Chemistry & Biology
Article

Structure-Based Redesign of GST A2-2 for Enhanced Catalytic Efficiency with Azathioprine

Wei Zhang,[1] Olof Modén,[1] Kaspars Tars,[2] and Bengt Mannervik[1,3,*]
[1]Department of Biochemistry and Organic Chemistry, Uppsala University, BMC, Box 576, SE-75123 Uppsala, Sweden
[2]Biomedical Research and Study Center, LV-1067 Riga, Latvia
[3]Department of Neurochemistry, Stockholm University, SE-10691 Stockholm, Sweden
*Correspondence: bengt.mannervik@biorg.uu.se
DOI 10.1016/j.chembiol.2012.01.021

SUMMARY

Glutathione transferase (GST) A2-2 is the most efficient human enzyme in the biotransformation of the prodrug azathioprine (Aza). The activation of Aza has therapeutic potential for possible use of GSTs in targeted enzyme-prodrug treatment of diseases. Based on the assumed catalytic mechanism and computational docking of Aza to the active site of the enzyme, active-site residues were selected for construction of focused mutant libraries, which were thereafter screened for Aza activity. Mutants with elevated Aza activity were identified, DNA sequenced, and the proteins purified. The two most active mutants showed up to 70-fold higher catalytic efficiency than the parental GST A2-2. The structure of the most active triple mutant (L107G/L108D/F222H) enzyme was determined by X-ray crystallography demonstrating significant changes in the topography of the active site facilitating productive binding of Aza as a substrate.

INTRODUCTION

Enzymes are nature's catalysts evolved for biochemical reactions. Even though natural enzymes are adapted to their particular function in vivo, their use in industrial or pharmaceutical applications often requires a redesign for proper performance. Moreover, owing to the complexity of enzyme catalysis (Hammes-Schiffer and Benkovic, 2006), enzyme engineering has been considered as one of the most challenging problems of contemporary biochemistry. Among all methodologies for enzyme engineering, directed evolution is probably the most successful and most frequently used approach. By subjecting members of a mutant library to the power of Darwinian evolution, enhanced enzyme performance can be achieved by conventional directed evolution strategies even in the absence of structural information. Despite extensive application of directed enzyme evolution for tailoring biocatalysts for organic synthesis or biosynthesis of high-value pharmaceutical intermediates (Zhang et al., 2010a; Schmidt-Dannert et al., 2000; Savile et al., 2010; Ran et al., 2004; Bartsch et al., 2008), production of alternative biofuels (Fasan et al., 2011; Atsumi and Liao, 2008), development of laundry detergent additives (Cherry et al., 1999), and cancer therapy (Chen et al., 2008; Liu et al., 2009), several fundamental questions with regard to directed evolution still need to be addressed.

Directed evolution comprises iterative cycles of randomization and screening. Theoretically, generation of molecular diversity by introducing point mutations is not a technical constraint for directed evolution. Various attempts have been tried to generate libraries with mutation rates exceeding 30 mutations per protein using error-prone PCR (Zaccolo and Gherardi, 1999). The resulting immense number of library members therefore has increased the demand for high-throughput enzyme assays. Features of all available high-throughput screening or selection methods have been reviewed in detail elsewhere (Lin and Cornish, 2002; Aharoni and Tawfik, 2009). However, the success of these approaches is partly dependent on convenient selection or screening methods based on binding affinity or fluorescence detection. There is still no general solution to the assay of catalytic turnover numbers via high-throughput assays. An assay directly based on monitoring product formation is still the typical scenario, representing the most significant bottleneck in directed evolution (Dietrich et al., 2010).

Smarter strategies for construction of small mutant library with high quality, therefore, are receiving increasing attention (Shivange et al., 2009; Reetz et al., 2009). Inspired by de novo computational enzyme design (Jiang et al., 2008), reaction mechanisms can be coupled with directed evolution so that small focused mutant libraries rationalized by reaction mechanisms can be generated to reduce the screening effort.

In the present work, we tested the utility of such a strategy for enhancing the efficiency of a glutathione transferase (GST) in the bioactivation of azathioprine (Aza). Activation of Aza has medical interest in possible applications of targeted enzyme-prodrug therapy of diseases. Structural modeling of the enzyme-substrate complex was obtained by molecular docking of Aza to the protein. After outlining the possible catalytic mechanism of the reaction, enzyme-substrate interactions assumed to stabilize the transition state were considered. On the basis of this analysis three residues were selected for simultaneous mutagenesis. A corresponding triple mutant library, randomized with reduced amino acid alphabets (two amino acid residues randomized with NDT [D: A/G/T] degenerate oligonucleotides and one other residue randomized with YHC degeneracy [Y: C/T; H: A/C/T]), with a theoretical size of 864 members (12*12*6) was generated and screened for Aza activity. One of the evolved GST variants displayed a 70-fold increase in

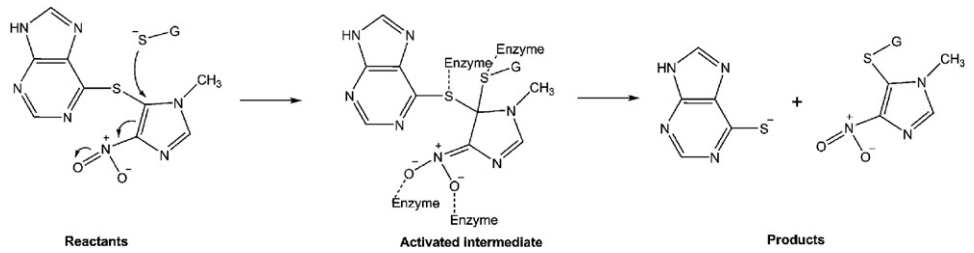

Figure 1. Reaction of Azathioprine with Glutathione Catalyzed by GSTs

The toxic antimetabolite 6-mercaptopurine is released via a tetrahedral intermediate formed by nucleophilic attack of the glutathione thiolate (GS^-). The reaction is presumed to be promoted by H-bonding or other polar interactions indicated by dashed lines to the intermediate.
See also Figure S1A.

catalytic efficiency relative to the starting-point GST. Other active-site residues of the parental enzyme in close proximity to the substrate were further randomized to assess their importance for activity with the targeted substrate. However, not a single mutant with enhanced Aza activity was obtained. Finally, the crystal structure of the new GST mutant was determined, demonstrating an altered topology of the active site that facilitates productive binding of the substrate.

