UNIVERSITY OF LONDON THESIS

Degree phd Year 2007 Name of Author TARIK SENUSSI

Directed Evolution to Modify the Substrate Specificity of Transketolase, a Carbon-Carbon Bond-Forming Enzyme

A Thesis Submitted for the Degree of
Doctor of Philosophy
to the
University of London

Tarik O.Senussi

Department of Biochemical Engineering
University College London

2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I, Tarik Omar Senussi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank everyone who assisted me during the course of my PhD research. In particular would like to thanks my supervisor, Dr. Paul Dalby and my advisor Dr. Gary Lye. Without their valuable input, support and feedback this work would not have been possible.

I would also like to thank member of the Bioconversion-Chemistry-Engineering Interface program (BiCE), department of chemistry (particularly Dr Mark Smith) and fellow researchers' at foster court (particularly Sean Costelloe and Dr. Edward Hibbert).

I would also like to thank my friends and family for their support and encouragement over the course of this PhD.

Thanks also to the funding from the UK Biotechnology and Biological Sciences Research Council (BBSRC).

Abstract

Transketolase (TK) (E.C. 2.2.1.1) has significant potential as a biocatalyst in the production of pharmaceuticals and fine chemicals, catalysing the irreversible and stereospecific transfer of a C2 (1,2-dihydroxyethyl) moiety from the ketol donor substrate β-hydroxypyruvate (β-HPA) to a wide range of aldehyde acceptor substrates, such as glycoaldehyde (GA). Commercial application of TK is restricted by the limited availability and expense of β-HPA as well as its limited activity towards novel substrates. This project describes efforts to generate and identify variants of *E. coli* TK capable of accepting novel ketol donors and/or aldehyde acceptors. Variants were prepared by saturation site directed mutagenesis (SSDM), and characterised, using novel high-throughput HPLC and TLC screens. The model β-HPA and GA reaction, as well as a range of novel acceptors, such as propionaldehyde, benzaldehyde and hydroxybenzaldehyde were examined.

HPLC assays were also developed for all aldehyde substrates, for the detailed analysis of enzyme variants identified from libraries by rapid HPLC or TLC. During such analysis a variety of buffers were tested for suitability in novel screens. Mops and Hepes were both found to be capable of substrate conversion in the absence of transketolase, hence the discovery of the first ever mimetic reaction for transketolase.

Two techniques were used to identify residues to target for random mutagenesis: phylogenetic library design (10 sites determined by S. Costelloe); and structural library design (10 site library generated and screened in collaboration with E. Hibbert). HPLC and TLC analysis identified variants

with improved reaction rates towards the model HPA and GA reaction (A29E, A29D, I189Y and 461S), and also towards propionaldehyde (H26A, H26T, H26K, R358I, H461S, D469S, and D469T), but failed to definitively identify a mutant capable of activity towards benzaldehyde.

Abbreviations

ACN	acetonitrile solution (HPLC grade)
A_n	absorbance at *n*nm
Amp^+	denotes the inclusion of 150mg.l^{-1} ampicillin
amp^R	ampicillin resistance gene
AU	absorbance unit
BAL	benzaldehyde
BKD	benzaldehyde keto-diol product
bp	base-pair
BrPA	bromopyruvate
CASTing	combinatorial active-site saturation
CLERY	combinatorial libraries enhanced by recombination in yeast
DHETPP	dihydroxyethylthiamine pyrophosphate
DHFR	dihydrofolate reductase
DNA	deoxyribonucleic acid
DNAP	DNA polymerase
2,4-DNP	2,4-dinitrophenylhydrazene
dsDNA	double-stranded DNA
DTT	dithiothreitol
DXP	1-deoxy-D-xylulose-5-phosphate
EC	Enzyme Commission
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EpPCR	error-prone PCR
FACS	Fluorescence Activated Cell Sorting
FDA	Federal Drugs Agency
fepPCR	focused error-prone PCR
FLINT	fluorescent intensity assay
FlPA	fluoropyruvic acid
GAL	glycoaldehyde
GAP	glyceraldehyde-3-phosphate
Gly-Gly	glycine-glycine buffer
Hepes	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
β-HPA	β-hydroxypyruvate
HPLC	high performance liquid chromatography
HSPA	mercaptopyruvic acid
IGPS	indole-3-glycerol-phosphate synthase
IL-2	interleukin-2
ISPR	*in situ* product removal
ITCHY	incremental truncation from the creation of hybrid enzymes
kat	katal: 1 mole of product formed per second
kat.m^{-3}	katals per cubic metre
k_{cat}	catalytic constant (or turnover number)
K_d	dissociation constant
K_m	Michaelis constant
LB	Luria Bertani
LFER	linear free energy relationship

MEGAWHOP	megaprimer PCR of whole plasmid
MePA	ketobutyric acid
Mops	3-(N-Morpholino)propanesulfonic acid
mpv	mutations per variant
M_r	relative molecular mass
MSDM	"multi" site-directed mutagenesis
MTEH	5-(2-methylthioethyl)-hydantoin
NADPH	nicotinamide adenine dinucleotide, reduced form
OD_n	optical density at *n*nm
ODU	optical density unit
OPPP	oxidative pentose phosphate pathway
PAL	propionaldehyde
PCR	polymerase chain reaction
PDA	protein data automation
PDB	RCSB Protein Data Bank
pKa	negative logarithm of the acid dissociation constant, Ka
PKD	propionaldehyde Keto-diol product
PMA	phosphomolybdic acid
PPP	pentose phosphate pathway
pQR791	unmodified 'WT' his-tagged DNA library template plasmid
Pyr	Pyruvate
RACHITT	random chimeragenesis on transient templates
rpm	revolutions per minute
R5P	ribulose-5-phosphate
RPR	random-priming recombination
scFv	antibody single-chain Fv fragment
SCOPE	structure-based combinatorial protein engineering
SDM	site-directed mutagenesis
SSDM	saturation site-directed mutagenesis
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELEX	systematic evolution of ligands by exponential enrichment
SHIPREC	sequence homology-independent protein recombination
StEP	staggered extension process
TAE	Tris·acetate EDTA
TBE	Tris·borate EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	trifluoroacetic acid
ThDP	thiamine diphosphate
Tkt	*Escherichia coli* transketolase gene
TKTL	transketolase like transcript gene
TLC	thin layer chromatography
T_m	melting temperature
TPP	thiamine pyrophosphate
Tris	tris(hydroxymethyl)aminomethane
UCL	University College London
UV	ultraviolet
V_{max}	maximal velocity
X-5-P	xylulose-5-phosphate

Table of contents

List of figures

List of tables

Chapter 1 – Introduction

1.1 Enzyme Structure

1.1.1 Introduction

Enzymes are proteins composed of amino acids linked by peptide bonds into a specific sequence. These peptide bonds are formed by a condensation reaction between the carboxy group of one amino acid with the amino group of another. There are twenty amino acids available for use in the standard repertoire of a cell. Subsequent post-translational modifications (glycosylation, methylation and phosphorylation among others), the addition of prosthetic groups (metal ions, heme groups etc) and binding of cofactors are also possible. All of these variables together provide enzymes with up to 200 chemical groups at their disposal (Brandon and Tooze, 1991).

Some functional proteins have been found to lack a folded structure under physiological conditions, however all known active enzymes have a well defined structure. The conformation that a specific protein adopts is dependent primarily on sequence and as a result of the interaction of the amino acids in the polypeptides present. The primary structure is equivalent to the amino acid sequence and is independent of spatial arrangement.

1.1.2 Secondary structure

Secondary structure refers to the arrangement of the polypeptide chain into relatively regular hydrogen-bonded structures, which map to specific regions of Ramachandran space. Two such structures are the α-helix and β-pleated sheet. An alpha-helix is a tight helix formed out of the polypeptide chain. The polypeptide main-chain makes up the central core structure, and the side chains extend out and away from the helix. The CO group of one amino acid (n) is hydrogen bonded to the NH group of the amino acid four residues away (n +4). In this way every CO and NH group of the backbone is

hydrogen bonded. α-helices are most commonly made up of hydrophillic amino acids, because hydrogen bonds are generally the strongest attraction possible between such amino acids. β-pleated strands are the other type of secondary structure.

1.1.3 Supersecondary structure

The next level of interaction the peptide uses to define its structure is known as supersecondary structures. Motifs, e.g. greek keys and α-helix bundles, are commonly observed topologies made up of multiple super-secondary structures in various combinations. Examples of the types of elements seen in supersecondary structure are the β-β unit, composed of two anti-parallel strands connected by a hairpin; the β-α-β unit, two parallel strands separated by an antiparallel alpha helix, with 2 hairpins connecting the three secondary structures and α-α hairpin. Domains must also be considered and these are independently folded units of structure that frequently have discrete biological function(s) in their own right (e.g. DNA-binding homeodomains). It has been proposed that new enzyme and protein functions often arise through the combination of different domains with their own individual function (Orengo and Thornton, 2005).

1.1.4 Tertiary and Quaternary structure

Tertiary structure can be defined as the three-dimensional configuration of the entire folded polypeptide chain. These result from interactions between residues, secondary structured regions, motifs and domains that occur after the folding process. Quaternary structure refers to the spatial arrangement of the subunits composed of different polypeptide chains (Branden and Tooze, 1991). As a result, quaternary structure is only observed in multiple subunit/polypeptide complexes.

1.2 Enzyme Engineering

It is possible to extensively modify the structure of enzymes, by manipulating the genes that encode them. A widely used example of this is side directed mutagenesis (SDM) which facilitates the mutation of a specific residue to another of choice (Farber and Petsko, 1990). The site-directed mutagenesis of enzymes has provided means to achieve a greater understanding of the mechanisms underlying the catalysis/binding and is essential if there is any hope to engineer and adapt them for specific and novel functions. There are also a variety of error prone-PCR methods which facilitate incorporation of random mutations into the sequence of choice; some control can be made (transitions and/or transversions) by changing bias. Gene shuffling can also be used in a homology dependent or independent fashion. However in spite of the myriad techniques which can be used to adapt existing enzymes, the changes introduced are often functionally deleterious, due to the complexity and marginal stability of enzymes, which in recent years directed evolution has had some success in overcoming.

