UPTEC X 07 066
Examensarbete 20 p
Januari 2008

Structural determinants for azathioprine activity of glutathione transferase identified by screening of mutant libraries

Olof Modén

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 07 066	**Date of issue 2008-01**

Author

Olof Modén

Title (English)

Structural determinants for azathioprine activity of glutathione transferase identified by screening of mutant libraries

Title (Swedish)

Abstract

Screening of a glutathione transferase (GST) alpha class library identified chimeras with varying azathioprine activity and a structure-function relationship analysis ascribed part of the C-terminal helix as a determining factor for azathioprine activity. Azathioprine was recently shown to be activated by GSTs and enzymes with improved activity towards the prodrug azathioprine could be used in gene transfer applications. The library was constructed by shuffling of five parents, as well as a GST library with random H-site mutations. Screening of 1570 clones was performed by spectroscopic measurement of lysates at 320 nm during the reaction with azathioprine and a high-throughput method using 96-well plates for cultivation was optimized. Ten examined chimeras, with on average three parents, three crossovers and less than one point mutation, were purified by IMAC and their activities with azathioprine were shown not to be correlated with that of MCB, another GST substrate. The six most active enzymes had more than 20 % of the hA2-2 activity and a sequence identity of 85-99 %, with DNA originating from all parents. Thus, the first generation contained GSTs with enough activity and variation for continuing with a subsequent generation. Of nine recombined segments, the sequence of one segment was differing between the six most active enzymes and the other ten, and this segment, at positions 207-219, forms most of the C-terminal helix, with residues 208, 213 and 216 forming one wall of the H-site. By keeping this region constant, library sizes in later steps of directed evolution can be decreased while increasing the fraction of active GSTs. Alternatively, these key residues can be targeted for mutagenesis.

Keywords

Directed evolution, gene shuffling, glutathione transferase, prodrugs, azathioprine

Supervisors

Bengt Mannervik
Uppsala Universitet

Scientific reviewer

Per Jemth
Uppsala Universitet

Project name	Sponsors
Language **English**	Security 1 year (until Jan 2009)
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **41**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Structural determinants for azathioprine activity of glutathione transferase identified by screening of mutant libraries

Olof Modén

Sammanfattning

Riktad evolution är en samling metoder som härmar den naturliga evolutionen genom att göra stora bibliotek av gener med variation och sedan bland dessa leta efter enzymer med önskvärda egenskaper. Det tar ofta flera upprepningar eller generationer innan man når sitt mål. Detta projekt gick ut på att leta efter glutation transferaser (GST) aktiva med läkemedlet azatioprin (Imurel), som nyligen visat sig aktiveras i kroppen av olika GST. Enzymer med hög aktivitet med azatioprin skulle kunna användas för att selektivt ta bort oönskade celler som uttrycker enzymet. Biblioteket som användes bestod av en korsning av fem GST från olika arter, samt ett GST med slumpvis förändringar i tio aminosyror som binder substratet. Fördelen med korsningar av besläktade gener är att de är lika föräldragenerationens gener i de regioner som är viktiga för enzymets stabilitet och funktion. Azatioprin används idag som en dämpare av immunförsvaret bland annat under organtransplantation och aktiveras av GST genom att reagera med glutation, en antioxidant naturligt förekommande i våra celler. Ljusigenomsläppligheten förändras då och reaktionen kan följas vid en våglängd på 320 nm. Odling och mätning av aktiviteten i 96-hålsplattor optimerades och 1570 lysat med enzym undersöktes. Tio varianter av GST renades och de sex mest aktiva enzymerna visade sig ha en likhet i sekvensen av aminosyror på ner till 85 % samt en aktivitet på mer än 20 % av den högsta aktiviteten, vilket förmodligen är tillräckligt med variation och aktivitet för att tillverka ytterligare generationer. Bland dessa sex fanns en gemensam del som skiljde dem från de mindre aktiva enzymerna. Detta segment bildar sista alfahelixen i enzymet samt en av de tre väggarna som omger substratet. Framtida riktad evolution av GST kan förbättras genom att utnyttja att detta segment är extra viktigt för aktiviteten med azatioprin.

Examensarbete 20 p i Civilingenjörsprogrammet Molekylär Bioteknik

Uppsala universitet april 2007

Table of contents

Abbreviations

AATase	aspartate aminotransferase
AZA	azathioprine, 6-[(1-methyl-4-nitro-5-imidazolyl)thio]purine
CDNB	1-chloro-2,4-dinitrobenzene
DMSO	dimethyl sulfoxide
ESI-MS	Electrospray ionization - mass spectrometry
GSH	glutathione
GST	glutathione transferase
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani medium
MCB	monochlorobimane
MTT	methyl-thiazolyl tetrazolium salt
NO	nitric oxide
RP-HPLC	Reversed phase - high performance liquid chromatography
SCOP	Structural classification of proteins
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TATase	tyrosine aminotransferase

Prefixes in GST names:

b	*Bos taurus* (cow)
h	*Homo sapiens* (human)
r	*Rattus norvegicus* (rat)
A	enzyme from alpha class or alpha class library
M	enzyme from mu class

1. Introduction

1.1 Glutathione transferase (GST)

Glutathione transferase is an enzyme that transfers glutathionyl to a second substrate. The former name glutathione S-transferase might be misleading since the sulphur atom is not transferred, but the abbreviation GST is still valid in the meaning GS^- transferase. (Mannervik *et al*., 2005) The enzyme catalyzes the general reaction

(1) $$GS{-}H + R{-}X \longrightarrow GS{-}R + H{-}X$$

where R is an aliphatic, aromatic or heterocyclic group and X is a sulfate, nitrile or halide group. Glutathione transferases belong to the enzyme classification group transferases (EC 2.1.5.18). However, the reactions catalyzed can be substitutions, additions or isomerizations, depending on the nature of the enzyme and the second substrate. Generally the substrates are hydrophobic and as a consequence of that, there are mainly hydrophobic residues in the H-site (binding site for the second substrate).

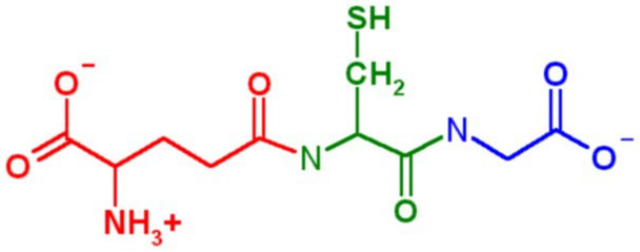


Fig 1.1. Glutathione (GSH), γ-L-Glu-L-Cys-Gly, is a peptide consisting of three amino acids.

Glutathione (fig 1.1) is a tripeptide and acts as a reducing agent. It is present in almost every cell, in mammalian cells at concentrations up to 10 mM (Eklund *et al*., 2006). The reaction between glutathione and a second substrate is often catalyzed by GSTs and the binding site for glutathione (G-site) is conserved.

The GSTs are promiscuous, meaning that each enzyme can accept many different substrates at the H-site. The promiscuity combined with the broad arsenal of different GSTs provide the cell with a possibility to conjugate glutathione to a wide range of substrates. It also provides the researchers a good model enzyme for studying evolution of substrate specificities and enzymatic mechanisms. One example is the enzymatic quasi-species found in the shuffling of two GSTs, *hM1* and *hM2*, showing how new properties might emerge during evolution (Emrén *et al*., 2006).

1.2 Phase 2 enzymes

The biological role of GSTs is, as a phase 2 enzyme, to protect against endogenous or exogenous compounds. Williams suggested in 1967 that phase 1 enzymes add functional groups to xenobiotics and that phase 2 enzymes conjugate the xenobiotics to for example glutathione (Williams, 1967). That makes them less toxic, more soluble and easier to get rid of from the cell and the body. The phase 2 genes are thus an important part of the defense against oxidative stress and the initiation of cancer. Talalay has shown that the highly

inducible phase 2 genes can be upregulated by various molecules found in plants, especially broccoli sprouts (Dinkova-Kostova *et al*., 2007). However, this protective effect of GSTs can also lead to drug resistance in pathogenic organisms carrying GSTs. For example *Plasmodium falciparum* has one GST and the anti-malaria drugs chloroquine and methylene blue has been seen to influence the glutathione metabolism (Deponte and Becker, 2005).

1.3 Four protein families with glutathione transferase activity

Based upon sequence identity, the cytosolic GSTs are divided into at least eight classes, named alpha, mu, omega, phi, pi, sigma, theta and zeta (Pearson, 2005). In the human genome the alpha and mu classes hold several members, while the other classes only have one or two members, except for the phi class which is not present in mammals at all. The structure of a cytosolic GST can be viewed in fig 1.2. There are also mitochondrial GSTs, called the kappa class. It is not certain whether this class has a different evolutionary origin than the cytosolic GSTs, even though they both share the N-terminal thioredoxin domain. The membrane bound GSTs (MAPEG, membrane-associated proteins in eicosanoid and glutathione metabolism) have three representatives in humans and are smaller than the cytosolic and mitochondrial GSTs. The last family is the fosfomycin/glyoxalase family including antibiotic resistance proteins. All four families have homologues in both prokaryotes and eukaryotes.

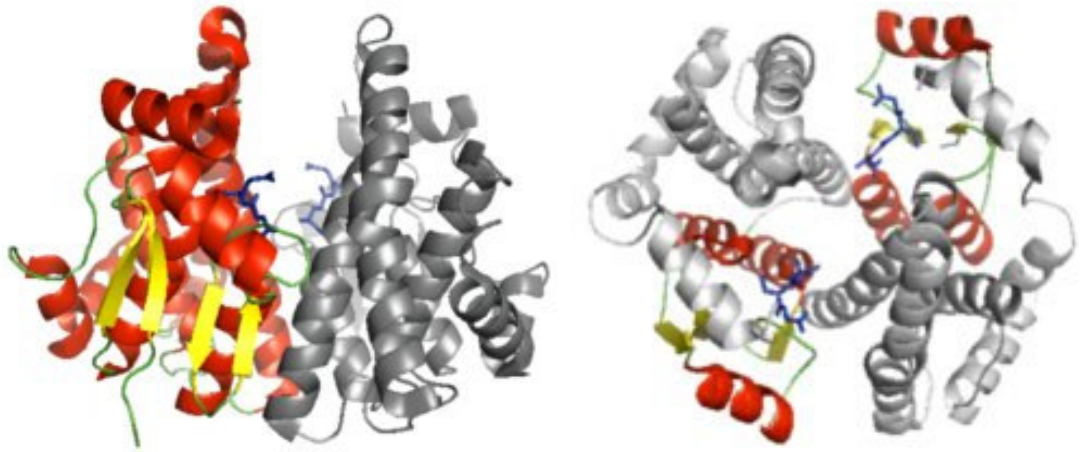


Fig 1.2. Left: Crystal structure of hA1-1, a cytosolic GST (PDB entry 1PKW). One subunit is coloured gray, while the other is coloured such that alpha helices are red, beta sheets yellow and loops green. Two bound glutathione molecules, one in each active site, are coloured blue. Right: The same enzyme viewed from "above" and coloured differently. The N-terminal thioredoxin domain consists of the secondary structure elements βαβαββα (coloured with helices as red, sheets yellow and loops green), while the C-terminal part has five helices (coloured gray). Glutathione molecules are blue. Images generated by PyMOL (version 0.98, DeLano Scientific LLC).

1.4 GST alpha class

The GST alpha class has four human members and the genes are named *hA1*, *hA2*, *hA3* and *hA4*. The enzymes are called hA1-1, hA2-2, hA3-3 and hA4-4. There are reports of heterodimers between hA1 and hA2, then named hA1-2. There is also an additional human gene named *hA5* and it is not known whether it is ever transcribed into RNA. Aside from *hA5* there are several remnants of GSTs located in the same region on chromosome 6. If they have been encoding functional enzymes or are leftovers from the evolution of the alpha class GSTs is not known. The GSTs are divided into classes based on sequence identity and within the human alpha class, *hA4* has only about 50 % amino acid identity to the other genes, which are more similar to each other (see table 1.1). This is also reflected in the intron sequences which are very similar for all the genes except for *hA4*. *hA1*, *hA2* and *hA3* are expressed mostly in

liver and kidney but also in the pancreas, lung and prostate. *hA4* is more evenly expressed in all tissue types except bone marrow (Expression data from GeneCards Version 2.34u1, July 9, 2006). The genes that will be focused on in this study are *hA1*, *hA2*, *hA3* as well as *rA2* and *rA3* from *Rattus norvegicus* and *bA1* from *Bos taurus*. Fig 1.3 shows these genes as well as other predicted alpha class genes (Data from ensembl.org, version 38, 3rd of May 2006, Hubbard *et al*., 2007). The selected genes cover a large part of the diversity in the alpha class but not the groups with *hA4* or *rA4* or the genes from *Gallus gallus* (chicken), *Monodelphis domestica* (gray short-tailed opossum) and *Canis familiaris* (dog). The GST from *Tetraodon nigroviridis* (green spotted puffer fish) is included in the alignment as an outgroup.

Table 1.1. Pairwise amino acid sequence identity for the human alpha class GSTs.

Amino acid sequence identity (%)					
	hA1	**hA2**	**hA3**	**hA4**	**hA5**
hA1					
hA2	95.0				
hA3	90.5	88.7			
hA4	53.6	54.1	53.2		
hA5	90.1	88.7	85.6	53.2	

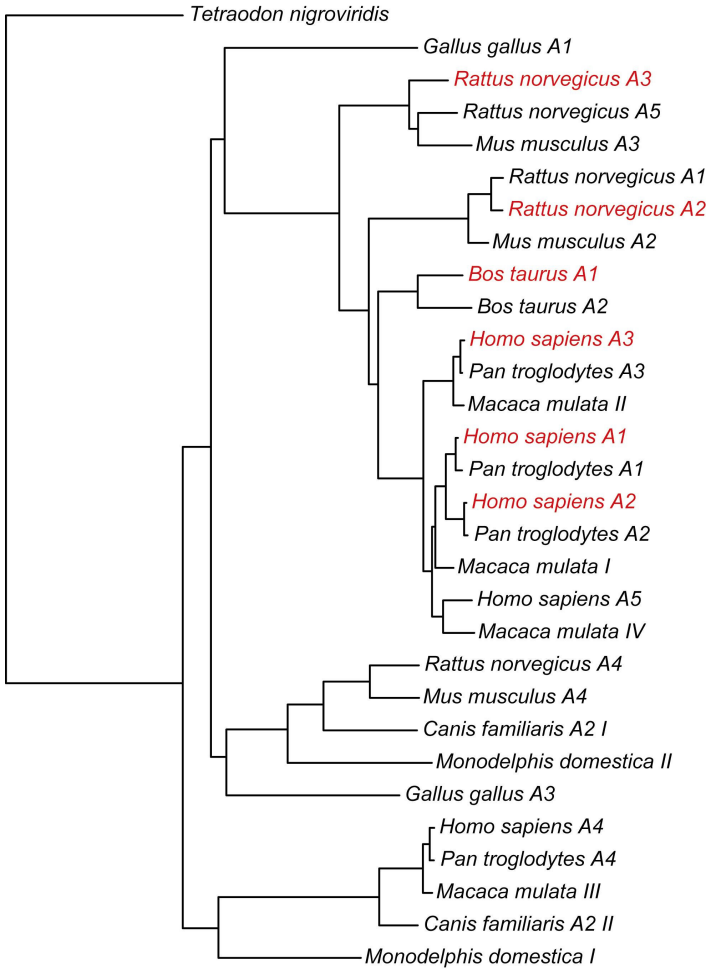


Fig 1.3. Phylogenetic tree based on the DNA sequences of 30 alpha class GSTs. This study concerns the genes marked as red. Several of the predicted genes have no proper names assigned yet and are only numbered according to their order of appearance in the database.