RESULTS

Design of a Focused GST Library

The H-site of human GST A2-2 is composed of 13 amino acid residues. We hypothesized that simultaneous introduction of multiple mutations in the active site might introduce novel catalytic components specifically for the desired substrate. When we examined the composition of the H-site, proline-110 seemed to be broadly conserved among Alpha class GST isoenzymes. One naturally occurring allelic GST variant (GST A2-2*E) containing a mutation at this position happened to be 3.7-fold more active with Aza than the allelic variants presenting proline at position 110 (Zhang et al., 2010b). Therefore, we used GST A2-2*E (A2*E) as a parent for further evolution.

The GST-mediated conjugation of GSH with azathioprine is an aromatic substitution (S_NAr) involving nucleophilic attack by the sulfur of GSH on the electrophilic center at carbon 5 of the imidazole ring (Figure 1). We considered two possible mechanisms to enhance the reaction rate: (1) increasing the partial positive charge of carbon 5 by increasing the resonance effect of the adjacent nitro group, or (2) stabilizing the reaction intermediates by interaction with the sulfur atom of Aza or the thiolate group of glutathione (Figure 1). To achieve such effects, novel interactions between side chains of different active-site residues and the nitro group or sulfur atoms were aimed for by introducing novel residues in the active site.

Computer modeling of the binding of Aza to the H-site of human GST A2-2 was performed, and all docking poses considered reasonable for catalysis were subjected to further analysis (Figure S1A available online). Candidate residues were identified on the basis of proximity to the nitro group of the imidazole moiety of Aza, the sulfur linkage between the two aromatic rings, or the sulfur of glutathione. For various docking poses, Leu-107, Leu-108, Phe-111, and Phe-222 of the active-site residues were in proximity to the above-mentioned atoms within a distance of 5 Å. Three focused libraries were constructed with reduced amino acid alphabets: one targeting three positions 107, 108, and 222, encoding Arg, Asn, Asp, Cys, Gly, His, Ile, Leu, Phe, Ser, Tyr, Val (with NDT degeneracy) at each of positions 107 and 108, and His, Leu, Phe, Pro, Ser, Tyr (with YHC codon randomization) at position 222. The second and third libraries explored the impact of positions 111 and 213 by separate libraries based on NDT codon randomization.

Construction of a Triple-Point Mutant Library and Preliminary Screening for Aza Activity

A dual-tube megaprimer approach was used for focused mutagenesis (Tseng et al., 2008). By randomly sequencing more than 40 clones derived from the constructed library, not a single clone corresponding to the wild-type gene was found, indicating that adequate variability of the mutations had been achieved. Thereafter, the constructed library was expressed in *Escherichia coli* and subjected to screening for Aza activity.

Lysates of 1901 randomly picked bacterial colonies, and 19 colonies of the wild-type gene as positive controls, were prepared and assayed with Aza. The experimental coverage of the mutant library was estimated by Monte Carlo simulation to be 89% ± 1% (768 ± 8 unique variants of the 864 possible variants). The activity of the starting GST with Aza was 16.0 ± 4.0 ($10^{-3}\Delta A_{320}$/min). The frequency distribution of activity was skewed to the right with the highest frequency occurring around the wild-type value, indicating that a randomly chosen mutant would be more prone to have increased activity with Aza (Figure 2). 44.2% (840 out of 1901) transformants exhibited enhanced activity (the cutoff value was set at mean + 3 SD) with Aza compared to the starting enzyme. Overrepresentation of active variants indicated that a high coverage of the relevant protein sequence space was achieved in the screening process.

Thirteen variants with activities >200 ($10^{-3}\Delta A_{320}$/min) were sequenced to explore the mutations affording enhanced activity with azathioprine. Sequence information on these mutants can be found in Table S2. All of the 13 mutants contained targeted residue alterations and no spurious mutations. However, one of the isolates (TP10_12E) turned out to be a mixture of two distinct GST sequences containing the triple mutations L107G/L108D/F222P (mutant GDP) and L107G/L108D/F222H (mutant GDH), both of which were represented separately among the most active mutants identified (Table S2). Substitutions occurring in the sequences of the most active variants, largely led to charged or hydrophilic side chains.

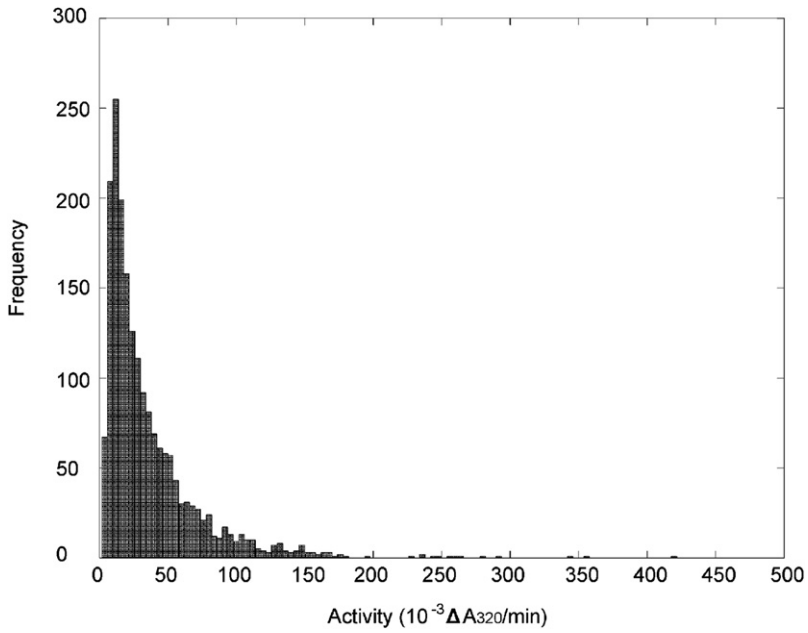

Figure 2. Frequency Distribution of Aza activity in Bacterial Lysates of the GST Mutant Library Screened

Activities are given in absorbance units as measured in 300 μl reaction volumes in microplates. The histogram was constructed with a bin interval of 4 ($10^{-3}\Delta A_{320}$/min activity units); the activity of the parental GST A2-2*E was 16.0 ± 4.0 units.
See also Table S2.