1.2.1 Rational Design

This approach uses current scientific knowledge of, and reasoning about, a protein in order to design desired properties into a polypeptide sequence. At the centre of the design approach is the "design cycle," in which theory and experiment alternate. The starting point is the development of a molecular model, based on rules of protein structure and function or structural information, combined with an algorithm for applying these. This is followed by experimental construction and analysis of the properties of the designed protein. If the experimental outcome is failure or partial success, then a next iteration of the design cycle is started in which additional complexity is introduced, rules and parameters are refined, or the algorithms for applying them are modified. Rational protein design or engineering specific amino acid

substitutions can also be used to modify proteins. This technique is not very widely used since it can be laborious. The problems associated with rational design are that usually knowledge of the structure and the relationship between its sequences is necessary. In recent years NMR spectroscopy has improved our understanding of protein structure and the increasing number of sequences available makes this analysis easier. This has enabled prediction of how to increase selectivity, stability and activity of the enzyme of choice via molecular modelling. Even if there is no direct structural information available, a homologous enzyme structure can be used as a model (when sequence homology is greater than 30%). A number of studies have successfully implemented the rational design of protein structures and design of altered properties such as protein stability, and have been reviewed extensively (Bornscheuer and Pohl, 2001).

1.2.2 Directed Evolution (DE)

Directed evolution has in recent years been the method of choice and facilitated the introduction of changed or novel properties into existing enzymes. A number of enzyme properties can be modified such as improving stability, substrate specificity, altering enantioselectivity and sometimes introducing properties not seen in the parent enzyme.

Diversity is typically introduced at a genetic level by employing one or more of the methods described below in more detail. It is then followed up by screening or selection of favourable traits or functional features in the adapted enzymes. Clearly for this to work well, it relies on the availability of precise/appropriate selection or screening methods to identify desired variants. In nature the viability of new variants and their ability to produce viable offspring will determine their fitness. However for directed evolution, the availability of a viable screening technique has often been a limiting factor to how widely the technique can been applied.

Directed evolution has been applied successfully in recent years to enhance different features of an enzyme as well as overall performance, such as altering substrate specificity (Glieder *et al.*, 2002; Meyer *et al.*, 2002),

enantioselectivity (May *et al.*, 2000; Reetz et *al.*, 1997 and 2000), improving thermal stability (Gonzalez-Blasco *et al.*, 2000; Merz *et al.*, 2000) as well as adapting other variables such as optimum pH (Bessler *et al.*, 2003; Eijsink *et al.*, 2005). There are different methods for achieving directed evolution to modify existing enzymes functions and/or to introduce non-native enzyme activities. The methods used to create enzyme variant libraries fall into the two main categories of recombinative and non-recombinative mutagenesis.

1.2.3 Recombinative strategies

DNA shuffling involves the molecular mixing of similar or randomly mutagenised genes to develop chimeric recombinants, and their selection to obtain enhanced traits for the desired properties. The initial technique involved random partial digestion of gene fragments with DNase I, before recombining the fragments randomly in a PCR amplification step (Stemmer 1994). DNA shuffling is not restricted to mutants of the same gene as it can also be performed with homologous genes from different species. Ultimately, beneficial mutants can be identified and genes reshuffled against each other or by recombination with the wild type to effectively eliminate neutral and deleterious mutations.

The staggered Extension Process StEP is another technique based around PCR reactions (Zhao *et al.*, 1998). It involves very short annealing and extension steps, which facilitate formation of premature extension products, and eliminates the need for DNase I digestion. Consecutive cycles can allow formation of truncated strands which can anneal randomly to the parental strand enabling the information on different parent strands to combine.

Random Priming Recombination RPR is another alternative technique to DNA shuffling (Shao *et al.*, 1997). Random primers are used to generate short DNA fragments which are complementary to different regions of the target DNA. The fragment often has mutations arising from misincorporation or priming of bases. When integrated these lead to recombination via reassembly of the full length gene. The random chimeragenesis on transient templates (RACHITT) technique relies on the trimming, ordering and joining of

randomly cleaved single-stranded parental gene fragments annealed onto a transient full-length single stranded template (Coco, 2003). This technique enables high recombination frequency and chimeric gene products. There are also some non-homologous recombination techniques. SHIPREC (sequence homology-independent protein recombination) creates libraries of single crossover hybrids of unrelated or distantly related proteins (Udit et *al.*, 2003). Incremental truncation from the creation of hybrid enzymes (ITCHY) is a technique which uses nucleotide triphosphate analogues to create incremental truncation libraries (Lutz *et al.*, 2001a). SCRATCHY is a modification of ITCHY which combines it with DNA shuffling (Lutz *et al.*, 2001b) to help overcome the drawback of only single crossovers occurring between the parent genes as can occur in ITCHY and SHIPREC.

1.2.4 Non-recombinative strategies

Error-prone PCR (EpPCR) is a technique that facilitates random point mutagenesis. This is a modified form of PCR where the DNA amplification is not done to the same high fidelity. EpPCR takes advantage of base mis-incorporation by increasing mutation rate by varying magnesium and manganese (Mg^{2+}/Mn^{2+}) or dGTP concentrations. Varying the concentrations of one or both of these can increase the mutation rate to 2-5 base substitutions per gene.

An alternative to EpPCR for non-recombinative mutagenesis is the use of bacterial mutator strains. Wild-type *E. coli* strains exhibit a spontaneous mutation frequency of about 2.5×10^{-4} per 1000 nucleotides of DNA propagated on a pBluescript-like plasmid after 30 generations of growth. Strains with compromised DNA-repair pathways exhibit much higher spontaneous mutation rates (Miller, 1992). Stratagene Ltd. has created a commercial strain of *E. coli* – XL1-Red – that contains mutations in three independent DNA repair pathways (*mutD*, *mutS*, and *mutT*) (Greener *et al.*, 1997). This strain exhibits a spontaneous mutation frequency of 0.5 mutations per 1000 nucleotides of DNA. To generate a mutant library, a gene is cloned into a suitable plasmid, transformed into *E. coli* XL1-Red, and propagated.

Unfortunately, if the target sequence is small (<100 bp), or if multiple mutations are required, the number of generations needed for propagation of the plasmid becomes impractically high. In addition, due to the lack of DNA repair pathways, the strain is fairly unstable and cannot be propagated for long periods.

Even though laboratory evolution aims to mimic the natural processes of mutation, recombination and selection there will be differences between the two processes. Directed evolution is a guided process with a target in mind, resulting in mainly adaptive mutations in the protein, while natural evolution accumulates adaptive mutations due to selective pressure and neutral mutations which do not affect 'fitness'.

1.2.5 Evolutionary strategies

The number of possible permutations of an amino acid sequence balloons with sequence length. As there are 20 natural amino acids available at each position in a sequence, an enzyme of 300 amino acids has a total of 20^{300} possible mutants. This "sequence space" is mostly empty of function and, from a practical point of view, infinite. As complete inspection of this space is impossible, it must be sampled in a manner that suits a particular directed evolution project. Non-recombinative and homologous recombination methods are ideally suited for the generation, diversification, and optimization of local protein space (Bogarad and Deem, 1999; Miller and Orgel, 1974). In other words, DNA base substitutions are best for "fine-tuning" the activity or specificity of an enzyme. Non-homologous recombination, on the other hand, permits the creation of productive tertiary folds by way of the novel juxtaposition of encoded structures (Bogarad and Deem, 1999). This approach is, therefore, suited to the evolution of entirely novel protein functionality.

Evolution is often referred to as a hill-climbing exercise in the "fitness landscape" of sequence space (Kauffman, 1993). The individual fitness levels of the protein sequences in sequence space make up this landscape. "Tweaking" an enzyme's activity or specificity by non-recombinative methods

or homologous recombination is akin to climbing the local optimum. Beneficial point mutations are very rare, so most paths lead downwards (decreasing fitness). Furthermore, the probability that one will find oneself at a higher fitness point decreases rapidly as the number of simultaneous point mutations increases (Arnold, 1998). Consequently, an uphill climb in this landscape is more likely to be successful if it can take place in small steps (a single amino acid substitution per generation). While the global optimum may never be achieved, the accumulated improvements yielded by this approach may nonetheless yield a highly successful result.

An alternative to this up-hill walk by single amino acid substitutions is to subject a small region of the gene to a higher level of random mutation (Figure 1.1). This approach may yield novel combinations of substitutions that operate synergistically. Also, careful selection of a targeted region could lead to a smaller proportion of the scrutinised sequence space being functionally-derelict. Voigt and co-workers have described computational tools for identifying amino acids in a given protein that are tolerant to mutation (Voigt *et al.*, 2001a; Voigt *et al.*, 2001b). This concept of pre-screening libraries *in silico* has been taken a step further by Protein Data Automation (PDA) technology which eliminates sequences that are incompatible with the protein fold (Hayes *et al.*, 2002). After selecting 19 residues near to the active site of β-lactamase, Hayes and co-workers used PDA to trim the 7×10^{23} possible sequences down to just 172800. The remaining variants were constructed and experimentally screened in a single round to produce entirely novel mutations and an incredible 1280-fold increase in cefotaxime resistance.

