Diverging catalytic capacities and selectivity profiles with haloalkane substrates of chimeric alpha class glutathione transferases

Sanela Kurtovic, Abeer Shokeer and Bengt Mannervik

Department of Biochemistry and Organic Chemistry, Uppsala University, Box 576, SE-75123 Uppsala, Sweden

1To whom correspondence should be addressed.

E-mail: bengt.mannervik@biorg.uu.se

Six homologous Alpha class glutathione transferases of human, bovine, and rat origins were hybridized by means of DNA shuffling. The chimeric mutants were compared with the parental enzymes in their activities with several alkyl iodides. In order to facilitate a multivariate analysis of relationships between substrates and enzyme activities, three descriptors were introduced: ‘specific catalytic capacity’, ‘substrate selectivity’, and ‘unit-scaled substrate selectivity’. In some cases the purified mutants showed higher specific activity with a certain alkyl iodide than any of the parental enzymes. However, the overriding effect of DNA shuffling was the generation of chimeras with altered substrate selectivity profiles and catalytic capacities. The altered substrate selectivity profiles of some mutants could be rationalized by changes of the substrate-binding residues in the active site of the enzyme. However, in four of the isolated mutants all active-site residues were found identical with those of rat GST A2-2, even though their substrate specificity profiles were significantly different. Clearly, amino acid residues distant from firstsphere interactions with the substrate influence the catalytic activity. These results are relevant both to the understanding how functional properties may develop in natural enzyme evolution and in the tailoring of novel functions in protein engineering.

Keywords: DNA shuffling/glutathione transferase/multivariate data analysis/mutant library/substrate selectivity

Introduction

Directed evolution of protein function has gained much attention as a powerful tool for evolving new or improved properties (Powell et al., 2001; Yuan et al., 2005). Numerous studies have been reported where improved catalytic activity (Ness et al., 1999), thermal stability (Minagawa et al., 2007), acid tolerance (Patnaik et al., 2002), or enantioselectivity (Rui et al., 2005) have been achieved. Evolving enzymes are known to display promiscuous functions or broad substrate specificities, a feature believed to be important in gaining novel functions, both in vivo and in vitro (Aharoni et al., 2005; Khersonsky et al., 2006). Thus, in recent years, the importance of monitoring alterations of substrate selectivity profiles in directed evolution of proteins has been recognized (Hansson et al., 1999; Goddard and Reymond, 2004; Grognum and Reymond, 2004; Varadarajan et al., 2005).

Haloalkanes are important chemicals used as pharmaceuticals, herbicides, fungicides, insecticides, flame retardants, intermediates in organic synthesis, solvents, etc. (Fetzner and Lingens, 1994; van Pee and Unversucht, 2003). For example, 1,2-dihaloethanes (vicinal dihalogeno alkanes), such as 1,2-dibromoethane, have been widely used as fumigants and antiknock agents in leaded gasoline. These chemicals have the potential to induce toxicity leading to deaths among humans exposed to very high concentrations. In model studies, induction of tumors at different sites in rats has been observed (Anders, 2004). 1,2-Dibromoethane has also been shown to be a mutagen in many biological systems, including the Ames Salmonella typhimurium test, Neurospora crassa, Drosophila melanogaster, and mammalian cells (cf. Josephy et al., 2006). The industrial and agricultural use of 1,2-dibromoethane has been restricted for these reasons. Other haloalkanes, such as methylene chloride, chloroform, iodofrom, and carbon tetrachloride have also been employed in large quantities for both industrial and domestic use, until their toxicities and carcinogenicities were discovered (Burek et al., 1984; Nitschke et al., 1988; Weber et al., 2003).

Glutathione transferases (GSTs) catalyze the conjugation of the cellular tripeptide glutathione (GSH) with electrophilic substrates of both endogenous and exogenous origins. GSTs provide the cells in this way with necessary protection against many different cytotoxins (Josephy and Mannervik, 2006). On the basis of their sequence similarities, soluble mammalian GSTs can be divided into seven different classes designated, alpha, mu, omega, pi, sigma, theta and zeta (Mannervik et al., 2005). Cytosolic GSTs are dimeric enzymes composed of two equal-sized subunits that can form homo- or heterodimers. Each subunit is made up of two domains, an N-terminal a/b thioredoxin-like fold and a C-terminal all a-helical structure. The N-terminal domain harbors the highly conserved G-site providing the interactions needed for the tripeptide GSH to bind. The C-terminal domain provides important interactions for binding of the second electrophilic substrate to its cognate H-site. The H-site is not highly conserved among the GSTs and provides these enzymes with the ability to recognize and detoxify a highly varied assortment of electrophiles (Josephy and Mannervik, 2006).

GSTs catalyze the conjugation of alkyl halides with GSH, a reaction that can have either a detoxication or a bioactivation outcome depending on the substrate (Guengerich, 2005). The general mechanism is a substitution reaction in which a halide ion is displaced by the thiolate of GSH, and if the substrate contains a second halogen the product may detoxify a highly varied assortment of electrophiles (Josephy and Mannervik, 2006).