Additional Libraries of Single-Point Mutations

Two additional mutant libraries targeting active-site residues 111 and 213, one at a time, were constructed with NDT degeneracy of the codons. Neither mutations of one nor the other in the A2*E wild-type background gave rise to any GST mutants with enhanced Aza activity in bacterial lysates.

Kinetic Characterization of Selected GST Variants

Nine variants of the above-mentioned sequenced GST proteins, as well as three additional variants chosen based on end-point assays, were expressed and purified. Enzyme reaction rates with Aza in bacterial lysates were from 178 to 336 ($10^{-3}\Delta A_{320}$/min) for all variants examined. All purified protein variants exhibited enhanced specific activities by a factor from 7 to 56 compared to the parental enzyme A2*E. However, the ratio between lysate activity to specific activity of purified GST mutants varied up to 10-fold among the 12 variants investigated, even though 9 out of 12 variants exhibited a similar ratio as that of the parental GST. Apparently, expression levels of the enzyme protein varied among the mutants (Figure 3). Although the observed increases in lysate activities stemmed from differences both in intrinsic enzyme activities and expression levels, the main factor was the enhanced specific activities. The L107C/L108S/F222H mutant (CSH, designating the new residues of the mutated positions in one-letter amino acid code) and the L107C/L108G/F222H mutant (CGH) were two GST variants with enhanced expression levels, while their specific activities were only moderately increased. A third mutant GDH displayed a ~50% decrease in expression compared to that of the parental enzyme. The two most active variants (GDH and GDP) differed only in one amino acid residue, both having Gly and Asp at positions 107 and 108, respectively. The purified GDH exhibited a specific activity of 128 μmol min^{-1} mg^{-1} with Aza, approaching the upper limit of known activities of the GST scaffold (Table 1).

A2*E as well as the two most active variants GDH and GDP were then subjected to steady-state kinetic studies, in which the mutants showed 70- and 40-fold enhanced catalytic efficiencies with Aza, respectively (Table 2). The gains in catalytic efficiency stemmed from both increased k_{cat} and decreased K_M, as an indication of enhanced substrate recognition.

Thermal Inactivation and pH-Rate Profiles of the Selected GST Variants

Thermal inactivation and effect of pH on enzyme activity of the parental A2*E, GDP and GDH were also investigated (Figure 4A). Thermal inactivation of A2*E and GDP displayed half-lives estimated at 21 and 12 days at 37°C, respectively. In contrast, thermal denaturation of GDH displayed two phases, with a slow phase resulting in a loss of ~50% of activity after incubation for 8 days. These results evidenced a negative tradeoff between the enhanced activity and thermostability in the evolved enzymes, which is often observed in laboratory enzyme evolution (Romero and Arnold, 2009).

The pH-rate profiles of all GST variants investigated largely coincided, with maximum activity exhibited around pH 7 (Figure 4B). In particular, the similarity between the profiles of A2*E and GDH at low pH values suggested that ionization of His-222 is not a contributing factor to the elevated activity of GDH.

Structural Studies of the Evolved GST Mutant

The GDH mutant was crystallized with the imidazole conjugate of glutathione (S-1-methyl-4-nitro-5-imidazolyl-glutathione) resulting from the reaction with Aza followed by X-ray analysis (Table S3). The crystal unit cell contained two dimeric GST structures showing expected similarities with the GST A2-2 enzyme previously studied (Tars et al., 2010). However, the human enzyme occurs in five allelic variants (Zhang et al., 2010b) and the previous structure represents the C allele (GST A2*C), which is 3.7-fold less active with Aza than the A2*E variant chosen for mutagenesis in the present study. Comparison of the two structures shows that Pro-110 in A2*C causes the polypeptide chain to restrict the active-site cavity to a smaller volume than that of the GDH mutant derived from A2*E which features Ser-110 in the same position (Figure 5). Apparently, this structural difference underlies the divergence between the allelic variants in the activities with Aza. There is also a difference in position 112 (Thr and Ser in allelic variants C and E, respectively), but this conservative replacement of a side chain pointing away from the active site is not likely to influence the activity.

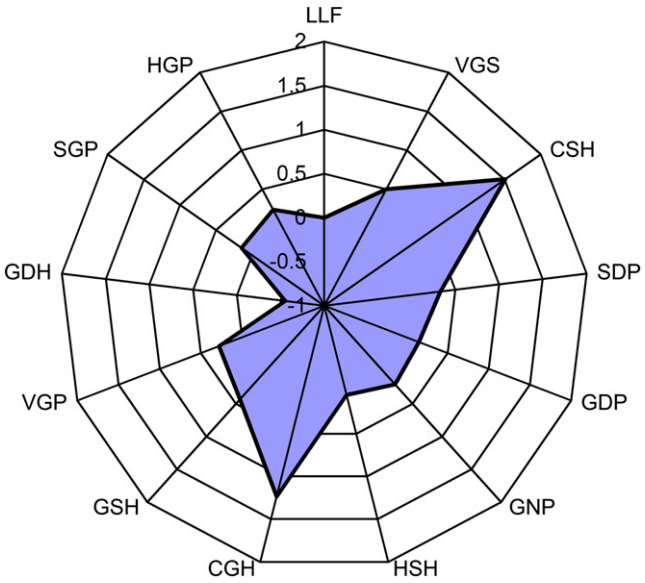

Figure 3. Radar Chart Illustrating Expression Levels of GST Mutants in *E. coli* Relative to That of the Parental GST A2-2*E

The three letter designations identify the amino acids (in one-letter code) in positions 107, 108, and 222 of the primary structure. The length of a spoke is proportional to the natural logarithm of the expression relative to that of A2*E (LLF). The values were calculated from the ratio between lysate activity to specific activity for each variant.