Technique	Description and merits	Reference
Non-recombinative		
Error-prone PCR (EpPCR)	A gene is amplified by PCR under conditions that encourage the random misincorporation of nucleotides	Zhou *et al.*, 1991
Mutator strain: *E. coli* XL1-Red	Mutations in three independent DNA repair pathways cause high levels of spontaneous mutation in an introduced plasmid	Greener *et al.*, 1997
Random oligonucleotide mutagenesis	Random mutations, encoded on a synthetic oligonucleotide, are incorporated into a specific region of a gene by ligation	Sneeden and Loeb, 2003
Homologous recombination		
DNA-shuffling	One or more homologous genes are fragmented with DNase I. Full-length hybrid genes are reconstructed from these fragments in a PCR-like reaction	Stemmer, 1994
Staggered extension process (StEP)	Simple and efficient variation of DNA-shuffling based entirely on PCR	Zhao *et al.*, 1998
Random-priming recombination (RPR)	Variation of DNA-shuffling where every nucleotide has an equal probability of being mutated	Shao *et al.*, 1998
Combinatorial libraries enhanced by recombination in yeast (CLERY)	*In vitro* family-shuffling followed by *in vivo* homologous recombination in yeast. High complexity hybrids but low fitness	Abecassis *et al.*, 2000
Random chimeragenesis on transient templates (RACHITT)	Randomly-cleaved parental gene fragments are annealed to a transient full-length template. 100% chimerical gene products are generated	Coco *et al.*, 2001
Non-homologous recombination		
Sequence homology-independent protein recombination (SHIPREC)	Two non-homologous parent genes are fused end-to-end and randomly fragmented by DNase I. Parent-sized fragments are isolated and circularised. Single-crossover hybrids are thus generated	Sieber *et al.*, 2001
Incremental truncation for the creation of hybrid enzymes (ITCHY)	Nucleotide triphosphate analogues are used to create incremental truncation libraries. Single-crossover hybrids of unrelated or distantly related proteins are generated	Lutz *et al.*, 2001a
ITCHY with DNA-shuffling (SCRATCHY)	Same as ITCHY but capable of generating multiple-crossover hybrids	Lutz *et al.*, 2001b
Structure-based combinatorial protein engineering (SCOPE)	In a series of PCRs, "hybrid oligonucleotides" act as surrogate introns to direct the assembly of coding segments from two non-homologous genes. Multiple crossovers occur in each hybrid	O'Maille *et al.*, 2002

Table 1.1 Current techniques for generating libraries of mutants for directed evolution.

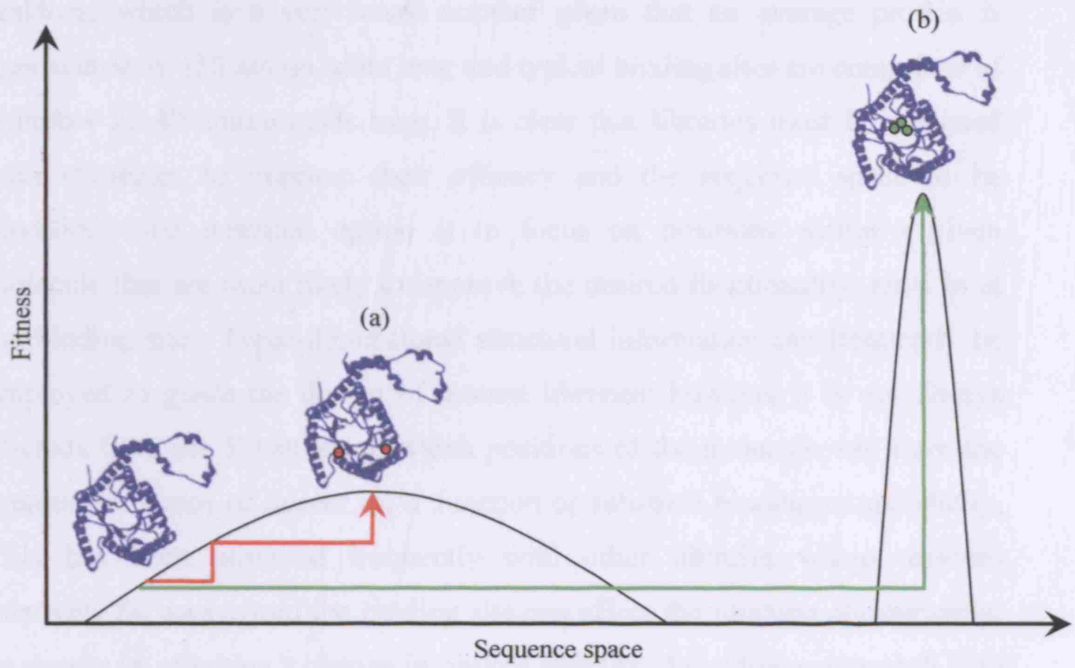


Figure 1.1 Simplified representation of evolutionary progressions by: (a) repeated single amino acid substitutions; and (b) intense mutation of a small region of the gene. It is possible for method (b) to yield highly successful mutants with substituted residues that work together. Note that, in reality, sequence space has many more dimensions.

1.3 Library size and coverage

1.3.1 Library size

There are practical limits to the achievable size of display libraries. Cell-based and phage display libraries are unable to handle more than 10^{11} library members and cell free ribosomal displays can handle no more than 10^{14} variants. Direct synthesis of random oligonucleotides is easy to do but is not the most effective way to search sequence space. Complete randomization of 24 nucleotide position leads to 4^{24} distinct possible library members, or a potential library size of 10^{14}. This only corresponds to the randomisation of eight amino acid positions, but redundant and stop codons can be eliminated if desired. If produced with perfect accuracy, the theoretical limit of library size thus corresponds to the complete randomisation of only 11 amino acid

residues, which is a very small number given that an average protein is approximately 150 amino acids long and typical binding sites are comprised of stretches 20-40 amino acids long. It is clear that libraries must be designed with strategies to improve their efficacy and the sequence space to be modified. One potential option is to focus on positions within a given molecule that are most likely to improve the desired functionality, such as at the binding site. Three-dimensional structural information can frequently be employed to guide the design of mutant libraries; however it is not always obvious from the 3D structure which positions of the molecule will have the greatest influence or impact on it function or substrate binding or specificity. This has been observed frequently with other libraries where residues relatively far away from the binding site can affect the function allosterically, or simply by affecting a change in protein stability. An extreme approach is to attempt to introduce random mutations throughout the protein gene of interest, by either error-prone PCR (EpPCR) or an *E. coli* strain with deleterious mutations in it DNA repair mechanisms (E.g. XL1-Red). Although the coverage on the sequence space is low and frequently non-viable mutants are produced this has been used when little functional information is available. Gene shuffling is also a potentially useful technique when multiple copies of a gene are available. These can then be amplified and randomly fragmented before being reassembled in novel ways producing chimeric proteins. It is also a potentially powerful technique of library design as useful or interesting progeny can be selected and re-shuffled enabling potential accumulation of beneficial mutations which can have powerful results. However shuffling and randomised mutation approaches such as EpPCR or mutator strain techniques depending on the frequency of mutation and the gene involved can result in very large libraries much of which may be deleterious or redundant mutations. As a result saturation site directed mutagenesis and use of structural and phylogenetic analysis as well as predictive forcefields should be employed together when possible to enable production of optimised libraries with potentially good candidate sites for introducing desired traits or improvements.

1.3.2 Library size and redundancy of mutation

An obvious limitation to any mutation strategy is the ability to screen the resulting library with the necessary speed and accuracy. The quality and speed of the screen to be used are key to successful analysis of a large library of promising variants. The method used for the directed evolution will also affect the size of the library, and the ability to screen it, as the frequency of mutation over a gene will significantly alter the number of possible variants that are likely to be produced. Whole gene mutation will create libraries of much greater size than a more focused mutagenic method and although some more variants of interest might be accessible as a result of the greater library size they would be very hard to find if a high throughput screen was not available. While whole gene mutation can provide novel mutations and enzyme improvements, many of the mutations will be deleterious and these libraries may have a lot of redundancy. This is undesirable as more time will be spent on screening variants many of which may be disrupted functionally by the mutations. By focusing mutagenesis to sites of interest, such as active site residues which are most likely to cause substrate specificity changes (Dalby, 2003), the size of libraries that need to be screened and the redundancy encountered may be significantly reduced. The current opinions for library production are generally in favour of smaller, better 'quality' libraries (Lutz and Patrick, 2004). Other ways of reducing redundancy and library sizes are modelling of deleterious and beneficial mutations using computational forcefields. These are often simulations based on models which allow calculation of each residue's structural tolerance. These types of analyses as well as phylogenetic analyses can greatly reduce the search space for directed protein evolution and as a result, the redundancy encountered in library production also significantly decreases.

1.4 Screening and selection

The ability to screen or select desired mutants from a constructed library will determine the size and quality of the libraries that need to be produced. A balance must be struck between the size and quality of the libraries and the time taken to screen and identify possible mutants of interest; as a result this is a factor that must be considered when determining library size and the construction methods used.

1.4.1 Screening Strategies

Following the creation of a mutant library, the mutant genes are expressed and a screening or selection step is used to isolate the best variant(s) for the next cycle. "Screening" is the process of identifying a desired member of a library in the presence of all other members; for example, each variant is assayed for improved catalytic activity under a specific set of conditions. With automation, approximately 10^4 mutant proteins can be studied by screening (Reetz and Jaeger, 1999). If "selection" is applied instead, then only the desired member of a library appears; for example, as a viable microbial clone. Selection processes allow the experimenter to examine approximately 10^8 individual variants – suggesting an increased chance of finding the desired enzyme – but have severe drawbacks. True selections in which only those clones carrying a desired improvement survive or grow faster are rare, usually highly specific to one problem, and difficult to implement productively in directed evolution (Reetz and Jaeger, 1999). When available and validated, however, they can deliver dramatic results.

Recombinant antibody fragments are unstable in the reducing environment of the cytoplasm because the intradomain disulphide bonds are not formed. Martineau and co-workers (Martineau *et al.*, 1998) have described a directed evolution experiment with a selection procedure that linked correct folding of a recombinant antibody fragment to cell survival. An antibody single-chain Fv fragment (scFv) that binds and activates an inactive mutant β-galactosidase was isolated. The gene encoding this scFv fragment was subjected to random mutation by EpPCR and the resulting library was co-expressed with the mutant β-galactosidase gene in *lac*$^-$ bacteria. By plating on

limiting lactose, antibody mutants with improved folding were selected. After four successive rounds of mutation and selection, a variant was isolated that was expressed correctly in the bacterial cytoplasm with an excellent yield ($3.1 g.l^{-1}$ in a fermentor).

A screen is required when the desired activity or feature cannot be easily linked to cell survival or growth. Screens are more versatile than selections, but at a cost to throughput. Reactions that take place in microwell plates – allowing for spectrophotometric detection of the products – are still the method of choice in the development of screens.

