Beilstein-Institut ESCEC, September 23rd – 26th, 2007, Rüdesheim/Rhein, Germany

Functional Genomics in *Escherichia coli*: Experimental Approaches for the Assignment of Enzyme Function

Nina V. Stourman[1], Megan C. Wadington[1], Matthew R. Schaab[1], Holly J. Atkinson[2], Patricia C. Babbitt[3], Richard N. Armstrong[1]*

[1]Departments of Biochemistry and Chemistry, Center in Molecular Toxicology, and the Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, TN, 37232 – 0146, U.S.A.

[2]Program in Biological & Medical Informatics, University of California, San Francisco, CA 94158 – 2330, U.S.A.

[3]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry and California Institute for Quantitative Biosciences, University of California, San Francisco, CA 94158 – 2330, U.S.A.

E-Mail: *r.armstrong@vanderbilt.edu

Received: 11th February 2008 / Published: 20th August 2008

Abstract

A major challenge in biochemistry is to understand the functional genomics of organisms. This is a staggering problem when one considers the fact that almost 40% of the genes in one of the best-understood organisms in the biosphere, *Escherichia coli*, have no experimentally verified function. In this paper we address the challenge of, and criteria for, assigning protein function in the context of the glutathione (GSH) transferase paralogues encoded in the *E. coli* genome. The *E. coli* genome harbors genes encoding nine GSH transferase homologues including YliJ, YncG, Gst, YfcF, YfcG, YghU, SspA and YibF as well as the membrane-bound enzyme YecN. Amazingly, only one of these genes has a reasonably well-defined function and it does NOT encode a protein with GSH transferase activity but rather a transcription factor, stringent starvation protein A, SspA.

Introduction

The protein world, as we currently understand it, is composed in large part, of groups of proteins commonly called superfamilies that share similarities in sequence and three-dimensional structure but can diverge considerably with respect to biological function. One central question in modern biology is the extent of functional diversity that can be realized in a given protein superfamily. A different but related question is, how can we experimentally define the biological function(s) of all genes in a given organism? This has yet to be done for any single organism. In this paper we illustrate some experimental approaches that, when applied in parallel, are designed to reveal the functions of members of the GSH transferase superfamily in *E. coli*.

Glutathione is the predominant redox-active thiol in most aerobic organisms where it plays a fundamental role in metabolic, catabolic and redox chemistry. GSH transferases are enzymes that typically catalyze the addition of GSH to electrophilic acceptors as illustrated in Scheme 1 [1]. In microorganisms, these enzymes often participate in the catabolism of xenobiotic molecules [2 – 5]. The name, however, obscures the diverse impact that this group of proteins has in mammalian and microbial biology. In the last several years it has become apparent that members of this superfamily also perform other very diverse functions that are not at all related to the reaction illustrated in Scheme 1. These other functions include the regulation of transcription [6 – 10] and translation [11] and the intracellular transport of ions [12, 13].

Scheme 1: $GSH + R\text{-}X \rightarrow GSR + X^- + H^+$

The canonical or soluble GSH transferases are typically dimeric proteins where each subunit is composed of a N-terminal thioredoxin-like domain and a C-terminal α-helical domain as illustrated in Figure 1. The eight GSH transferase homologues encoded in the *E. coli* genome share these same structural characteristics based on available structures determined to date or inferred from sequence alignments. The *E. coli* paralogues are also defined by a consensus sequence of 17 residues, most of which are involved in the hydrophobic core of the thioredoxin domain [14]. Only two of the seventeen conserved residues are implicated in the binding of GSH, a fact that suggests that some of the paralogues may not bind or utilize GSH.