In spite of the fact that the crystals of GDH were obtained with S-1-methyl-4-nitro-5-imidazolyl-glutathione, only the glutathione moiety was clearly visible in the electron density. Its position and conformation were highly similar to that of glutathione in the A2*C enzyme, in basic agreement with the coordinates used for the modeling prior to mutagenesis (Figure S1A). It is likely that the imidazolyl portion of the glutathione conjugate is not bound in a fixed position, and is therefore not observable in the active site.

His-222 in GDH is located in essentially the same position as Phe-222 in wild-type GST A2-2, but in both proteins residue 222 is more mobile than other active-site residues, as judged by less well-defined electron density. His-222 is too far away to directly influence the reactivity of the glutathione sulfur atom, consistent with the unchanged pH-activity profile of GDH in comparison with the wild-type enzyme (Figure 4B).

Residues 107 and 108 in GDH demonstrate the most conspicuous differences from the wild-type GST, which features Leu-107 and Leu-108. In the wild-type enzyme the side chains of both residues point toward bound hydrophobic substrates. Replacement of Leu-107 by Gly removes the bulky hydrophobic side chain and provides additional space in the active-site cavity. An even more dramatic alteration of the active site occurs in position 108 where the polar Asp-108 is introduced. The side chain of Asp-108 in the GDH structure is directed toward Lys-120, apparently making an ionic bond with its side-chain ammonium group. In this position the carboxyl group of Asp-108 is far removed from the sulfur of glutathione, but creates additional room for substrate binding.

Table 1. Molecular Diversity and Specific Activities of the 12 Selected GST Variants

Bacterial Clone	Residues at the Mutagenized Positions 107 108 222	Specific Activity (μmol mg^{-1} min^{-1})
A2*E	L L F	2.3 ± 0.1
TP01_1F	V G S	59 ± 3
TP02_1E	C S H	16 ± 1
TP03_10D	S D P	49 ± 1
TP04_3H	G D P	86 ± 1
TP04_9H	G N P	48 ± 2
TP08_6A	H S H	49 ± 2
TP09_3G	C G H	19 ± 1
TP10_6B	G S H	34 ± 1
TP14_4F	V G P	50 ± 1
TP15_6D	G D H	128 ± 4
TP15_12G	S G P	64 ± 5
TP16_1E	H G P	65 ± 5

For oligonucleotides used in generating the library (see also Table S1).

DISCUSSION

One of the most intriguing problems in biochemistry is enzyme engineering, not only because of numerous technical challenges for rational design and directed evolution, but also due to extensive applications in medicine, synthetic chemistry, pharmaceutical development and agricultural sciences. In spite of recent achievements in computational enzyme (re)design (Jiang et al., 2008; Murphy et al., 2009; Röthlisberger et al., 2008; Siegel et al., 2010), the intricacy of predicting allosteric effects and taking structure dynamics and substrate/product affinities into consideration imposes limitations to the performance of computational (re)design with regard to rate enhancement as compared to that of directed evolution (Golynskiy and Seelig, 2010). In contrast, by randomizing a rigid secondary element in the interface between the FAD-NADP domains of a thermostable Baeyer-Villiger monooxygenase (PAMO), Reetz and co-workers successfully introduced allosterically induced domain movements to expand the substrate specificity of the enzyme by a directed evolution approach (Wu et al., 2010). Although large libraries allow extensive exploration of sequence space, which may facilitate identification of desired enzyme variants, screening efforts currently constitute the major bottleneck for directed enzyme evolution. Despite recent progress including in vitro display techniques, in vivo selection systems, in vitro

Table 2. Catalytic Efficiency of Selected Variants with Aza

Aza	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (mM^{-1} s^{-1})
A2*E	12.7 ± 1.6	3.1 ± 0.5	4.1
GDH	145 ± 9	0.50 ± 0.06	292
GDP	86 ± 4	0.48 ± 0.04	180

Data are represented as mean ± standard deviation and are based on triplicate measurements.
The k_{cat} and k_{cat}/K_M values are expressed per subunit of the dimeric enzyme.

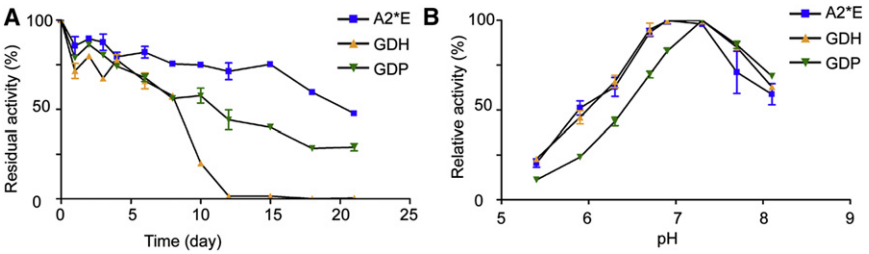

Figure 4. Thermal Inactivation and pH-Rate Profiles of Wild-Type GST A2-2*E and Mutants GDH and GDP

(A) Thermal inactivation test progress of GST variants. Enzyme at a concentration of 0.1 mg/ml was incubated in 100 mM sodium phosphate (pH 7.4) in the presence of 1 mM GSH at 37°C. Aliquots of the incubation mixture were taken at different time points and assayed for residual specific activity. The error bars represent the standard deviation based on three replicates.

