

Analysis of the glutathione S-transferase (GST) gene family

Daniel W. Nebert^{1*} and Vasilis Vasiliou²

¹Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH 45267-0056, USA

²Molecular Toxicology and Environmental Health Sciences Program, Department of Pharmaceutical Sciences, University of Colorado Health Sciences Center, Denver, CO 80262, USA

*Correspondence to: Tel: +1 513 558 4347; Fax: +1 513 558 3562; E-mail: dan.nebert@uc.edu

Date received (in revised form): 4th August 2004

Abstract

The glutathione S-transferase (GST) gene family encodes genes that are critical for certain life processes, as well as for detoxication and toxification mechanisms, via conjugation of reduced glutathione (GSH) with numerous substrates such as pharmaceuticals and environmental pollutants. The GST genes are upregulated in response to oxidative stress and are inexplicably overexpressed in many tumours, leading to problems during cancer chemotherapy. An analysis of the GST gene family in the Human Genome Organization-sponsored Human Gene Nomenclature Committee database showed 21 putatively functional genes. Upon closer examination, however, GST-kappa I (*GSTK1*), prostaglandin E synthase (*PTGES*) and three microsomal GSTs (*MGST1*, *MGST2*, *MGST3*) were determined as encoding membrane-bound enzymes having GST-like activity, but these genes are not evolutionarily related to the GST gene family. It is concluded that the complete GST gene family comprises 16 genes in six subfamilies — alpha (*GSTA*), mu (*GSTM*), omega (*GSTO*), pi (*GSTP*), theta (*GSTT*) and zeta (*GSTZ*).

Keywords: human genome, glutathione S-transferase gene family, microsomal glutathione S-transferases, prostaglandin E synthase, MAPEG family, DsbA-like thioredoxin domain

Introduction

One goal of this 'Update on Genome Completion and Annotations' series^{1,2} has been to select a gene, or gene family, check for accuracy in the databases, and then help to suggest ways to correct any nomenclature problems that might exist. The glutathione S-transferases (GSTs) represent an important group of enzymes which detoxify both endogenous compounds and foreign chemicals such as pharmaceuticals and environmental pollutants. Although a large number of reviews about this important enzyme family have appeared,^{3–11} there continues to be considerable confusion in the field with regard to the naming and classification of these genes and gene products.

Homologous genes, having a common ancestral origin 2 billion years ago or more, can be identified more readily, if they are designated with a stem (or root) symbol. A root symbol is very much encouraged by the Human Gene Nomenclature Committee (HGNC) as the basis for a hierarchical series of genes (eg for the *ABC* family, subfamily *A*, *ABCA1*, *ABCA2*, *ABCA3*, *ABCA4*) that are either the result of evolutionary divergence of an ancient ancestral gene, or have conserved functions — via pathways, interactions or protein domains. Such a root symbol allows the easy

identification of other related members in both database searches and the literature.

Homologous regions of 15–25 per cent of nucleotides or amino acids can be detected by the various alignment programs, denoting divergence from an ancestral gene; a small almost-invariant DNA motif or protein domain — functioning as an enzyme active-site, cofactor docking site or ligand-binding site — is further evidence of divergence from an ancestral gene. One of the earliest examples of this nomenclature approach for homologous genes was the cytochrome P450 (*CYP*) gene superfamily, in which it was agreed that approximately 40 per cent or more amino acid similarity allows two members to be placed in the same family and about 55 per cent or greater similarity allows two members to be assigned to the same subfamily.¹ These cut-off values follow the original recommendations by Margaret Dayhoff. At present, more than 130 additional gene superfamilies and large gene families have since followed this same format.¹