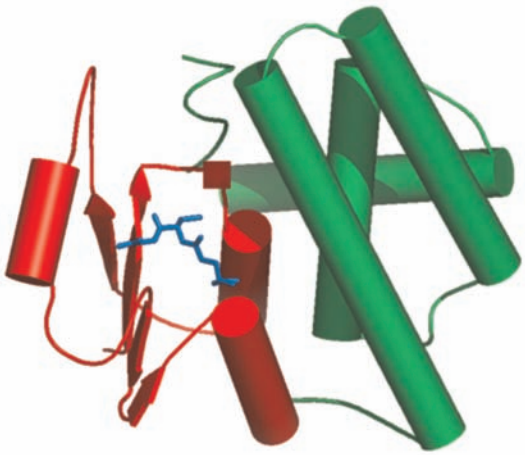

Figure 1. Diagram of the structure of a typical GSH transferase subunit. The all α-helical domain is shown in green and the thioredoxin domain is shown in red. Glutathione (shown in blue) is bound principally by the thioredoxin domain.

The question then arises, how can the biological functions of proteins with unknown function be divined? The intellectual and experimental approaches to this type of problem are numerous and must be used in concert, inasmuch as no single approach is likely to be definitive. Our combination of approaches includes; (i) analysis using informatics and genome context, (ii) response of gene expression to environmental stress, (iii) phenotypic response to gene knockouts, (iv) a search for protein partners, (v) structural biology, and of course (vi) functional assays of proteins. Other possible experimental avenues include metabolomics, proteomics and genetics. The point of this article is to illustrate how we are approaching this problem in the context of the GSH transferase superfamily in the biological context of *E. coli*.

The GSH transferase homologues encoded in the genome of *Escherichia coli* are quite diverse within the superfamily. A Cytoscape analysis [15] (Figure 2) of sequence similarity of both microbial and eukaryotic GSH transferase homologues indicates that at least five families are represented by the eight proteins including HSP26 (YibF, sspA); zeta, beta or tau (YfcF), beta (gst, YncG); phi or other (YliJ); and theta or other (YghU and YfcG). The diversity is remarkable and suggests a significant functional diversity as well. It is also remarkable that several nodes for the *E. coli* paralogues (YncG, YfcF, gst, and YghU) lie on the periphery of one main cluster and none are associated with the cluster that includes the alpha, mu, pi and sigma subfamilies that are common in eukaryotes.

