diffusion and commercially available crystallization screens. Both the native enzyme as well as the enzyme in the presence of 10 mM GSH was tested. Diffraction quality crystals of the enzyme in the presence of GSH were obtained with a reservoir solution of 1.6 M ammonium sulfate, 0.1 M MES (pH 6.5), and 10% v/v 1,4-dioxane. The drop contained a 1:1 mixture of YqjG (66 mg/mL) with 10 mM GSH, in a buffer containing 20 mM KH_2PO_4 (pH 7.0), 1 mM DTT and 1 mM EDTA. Crystals grew within 72 hours and belonged to the $P3_12_1$ space group with unit cell dimensions $a = b = 147.9$ Å, $c = 108.7$

Å. The asymmetric unit contained the complete YqjG dimer. The crystals were transferred to a cryo-protection solution containing 30% v/v glycerol, 1.6 M ammonium sulfate, 0.1 M MES (pH 6.5), and 2% v/v 1,4-dioxane before freezing.

X-ray Data Collection and Processing

The X-ray data from flash-frozen crystals were collected at the Life Sciences Collaborative Access Team Beamline at the Advanced Photon Source, Argonne National Laboratory, Argonne, IL. The native data were collected on 21-ID-G while the SAD data was collected on 21-ID-D. Each data set were collected from a single crystal. Native and SAD data sets were processed using HKL2000 [90]. The data collection statistics are presented in Table IV-1.

Structure Determination and Refinement

Data from the peak wavelength were output using the 'no merge original index' option in HKL2000 and used as input to the program SOLVE [108] using data to 2.48 Å. SOLVE was able to locate all eight selenium positions in the asymmetric unit and RESOLVE [80] was able to build approximately 81% of the model. After one round of refinement in refmac5 [82] the model phases were used as initial phases for ARP/wARP [94] and extended using the inflection data to 2.17 Å. ARP/warp was able to build the model to 98% completeness. The model was completed by cycles of examining and building in the graphics display program COOT [83], and refining in REFMAC5. The model included 5,023 non-hydrogen protein atoms, 2 GSH molecules (40 atoms) 244 water molecules and 9 sulfur molecules. Better electron density was observed in chain A (average B value

Parameters	Values	
	SAD phasing data	Native data
Space group	$P3_121$	$P3_121$
Unit cell dimensions (Å)	a = b = 147.9, c = 108.7	a = b = 147.2, c = 108.9
Wavelength of data collection (Å)	0.97938	0.97856
Resolution limits (Å)	50-2.45 (2.49-2.45)	50-2.17 (2.21-2.17)
No. independent reflections	50,652	71,532
Completeness (%)	100.0 (100.0)	99.4 (92.2)
Redundancy	22.6 (22.6)	5.3 (2.9)
Average *I*/average $\sigma(I)$	29.8 (5.6)	18.5 (2.1)
R_{sym} (%)	0.105 (0.736)	0.069 (0.488)

Table IV.1: X-ray data collection statistics for the structure of YqjG. Values for the highest resolution bin are given in parentheses.

for chain A = 38.8, chain B = 47.4). The A chain is missing residue 1 and residues 11-28 as well as residue 328, the C-terminal amino acid. The B chain is missing residues 1-2, 12-29 and 328. Relevant refinement statistics are presented in Table IV-2.

Cytoscape Analysis of the Glutathione Transferase Superfamily

Full-length sequences were obtained for a nonredundant set of the members of the Pfam GST_N and GST_C families. All sequences with lengths under 150 amino acids were removed. Each sequence in this set was used as a query against a custom database containing only the sequences in the set. BLAST e-values of $1e^{-14}$ or more significant were used to create a cytoscape metanode network. Each node in the network represents a set of sequences where each sequence in the metanode is connected to at least one other by an e-value of $1e^{-100}$ or more significant. Each edge in the network represents the most significant BLAST connection between any two sequences in the appropriate metanodes. The resulting networks were visualized in Cytoscape version 2.7.0 using the organic layout.

Results and Discussion

Characterization of the substrate binding and enzyme activity of YqjG

Full length, native, YqjG protein was expressed in *E. coli* and purified using traditional methods. The dissociation constant for GSH was determined by fluorescence titration (K_d = 0.16 ± 0.04 mM). The GSH transferase activity of YqjG was assayed with the model substrate, CDNB. Very little activity was observed with YqjG over the spontaneous reaction, preventing the measurement of steady-state kinetic parameters.

Parameters	Values
Resolution limits (Å)	42.5-2.2
R-factor (overall) (%)/No. of reflections	18.0/65002
R-factor (free) (%)/No. of reflections	21.2/3293
No. of protein atoms	5003
No. of heteroatoms[a]	329
Average B values	
Protein atoms ($Å^2$) chain A, chain B	38.8, 47.4
Glutathione ($Å^2$)	63.6
Solvent ($Å^2$)[b]	50.2
Weighted root-mean-square deviation of ideality	
Bond lengths (Å)	0.016
Bond angles (deg)	1.8

[a] These include 244 water molecules, 2 glutathione molecules and 9 sulfates.

[b] Solvent includes 244 water molecules and 9 sulfates.

Table IV.2: Least-squares refinement statistics for the structure of YqjG. Values for the highest resolution bin are given in parentheses.

This lack of GSH transferase activity for YqjG with CDNB in our assay agrees with those conducted by Xun, L [97]. The GSH-dependent peroxidase activity of YqjG was assessed with cumene hydroperoxide using a coupled assay. No activity was observed. Initial experiments indicated that the enzyme possesses modest disulfide-bond oxidoreductase activity. This activity was surveyed with a panel of disulfide substrates. Kinetic constants were determined for hydroxyethyl disulfide and cystine as summarized in Table IV-3, little if any activity was observed with fufuryl, D-penicillamine or D-pantethine disulfide.

Structural Characterization of YqjG

The 2.2 Å resolution structure of YqjG in the presence of GSH was determined by SAD phasing. A ribbon representation of the structure of the dimer of YqjG is shown in Figure IV-2. The protein adopts a typical GSH transferase fold with an N-terminal thioredoxin-like domain and a C-terminal all α-helical domain. The sequence of YqjG is much longer (approximately 100 amino acids) than other *E. coli* GSH transferase paralogs. These additional amino acids are located at the N-terminus, between the second β-sheet and second α-helix of the thioredoxin domain and at the C-terminus of the protein. At the N-terminus residues 2-10 form a β-turn structure. Residues 11-29 are disordered, no electron density was observed for this region of the protein. Residues 29-53 form a random coil structure that lies on the surface of the protein and contacts the C-terminal most β-sheet and α-helix of the thioredoxin domain. Therefore, Arg-54 is the first amino acid of β-sheet 1 in the thioredoxin domain. There is a 30 amino acid extension after the second β-sheet and before α-helix 2 of the thioredoxin domain. At the

	Hydroxyethyl disulfide	Cystine
k_{cat} (sec^{-1})	1.2 ± 0.1	0.36 ± 0.08
$K_M^{disulfide}$ (mM)	0.29 ± 0.11	0.12 ± 0.05
$k_{cat}/K_M^{disulfide}$	(4.0 ± 1.6) x 10^3	(3.0 ± 1.3) x 10^3

Table IV.3. Steady state kinetic parameters for the reduction of 2-hydroxyethyldisulfide and cystine by YqjG (0.5 μM).

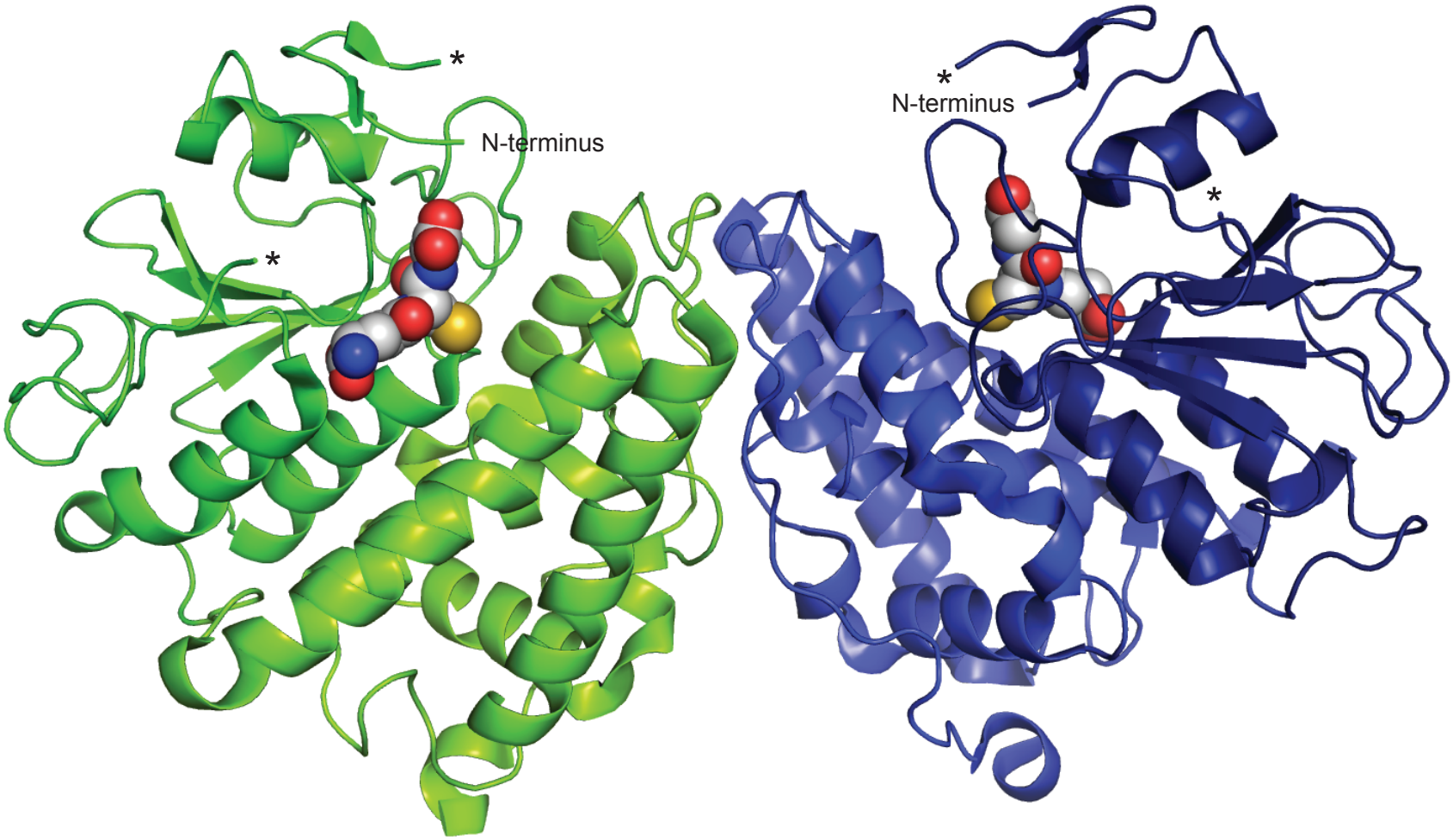


Figure IV-2. Ribbon diagram of the structure of the dimer of YqjG with one molecule of GSH bound in each active site. The two subunits are shown in green and blue, respectively. The molecules of GSH are shown in space filling representation. The astricks denote the continuation of the structure after the missing residues in the N-terminus.

C-terminus is a 40 amino acid extension that includes an additional 10 amino acid α-helix, is similar to the additional helix observed in omega, tau and delta class GSH transferases.

One molecule of GSH is bound to YqjG per subunit. The details of the potential hydrogen bonding interactions between the protein and GSH are shown in Figure IV-3. The side chain carboxylate of Glu-148 and the hydroxyl group of Ser-149 form several hydrogen bonding interactions with the γ-glutamyl portion of GSH. This GSH recognition motif is conserved in many GSH transferases (reference Figure II-5) The glycine carboxylate of GSH forms hydrogen bonding interactions with arginine 130. The ε^1 nitrogen of Trp-96 hydrogen bonds to the carbonyl oxygen of the cysteine moiety. Use of a tryptophan in this capacity is not typical of most GSH transferases. However, multiple sequence alignments show that a tryptophan residue is highly conserved in this subfamily of proteins and may be a shared mechanism of GSH binding in this new class [103].