A simple spectrophotometric approach for following the course of a reaction that depletes or generates protons is to include a pH indicator in the reaction. May and co-workers (May *et al.*, 2000) used this tactic in a screen for enantioselectivity. A library of mutants was created from a D-selective hydantoinase gene using EpPCR. 10000 variants were grown up and incubated with L-5-(2-methylthioethyl)-hydantoin (L-MTEH) and D-MTEH separately. The rate of each reaction was measured by following the change in A_{580} of the pH indicator cresol red as it responded to changes in pH. This simple screen identified several mutants with novel L-selectivity. Further genetic manipulation of these variants improved performance and yielded an enzyme with possible application in the industrial synthesis of the amino acid L-methionine.

Aptamers consist of ssDNA or RNA obtained from a random oligonucleotide library by Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Tuerk and Gold 1990; Daniels *et al.*, 2003). SELEX is a technique used to screen large combinatorial libraries of oligonucleotides by an iterative process of *in vitro* selection and amplification. The combinatorial libraries based on replicable oligonucleotides offer the convenience of amplification, and also make the screening process fast and easy. In the selection procedure, the library is incubated with a target of interest which is often fixed on a solid vector such as beads, nitrocellulose membranes or columns. The population of unbound sequences is washed away and the bonded sequences are isolated and amplified to obtain a smaller, but more enriched library for the next selection round. The enrichment efficiency can

be detected by several binding assays such as flow cytometry techniques. Once binding saturation is achieved after several rounds of selections, the selected oligonucleotides are cloned and sequenced to obtain the single sequence, which is called the aptamer. The length of an aptamer is usually 50–80 base pairs. The relative small molecular weight makes it suitable to apply to the molecular reorganization. Aptamers could fold into a three-dimensional structure based on their nucleic acid sequence to bind with the target. The dissociation constants (*Kd*) between the aptamer and its target is often in the nano molar or even pico molar range which shows a very strong specific interaction. With the advantage of being non-immunogenic; and showing rapid kidney clearance and easy plasticity, aptamers are being used on both diagnostic and therapeutic applications. Aptamers with high affinity and specificity for cells and tissues have also been produced, demonstrating that complex targets, including purified proteins, amino acids, and cells are suitable for use with the SELEX procedure.

Recently aptamers produced from a library after 10 rounds of selection for high affinity for osteoblasts were PCR-amplified. The aptamer was used immobilized on streptavidin coated plates and then incubated with a solution of osteoblasts with 3 ½ times more cells being attached than with non aptamer coated surfaces (Guo *et al.*, 2005). The aim of this work is to increase the cytocompatibility of medical implant and as a result potentially causing a reduction in the impaired cell adhesion and a reduced ability to grow encountered with uncoated implants.

Another emerging screening method is based on the "cell immobilized on adsorbed bead" approach (Francisco *et al.*, 1992, Georgiou *et al.*, 1997) : the cell population to be screened is mixed with an excess of medium pre-equilibrated polyacrylamide beads, chemically derivatized to effect quantitative cell immobilization by adsorption. The resulting bead population, comprising of single cell on a bead or blank beads, is then immobilized on a solid glass support. After removal of the freely flowing liquid, the cells immobilized on the adsorbed beads are allowed to grow into micro colonies, utilizing the medium retained within the supporting hydrogel matrix. These colonies are subsequently equilibrated with chromogenic or fluorogenic

substrate and screening is carried out under a stereomicroscope, resulting in readily retrieved of the most active colonies. This technique may be particularly useful when the screened mutants are expressed and displayed on the cell surface, providing an active and homogeneous ''naturally immobilized'' enzyme population with minimal substrate diffusion limitations. This technique has been used to screen large bacterial libraries for desirable mutant proteins in this case catalytic activity (Freeman *et al.*, 2004)

1.4.2 Selection Strategies

Two main types of *in vitro* selection systems can be distinguished. The first exploits the physical relationship/proximity between messenger RNA (mRNA) and nascent polypeptides, during translation, to couple genotype and phenotype (Martineau *et al.*, 1998). The second imitates the compartmentalisation of living cells by performing translation and selection within a water-in-oil emulsion (Tawfik and Griffiths 1998).

In contrast to *in vivo* selections, in vitro technologies can handle libraries with up to 10^{14} members, depending on the scale of the in vitro translation used. This enables an increase in the diversity of sequences and thus improves the likelihood that a successful mutant will be isolated. The system used will depend on the desired end result, however *in vitro* systems are important when bacteria or fungi fail to express desired proteins well, such as folding, transport, membrane insertion and provide significant *in vivo* selection pressure against certain proteins and scaffolds. These pressures are likely to be significantly reduced *in vitro* where efficient translation is the only requirement. The main drawback with these systems is that they can become complicated methodologically.

1.4.3 In vivo selection

In vivo selections are strategies which make use of existing cell machinery to select for desired mutants. The cell is used as a compartment

within which the desired protein(s) are expressed and perform their catalytic functions, and which will ensure the propagation of the gene encoding the desired protein(s) by means of cellular replication. One method commonly used is employing an auxotrophic cell line as host of choice (Joerger *et al.*, 2003) and or to link the activity of the protein to be evolved to the survival of the cell (Zaccolo *et al.* 1996).

In vivo systems can and have been used with success, but they have inherent limitations. The cell may find a way to overcome the 'problem' presented to it by the in vivo selection by making use of large amount of genomic DNA available to circumvent the need to use the desired protein or survive in its absence. Another problem is that clearly only functions that can be linked to cell survival can be selected for using this methodology. A problem also encountered with these systems is that often it is not possible to select for enzymes with low activity, as a certain threshold of minimum activity is likely to be needed for the cell to survive.

1.4.4 In Vitro selection

Two such strategies that have been broadly used in recent years are phage display and colorimetric and flourimetric assays. In phage display the protein of interest is fused to a phage-coat protein, which results in the display of one or more copies of the fusion protein on the surface of the filamentous bacteriophage. Molecules are selected for by affinity panning on immobilised substrates, and the resulting phages are mixed with bacteria to produce progeny phage for subsequent rounds of panning (Rader and Barbas,. 1997). Colourimetric/flourimetric assays utilise bacteria or other host to translate and express the desired protein, while using chromogenic/fluorogenic products either to screen colonies of bacteria or using Fluorescence Activated Cell Sorting (FACS) to sort repertoires of genes or bacteria compartmentalised in double emulsions (Bernath *et al.*, 2004).

As a results of the in vivo step employed in these strategies the maximum library size is limited by the transformation/transfection limit of the organisms involved (10^7-10^{11} cells/μg DNA for *E.coli* and 10^7-10^8 cells/μg of DNA for

yeast). In addition to limiting the library size, the in vivo step may apply only to evolving enzymes that do not interfere substantially with the cellular metabolism and can be distinguished from the background of the other cell processes.

A mixture of these strategies will be applied to a model enzyme in this case transketolase (TK) to investigate the best way(s) to modify enzymatic activity a structural based approach or a phylogenetic based DE strategy.

1.5 Improving directed evolution strategies

Several successful strategies for directed evolution have been demonstrated, with the most recent examples aimed at making this technique more efficient. Higher efficiency is obtained by reducing the number of redundant variants in a library, or by increasing the potential for improvement over wild-type activity. These two are not independent as less redundancy is expected to allow greater potential for improvement using the same overall library size.

1.5.1 Mutation rate

One method for improving directed evolution efficiency is to increase the mutation rate of, for example EpPCR, to permit the synergistic effects of multiple mutations to be identified more frequently (Zaccolo and Gherardi, 1999). Libraries constructed using high-mutation frequencies have been shown to yield a higher than expected number of functional variants given that this also leads to an increased likelihood of negating positive mutations, or non-functional variants. However, it was shown recently that an increased mutation rate leads to more unique variants in a library, whereas single mutant libraries often contain many copies of the same variant (Drummond *et al.*, 2005).

1.5.2 Targeted mutagenesis

Another strategy, that has the potential to obtain more useful variants from restricted library sizes, is to focus mutations in regions of the enzyme more likely to result in beneficial mutations, for example the active site (Shinkai *et al.*, 2001), or where hot-spots have been identified by error-prone PCR (Miyazaki and Arnold, 1999). Recent examples of non-targeted directed evolution, for improvement of properties traditionally associated with the active site, were producing the majority of mutations in regions that contribute to substrate binding, catalysis or the conformation and dynamics of the active site environment (Dalby, 2003) (Morley and Kazlauskas, 2005). However, not all enzyme properties can be expected to improve due to active-site mutations alone. For example, thermostability was shown to be improved equally by mutations close to and distant from the active site.

The targeting of saturation mutagenesis, either experimentally or *in silico*, to active-site residues chosen solely on a structural basis for their proximity to substrate or product has led to the successful improvement of enzyme activities (Hayes *et al.*, 2002), altered substrate specificity (Santoro and Schultz, 2002;Wang *et al.*, 2001;Ting *et al.*, 2001;Ashworth *et al.*, 2006), and altered catalytic reactions (Peimbert and Segovia, 2003;Goud *et al.*, 2001;Dwyer *et al.*, 2004).

The overall benefit of targeted mutagenesis has been systematically and directly compared to fully random mutagenesis in a number of studies, including the improved enantioselectivity of the *P. fluorescens* esterase (Park et al., 2005), and the improved activity of dihydrofolate reductase (DHFR) (Schmitzer *et al.*, 2004). The most striking example has compared the site-saturation mutagenesis of three carefully chosen active-site residues in *E. coli* β-galactosidase, to a previous DNA shuffling experiment for the same enzyme (Parikh and Matsumura, 2005). While DNA shuffling enhanced the k_{cat}/K_m for β-fucosidase activity by 10-fold after seven rounds, saturation mutagenesis resulted in a 180-fold improvement in a single round.

Simultaneous mutagenesis of typically up to eight residues can be used, requiring powerful selection-based methods to identify useful variants (Santoro and Schultz, 2002;Wang *et al.*, 2001;Peimbert and Segovia, 2003).