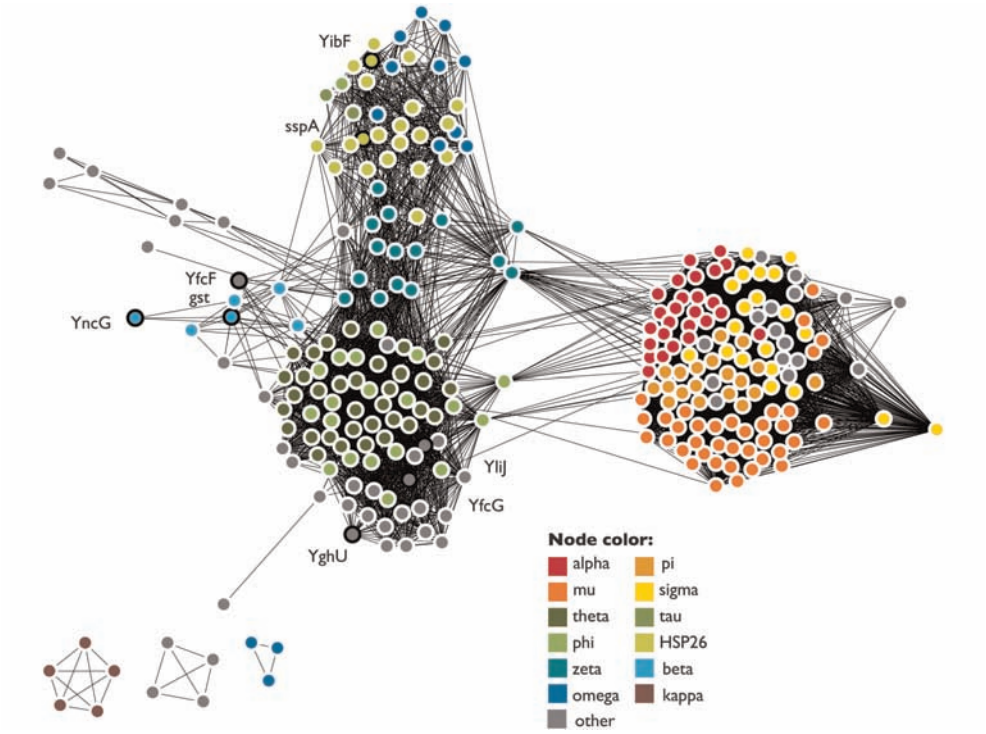

Figure 2. Sequence similarity network of 272 sequences longer than 80 residues from the SWISS-PROT GST superfamily, as well as proteins containing GST-like domains. For each pair of sequences, an edge is shown if the BLAST E-value for the pairwise alignment is better than 1×10^{-9}. Edges at this limiting E-value had a median residue identity of 25% over an alignment length of 180 amino acids. If two sequences are less similar, their nodes are not connected in this network. A thick black border denotes the *E. coli* GSH transferase paralogues. Node color indicates the family classification by SWISS-PROT.

Materials and Methods

Reduced glutathione (GSH), all buffer salts and other chemicals were obtained from commercial sources. Glutathionylspermidine (GspSH) was synthesized by incubating the C55A mutant of the *E. coli* glutathionylspermidine synthetase/amidase with GSH, ATP and spermidine. The GspSH was purified by ion exchange chromatography. The details of the synthesis will be published at a later date.

Chromosomal gene disruptions or knockouts in *E. coli* genome were made by the Wanner method [16] as follows. *E. coli* BW25113-*pKD46* cells were grown overnight at 30 °C in 5 mL SOC media containing ampicillin (100 µg/mL). The culture was diluted 100-fold into 20 mL of fresh SOB media supplemented with 1 mM arabinose and grown at 30 °C to

OD_{600}= 0.6. Cells were made electrocompetent by four consecutive washes with decreasing volumes of ice-cold 10% sterile glycerol. The final pellet was suspended in 40 µL of cold 10% sterile glycerol, flash-frozen and stored at – 80 °C. DNA fragments used for the gene disruption were amplified by PCR from the pKD3 vector carrying the chloramphenicol resistance gene. Each primer consisted of 40-nt homologous to the gene to be eliminated and 20-nt priming sequence for pKD3. Electroporation was done in 1 mm cuvettes using ElectroCell Manipulator BTX ECM 399 (BTX Harvard Apparatus, Holliston, MA) with charging voltage 1.4 kV and 5 ms pulse length. Electroporation cuvettes were chilled on ice before the addition of 40 µL of electrocompetent cells mixed with 2 µL of gel-purified PCR product containing 10 to 30 ng of DNA. Immediately after the pulse the cells were transferred to culture tubes containing 960 µL of SOC media and incubated at 37 °C for 1 hr. A 400 µL aliquot of the culture was plated on the LB/chloramphenicol plates. Plates were kept in the incubator at 37 °C for up to 24 hr allowing colonies to grow. The deletion of the gene was confirmed by colony PCR using specific gene primers. The primers for the *yghU* knockout using pKD3 were:

Forward: ATACTTATCA GCCCGCGAAA GTCTGGACGT GGGATAAATC *GTGTAGGCTGGAGCTGCTTC*
Reverse: TGACGCTTAT CTTCCGTATT CGTCTCGAAA TCACTGGCGT *CATATGAATATCCTCCTTAGT*
Specific primers for *yghU* (amplicon size 318 bp) were:
Forward: TATCTGGCGAGAAATTTGG
Reverse: CTCAGCGGCATCATACACAC
Specific primers for *gss* knockout using pKD3 were:
Forward: CAAAGGAACG ACCAGCCAGG ATGCCCCGTT CGGGACATTA *CATATGAATATCCTCCTTAGT*
Reverse: CTCTTTTTTG ATGACCAGTG ATTCATCACC GCGCAAACAC *GTGTAGGCTGGAGCTGCTTC*
Specific primers for *gss* (amplicon size 353 bp) were:
Forward: AAGTCCGTATTGCGGAACAG
Reverse: GGCACTCTCGGTAATGGTGT