(B) pH activity profiles indicating neutral pH optima of GST variants. Enzymes were assayed in the standard system for specific activity, with the exception that the pH of the 100 mM sodium buffer was varied. The error bars represent the standard deviation based on three replicates.

vacuolar compartmentation, it is still true that catalytic turnover-based screening methods directly monitoring product formation are typically more versatile. In order to reduce the screening effort of simultaneous multisite saturation mutagenesis, the group of Bornscheuer proposed randomization of targeted residues with the most frequently naturally occurring amino acids in the quest for thermostabilized *Pseudomonas fluorescens* esterase (Jochens et al., 2010). Generation of such small high-quality libraries for ease of screening is therefore starting to gain increased attention as a means to exploring sequence space more efficiently (Alcolombri et al., 2011; Jäckel et al., 2008).

In this work, we integrated molecular docking and reduced amino acid alphabets for enzyme engineering. By examining all possible substrate orientations (Figure S1A), we identified three hot spots (positions 107, 108, and 222) for randomization on the basis of the underlying catalytic mechanism (Figure 1). Simultaneous randomization of the three positions for a complete set of amino acids using NNK codon degeneracy would result in a mutant library of 32^3 members, which would be very time consuming with regard to the screening effort in the absence of high-throughput methods (even disregarding the need for oversampling to achieve a necessary coverage of protein sequence space). The merits of using NDT randomization in combinatorial protein engineering have been demonstrated by Reetz et al. (2008, 2009). Consequently, residues 107 and 108 were randomized with NDT degeneracy. A previous study indicated an important role of the bulky Phe-222 for catalysis (Nilsson et al., 2002), and residue 222 was therefore randomized with YHC degeneracy, mainly leading to large side chains at this position. In addition, the chosen codon degeneracies limited the number of possible combinations to 864, thereby facilitating the screening. By such randomization of three residues of the 13 active-site residues in the parental GST, followed by screening, we successfully enhanced GST activity with Aza by 70-fold.

The rate-limiting step and the transition state of the GST-catalyzed reaction with Aza are unknown, even though the tetrahedral intermediate shown in Figure 1 should exist on the reaction trajectory. Elucidation of the detailed mechanism would require further investigations. The most obvious consequence of the mutations is an enlarged H-site, which facilitates accommodation of the purine ring of Aza. The crystal structure of mutant GDH does not support our postulate that the mutated residues directly polarize functional groups of the imidazole moiety of Aza and thereby enhance catalytic efficiency. The side chains of the four mutated amino acids are too far removed from the reaction center for direct interactions. Molecular docking of Aza to mutant GDH indicates possible interactions between Arg-15 and Glu-104 with Aza, but the analysis does not imply any direct interaction between the side chains of the four mutated amino acids in GDH and Aza (Figure S1B). However, it should be noted that a rigid model of protein mutant GDH was used in the docking analysis, and the dynamic behavior of the protein is thus not accounted for. It can be speculated that the negatively charged Glu-104, directly or indirectly, promotes Aza activation. The notion of a negative charge enhancing catalytic efficiency is also indicated by site-directed mutagenesis of Leu-108 in A2*E to Asp (data not shown). Furthermore, the back-mutation of Asp-108 in GDH to Leu (data not shown) has confirmed the activating effect of the negatively charged side chain. Possibly, the enhanced function of the active site is primarily based on electrostatic stabilization of the transition complex, as argued in general for enzyme catalysis by Warshel et al. (2006), and not by provision of novel catalytic groups. A 70-fold enhanced catalytic efficiency corresponds to a $\Delta\Delta G^{\ddagger}$ of 2.5 kcal mol^{-1} transition state stabilization. We are currently working on a model for the activated reaction intermediate and the transition state.

In retrospect, it would appear that our strategy of mutating residues in the required proximity to Aza involved the most reasonable choices of a total of seven active-site amino acids: Arg-15, Leu-107, Leu-108, Phe-111, Leu-213, Phe-220, and Phe-222. Arg-15 is functionally significant and conserved in alpha class GSTs (Björnestedt et al., 1995) and Phe-220 is part of the glutathione binding site (Bruns et al., 1999) and they were therefore excluded from mutagenesis. Phe-111 and Leu-213 in A2*E were separately randomized with NDT and NNK codon degeneracy, respectively, but only variants with diminished Aza activity were found after screening of >600 clones.

Early dead ends of directed enzyme evolution can occur not only because of evolutionary traps caused by loss of catalytic efficiency, but also may be due to accumulation of destabilizing mutations in the evolutionary paths. Enzymes are only marginally stable with ΔG_{fold} of about −5 to −15 kcal mol^{-1} (Williams et al., 2006; Fersht, 1999). Moreover, the flux of destabilizing mutations is considered > 20-fold larger than that of stabilizing mutations (Tokuriki and Tawfik, 2009). In other words, from the perspective of thermodynamic stability only a handful of mutations should be introduced simultaneously in order for the majority of the library members to be properly folded and structured. Thermodynamic effects of new-function mutations seem to be

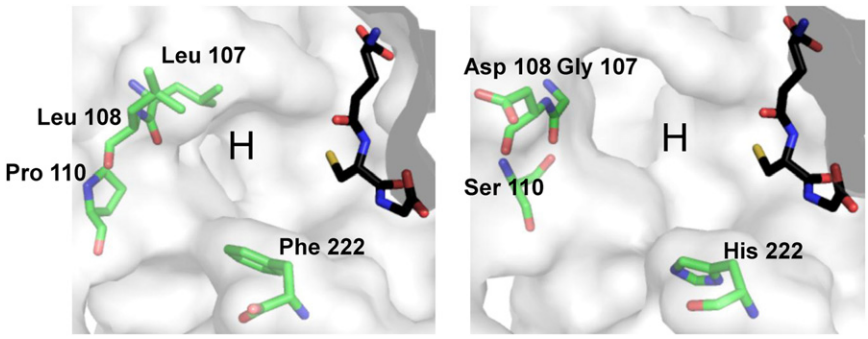

Figure 5. Comparison of the Active Sites of Allelic Variant C of Human GST A2-2 and Mutant GDH of Allelic Variant E

Glutathione (rendered as a stick model with black carbons) points its nucleophilic sulfur toward the H-site (marked "H"), which binds the electrophilic second substrate. Four residues (107, 108, 110, and 222) differing between the two structures (stick models with green carbons) define the H-site cavity, in which Aza is supposed to bind with the imidazole leaving group close to the attacking sulfur of glutathione. The large purine moiety of Aza has more space in the H-site of the GDH mutant since the bulky Leu side chains have been removed. Figure was made in Pymol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC). Active sites of allelic variant C of human GST A2-2 (left, PDB code 2WJU) and mutant GDH of allelic variant E (right, present study, PDB code 4ACS).
See also Figure S1B and Table S3.

scaffold-dependent (Tokuriki et al., 2008). In the case of human GST, the average value of $\Delta\Delta G_{fold}$ for each new-function mutation is 2.08 kcal mol^{-1} (Tokuriki et al., 2008). Accordingly, no more than five mutations would be tolerated for GST reengineering, in the absence of compensatory stabilizing mutations. Identifying the active-site residues crucial for catalysis is therefore a key to successful evolutionary design.