Biochemistry of the GST enzymes

The fundamental basis for all GST catalytic activities is the capacity of these enzymes to lower the pK_a of the sulfhydryl

group of reduced glutathione (GSH) from 9.0 in aqueous solution to about 6.5 when GSH is bound in the active site.¹² GSH exists as the thiolate (GS^-) anion at neutral pH when complexed with the GST enzyme. Catalysis by GST occurs through the combined capacity of the enzyme to promote GS^- formation and to bind hydrophobic electrophilic compounds at a closely adjacent site.³ The GSH-binding and the hydrophobic substrate-binding sites have been called the G- and H-sites, respectively.¹³ In the case of certain substrates (eg benzyl and phenethyl isothiocyanates, alkyl dihalides), GST can catalyse both the forward and reverse reactions, leading to increased toxicity rather than detoxication.³ The active cytosolic enzyme exists as a dimer of two subunits.^{3,4}

Evolution of the GST genes

GSTs are widely distributed in nature — from bacteria and yeast to plants and animals. Plant GSTs include the phi, tau, theta, zeta and lambda classes; the theta and zeta have counterparts in animals.^{4,5} The sigma and theta classes are abundant in non-vertebrate animals.³ There is significant homology between a class theta GST and a dichloromethane dehalogenase enzyme from the prokaryote *Methylobacterium*,¹⁴ suggesting that the ancestral progenitor for mammalian GSTs probably arose from the theta class.

The analysis in this review will focus only on human GST genes. Numerous polymorphisms exist in the human GST genes,^{10,11} including the complete absence of the *GSTM1* or the GST theta 1 gene — at frequencies as high as 20 per cent to 50 per cent in some populations. Given the absence in certain GST activities, one can see how this might lead to decreased detoxication of environmental carcinogens or chemotherapeutic agents and thus to clinical problems in patients lacking these genes. Evidence is also emerging that GST genes from some pathogens might exert immunomodulatory functions towards the immune system, involving separate profiles of cytokine gene transcription and different patterns of cell growth.⁶ Antioxidants, as well as oxidative stress, induce transcription of many of the GST genes,^{8,9} leading to increased protection of the cell against insult by environmental chemicals and drugs.

Cytosolic versus membrane-bound GSTs

Many of the GST reviews include membrane-bound as well as cytosolic enzymes.^{4,7} Microsomal GST¹⁵ and leukotriene C₄ synthase¹⁶ have been described as members of the GST family, although it has been noted³ that neither shares sequence identity with the cytosolic GSTs. It would therefore appear likely that these membrane-bound GST enzymes represent

examples of *convergent*, rather than divergent, evolution; at a particular point in time during evolution, Mother Nature required an enzyme to carry out such a membrane-bound catalytic reaction and assigned that task to an enzyme class different from that of the cytosolic GSTs.

The real GSTs have the two domains GST_N¹⁷ and GST_C.¹⁸ One or the other of these domains appears in a number of other proteins. This might explain why some other proteins exhibit GST-like activity.

HGNC search for GST genes

A quick search of the GST gene database (Table 1) showed 22 putatively functional genes and five pseudogenes. Upon closer inspection,¹⁹ it was determined that *GSTM1L* is a pseudogene — which changes the number to 21 functional genes and six pseudogenes. There is a cluster of five *GSTA* genes located at Chr 6p12; a cluster of five *GSTM* genes at 1p13; two *GST-omega* genes at 10q25.1; *GSTP1* at 11q13-qter; two *GST-theta* genes at 22q11.2; and a single gene *GSTZ1* at 14q24.3. Pseudogenes are often found at different chromosomal locations from the cluster of functional genes from which the pseudogenes originated. Interestingly, *GSTZ1* is identical to maleylacetoacetate isomerase, a key enzyme in tyrosine catabolism, catalysing the GSH-dependent isomerisation of maleylacetoacetate to fumarylacetoacetate.²⁰

It will be shown that the remaining five genes — *GSTK1*, *PTGES* and the three microsomal GSTs — are not evolutionarily part of the GST gene family. Phylogenetic analysis (Figure 1) places these five genes at one edge of the putative evolutionary tree, at almost the same distance as the *GSTW*, *GSTZ1* and *GSTQ* genes at the other edge. The tree-making algorithm, however, scores sequence similarity between each pair of protein sequences. What the tree shows is that the omega, zeta and theta GSTs are almost as dissimilar to the other more typical GSTs as are the *GSTK1*, *PTGES* and three microsomal GST proteins.