In cases where selection is not possible, automated screening techniques must be used instead, and consequently library sizes are restricted to a few thousand variants. This in turn limits the targeted saturation mutagenesis to only single or double mutants. Double mutation in a pairwise manner has been attempted where residues close in sequence were simultaneously randomised in a technique dubbed CASTing (combinatorial active-site saturation) (Reetz *et al.*, 2005), and with which improved variants, including double mutants, of a lipase from *Pseudomonas aeruginosa* were obtained on substrates not previously accepted by the wild-type enzyme.

1.5.3 Prediction of sites tolerant to mutagenesis

Predictive methods, other than structural analysis, for guiding the location of targeted mutagenesis would be extremely valuable. Tolerance to mutations, and therefore, the ability to evolve activities that otherwise destabilise the enzyme, increases with protein stability (Bloom *et al.*, 2006). Previous analyses of directed evolution experiments have shown that mutated sites resulting in improvements, are also frequently those that are tolerant to mutation as determined by computational modelling of the protein thermostability (Voigt *et al.*, 2001a;Voigt *et al.*, 2001b). Interestingly, mutation of enzyme active-sites which reduce their activity, frequently lead to an increase in protein stability, presumably due to release of the strain imposed on the active-site structure that is imposed by functional requirements (Shoichet *et al.*, 1995).

1.5.4 Synergy of mutations

One extra potential benefit of targeted mutagenesis is an increased likelihood of obtaining synergistic mutations. It has been predicted previously that at low mutation rates, the probability of a beneficial mutation occurring at a highly coupled residue decreases rapidly as the fitness of the parent increases (Voigt *et al.*, 2001a). Such mutations are only found by current iterative directed evolution methods when each mutation contributes independently to

fitness and are then recombined, or alternatively when at least the first mutation results in an increase in activity. This results from the practical limitation of imposing a positive selection pressure at each round of library screening. Relaxation of this requirement may allow neutral and even some deleterious mutations to accumulate, as occurs in nature, leading to the possibility of observing synergistically coupled mutations during later evolution. This effect is postulated by the authors to explain their observation of poor correlation between natural sequence entropy and their calculated site entropy, which imposes a positive selection pressure in terms of maintained thermostability. Recently, Weinreich and co-workers (Weinreich *et al.*, 2006) showed that where positive selection is a pre-requisite, then accessible evolutionary trajectories towards a final fitness peak tend to be limited to only the few most probable ones (18/120 possible trajectories for evolving an increased activity of TEM β-lactamase towards cefotaxime). However, the authors note that as microbes are exposed to a spatial and temporal diversity of selection pressure a wider range of trajectories are more probable in Nature. This includes those for which synergistic gains are obtained from one or more mutations that can each have negative effects when occurring independently. However, as directed evolution experiments with a positive selection pressure preclude the observation of synergy from independently deleterious mutations, simultaneous mutation in a single round of selection is required. Targeted mutagenesis makes simultaneous mutation more practical, and also therefore the search for synergy, as the size of a library containing all possible mutants is dramatically reduced. It currently remains unclear whether synergy can easily be obtained through pairwise mutation as previously attempted (Reetz *et al.*, 2005), as it may be that coupling events occur more frequently between two residues that are not adjacent in sequence.

1.5.5 Sequence entropy and phylogenetics

An alternative to using structural information to guide mutagenesis is to use sequence information. Sequence alignments have been used to identify residues deviating from the consensus and subsequently to construct libraries

in which these residues are simultaneously mutated back (Amin *et al.*, 2004;Flores and Ellington, 2005) to obtain improved thermostability or altered cofactor specificity. Natural sequence variation has also been previously maximised in an artificially created DNA shuffling library (Ness *et al.*, 2002). While successfully maximising genetic recombination, this method potentially limits the range of beneficial mutations that are not observed in Nature.

More recently, it has been shown that for the improvement of EGF, EGFR, IL-2 and subtilisin, that directed evolution is biased towards residues that are phylogenetically variant in Nature (Murase *et al*, 2003; Sato *et al*, 2004; Rao *et al.*, 2004&2005; Lehmann *et al.*, 2000&2002), and improved EGF variants were especially obtained from variants containing residues found in orthologous EGF species (Cochran *et al.*, 2004). Loss of function was observed for mutation of highly conserved residues. More generally, from EpPCR experiments, 40-80% of improved mutations from IL-2, EGF, EGFR and subtilisin were also present in at least one ortholog.

1.5.6 Focused Directed Evolution

The ability of enzymes to catalyse reactions with exquisite control of regio- and stereochemistry as well as enabling up to 10^{17} –fold rate acceleration clearly shows why their use is increasing in many areas including synthetic chemistry and diagnostics. The acceptance and continued use of enzymes in chemical synthesis will depend on their generality as catalysts, in particular substrate range. Even if an enzyme is found with the correct substrate activity, other characteristics will often have to be improved also to make it applicable in industrial process such as enzyme stability temperature and pH. Directed evolution is central to producing enzymes suitable for industrial processes and or improving existing ones for more efficient process. Recent theories however point to a need to produce libraries which are easily screened and are of better quality, less redundancy and a more informed choice of residues to target. A combination of computational modelling, bioinformatics and genetic tools for creating smart libraries is the way forward (Dalby, 2003).

1.5.7 Promiscuity

Catalytic promiscuity in enzymes (also know as moonlighting activity, cross-reactivity or substrate ambiguity) is the ability of their active sites to catalyse distinctly different chemical transformations. These transformations may differ in the functional group involved, and as a result the type of bond formed or cleaved during the reaction and/or may differ in the catalytic mechanism or path of bond breaking or formation. Recent evidence has shown that such catalytic promiscuity exist not just among a few enzymes, but is rather common (Copley, 2003; Bornscheuer and Kazlauskas, 2004). The promiscuous activity of many enzymes is often at much lower levels than their 'natural' reactions, however often the promiscuous activity of enzymes is similar to the natural activity of an evolutionarily related enzyme. These observations of similar activities have led to theories that nature utilises these promiscuous activities as starting points for evolution of new enzymes with gene duplication allowing this to occur when the parental enzyme's function is vital to the organism's survival. This idea was first suggested by Jensen (Jensen, 1976) that, under changing environments, promiscuous activity in existing proteins can endow the organism with a selective advantage and thereby enable its survival and further evolution. However direct evidence for this remains scarce, but exploiting promiscuous activity of enzymes clearly has potential not only for modifying and improving alternative substrate specificities but potentially engineering novel enzymes using promiscuous activities as a basis.

1.5.8 Substrate walking concept

The concept of substrate walking is schematically shown in Figure 1.2, and is discussed in more detail in Chapters 3 and 5. It could be viewed as an application of promiscuous enzyme activity as a method to direct the substrate chosen to be engineered for improved or novel activity as well as looking at

evolutionarily related enzymes to find which residues when modified might improve the desired activity.

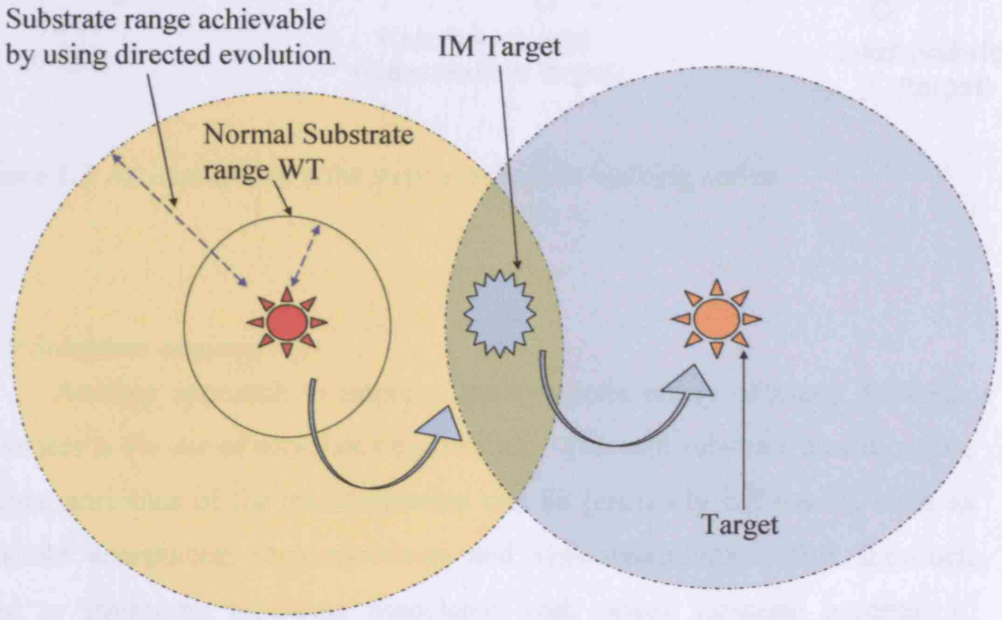


Figure 1.2 Substrate walking concept.

It is possible that the intermediate target, as shown in Figure 1.3 could be considered a promiscuous substrate (as the boundaries shown as in Figure 1.2, are not strictly defined), but it would need to be useable at low levels like those seen naturally under deleterious conditions for natural organisms if an intermediate target is found with low levels of activity. This target could then be used as a stepwise process to a more desirable or complex novel substrate. The potential range of the substrate specificities that might be introduced using substrate specificity (in conjunction with directed evolution) needs further investigation.

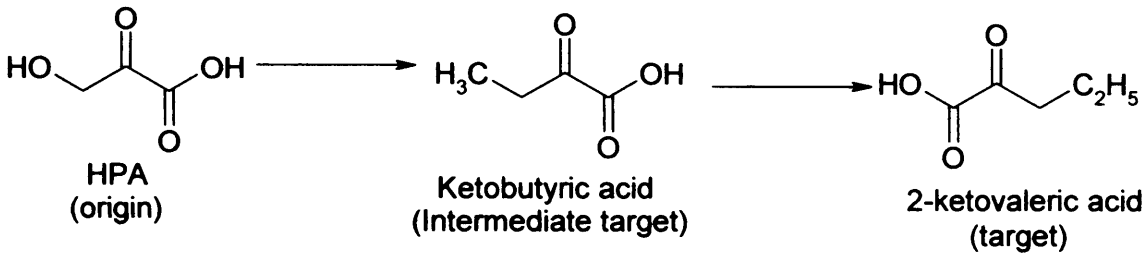


Figure 1.3 An example of a theoretical substrate walking series.