SIGNIFICANCE

The current work has integrated the assumed reaction mechanism and molecular docking into a general semirational enzyme engineering approach. As a result, high frequency of mutant variants with enhanced activity was achieved. By simultaneously introducing three mutations, an enzyme variant was discovered exhibiting a specific activity and k_{cat} value approaching the upper limits of these parameters known for the GST scaffold. Moreover, the approach was experimentally proven to be more successful for the improvement of enzyme performance than intuitively choosing active-site residues such as residues 111 and 213 in proximity to the substrate. Considering the importance of pinpointing crucial residues for enhancement of enzyme performance, we believe that such a semirational approach is a valuable strategy for enzyme engineering. In our application, a GST with the targeted enhanced efficiency in the activation of the prodrug Aza was accomplished.

EXPERIMENTAL PROCEDURES

Materials

Pfu DNA polymerase and DpnI restriction nuclease were purchased from Stratagene and Fermentas, respectively. QIAquick gel extraction kit was from QIAGEN. Oligonucleotides used were ordered from Thermo Fisher Scientific. *E. coli* XL1-Blue obtained from Stratagene was used for protein expression. All chemical substrates were purchased from Sigma-Aldrich.

Automated Docking

Azathioprine input file was generated at DockingServer, with ligand partial charged and geometry refined by semiempirical methods (PM6) (Bikadi and Hazai, 2009). A structural model for GST A2-2*E (A2*E) was created from the structure of GST A2-2*C (PDB code: 2WJU; Tars et al., 2010) using the Swiss-Pdb Viewer software (Guex and Peitsch, 1997). Hydrogen atoms were added to the holo structure of A2*E using Auto Dock Tools 1.5.1 (Morris et al., 1998). A grid box of 50 × 50 × 50 points spaced at 0.375 Å, centered on the center of H-site covering the hydrophobic ligand-binding site, was constructed for the docking of azathioprine. Using AutoDock 4.0, a combined global and local search algorithm through a Lamarckian genetic algorithm (LGA) was used to determine the docked conformations. Two hundred runs were performed for genetic algorithms, with a population size of 250, 540,000 as maximum number of generations, and other parameters with default values. The default parameters were used for local search. To analyze the docked conformations, the cluster tolerance of RMSD was 0.5 Å, so only conformations within this restriction would be subgrouped into the same cluster.

Construction of a Multiple Site-Directed Mutagenesis Library

Plasmid pGΔETac_NHis_hA2*E was used as the template to construct a focused mutagenesis library. Detailed information about primers can be found in Supplemental Information (Table S1).

A dual-tube approach was used for the construction of the mutant library (Tseng et al., 2008). The first-stage amplification reaction was performed in a total volume of 100 μl, containing 40 ng template plasmids, 10 pmol of oligonucleotides, 10 nmol of dNTP and 2.5 U of *Pfu* DNA polymerase. After initial denaturation at 95°C for 5 min, the first-stage amplification was carried out by 24 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 60 s, and final extension at 72°C for 10 min. For further extension from PCR product as megaprimers with desired mutations to whole plasmid, the reaction was performed in a total volume of 50 μl, containing 40 ng of template plasmids, 10 μl of purified PCR product from the first-stage amplification as primers, 10 nmol of dNTP and 1.25 U of *Pfu* DNA polymerases. After denaturation of dsDNA at 95°C for 5 min, 30 cycles of denaturation at 95°C for 30 s, annealing at 68°C for 30 s, and extension at 72°C for 4 min were carried out, followed by final extension at 72°C for 10 min. After DpnI digestion of the template plasmids at 37°C for 8 hr, 2 μl of DpnI-treated DNA was electroporated into *E. coli* XL1-Blue, and transformants were selected on ampicillin-containing (100 μg/ml) LB plates.

Preparation of Bacterial Lysates

A total of 1,920 single colonies (including 19 parental colonies as positive controls) were randomly picked and incubated in 96-well microtiter plates with 250 μl LB medium supplemented with 100 μg/ml ampicillin. After incubation at 37°C for 16 hr, and centrifugation at 500 rpm for 3 min to avoid potential contamination, colonies from the bacterial cultures were diluted 100-fold into 400 μl/well plates containing 300 μl 2TY supplemented with 100 μg/ml of ampicillin. Bacteria were grown for protein expression at 37°C for 17 hr in the absence of IPTG. For complete lysis, cultures were pelleted, resuspended in 150 μl 100 mM sodium phosphate buffer (NaPi [pH 7.0], with 0.2 mg/ml lysozyme), then incubated on ice for 1 hr, followed by three cycles of freezing/thawing (−80°C /37°C). The lysates then were centrifuged and the supernatants were used for activity measurement in microtiter plates. The experimental coverage of the library was estimated by Monte Carlo simulation using 10,000 iterations using a commercial software package (MATLAB 7.5, The MathWorks Inc., Natick, MA, 2007) and presented as mean ± standard deviation.