By means of CLUSTAL alignment,²¹ consensus GST_N and GST_C domains were found, plus significant stretches of sequence alignment, for the above-mentioned 16 GST genes (Figure 2), but none of these were found in the other five genes (data not illustrated). Upon further analysis, it was discovered that the microsomal GSTs, as well as prostaglandin E synthase, belong to the membrane-associated proteins in eicosanoid and glutathione metabolism (MAPEG) gene family (pfam01124). Due to structural similarities in the active sites of 5-lipoxygenase-activating protein (FLAP), leukotriene C₄ synthase and prostaglandin E synthase, substrates for each enzyme can compete with one another and modulate synthetic activity.²² By contrast, *GSTK1* has a bacterial disulphide-bond-A (DsbA)-like thioredoxin domain (pfam01323) and is a member of a diverse set of proteins

Table 1. List of human *GST* putatively functional genes.

Approved gene symbol	Approved gene name	Chromosomal location
<i>GSTA1</i>	Glutathione S-transferase (alpha) A1	6p12
<i>GSTA2</i>	Glutathione S-transferase A2	6p12.2
<i>GSTA3</i>	Glutathione S-transferase A3	6p12
<i>GSTA4</i>	Glutathione S-transferase A4	6p12
<i>GSTA5</i>	Glutathione S-transferase A5	6p12.1
<i>GSTAP1</i>	Glutathione S-transferase A pseudogene 1	Reserved
<i>GSTAP2</i>	Glutathione S-transferase A pseudogene 2	Reserved
<i>GSTK1</i>	Glutathione S-transferase kappa 1	7q35
<i>GSTM1</i>	Glutathione S-transferase (mu) M1	1p13.3
<i>GSTM1L (GSTM1P*)</i>	Glutathione S-transferase 'M1-like' pseudogene	3p25.2
<i>GSTM2</i>	Glutathione S-transferase M2 (muscle)	1p13
<i>GSTM3</i>	Glutathione S-transferase M3 (brain)	1p13.3
<i>GSTM3P</i>	Glutathione S-transferase M3 pseudogene	20
<i>GSTM4</i>	Glutathione S-transferase M4	1p13.3
<i>GSTM5</i>	Glutathione S-transferase M5	1p13.3
<i>GSTO1</i>	Glutathione S-transferase omega 1	10q25.1
<i>GSTO2</i>	Glutathione S-transferase omega 2	10q25.1
<i>GSTO3P1</i>	Glutathione S-transferase omega 3 pseudogene	3q22.1
<i>GSTP1</i>	Glutathione S-transferase (pi) P1	11q13-qter
<i>GSTPP</i>	Glutathione S-transferase pi pseudogene	12q13-q14
<i>GSTT1</i>	Glutathione S-transferase theta 1	22q11.2
<i>GSTT2</i>	Glutathione S-transferase theta 2	22q11.2
<i>GSTZ1</i>	Glutathione S-transferase (zeta) Z1 (maleylacetoacetate isomerase)	14q24.3
<i>MGST1</i>	Microsomal glutathione S-transferase 1	12p12.3-p12.1
<i>MGST2</i>	Microsomal glutathione S-transferase 2	4q28-q31
<i>MGST3</i>	Microsomal glutathione S-transferase 3	1q23
<i>PTGES</i>	Prostaglandin E synthase	9q34.3

* Correct gene name, according to the rules listed under <http://www.gene.ucl.ac.uk/nomenclature/>.

with a thioredoxin-like structure (pfam00085).²³ It therefore appears that *GST-kappa* has been misnamed in protein sequence databases, because it is clearly not a member of the *GST* gene family. Evolutionarily speaking, neither are the *PTGES* nor the three microsomal *GST* genes members of the *GST* gene family.