1.5.9 Substrate engineering

Another approach to improve the synthetic utility of many desirable substrates is the use of substrate engineering. Through substrate modification, various attributes of the transformation can be positively influenced, such as substrate acceptance, regioselectivity and stereoselectivity. This approach aims to overcome problems associated with novel substrate acceptance, undesired side reactions and selectivity. Substrate engineering has been shown to be a valuable method to direct and enhance the regioselectivity and stereoselectivity of enzymatic hydroxylation (de Raat and Griengl, 2002) engineering was applied to the various substrates including cycloalkanes, carboxylic acids, ketones, amines, amides and alcohols. In addition, the range of accepted substrates can be broadened, thereby improving the synthetic utility of the hydroxylation process. Another novel approach is the engineering of enzyme activity to be substrate assisted, and therefore specific to that substrate. The catalytic group from an enzyme is first removed by site-directed mutagenesis causing inactivation. Activity is then partially restored by substrates containing the missing catalytic functional group (Carter and Wells, 1987; Corey *et al.*, 1995). Another group who employed substrate engineering in conjunction with mutagenesis worked on nucleoside hydrolase which binds the purine base of the substrate between the aromatic side chains of two tryptophans. They were able to show through mutagenesis and substrate engineering a raise in *pKa* by several units (Loverix *et al.*, 2005).

1.6 Transketolase

Transketolase (TK) (EC 2.2.1.1) is a thiamine-diphosphate-dependent enzyme of the transferase class that catalyzes the conversion of sedoheptulose 7-phosphate and D-glyceraldehyde 3-phosphate to D-ribose 5-phosphate and D-xylulose 5-phosphate in the pentose phosphate pathway. TK catalyzes reactions in the Calvin cycle and the oxidative pentose phosphate pathway (OPPP) and produces erythrose-4-phosphate, which is a precursor for the shikimate pathway leading to phenylpropanoid metabolism.

In vitro, transketolase could be considered somewhat promiscuous and will accept a wide range of substrates. *In vivo*, the enzyme is found in the non-oxidative branch of the pentose phosphate pathway where it, together with transaldolase (EC 2.2.1.2), creates a link to glycolysis. It catalyses two main reactions: (a) the conversion of D-xylulose-5-phosphate and D-ribose-5-phosphate to D-sedulose-7-phosphate and D-glyceraldehyde-3-phosphate; and (b) the conversion of D-xylulose-5-phosphate and D-erythose-4-phosphate to D-fructose-6-phosphate and D-glyceraldehyde-3-phosphate.

Transketolase has been purified from a number of sources, initially from the yeast *Saccharomyces cerevisiae* (de la Haba *et al.*, 1955), and subsequently isolated from other sources including: rat liver (Horecker *et al.*, 1956), spinach (Horecker *et al.*, 1956), pig liver (Simpson, 1960), the bacterium *Lactobacillus pentosus* (Racker, 1961), the fungus *Torula sp.* (Racker, 1961), rabbit liver (Racker, 1961), the yeast *Saccharomyces carlsbergensis* (Racker, 1961), the bacterium *Alcaligenes faecalis* (Domagk and Horecker, 1965), the yeast *Candida utilis* (Kiely *et al.*, 1969), human erythrocytes (Heinrich and Wiss, 1971), mouse brain (Blass *et al.*, 1982), wheat (Murphy and Walker, 1982), human leukocytes (Mocali and Paoletti, 1989), and the bacterium *Escherichia coli* (Sprenger, 1991).

1.6.1 TK structure Created from the structure file 1QGD.pdb

TK is a homodimeric protein of two 75kDa subunits, and has been determined to be made up of three domains by X-ray crystallography as shown in Figure 1.4. These domains are the N-terminal or PP domain, the central

Pyr-domain and the C-terminal domain, and are arranged as shown in Figure 1.5.

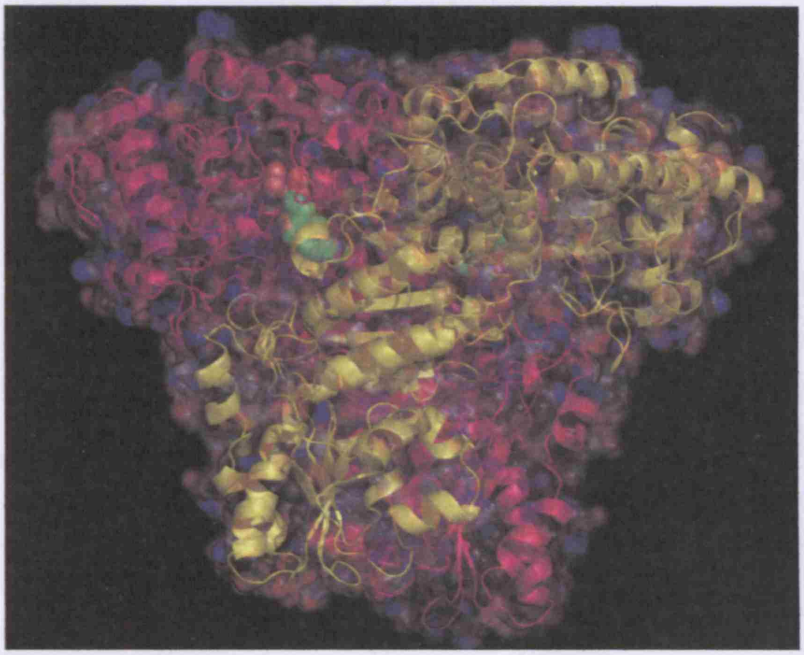


Figure 1.4 Overall structure of TK enzyme

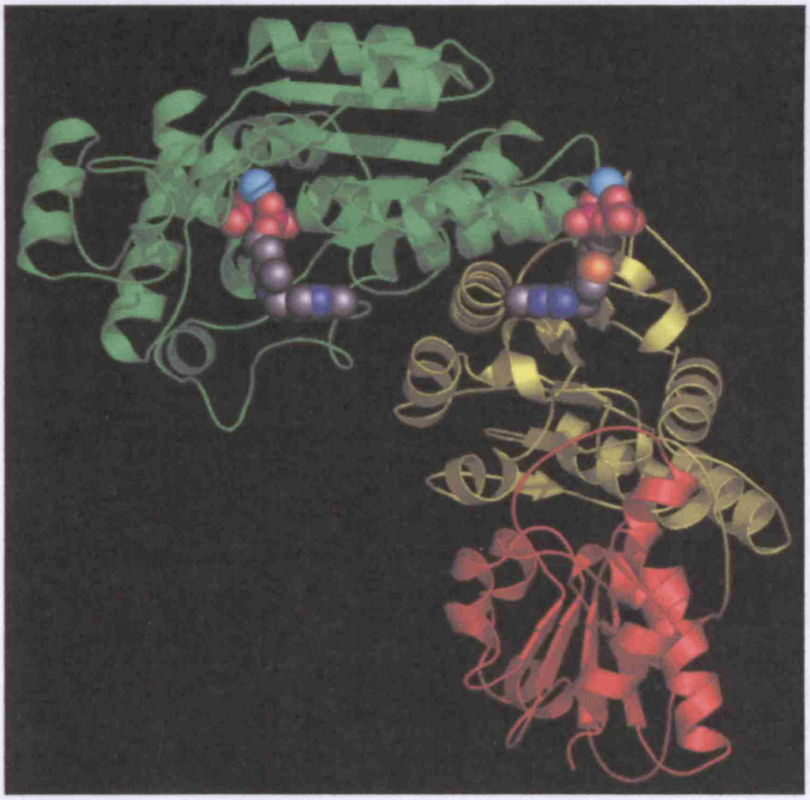


Figure 1.5 Ribbon diagram of a single subunit of *E. coli* transketolase. Two TPP molecules, which bind to the PP- and Pyr-domains, are shown as CPK space-filling models. The divalent metal ion required for catalytic activity, in

this case Ca^{2+}, is highlighted in cyan. (Created from the structure file 1QGD.pdb)

1.6.2 TK Substrate binding

TK has been widely studied for many years and investigation into the range of substrates it is able to accept has been carried out although mainly focused on the aldehyde side of the reaction. Some examples of the different aldehydes which have been tested and found to be suitable substrates with rates compared to glycolaldehyde (GA) are shown below in Table 1.2, and were constructed from data taken from (Bolte *et al.*, 1987, Demuynck *et al.*, 1991, Dalmas *et al.*, 1993, Sprenger *et al.*, 1995).

Name	Type of aldehyde	Concentration (mM)	Vrel	Km (M)	Vmax (U/mg)
Propionaldehyde		100	0.05	n/d	n/d
Butyraldehyde		100	0.11	n/d	n/d
Glycoaldehyde		100	1	$14x10^{-3}$	50.4
Methoxyacetaldehyde	Aliphatic aldehydes	100	0.32	n/d	n/d
D-Glyceraldehyde 3-phosphate	Hydrophilic	100	0.44	$2.1x10^{-3}$	100
D-Erythrose	aldehydes	60	0.84	$150x10^{-3}$	75
Benzaldehyde		100	0.11	n/d	n/d
2-Hydroxy-phenyl		100	0.28	n/d	n/d
2,3-Dihidroxy-phenyl	Aromatic aldehydes	100	0.32	n/d	n/d
2-Pyrrolyl		100	0.21	n/d	n/d
2-Furanyl		100	0.11	n/d	n/d
2-Thiophenyl		100	0.32	n/d	n/d
2-Pyridyl	Heteroaromatic	150	0.13	n/d	n/d
3-Pyridyl	aldehydes	150	0.13	n/d	n/d
3-Methyl crotonaldeyde	Unsaturated aldehydes	100	0.11	n/d	n/d

Table 1.2 Previously studied TK aldehyde acceptor substrates and their classifications. Some of the aldehydes shown here are also unsaturated such as benzaldehyde, hydroxyl benzaldehyde and the heterocyclic aldehydes shown in the table above. This demonstrates that the active site of TK is sufficiently exposed to accept large steric groups as substrates, at least for the aldehydes.