Gene expression levels of the GSH transferase homologues were quantified by the reverse transcriptase polymerase-chain-reaction (RT-PCR) as follows. Total RNA was purified from 5 mL of *E. coli* cells with RNeasy Mini Kit (Qiagen Inc,, Valencia, CA) according to the manufacturer's protocol with an additional RNA clean-up step after the treatment with DNA-free kit (Ambion, Foster City, CA). After this procedure, the RNA was essentially free of genomic DNA. Total RNA was quantified by measuring absorbance at 260 nm. RT-PCR reactions were performed using SuperScript II reverse transcriptase from Invitrogene (Carlsbad, CA) and random hexamer primers following the manufacturer's protocol. For each reaction 1 µg of purified RNA was used. cDNA obtained in the first step was used for the following PCR reactions with the specific gene primers which was carried at 55 °C in the

linear range of amplification (22 – 25 cycles). A 10 μL aliquot from each reaction was run on 1.2% agarose gel containing ethidium bromide. The gel images were acquired with Molecular Imager Gel Doc XR System (BioRad) and analyzed with Quantity One 1-D analysis software and Adobe Photoshop.

Thiol analysis of *E. coli* was obtained as follows. An overnight culture grown of BW251137 in LB at 37 °C was diluted 100-fold into fresh LB or MM9 medium containing 0.4% glucose. For aerobic conditions cultures were grown in the shaker at 37 °C to OD 0.6 for exponential growth or overnight for stationary phase. For anaerobic conditions the cells were grown in the tubes with tight-screw caps filled with the media to the top and harvested without opening the tube. Cells were harvested by centrifugation at 6,000 g for 10 min. The pellet was washed with ice cold PBS, resuspended in ice cold 10% TCA, incubated for 10 min on ice and centrifuged at 10,000 x g for 10 min. For derivatization of thiols, 5 μL of the supernatant was mixed with 50 μL of 100 mM potassium phosphate buffer, pH 7.0 and 5 μL of 0.5 mM solution of monobromobimame in acetonitrile. After incubation for 40 min in the dark, 10 μL of the reaction mixture was injected into C 18 reverse phase HPLC column for analysis with a Varian Analytical Instruments (Walnut Creek, CA) HPLC system equipped with a Dynamax (Rainin Instrument Company, Inc.,Woburn, MA) fluorescence detector tuned to excitation at 380 nm and emission at 480 nm. The elution buffer was 140 mM ammonium acetate, pH 5.0 with the gradient of acetonitrile (15 – 25%) over 20 min.

Thiol Substrates in *Escherichia coli*

One of the first issues in elucidating the functions of the GSH transferase paralogues in *E. coli* is whether they interact with or utilize GSH. This question is complicated by the fact that there are two major forms of GSH in *E. coli*; GSH itself and glutathionylspermidine (GspSH), a condensation product of GSH and spermidine [17]. It is known that GSH is the predominant thiol under aerobic growth in log phase. However, in late stationary phase and particularly under anaerobic conditions most of the thiol is found as GspSH [18].