Greek-to-Latin alphabetic conversions

Finally, of the six *GST* subfamilies, two of these are misnamed in the HGNC database, according to its own guidelines (Table 1). The two functional genes and one pseudogene of

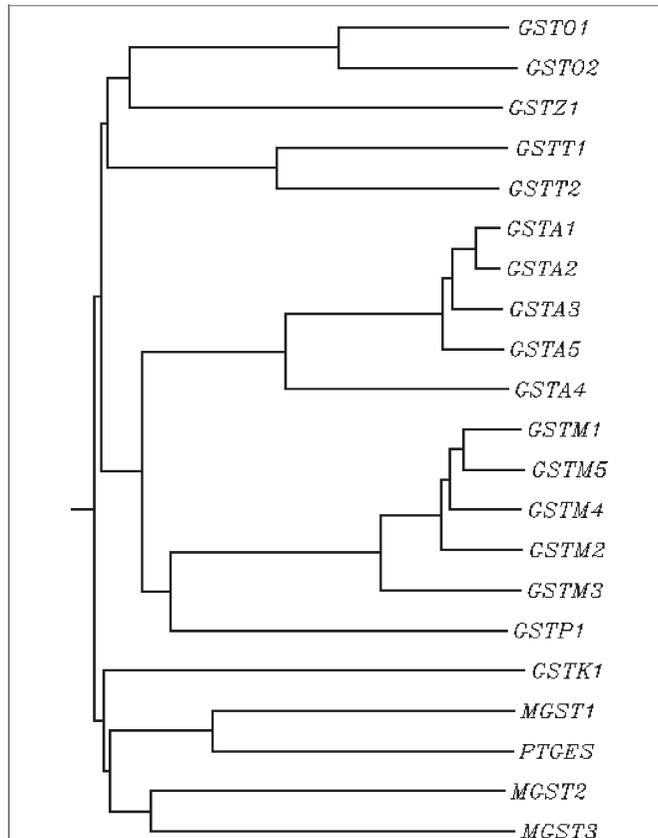


Figure 1. Dendrogram of the 21 genes under discussion. In order to decrease clutter and confusion, it has been chosen not to include three alternative splice variants of *GSTM4*, two variants of *GSTM1* and three variants of *GSTZ1* in construction of this tree. The X-axis denotes evolutionary time, which is not precise but extends over hundreds of millions of years.

the omega class should correctly be named *GSTW1*, *GSTW2* and *GSTW3P1*, respectively; the symbol ‘W’ stands for ‘omega’, whereas the symbol ‘O’ stands for ‘omicron’. Similarly, the two functional genes of the theta class should correctly be named *GSTQ1* and *GSTQ2*, because the symbol ‘Q’ stands for ‘theta’, whereas the symbol ‘T’ stands for ‘tau’. Plants contain *GST* tau genes.^{4,5}

The HGNC addressed this ‘Greek letter’ issue in relation to the *GST* genes. *GSTT1* and *GSTT2* were approved in 1994, in line with a request from Board’s laboratory²⁴ and have been widely used ever since, with *GSTT1* especially appearing in hundreds of references listed in PubMed. Likewise, Board’s group published work about the *GSTO1*, *GSTO2* and *GSTOP3* genes,²⁵ which were approved by the HGNC in 2003. HGNC therefore concluded:

‘Although we do indeed have guidelines for Greek letter conversions, we also aim to serve the community by providing a useable and used nomenclature. It would seem to us to be somewhat pedantic to change the symbols for these five genes, all of which are being widely used in publications, simply because they