Protein Engineering, Design & Selection vol. 21 no. 5 pp. 329–341, 2008
Published online March 19, 2008 doi:10.1093/protein/gzn010
will react with nucleophilic groups in DNA and proteins. The reactions with haloalkanes are catalyzed by several different classes of mammalian GSTs (Cmarik et al., 1990). The theta class GSTs are thought to be the most active enzymes, at least with small haloalkanes (Meyer et al., 1991; Thier et al., 1993; Ploemen et al., 1995; Shokeer et al., 2005).

In the present investigation six alpha class GSTs have been subjected to DNA shuffling in order to evolve chimeras with altered substrate specificity profiles towards several iodoalkanes, including both moniodoalkanes and diiodoalkanes. Multivariate analysis of the kinetic data and suitable descriptors of the relationships among functional properties of the enzyme variants were introduced. The results demonstrate that it is possible to hybridize GST structures from the same class to obtain altered catalytic capacities and diverging substrate selectivity profiles with compounds undergoing essentially the same chemical reaction as well as being similar in structure. The results are relevant both to the understanding of molecular evolution and to the engineering of enzymes with novel substrate selectivities for applications in bioremediation and monitoring of waste streams and other environmental sites.

Materials and methods

Library construction

An alpha class GST mutant library (Kurtovic et al., 2008) was created by shuffling of cDNAs encoding hA2, hA3, bA1, rA2, rA3 and an hA1 mutant library (Widersten and Mannervik, 1995). The purified library DNA was ligated into EcoRI and SalI digested pGDeTac vector. Transformation of electrocompetent Escherichia coli (E. coli) XL-1 Blue cells (Stratagene, La Jolla, CA, USA) was carried out by electroporation with the resulting plasmid construct.

Preparation of bacterial lysates

Colonies from the library were randomly picked from agar plates. A total of 407 mutants as well as parental enzymes were grown over night at 37°C in a total volume of 2 ml LB medium with ampicillin (100 μg/ml). The bacterial cultures were thereafter diluted 100-fold in a total volume of 10 ml 2TY supplemented with 100 μg/ml of ampicillin and grown at 37°C for an additional 2 h before the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.2 mM, whereupon the bacterial culture was allowed to grow for additional 16 h. The bacteria were then harvested by centrifugation at 4°C for 10 min at 3500 r.p.m. Each bacterial cell pellet was resuspended in 250 μl of 0.1 M sodium phosphate buffer, pH 7.4, and lysed by treatment with lysozyme (0.2 mg/ml) on ice for 60 min. Freezing at −80°C and thawing at 37°C were performed three times. After the final thawing, the suspension was centrifuged at 4°C for 30 min at 15 000 g. The supernatants were collected and stored at −80°C.

Screening of the bacterial lysates with iodoalkanes using a colorimetric end-point assay

Screening of 407 bacterial lysates was performed with six haloalkanes, 1,2-diiodoethane (diiodoethane), 1,4-diiodobutane (diiodobutane), 1,6-diiodohexane (diiodohexane), 1,8-diiodooctane (diiodooctane), 1-iodohexane (iodohexane), and 3-iodopropylbenzene (3-IPB) utilizing a colorimetric end-point assay (Kurtovic et al., 2007). Structures of the substrates used in this study are shown in Fig. 1. The assay is based on iodide ion release by conjugation with GSH. Iodide, when oxidized, will react with starch and give rise to blue color development. The intensity of the blue color was scored by eye on a digital scale of 0−5. The scoring was performed by two investigators independently. Lysates of bacteria expressing parental enzymes were assayed in parallel to lysates of the chimeras. All measurements were performed in 100 mM sodium phosphate buffer, pH 7.4 in a total volume of 100 μl.

![Fig. 1. Iodoalkane substrates used in the screening of the GST mutant library.](https://academic.oup.com/peds/article-abstract/21/5/329/1555534)}
The iodide-releasing reactions were carried out with 2.5 mM GSH and 2.5 mM iodoalkane, except for 1,2-diiodoethane, which was used at 0.2 mM concentration. The reaction was left to run at 30°C for 5 h for all iodoalkanes except 1,2-diiodoethane, which was incubated for 15 min. The reaction was terminated by addition of 50 μl starch solution prior to adding 100 μl of 2%(v/v) H₂O₂ in 2M HCl. Blue color developed immediately after addition of the oxidant (Kurtovic et al., 2007).

DNA sequencing
Eight mutants were chosen for further studies based on the screening results. DNA was purified by using the Maxiprep DNA purification kit (QIAGEN). The concentration of the purified DNA was assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., USA). Purified DNA was sequenced by Uppsala Genome Center, Rudbeck Laboratory, using a Mega BACE 1000 instrument (GE Healthcare, Uppsala, Sweden).

Purification of His-tagged GST variants
Selected clones from the screening were used for expression of the variant GSTs in E. coli XL1-Blue as described by Johansson et al. (1999) and purified as follows. The bacterial cells were harvested, resuspended in binding buffer (20 mM sodium phosphate pH 7.4 containing 500 mM NaCl, and 20 mM imidazole), and lysed by treatment with lysozyme (0.2 mg/ml) on ice for 60 min followed by sonication. After centrifugation at 30 000g for 30 min at 4.0°C, the supernatant fraction containing the recombinant enzyme was loaded on a Ni-IMAC affinity column (His GraviTrapTM, GE Healthcare). Unbound protein was washed out with 10 ml binding buffer. The recombinant GST was eluted with 3 ml binding buffer containing 500 mM imidazole. The purified enzyme was dialyzed against 100 mM sodium phosphate buffer, pH 7.4. The homogeneity of the pooled material was checked with SDS–PAGE, using 12.5% (w/v) polyacrylamide gel. Protein bands were visualized by staining with Coomassie Brilliant Blue R-250. The purified enzyme was stored at −80°C. The protein contents were measured with the Bio-Rad protein assay (Bio-Rad).