1.5 Steady state kinetics

Glutathione transferases catalyze bisubstrate reactions. However, if the reaction takes place with one of the substrates in a high concentration where it reaches saturation of the enzyme, it can be treated as a monosubstrate reaction, simply described as

(2) $E + S \rightleftharpoons ES \longrightarrow E + P$

When mixing enzyme (E) and substrate (S), the enzyme-substrate complex (ES) accumulates and reaches an almost constant concentration after a very short moment, called the pre-steady state. In the steady state [ES] is almost constant while [S] decreases as the product P is formed. The initial velocity, v, during the steady state is related to the initial substrate concentration, [S], according to the Michaelis-Menten equation, $v = V_{max}*[S] / (K_M + [S])$, with $V_{max} = k_{cat}*[E]_t$, where $[E]_t = [E] + [ES]$ is the total enzyme concentration (see fig 1.4). Inhibition or cooperativity will give rise to deviations from this curve.

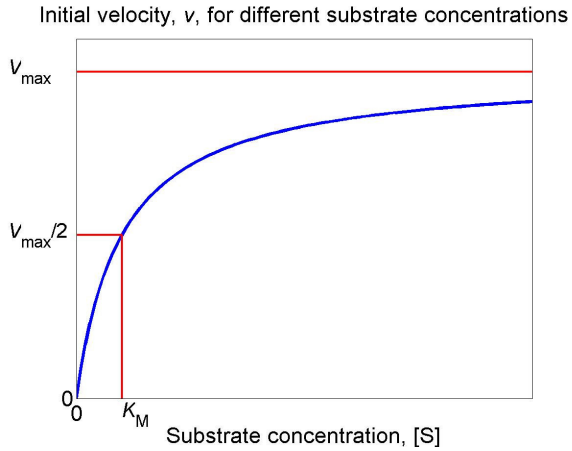


Fig 1.4. Saturation curve for a monosubstrate reaction. The Michaelis-Menten equation relates the initial velocity to the substrate concentration.

The reaction for a certain enzyme with a substrate can this way be characterized by the Michaelis-Menten constant K_M and the turnover number k_{cat}. K_M is the substrate concentration at which the velocity is half of its maximum. A higher K_M means that the enzyme needs a higher [S] to get the same velocity as an enzyme with a lower K_M. k_{cat} describes the maximum performance under saturating conditions or the number of substrate molecules "turned over" per enzyme molecule per second. However, most interesting from a biological point of view is the ratio k_{cat}/K_M, called the enzyme efficiency or specificity, since it describes the performance at low substrate concentrations, which is the usual case *in vivo*.

1.6 Azathioprine and prodrugs

Prodrugs and GSTs

While *hA3* and *bA1* are involved in steroidogenesis (Raffalli-Mathieu *et al*., 2007), *hA1*, *hA2* and *hM1* from the mu class have been reported to be involved in the bioactivation of azathioprine (Eklund *et al*., 2006), which was earlier used as an anti-cancer drug and today is used as an immunosuppressor. Azathioprine is a prodrug, which means that it must be activated in the body in order to become functional. The point of prodrugs is to prolong its lifetime in the body or to activate the drug only in the tissues where the activating enzymes are expressed. GSTs can be utilized as prodrug activators since they conjugate glutathione to

electrophilic substrates. However, in this case the product will be functional or lethal to the cell, instead of being rendered harmless.

6-MP and azathioprine

6-mercaptopurine (6-MP) was synthesized in the 1950s and was designed as a substituted base analogue. As such it was activated by several enzymes and in the end incorporated into the DNA, generating antimetabolic effects. However, the activation was too fast with severe side effects and Hitchings and Elion synthesized azathioprine by adding 1-methyl-4-nitro-5-imidazolyl to 6-MP (Elion *et al*., 1989), which then could be removed *in vivo* by conjugation with glutathione. The final result is a slower activation in the body. Azathioprine gives rise to apoptotic or antimetabolic effects, and this is probably because of a combination of substituted nucleotide incorporation and glutathione depletion, but it is not known what role the released GS-imidazole moiety play. One study showed that azathioprine decreased the metabolic activity, as measured by MTT reduction, and decreased the level of glutathione, while 6-MP did neither of them (Menor *et al*., 2004). Another work suggested that other imidazole derivatives, not connected to 6-MP, also had inhibitory effects (Crawford *et al*., 1996). Yet another study showed that azathioprine but not 6-MP decreased the expression of inducible nitric oxide synthase and thus decreased NO production, giving a potential cause of the anti-inflammatory effects of azathioprine (Moeslinger *et al*., 2006), however without considering the effects of glutathione depletion.

Reaction and polymorphism

Figure 1.5 shows the nucleophilic attack of the deprotonated glutathione on the electrophilic substrate azathioprine, a reaction catalyzed by GSTs, especially hA1-1, hA2-2 and hM1-1. This reaction was earlier thought to be non-enzymatical but Eklund showed that it is about 100 times faster with enzymes, considering the enzyme concentration in the cells (Eklund *et al*., 2006). There is polymorphism within these genes, leading to differential activation and thus side effects because of too fast activation or glutathione depletion. The largest difference is that some individuals lack *hM1* while others have two copies. Gene expression is also differing between individuals, 6-15 fold difference for *hA2* and 17-30 fold for *hA1*. *hA2* has two point mutations confirmed by RP-HPLC followed by ESI-MS, namely T112S and E210A (Coles *et al*., 2000), but it is not known how they affect the activity with azathioprine. Polymorphism is also seen in thiopurine methyltransferase, which inactivates 6-MP by methylation. One issue to keep in mind when testing the prodrug on animals is that even the alpha class GSTs have diverged so that the enzymes in different species, for example human and rat, do not have the same activity profiles, and thus *in vivo* activation might occur on a different time scale in different species.

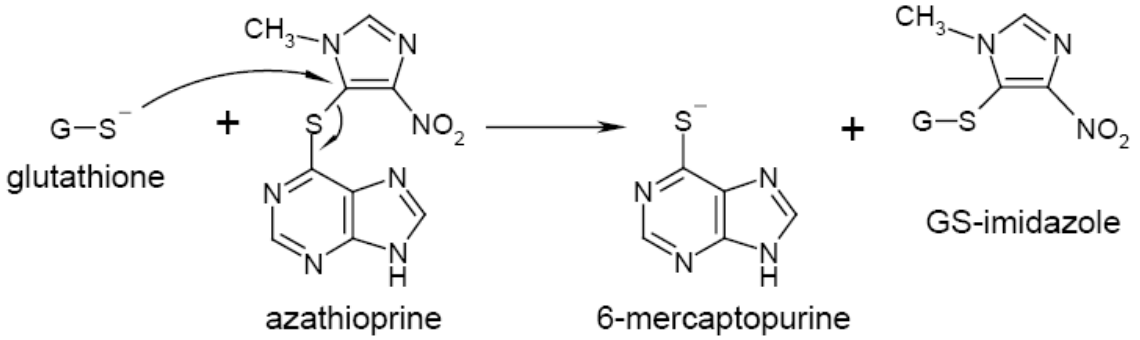


Fig 1.5. Conjugation of glutathione and azathioprine forming 6-MP and a GS-imidazole moiety. The sulphur on the deprotonated glutathione attacks the electrophilic 5' carbon on the imidazole part of azathioprine.

Immunosuppression and side effects
The locations where the different steps of activation of the azathioprine derivatives occur are not fully characterized. However, lymphocytes do not use degradation of DNA and RNA for production of new nucleotides, and are thus more sensitive to the effects of nucleotide substitutes (Patel *et al*., 2006). Azathioprine or its metabolites may also be involved in blocking of the costimulatory signal, inducing apoptosis instead of T-cell activation. Today azathioprine is mainly used as an immunosuppressor during kidney transplantation, preventing rejection, and for refractory, severe rheumatoid arthritis. Azathioprine also works on inflammatory bowel disease (IBD, Crohn's disease and ulcerative colitis), multiple sclerosis, myasthenia gravis, malignancies and autoimmune conditions. Side effects are gastrointestinal, from nausea to diarrhoea, commonly occurring within the first ten days of treatment, but there are also case reports of nausea, vomiting, anorexia, diarrhoea and fever only hours after a single dose. Most concerning however is the risk for developing malignancy, since mutations from incorporation without apoptosis might induce tumour formation. This risk is higher with heavier immunosuppression, e.g. during transplantation.

Activity and application
Most enzymes used as prodrug activators have low activity with the prodrug since the reaction should not be too fast. This is also the case with the GSTs and azathioprine. hA2-2 is the most active enzyme with azathioprine, with a k_{cat} of 5.2 s^{-1} and k_{cat}/K_M of $1.3*10^3$ $s^{-1}M^{-1}$. This is lower than GSTs with other substrates, having a k_{cat} of 100-1000 s^{-1}, and much lower than the most efficient enzymes found in nature, where the reaction is only limited by diffusion of the substrate molecules, having a k_{cat}/K_M of about 10^9 $s^{-1}M^{-1}$. One idea of this project was to find out if the examined libraries contained GSTs with activities towards azathioprine. If a GST with high activity could be successfully generated, the enzyme could be used as a negative selectable marker, for example during embryonic development research or to remove inserted cells after a successful temporal gene therapy.

1.7 Directed evolution

Enzyme engineering
Directed evolution is a collection of methods that mimic the natural evolution in producing proteins with new properties. The generated proteins can achieve new or changed functions, activities, selectivities or temperature or solvent dependencies, but the methods can also give insight into the processes of natural evolution. The products of enzyme engineering have many applications, e.g. drugs, washing powder, industrial biocatalysis such as production of biofuels or chemicals using less toxic substances, genetically modified food and various other biotechnological applications.

Variation and selection
In nature, genetic variation appears through point mutations, recombinations and insertions and deletions of pieces of DNA. This leads to phenotype differences and the organisms that survive and reproduce successfully will transfer their DNA to subsequent generations. Directed evolution needs variation too, in the form of random point mutations, e.g. from an error-prone polymerase, site-directed mutagenesis or several similar but not identical genes. The best enzymes according to the chosen criteria are then found by selection or screening methods, and these are used as starting material for subsequent generations. The library of genes in each generation can range in size from a few hundred up to billions of unique genes. Some methods are set up to test all at once, for example by transforming a library with variation of genes for antibiotic resistance into *E. coli* and then cultivating them in a medium

with addition of the antibiotic, while other methods involve that the enzymes, purified or in lysates, have to be tested one at the time.

Gene shuffling

Homologous DNA shuffling is a method where similar genes are recombined with the conserved regions of the genes as crossover locations (Stemmer, 1994), see fig 1.6. One advantage is that high variation can be achieved while keeping the conserved regions identical and these regions are expected to be conserved because they are needed for folding or functioning of the enzyme. Rational approaches to enzyme engineering often focus on mutating one or a few amino acids in the active site, but combinatorial methods also generate variation in the overall structure of the enzyme, which might change the positioning of the active site residues or the dynamics of the enzyme during catalysis. Large library sizes are often used because it is not known beforehand what changes are needed. There are also methods to create recombined libraries from non-homologous genes or with specific crossover locations (Hiraga and Arnold, 2003). The question appears which building blocks of the enzymes should preferably be shuffled and the SCHEMA energy function strategy is one approach to answer that question, utilizing a simple algorithm that counts the number of broken interactions in chimeras compared to their parent genes, producing libraries with many mutations but still a large fraction of folded enzymes (Meyer *et al*., 2003). Correlated or coevolved residues are however probably better descriptors of interacting residues than contacting pairs and these might form a network linking the active site with distant residues (Suel *et al*., 2003, Saraf *et al*., 2003). Other researchers focus on improving the shuffling on a DNA level and predicting the crossovers in homologous shuffling (Moore *et al*., 2001, Maheshri and Schaffer, 2003) and combining this with predictions of foldability or function may be fruitful.

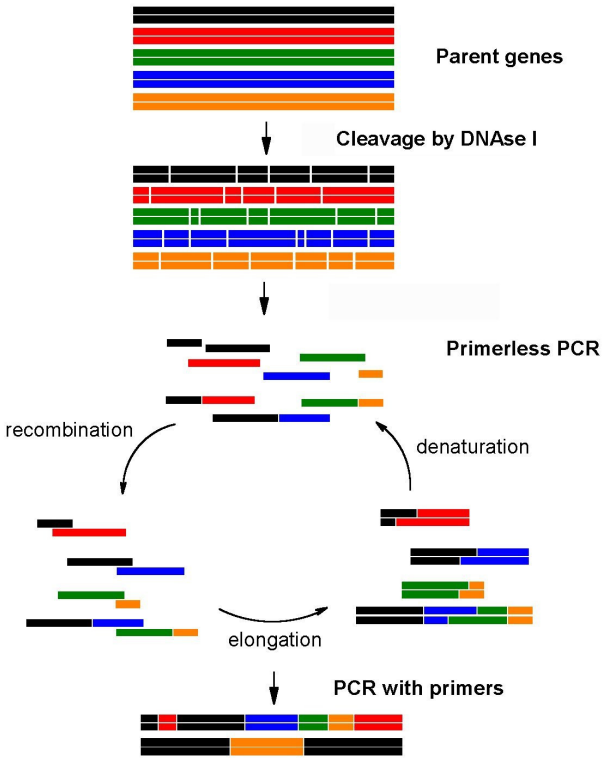


Fig 1.6. Gene shuffling of homologous sequences generates chimeras. The gene pieces themselves act as primers in the annealing step of the PCR cycle, resulting in recombination.