Screening Assay with Azathioprine

For the screening of GST activity with azathioprine, 5 μl centrifuged bacterial lysate, 1 mM GSH (final concentration), 0.2 mM azathioprine (final concentration) and NaPi buffer (100 mM [pH 7.0]) were added to a total volume of 250 μl in microtiter plate wells. The change in absorbance at 320 nm was directly proportional to GST activity with azathioprine, and the activity of 5 μl lysate was expressed in the unit of $10^{-3}\Delta A_{320}$/min.

Protein Expression and Purification

For cytosolic overexpression of GST variants in *E. coli*, overnight cultures were transferred into 500 ml 2TY medium with 100 μg/ml ampicillin at a 1:100 dilution, and grown to an OD600 of 0.3 before induction for 14 hr with 0.2 mM IPTG. Cells were pelleted and lysed by lysozyme at a final concentration of 0.2 mg/ml for 60 min on ice, then subjected for sonication for complete lysis. After centrifugation at 27,200 × *g* for 60 min at 4°C, supernatants were loaded on His GraviTrap columns for purification. The purified enzymes were dialyzed overnight against Tris-EDTA-mercaptoethanol (TEM) buffer (50 mM Tris, 0.5 mM EDTA, 50 mM NaCl, 1 mM 2-mercaptoethanol, 5% glycerol [pH 7.4]). Protein purity was then verified by SDS-PAGE. The purified protein concentration was determined through Bradford assay with the protein assay reagents from Bio-Rad, with bovine serum albumin as a protein standard.

Determination of Specific Activity and Kinetic Parameters

Detailed description of the spectrophotometric enzymatic assays has been published previously (Zhang et al., 2010b; Eklund et al., 2006). Kinetic parameters were determined by fitting the Michaelis-Menten equation to initial velocity data using GraphPad Prism 5. The k_{cat} values are expressed per subunit of the dimeric enzyme.

Crystallization and Structure Determination of the GDH Mutant

The mutant GDH was crystallized by the hanging-drop vapor-diffusion method. The mutant protein was purified as described above. After buffer replacement to 10 mM Tris-HCl (pH 7.8), the protein was concentrated by ultrafiltration. The final solution for crystallization contained protein at 10 mg/ml and 5 mM S-1-methyl-4-nitro-5-imidazolyl-glutathione in 10 mM Tris-HCl (pH 7.8). The glutathione S-conjugate was synthesized from 5-chloro-1-methyl-4-nitroimidazole and glutathione essentially as described for S-hexyl-glutathione (Vince et al., 1971; Aronsson and Mannervik, 1977). Crystals were grown by mixing equal volumes (4 μl) of protein – glutathione S-conjugate mixture and reservoir solution containing 9%–14% PEG 4000 and 2 mM DTT in 100 mM Tris-HCl (pH 7.8) with additional octyl D-β- glucopyranoside to a final concentration of 0.1% (w/v). Diffraction data were collected from beamline ID14-2 at the ESRF synchrotron (Grenoble, France). Data were indexed with MOSFLM (Leslie, 1992) and scaled using SCALA (Evans, 1997). The structure was solved by molecular replacement in MOLREP (Vagin and Teplyakov, 1997) using native GST A2-2 coordinates (PDB code 2WJU; Tars et al., 2010) as a starting model. The space group was $P2_1$ with unit cell dimensions a = 48.6 Å, b = 94.8 Å, c = 113.7 Å and β = 92.8°. There were two GST dimers in the unit cell. Mutated residues were built in COOT (Emsley et al., 2010) and structure refined with REFMAC (Murshudov et al., 1997). Processing, refinement and validation statistics are shown in Table S3. Coordinates and structure factors have been deposited in the PDB databank with accession code 4ACS.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and three tables and can be found with this article online at doi:10.1016/j.chembiol.2012.01.021.

ACKNOWLEDGMENTS

We thank Birgit Olin for expert assistance including the synthesis of S-1-methyl-4-nitro-5-imidazolyl-glutathione and crystallization of the enzyme, and Magnus Johansson for collecting diffraction data. The work was supported by grants from the Swedish Cancer Society, the Swedish Research Council and European Regional Development Foundation grant 2DP/2.1.1.1.0/10/APIA/VIAA/052.

Received: September 12, 2011
Revised: January 8, 2012
Accepted: January 10, 2012
Published: March 22, 2012

REFERENCES

Aharoni, A., and Tawfik, D.S. (2009). In vitro compartmentalization (IVC) and other high-throughput screens of enzyme libraries. In Protein Engineering Handbook, S. Lutz and U.T. Bornscheuer, eds. (Weinheim, Germany: Wiley-VCH Verlag GmbH), pp. 649–668.

Alcolombri, U., Elias, M., and Tawfik, D.S. (2011). Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. J. Mol. Biol. *411*, 837–853.

Aronsson, A.C., and Mannervik, B. (1977). Characterization of glyoxalase I purified from pig erythrocytes by affinity chromatography. Biochem. J. *165*, 503–509.

Atsumi, S., and Liao, J.C. (2008). Directed evolution of Methanococcus jannaschii citramalate synthase for biosynthesis of 1-propanol and 1-butanol by Escherichia coli. Appl. Environ. Microbiol. *74*, 7802–7808.

Bartsch, S., Kourist, R., and Bornscheuer, U.T. (2008). Complete inversion of enantioselectivity towards acetylated tertiary alcohols by a double mutant of a Bacillus subtilis esterase. Angew. Chem. Int. Ed. *47*, 1508–1511.

Bikadi, Z., and Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. J. Cheminform. *1*, 15.

Björnestedt, R., Stenberg, G., Widersten, M., Board, P.G., Sinning, I., Jones, T.A., and Mannervik, B. (1995). Functional significance of arginine 15 in the active site of human class alpha glutathione transferase A1-1. J. Mol. Biol. *247*, 765–773.

Bruns, C.M., Hubatsch, I., Ridderström, M., Mannervik, B., and Tainer, J.A. (1999). Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. J. Mol. Biol. *288*, 427–439.

Chen, K.C., Wu, C.H., Chang, C.Y., Lu, W.C., Tseng, Q., Prijovich, Z.M., Schechinger, W., Liaw, Y.C., Leu, Y.L., and Roffler, S.R. (2008). Directed evolution of a lysosomal enzyme with enhanced activity at neutral pH by mammalian cell-surface display. Chem. Biol. *15*, 1277–1286.