Specific activity measurements of purified variants
The conjugation of GSH with iodoalkanes catalyzed by alpha class GST variants was measured spectrophotometrically by monitoring iodide ion release at 226 nm (Shokeer et al., 2005). The reaction was initiated by addition of the iodoalkane to 1.0 mM GSH and GST in 100 mM sodium phosphate buffer, pH 7.4 at 30°C. The substrates used were diiodoethane, diiodobutane, diiodohexane, iodohexane, and 3-IPB. The concentration of diiodohexane and 3-IPB used in the assay was 0.05 mM and the concentration of diiodobutane, diiodoethane, and iodohexane was 0.1 mM. The net extinction coefficient for iodide release in the GSH conjugation is $Δε_{226} = 4.5$ mM$^{-1}$ cm$^{-1}$ (Shokeer et al., 2005).

Data analysis
The biplot and dendrogram analyses were made using Simfit 5.7.6 (http://www.simfit.man.ac.uk). The principal component analysis (PCA) was made using Simca P+11 (Umetrics) software. The lyase data were normalized to unit variance with regard to substrates and mean centered prior to PCA. The specific activity data were not normalized prior to dendrogram analysis or catalytic capacity determinations. In other analyses standardizations were used as indicated in the text and figure legends.

Results
In the present study cDNA encoding three human class GSTs, hA1, hA2 and hA3, one bovine GST, bA1, as well as two rat GSTs, rA2 and rA3, were subjected to DNA shuffling (Kurtovic et al., 2008). The resulting library of GST chimeras was expressed in E. coli, and 407 bacterial clones were picked and screened with six different iodoalkane substrates, diiodoethane, diiodobutane, diiodohexane, diiodooc-tane, iodohexane and 3-IPB (Fig. 1). The iodide release was scored on a digital activity scale of 0–5. The functional analysis of the GST activities in bacterial lysates clearly demonstrated that the chimeras have diverging substrate selectivity profiles with the different iodoalkane substrates (Table I).

PCA of activity data unravels diverging groupings in the GST mutant library
Multidimensional data obtained by assaying several enzyme variants with a set of alternative substrates (cf. Tables I and II) can be considered as a set of enzyme vectors, where the activities in a particular row define the position of the vector in substrate-activity space. Enzyme vectors pointing in the same direction indicate similar functional properties of the proteins, whereas diverging vectors define enzymes with different properties. The columns corresponding to the n alternative substrates provide the coordinates of the enzyme vectors in n-dimensional space. Data of higher dimensions can be divided into subsets that allow visualization in two- or three-dimensional substrate plots (Broo et al., 2002). However, the simultaneous contributions of many substrates call for multivariate data analysis to define the major directions of the vectors and the influences of the individual substrates. If the rows in Tables I or II (i.e. coordinates of the vectors) are regarded as defining points of an enzyme cluster in n-dimensional activity space, principal component (PC) analysis describes the variability of the cluster in consecutive orthogonal directions of decreasing magnitude. PC1 defines the direction of highest variability, PC2 the next highest, etc. The PCs are calculated from the variance-covariance matrix of the data and are linear combinations of the variables provided by the different substrates. Score plots show the projection of PCs in two dimensions, and corresponding loading plots display the influences of the independent variables (i.e. activities with alternative substrates) on the directions of the PC axes (Krzanowski, 2000). Comparisons of the score plots of enzymes with plots of the loadings provide information about associations between enzymes and their activities with the respective substrates.

Applied to the current study of 413 bacterial lysates, the activities with the alternative substrates were subjected to PCA in five-dimensional substrate-activity space (Fig. 2). In the analysis the data obtained with diiodoethane were excluded, because the non-enzymatic background reaction could not be measured accurately due to the high reactivity of the substrate. The score plot of PC1 and PC2 shows that the parental GSTs, except rGST A2-2, are located close to the origin, whereas the mutants from the library diverge...
predominantly in the fourth quadrant, where rGST A2-2 is positioned (Fig. 2A). A smaller group deviates in the first quadrant, where no parental GST is found. Five mutants from the fourth quadrant (146, 197, 371, 408 and 419) and three mutants (244, 330 and 442) from the first quadrant appeared to have particularly distinctive properties, including long distances from the origin. The corresponding loading plot (Fig. 2B) shows that all substrates contribute to PC1, which can be regarded as a measure of overall enzyme activity. In the dimension of PC2, diiodooctane and diiodobutane are most influential in positioning the mutants in the first and fourth quadrants, respectively. The loadings of iodo-hexane, 3-IPB and diiodohexane are grouped more closely together near the PC1 axis and have only a minor influence on PC2. In the score plot of PC2 versus PC3, it is evident that the five mutants with functional similarities to rGST A2-2 still group with this enzyme in a single quadrant, and that the three mutants 244, 330 and 442 are clustered in another quadrant (Fig. 2C). The corresponding loading-plot shows that in the PC2/PC3 plane, 3-IPB and diiodoctane are the primary contributors to the separation of the two main clusters (Fig. 2D). A third cluster diverging from the parental GSTs in the PC2/PC3 plane is governed by the activity with diiodobutane. Thus, the PCA implies that many of the chimeras have separated from the parental enzymes in functional space and gained diverging substrate selectivity profiles compared with the parental enzymes.