GspSH is formed by the enzyme glutathionylspermidine synthetase/amidase (GSS), which is a bifunctional enzyme that catalyzes the ATP-dependent condensation of the glycylcarboxylate of GSH with N1 (or the short arm) of spermidine as illustrated in Figure 3. The enzyme also catalyzes the hydrolysis of GspSH to give GSH and spermidine. The two opposing activities of the enzyme obviously need to be regulated but the mechanism of that regulation is not known [19, 20]. The fact that the *yghU* gene is located adjacent to the gene encoding GSS led us to the initial hypothesis that YghU might be a protein that regulates GSS activity [14]. A number of experiments using purified GSS and YghU revealed that YghU had no detectable influence on the activity of GSS either in the forward or reverse reactions.

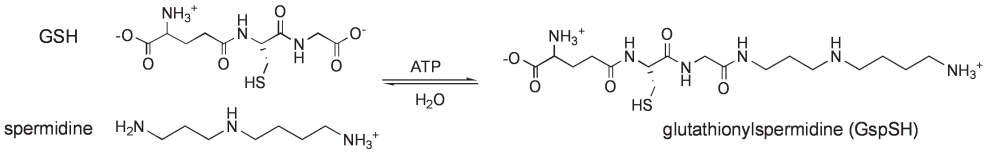

Figure 3. Reactions catalyzed by glutathionylspermidine synthetase/amidase (GSS).

Gene Knockouts

In order to examine the effect of YghU on GSS in a cellular assay we disrupted each gene in separate experiments. The loss of the gene is illustrated in Figure 4 where the absence of the *yghU* or the *gss* message is clear from the RT-PCR experiment as compared to the wild–type organism. The effect of the disruptions on the levels of GSH and GspSH under anaerobic conditions is illustrated in Figure 5. As anticipated from the published literature [18], in minimal media under anaerobic conditions, GspSH was the predominant thiol (81%). The *yghu* gene knockout decreased the amount of GspSH to 65% of the thiol total which is essentially within the error of the experimental measurement. Disruption of the *gss* gene essentially eliminated the GspSH (< 5%) from the thiol pool. The conclusion from these results is that the YghU protein does not regulate the GSH/GspSH tone in the cell to a significant extent under these conditions.

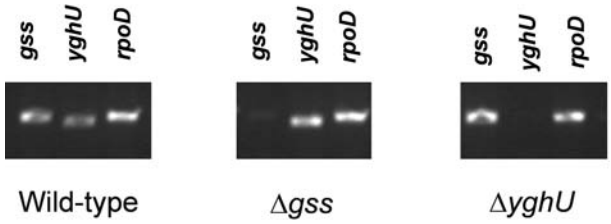

Figure 4. Reverse transcriptase polymerase chain reaction (RT-PCR) analysis of the messenger RNA for glutathionylspermidine synthetase/amidsase (gss) and YghU in wild-type *E. coli* cells BW25113 (left), *E. coli* cells BW25113 (D*gss*) where the *gss* gene has been disrupted (middle) and *E. coli* cells BW25113 (D*yghU*) where the *yghU* gene has been disrupted (right). Note the absence of messenger RNA in the two knockouts.