In vitro evolution
GSTs are promiscuous enzymes and it has been shown in other enzymes that the promiscuous functions can evolve *in vitro* without compromising the original function, which might be more robust (Aharoni *et al*., 2005). It is also hypothesized that enzymes at first are more general, with a broader range of substrates, and that it is followed by specialization, and reversed evolution to broader specificity can also be performed *in vitro* (Matsumura and Ellington, 2001). A combined approach using both rationality and directed evolution reverted the specificity of a metallohydrolase scaffold to a more undifferentiated fold followed by combinatorial insertion of predesigned loops with different properties generating, with the addition of random point mutations, specialization in a new direction (Park *et al*., 2006). Tyrosine aminotransferases (TATases) have probably evolved from aspartate aminotransferases (AATases) and this evolutionary process was replicated by directed evolution, resulting in new TATases with few mutations in regions conserved in both the parent AATases and the natural TATases and with more mutations in regions where the parents are conserved but the natural TATases carry variation (Rothman and Kirsch, 2003).

Large scale analysis
Bioinformatical approaches examining whole genomes or all known protein structures give clues to the evolution of proteins. A clustering of protein structures showed that newer proteins are smaller and often belong to the SCOP classes of α (alpha helices only), β (beta sheets only) and $\alpha+\beta$ (random mix of α and β) while older proteins are larger with the α/β class (organized combinations of alpha helices and beta sheets) as the dominant class (Choi and Kim, 2006). One theory of enzyme evolution is exon shuffling, which means not only recombination between corresponding exons in homologous genes, but rather the view of exons as building blocks that can be collected from different genes and combined into new functional genes. A competing or complementing way of generating new genes is by gene duplication followed by repeated point mutations and smaller insertions and deletions. One study that compared the rate of gene duplication to the rate of shuffling of dissimilar genes showed that gene duplication was much more frequent, at least among those genes that remained in the population, which might be logical since new genes created by exon shuffling probably also would need to be accompanied by gene duplication in order to keep the original functions (Conant and Wagner, 2005). The study was however limited to genes conserved among several species and these might mutate more slowly. The crossovers found could at least be correlated to structural domain borders but not specifically exon borders. Another work concerning exons showed that new exons most frequently reside in the untranslated regions and are alternatively spliced with low inclusion rates (Zhang and Chasin, 2006). This way they do not disturb the normal function, and older exons, which probably contributed with functionality, are included with higher rates. Surprisingly, many of these new exons consist of repeated sequences, especially Alu-repeats in humans, which can move in the genome and be exonized by a few mutations. The authors conclude that the most important contributions to new exons are not shuffled from other genes, but new repeat-like sequences that with time gain function and inclusion by splicing.

Selection methods
Selection methods based on the survival of the host organism are difficult to implement for non-trivial applications and are mostly used for studying the library generation methods themselves. However, other selection methods exist. In phage display, the enzymes will be displayed on the surface of the phage particle, available for selection of binding properties, while the DNA resides inside the phage particle (Scott and Smith, 1990). In mRNA display, the translated RNA will remain attached to the emerging polypeptide chain via a puromycin

linker and the DNA of the polypeptides, selected by affinity chromatography, is retrieved by the use of reverse transcriptase (Roberts and Szostak, 1997). *In vitro* compartmentalization uses water-in-oil emulsion to make many separate transcription-translation systems, where the substrate and thus the product, if there is a reaction, is attached to the DNA template which can be selectively enriched and characterized (Tawfik and Griffiths, 1998).

Amino acids and codons

Research has been done to increase the range of functionality in enzymes by incorporation of new amino acids, e.g. by post-translational modification of cysteine (Hegazy *et al*., 2006) or by loading of natural or mutant aminoacyl tRNA synthetases with unnatural amino acids (Hartman *et al*., 2006). Others have produced enzymes with as few amino acids as possible, e.g. the SH3 domain with five amino acids only (Riddle *et al*., 1997). The amino acids Ala, Gly, Ser, Asp, Asn and perhaps Val are proposed to be the first amino acids on earth, based on GC-content in codons, short simple synthetic pathways, association with aminoacyl-tRNA synthetase class II, observation in prebiotic experiments and inclusion in secondary structure elements, especially beta sheets (Klipcan and Safro, 2004). Using several approaches, e.g. the exchange cost also found in scoring matrices used for multiple alignment, the minimum size of the amino acid alphabet is proposed to be around ten, grouping them for example into aromatic, polar/acidic, basic and different groups of hydrophobic with cysteine, proline and glycine as single amino acid groups (Murphy *et al*., 2000, Fan and Wang, 2003). The information about the amino acids and their codons can be utilized when producing libraries with random point mutations, if smaller library sizes are desirable or when some properties but not all are wanted.

1.8 Aims

One aim of the project was to optimize the azathioprine screening procedure. The first generations of two libraries, created using gene shuffling and site-directed random mutagenesis, were then screened for GSTs with azathioprine activity. If a GST with high activity could be generated, the enzyme could be utilized as a negative selectable marker. Interesting mutants were sequenced, purified and kinetically characterized followed by a structure-function analysis. Information obtained could be used to compare the different approaches of directed evolution, to improve our knowledge of azathioprine activation and to guide future experiments, for example to determine the parents of subsequent generations in the gene shuffling.

2. Method

2.1 Library preparation

The alpha class library was constructed previously by Sanela Kurtovic (data not yet published). The glutathione transferase genes used in the gene shuffling were *hA2*, *hA3*, *bA1*, *rA2* and *rA3*, with N-terminal 6xHis-tags, as well as the *hA1*-library earlier produced by Mikael Widersten (Widersten and Mannervik, 1995). The *hA1*-library contains the *hA1* gene with random mutations in ten amino acid residues positioned around the second substrate at the H-site. The parents were integrated into the pGΔETac vector (Widersten and Mannervik, 1995) with a six-His-tag inserted at the N-terminal, for later purification of enzymes using immobilized metal ion affinity chromatography (IMAC, Porath *et al*., 1975). A schematic picture of the gene shuffling performed can be viewed in fig 1.6. The libraries were transformed into *E. coli* cells (XL1-Blue, Stratagene) by electroporation, and spread on LB agar plates with ampicillin, where single colonies could be picked for screening.

2.2 Screening

The alpha class library was screened in two separate rounds. The screening process involves lysate preparation followed by activity measurement with azathioprine. The first round of lysate preparation was done by Sanela Kurtovic and Abeer Shokeer, using a culture volume of 10 ml in Falcon tubes. This generated enough lysate, 300 µl, from each colony for measurement with several substrates and 5 µl of the lysate was enough for measurement with azathioprine.

In the second round of screening, the lysate preparation was optimized for 96-well plates. By using two plates, 192 clones could be screened simultaneously, compared to 24 or 48 for the preparation in Falcon tubes. 90 µl of lysate was collected from each well and 40 µl was used in the activity measurement, resulting in roughly the same signal for hA2-2 compared to the non-enzymatic reaction as in the Falcon preparation using 5 µl.

Lysate preparation in 96-well plates started by picking colonies from the transformed libraries. If they were not fresh, they were grown overnight on LB agar plates with ampicillin. Two wells on each plate contained colonies with *hA2* as a positive control. The cultures were then grown overnight at 37 °C in 300 µl LB media with ampicillin (100 µg/ml), at 200 rpm in the plate incubator. The plates (Nunc, 267334) were covered with pierced aluminium foil in order to minimize contamination due to the condensation. Next day, the cultures were diluted 100 times in 2TY media with ampicillin (100 µg/ml) and grown for about 21 hours. In the Falcon preparation and when purifying, IPTG is added after a few hours to induce expression of the protein, but in this case IPTG is replaced by a longer growth time, giving similar expression at lower cost and effort. The bacteria were harvested by centrifugation for 10 min at 3500 rpm, 4 °C, and the pellets resolved in 150 µl 0.1 M sodium phosphate buffer pH 7.0 with 0.2 mg/ml of lysozyme. After leaving on ice for an hour the cells were lyzed by freeze thawing using three cycles of -80 °C (10 min) and 37 °C (4 min). The plates were centrifuged for 30 min at 3500 rpm, at 4 °C. The pellets were removed by loosening them and taking up 45 µl of pellet and lysate from each well. 90 µl of pellet-free lysate were then collected and stored in -80 °C or used directly for measurement of azathioprine activity.

The activity measurement was performed in a total volume of 250 µl in 96-well plates (Nunc, 269620). The product formation is coupled to an increase of the absorbance at 320 nm (Eklund *et al*., 2006). The initial reaction rate at 30 °C was studied and compared to the non-enzymatic reaction. The reaction rate was calculated from the change in absorbance according to the Beer-Lambert law, A = c*k*l, where A is absorbance, c is concentration, k is the extinction coefficient for azathioprine at that wavelength (=16000 $M^{-1}cm^{-1}$) and l is the path length. A SpectraMax Plus 384 plate reader was used for absorbance measurement and SoftMax Pro 3.12 for data analysis. From plate preparation, 40 µl of lysate was used and from Falcon preparation, 5 µl lysate was used. Glutathione (5 µl, 50 mM) in buffer was added for a final concentration of 1 mM. Azathioprine (5 µl 10 mM) in DMSO was added for a final concentration of 0.2 mM. The sample was mixed before measurement in the plate reader. The buffer used was 0.1 M sodium phosphate pH 7.0. When measuring specific activity or steady state kinetics, pH 7.4 was used, but the pH during screening was lowered to 7.0 in order to make glutathione more protonated and thus decrease the non-enzymatic reaction rate. Since the path length during plate reader measurement is not only depending on the volume but also on the surface tension of the sample, the wells with lysate added had more liquid closer to the edges of the wells and thus shorter path lengths in the middle where the absorbance was determined. Because of this the non-enzymatic reaction rate was adjusted afterwards by a factor of 0.856, which was determined in the same way as described later in section 2.6 (Steady state kinetics). The initial rate of the reaction was calculated as an average during an interval, which was chosen as early as possible while still avoiding artifacts that are more commonly occurring within the first 100 seconds. This interval was chosen to be 100-300 seconds from the start, but when not suitable, e.g. if bubbles had distorted the measurement in this region, another interval was chosen. Since the reaction rate slowly increased with time, which might have been an effect of the temperature rising, later intervals would give higher values.

The first round of screening, using lysate preparation in Falcon tubes, covered 456 clones, while the second round in 96-well plates covered 1114 clones. Interesting colonies were grown for sequence determination and if still interesting, the enzymes were purified and kinetically characterized. The *hA1*-library was screened in one round, covering 384 clones, with lysate prepared by Ronnie Jansson in the same way as the previously described Falcon preparation.

If the colonies were dead when they should be picked for cultivation and analysis, they were revived from the agar plate, by purifying their DNA using Viogene Mini-M Plasmid DNA Extraction System, and transforming them into electrocompetent *E. coli* XL1-Blue cells. 5 µl of DNA with a concentration of 1-2 ng/µl was mixed with 50 µl of electrocompetent cells and transformed by electroporation. The bacteria were grown in 400 µl 2TY for one hour and then spread on two agar plates with ampicillin. Cultures of interest were stored as freeze cultures, by mixing culture with glycerol for a concentration of 15 %.

2.3 Sequencing

Interesting clones were cultured in order to produce DNA for sequencing. Plasmid preparation was done from growth volumes of 2 ml or 3x2 ml using Viogene Mini-M Plasmid DNA Extraction System, vacuum method. For better sequencing performance, QIAGEN Plasmid Maxi Protocol was used on cultivation volumes of 500 ml. The concentration of the purified DNA was assessed using NanoDrop. The DNA was sequenced by Ulla Gustafson at the Department of Animal Breeding and Genetics, using existing primers.

Sequence data was evaluated using Chromas 2.31 (Technelysium Pty Ltd). Multiple alignments were performed using emboss:emma (Clustal W 1.83), followed by manual correction in some cases, and visualized using emboss:showalign (Rice *et al*., 2000). Guide trees from the multiple alignments were used as simple similarity trees, for example to display the similarities among chimeras and parents in fig 3.8, also with the help of emboss:fdrawtree. A phylogenetic tree was constructed, without bootstrap sampling, from a multiple alignment of 30 alpha class GSTs (fig 1.3), using emboss:distmat (Tamura correction) to create a distance matrix, emboss:fneighbor (Neighbor-joining method, Saitou and Nei, 1987) to produce phylogenies, phylip3.66:retree to place the outgroup at the root (Felsenstein, 1989) and emboss:fdrawgram to visualize the tree. The sequences of the chimeras and parents were compared to their respective enzymatical activities with azathioprine and analyzed both using the whole sequence and specifically the H-site residues which are expected to bind the substrate.

2.4 Purification

His-tagged GSTs were purified using IMAC (Porath *et al*., 1975). A smaller culture was diluted into 500 ml of 2TY medium with 100 µg/ml ampicillin. When OD_{600} = 0.3, expression was induced by IPTG (0.2 mM). After overnight growth at 37 °C, the bacteria were centrifuged and lysed by lysozyme (0.2 mg/ml) and sonication. The sample was added to a His GraviTrap™ column (GE Healthcare, Amersham Biosciences AB, containing precharged Ni Sepharose™ 6 Fast Flow), with a binding buffer of 20 mM sodium phosphate, 500 mM NaCl, 20 or 50 mM imidazole and pH 7.4, and eluted by an elution buffer of 20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole and pH 7.4. Dialysis was performed twice overnight with 0.1 M sodium phosphate buffer having pH 7.0 or 7.4. Concentrations were first determined by NanoDrop measurement and later adjusted by a Bio-Rad Protein Assay (bovine serum albumin standard) in plates. The purity was assessed by SDS-PAGE, with 12.8 % total acrylamide and 0.34 % total bis-acrylamide in the separating gel and 3.8 % and 0.10 % in the stacking gel. The parents hA1-1, hA3-3, bA1-1, rA2-2 and rA3-3 were purified and provided by others.