Interestingly, additional density (present at 4σ in an F_0-F_c map) was observed near the bound GSH in the active site of the enzyme (Figure IV-4). We were not able to fit this density to anything in the protein buffer or crystallization conditions. Mass spectrometry experiments at the Institute of Genomic Biology (University of Illinois) have thus far failed to positively identify the unknown ligand. Using MALDI mass spectrometry, the protein can be observed in 4 states: (1) an apo form, (2) a +150 Da form, (3) a +305 Da (GSH) form and (4) a smaller form that appears to be an apo, +150, +305 Da form. However, neither MALDI nor ESI experiments have been successful in elucidating a small molecule in the 150 Da range that is liberated by denaturing the protein. However,

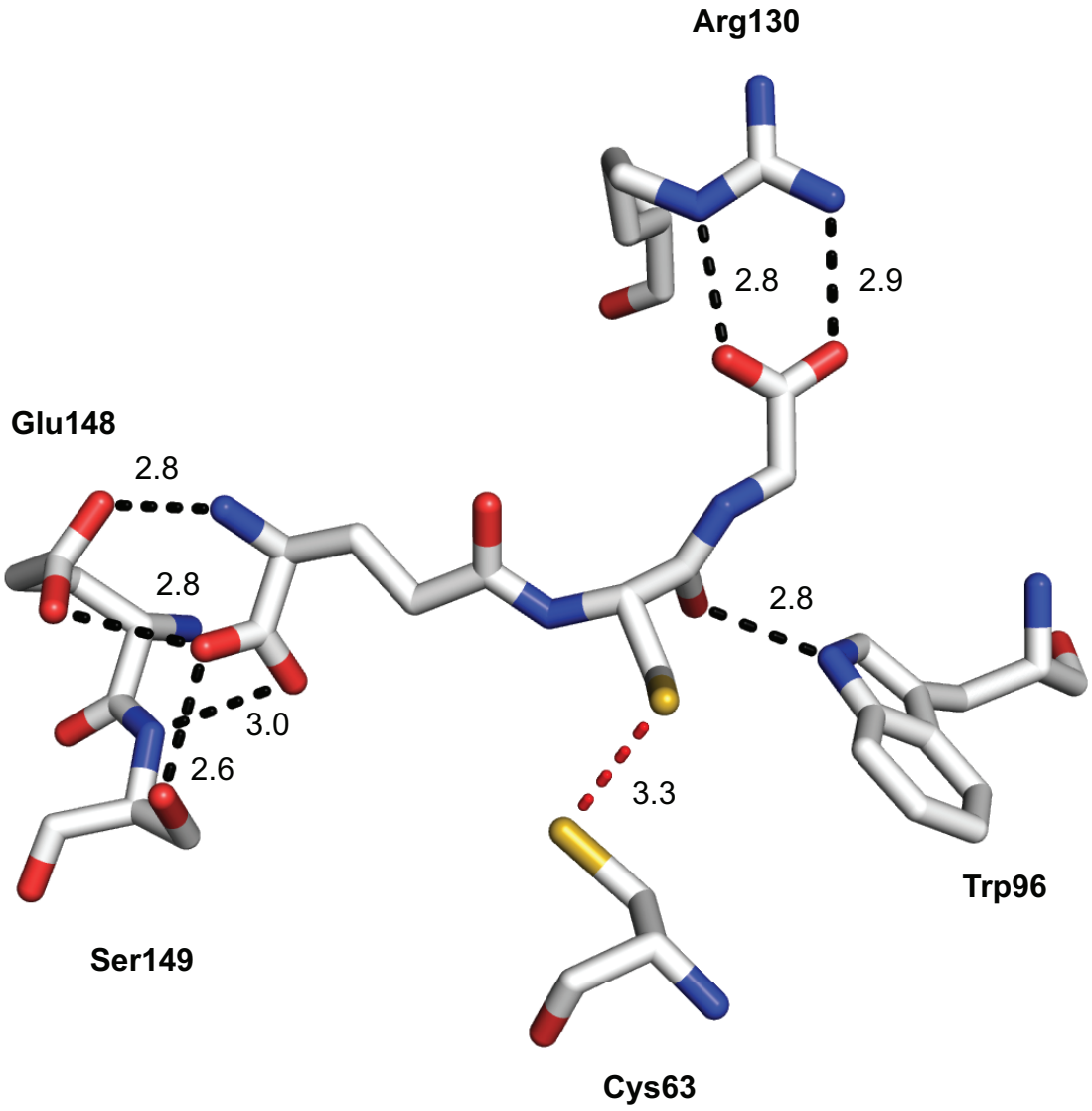


Figure IV-3 Details of the GSH binding site of YqjG. Potential hydrogen bonding interactions are shown as dashed black lines with distances shown in angstroms (Å). The distance between the active site cysteine 63 and the sulfur of GSH is also indicated.

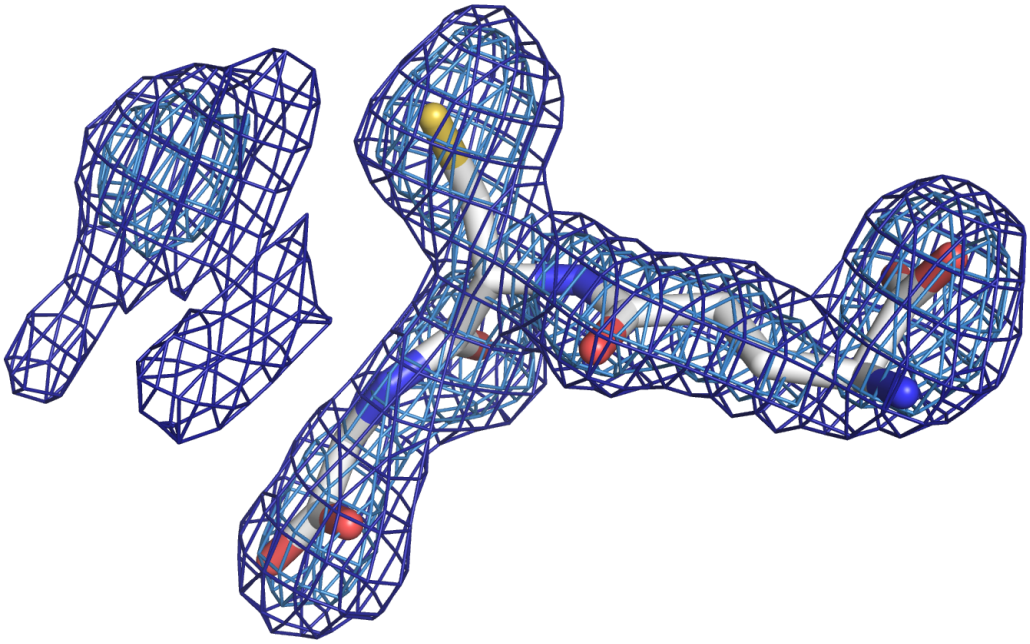


Figure IV-4 Omit map of the electron density for the molecule of GSH and additional, unknown density observed in the active site of YqjG. The map, contoured at 4σ (outer, dark blue cage) and at 2.5σ (inner, light blue cage), was calculated with coefficients of the form F_o-F_c in which the observed and calculated structure factor amplitudes were calculated from the model lacking the coordinates for the molecule of GSH. The model for GSH is shown in stick representation.

an interesting consequence of this analysis has been identification of a low-occupancy protein-GSH adduct, most likely at the active site Cys-63. Under denaturing conditions, a lower intensity set of peaks was observed that correspond to the apo protein mass plus GSH minus $2H^+$. These data suggests a covalent protein-GSH linkage that may indicate a catalytic role for Cys-63 (Figure IV-5).

Certainly, the most interesting feature of the YqjG structure is the dimer interface. Most canonical, soluble GSH transferases are dimeric proteins, which utilize residues along two helices (one from thioredoxin domain, one from all α-helical domain) for the formation of the dimer interface (Figure IV-6) [9]. In contrast, YqjG utilizes amino acids located on loops between helices in the all α-helical domain and at the C-terminal tail of the protein to form the dimer interface (Figure IV-7). This results in a splayed dimeric structure in which the α-helical domains are situated at the dimer interface and the thioredoxin domains are at opposite ends of each other. There are now three available structures of GSH transferase proteins that form this unique dimer interface, the Q8NR03 protein from *Corynebacterium glutamicum* (PDB code: 3M1G, unpublished), the PhGSTO1 protein from *Phanerochaete chrysosporium* (PDB code: 3PPU, reference 103) and YqjG (Figure IV-8).

Analysis of these three structures reveals the most important features of the GSH binding site and dimer interface in this novel structural class. Multiple sequence alignments of the YqjG, Q8NR03 and B3VQJ7 (PhGSTO1), sequences were created and amino acid conservation was mapped onto the YqjG crystal structure (Figure IV-9). A high degree of conservation is present in the GSH-binding site. This includes the "ACPWA" GSH binding motif and the catalytic cysteine, which is necessary for disulfide

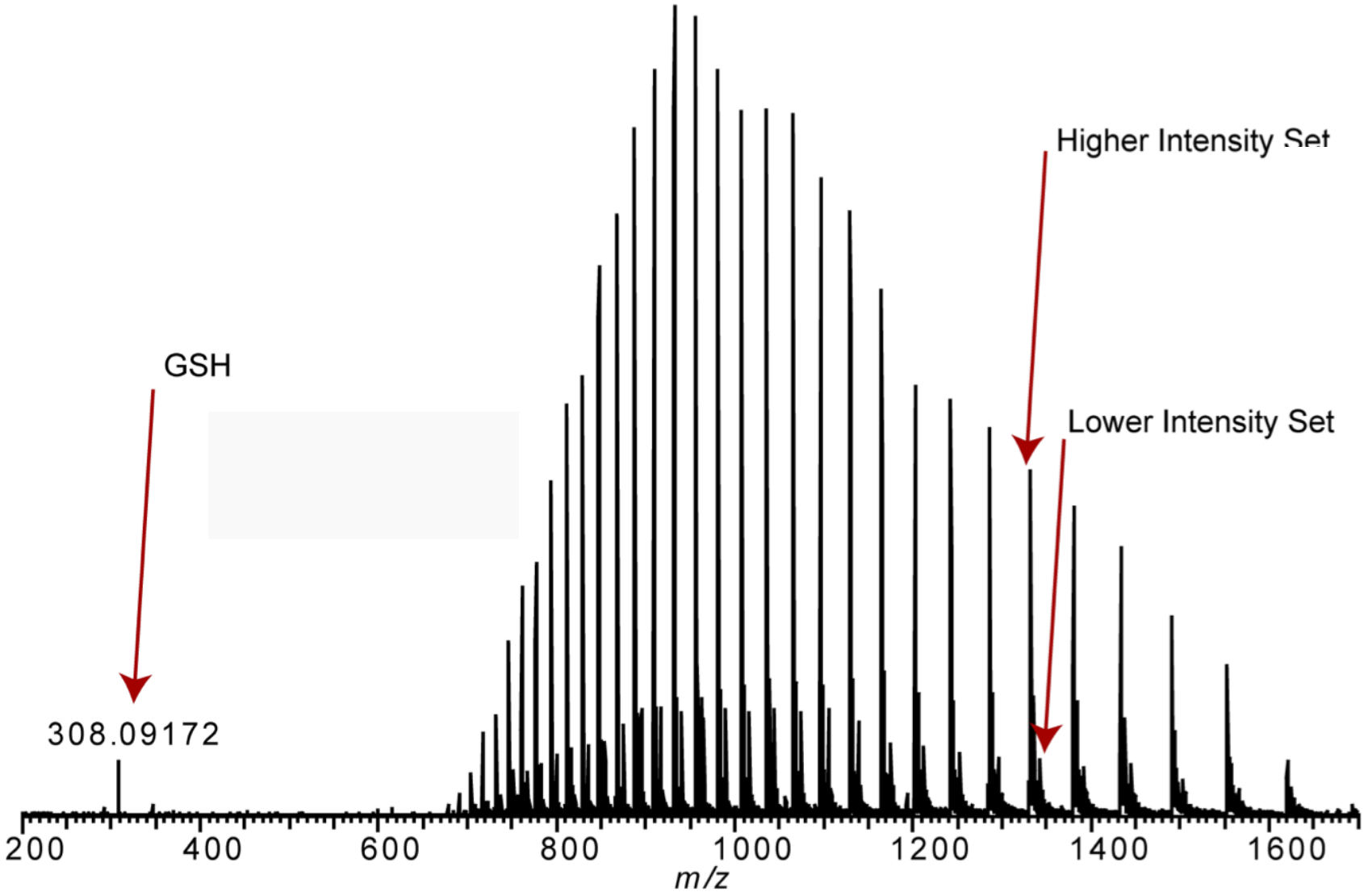


Figure IV-5. Electrosptray ionization mass spectrum of YqjG. Lower intensity peaks correspond to a shifted mass distribution of 306 m/z (+GSH, -2H), indicating a low occupancy protein-GSH adduct. Data collected by Brad Evans PhD, Institute for Genomic Biology, University of Illinois.