Mutants in the library are composed of chimeric sequences

Primary protein structures (Fig. 3) were determined by DNA sequence analysis of the eight mutants chosen. In the DNA sequences there were between 2 and 7 crossovers, 4.6 on an average, corresponding to recombinations among the parental GST sequences. Translated into protein primary structures the chimeric GSTs can be considered as composed of sequences from between 2 and 6 different parental enzymes. The different library variants are to a large extent composed of rGST A2-2 sequence elements, in particular, in the region embracing amino acid residues 100–222 covering most of the C-terminal domain. However, all other parental enzymes are represented in the chimeric sequences as well.

Specific activities of purified chimeric GSTs reveal diverging substrate specificities

In lysates measurements, confounding factors such as expressivity, solubility and stability influence the outcome of the analysis. The results of the lysate screening suggested that all the chosen chimeras had higher activities with some substrates than the parental enzymes and that they also showed different substrate selectivity profiles. This was fully confirmed with the purified chimeras with regard to different substrate selectivity profiles. However, higher specific
activities were not noted for all the chosen mutants and not to the extent indicated by the lysate data (cf. Tables I and II). The differences between lysate data and specific activity measurements result from different expression levels among the chimeras and parental enzymes. This is strongly indicated when measuring the mutants from the library with the ‘general’ GST substrate, 1-chloro-2,4-dinitrobenzene, which also shows much higher activity with the chosen chimeras in lysates, but not in purified form (data not shown).

For this reason, the eight chimeras chosen were purified and subjected to specific activity measurements. The mutants represented variants with diverging substrate selectivities as well as higher enzymatic activities. The purified chimeras, as well as the parental GSTs, were assayed with five iodoalkanes: diiodoethane, diiodobutane, diiodohexane, iodohexane and 3-IPB (Table II). Specific activities of the purified enzymes were not determined with diiodooctane, because of the low solubility of the substrate. In purified form, two of the eight chimeras, mutants 244 and 419, were more active with at least one substrate than any of the parental enzymes.

Mutant 419 is more active with three substrates, 3-IPB, iodohexane and diiodobutane. On the other hand, this chimera has suppressed specific activities with diiodohexane and diiodoethane compared with the parental enzymes displaying the highest specific activities with these substrates. Mutant 244 shows overall the same activity profile. With iodohexane, mutant 244 gives the highest activity of any of the GSTs tested (Table II). However, the activities of mutant 244 with all other substrates are lower than those of mutant 419. It is thus clear that the chimeras show diverging substrate selectivity profiles compared with the parental enzymes and that some mutants have differentially enhanced activities.

Specific catalytic capacities and substrate selectivities of the GST variants

In Fig. 4A the absolute values of the specific activities of chimeras together with parental enzymes are visualized in a bar plot. The length of a bar represents the overall catalytic capacity with the substrates used per mg of the studied enzymes. By this criterion, four of the parental GSTs are
more active than all the chimeras, with the exception of mutant 371. However, most of the activity is accounted for by one substrate, i.e. diiodoethane. In contrast, when the specific activities of a given enzyme are transformed into relative contributions (Fig. 4B), the substrate selectivity profile of each individual enzyme becomes detectable. It can be seen that hGST A2-2 is the most selective of the parental enzymes, displaying most of its activity with diiodoethane and very little with the other substrates. In contrast, rGST A2-2 is the least specific of the parental enzymes. The other parental GSTs are ranked between hGST A2-2 and rGST A2-2 in their substrate selectivities. hGST A1-1 and hGST A2-2 have high similarity in their primary structures (95% sequence identity), but in their functional profiles differences are clear. Thus, in spite the high sequence similarity, hGST A1-1 has broader functional profile than hGST A2-2. Most of the chosen chimeras display broad substrate specificities and are consequently similar to rGST A2-2 in their functions. However, mutants 371 and 408 have narrower substrate selectivity profiles than rGST A2-2, thus being more divergent in their functions compared with the other purified chimeras (Fig. 4B).

Multivariate biplot analysis of GST activities with alternative substrates

Each variant GST can be represented as a vector in the five-dimensional substrate-activity space and their relationships are reflected in directions and lengths of the vectors. However, for visualization, the five-dimensional data can be projected down to two dimensions by being subjected to a biplot analysis (Fig. 5). A multivariate biplot is a two-dimensional representation of a data matrix where each row is represented as a vector and each column is represented as a vector and chosen so that any element of the matrix is exactly the inner product of the vectors (Gabriel, 1971). Thus, the influence of substrate vectors on the distribution of enzyme vectors is visualized. In the analysis of untransformed specific activity values 98% of variation was accounted for in the two dimensions, and the biplot therefore accurately depicts relationships among the vectors. The length of an enzyme vector reflects the specific activity of that particular chimera, and the distribution of the substrate vectors indicate the association between the activities given with the substrates and the juxtaposed enzyme vectors. In Fig. 5 it is thus clear that hGST A2-2 has the highest specific activity with diiodoethane, while rGST A3-3, still dominated by the same substrate, has a lower value. Diiodoethane is dominating the analysis, because it is the most active substrate, while the vectors of the other substrates are essentially perpendicular to that of diiodoethane (Fig. 5). A similar grouping of the substrates based on the specific activities of the GST variants also appears in a dendrogram analysis (Table II). The directions of the substrate vectors reflect substrate selectivity, arcing from hGST A2-2 with high selectivity for diiodoethane, to rGST A2-2, which has a broader substrate selectivity profile, where diiodobutane, diiodohexane, iodohexane and 3-IPB are clearly represented. The other parental GSTs have intermediate substrate specificities ranging between those of hGST A2-2 and rGST A2-2. The directions of the vectors of all mutants, except 371, are