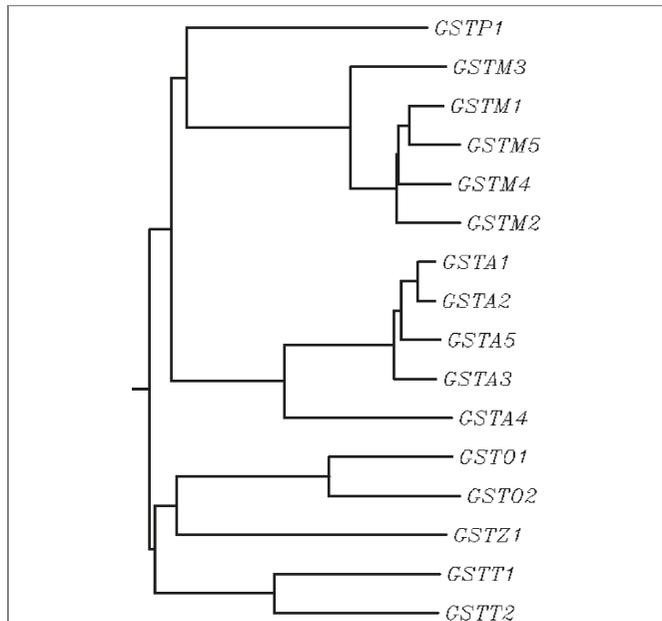


Figure 2. Dendrogram of the 16 *GST* genes that are *bona fide* members of the *GST* family. Again, for the sake of simplicity, the additional splice variants of *GSTM4*, *GSTM1* and *GSTZ1* have been excluded. The X-axis denotes evolutionary time, which is not precise but extends over hundreds of millions of years.

did not conform to a guidance conversion table. In a similar manner, we usually use ‘G’ for gamma, but sometimes ‘C’ has been used instead, because this is taken as the third-letter-of-the-alphabet equivalent, eg laminin gamma-2 encoded by the LAMC2 gene.²⁶ Hence, we do not see a need to change these glutathione S-transferase symbols. We realize that people can become upset by nomenclature changes, and we believe that a working nomenclature system is more desirable than a perfect one’.

Since mouse nomenclature follows that of human, the Mouse Genomic Nomenclature Committee (MGNC) will similarly stay with these same symbols for the orthologous genes. Both HGNC and MGNC continue to work closely with experts in the field, and the committees certainly make changes to the nomenclature, based on information from the experts when necessary. In most instances, changes will be made if they are necessary in order to promote accuracy and consistency.

Conclusions

The *GST* gene family comprises 16 genes in six subfamilies. Several problems were found in the HGNC listings and nomenclature for the *GST* gene family. First, *GSTM1L* is a pseudogene. Secondly, there are five additional genes included (*GSTK1*, *MGST1*, *MGST2*, *MGST3* and *PTGES*) that encode membrane-bound enzymes having *GST*-like activity but which are not evolutionarily related to the 16 true *GST* genes. Thirdly, according to the Human Genome

Organization HGNC's own rules, the GST-omega subfamily should include 'W' for omega — instead of 'O', which is reserved for omicron — and the GST-theta subfamily should include 'Q' for theta — instead of 'T', which is reserved for tau. And plants have a GST-tau subfamily. The present authors' analysis of the GST gene family simply underscores some of the problems encountered in the various databases. Similar nomenclature problems were seen with the mouse *Gst* genes (not shown). The authors estimate that it will take many years before all of the bumps and wrinkles can be ironed out of the nomenclature systems for human and mouse genes and gene families.

Acknowledgments

The writing of this article was funded, in part, by NIH grants P30 ES06096 (D.W.N.) and R01 EY11490 (V.V.).