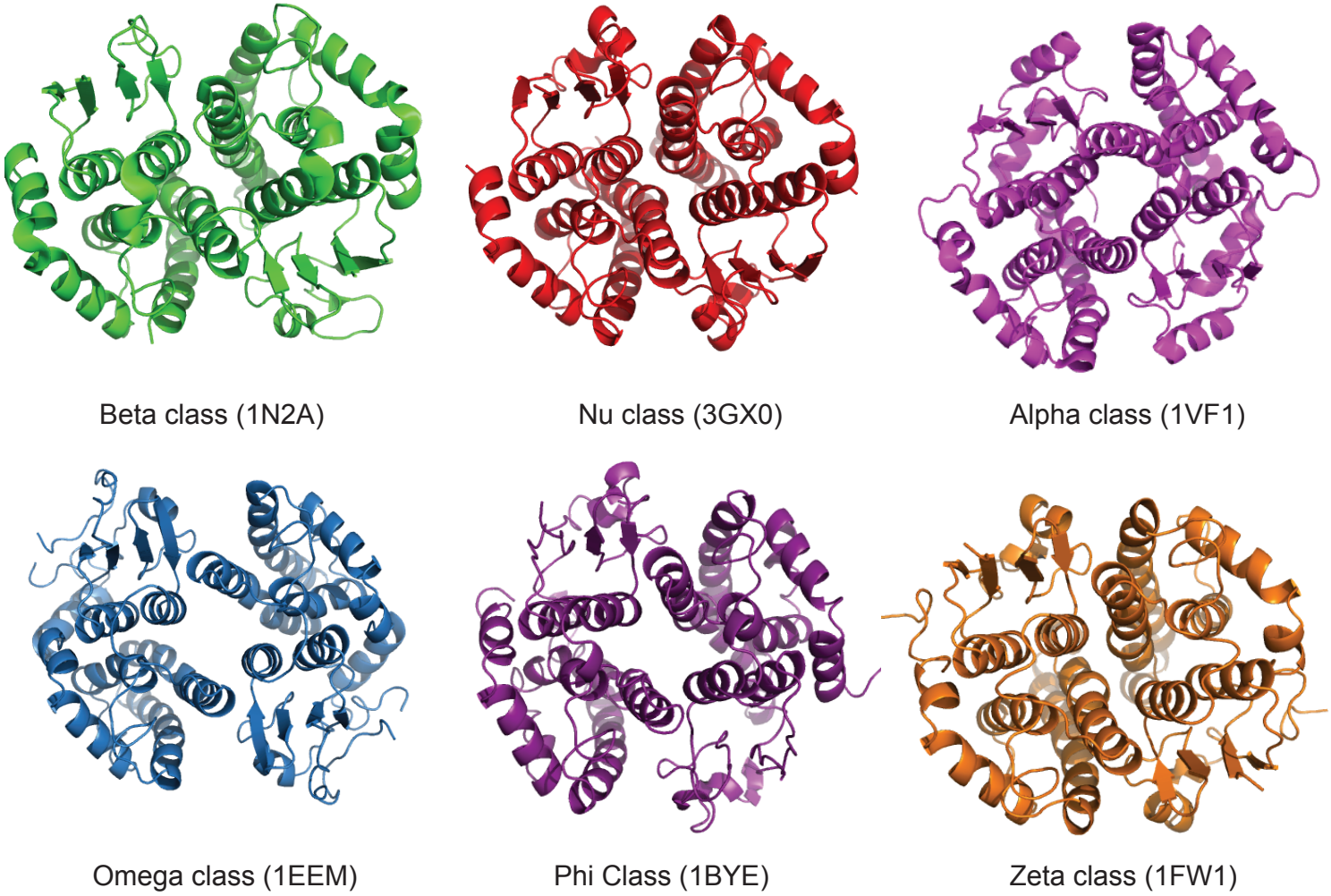


Figure IV-6. 'Traditional' dimer interface formation in several classes of glutathione transferases. The 4-helix bundle at the dimer interface is created from the last α-helix of the thioredoxin domain and the first alpha-helix of the all α–helical domain from each monomer.

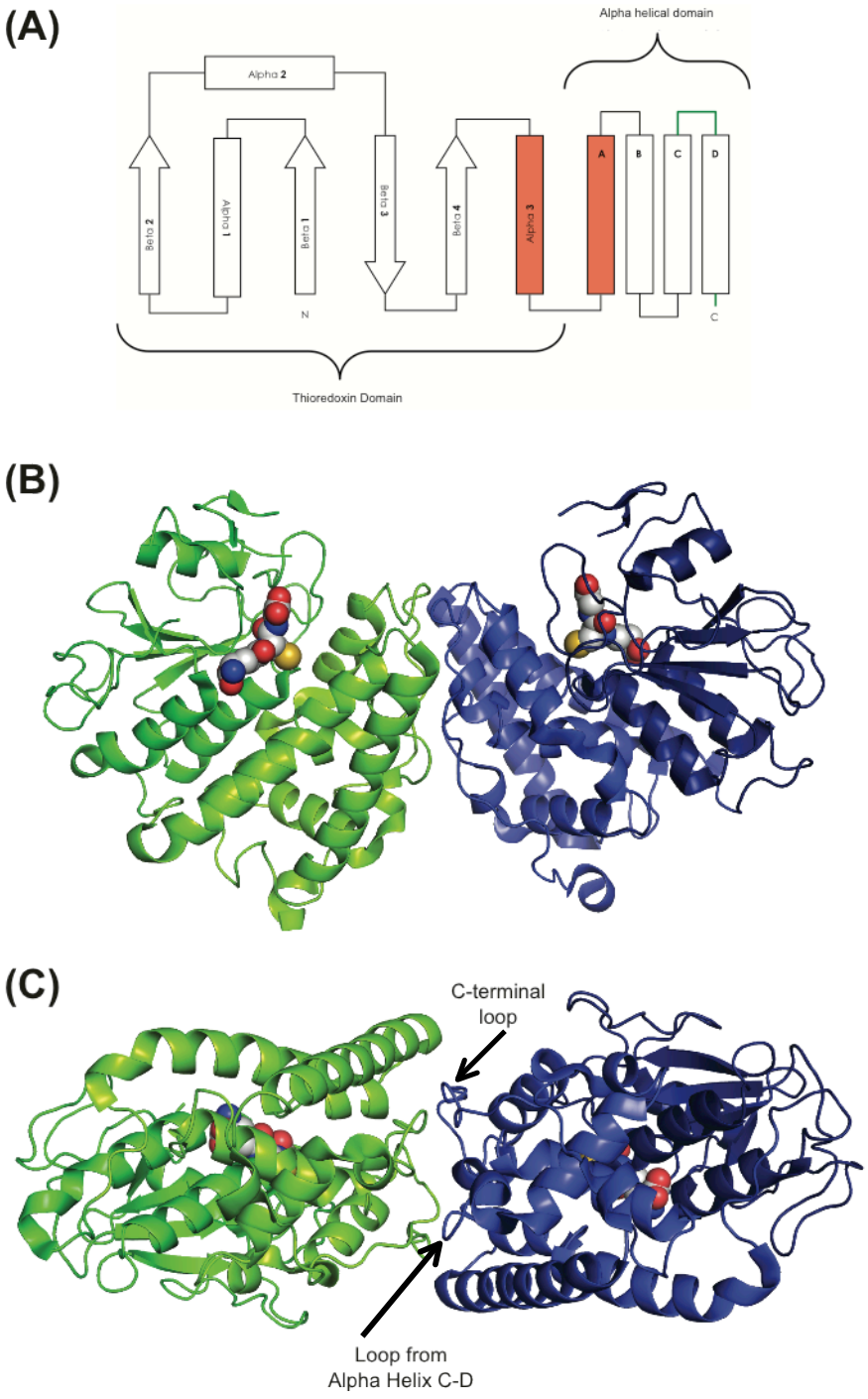


Figure IV-7 (A) Topology map of the glutathione transferase fold. Alpha helices highlighted in red typically form the dimer interface. Loops highlighted in green form the dimer interface in the YqjG, PcGSTO1, and 3M1G structures. **(B)** Structure of YqjG **(C)** Dimer interface formation of YqjG. The structure has been rotated into the plane of the paper 90°.