Fig. 3. Amino acid sequence alignment including all parental GSTs and selected mutants from the library. G-site residues are shaded in gray and H-site residues are identified in boxes. Amino acid residues identical to those in hGST A2-2 are represented as dots.

S.Kurtovic et al.

334
similar to that of rGST A2-2 thus indicating broad substrate selectivity profiles. Mutant 371, on the other hand, shows a more restricted substrate selectivity profile (Fig. 4B) being located quite close to hGST A2-2 and rGST A3-3 in the biplot (Fig. 5).

Imminent properties of library chimeras become highlighted when specific activities are scaled to unit length

The large impact on the analysis of the relatively high specific activity of diiodoethane can be overcome by scaling the contribution of each substrate to unit length. In this way, each substrate will contribute equally to the analysis (Fig. 6). The scaled specific activities of the chimeras and the parental enzymes are shown in a three-dimensional column plot (Fig. 6A). The mean coefficient of variation overall was estimated at 8%. For the individual substrates the mean coefficients of variation were as follows: diiodoethane, 6.7%; diiodobutane, 7.1%; diiodohexane, 9.0%; iodohexane, 7.6% and 3-IPB, 9.3%. Figure 6B shows a biplot based on the scaled specific activities. In this analysis, hGSTs A1-1, A2-2 and A3-3 join chimera 371 with diiodoethane and diiodohexane as signature substrates. Parental enzyme rGST A2-2 groups with mutants 146, 197, 419 and 442 in another direction. rGST A3-3 and bGST A1-1 come together with chimeras 330 and 408 in a direction in between the other two groupings. However, when compared individually within a group, it is clear that the chimeras show diverging substrate selectivity profiles (Fig. 6A), which can remain undetected in an overall biplot analysis. Mutants 146, 197, 244 and 330 group together with rGST A2-2 in the biplot analysis, but show clearly diverging substrate specificities in Fig. 6A. For example, mutant 197 has gained higher specific activities with diiodobutane and 3-IPB compared with rGST A2-2, but kept the other activities at the same level as rGST A2-2.

Fig. 4. Stacked bar plot representation of the specific activities of purified GSTs measured with five alternative iodoalkane substrates. The data were obtained with the parental enzymes as well as with selected mutants from the alpha class library. (A) Colored bars represent absolute specific activities. (B) Colored segments represent the percentage of specific activity of each substrate as a fraction of the sum of specific activities for a given enzyme variant.
Mutant 146 has the activities with diiodoethane and iodohexane suppressed, but the other activities at the same level as rGST A2-2. Mutant 330, on the other hand, has almost all of its activities suppressed compared with rGST A2-2. The differences observed between the individual chimeras belonging to the same group in the biplot analysis, are statistically significant as evidenced by small standard deviations of the activities with the alternative substrates (Fig. 6A). In order to exclude that the differences noted between the purified chimeras could be ascribed to low accuracy in the data, an analysis with eight dissimilar electrophilic GST substrates was performed (data not shown). This analysis clearly demonstrated that the chosen chimeras are diverging in their substrate selectivity profiles also with these diverse substrates, verifying that the difference in substrate selectivities of the chosen chimeras indeed is real.
**Discussion**

**Multivariate analysis as a tool for monitoring enzyme evolution**

Protein evolution involves the tailoring of enzymes for enhanced catalytic efficiencies, novel functions, as well as altered substrate selectivity profiles. In biological systems catalysts evolve to effect chemical reactions and to control metabolic pathways. In biotechnology, functional and physical properties of enzymes are optimized for the intended applications. In both instances the utility of the enzymatic properties should be assessed in relation to the ambient conditions, in particular, the presence of compounds that could serve as alternative substrates. An enzyme in a living cell will encounter thousands of different molecules but utilize only a limited number of them as substrates. For biosynthetic pathways high substrate selectivity is desirable, but for catabolic processes broad acceptance of substrates is more favorable. Biotechnical applications may present alternative substrates one at a time, such that discrimination among numerous compounds lacks the significance it may have in biology, even though high stereospecificity may be important. In all cases, what matters is the set of potential substrates that the enzyme is exposed to during its lifetime, be it in a cellular compartment or in a chemical reactor. The issues of enzyme promiscuity and broad substrate selectivity profiles, both important concepts in enzyme evolution, therefore have to be assessed in relation to a defined and relevant subset of all imaginable substrates. Enzymes involved in cellular detoxication, such as cytochrome P450s and GSTs, naturally are active with hundreds or thousands of chemical compounds, whereas enzymes in, e.g., the citric acid cycle have high substrate specificities with naturally occurring cell constituents. Nevertheless, in protein engineering and directed evolution substrate specificities may be relaxed to make enzymes more ambiguous in their substrate selectivities (O’Loughlin *et al.*, 2006). In order to monitor the evolution of families of genetically related enzymes, it is useful to consider the functional properties in multidimensional substrate-activity space. Both in natural and directed evolution, large numbers of mutant enzymes are tested with alternative substrates. Multivariate analysis of relevant activities can reveal relationships between substrates and enzymes that illuminate the evolution of novel functions and guide the engineering of enzymes for novel properties. For this purpose we have defined the following descriptors that apply to the analysis of variant enzymes tested with a defined set of substrates.