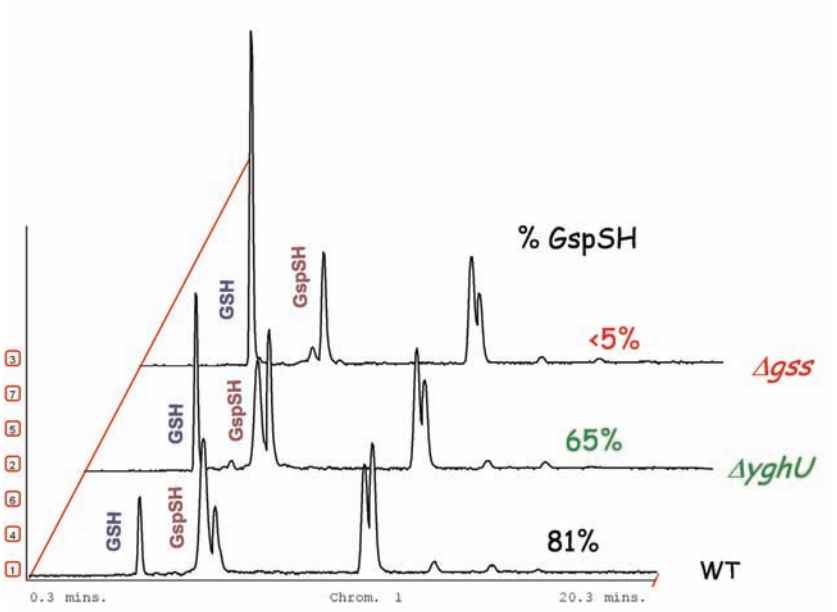

Figure 5. Analysis of the thiol content of *E. coli* cells BW25113 grown in M9 minimal media to stationary phase under anaerobic conditions. The bottom trace (wt) is for wild-type cells. The middle (D*yghU*) and top (D*gss*) traces are results for the *yghU* and *gss* gene knockouts, respectively. The unlabeled peaks are due to decomposition products of the fluorescence reagent.

Gene Expression Levels

The measurement of gene expression levels offers another view of how proteins influence the biology of a cell under particular environmental conditions. The conditions can represent points in the growth of the organism, nutrient status, or physical or chemical stress. Gene expression can be measured either by mRNA levels in a cell or by direct measurement of protein expression levels. The semi-quantitative measurement of mRNA levels by RT-PCR is a cost effective way of examining gene expression levels. Figure 6 illustrates the gene expression levels for all eight GSH transferase paralogues and glutathionylspermidine synthetase/amidase under normal aerobic growth conditions.

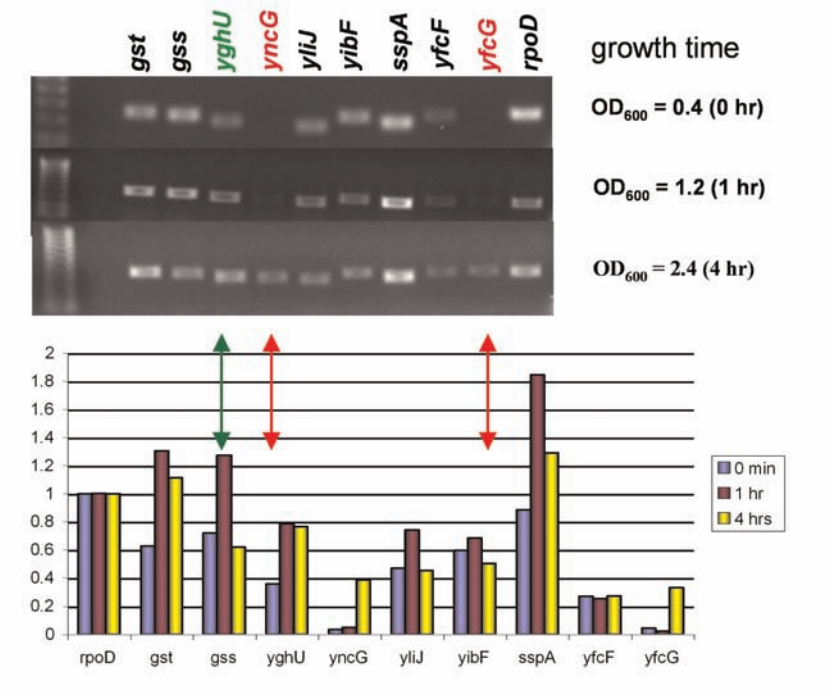

Figure 6. mRNA levels for *gss* and the eight GSH transferase paralogues as a function of growth time in LB media and determined by RT-PCR analysis. The bottom panel shows the mRNA levels normalized to the *rpoD* message. Note that two of the GSH transferase paralogues, *yncG* and *yfcG* (red arrows), exhibit large = 10-fold increases in message in late stationary phase while the *yghU* (green arrow) gene exhibits robust expression through out growth.