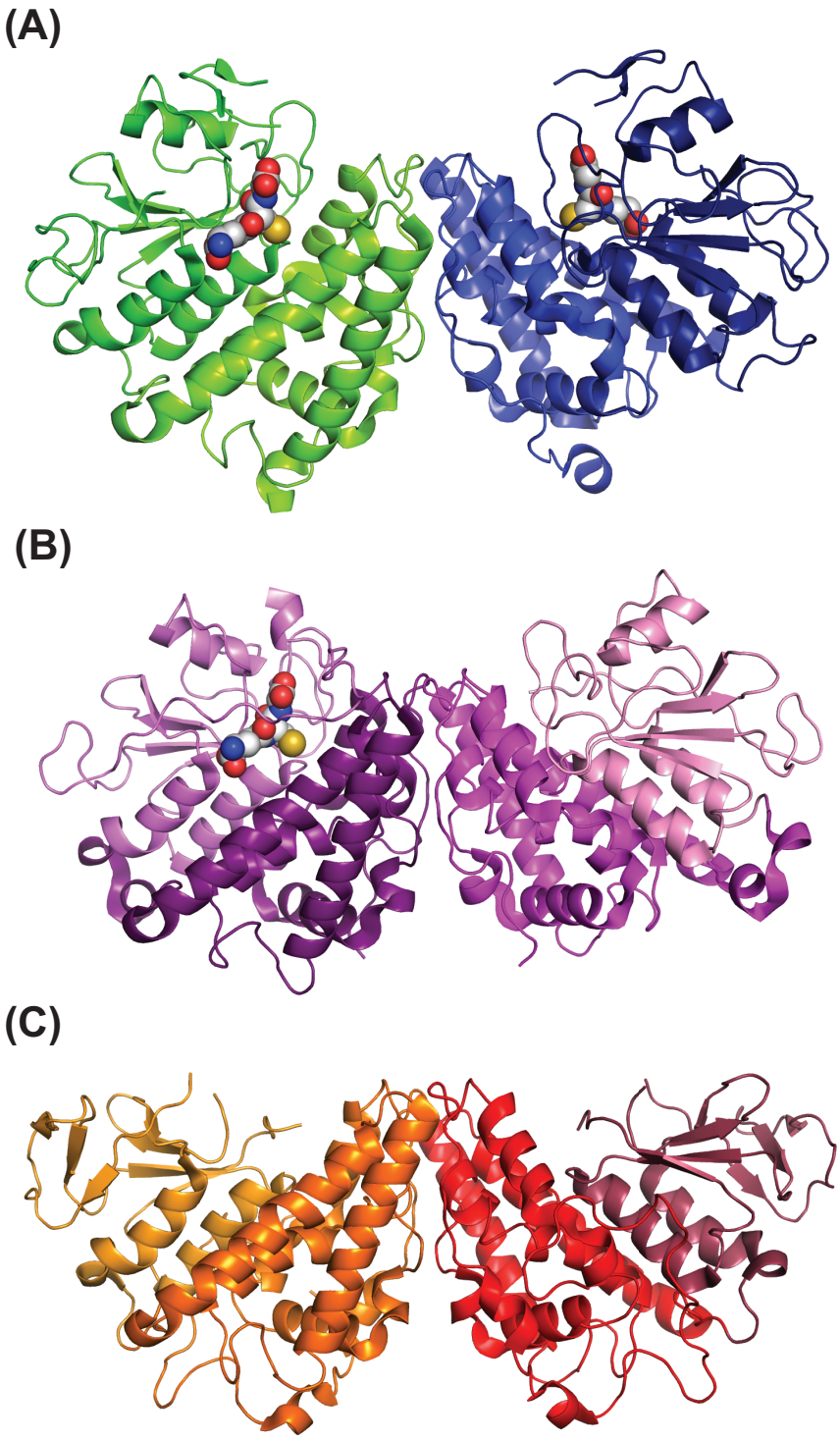
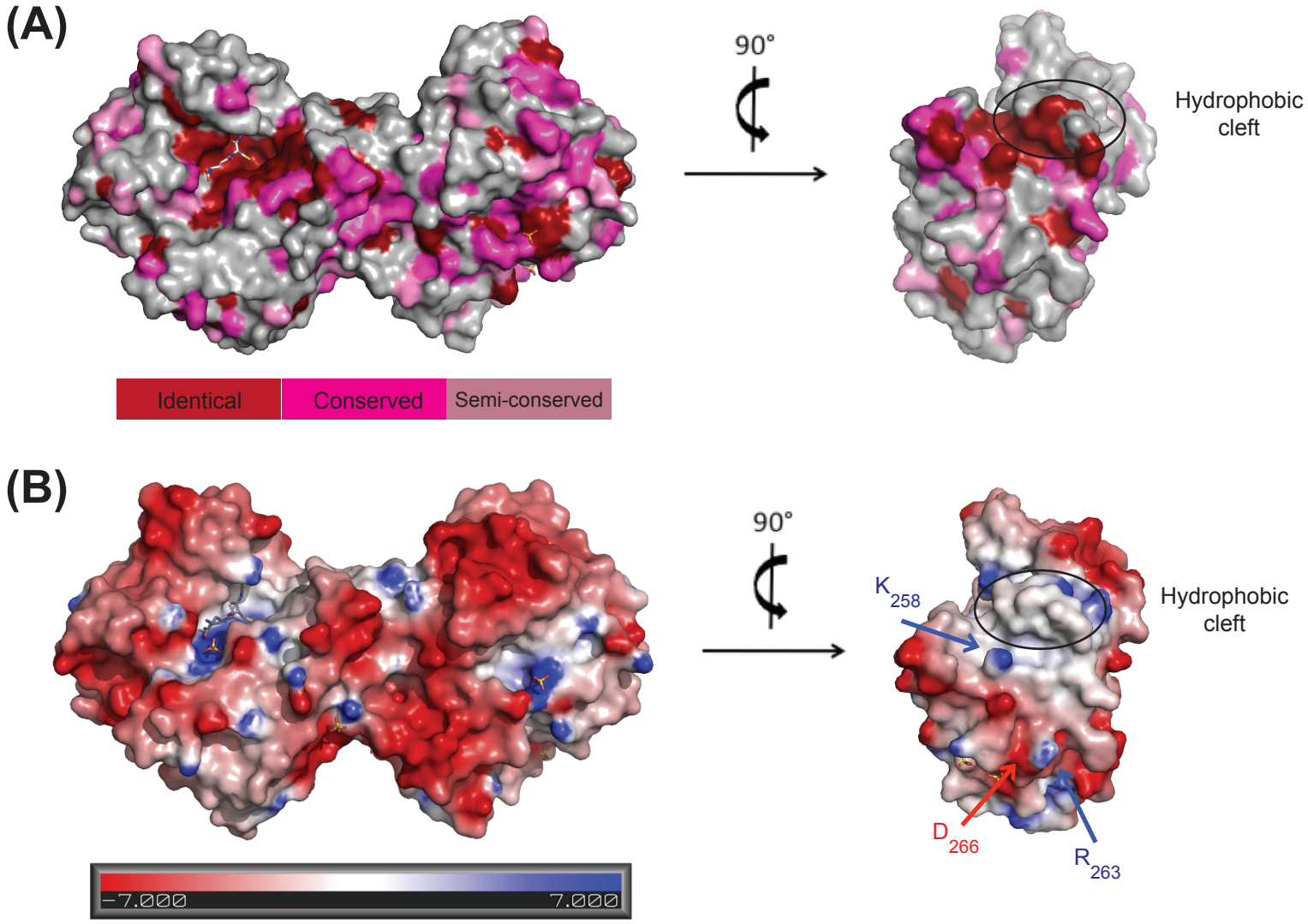


Figure IV-8 Comparison of the three structures in the 'novel' class of glutathione transferases. **(A)** The structure of YqjG from *Escherichia coli* (PDB code: 3R3E) **(B)** The structure of the PhGSTO1 protein from *Phanerochaete chrysosporium* (PDB code: 3PPU, reference 103) **(C)** The structure of the Q8NR03 protein from *Corynebacterium glutamicum* (PDB code: 3M1G).


(A)
90°
Hydrophobic cleft
Identical
Conserved
Semi-conserved
(B)
90°
K258
Hydrophobic cleft
D266
R263
−7.000
7.000

(C)

```
CLUSTAL 2.1 multiple sequence alignment


YQJG_ECOLI        -----MGQLIDGVWHDTWYDTK-----------STGGKFQRSASAFR-- 31
B3VQJ7_PHACH      MTTICLRHLSSGFRRQLLQNNTRVYRSFPARMSFTTGSTLQHQEAQLQGA 50
Q8NR03_CORGL      -----MANTSSDWAGAPQNAS-------------ADGEFVRDTNYID-- 29
                       : :  ..        .              :.. : :.   :


YQJG_ECOLI        -----NWLTADGAPGP------------------------------TGTG 46
B3VQJ7_PHACH      KDETVTQLEQRAAGGASHELQSDISKMKTEDDGSFKRKAASFRNWIQPNG 100
Q8NR03_CORGL      -----DRIVADVPAGSEPIAQE-----------------------DGTF 50
                        :    . *.                               .


YQJG_ECOLI        GFIAEKDRYHLYVSLACPWAHRTLIMRKLKGLEPFISVSVVNPLMLENGW 96
B3VQJ7_PHACH      DFTPEKGRYHLYVSYACPWATRTLIVRKLKGLEDFIGVTVVSPRMGSNGW 150
Q8NR03_CORGL      HWPVEAGRYRLVAARACPWAHRTVITRRLLGLENVISLGLTGPTHDVRSW 100
                   :  * .**:* .: ***** **:* *:* *** .*.: :..*    ..*
                                GSH
                              Binding

YQJG_ECOLI        TFD--DSFPGATGDTLYQNEFLYQLYLHADPHYSGRVTVPVLWDKKNHTI 144
B3VQJ7_PHACH      PFANVDPFPAADSDPLNNAQHVKDLYLKVKPDYDGRFTVPVLWDKHTGTI 200
Q8NR03_CORGL      TFD----LDPNHLDPVLQIPRLQDAYFNRFPDYPRGITVPALVEESSKKV 146
                  .*     :     *.: :   : : *::  *.*   .***.* :: . .:


YQJG_ECOLI        VSNESAEIIRMFNTAFDALG--AKAG-DYYPPALQTKIDELNGWIYDTVN 191
B3VQJ7_PHACH      VNNESSEIIRMFNTAFNHLLPEDKAKLDLYPESLRAKIDEVNDWVYDTVN 250
Q8NR03_CORGL      VTNDYPSITIDFNLEWKQFHR--EGAPNLYPAELREEMAPVMKRIFTEVN 194
                  *.*: ..*   **  :. :    :.  : **  *: ::  :   ::  **


YQJG_ECOLI        NGVYKAGFATSQEAYDEAVAKVFESLARLEQILGQHRYLTGNQLTEADIR 241
B3VQJ7_PHACH      NGVYKSGFASTQKAYEAAVIPLFESLDRLEKMLEGQDYLIGGQLTEADIR 300
Q8NR03_CORGL      NGVYRTGFAGSQEAHNEAYKRLWVALDWLEDRLSTRRYLMGDHITEADIR 244
                  ****::*** :*:*:: *   :: :*  **. *  : ** *.::******
                             +/

YQJG_ECOLI        LWTTLVRFDPVYVTHFKCDKHRISD-YLNLYGFLRDIYQ-MPGIAETVNF 289
B3VQJ7_PHACH      LFVTIVRFDPVYVTHFKCNLRTIRDGYPNLHRWMRKLYWGNPAFKDTCNF 350
Q8NR03_CORGL      LYPTLVRFDAVYHGHFKCGRNKITE-MPNLWGYLRDLFQ-TPGFGDTTDF 292
                  *: *:****.**  ****. .  * :   **  ::*.::   *.: :* :*
                             -/-/  - +
                                    Ionic

                            /  /  // ///
YQJG_ECOLI        DHIRNHYFRSHKTINPTGIISIGP-WQDLDEPHGRDVRFG---------- 328
B3VQJ7_PHACH      EHIKTHYFWSHTFINPHRIVPIGP-IPDILPLDA---------------- 383
Q8NR03_CORGL      TEIKQHYYITHAEINPTRIVPVGPDLSGFATPHGREKLGGSPFAEGVTLP 342
                   .*: **: :*  ***  *:.:**   .:   ..
                             Hydrophobic cleft

YQJG_ECOLI        -----------------
B3VQJ7_PHACH      -----------------
Q8NR03_CORGL      GPIPAGEEVKNPEPFQK 359
```

Figure IV-9. Sequence conservation and electrostatic potential for the three structurally defined 'novel' class proteins. **(A)** Conservation of the YqjG, Q8NR03, and B3VQJ7 sequences, mapped onto the YqjG crystal structure. A high degree of sequence conservation is observed in the GSH binding site (left panel) and at the hydrophobic cleft of the dimer interface (right panel). **(B)** Electrostatic potential surface map for the YqjG structure created using the program DelPhi. Conserved, charged residues at the dimer interface are highlighted in the right panel. (**C)** Multiple sequence alignment of the YqjG, Q8NR03, and B3VQJ7 sequences created using ClustalW2. Highly conserved regions of sequence involved in GSH binding are highlighted in yellow. Interactions at the dimer interface are highlighted in blue and red (hydrogen bonding), green (ionic) and grey (hydrophobic).

bond reductase activity in the PhGSTO1 protein. The conserved glutamate-serine recognition motif, which is responsible for γ-glutamyl recognition of GSH is present in the YqjG and PhGSTO1 structures but missing in the Q8NR03 sequence. A 'hydrophobic cleft' of highly conserved hydrophobic isoleucine, valine, glycine and proline residues forms the top of the dimer interface. An electrostatic surface potential map for the YqjG structure was created using the DelPhi server (Figure IV.10). The map highlights the ionic interaction between Arg-263 and Asp-266 at the base of the dimer interface. In the Q8NR03 structures, this ionic interaction occurs between Lys-266 and Glu-269. No ionic interaction occurs in the PhGSTO1 structure. Conserved hydrogen bonding interactions at the dimer interface occurs between a glutamine and isoleucine and a lysine and an asparagine amino acids in all three structures.

Bioinformatics Analysis of this New Class

Analysis of the global GSH transferase network clearly demonstrates that YqjG belongs to a new class of GSH transferases (Figure IV-10). In total, we have identified 503 sequences (~5% total GSH transferase sequence universe) belonging to this new class. Enzymes in this cluster have described activities as disulfide bond oxido-reductases, *S*-glutathionyl-(chloro) hydroquinone reductases and *S*-(phenylacyl)glutathione reductases [97,103].