References

- Nebert, D.W. and Wain, H.M. (2003), 'Update on human genome completion and annotations: Gene nomenclature', *Hum. Genomics* Vol. 1, pp. 66–71.
- Nebert, D.W., Sophos, N.A., Vasiliou, V. and Nelson, D.R. (2004), 'Update on human genome completion and annotations: Cyclophilin nomenclature problems, or a visit from the sequence police', *Hum. Genomics* Vol. 1, pp. 381–388.
- Hayes, J.D. and Pulford, D.J. (1995), 'The glutathione S-transferase gene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance', *Crit. Rev. Biochem. Mol. Biol.* Vol. 30, pp. 445–600.
- Sheehan, D., Meade, G., Foley, V.M. and Dowd, C.A. (2001), 'Structure, function and evolution of glutathione transferases: Implications for classification of non-mammalian members of an ancient enzyme superfamily', *Biochem. J.* Vol. 360, pp. 1–16.
- Dixon, D.P., Laphorn, A. and Edwards, R. (2002), 'Plant glutathione transferases', *Genome Biol.* Vol. 3, pp. S3004.
- Ouassii, A., Ouassii, M. and Sereno, D. (2002), 'Glutathione S-transferases and related proteins from pathogenic human parasites behave as immunomodulatory factors', *Immunol. Lett.* Vol. 81, pp. 159–164.
- Murakami, M., Nakatani, Y., Tanioka, T. and Kudo, I. (2002), 'Prostaglandin E synthase', *Prostaglandins Other Lipid Mediat.* Vol. 68–69, pp. 383–399.
- Owuor, E.D. and Kong, A.N. (2002), 'Antioxidants- and oxidants-regulated signal transduction pathways', *Biochem. Pharmacol.* Vol. 64, pp. 765–770.
- Rinaldi, R., Eliasson, E., Swedmark, S. and Morgenstern, R. (2002), 'Reactive intermediates and the dynamics of glutathione transferases', *Drug Metab. Dispos.* Vol. 30, pp. 1053–1058.
- Townsend, D. and Tew, K. (2003), 'Cancer drugs, genetic variation, and the glutathione-S-transferase gene family', *Am. J. Pharmacogenomics* Vol. 3, pp. 157–172.
- Coles, B.F. and Kadlubar, F.F. (2003), 'Detoxification of electrophilic compounds by glutathione S-transferase catalysis: Determinants of individual response to chemical carcinogens and chemotherapeutic drugs?', *Biofactors* Vol. 17, pp. 115–130.
- Armstrong, R.N. (1994), 'Glutathione S-transferases: Structure and mechanism of an archaetypical detoxication enzyme', *Adv. Enzymol. Relat. Areas Mol. Biol.* Vol. 69, pp. 1–44.
- Mannervik, B. (1985), 'The isoenzymes of glutathione transferase', *Adv. Enzymol. Relat. Areas Mol. Biol.* Vol. 57, pp. 357–417.
- La Roche, S.D. and Leisinger, T. (1990), 'Sequence analysis and expression of the bacterial dichloromethane dehalogenase structural gene, a member of the glutathione S-transferase gene superfamily', *J. Bacteriol.* Vol. 172, pp. 164–171.
- Morgenstern, R. and DePierre, J.W. (1983), 'Microsomal glutathione transferase. Purification in unactivated form and further characterization of the activation process, substrate specificity, and amino acid composition', *Eur. J. Biochem.* Vol. 134, pp. 591–597.
- Nicholson, D.W., Ali, A., Vaillancourt, J.P. et al. (1993), 'Purification to homogeneity and the N-terminal sequence of human leukotriene C₄ synthase: A homodimeric glutathione S-transferase composed of 18-kDa subunits', *Proc. Natl. Acad. Sci. USA* Vol. 90, pp. 2015–2019.
- <http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00043>.
- <http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF02798>.
- http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=default&list_uids=391511#ubor3_RefSeq.
- Polekhina, G., Board, P.G., Blackburn, A.C. and Parker, M.W. (2001), 'Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity', *Biochemistry* Vol. 40, pp. 1567–1576.
- CLUSTAL W alignment for DNA and proteins. <http://www.ebi.ac.uk/clustalw/>.
- <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=pfam01124&version=v2.00>.
- <http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=pfam01323&version=v2.00>.
- Tan, K.L., Webb, G.C., Baker, R.T. and Board, P.G. (1995), 'Molecular cloning of a cDNA and chromosomal localization of a human theta-class glutathione S-transferase gene (*GSTT2*) to chromosome 22', *Genomics* Vol. 25, pp. 381–387.
- Board, P.G., Coggan, M., Chelvanayagam, G. et al. (2000), 'Identification, characterization, and crystal structure of the omega class glutathione transferases', *J. Biol. Chem.* Vol. 275, pp. 24798–24806.
- Pulkkinen, L., Christiano, A.M., Airenne, T. et al. (1994), 'Mutations in the gamma-2 chain gene (*LAMC2*) of kalinin/laminin-5 in the junctional forms of epidermolysis bullosa', *Nat. Genet.* Vol. 6, pp. 293–297.