(i) ‘Specific catalytic capacity’ is the sum of enzymatic activities of a given enzyme measured separately with alternative substrates (Fig. 7A). This scalar value can be specified as the sum of specific activities, $k_{\text{cat}}$ or $k_{\text{cat}}/K_m$ values, or some other measure of enzymatic activity. Catalytic capacity is defined by the set of substrates analyzed and the ambient conditions of the assay. The specific catalytic capacity is useful for the comparison of enzymes and reflects the magnitude of catalytic competence in relation to all substrates considered. The specific catalytic capacity multiplied with the amount of enzyme present gives the total catalytic capacity available. An enzyme with high catalytic capacity will efficiently promote the chemical transformation of one or several substrates.

(ii) ‘Substrate selectivity’ is an account of the catalytic discrimination among the alternative substrates to which a given enzyme is exposed. This account is a vector, the components of which can be represented by the fractions of activities with the alternative substrates in comparison with the sum of all activities (Fig. 7B). Substrate selectivity is a direction in substrate-activity space indicating to what extent the enzyme will preferentially act on one or several substrates in comparison with others. Clearly, even an enzyme with low specific catalytic capacity can be valuable if it has high selectivity, since an increased amount of enzyme will enhance the rate of the catalyzed reaction without influencing the selectivity. Expressed in $k_{\text{cat}}/K_m$ values the fluxes of competing reactions can be calculated as the concentration of a particular substrate multiplied by the corresponding $k_{\text{cat}}/K_m$ value for each of the alternative substrates (Fersht, 1999). When high selectivity of an enzyme is required, the fluxes of undesired alternative reactions should be close to zero.

(iii) ‘Unit-scaled substrate selectivity’ is another vector useful for the comparison of variant enzymes. In this case, the comparison is based not only on the substrates under consideration, but also on the current ensemble of enzyme variants compared. It is generally true that novel activities emerging in evolution are minute in comparison with the already established activities. The direct comparison of substrate selectivities of different enzyme variants may therefore not recognize the emergence of novel properties, because they are overshadowed by the well-established activities. In order to overcome this inequity each substrate vector composed of the activities of all enzymes assayed with a given substrate is scaled to unit length. Figure 7C demonstrates how two substrate vectors are either stretched out or suppressed to a unit circle (or unit hypersphere in the higher dimensions afforded by additional substrates). By this scaling the components of the enzyme vectors will allow every substrate to contribute equally to the analysis regardless of the absolute activity. The scaling procedure is different from the standardization (ii) above, because it is dependent on the enzyme ensemble analyzed. In this way, a dormant enzymatic activity can be recognized and evolved into higher activity or even a diverging function compared with the parental enzymes.

**Multivariate analysis applied to the screening of a GST library with iodoalkane substrates**

In this study, a GST mutant library was screened with different iodoalkanes, including both diiodoalkanes and monoiodoalkanes (Fig. 1). Enzymes with clearly diverging substrate selectivities were recognized in the analysis of lysates of bacterial clones derived from the library (Table I and Fig. 2). Many of the GST variants gave higher values than the parental GSTs, as indicated by their relatively longer distance from the origin in Fig. 2. However, to a large degree the higher activities were due to enhanced expression in comparison with the parental enzymes (see Results). Studies of purified proteins were therefore essential, and the specific activities with five alternative iodoalkane substrates were determined (Table II).
In the group of six parental enzymes, hGST A2-2 showed the highest specific catalytic capacity, whereas the rat enzyme rGST A2-2 had the lowest specific catalytic capacity (Fig. 4A). In terms of substrate selectivity, hGST A2-2 also displayed the highest discrimination in favor of diodoethane against the other iodoalkanes, while rGST A2-2 was the least selective enzyme (Fig. 4B). Diodoethane is the intrinsically most reactive compound, and hGST A2-2 had the highest activity with this substrate (Table II). However, the high substrate selectivity of hGST A2-2 in favor of diodoethane is not due to this high activity as shown by the relative activities with alternative substrates (Fig. 4B). The other parental GSTs range in between hGST A2-2 and rGST A2-2 with respect to both catalytic capacity and substrate selectivity (Fig. 4). This ranking is also demonstrated in the biplots of the vectors of enzyme activities of both unscaled (Fig. 5) and scaled data (Fig. 6B).