These sequences were analyzed for taxonomic distribution (Figure IV-11). Approximately 76% (350 total) of these sequences are bacterial, the majority belonging to the proteobacteria phylum (262 sequnces, 52% of the total). About 10% of the sequences are from cyanobacteria. Eukaryotic sequences account for 23% (115 total).

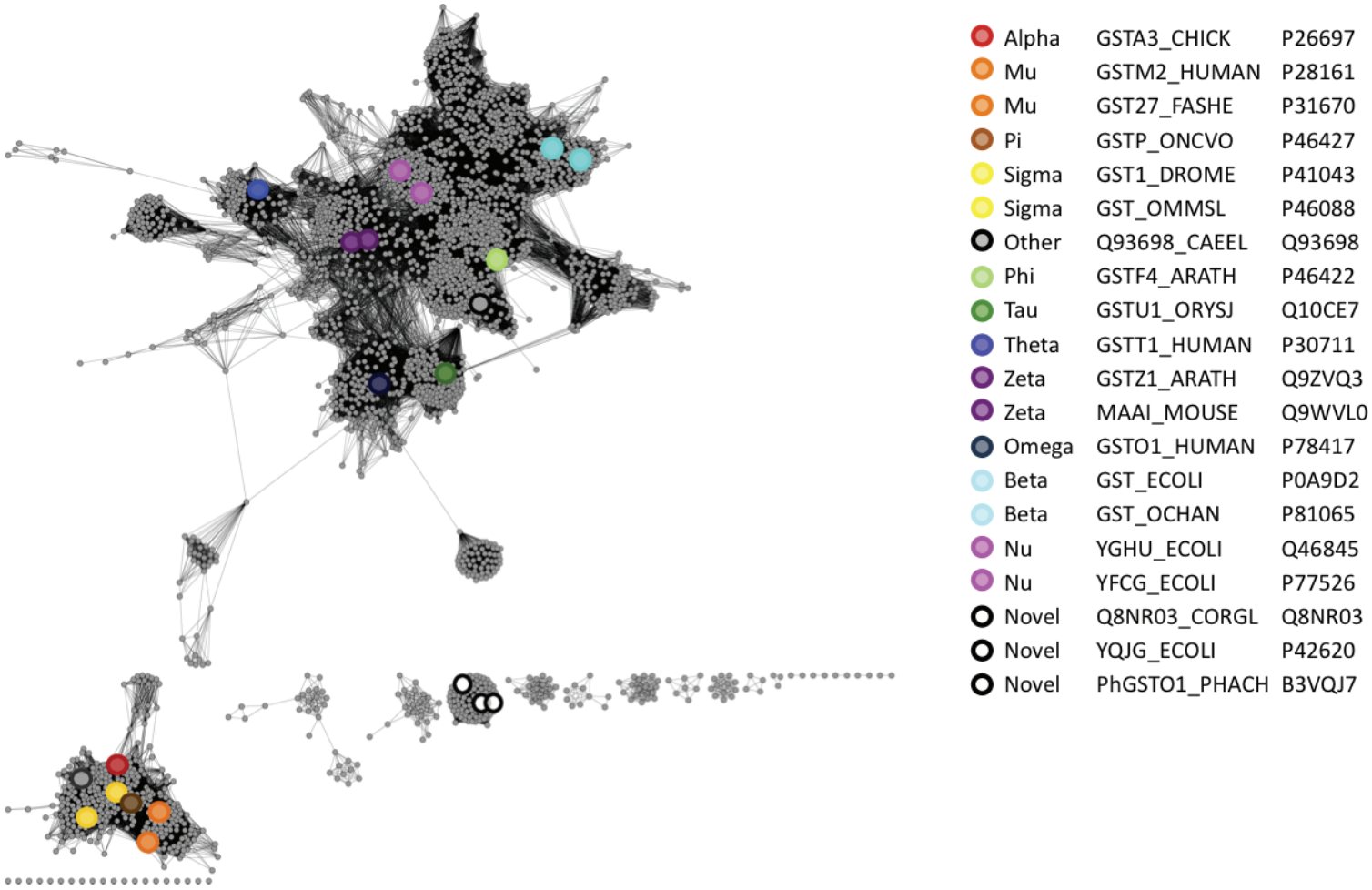


Figure IV-10. YqjG belongs to a novel and unique cluster of glutathione transferases that are distinct from other classes. Overall sequence similarity network containing 2,851 sequences and 97,144 edges. Edges represent BLAST E values of 10-18 or more stringent. Large nodes are colored by the classification of the amino acid sequences in SWISS-PROT and indicate a representative member of each subgroup as indicated in the legend.

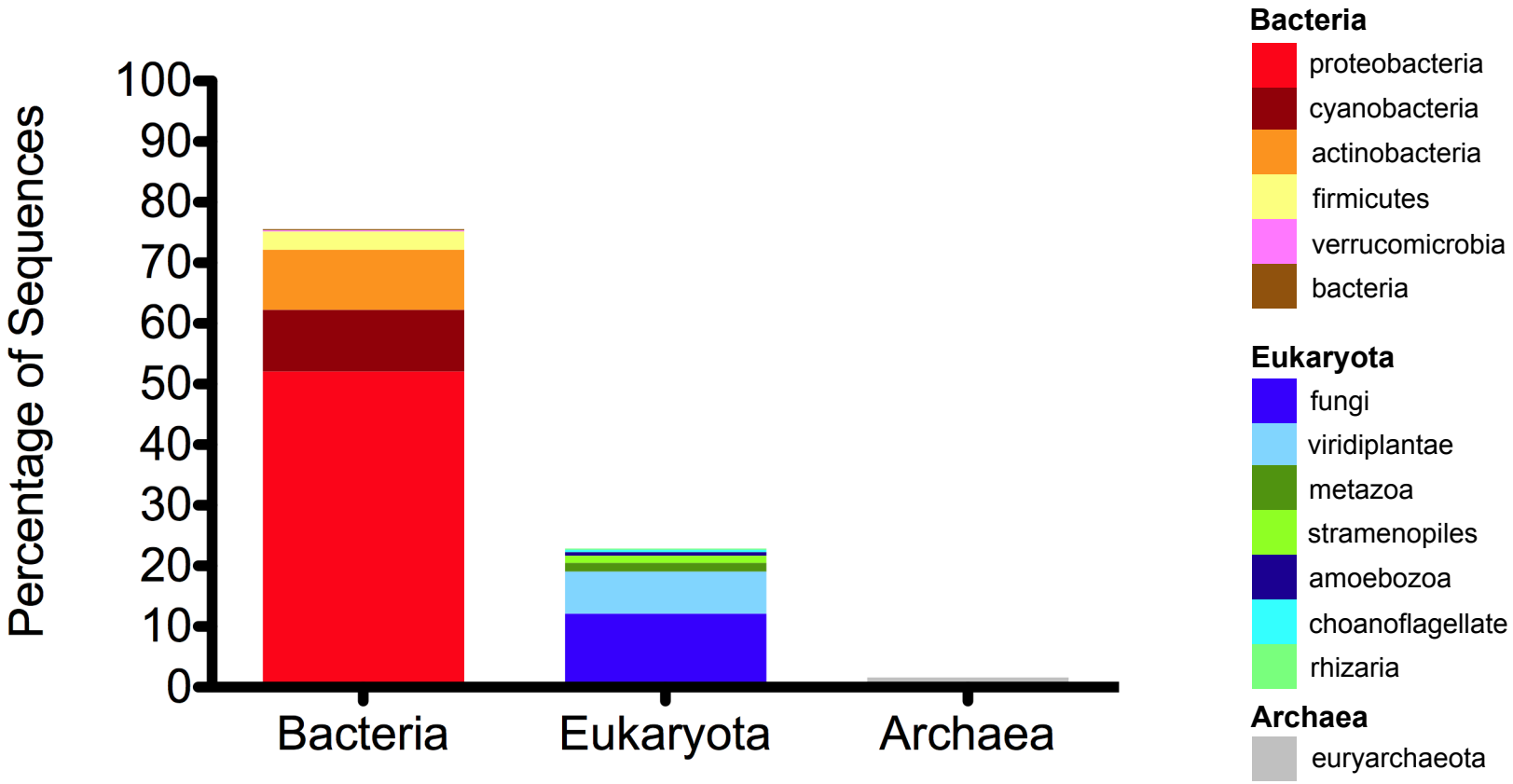


Figure IV-11. Taxonomic distribution of 503 sequences in the 'novel' class cluster. Approximately 76% of these sequences are bacterial, 23% are eukaryotic and < 2% are archaeal.

Eukaryotic sequences are composed primarily of fungi (53%, 61 total) and viridiplanta (7%, 35 total). Only eight sequences were from archaea, the euryarchaeota phylum.

Important aspects in defining this new class are the features of the dimer interface. All three structurally defined proteins in this region form a splayed dimer that has not been observed previously. To address how this interface differs from the classical 'four helix bundle,' a structure based sequence alignment was created from 17 structures representing 12 different GSH transferase classes (Figure IV-12). These structures were compared to the three structures from the novel class. In the traditional dimer, α-helix 3, the terminal helix of the thioredoxin domain, interacts with helix A (the first helix) of the all α-helical domain (reference Figure IV-7). These interactions are predominantly, though not exclusively, hydrophobic. Accordingly, α-helix 3 in the traditional dimer is rich with valine, leucine, isoleucine, and phenylalanine residues. However, α-helix 3 of the new class, which is found on the surface of the protein, contains primarily polar and polar charged amino acid residues (ie. asparagine, glutamate, arginine, serine). Alpha-helix A from the all α-helical domain varies significantly in length and amount of contact with α-helix 3. Further analysis of this region will be necessary to determine all the interactions of this helix in the formation of the traditional dimer.

Perhaps, what is more significant than the lack of conservation between the amino acids in the 'traditional' versus the 'novel' dimer, is the lack of equivalent residues. The C-termini of proteins in the new class form critical hydrophobic contacts at the dimer interface but lack equivalent residues to the other families. Phe-257 and Lys-258, which are invariant at the dimer interface of the new class, do not have equivalent residues. This may be due to the fact that sequences in the new class are much longer (average length

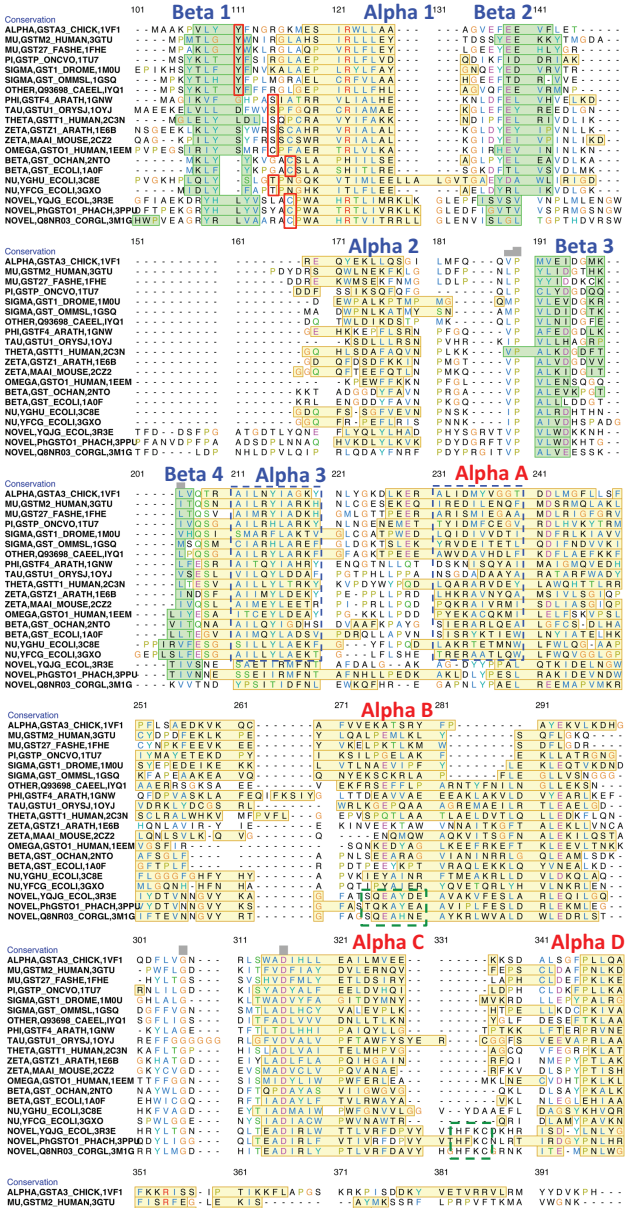


Figure IV-12. Structure based sequence alignment highlights the differences between the traditional four-helix bundle dimer and 'novel', splayed dimer. Residues in alpha helices have a yellow background and those in beta strands have a green background. The residue that interacts directly with the nucleophilic sulfur of GSH is highlighted in red, residues involved in dimer interface formation in the traditional dimer are highlighted in blue, while residues involved in the splayed dimer are highlighted in green.