To date not many ketols have been tested for activity to establish TK specificity for those types of substrates. A summary of known ketol donor substrates for TK is shown in Table 1.3.

Name of ketol	*Km* (M)	*Vmax*
β-Hydroxypyruvate	$18X10^{-3}$	60
Xylulose 5-phosphate	$160X10^{-6}$	≥110
Fructose 6-phosphate	$1.1X10^{-3}$	62
Seduheptulose 7-phosphate	$4X10^{-3}$	6

Table 1.3 Previously studied TK ketol donor substrates and their classifications (Sprenger *et al.*, 1995).

Five other ketol donors have been selected to be screened for activity with TK, and these will be compared to the well characterised substrate hydroxypyruvate (HPA) in Chapter 3 of this thesis. The five chosen substrates are ketobutyric acid, fluoropyruvate, pyruvic acid, mercaptopyruvic acid (Na salt) and bromopyruvic acid, which all share structural similarity to HPA.

1.6.3 TK reaction mechanism

TK facilitates the reversible transfer of a two-carbon glycolaldehyde fragment from keto-sugars to the C-1 aldehyde group of aldo-sugars (Schenk *et al.*, 1998) , allowing *in vitro* synthesis of compounds with new C-C bond and a chiral hydroxyl group. Examples of the natural reactions carried out by TK are shown in Figure 1.7.

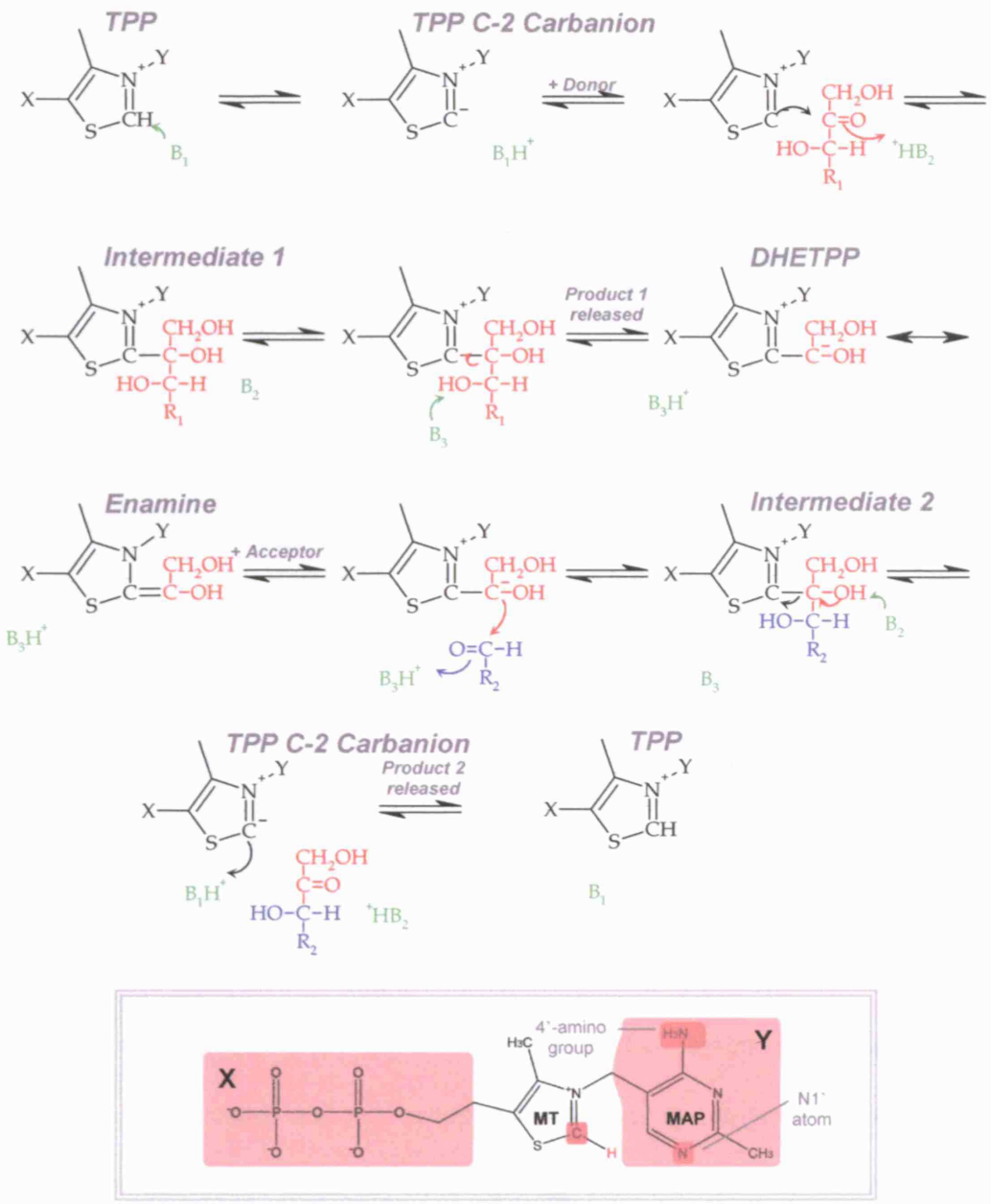


Figure 1.6: Reaction mechanism of transketolase. The donor substrate (a ketose) and the acceptor substrate (an aldose) are coloured red and blue respectively. B1 is most likely to be the 4'-imino group of the TPP pyrimidine ring. This group is also a candidate for B2; his473 is another possibility. His26 and his261 are the two candidates for B3. In the mechanism diagram, the TPP molecule is simplified. regions labelled X and Y are shown in the full molecule diagram for TPP in the grey box. Also highlighted are the N1' atom, the 4'-amino group and the C-2 atom of the MT ring. Adapted from Wikner et al 1995.

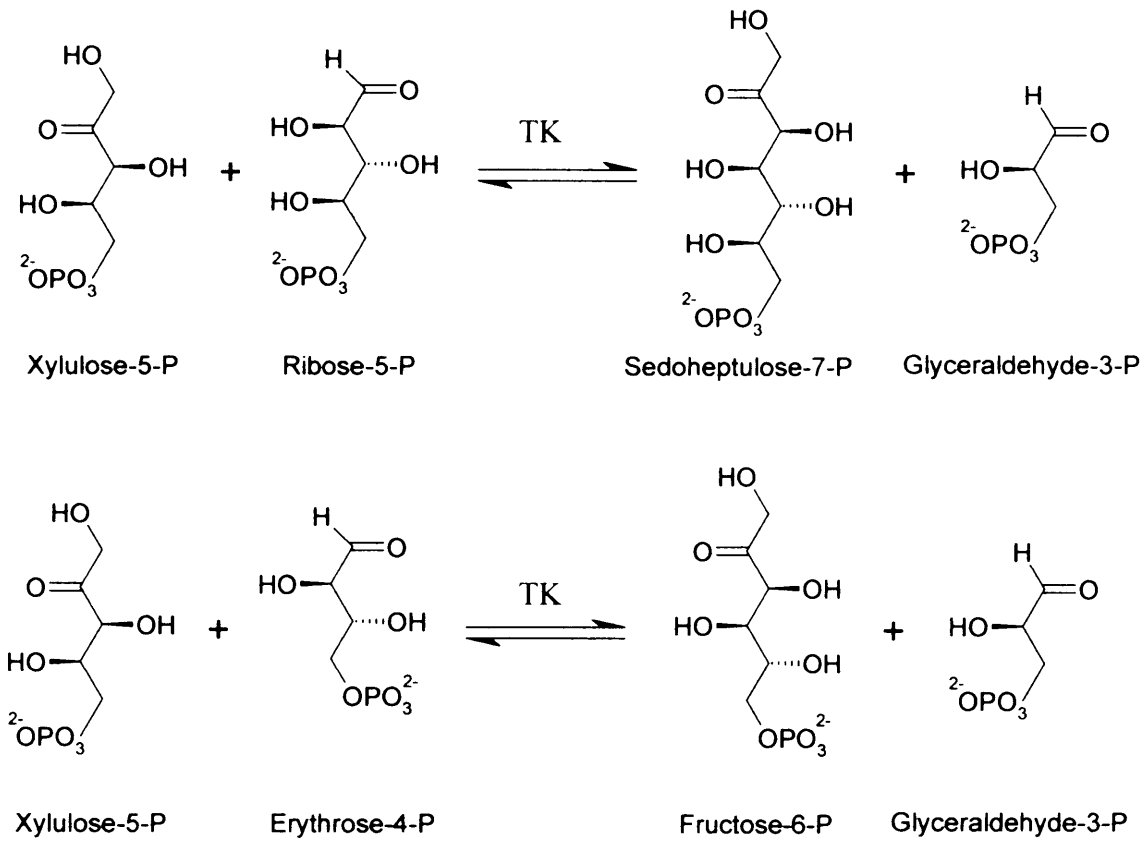
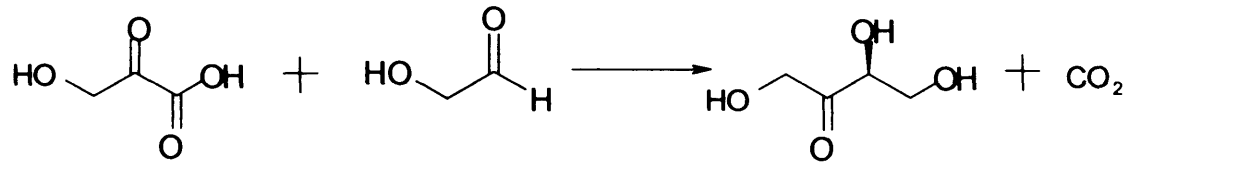


Figure 1.7 Natural TK enzymatic reactions

As previously discussed transketolase *in vitro*, could be considered somewhat promiscuous and will accept a wide range of substrates. The main non-natural substrates used for investigating TK function and activity are glycolaldehyde (GA) and β-hydroxypyruvic acid (HPA). These molecular mimics differ from the natural substrates in that the HPA results in an irreversible reaction due to the evolution of carbon dioxide (Datta and Racker, 1958), as shown in Figure 1.7. The irreversible nature of this non-natural reaction means it can be readily used for experimental research using TK where reversible reactions could pose a problem to analyse or quantify results.