2.5 Specific activity

Azathioprine activity was assayed spectrophotometrically, with an extinction coefficient of 16000 $M^{-1}cm^{-1}$ at 320 nm, using SpectraMax Plus 384 and SoftMax Pro 3.12. The concentration of azathioprine was 0.2 mM (in DMSO), glutathione 1.0 mM (in buffer, pH 7.4) and the sodium phosphate buffer 0.1 M (with pH 7.4 and 0.1 mM EDTA). The reaction took place at 30 °C and the buffer was preheated to 30 °C. The total volume was 500 µl in a quartz cuvette or 250 µl in a plate, and the volume of enzyme solution was 10 µl or 5 µl respectively. The initial rate was measured as an average during the first minute with the non-enzymatic reaction rate subtracted, and the activity was calculated according to the Lambert Beer law A = c*k*l, where A is absorbance, c is concentration and k is the extinction coefficient. (Eklund *et al*., 2006)

CDNB activity was also assayed spectrophotometrically, with an extinction coefficient of 9600 $M^{-1}cm^{-1}$ at 340 nm. The concentration of CDNB was 0.2 mM (in 95 % ethanol), glutathione 1.0 mM (in buffer, pH 6.5) and the sodium phosphate buffer 0.1 M (with pH 6.5). The reaction took place at 30 °C and the buffer was preheated to 30 °C. The total volume was 1000 µl in a quartz cuvette, and the volume of enzyme solution was 20 µl. The initial rate was measured as an average during the first minute with the non-enzymatic reaction rate subtracted. (Ivarsson *et al*., 2003)

MCB activity was assayed fluorometrically, with an excitation wavelength of 355 nm and an emission wavelength of 500 nm, using Fluoroskan Ascent Type 374 (Thermo Labsystems Oy) and Ascent Software v 2.6. The concentration of MCB was 40 µM (in acetonitrile), glutathione 10 mM (in buffer, pH 6.5) and the sodium phosphate buffer 0.1 M (with pH 6.5). The total volume was 100 µl, and the volume of enzyme solution was 10 µl. The enzyme solutions were diluted to generate similar signals, of about 0.3-0.6 arbitrary units per minute. The samples were incubated at 30 °C for 5 minutes before MCB addition, and the initial rate of fluorescence increase was measured as an average during one minute, also at 30 °C. The non-enzymatic reaction rate, although very small, was subtracted, and the activities presented as percent of the most active enzyme. (Eklund *et al*., 2002)

2.6 Steady state kinetics

The absorbance of azathioprine at 320 nm was too high for measurement at near-saturating concentrations in a 0.5 cm cuvette. The extinction coefficient had been determined earlier for 320 nm (16000 $M^{-1}cm^{-1}$) since the difference in absorbance between substrate and product was high and fairly constant in this region, which is favorable if wavelength drift occurs. In order to make saturation curves, a shorter path length should be used or a new wavelength should be chosen. In this case, a shorter path length was obtained by using a small volume of 40 µl in a plate reader spectrophotometer, resulting in a path length of about 1 mm.

Because of the meniscus, samples with the same volume can give rise to different path lengths. To make the path lengths of the enzymatic and non-enzymatic reaction equal, Triton X100 was added at a concentration low enough to still ensure full enzymatic activity. This was later seen to be an unnecessary step, since the path length can be measured and adjusted for. The path length can be measured either using the absorbance of azathioprine at different concentrations or using the absorbance peak of water at around 1000 nm (SOFTmax® PRO User's Manual 4-20, Molecular Devices Corporation, 2000). The latter gives the path length d = $(OD_{1000} - OD_{900})/k$, and the constant k was determined by measurement of the solutions in cuvettes with known path length.

The path length of about 1 mm allowed for azathioprine concentrations ranging from 0.025 to 3.2 mM, with a total DMSO concentration of 4 % ensuring solubility. Concentrations were for glutathione 1 mM (in buffer), for enzyme 0.01 mg/ml and for Triton X100 0.002 % and the reaction took place in sodium phosphate buffer 0.1 M with EDTA 0.1 mM, pH 7.4. For each substrate concentration, two times four samples were run, both for the enzymatic and the non-enzymatic reaction. These were all placed in the same column on a plate to avoid systematic errors from absorbance differences between rows. The reaction took place during 1 minute followed by an endpoint measurement, in order to retrieve the absorbance at 900 and 1000 nm to calculate the path length as an average of all samples of the same kind. A SpectraMax Plus 384 plate reader and SoftMax Pro 3.12 were used.

Non-linear regression was performed to fit the initial rate data to the Michaelis-Menten equation, using Prism GraphPad or SimFit. k_{cat}/K_M were calculated either from k_{cat} and K_M or from the slope of the saturation curve at low substrate concentrations.

2.7 Screening improvements

MCB correlation

There exists a method to screen for active GSTs by spraying agar plates with MCB (Eklund *et al*., 2002). The enzymatic reaction with MCB give rise to a fluorescing product and the colonies with active GSTs can be identified and picked for further screening. The activity

with MCB was measured for 14 purified GSTs and a test of correlation with azathioprine activity was performed.

Azathioprine-agar plates
Another idea is to test if bacteria with an azathioprine active enzyme are more or less able to survive and grow in the presence of azathioprine. If bacteria with active enzymes do not grow, then a negative selection system can be made. A transformed library could then, after a replica is made, be transferred to a plate with azathioprine and the bacteria that contain active enzymes would not grow.

Three tests were performed to explore this idea. The first was with bacteria in solution to find an azathioprine concentration that inhibited the growth. A too high concentration would probably kill all cells while all would survive in a too low concentration. In order to differentiate on enzymatic activity, a concentration where only some are inhibited should be found. Besides, there would be aggregation if the concentration of azathioprine was too high while the total DMSO concentration was low. DMSO and ethanol are typical co-solvents, but ethanol is known to induce expression and DMSO is also shown to affect the expression of several enzymes in the oyster mushroom *Pleurotus ostreatus* (Shah *et al*., 2006). At the choice of 1 % DMSO, the maximum azathioprine concentration is determined to be 0.8 mM. Concentrations of 0-0.8 mM in 1 % DMSO as well as 0 % DMSO were used when growing bacteria with active (hA2) and inactive enzymes (A260) in 0.5 ml medium (2TY + 100 µg/ml ampicillin) for 16 h in 37 °C. The results of growth were determined visually.

The second test involved azathioprine-agar plates. They were made with LB medium, 0.8 mM azathioprine, 1 % DMSO and 100 µg/ml ampicillin. Single colonies with active (hA2, A1362) and inactive enzymes (A260) were streaked on the plates and incubated overnight at 37 °C.

In the third test, the number of bacteria for two different clones (A260 and hA2) was controlled by OD_{600} measurements, to ensure an equal starting situation. One solution of each kind, that both had an OD_{600} of 0.26, were diluted 1x, 100x, 10000x and 1000000x. 10 µl of each solution were spread on both normal agar plates and azathioprine-agar plates, and incubated overnight at 37 °C. The dilutions made it possible to count the total number of bacteria and calculate their survival ratio on the azathioprine-agar plates.

2.8 Comparative modeling

Because of the sequence similarities and the expected conservation of the fold within the alpha class GSTs, the structures of the parent enzymes without a known structure as well as interesting chimeras can be approximated using comparative modeling. One should have in mind that the structures produced are often more similar to the template structure used than to the true structure of the modeled protein and that refinement using molecular dynamics or similar would not necessarily generate a more relevant structure than by keeping the coordinates of the unchanged backbone. Even though the protein is dynamic in a physiological environment a crystallized protein can still bind the substrates and often be catalytically active. However, the information obtained from comparative modeling can at least be used for visualizing the exchanged amino acids and allow for speculations of their importance on the protein function.

SWISS-MODEL provides a net-based interface to a modeling server as well as through the molecule visualization tool DeepView/Swiss-PdbViewer (Schwede *et al*., 2003, Guex and Peitsch, 1997, Peitsch, 1995). SWISS-MODEL uses the program BLAST-P2 to find similar

sequences, SIM to find template structures, ProModII to generate a model and Gromos96 to energy minimize the model. ProModII superimposes the template structures, generates averaged coordinates, rebuilds lacking loops, corrects the backbone and sidechains, verifies the model, checks the packing and refines the structure using energy minimization. Loops are rebuilt by fitting fragments derived from the Brookhaven Data Bank. The first two programs can be excluded by providing the sequence alignment and template structure manually. This way one subunit of hA2-2 was modeled using the coordinates from a crystal structure of hA1-1 bound to glutathione, from the ExPDB template 1pkwA based on the PDB entry 1PKW (Grahn *et al.*, 2006).

3. Results

3.1 Screening and characterization

Screening

The screening of the alpha class library resulted in 8 clones out of 456 from the first round and 16 clones out of 1114 from the second round that were analyzed further. The distribution of the reaction rates during the second round of screening can be seen in fig 3.1. The positive control (*hA2*, 29 values from different preparations) generated screening activity values of 14.6-37.8 $10^3 dA_{320}$/min with an average of 27.0 (SD 5.5) $10^3 dA_{320}$/min. The non-enzymatic reaction had a reaction rate of 7.4 (SD 0.4) $10^3 dA_{320}$/min. An average of all lysate reaction rates were 7.6 (SD 2.3) $10^3 dA_{320}$/min, while all lysates with a reaction rate under 12.25 $10^3 dA_{320}$/min had an average of 7.44 (SD 1.39) $10^3 dA_{320}$/min and a median of 7.32 $10^3 dA_{320}$/min.

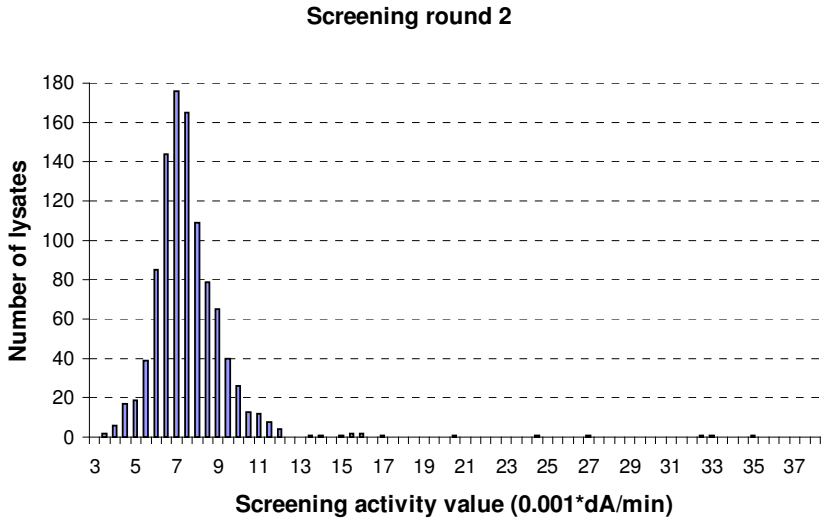


Fig 3.1. The distribution of reaction rates in the second screening. Positive controls not included.

The screening of the *hA1*-library resulted in only one enzyme with high activity, and this was later identified as the parent hA1-1.

Sequencing

Plasmid preparation using 2 ml growth media resulted in DNA concentrations of 30-60 ng/µl, and when combining 3x2 ml into each column 50-100 ng/µl was obtained. Preparation using 500 ml growth media resulted in DNA concentrations of 250-4000 ng/µl and had to be diluted to 200-500 ng/µl before sequence determination.

Of the eight clones examined from the first round, six were *hA2* and the two with low screening reaction rates were chimeras. Since the sequencing results in some cases were ambiguous, the activity profiles of purified enzymes with azathioprine and CDNB were determined and the identities of the clones as *hA2* were confirmed. The second round generated eight chimeras as well as one *hA1* and seven *hA2*. In total, 24 clones were chosen after a screening of 1570 clones from the alpha class library, and of those 14 were parents while 10 were chimeras. The sequences of the ten chimeras and the six parents are presented

in appendix A. All chimeras and parents, except for *bA1*, also carry the 6xHis-tag not shown in the list of sequences in appendix A.

Purification

Fig 3.2 shows that the main components of the lysates were GSTs and that probably both monomer and dimer was seen on the SDS-PAGE gel. The ten chimeras were purified but most still contained contaminants, as seen in fig 3.3. The bands of different GSTs and possibly the dimers were not on the exact same level, which is in agreement with earlier observations.

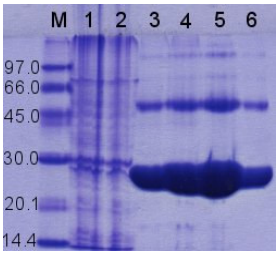


Fig 3.2. SDS-PAGE gel of marker, lysates and purified GSTs. M: molecular weight marker. Lanes 1 and 2: lysate. Lanes 3-6: Overload of purified hA2-2.

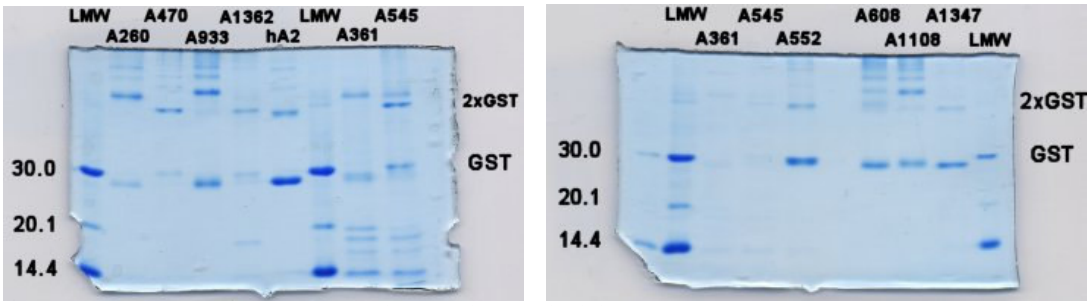


Fig 3.3. SDS-PAGE gel of marker and purified GSTs, both hA2-2 and chimeras from the alpha class library. LMW = molecular weight marker. Left: About 5 µg of the five first samples, which were purified with 50 mM imidazole in the binding buffer. About 25-30 µg of A361 and A545, which were purified with 20 mM in the binding buffer. Right: About 5 µg of each sample, which were purified with 20 mM imidazole in the binding buffer.

Specific activity

The specific activities with azathioprine of the parents and the chosen chimeras were determined. The enzymes were divided into three groups according to activity to simplify the analysis, see fig 3.4, with 6 highly active, 5 with low activity and 5 with almost no activity.