357 amino acids) than the traditional classes (average length 223 amino acids). Proteins in this novel class may form new structural elements that facilitate the novel dimer interface, rather than repurposing conserved elements of sequence.

To extend this analysis further, we created a multiple sequence alignment between the 3 structures of the novel class and the other 500 sequences in this cluster (Figure IV-13). The consensus sequences for this alignment is shown at the bottom, and positions that are highly conserved (≥ 80%) are highlighted in red. Regions of sequence known to be involved in the dimer interface are underlined in black. While many amino acids are highly conserved in the C-terminus of these proteins, there is a lack of absolute sequence conservation in the regions known to be involved in the dimer interface. Presently, we cannot conclude that all proteins in this new class will adopt the splayed dimer. Considering our lack of structural information (3 structures for ≥ 503 sequences, <1% coverage), additional information will be necessary to determine if the splayed dimer is a hallmark of this new class.

Conclusions

YqjG belongs to a unique cluster of GSH transferases that has not been previously characterized. The protein exhibits modest disulfide bond reductase activity and may have glutathionyl-reductase activity with other substrates. The protein adopts a classical GSH transferase fold but a unique splayed dimer. Proteins in this class are primarily from bacteria, although some are from eukaryotes including fungi, and may constitute a novel class of GSH transferases.

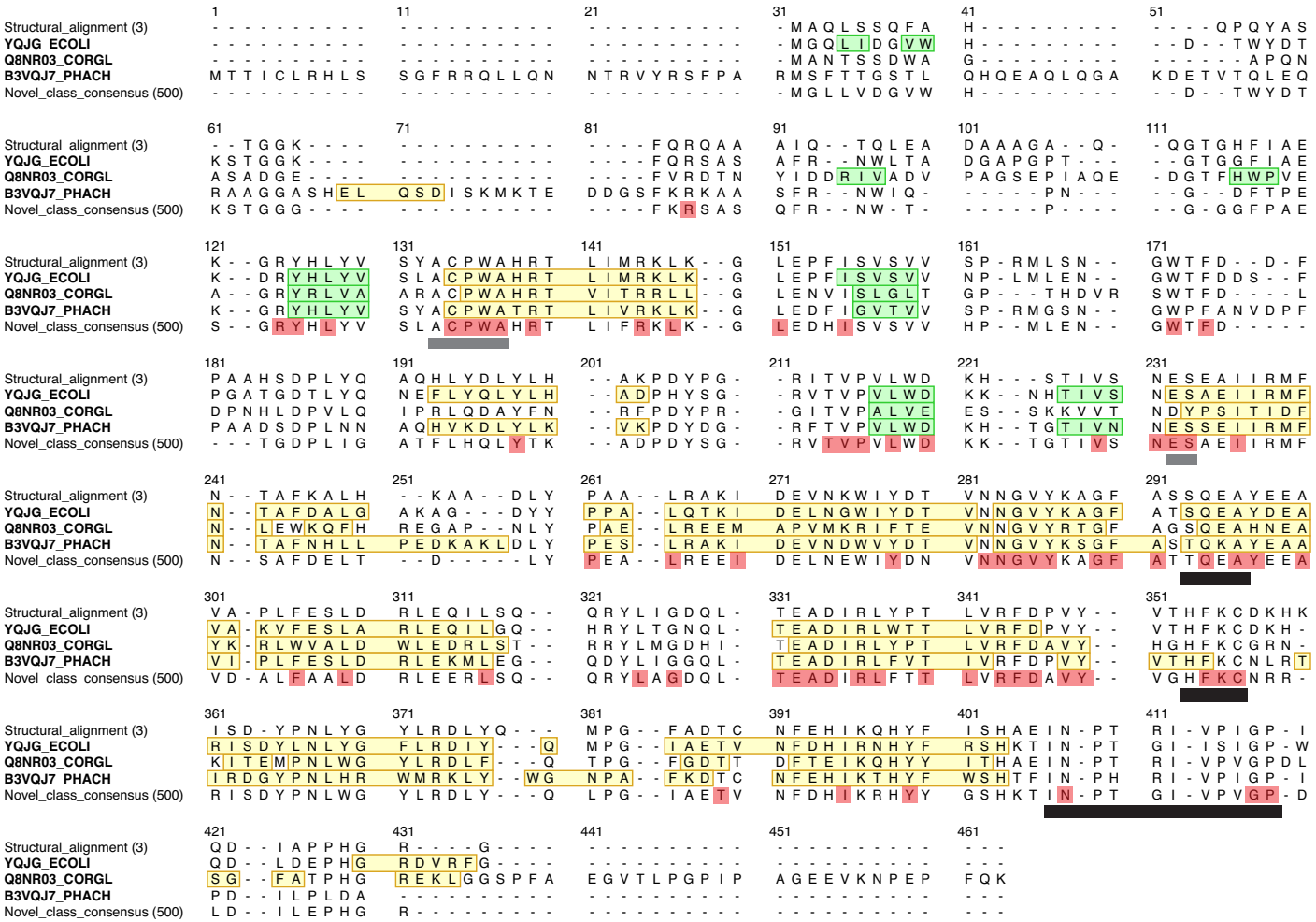


Figure IV-13. Multiple sequence alignment of the YqjG, Q8NR03, and B3VQJ7 sequences compared to the consensus sequence of 500 similar sequences reveal regions of conservation in the C-terminal domain. Residues that are ≥80% conserved are heighted in red. Regions involved in GSH binding are underlined in grey and regions known to be involved in the dimer interface in the 3 known structures are underlined in black.

Future Studies

Proteins in the novel class of GSH transferases described in this report have activities as *S*-glutathionyl-(chloro) hydroquinone reductases and S-(phenylacyl) hydroquinone reductases [97,103]. These include the PcpF protein from *Sphingobium chlorophenolicum*. PcpF is part of the PCP gene cluster, which allows the bacterium to mineralize the environmental pollutant pentachlorophenol. Specifically, the role of PcpF is to remove the GSH moiety from GSH conjugates of TriHCQ and DiHCQ. Specific activities for the PcpF protein and YqjG from *E. coli* and *R. eutropha* have been reported [97]. However, the *E. coli* yqjG gene is not in a gene cluster and the bacterium cannot mineralize PCP. Therefore, TriHCQ and DiHCQ are not likely to be the physiological substrates for YqjG. Another protein in this class, PhGSTO1 from *Phanerochaete chrysosporium*, has been shown to deglutathionylate *S*-glutathionyl-p-benzohydroquinone and *S*-glutathionyl-naphthoquinone (thiodione) [103]. In both cases the active site cysteine is essential for catalysis [97,103].

To search for physiologically relevant substrates for YqjG, we interrogated the EcoCyc database for hydroquinone molecules in *E. coli* [109]. Some hydroquinone metabolites include 5-hydroxy-1,4-naphthoquinone (TCA cycle intermediate), 1,4-napthoquinone, gentisate (hydroquinonecarboxylic acid), several menaquinone (vitamin K) derived substrates and several ubiquinone (coenzyme Q10) substrates. These hydroquinones can be readily oxidized to the corresponding benzoquinones in air at neutral pH, which can react with GSH via Michael addition to the corresponding GS-hydroquinone. Future studies will focus on determining the physiologically relevant substrates of YqjG in *E. coli*.

CHAPTER V

CONCLUSIONS AND FUTURE STUDIES

This study focused on exploring new functional and structural space in the GSH transferase superfamily from *E. coli* K-12. The structural and functional properties of YghU and YfcG indicate they are members of a new subfamily of GSH transferases. These proteins possess poor GSH transferase activity with electrophilic substrates but demonstrate excellent disulfide bond oxido-reductase activity with hydroxyethyl disulfide (k_{cat} = 74 ± 6 s^{-1} and k_{cat}/K_M^{GSH} = (6.4 ± 1.3) x 10^4 $M^{-1}s^{-1}$ for YghU). The crystal structure of YfcG (2.3 Å resolution) was determined from crystals grown in the presence of GSH. Surprisingly, a molecule of GSSG was observed in the active site. GSSG binds the enzyme 100x more tightly than GSH (K_d^{GSH} = 330 ± 10 μM, K_d^{GSSG} = 2.4 ± 0.5 μM). The crystal structure of YghU (1.5 Å resolution) reveals the protein binds two molecules of GSH in each active site, which is consistent with biphasic equilibrium data. (K_{d1} = 0.07 ± 0.03 mM, K_{d2} = 1.3 ± 0.2 mM). A superposition of the YfcG and YghU structures reveals remarkable similarity between the two enzymes (RMSD of 0.817 Å for 177 structurally equivalent α-carbons), particularly with regard to the 2 GSH/GSSG molecules in each. In both structures the hydroxyl group of a threonine residue is within hydrogen bonding distance to the sulfur of GSH1 (YghU) or the disulfide of GSSG (YfcG). These represent the first structurally confirmed cases of a threonine residue used in this capacity. We hypothesize that the two structures represent reduced and oxidized forms of GSH-dependent disulfide-bond oxidoreductases.

Using a network analysis we identified 166 sequences related to YfcG and YghU, including several Ure2 proteins from yeast. These sequences may represent a new class of GSH transferases, which possess GSH-dependent peroxidase and disulfide-bond reductase activities and are characterized by a marked reduction in cysteine content. Proteins in the nu-class may represent a new category of disulfide-bond oxidoreductases that operate through a cysteine-independent mechanism and thereby circumvent oxidative damage to protein thiols that would normally occur under conditions of oxidative stress. However, the mechanism of thiol(s) binding may differ in this class of proteins and the exact substrates and the *in vivo* activity of the protein remains to be elucidated.

The functional and structural properties of YqjG, a GSH transferase homolog from *E. coli* were also explored. YqjG is annotated in UniProt as a member of the omega class of GSH tranferase proteins. However, using bioinformatics analysis of the global GSH transferase superfamily we determined that the YqjG protein is more similar to a subset of ~500 other proteins that are distinct from the omega class and all other previously characterized classes of GSH transferases. The crystal structure of YqjG at 2.2 Å resolution reveals that the protein adopts a classical GSH transferase fold but a unique splayed dimer. Most canonical, soluble GSH transferases are dimeric proteins, which utilize residues along two helices (one from thioredoxin domain, one from all α-helical domain) for the formation of the dimer interface. In contrast, YqjG utilizes amino acids located on loops between helices in the all α-helical domain and at in the C-terminal tail of the protein to form the dimer interface. Three structurally similar proteins have been described. These proteins have activities as disulfide bond oxido-reductases, *S*-glutathionyl-(chloro) hydroquinone reductases and *S*-(phenylacyl)glutathione reductases

[97,103]. YqjG has modest disulfide bond reductase activity with 2-hydroxyethyldisulfide and cysteine. Possible hydroquinone substrates for YqjG in *E. coli* are currently being investigated.