The normalized gene expression levels reveal only modest (= 2-fold changes) in gene expression as a function of growth time for most of the genes. The two exceptions are the *yncG* and *yfcG* genes that exhibit 10 to 30-fold increases in expression in late stationary phase. Three interesting observations can be made from these data. The first is that the increased expression level of both genes coincides with the elevation of GspSH in late stationary phase suggesting that there may be a connection between GspSH and the YncG and YfcG proteins. The second observation is that YncG and YfcG are distinctly different with respect to the GSH transferase families to which they belong (Figure 2). The enhanced expression of these two genes in late stationary phase has not been reported in microarray data. In contrast, *yfcG* has been reported to be down-regulated by the absence of the transcriptional regulator Fis (factor for inversion stimulation) in stationary phase [21].

If the *yncG* and *yfcG* genes are expressed in late stationary phase when the synthesis of GspSH takes place, it might be expected that the gene products would preferentially interact with GspSH as opposed to GSH. In preliminary work, we have found, by fluorescence

titration, that the YfcG protein binds GspSH 10-fold more tightly (K_d= 29 ± 7 μM) than it does GSH (K_d = 329 ± 6 μM). It would then appear that YfcG has a significant preference for binding GspSH and we conclude that the protein plays some role in the biochemistry of GspSH.

Criteria for the Assignment of Enzyme Function

The criteria for the assignment for protein or enzyme function are as varied as the function of any given protein is complicated. At a minimum, a protein needs to be characterized with respect to what other molecules it interacts with, including small molecules or substrates and other macromolecules. The latter can be accomplished with various types of pull-down assays using the protein of interest as bait. Ideally, the molecular interactions with large or small molecules should be characterized in as much structural detail as possible by X-ray crystallography or NMR spectroscopy.

The temporal or environmental influence on gene expression levels is also often a valuable piece of information as demonstrated above. The viability or sensitivity of an organism to gene knockouts can also reveal essential clues as to the biological role of a particular protein. These clues can be detected by with a variety of techniques including metabolomics (the appearance or loss of metabolites), proteomics (the appearance or loss of specific proteins) and genetics (the interaction of one gene with another). Needless to say, the stringency of the criteria for defining enzyme function can vary enormously from simply elucidating what kind of reaction an enzyme catalyzes to more global questions as to why a particular reaction is important to a given organism under specific circumstances.

Acknowledgement

This work was supported by National Institutes of Health Grants R01 GM030910, T32 ES 007028 and T32 GM008320, P30 ES 000267 and R01 GM060595.

References

[1] Armstrong, R.N. (1995) Structure, Catalytic Mechanism and Evolution of the Glutathione Transferases. *Chem. Res. Toxicol.* **10**:2 – 18.

[2] Vuilleumier, S., Pagni, M. (2001) The elusive roles of bacterial glutathione S-transferases: New lessons from genomes. *Appl. Microbiol. Biotechnol.* **58**:138 – 146.

[3] Oakley, A.J. (2005) Glutathione transferases: new functions. *Curr. Opin. Struct. Biol.* **15**:716 – 723.

[4] Stourman, N.V., Rose, J.A., Vuilleumier, S., Armstrong, R.N. (2003) Catalytic mechanism of dichloromethane dehalogenase from *Methylophilus* sp. strain DM 11, *Biochemistry* **42**:11048 – 11056.

[5] Thompson, L.C., Ladner, J.E., Codreanu, S., Harp, J., Gilliland, G.L., Armstrong, R.N. (2007) 2-Hydroxychromene-2-carboxylic acid isomerase: a Kappa class glutathione transferase from *Pseudomonas putida*. *Biochemistry* **46**:6710 – 6722.

[6] Williams, M.D., Ouyang, T.X., Flickinger, M.C. (1994) 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.* **11**:1029 – 1060.