Cherry, J.R., Lamsa, M.H., Schneider, P., Vind, J., Svendsen, A., Jones, A., and Pedersen, A.H. (1999). Directed evolution of a fungal peroxidase. Nat. Biotechnol. *17*, 379–384.

Dietrich, J.A., McKee, A.E., and Keasling, J.D. (2010). High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu. Rev. Biochem. *79*, 563–590.

Eklund, B.I., Moberg, M., Bergquist, J., and Mannervik, B. (2006). Divergent activities of human glutathione transferases in the bioactivation of azathioprine. Mol. Pharmacol. *70*, 747–754.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Evans, P.R. (1997). SCALA. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography *33*, 22–24.

Fasan, R., Crook, N.C., Peters, M.W., Meinhold, P., Buelter, T., Landwehr, M., Cirino, P.C., and Arnold, F.H. (2011). Improved product-per-glucose yields in P450-dependent propane biotransformations using engineered Escherichia coli. Biotechnol. Bioeng. *108*, 500–510.

Fersht, A. (1999). Structure and Mechanism in Protein Science: Guide to Enzyme Catalysis and Protein Folding (New York: W.H. Freeman).

Golynskiy, M.V., and Seelig, B. (2010). De novo enzymes: from computational design to mRNA display. Trends Biotechnol. *28*, 340–345.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714–2723.

Hammes-Schiffer, S., and Benkovic, S.J. (2006). Relating protein motion to catalysis. Annu. Rev. Biochem. *75*, 519–541.

Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F., 3rd., et al. (2008). De novo computational design of retro-aldol enzymes. Science *319*, 1387–1391.

Jochens, H., Aerts, D., and Bornscheuer, U.T. (2010). Thermostabilization of an esterase by alignment-guided focussed directed evolution. Protein Eng. Des. Sel. *23*, 903–909.

Jäckel, C., Kast, P., and Hilvert, D. (2008). Protein design by directed evolution. Annu Rev Biophys *37*, 153–173.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography *26*.

Lin, H., and Cornish, V.W. (2002). Screening and selection methods for large-scale analysis of protein function. Angew. Chem. Int. Ed. *41*, 4402–4425.

Liu, L., Li, Y., Liotta, D., and Lutz, S. (2009). Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. Nucleic Acids Res. *37*, 4472–4481.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. *19*, 1639–1662.

Murphy, P.M., Bolduc, J.M., Gallaher, J.L., Stoddard, B.L., and Baker, D. (2009). Alteration of enzyme specificity by computational loop remodeling and design. Proc. Natl. Acad. Sci. USA *106*, 9215–9220.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nilsson, L.O., Edalat, M., Pettersson, P.L., and Mannervik, B. (2002). Aromatic residues in the C-terminal region of glutathione transferase A1-1 influence rate-determining steps in the catalytic mechanism. Biochim. Biophys. Acta *1598*, 199–205.

Ran, N., Draths, K.M., and Frost, J.W. (2004). Creation of a shikimate pathway variant. J. Am. Chem. Soc. *126*, 6856–6857.

Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the numbers problem in directed evolution. ChemBioChem *9*, 1797–1804.

Reetz, M.T., Kahakeaw, D., and Sanchis, J. (2009). Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis. Mol. Biosyst. *5*, 115–122.

Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. *10*, 866–876.

Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., et al. (2008). Kemp elimination catalysts by computational enzyme design. Nature *453*, 190–195.

Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., et al. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science *329*, 305–309.

Schmidt-Dannert, C., Umeno, D., and Arnold, F.H. (2000). Molecular breeding of carotenoid biosynthetic pathways. Nat. Biotechnol. *18*, 750–753.

Shivange, A.V., Marienhagen, J., Mundhada, H., Schenk, A., and Schwaneberg, U. (2009). Advances in generating functional diversity for directed protein evolution. Curr. Opin. Chem. Biol. *13*, 19–25.

Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., St Clair, J.L., Gallaher, J.L., Hilvert, D., Gelb, M.H., Stoddard, B.L., et al. (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science *329*, 309–313.

Tars, K., Olin, B., and Mannervik, B. (2010). Structural basis for featuring of steroid isomerase activity in alpha class glutathione transferases. J. Mol. Biol. *397*, 332–340.

Tokuriki, N., and Tawfik, D.S. (2009). Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. *19*, 596–604.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How protein stability and new functions trade off. PLoS Comput. Biol. *4*, e1000002.

Tseng, W.C., Lin, J.W., Wei, T.Y., and Fang, T.Y. (2008). A novel megaprimed and ligase-free, PCR-based, site-directed mutagenesis method. Anal. Biochem. *375*, 376–378.

Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H., and Olsson, M.H. (2006). Electrostatic basis for enzyme catalysis. Chem. Rev. *106*, 3210–3235.

Williams, P.D., Pollock, D.D., and Goldstein, R.A. (2006). Functionality and the evolution of marginal stability in proteins: inferences from lattice simulations. Evol. Bioinform. Online *2*, 91–101.

Wu, S., Acevedo, J.P., and Reetz, M.T. (2010). Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. Proc. Natl. Acad. Sci. USA *107*, 2775–2780.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J. Appl. Cryst. *30*, 1022–1025.

Vince, R., Daluge, S., and Wadd, W.B. (1971). Studies on the inhibition of glyoxalase I by S-substituted glutathiones. J. Med. Chem. *14*, 402–404.

Zaccolo, M., and Gherardi, E. (1999). The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. J. Mol. Biol. *285*, 775–783.

Zhang, K., Li, H., Cho, K.M., and Liao, J.C. (2010a). Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. Proc. Natl. Acad. Sci. USA *107*, 6234–6239.

Zhang, W., Modén, O., and Mannervik, B. (2010b). Differences among allelic variants of human glutathione transferase A2-2 in the activation of azathioprine. Chem. Biol. Interact. *186*, 110–117.