The overarching goal of this project is the assignment of *in vivo* physiological function to members of the GSH transferase superfamily in *E. coli* K-12. Several strategies are required for the accurate assignment of function to proteins of unknown function. These approaches include; 1) analysis using informatics and sequence similarity, 2) co-localization of genes providing operon/metabolic context, 3) transcriptional analysis, 4) phenotypic response to gene knockouts, 5) structural biology, 6) and functional assays of the protein [73,105]. The work described in this report has focused on functional and structural studies of three GSH transferase homologs from *E. coli*. Because the accurate assignment of an *in vivo* physiological function cannot rely solely on *in vitro* catalytic activity, future studies will focus of phenotypic screening of knock out strains and metabolic analysis. Many of these experiments will be done in collaboration with Dr. John Cronan at the University of Illinois as part of the Enzyme Function Initiative (enzymefunction.org).

A key component to the *in vivo* assignment of enzyme function is the analysis of phenotypes in knock out strains. The Keio collection contains deletion strains for all non-essential genes in *E. coli*. In addition, a double knock out of the *yfcG* and *yghU* genes has been made. Using a plate reader, high throughput analysis of these deletion strains under a variety of growth conditions, including anaerobically, can be performed. Custom plates containing reductive and oxidative stress agents can be used to integrate enzyme function. The read out for cell viability in these assays is a color change that occurs upon

reduction of a dye by intracellular NADH. Therefore metabolic activity of the cells is measured not cell density. Preliminary experiments indicate that the Δ*yghU* strain undergoes changes in cell growth in the presence of diamide, a small oxidant probe that changes the thiol status of the organism.

Experimental approaches to functionally characterize microbial metabolism are needed for the discovery and assignment of enzyme function [110]. The high sensitivity of mass spectrometry combined with its ability to detect and quantify a wide range of metabolites from complex mixtures makes it well suited for functional genomics studies. Often, metabolites are the final downstream products of enzymatic processes and can be used to assign or validate functional annotations related to enzyme activities [110]. Using metabolic profiling one can identify large numbers of changing metabolites between strains over expressing a protein target and strains with a knock out (null) of the target. A critical step in interpreting these metabolic profiles is the identification of unknown metabolites. Microorganism specific databases like BioCyc [111] and KEGG [112] allow comparison of experimentally determined metabolic profiles against predicted ones. For the YfcG and YghU proteins, the metabolic profiles of the knock-out strains and over-expression strains will be collected and differences in the redox status or modifications of small molecules or proteins with GSH will be investigated.

A particularly interesting future study is investigation of the subcellular localization of the YghU protein. YghU contains a 50 amino acid N-terminal extension, which does not contain a traditional signal sequence for export by the general secretory (Sec) pathway or the twin-arginine translocation (Tat) system. However, the protein does score relatively high in the SecretomeP program [113], which scores proteins based on

classical and leaderless secretion of proteins. The beta class GSH transferase from *O. anthropi* (OaGST) was found in the periplasm even though no signal sequence could be identified [114]. The mechanism of OaGST transport (and possibly other GSH transferases) into the periplasm is currently uncharacterized. The discovery of a GSH transporter in *E. coli* and the description of GSH in the periplasm at biologically significant concentrations indicates that GSH transferases could display activity in this cellular compartment [115,116]. The localization of YghU will be investigated by construction of YghU and β-galactosidase (LacZ) gene fusions [117].

REFERENCES

1. Bond JA. *Comprehensive Toxicology: General Principles, Volume 1*. Pergamon; 1999:354.

2. Jakoby WB. *Enzymatic Basis of Detoxication*. Academic Pr; 1980:369.

3. Armstrong RN. Enzyme-catalyzed detoxication reactions: mechanisms and stereochemistry. *CRC Crit. Rev. Biochem.* 1987;22(1):39-88.

4. Hayes AW. *Principles and methods of toxicology*. Informa HealthCare; 2007:2270.
5. Guengerich FP. Cytochrome p450 and chemical toxicology. *Chem. Res. Toxicol.* 2008;21(1):70-83.

6. Guengerich FP. *Comprehensive Toxicology: Biotransformation, Volume Volume 3*. Pergamon; 1999:528.

7. Booth J, Boyland E, Sims P. An enzyme from rat liver catalysing conjugations with glutathione. *Biochemical Journal*. 1961;79:516-524.

8. Combes B, Stakelum GS. A liver enzyme that conjugates sulfobromophthalein sodium with glutathione. *J. Clin. Invest.* 1961;40:981-988.

9. Armstrong RN. Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* 1997;10(1):2-18.

10. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* 2005;45:51-88.

11. Sau A, Pellizzari Tregno F, Valentino F, Federici G, Caccuri AM. Glutathione transferases and development of new principles to overcome drug resistance. *Arch. Biochem. Biophys.* 2010;500(2):116-122.

12. Sheehan D, Dowd CA, Meade G, Foley VM. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*. 2001;360(Pt 1):1.

13. Pearson WR. Phylogenies of glutathione transferase families. *Meth. Enzymol.* 2005;401:186-204.

14. Armstrong RN. Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry*. 2000;39(45):13625-13632.

15. Gaballa A, Newton GL, Antelmann H, et al. Biosynthesis and functions of

bacillithiol, a major low-molecular-weight thiol in Bacilli. *Proc. Natl. Acad. Sci. U.S.A.* 2010;107(14):6482-6486.

16. Fillgrove KL, Pakhomova S, Schaab MR, Newcomer ME, Armstrong RN. Structure and mechanism of the genomically encoded fosfomycin resistance protein, FosX, from Listeria monocytogenes. *Biochemistry*. 2007;46(27):8110-8120.

17. Ladner JE, Parsons JF, Rife CL, Gilliland GL, Armstrong RN. Parallel evolutionary pathways for glutathione transferases: structure and mechanism of the mitochondrial class kappa enzyme rGSTK1-1. *Biochemistry*. 2004;43(2):352-361.

18. Thompson LC, Ladner JE, Codreanu SG, et al. 2-Hydroxychromene-2-carboxylic acid isomerase: a kappa class glutathione transferase from Pseudomonas putida. *Biochemistry*. 2007;46(23):6710-6722.

19. Jakobsson PJ, Morgenstern R, Mancini J, Ford-Hutchinson A, Persson B. Common structural features of MAPEG -- a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Sci.* 1999;8(3):689-692.

20. Allocati N, Federici L, Masulli M, Di Ilio C. Glutathione transferases in bacteria. *FEBS J.* 2009;276(1):58-75.

21. Dirr H, Reinemer P, Huber R. X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur. J. Biochem.* 1994;220(3):645-661.

22. Oakley A. Glutathione transferases: a structural perspective. *Drug Metab Rev*. 2011;43(2):138-151.

23. Atkinson HJ, Babbitt PC. Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry*. 2009;48(46):11108-11116.

24. Whitbread AK, Masoumi A, Tetlow N, et al. Characterization of the omega class of glutathione transferases. *Meth. Enzymol.* 2005;401:78-99.

25. Board PG, Anders MW. Human glutathione transferase zeta. *Meth. Enzymol.* 2005;401:61-77.

26. Board PG, Taylor MC, Coggan M, et al. Clarification of the role of key active site residues of glutathione transferase zeta/maleylacetoacetate isomerase by a new spectrophotometric technique. *Biochem. J.* 2003;374(Pt 3):731-737.

27. Harrop SJ, DeMaere MZ, Fairlie WD, et al. Crystal structure of a soluble form of the intracellular chloride ion channel CLIC1 (NCC27) at 1.4-A resolution. *J. Biol. Chem.* 2001;276(48):44993-45000.

28. Littler DR, Harrop SJ, Fairlie WD, et al. The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. *J. Biol. Chem.* 2004;279(10):9298-9305.

29. Shishido T. Glutathione *S*-transferase from *Escherichia coli*. *Agricultural and Biological Chemistry*. 1981;45:2951-2953.

30. Allocati N, Federici L, Masulli M, Favaloro B, Di Ilio C. Cysteine 10 is critical for the activity of Ochrobactrum anthropi glutathione transferase and its mutation to alanine causes the preferential binding of glutathione to the H-site. *Proteins*. 2008;71(1):16-23.

31. Di Ilio C, Aceto A, Piccolomini R, et al. Purification and characterization of three forms of glutathione transferase from Proteus mirabilis. *Biochem. J.* 1988;255(3):971-975.

32. Iizuka M, Inoue Y, Murata K, Kimura A. Purification and some properties of glutathione S-transferase from Escherichia coli B. *J. Bacteriol.* 1989;171(11):6039-6042.

33. Tocheva EI, Fortin PD, Eltis LD, Murphy MEP. Structures of ternary complexes of BphK, a bacterial glutathione S-transferase that reductively dechlorinates polychlorinated biphenyl metabolites. *J. Biol. Chem.* 2006;281(41):30933-30940.

34. Federici L, Masulli M, Bonivento D, et al. Role of Ser11 in the stabilization of the structure of Ochrobactrum anthropi glutathione transferase. *Biochem. J.* 2007;403(2):267-274.

35. Rossjohn J, Polekhina G, Feil SC, et al. A mixed disulfide bond in bacterial glutathione transferase: functional and evolutionary implications. *Structure*. 1998;6(6):721-734.

36. Inoue H, Nishida M, Takahashi K. Effects of Cys10 mutation to Ala in glutathione transferase from *Escherichia coli*. *Journal of Organometallic Chemistry*. 2000;611(1):593-595.

37. Nishida M, Harada S, Noguchi S, et al. Three-dimensional structure of Escherichia coli glutathione S-transferase complexed with glutathione sulfonate: catalytic roles of Cys10 and His106. *J. Mol. Biol.* 1998;281(1):135-147.

38. Casalone E, Allocati N, Ceccarelli I, et al. Site-directed mutagenesis of the Proteus mirabilis glutathione transferase B1-1 G-site. *FEBS Lett.* 1998;423(2):122-124.

39. Allocati N, Casalone E, Masulli M, et al. Evaluation of the role of two conserved active-site residues in beta class glutathione S-transferases. *Biochem. J.* 2000;351 Pt 2:341-346.

40. Favaloro B, Tamburro A, Trofino MA, et al. Modulation of the glutathione S-transferase in Ochrobactrum anthropi: function of xenobiotic substrates and other forms of stress. *Biochem. J.* 2000;346 Pt 2:553-559.

41. Fahey RC, Brown WC, Adams WB, Worsham MB. Occurrence of glutathione in bacteria. *J. Bacteriol.* 1978;133(3):1126-1129.

42. Fahey RC, Sundquist AR. Evolution of glutathione metabolism. *Adv. Enzymol. Relat. Areas Mol. Biol.* 1991;64:1-53.

43. Wiktelius E, Stenberg G. Novel class of glutathione transferases from cyanobacteria exhibit high catalytic activities towards naturally occurring isothiocyanates. *Biochemical Journal*. 2007;406:115-123.

44. Scholtz R, Wackett LP, Egli C, Cook AM, Leisinger T. Dichloromethane dehalogenase with improved catalytic activity isolated from a fast-growing dichloromethane-utilizing bacterium. *J. Bacteriol.* 1988;170(12):5698-5704.

45. Vuilleumier S, Ivos N, Dean M, Leisinger T. Sequence variation in dichloromethane dehalogenases/glutathione S-transferases. *Microbiology (Reading, Engl.)*. 2001;147(Pt 3):611-619.

46. Stourman NV, Rose JH, Vuilleumier S, Armstrong RN. Catalytic mechanism of dichloromethane dehalogenase from Methylophilus sp. strain DM11. *Biochemistry*. 2003;42(37):11048-11056.

47. Vuilleumier S, Leisinger T. Protein engineering studies of dichloromethane dehalogenase/glutathione S-transferase from Methylophilus sp. strain DM11. Ser12 but not Tyr6 is required for enzyme activity. *Eur. J. Biochem.* 1996;239(2):410-417.

48. McCarthy DL, Louie DF, Copley SD. Identification of a Covalent Intermediate between Glutathione and Cysteine13 Formed during Catalysis by Tetrachlorohydroquinone Dehalogenase. *J. Am. Chem. Soc.* 1997;119(46):11337-11338.

49. McCarthy DL, Claude AA, Copley SD. In vivo levels of chlorinated hydroquinones in a pentachlorophenol-degrading bacterium. *Appl. Environ. Microbiol.* 1997;63(5):1883-1888.