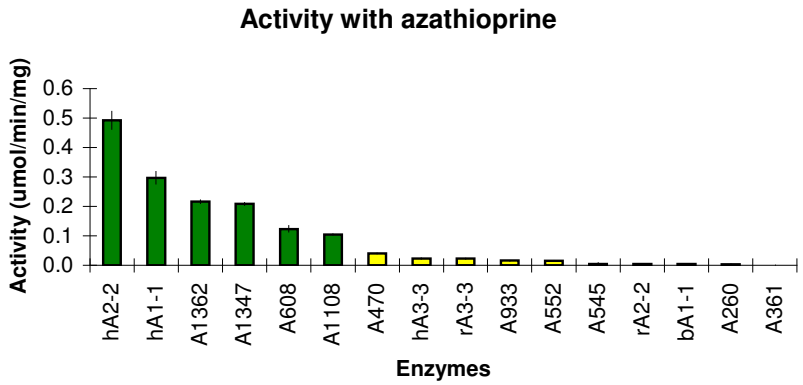


Fig 3.4. Specific activity of the alpha class parents and ten chimeras as purified enzymes. The enzymes are grouped based on their activity as "high activity" (green), "low activity" (yellow) or "almost no activity" (red). The thin bars represent the standard deviations.

Steady state kinetics

Saturation curves were determined for the three enzymes with highest activity towards azathioprine and the calculated constants can be viewed in table 3.1. A saturation curve for hA2-2 with azathioprine is shown in fig 3.5.

Table 3.1. Kinetic constants regarding azathioprine for the three most active enzymes.

enzyme	specific activity (umol/min/mg)	standard deviation	% of hA2-2 activity	kcat (1/s)	Km (mM)	kcat/Km (1/s/mM)
hA2-2	0,49	0,04	100	4,0	3,0	1,35
hA1-1	0,30	0,02	60	8,2	11,7	0,70
A1362	0,22	0,01	44	2,5	4,1	0,61

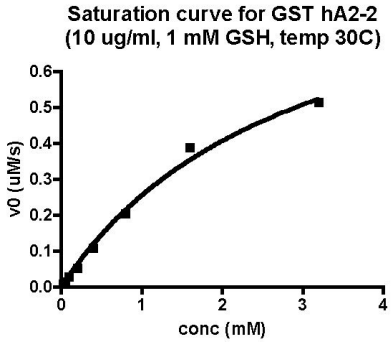


Fig 3.5. Saturation curve for GST hA2-2.

3.2 Library and chimera analysis

Distribution of H-site residues on enzymes with different activity

Ten of the twelve H-site residues, that surround the substrate binding pocket, differ among the enzymes, see table 3.2.

Table 3.2. Amino acids at the H-site positions for the parents and ten chimeras, together with the azathioprine (AZA) activity as percent of the highest activity. The H-site residues are coloured based on the specific activity of the enzyme as green ("high activity"), yellow ("low activity") or red ("almost no activity"). The names of the parents have a blue background.

		H-site residue position											
Enzyme	**AZA activity**	**10**	**12**	**14**	**104**	**107**	**108**	**110**	**111**	**208**	**213**	**216**	**222**
hA2	100 %	S	I	G	E	L	L	P	F	M	L	S	F
hA1	60 %	F	A	G	E	L	L	P	V	M	L	A	F
A1362	44 %	S	I	G	E	L	H	P	Y	M	L	S	F
A1347	42 %	S	I	G	E	L	L	P	F	M	L	S	-
A608	25 %	S	I	G	E	M	H	P	L	M	L	S	F
A1108	21 %	S	I	G	E	M	H	P	L	M	L	S	F
A470	8 %	F	G	G	E	L	H	P	Y	E	V	A	-
rA3	5 %	F	G	G	E	L	H	P	Y	E	V	A	-
hA3	5 %	F	G	G	E	L	L	P	L	A	L	A	F
A933	3 %	S	I	G	E	M	H	P	L	A	L	A	F
A552	3 %	S	I	G	E	L	H	P	Y	A	L	A	F
A545	1 %	S	I	G	E	L	H	P	Y	E	V	A	-
bA1	1 %	F	G	G	E	M	H	P	L	T	I	A	F
rA2	1 %	F	A	G	E	I	Q	V	I	M	I	A	F
A260	1 %	S	I	G	E	M	H	P	L	A	L	A	F
A361	0 %	S	I	G	E	L	H	P	Y	M	I	A	F

Whole sequence comparison

Fig 3.6 shows the sequences of parents and chimeras, divided into conserved and variable regions. The chimeras are colour coded as the parents to display the origin of each region. The origin was based on amino acid identity but was not obvious in all cases. In some regions, the DNA sequence could determine the origin. Sometimes a region had the same DNA sequence in several of the parents and then proximity to a region with known origin decided the origin of this region as well. For very similar sequences, especially *hA1* and *hA2*, the origin simply had to be chosen as one of them, and in this case *hA2* was always chosen when it was not clear from the sequence. The ten chimeras had in total six point mutations in the amino acid sequences, leading to 0.6 point mutations per chimera or about 0.27 % mutations per amino acid. There were also three amino acid mutations (one in A1362 and two in A1108) that could either be point mutations caused by the polymerase or mutations caused by recombination, later called "unclear point mutations". On average there were 3.2 parents per chimera and 3.6 crossovers, leading to 4.6 different segments per chimera. However, if the unclear point mutations were considered point mutations instead of recombinations, there would be on average 0.9 point mutations, 3.0 parents and 3.2 crossovers (4.2 segments) per chimera.

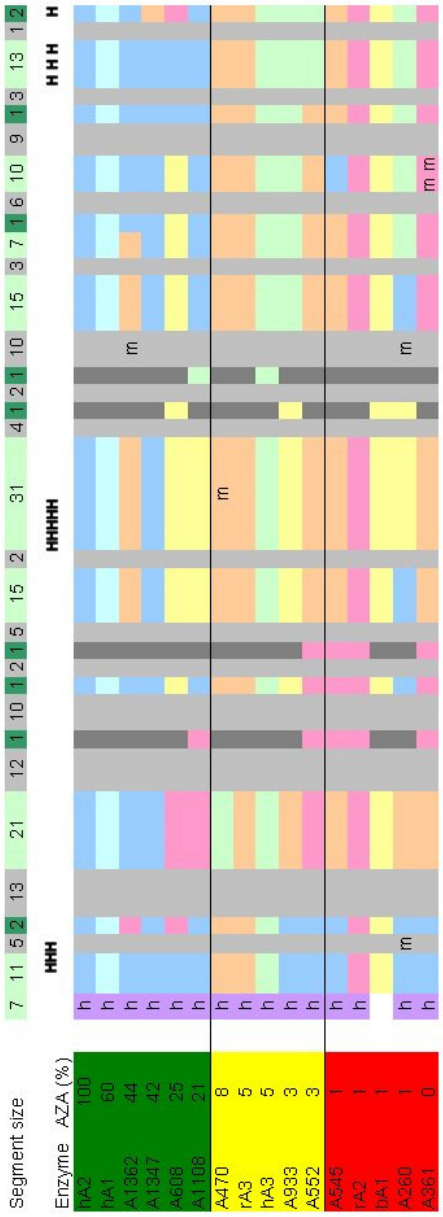


Fig 3.6. Sequences of parents and ten chimeras, divided into conserved and variable regions. The segments of chimeras are colour coded to show the origin of their segments, with light blue (hA1), blue (hA2), green (hA3), yellow (bA1), red (rA2) and orange (rA3). Gray segments are conserved among the parents. Dark gray segments are identical in all but one parent. "m" means one point mutation. "h" means His-tag. "H" represents approximate position of a H-site residue. The enzyme names are colour coded according to the specific activity of the enzymes with azathioprine as green ("high activity"), yellow ("low activity") or red ("almost no activity"). The activity with azathioprine (AZA) is shown as percent of hA2-2.

In fig 3.7, it can be seen that *hA2* dominated the origin of the group with active enzymes. Aside from a single amino acid, only four of the five parents were represented in this group. rA3 and hA3 dominated the low-active enzymes and rA3, rA2 and bA1 dominated the almost inactive enzymes.

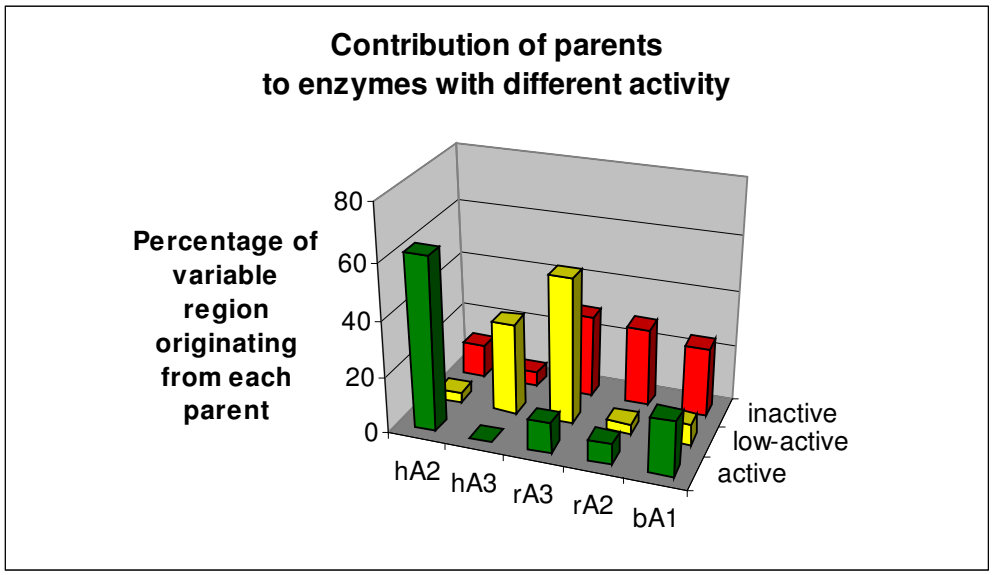


Fig 3.7. The percentage of the non-conserved region originating from each parent for enzymes with different specific activities, grouped and coloured as in fig 3.4.

The sequence similarity of the chimeras and parents were represented as the number of identical amino acids when compared to the closest parent, and this value was ranging from 83 % to 99 %. A similarity tree was also constructed, see fig 3.8.

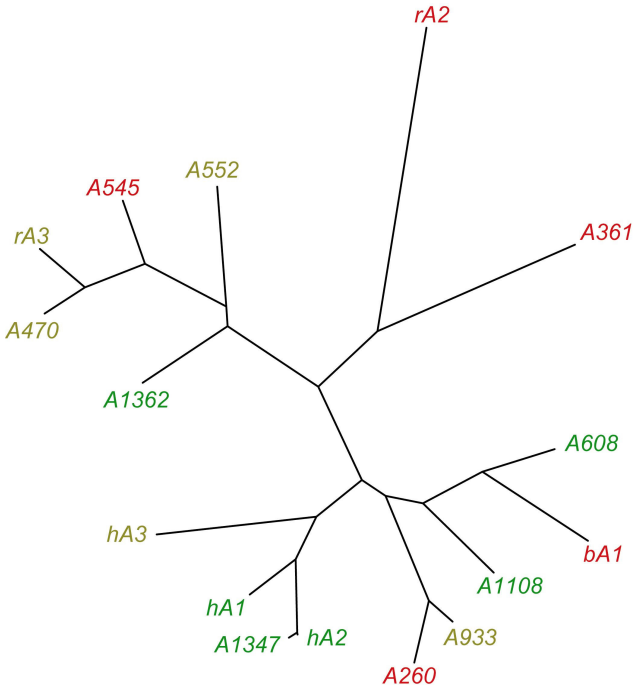


Fig 3.8. Similarity tree with the parents and ten chimeras based on amino acid sequences. The gene names are colour coded according to the specific activity of the enzymes with azathioprine as green ("high activity"), yellow ("low activity") or red ("almost no activity").

Analysis of variable regions

To enable an analysis of the main segments that were shuffled and how they contributed to the enzymatic activity, the information in fig 3.6 needed simplification. In fig 3.9 the

sequences of the parents and chimeras are presented without conserved regions and single variable sites, showing only the main variable segments. Besides, segments from different parents are colour coded with the same colour if their sequences are identical, and segments that only differ by one mutation are described with the same colour as well, but with an "m" for mutation. This way, the similarities among the enzymes are emphasized.

Segment		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
From residue		2	33	86	103	152	170	184	207	221
To residue		12	53	100	133	166	176	193	219	222
Segment size		11	21	15	31	15	7	10	13	2
Max. num. of mut.		6	13	4	20	9	4	6	8	2
Enzyme	**AZA (%)**									
hA2	100									
hA1	60			m					m	
A1362	44									
A1347	42									
A608	25			m			m			m
A1108	21			m						
A470	8				m					
rA3	5									
hA3	5					m				
A933	3			m		m				
A552	3									
A545	1									
rA2	1									m
bA1	1	m		m			m			m
A260	1									
A361	0							mm		m

Fig 3.9. Nine variable segments within the parents and chimeras. Conserved regions and single variable sites are left out. The colour of the segments are originally the same as of the parent it originates from, namely light blue (hA1), blue (hA2), green (hA3), yellow (bA1), red (rA2) and orange (rA3), but identical segments that originate from different parents are changed to the same colour to emphasize the similarities. If segments from two parents differ by only one mutation, they are also coloured the same but with the addition of the letter "m". A361 differ from its parent with two amino acids in one segment, thus shown as "mm" in this segment. The maximum number of amino acids that differ among the enzymes in a specific segment is presented. The enzyme names are colour coded according to the specific activity of the enzymes with azathioprine (AZA) as green ("high activity"), yellow ("low activity") or red ("almost no activity").

3.3 Screening improvements

MCB correlation

There was no clear correlation between azathioprine and MCB activity ($R^2 = 0.02$), see fig 3.10. The results in fig 3.11 show that the azathioprine active enzymes rather were a subgroup of the MCB active enzymes. The alpha class parents hA1-1, hA3-3, bA1-1 and rA2-2 all have high activity with MCB, but not the parents hA2-2 and rA3-3.

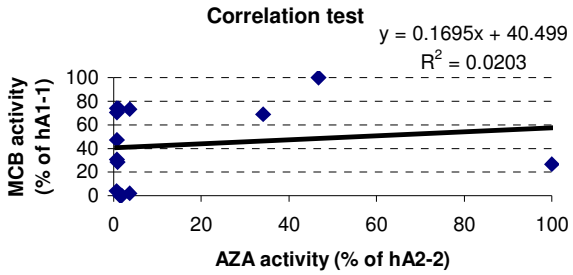


Fig 3.10. Correlation test between azathioprine (AZA) and MCB specific activity.