Among the GST library members, mutant 371 had both the highest specific catalytic capacity and the highest selectivity for the most active substrate diodoethane (Fig. 4). However, in both respects the mutant was inferior to hGST A2-2. The second-highest catalytic capacity was shown by mutant 408, which also had higher substrate selectivity than the other mutants, with the exception of mutant 371. The lowest specific catalytic capacity was shown by mutant 330, which like mutants 146, 197 and 244 had the least restricted substrate selectivity (Fig. 4). Mutants 146, 197, 244 and 330 were the variants most similar to the parental rGST A2-2, having both low activity and low selectivity. It would appear that this variant, shown as a dendrogram (Table II), demonstrates the cluster analysis of the specific activities as a function of GST regions of the H-site are composed of amino acid residues as different magnitudes of activities with the alternative substrates. However, projection of the unit-scaled activity data in a biplot reveals that mutant 419 distinguishes itself from the other GST variants by having high activity with both 3-IPB and iodohexane. Obviously, examination of Table II demonstrates that mutant 419 has the highest specific activity of all purified enzymes with both 3-IPB and iodohexane, as well as with diiodobutane. Nevertheless, difficulties in visualizing the distinction of multi-substrate preferences among enzyme variants arise from the high dimensionality of the enzyme vectors and the different magnitudes of activities with the alternative substrates. However, projection of the unit-scaled activity data in a biplot reveals that mutant 419 distinguishes itself from the other GST variants by having high activity with both 3-IPB and iodohexane in comparison with the other substrates (Fig. 6B). In other words, the length of the substrate-activity vector of mutant 419 exceeds that of other vectors pointing in the same general direction of substrate-activity space. Thus, scaling of substrate-activity data to unit length facilitates comparisons of the mutants with the parental enzymes.

**Reaction mechanism of diiodoalkanes with GSH catalyzed by GSTs**

The fact that several of the substrates are diiodoalkanes suggests the possibility that they may form more than one product. However, the diiodoalkanes have a symmetrical structure, such that substitution of one of the iodine atoms by a glutathionyl residue would give rise to the same product irrespective of which iodine that was displaced. Substitution of both iodine atoms is a real possibility, but the second substitution would occur in a separate step following the initial monosubstitution. The structure of hGST A1-1 in complex with S-hexylglutathione, a conjugate representing the structures of the products of the iodoalkanes, has been determined (Le Trong et al., 2002), and does not indicate any option for binding of a second glutathione molecule to the active site. Therefore, a succeeding slow non-enzymatic reaction is the most likely route to a bis-glutathionyl-alkane. Diodoalkanes can also form cyclic sulfonium ions by an intramolecular electrophilic attack on the sulfur of glutathione (Guengerich, 2005). However, this secondary substitution is slow in comparison with the enzyme-catalyzed displacement of the first iodide ion. We therefore conclude that the GST activities based on the initial rates of the reactions represent monosubstitutions of iodine even with the diiodoalkane substrates. This conclusion is supported by HPLC analyses, which do not indicate more than one conjugation product in the limited time frame of the reaction (data not shown).

**Structure-activity relationships of purified mutants from the alpha class GST library**

The H-site of GSTs can be divided into three different regions, designated A, B and C (Fig. 8). The different regions of the H-site are composed of amino acid residues as follows, region A: 10, 12 and 14; region B: 104, 107, 108, 110 and 111; region C: 208, 213, 216 and 222 (Kurtovic et al., 2008). Four mutants have altered structural elements affecting the H-site, while four mutants have the same H-site residues as rGST A2-2 in the H-site. All GST variants, including the parental enzyme and the eight chimeras, have diverging substrate selectivity profiles with iodoalkanes (Figs 4, 5 and 6). Mutant 371 has the highest catalytic capacity of all mutants and a substrate specificity profile similar to hGST A2-2 (Fig. 4), but a total of five H-site residues representing all three regions differ from those in hGST A2-2 (Fig. 8A). Mutant 419 has the second highest catalytic capacity of the mutants and H-site residues identical to the rGST A2-2 residues, with the exception of an Ala12Gly replacement. However, its catalytic capacity is higher than that of rGST A2-2, which could possibly be related to the presence of Arg instead of Lys in the G-site. Mutant 408 has a lower catalytic capacity than bGST A1-1, but substrate selectivity similar to the bovine enzyme. The H-site regions A and C of mutant 408 and bGST A1-1 are identical, but region B of the mutant is identical to that of rGST A2-2. Mutant 442 is most similar in catalytic properties with rGST A2-2 among the parental enzymes. Their H-site regions B and C are identical, but residues 10 and 12 in region A are different (Fig. 8). The results suggest that the H-site residues are major determinants in the substrate selectivity profiles, and that the broad substrate specificities (low selectivities) of mutants 146, 197, 244, 330, 419 and 442, shared with rGST A2-2, are due to the extensive similarities in this binding site for the electrophilic substrate. This finding is in general agreement with the prior rational redesign of the H-site by site-directed mutations, which has created GSTs with targeted functions. hGST A1-1 was redesigned to mimic the first-sphere interactions of hGST A4-4 to afford high catalytic efficiency with alkenals (Nilsson et al., 2000), and hGST A2-2 was provided with
high steroid isomerase activity, which was not present in the wild-type enzyme (Pettersson et al., 2002).