50. van Hylckama Vlieg JET, Leemhuis H, Spelberg JHL, Janssen DB. Characterization of the Gene Cluster Involved in Isoprene Metabolism in Rhodococcus sp. Strain AD45. *J. Bacteriol.* 2000;182(7):1956.

51. Fall R, Copley SD. Bacterial sources and sinks of isoprene, a reactive atmospheric hydrocarbon. *Environ. Microbiol.* 2000;2(2):123-130.

52. van Hylckama Vlieg JE, Kingma J, Kruizinga W, Janssen DB. Purification of a

glutathione S-transferase and a glutathione conjugate-specific dehydrogenase involved in isoprene metabolism in Rhodococcus sp. strain AD45. *J. Bacteriol.* 1999;181(7):2094-2101.

53. Masai E, Katayama Y, Nishikawa S, Fukuda M. Characterization of Sphingomonas paucimobilis SYK-6 genes involved in degradation of lignin-related compounds. *J. Ind. Microbiol. Biotechnol.* 1999;23(4-5):364-373.

54. Masai E, Ichimura A, Sato Y, et al. Roles of the enantioselective glutathione S-transferases in cleavage of beta-aryl ether. *J. Bacteriol.* 2003;185(6):1768-1775.

55. Bresell A, Weinander R, Lundqvist G, et al. Bioinformatic and enzymatic characterization of the MAPEG superfamily. *FEBS J.* 2005;272(7):1688-1703.

56. Vuilleumier S. Bacterial glutathione S-transferases: what are they good for? *J. Bacteriol.* 1997;179(5):1431-1441.

57. Nishida M, Kong KH, Inoue H, Takahashi K. Molecular cloning and site-directed mutagenesis of glutathione S-transferase from Escherichia coli. The conserved tyrosyl residue near the N terminus is not essential for catalysis. *J. Biol. Chem.* 1994;269(51):32536-32541.

58. Rife CL, Parsons JF, Xiao G, Gilliland GL, Armstrong RN. Conserved structural elements in glutathione transferase homologues encoded in the genome of Escherichia coli. *Proteins*. 2003;53(4):777-782.

59. Hansen A, Gu Y, Li M, et al. Structural basis for the function of stringent starvation protein a as a transcription factor. *J. Biol. Chem.* 2005;280(17):17380-17391.

60. Ishihama A, Saitoh T. Subunits of RNA polymerase in function and structure. IX. Regulation of RNA polymerase activity by stringent starvation protein (SSP). *J. Mol. Biol.* 1979;129(4):517-530.

61. Williams MD, Ouyang TX, Flickinger MC. Starvation-induced expression of SspA and SspB: the effects of a null mutation in sspA on Escherichia coli protein synthesis and survival during growth and prolonged starvation. *Mol. Microbiol.* 1994;11(6):1029-1043.

62. Hansen A, Qiu Y, Yeh N, et al. SspA is required for acid resistance in stationary phase by downregulation of H-NS in Escherichia coli. *Mol. Microbiol.* 2005;56(3):719-734.

63. Kanai T, Takahashi K, Inoue H. Three distinct-type glutathione S-transferases from Escherichia coli important for defense against oxidative stress. *J. Biochem.* 2006;140(5):703-711.

64. Desai KK, Miller BG. Recruitment of genes and enzymes conferring resistance to the

nonnatural toxin bromoacetate. *Proc. Natl. Acad. Sci. U.S.A.* 2010;107(42):17968-17973.

65. Griffith OW, Mulcahy RT. The enzymes of glutathione synthesis: gamma-glutamylcysteine synthetase. *Adv. Enzymol. Relat. Areas Mol. Biol.* 1999;73:209-67, xii.

66. Lu SC. Regulation of glutathione synthesis. *Mol. Aspects Med.* 2009;30(1-2):42-59.

67. Bennett BD, Kimball EH, Gao M, et al. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* 2009;5(8):593-599.

68. Tabor H, Tabor CW. Isolation, characterization, and turnover of glutathionylspermidine from Escherichia coli. *J. Biol. Chem.* 1975;250(7):2648-2654.

69. Pai C, Wu H, Lin C, Wang AH. Structure and mechanism of Escherichia coli glutathionylspermidine amidase belonging to the family of cysteine; histidine-dependent amidohydrolases/peptidases. *Protein Sci.* 2011;20(3):557-566.

70. Bollinger JM, Kwon DS, Huisman GW, Kolter R, Walsh CT. Glutathionylspermidine metabolism in Escherichia coli. Purification, cloning, overproduction, and characterization of a bifunctional glutathionylspermidine synthetase/amidase. *J. Biol. Chem.* 1995;270(23):14031-14041.

71. Smith K, Borges A, Ariyanayagam MR, Fairlamb AH. Glutathionylspermidine metabolism in Escherichia coli. *Biochem. J.* 1995;312 (Pt 2):465-469.

72. Chiang B, Chen T, Pai C, et al. Protein S-thiolation by Glutathionylspermidine (Gsp): the role of Escherichia coli Gsp synthetASE/amidase in redox regulation. *J. Biol. Chem.* 2010;285(33):25345-25353.

73. Stourman NV, Wadington MC, Schaab MR, Atkinson HJ, Babbitt PC. Functional Genomics In Escherichia coli: Experimental Approaches for the Assignment of Enzyme Function Hicks MG, Kettner C, eds. *Proceedings of the 3rd International Beilstein Workshop on Experimental Standard Conditions of Enzyme Characterizations*. 2008:1-12.

74. Koonin EV, Mushegian AR, Tatusov RL, et al. Eukaryotic translation elongation factor 1 gamma contains a glutathione transferase domain--study of a diverse, ancient protein superfamily using motif search and structural modeling. *Protein Sci.* 1994;3(11):2045-2054.

75. Bai M, Zhou J, Perrett S. The yeast prion protein Ure2 shows glutathione peroxidase activity in both native and fibrillar forms. *J. Biol. Chem.* 2004;279(48):50025-50030.

76. Dulhunty A, Gage P, Curtis S, Chelvanayagam G, Board P. The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor

calcium release channel modulator. *J. Biol. Chem.* 2001;276(5):3319-3323.

77. Kosower NS, Kosower EM. Diamide: an oxidant probe for thiols. *Meth. Enzymol.* 1995;251:123-133.

78. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A. FFAS03: a server for profile--profile sequence alignments. *Nucleic Acids Res.* 2005;33:W284-8.

79. McCoy AJ, Grosse-Kunstleve RW, Adams PD, et al. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40(Pt 4):658-674.

80. Terwilliger TC. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D Biol. Crystallogr.* 2003;59(Pt 1):38-44.

81. Terwilliger TC. Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr. D Biol. Crystallogr.* 2003;59(Pt 1):45-49.

82. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 1997;53(3):240-255.

83. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 2004;60(Pt 12 Pt 1):2126-2132.

84. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007;35(Web Server issue):W375-83.

85. Flohe L, Gunzler WA. Assays of glutathione peroxidase. *Methods Enzymol*. 1984;105:114-121.

86. Vlamis-Gardikas A, Aslund F, Spyrou G, Bergman T, Holmgren A. Cloning, overexpression, and characterization of glutaredoxin 2, an atypical glutaredoxin from Escherichia coli. *J. Biol. Chem.* 1997;272(17):11236-11243.

87. Wadington MC, Ladner JE, Stourman NV, Harp JM, Armstrong RN. Analysis of the structure and function of YfcG from Escherichia coli reveals an efficient and unique disulfide bond reductase. *Biochemistry*. 2009;48(28):6559-6561.

88. Iacazio G, Langrand G, Baratti J, Buono G, Triantaphylides C. Preparative, enzymic synthesis of linoleic acid (13S)-hydroperoxide using soybean lipoxygenase-1. 1990.

89. Jakobsson PJ, Mancini JA, Riendeau D, Ford-Hutchinson AW. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. *J. Biol. Chem.* 1997;272(36):22934-22939.

90. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* 1997;276(Macromolecular Crystallography, Part A):307-326.

91. Sheldrick GM. Macromolecular phasing with SHELXE. *Zeitschrift für Kristallographie*. 2002;217(12-2002):644-650.

92. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* 2002;58(Pt 10 Pt 2):1772-1779.

93. Pape T, Schneider TR. HKL2MAP: a graphical user interface for macromolecular phasing with SHELXprograms. *J Appl Crystallogr*. 2004;37(5):843-844.

94. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. *Meth. Enzymol.* 2003;374:229-244.

95. McRee DE. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* 1999;125(2-3):156-165.

96. Wang X, Zhang Z, Perrett S. Characterization of the activity and folding of the glutathione transferase from Escherichia coli and the roles of residues Cys(10) and His(106). *Biochem. J.* 2009;417(1):55-64.

97. Xun L, Belchik SM, Xun R, et al. S-Glutathionyl-(chloro)hydroquinone reductases: a novel class of glutathione transferases. *Biochem. J.* 2010;428(3):419-427.

98. Yu Y, Wang H, Bai M, Perrett S. Flexibility of the Ure2 prion domain is important for amyloid fibril formation. *Biochem. J.* 2011;434(1):143-151.

99. Bousset L, Belrhali H, Melki R, Morera S. Crystal structures of the yeast prion Ure2p functional region in complex with glutathione and related compounds. *Biochemistry*. 2001;40(45):13564-13573.

100. Zhang Z, Perrett S. Novel glutaredoxin activity of the yeast prion protein Ure2 reveals a native-like dimer within fibrils. *J. Biol. Chem.* 2009;284(21):14058-14067.

101. Fahey RC, Hunt JS, Windham GC. On the cysteine and cystine content of proteins. *J Mol Evol*. 1977;10(2):155-160.

102. Miseta A. Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Molecular Biology and Evolution*. 2000.

103. Meux E, Prosper P, Ngadin A, et al. Glutathione Transferases of Phanerochaete chrysosporium: S-GLUTATHIONYL-P-HYDROQUINONE REDUCTASE BELONGS TO A NEW STRUCTURAL CLASS. *J. Biol. Chem.* 2011;286(11):9162-9173.

104. Frova C. Glutathione transferases in the genomics era: new insights and perspectives. *Biomol. Eng.* 2006;23(4):149-169.

105. Gerlt JA. Collaborative Center for an Enzyme Function Initiative (EFI). *Glue Grant Application, National Institute of General Medical Sciences*. 2007.

106. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2010;38(Database issue):D211-22.

107. Flohe L. Assays of glutathione peroxidase. *Meth. Enzymol.* 1984.

108. Terwilliger T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Rad*. 2003;11(1):49-52.

109. Keseler IM, Collado-Vides J, Santos-Zavaleta A, et al. EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Res.* 2011;39(Database issue):D583-90.

110. Baran R, Reindl W, Northen TR. Mass spectrometry based metabolomics and enzymatic assays for functional genomics. *Curr. Opin. Microbiol.* 2009;12(5):547-552.

111. Caspi R, Foerster H, Fulcher CA, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 2008;36(Database issue):D623-31.

112. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36(Database issue):D480-4.

113. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microbiol.* 2005;5:58.

114. Tamburro A, Robuffo I, Heipieper HJ, et al. Expression of glutathione S-transferase and peptide methionine sulphoxide reductase in Ochrobactrum anthropi is correlated to the production of reactive oxygen species caused by aromatic substrates. *FEMS Microbiol. Lett.* 2004;241(2):151-156.

115. Pittman MS, Robinson HC, Poole RK. A bacterial glutathione transporter (Escherichia coli CydDC) exports reductant to the periplasm. *J. Biol. Chem.* 2005;280(37):32254-32261.

116. Eser M, Masip L, Kadokura H, Georgiou G, Beckwith J. Disulfide bond formation by exported glutaredoxin indicates glutathione's presence in the E. coli periplasm. *Proc. Natl. Acad. Sci. U.S.A.* 2009;106(5):1572-1577.

117. van Beilen JB, Penninga D, Witholt B. Topology of the membrane-bound alkane hydroxylase of Pseudomonas oleovorans. *J. Biol. Chem.* 1992;267(13):9194-9201.