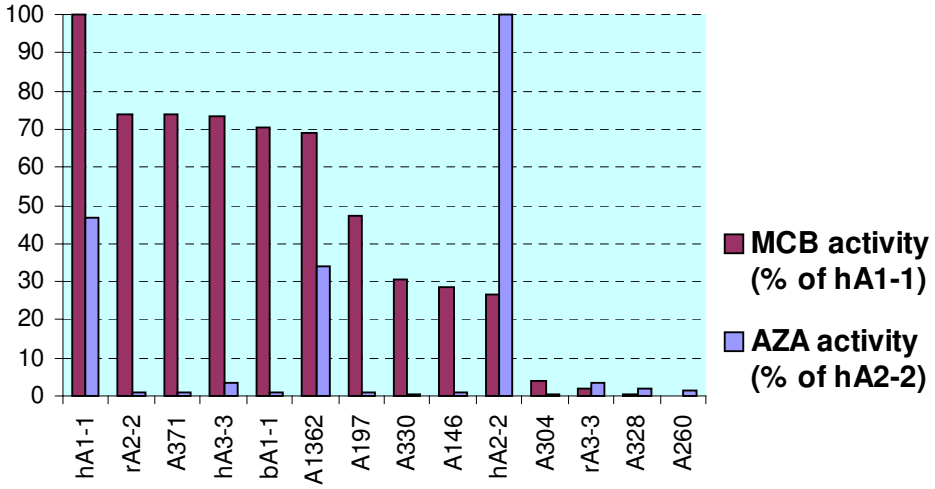


Fig 3.11. Specific activity of azathioprine (AZA) and MCB for selected enzymes.

Azathioprine-agar plates

There were three tests performed concerning the growth inhibiting effects of azathioprine on *E. coli* with GSTs. The result from the first test showed that an azathioprine concentration of 0.8 mM with 1 % DMSO inhibited the growth of both clones (A260 and hA2). There was still growth but not as much as in the other conditions. The clone with the inactive enzyme, A260, had grown slightly more than hA2 but it is not necessarily an effect of the enzymatic activity. The samples with 0 % DMSO, or 1 % DMSO with 0.08, 0.008 or 0.0008 mM azathioprine, had all grown much.

The second test showed that there is weaker growth on azathioprine-agar plates than on normal plates. This time, the bacteria expressing active enzyme (A1362) had grown more than those expressing A260.

The third test compared growth on agar plates with or without azathioprine. The two clones were grown to the same OD_{600} but the derivative was not controlled. When spread on plates, the dilution 1:10000 generated plates where the colonies could be counted, see table 3.3. For the inactive enzyme, A260, 10 µl with no dilution contained $10000*90 = 0.9*10^6$ bacteria, while for hA2 there were $1.4*10^6$ bacteria. The ratio of survival for the A260 clones in azathioprine was 4/90 = about 4 %, or if counting at the 100x dilution, the ratio would be (180/100) / 90 = about 2 %. For hA2 the survival ratio was 127/143 = about 89 %.

Table 3.3. Number of colonies seen in plates for different conditions, when using 10 µl with a 10000x dilution. One replicate per experiment condition.

Enzyme	A260	A260	hA2	hA2
AZA in plate	0.8 mM	-	0.8 mM	-
Colonies	4	90	127	143
Colony size	small	large	small	large
Survival ratio	4 %		89 %	

3.4 Comparative modeling

The structure of hA2-2 was modeled from coordinates of hA1-1. Of the eleven amino acid differences, eight are spread out as single amino acid changes, while three are located together in a loop between proline 110 and proline 114, flanking one wall of the H-site (see fig 3.12). Both prolines define the ends of one helix each and the ends of the loop, and this loop was rebuilt during the modeling.

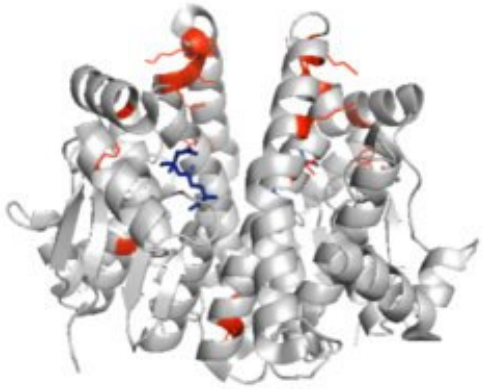


Fig 3.12. The eleven residues differing between hA1 and hA2 are shown in red in the structure of hA1-1 (PDB entry 1PKW). Bound glutathione is shown as blue. The loops that differ the most between the enzymes are located at the top. Image generated by PyMOL (version 0.98, DeLano Scientific LLC).

4. Discussion

4.1 Screening and characterization

Screening

One issue with the activity values obtained in the screening was that they were from measurements on lysates and not pure enzymes. For example the values of the positive control differ about 2-fold during measurement on 29 lysates with hA2-2 from 13 separate plate preparations. With measurement on pure enzymes one can easier distinguish the chimeras from the parents. 96-well plates with IMAC gel or affinity membranes could be utilized for purification during screening. However, the amount of pure enzyme obtained may be lower than that needed for measurements on several substrates. One could also note that the low screening values had an average of 7.44 $10^3 dA_{320}$/min and seemed to belong to a normal distribution. This value is almost the same as that obtained without enzyme. However, this average of 1008 measurements with lysates with very low activity might be an even better indicator of the non-enzymatic reaction rate.

Sequencing

The interpretations of the chromatograms were sometimes ambiguous, especially with DNA from preparations with a smaller cultivation volume, where the DNA concentration was a bit too low or the DNA perhaps was contaminated from overloading of the column. When two chromatograms from the same gene or DNA preparation showed a difference, for example one showing a parental sequence and the other a new mutant sequence, then the most probable interpretation was that it in fact was the parental sequence. This idea was confirmed when some sequences that first were interpreted as mutants were seen to have a completely parental sequence, while also the purified enzymes showed activity profiles similar to that of the parent. One region which often showed differences was the double stop codon, TAATAA, placed at the end of the coding region. Because of the proximity to the reverse primer the peaks in the chromatograms in this region were often very narrow and thus two peaks of the same kind were often joined into one, resulting in an apparent TATAA instead of TAATAA, which were often easier to see in the forward primer reaction chromatogram.

Purification

The purity of the selected enzymes were determined by SDS-PAGE to be satisfactory, regarding the analysis which focuses on the more active versus the less active enzymes. However, the denaturation during SDS-PAGE should be made more efficiently. The purifications that were successful with 50 mM imidazole in the binding buffer, resulted in a pure elute. But when lower imidazole concentrations must be used in order to get the protein to bind to the column, elution should be made with a gradient or in steps, to ensure a pure product.

Steady state kinetics

The enzyme specificities k_{cat}/K_M were obtained from non-linear regression. These values were about 50 % higher than if k_{cat}/K_M would be estimated from the slope of the initial rate at low substrate concentrations. If values were determined that way they would be similar to the values reported previously for hA2-2 and hA1-1 using that kind of estimation (Eklund *et al*., 2006).

4.2 Library and chimera analysis

Distribution of H-site residues on enzymes with different activity

The ten H-site residues that differed between the enzymes were located in four segments that remained unchanged during the gene shuffling. Each segment had 2-6 different amino acid compositions and the distribution of enzymes, with different activity, on these segments can be viewed in fig 4.1. A consensus sequence for active enzymes could be found. At H-site residues 208+213+216 the sequence M+L+S/A represented the active enzymes, see fig 4.2. The other segments did not show any clear relationships. At positions 107+108+110+111, the residues L+L+P+F/V represented three of the most active enzymes, but active enzymes could also have L+H+P+Y or M+H+P+L. This analysis presented one segment as the determining factor for azathioprine activity, but since all four segments correspond to the walls of the substrate binding pocket they all probably affect the activity. The GST genes of *bA2* and *hA5* also carry the M+L+A residues in the H-site positions 208+213+216, making it interesting to see if their corresponding enzymes also have activity with azathioprine. Most of the *hA4*-like alpha class GSTs on the other hand have completely different H-site residues in this segment.

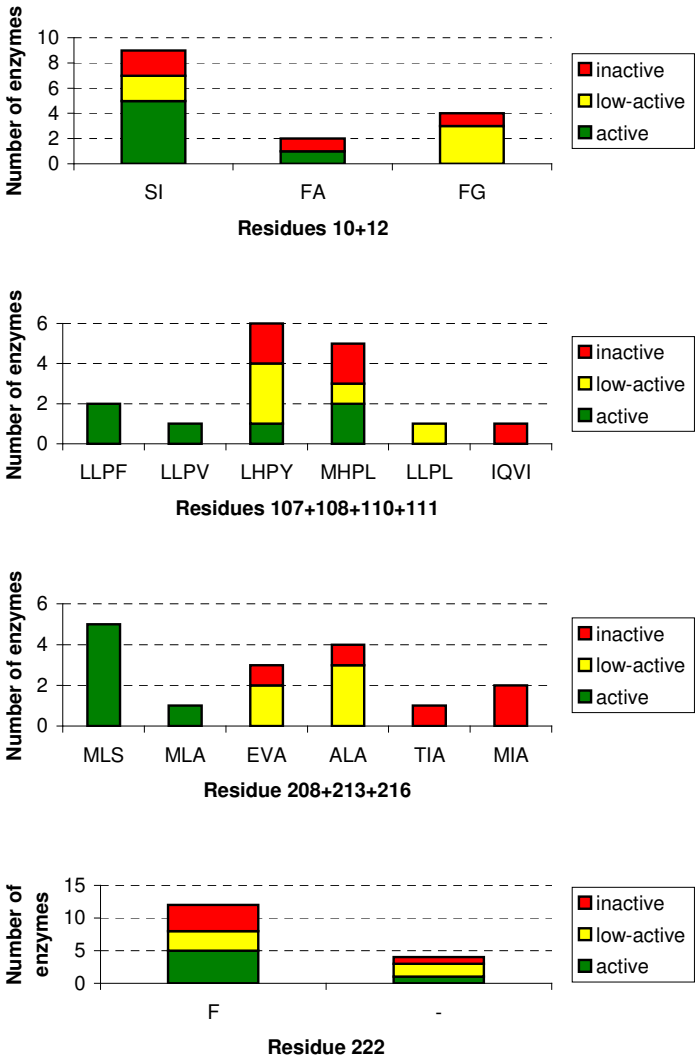


Fig 4.1. Distribution of enzymes, with different specific activities, on the four H-site residue segments. H-site positions 14 and 104 are not included since they are conserved among all examined enzymes.

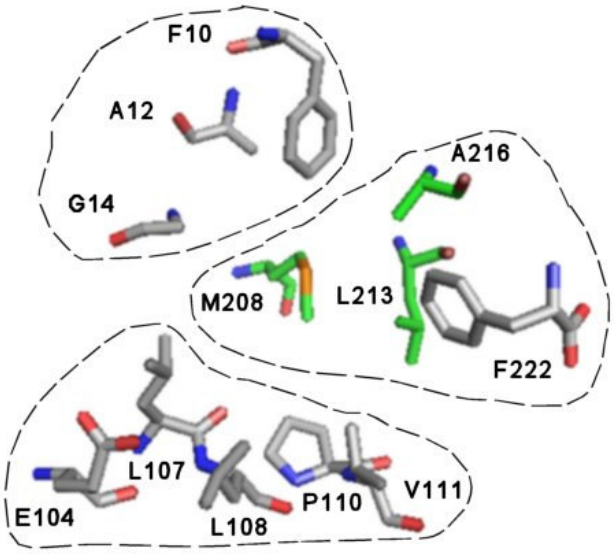


Fig 4.2. The H-site consists of twelve residues forming three walls (residues 10-14, 104-111 and 208-222, surrounded by dashed lines) of the substrate binding pocket. The H-site residues are viewed from the direction of the glutathione. Positions 208, 213 and 216 (coloured green) belonged to the same shuffled segment and this segment was identified as a structural determinant for azathioprine activity, with the active enzymes having amino acids M+L+S/A in these three H-site positions. Image generated by PyMOL (version 0.98, DeLano Scientific LLC) using the structure of hA1-1 (PDB entry 1PKW).

Whole sequence comparison

hA2 dominated the origin of enzymes with high activity towards azathioprine. Variations from the other parents were included in these enzymes as well and each chimera consisted of parts from on average three different parents. If the concentration of the active parent *hA2* had been larger during the gene shuffling there might have been a larger proportion of active chimeras, but there should have been more completely parental *hA2* found in the screening as well.

Analysis of variable regions

The sequences of the chimeras and alpha class library parents could be divided into nine main variable regions separated by conserved regions, if excluding a few single sites with variation. Fig 4.3 shows the number of different amino acids at each position in a multiple alignment of 32 known and predicted alpha class genes, namely those seen in fig 1.3 as well as two *Danio rerio* (zebrafish) GSTs. These nine variable regions correlated to the regions where the general alpha class genes were non-conserved, but two regions with variation in the general alpha class, at about position 65-80 and around position 145, were conserved within the alpha class library parents.

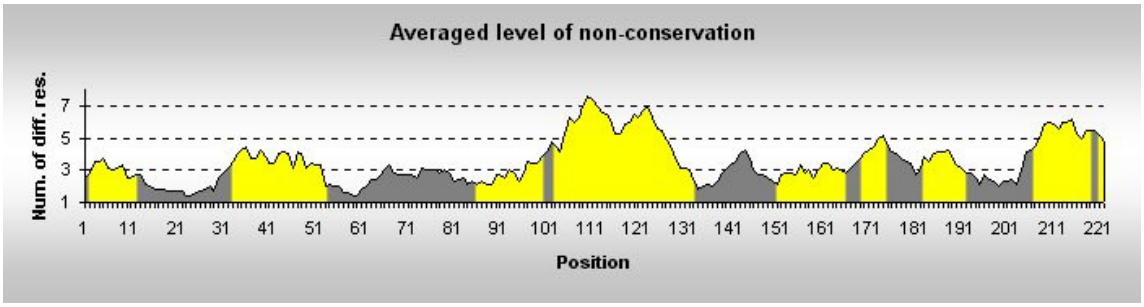


Fig 4.3. Comparison between the variation among several natural alpha class genes and among the chimeras and parents of the alpha class library. The level of non-conservation is the number of different residues available at each position in a multiple alignment of several known and predicted alpha class genes. The

averaged value for each position is then generated as an average of the level of non-conservation of that position and the six closest neighbours. In this way, a smoother curve is produced, highlighting the overall tendencies instead of the values of single positions. Highly conserved regions approach the value 1 while regions with more variation have a higher value. The nine main variable regions identified among the chimeras and parents of the alpha class library are shown in yellow, while the gray regions are conserved among them.