[7] Hansen, A.M., Qui, Y., Yeh, N., Blattner, F.R., Durfee, T., Jin, D.J. (2005) SspA is required for acid resistance in stationary phase by down regulation of H-NS in *Escherichia coli. Mol. Microbiol.* **56**:719 – 734.

[8] Hansen, A.M., Gu, Y., Li, M., Andrykovitch, M., Waugh, D.S., Jin, D.J., Ji, X. (2005) Structural basis for the function of stringent starvation protein A as a transcription factor. *J. Biol. Chem.* **280**:17380 – 17391.

[9] Bai, M., Zhou, J.M., Perett, S. (2003) The yeast prion protein Ure2 shows glutathione peroxidase activity in both native and fibrillar forms. *J. Biol. Chem.* **279**:50025 – 50030.

[10] http://www.ncbi.nlm.nih.gov/pubmed/16275904?ordinalpos=32&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum (2005) The transduction of the nitrogen regulation signal in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **102**:16537 – 16538.

[11] Koonin, E.V., Mushegian, A.R., Tatusov, R.L., Altschul, S.F., Bryant, S.H., Bork, P., Valencia, A. (1994) Eukaryotic translation elongation factor 1μa; contains a glutathione transferase domain-Study of a diverse, ancient protein superfamily using motif search and structural modeling. *Protein Sci.* **3**:2045 – 2054.

[12] Dulhunty, A., Gage, P., Curtis, S., Chelvanayagam, G., Board, P. (2001) The Glutathione Transferase Structural Family Includes a Nuclear Chloride Channel and a Ryanodine Receptor Calcium release Channel Modulator. *J. Biol. Chem.* **276**:3319 – 3323.

[13] Harrop, S.J., DeMaere, M.Z., Fairlie, W.D., Reztsova, T., Valenzuela, S.M., Mazzanti, M., Tonini, R., Qui, M.R., Jankova, L., Warton, K., Bauskin, A.R., Wu, W.M., Pankhurst, S., Campbell, T.J., Breit, S.N., Curmi, P.M.G. (2001) Crystal structure of the soluble form if the intracellular chloride Ion channel CLIC 1 (NC 27) at 1.4 Å resolution. *J. Biol. Chem.* **276**:44993 – 45000.

[14] Rife, C.L., Parsons, J.F., Xiao, G., Gilliland, G.L., Armstrong, R.N. (2003) Conserved structural elements in glutathione transferase homologues encoded in the genome of *Escherichia coli*. *Proteins: Struct. Func. Genetics* **53**:777 – 782.

[15] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage D., Amin N., Schwikowski, B., Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **13**:2498 – 2504.

[16] Datsenko, K.A., Wanner, B. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97**:6640 – 6645.

[17] Tabor, H., Tabor, C.W. (1975) Isolation, characterization and turnover of glutathionylspermidine from *Escherichia coli*. *J. Biol. Chem.* **250**:2649 – 2654.

[18] Smith, K., Borges, A., Ariyanayagam, M.R., Fairlamb, A.H. (1995) Glutathionylspermidine metabolism in *Escherichia coli*. *Biochem. J.* **312**:465 – 469.

[19] Bollinger, J.M. Jr., Kwon, D.S., Huisman, G.W. Walsh, C.T. (1995) Glutathionylspermidine metabolism in *Escherichia coli*. *J. Biol. Chem.* **270**:14031 – 14041.

[20] Lin, C.-H., Kwon, D.S., Bollinger, J.M. Walsh, C.T. (1997) Evidence for a glutathionyl-enzyme intermediate in the amidase activity of the bifunctional glutathionylspermidine synthetase/amidase from *Escherichia coli*. *Biochemistry* **36**:14930 – 14938.

[21] Bradley, M.D., Beach, M.B., Jason de Koning, A.P., Pratt, T.S., Osuna, R. (2007) Effects of Fis on *Escherichia coli* gene expression during different growth stages. *Microbiology* **153**:2922 – 2940.