However, close examination of mutants 146, 197, 244 and 330 shows that they do have diverging catalytic properties (Fig. 6 and Table II) in spite of the fact that all residues are identical with those of rGST A2-2 in the three H-site regions (A, B and C) (Fig. 8). Both higher and lower specific activities with, for example, iodohexane compared with rGST A2-2 are observed (Fig. 8).
A2-2 are represented with mutant 244 and 330 displaying higher and lower activities, respectively. Apparently the four chimeras are composed of primary-structure elements from different parental enzymes such that other parts of the structure differ among these GST variants (Fig. 3). All differences in primary structure among these four chimeras, as well as the differences from rGST A2-2, occur in the N-terminal portion of the sequence up to amino acid residue 82 (Fig. 3). Chimera 244 has the highest level of divergence from rGST A2-2 with 18 residues differing between the sequences, while chimera 330 has only two amino acids altered from rGST A2-2. The two other chimeras, 146 and 197 have 10 and 14 non-conserved amino acids compared with rGST A2-2, respectively. Most of the amino acids that diverge between the four chimeras and rGST A2-2 are situated in the a2-helix, the loop connecting B2-sheet and a2-helix, and the loop following a2-helix (Fig. 8). This is a structural region with relatively high mobility as indicated by the crystallographic B-factors (Sinning et al., 1993), and it is likely that the dynamics are coupled to the catalytic process. However, more incisive kinetics studies are required to clarify how the amino acid substitutions in this region modulate the catalytic activity in the GSTs with identical H-site residues.

Divergence of substrate selectivity profiles in protein evolution

Divergent properties of related enzymes have been studied through different approaches. An early application of multivariate analysis for classification of mammalian GSTs was based on divergent activities with alternative substrates as well as inhibition characteristics (Mannervik et al., 1985). In this investigation PCA afforded the classification of the GSTs into three classes, Alpha, Mu and Pi. The classification received independent support from structural data, which laid the foundation for complete description of the enzyme superfamily now established (Mannervik et al., 2005). The invention of DNA shuffling (Stemmer, 1994) and other recombinant methods has made possible the directed evolution and further understanding of the basis for the diversification of enzyme activities and the emergence of enzymes with novel properties in natural systems.

Many studies have addressed the challenge of evolving proteins with novel functions and altered substrate selectivity profiles. GSTs have been subjected to evolution of enzymes displaying diverging substrate specificities by means of site-directed mutagenesis as well as directed evolution by DNA shuffling (Hansson et al., 1999; Larsson et al., 2004; Emrén et al., 2006). Recently, P450 CYP1A subfamily members 1A1 and 1A2 were subjected to site-directed mutagenesis and DNA shuffling leading to novel variants displaying altered substrate selectivity profiles compared with the parents (Taly et al., 2007). Goddard and Reymond (2004) have addressed a problem of diverging enzyme activity fingerprints by screening a library of enzymes using a mixture of different substrates, a substrate cocktail that generated characteristic fingerprints for the enzymes. Groggnux and Reymond (2004) made functional fingerprints of lipases and esterases using an array of chiral fluorogenic esters of increasing chain lengths. Even different protease variants have been engineered to possess high catalytic activity and delicate substrate selectivity by development of a new screening strategy that employs both a selection substrate and two counterselection substrates that are quantified at the same time thereby giving the opportunity of selecting enzymes showing evolved catalytic activity with the desired substrate as well as suppressed efficiency with undesired counterparts (Varadarajan et al., 2005). A similar approach to diverging substrate selectivities has been adopted in the directed evolution of DNA cleaving enzymes, homing endonuclease I, applying both positive and negative selection pressure simultaneously (Doyon et al., 2006). Aspartate aminotransferases have also been subjected to directed evolution by applying a combination of DNA shuffling and functional selection leading to aminotransferases with novel substrate specificities (Rothman and Kirsch, 2003; Yano et al., 1998). Raillard et al. (2001) designed a library of different triazines displaying similarity in both structure and chemistry in order to screen DNA-shuffled variants of two triazine hydrolases. In this way, different substrate selectivity profiles resulting in novel functions in substrate-activity space were explored. A common feature of all these studies is the evolution of novel substrate selectivity profiles in families of enzymes catalyzing the same type of chemical reaction. In the present study the multivariate analysis has been refined to show how closely related enzymes differentially act on similar substrates, and illustrate the emergence of different catalytic capacities and substrate selectivities by DNA recombination.

Concluding remarks

Evolutionary processes involve the optimization of functional properties to match the current requirements. In the present investigation we have focused on catalytic properties including divergent activities with alternative substrates and substrate specificity profiles. However, in biological systems and for biotechnical applications properties other than kinetic parameters will also be important. Thermal stability, solubility, and expressivity are factors that influence the fitness of an enzyme for its tasks. In directed evolution experiments it is possible to guide the evolutionary process by monitoring populations of mutant enzymes, and select variants that best approximate the targeted properties. Like in natural evolution, protein engineering for altered functions proceeds stepwise by mutations of extant structures. The choice of parents for the next generation may be crucial for obtaining optimized functions and structural properties. For this purpose multivariate analysis facilitates rational design of directed evolution in the laboratory, and provides a more efficient trajectory to fitness than the random trial-and-error that is considered to drive the evolution of proteins in nature.

Supplementary data

Supplementary data (color images for figures 4, 6, 7, and 8) are available at PEDS online.

Funding

This work was supported by the Swedish Research Council and the Swedish Cancer Society.

Acknowledgements

The authors would like to thank Françoise Raffalli-Mathieu for the kind gift of purified bovine GST A1-1.