Several conclusions could be drawn from the simplified representation of the sequences in fig 3.9. One should remember that single point mutations that definitely could affect the activity were ignored when comparing only the large variable regions. However, only one of those point mutations was located in a H-site residue, namely S216A which differed *hA1* from *hA2*. If comparing the activities and sequences of hA1 and hA2, the importance of position 216 should absolutely be considered. For the other point mutations, their contribution to the activity was in this case neglected. When genes were very similar, a single of these nine variable segments could be responsible for the difference in activity between the enzymes. This was the case with the pairs A470 and rA3 as well as hA2 and A1347, see fig 4.4.

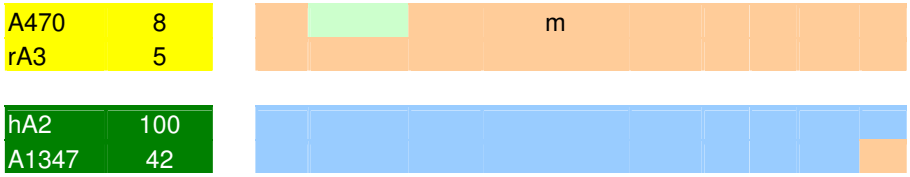


Fig 4.4. Two pairs of genes differing in only one of the nine variable segments. These images are extracted from fig 3.9 and can there be viewed in full context. Segment 2 separated A470 and rA3 and gave a two-fold difference in activity, while segment 9 separated hA2 and A1347 with a two-fold activity difference. The numbers represent the specific activities with azathioprine as percent of hA2-2. The effect of point mutations ("m") was in this case neglected.

Sometimes segments from the most active enzyme, hA2-2, were exchanged with segments from other parents, and these changes could give clues to the contribution to the high activity in hA2-2 from the different segments. When chimeras with a relatively higher activity had *hA2*-like segments replaced, when compared to the chimeras with relatively lower activity, segments that possibly could improve the activity of hA2-2 were identified. Four such segments, number 1, 3, 6 and 7, were found, see fig 4.5.

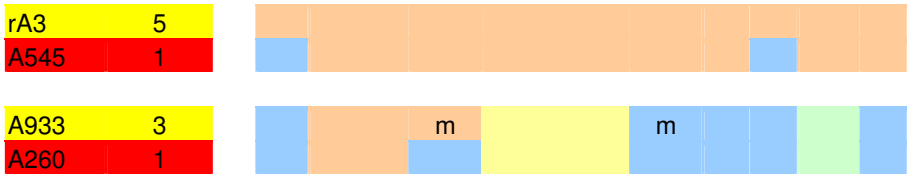


Fig 4.5. Several cases were found where hA2-originating segments resulted in lower activity when compared to enzymes with another origin of that segment. The numbers represent the specific activities with azathioprine as percent of hA2-2. These images are extracted from fig 3.9 and can there be viewed in full context.

When comparing all the sequences and the corresponding activities of the enzymes, as in fig 3.9, segment 8 looked like a determining factor of azathioprine activity. The six most active enzymes all had the *hA2*-originating segment or the similar *hA1*-originating segment there, while none of the others had any of them. Segments 1, 5, 7 and 9 had almost only *hA2*-originating segments among the six most active enzymes, but the *hA2*-segment was also present within the less active enzymes. Segments 2, 3, 4 and 6 had origins other than *hA2* and *hA1* among the six most active enzymes. Segment 8 (positions 207-219) and segment 9 make up the C-terminal helix and also one side of the H-site (see fig 4.6). Exchange of this helix

together with four other residues between *hA1* and *hA4* has earlier resulted in a change in activity profile from one enzyme to the other (Nilsson *et al*., 2000). The finding of segment 8 being important for the activity with azathioprine was also in agreement with the results from the H-site residue analysis.

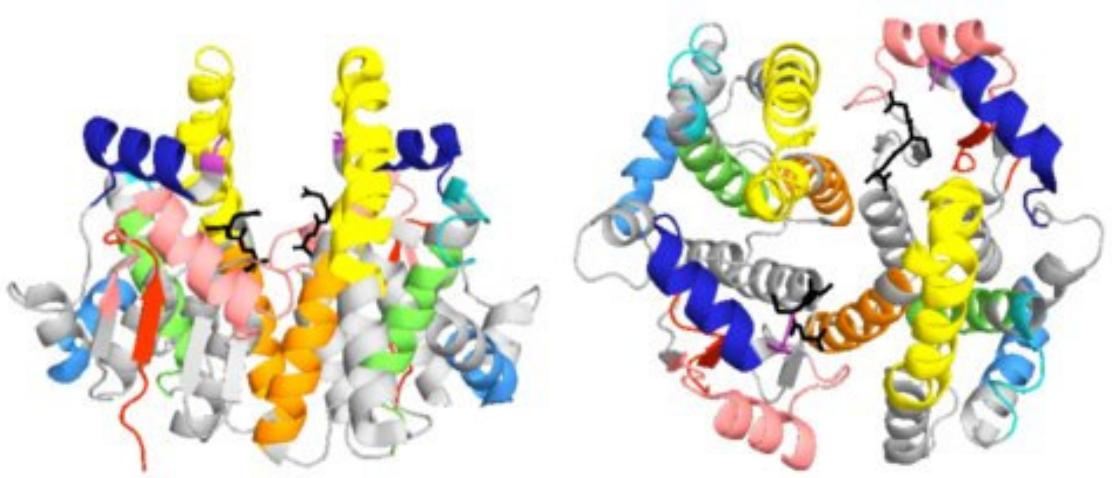


Fig 4.6. View of structure of hA1-1 (PDB entry 1PKW) from two angles. The nine main variable segments among the alpha class parents are coloured differently (1 = red, 2 = light red, 3 = orange, 4 = yellow, 5 = green, 6 = cyan, 7 = light blue, 8 = dark blue, 9 = purple). The H-site residues reside in segment 1 (red), 4 (yellow), 8 (dark blue) and 9 (purple), close to the glutathione (black). Image generated by PyMOL (version 0.98, DeLano Scientific LLC).

With knowledge of this potential azathioprine activity determining segment, one could progress by confirming this notion by replacing segment 8 in other enzymes. For subsequent shuffling generations one could either include only genes with segment 8 from *hA2* or at least over-represent it, or focus on improving this segment 8. For example the three H-site residues in segment 8 could be targeted for random mutagenesis resulting in a smaller library size than the *hA1*-library used in this project.

hA1-library

The *hA1*-library was included in the shuffling of the alpha class library, which otherwise consisted of *hA2*, *hA3*, *bA1*, *rA2* and *rA3*. None of the enzymes found in the screening of the alpha class library contained any segments of mutated *hA1* that could originate from the *hA1*-library. It was also of interest to screen the *hA1*-library on its own, especially since it is based on hA1-1, which has activity with azathioprine. The screening of the *hA1*-library generated only one positive hit, which later was found to be the parent hA1. This library had random codons (*NNN*) in ten H-site positions and if the codon distribution were completely random, the number of enzymes with a certain number of mutations and the number of enzymes with an early stop codon could be counted, see fig 4.7. Most mutants probably had many mutations in these important H-site residues and this variation made the library more suitable for selection methods than screening methods. Phage display has earlier been used to find functional mutants in this library (Hansson *et al*., 1997). Site-specific mutant libraries suitable for screening would rather have mutations in a few positions and perhaps not all codons in every position.

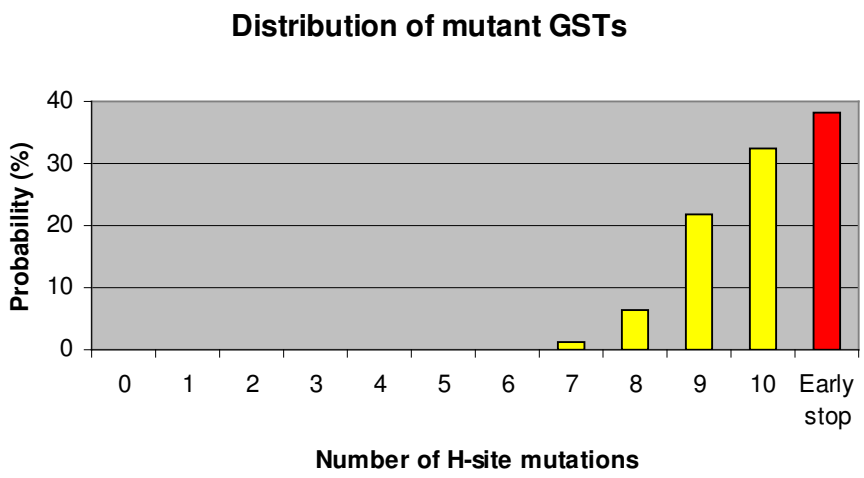


Fig 4.7. Distribution of mutants of the hA1-library with different number of mutations. The red bar represents the mutants with an early stop codon.

4.3 Screening improvements

The activity measurements with MCB showed that the activity profiles were different for the substrates MCB and azathioprine. However, one could also see that a MCB pre-screening that could identify functional GSTs probably would find azathioprine active enzymes as well. But the chimeras identified would probably to a large degree be based on the parents *hA1*, *hA3*, *bA1* and *rA2* and thus include many azathioprine negative chimeras and possibly also loose the *hA2*-like chimeras.

The tests of the growth inhibiting effects of azathioprine showed that it indeed inhibits the growth of *E. coli* with GSTs, at a concentration of 0.8 mM with 1 % DMSO, both in solution and in azathioprine-agar plates. However, it was not concluded how the enzymes contribute to this effect. The different clones certainly had a different survival rate in azathioprine, but it could also be an effect of different viability and phase status at the start of the experiment. The viability and status of the bacteria used should be the same when starting the experiments, demanding control of both OD_{600} and its derivative. Expression was not controlled and codon distribution in the gene may have affected the expression and thereby the growth rate. IPTG should perhaps be used. One could try separating the effects. Azathioprine gives rise to substituted nucleotide insertion, glutathione depletion and an imidazole moiety. With azathioprine and N-acetylcysteine, a glutathione precursor, the glutathione depletion can be removed. With 6-MP, the glutathione depletion and the imidazole moiety are removed, imitating a very efficient enzyme.

4.4 Comparative modeling

The structure of hA2-2 produced by comparative modeling gives clues to the differences between that enzyme and hA1-1. There are only eleven amino acid differences between the two enzymes. The R89K mutation, with hA1-1 used as reference, is located in the region where the two subunits connect but the residues are similar. Three mutations are in the H-site, A12I, V111F and A216S, and all of them have larger residues in hA2-2 than hA1-1. The whole loop between P110 and P114 is exchanged, which might slightly distort the positions of the adjoining helices and this could affect the active site since the end of the loop is located in the H-site. Investigations of more chimeras as well as structure determination of hA2-2 with azathioprine bound to the active site will provide more knowledge about the effect of

differences between the enzymes. A crystal structure would also provide validation for comparative models and improve the models in the form of a new template structure. Up to date there are only crystal structures of hA1-1 with or without different ligands and mutated variants of rA1-1 (Adman *et al*., 2001), see fig 4.8, and *rA1* is very similar to the parental gene *rA2*. Generally, comparative models should be able to be produced with quite good quality since there are no differences in the amino acid sequence that are considered to be insertions or deletions among the alpha class parental enzymes or the chimeras analyzed so far. If a crystal structure will not be available, it would be interesting to do ligand docking *in silico* to analyze which residues interact with azathioprine.

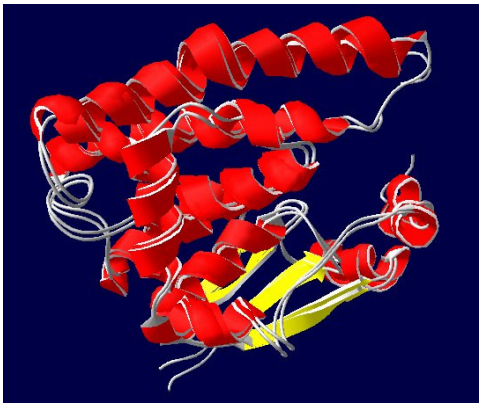


Fig 4.8. Structural similarity within the GST alpha class. Superimposition of hA1-1 (PDB entry 1PKW, chain A) and rA1-1 (PDB entry 1EV4, chain A). Image generated using DeepView/Swiss-PdbViewer 3.7.

5. Conclusions

- This study involved the screening of a broad GST alpha class library for azathioprine activity. GSTs with improved activity towards azathioprine could be used as negative selectable markers, for example for removing inserted cells after a successful temporal gene therapy.

- The sequence identity among the five parents ranged from 77 % to 91 % with only one having high activity with azathioprine. Included in this first library was also a second library, based on *hA1*, having random mutations in ten H-site positions.

- Screening of the library was performed by spectroscopic measurement of the lysate at 320 nm during the reaction with azathioprine and a high-throughput method using 96-well plates for cultivation was optimized.

- Screening of 400 clones from the *hA1*-library identified no azathioprine active GSTs, probably because of too much variation. This library would rather require selection methods.

- After screening of 1570 clones from the broad alpha class library, 10 chimeras were sequenced, purified by IMAC and kinetically characterized. The examined chimeras had on average 3 parents, 3 crossovers and less than one point mutation and their activities with azathioprine and MCB, another substrate for GSTs, were not correlated.

- The six most active enzymes towards azathioprine had a specific activity of more than 20 % of the hA2-2 activity and a sequence identity of 85-99 %. Of these, the four chimeras originated mostly from *hA2* but with all parents represented. Thus, the first generation contained GSTs with activity and variation enough to continue with a subsequent generation.

- The segment of amino acids 207-219 seemed to be a determining factor of azathioprine activity and forms the most C-terminal alpha helix and one wall of the substrate binding pocket, including the H-site residues in position 208, 213 and 216. By keeping this region constant, library sizes in later steps of directed evolution can be decreased while increasing the fraction of active GSTs. Alternatively these key residues can be targeted for mutation.

Acknowledgements

I especially want to thank my supervisors Bengt Mannervik and Sanela Kurtovic for all help and for sharing this project with me. I also want to thank Birgitta Eklund, Abeer Shokeer and Ronnie Jansson for help with the project, Birgit Olin and Françoise Raffalli-Mathieu for purified parents and all in our group and department for support.

References

Adman ET, Le Trong I, Stenkamp RE, Nieslanik BS, Dietze EC, Tai G, Ibarra C, Atkins WM. Localization of the C-terminus of rat glutathione S-transferase A1-1: crystal structure of mutants W21F and W21F/F220Y. *Proteins*. 2001; 42:192-200.

Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS. The 'evolvability' of promiscuous protein functions. *Nat Genet*. 2005; 37:73-76.

Choi IG, Kim SH. Evolution of protein structural classes and protein sequence families. *Proc Natl Acad Sci U S A*. 2006; 103:14056-14061.

Coles BF, Anderson KE, Doerge DR, Churchwell MI, Lang NP, Kadlubar FF. Quantitative analysis of interindividual variation of glutathione S-transferase expression in human pancreas and the ambiguity of correlating genotype with phenotype. *Cancer Res*. 2000; 60:573-579.

Conant GC, Wagner A. The rarity of gene shuffling in conserved genes. *Genome Biol*. 2005; 6:R50.

Crawford DJ, Maddocks JL, Jones DN, Szawlowski P. Rational design of novel immunosuppressive drugs: analogues of azathioprine lacking the 6-mercaptopurine substituent retain or have enhanced immunosuppressive effects. *J Med Chem*. 1996; 39:2690-2695.

Deponte M, Becker K. Glutathione S-transferase from malarial parasites: structural and functional aspects. *Methods Enzymol*. 2005; 401:241-253.

Dinkova-Kostova AT, Fahey JW, Wade KL, Jenkins SN, Shapiro TA, Fuchs EJ, Kerns ML, Talalay P. Induction of the phase 2 response in mouse and human skin by sulforaphane-containing broccoli sprout extracts. *Cancer Epidemiol Biomarkers Prev*. 2007; 16:847-851.

Eklund BI, Edalat M, Stenberg G, Mannervik B. Screening for recombinant glutathione transferases active with monochlorobimane. *Anal Biochem*. 2002; 309:102-108.

Eklund BI, Moberg M, Bergquist J, Mannervik B. Divergent activities of human glutathione transferases in the bioactivation of azathioprine. *Mol Pharmacol*. 2006; 70:747-754.

Elion GB. The purine path to chemotherapy. *Science*. 1989; 244:41-47.

Emrén LO, Kurtovic S, Runarsdottir A, Larsson AK, Mannervik B. Functionally diverging molecular quasi-species evolve by crossing two enzymes. *Proc Natl Acad Sci U S A*. 2006; 103:10866-10870.

Fan K, Wang W. What is the minimum number of letters required to fold a protein? *J Mol Biol*. 2003; 328:921-926.

Felsenstein, J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*. 1989; 5:164-166.

Grahn E, Novotny M, Jakobsson E, Gustafsson A, Grehn L, Olin B, Madsen D, Wahlberg M, Mannervik B, Kleywegt GJ. New crystal structures of human glutathione transferase A1-1 shed light on glutathione binding and the conformation of the C-terminal helix. *Acta Crystallogr D Biol Crystallogr*. 2006; 62:197-207.

Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997; 18:2714-2723.

Hansson LO, Widersten M, Mannervik B. Mechanism-based phage display selection of active-site mutants of human glutathione transferase A1-1 catalyzing S_NAr reactions. *Biochemistry*. 1997; 36:11252-11260.

Hartman MC, Josephson K, Szostak JW. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci U S A*. 2006; 103:4356-4361.

Hegazy UM, Hellman U, Mannervik B. Replacement surgery with unnatural amino acids in the lock-and-key joint of glutathione transferase subunits. *Chem Biol*. 2006; 13:929-936.

Hiraga K, Arnold FH. General method for sequence-independent site-directed chimeragenesis. *J Mol Biol*. 2003; 330:287-296.

Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, Meidl P, Ouverdin B, Parker A, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Severin J, Slater G, Smedley D, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wood M, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Flicek P, Kasprzyk A, Proctor G, Searle S, Smith J, Ureta-Vidal A, Birney E. Ensembl 2007. *Nucleic Acids Res*. 2007; 35:D610-617.

Ivarsson Y, Mackey AJ, Edalat M, Pearson WR, Mannervik B. Identification of residues in glutathione transferase capable of driving functional diversification in evolution. A novel approach to protein redesign. *J Biol Chem*. 2003; 278:8733-8738.

Klipcan L, Safro M. Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases. *J Theor Biol*. 2004; 228:389-396.

Maheshri N, Schaffer DV. Computational and experimental analysis of DNA shuffling. *Proc Natl Acad Sci U S A*. 2003; 100:3071-3076.

Mannervik B, Board PG, Hayes JD, Listowsky I, Pearson WR. Nomenclature for mammalian soluble glutathione transferases. *Methods Enzymol*. 2005; 401:1-8.

Matsumura I, Ellington AD. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J Mol Biol*. 2001; 305:331-339.

Menor C, Fernandez-Moreno MD, Fueyo JA, Escribano O, Olleros T, Arriaza E, Cara C, Lorusso M, Di Paola M, Roman ID, Guijarro LG. Azathioprine acts upon rat hepatocyte mitochondria and stress-activated protein kinases leading to necrosis: protective role of N-acetyl-L-cysteine. *J Pharmacol Exp Ther*. 2004; 311:668-676.

Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, Wang ZG, Arnold FH. Library analysis of SCHEMA-guided protein recombination. *Protein Sci*. 2003; 12:1686-1693.

Moeslinger T, Friedl R, Spieckermann PG. Inhibition of inducible nitric oxide synthesis by azathioprine in a macrophage cell line. *Life Sci*. 2006; 79:374-381.

Moore GL, Maranas CD, Lutz S, Benkovic SJ. Predicting crossover generation in DNA shuffling. *Proc Natl Acad Sci U S A*. 2001; 98:3226-3231.

Murphy LR, Wallqvist A, Levy RM. Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng*. 2000; 13:149-152.

Nilsson LO, Gustafsson A, Mannervik B. Redesign of substrate-selectivity determining modules of glutathione transferase A1-1 installs high catalytic efficiency with toxic alkenal products of lipid peroxidation. *Proc Natl Acad Sci U S A*. 2000; 97:9408-9412.

Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS. Design and evolution of new catalytic activity with an existing protein scaffold. *Science*. 2006; 311:535-538.

Patel AA, Swerlick RA, McCall CO. Azathioprine in dermatology: the past, the present, and the future. *J Am Acad Dermatol*. 2006; 55:369-389.

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

Peitsch MC. Protein modeling by E-mail. *Bio/Technology*. 1995; 13:658-660.

Porath J, Carlsson J, Olsson I, Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*. 1975; 258:598-599.

Raffalli-Mathieu F, Persson D, Mannervik B. Differences between bovine and human steroid double-bond isomerase activities of Alpha-class glutathione transferases selectively expressed in steroidogenic tissues. *Biochim Biophys Acta*. 2007; 1770:130-136.

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000; 16:276-277.

Riddle DS, Santiago JV, Bray-Hall ST, Doshi N, Grantcharova VP, Yi Q, Baker D. Functional rapidly folding proteins from simplified amino acid sequences. *Nat Struct Biol*. 1997; 4:805-809.

Roberts RW, Szostak JW. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A*. 1997; 94:12297-12302.

Rothman SC, Kirsch JF. How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *J Mol Biol*. 2003; 327:593-608.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4:406-425.

Saraf MC, Moore GL, Maranas CD. Using multiple sequence correlation analysis to characterize functionally important protein regions. Protein Eng. 2003; 16:397-406.

Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*. 2003; 31:3381-3385.

Scott JK, Smith GP. Searching for peptide ligands with an epitope library. *Science*. 1990; 249:386-390.

Shah V, Baldrian P, Eichlerova I, Dave R, Madamwar D, Nerud F, Gross R. Influence of dimethyl sulfoxide on extracellular enzyme production by Pleurotus ostreatus. *Biotechnol Lett*. 2006; 28:651-655.

Stemmer WP. Rapid evolution of a protein in vitro by DNA shuffling. *Nature*. 1994; 370:389-391.

Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol*. 2003; 10:59-69.

Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol*. 1998; 16:652-656.

Widersten M, Mannervik B. Glutathione transferases with novel active sites isolated by phage display from a library of random mutants. *J Mol Biol*. 1995; 250:115-122.

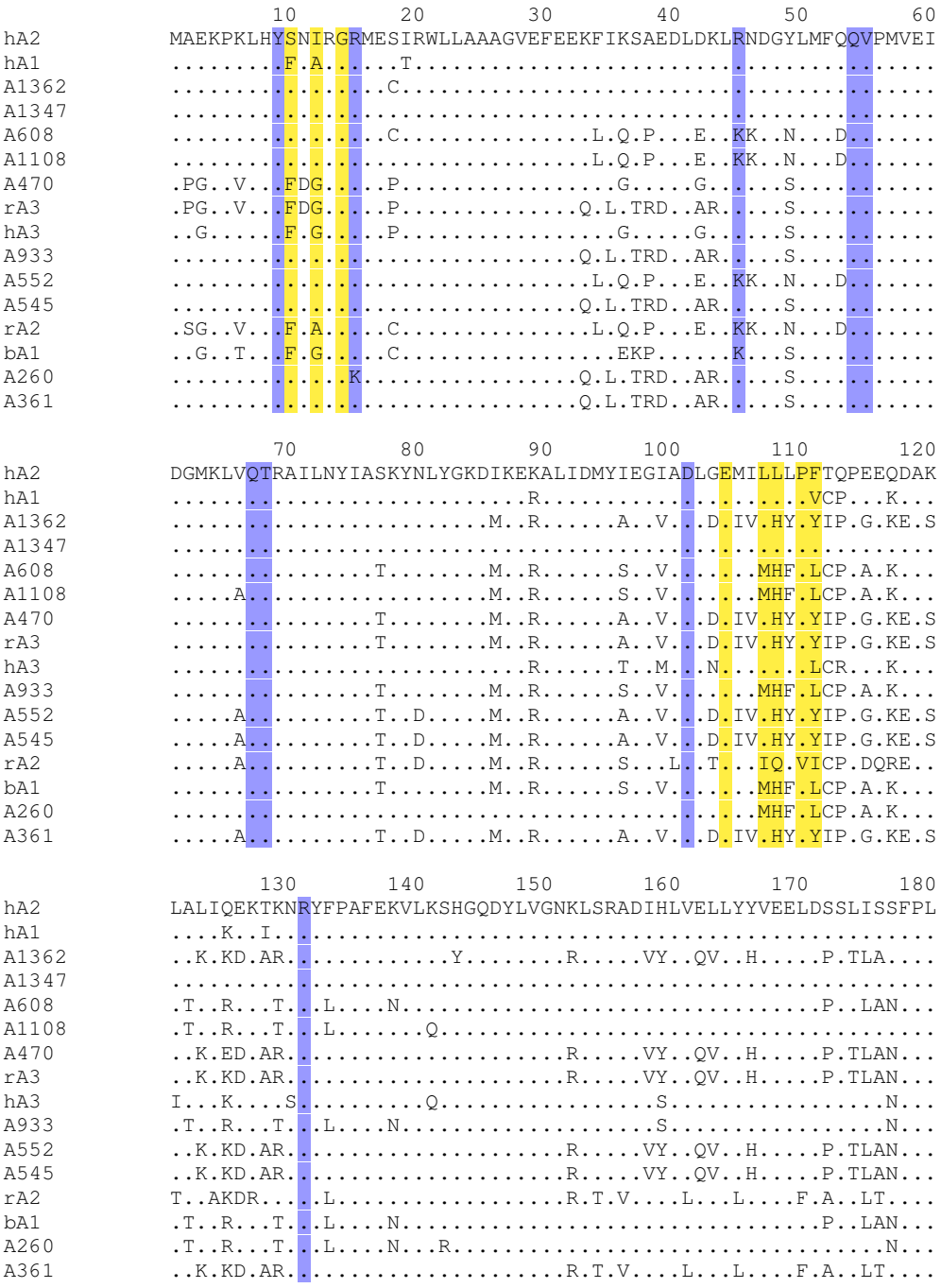
Williams RT. Comparative patterns of drug metabolism. *Fed Proc*. 1967; 26:1029-1039.

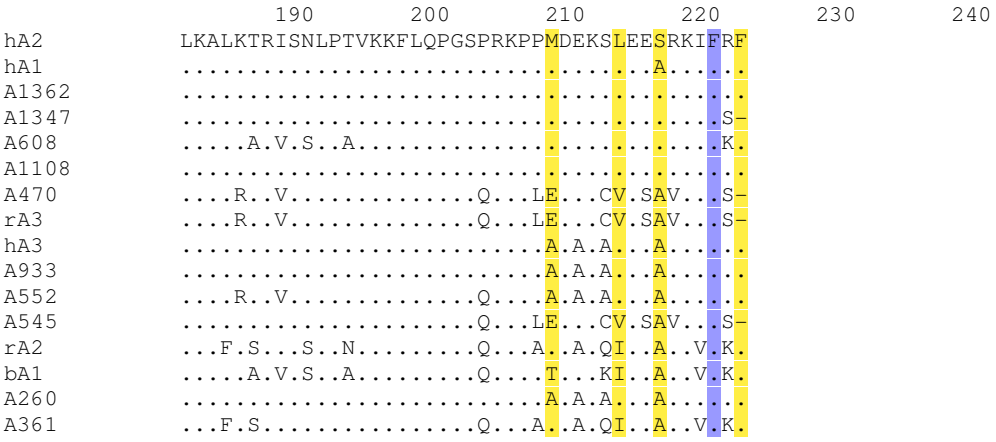
Zhang XH, Chasin LA. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci U S A*. 2006; 103:13427-13432.

Appendix

A. Sequences

Table A.1. Amino acid sequences of six parents and ten chimeras (6xHis-tags excluded). G-site residues are coloured blue and H-site residues yellow.




190 200 210 220 230 240
hA2 LKALKTRISNLPTVKKFLQPGSPRKPPMDEKSLEESRKIFRF
hA1 ..... A.....
A1362 .....
A1347 ..... S-
A608 .....A.V.S..A..... K.
A1108 .....
A470 ....R..V.....Q...LE...CV.SAV...S-
rA3 ....R..V.....Q...LE...CV.SAV...S-
hA3 .....A.A.A...A.....
A933 .....A.A.A...A.....
A552 ....R..V.....Q....A.A.A...A.....
A545 .....Q...LE...CV.SAV...S-
rA2 ...F.S...S..N.....Q...A..A.QI..A..V.K.
bA1 .....A.V.S..A.....Q....T...KI..A..V.K.
A260 .....A.A.A...A.....
A361 ...F.S.....Q...A..A.QI..A..V.K.