Hydroxypyruvate (Ketol donor) | Glycoaldahyde (Aldehyde acceptor) | Erythrulose | Carbon dioxide

Figure 1.8 Scheme for the TK enzyme catalysed irreversible reaction between β-hydroxypyruvic acid (HPA) and glycolaldehyde (GA).

1.6.4 Transketolase-based biotransformations versus chemical processes

The fact that transketolase will act on such a wide range of substrates advocates its use in the synthesis of fine chemicals and pharmaceuticals. Biotransformations have several key advantages over traditional chemistry-based routes to the same compounds. These advantages are from an industrial point of view, low environmental impact and simple process steps are both favourable characteristics of biotransformations. However, the greatest driving force for the increased use of biotransformations in industry, particularly in the pharmaceutical sector, is the potential to produce optically-pure products. The Food and Drug Administration (FDA), and equivalent agencies around the world, now demand pharmacological and toxicity data for each enantiomer of a chiral drug (FDA, 1992). These regulations have provided pharmaceutical companies with a strong incentive to develop single enantiomer drugs: providing separate data for two enantiomers massively increases costs. Companies also have the possibility of redeveloping previously licensed racemic drugs as single enantiomers to extend their patent protection (Stinson, 2000).

Chemical processes such as aldol condensation have the potential to produce similar chiral compounds to transketolase (Trost and Fleming, 1991;

Seyden-Penne, 1995), but only when chiral auxiliaries are used. These auxiliaries are necessary because chemical syntheses are themselves fundamentally incapable of creating new stereocentres. The expense and limited variety of these reagents make large-scale chemical processes more costly and less flexible than equivalent biotransformations. Many examples of potential syntheses using transketolase have been demonstrated and are summarised in Table 1.4. Recent focus has shifted to the transketolase from *E. coli*, rather than the yeast TK, as it is a suitable candidate for large-scale production (French *et al.*, 1996; Hobbs *et al.*, 1996; Bongs *et al.*, 1997). The enzyme is tolerant to cosolvent and seems to accept generic aldehydes with full stereocontrol (Morris *et al.*, 1996). Technical advances involve operation under unbuffered conditions (Morris *et al.*, 1996) or continuously in a membrane reactor (Bongs *et al.*, 1997) whereas product removal in situ has been demonstrated via in-line borate complexation (Chaunhan *et al.*, 1996). More recently approaches have been developed combining enzyme engineering techniques and screening in microwell systems that mimic the ideal conditions experienced in industrial biocatalytic processes, these are now being used to exploit the asymmetric carbon–carbon bond forming enzyme transketolase (Lye *et al.*, 2002, 2003; Burton *et al.*, 2002).

Substrate Donor	**Acceptor**	**Product**	**Reference**
β-HPA	Glycolaldehyde	L-Erythrulose	Bolte *et al.*, 1987
[2,3-$^{13}C_2$]hydroxy-pyruvate	D-Glyceraldehyde	D-[1,2-$^{13}C_2$]xylulose	Demuynck *et al.*, 1990
β-HPA	Racemic 3-hydroxy-4-oxobutyronitrile	2-Deoxy-L-threo-5-hexulosonitrile	Effenberger and Null, 1992
β-HPA	D-Ribose	D-Sedoheptulose	Dalmas and Demuynck, 1993
β-HPA	4-Deoxy-L-threose	6-Deoxy-L-sorbose	Hecquet *et al.*, 1996
L-Erythrulose	2-Deoxy-D-erythrose-4-phosphate	4-Deoxy-D-fructose-6-phosphate	Guérard *et al.*, 1999
Xylulose-5-phosphate	D-[5-^{14}C, 5-^{3}H] ribose-5-phosphate	[7-^{14}C, 7-^{3}H]sedo heptulose-7-phosphate	Lee *et al.*, 1999
β-HPA	D-Glyceraldehyde-3-phosphate	D-Xylulose-5-phosphate	Zimmermann *et al.*, 1999
β-HPA	3-*O*-benzyl-glyceraldehyde	5-*O*-benzyl-D-xylulose	Humphrey *et al.*, 2000

Table 1.4 A selection of fine chemicals and drug precursors that have been synthesised using transketolase.

1.6.4.1 Transketolase as a directed evolution target.

TK is an ideal candidate for enzyme engineering since the high homology between species means that the diverse sub-specificities are due to a small number of variable amino acids. As a result saturation mutagenesis libraries are likely to have a significant impact on the specificity of the enzyme whilst also enabling the production of better quality libraries with less likelihood of redundancy of mutation and inactivation of the enzyme.

1.6.5 Inhibition of transketolase

1.6.5.1 Substrate binding analogues

Several substrate binding analogues have been studied previously, such as arabinose-5-phosphate which occurs naturally in *E. coli* and other organisms and is accepted as a substrate for yeast TK, but not *E. coli* where it acts as a competitive inhibitor. The *Ki for* D-arabinose 5-phosphate was found to be around 6 mM when studied at a concentration of 2.5 mM (Sprenger *et al.*, 1995). The effect of a natural analogue of transketolase substrates on the catalytic activity of the TK enzyme has also been investigated. Hydroxyphenylpyruvate demonstrates a reversible competitive inhibition of transketolase with respect to substrate, though it was also able to displace the thiamine diphosphate cofactor from holotransketolase. It is claimed by the authors that the data suggests that hydroxyphenylpyruvate participates in the regulation of tyrosine biosynthesis by influencing the catalytic activity of transketolase (Solovjeva *et al.*, 1999).

1.6.5.2 Competitive inhibitors (potential targets for anti-cancer compounds)

The transketolase enzyme reactions and other reactions of the pentose-phosphate pathway (PPP) allow glucose conversion to ribose-5-phosphate

used in nucleic acid synthesis, and also generates NADPH, a reducing agent required for biosynthesis reactions. Both of these products of the PPP are required for growing tumour cells. In addition, the non-oxidative part of the PPP, controlled by TK, allows anaerobic glucose degradation. As anaerobic conditions are often present in tumours, limiting the occurrence of these conditions can slow their growth. Tumours ferment glucose to lactate even in the presence of oxygen (so-called aerobic glycolysis, or the Warburg effect). The PPP allows glucose conversion to ribose for nucleic acid synthesis and glucose degradation to lactate. More recently, evidence has also emerged implicating a mutated TK transcript being upregulated in tumours. Transketolase like transcript-1 TKTL1 was found to be upregulated whereas TK and TKTL2, a second TK like transcript, were found not to be upregulated in colon and urothelial tumour tissues tested. Emerging evidence has also implicated TK and its transcripts in diabetes and neurodegenerative diseases (Langbein *et al.*, 2006; Johannes *et al.*, 2005). Clearly, TK and its mutant transcripts, and also other PPP enzymes have potential as anti-cancer targets to limit cell proliferation. Indeed, the nonoxidative PPP has been recently shown to contribute more than 70% of nucleic acid ribose synthesis in tumour cells (Du *et al.*, 2004); showing that TK could be targeted as a means to slow tumour growth/cell division.

Thiamine deficiency frequently occurs in patients with advanced cancer and therefore thiamine supplementation is used as nutritional support. Thiamine (vitamin B1) is metabolized to thiamine pyrophosphate, the cofactor of transketolase which controls the ribose synthesis necessary for cell replication. As a result TK is clearly a potential target for cancer treatment, but it is also important to determine whether the benefits of thiamine supplementation outweigh the risks of tumour proliferation. Recently oxythiamine (TK inhibitor) has been used to target pentose cycle causing in decreased non-oxidative pentose cycle flux which results in G1 cell cycle arrest. The treatment of cancers with TK inhibitors such as oxythiamine has shown promising results to date given that some data has been shown that oxythiamine causes a greater reduction in tumour TK than that of liver in animals tested thus far (Comýn-Anduix *et al.*, 2001). This is clearly important

as the liver is the most metabolically active tissue in the body and as a result would be susceptible to potential damage by anti-thiamine TK inhibitors. Many other compounds are also being explored and tested for their ability to disrupt TK and or the PPP pathway in tumour cells. For example, a plant isoflavonoid, Genistein, acts by targeting the nonoxidative steps of the pentose cycle, and Avemar (commercial name) is a complex mixture of biologically active molecules produced from fermented wheatgerm which has been shown to reduce tumour growth by up to 50% (Comýn-Anduix *et al.*, 2002). It has also been shown to have an inhibitory effect on metastasis formation. However a recent group (Du *et al.*, 2004) claimed all of these were either not highly specific for TK or weak inhibitors of TK. The group claims to have found two novel scaffolds after a high-throughput library screening using a fluorescent intensity (FLINT) assay, which show high selectivity, ideal physiochemical properties and low molecular weight. The authors believe that this work sets the stage for the development of TK inhibitors as anti-cancer treatments.

1.6.6 Transketolase and Disease

TK has been implicated in a number of diseases. For example, reductions in TK activity have been implicated in cases of Chronic Uremia (Sterzel *et al.*, 1977), similar to those caused by thiamine deficiency (beri beri). Modifications to TK appear in cultured dermal fibroblasts from Alzheimer's patients suggesting that this could be used as a marker for the disease (Schenk et al., 1998). Erythrocyte TK can also be used as an indicator for many diseases such as pernicious anaemia (Kjosen and Seim, 1977) or nutritional deficiencies. Two such are Korsakoff's psychosis or Wernicke's encephalopathy, both of which can arise from alcoholism. The causes are usually poor eating habits, but deleterious affects to binding ability of TK have also been implicated in these cases after prolonged alcohol exposure meaning that what would have previously qualified as normal thiamine intake is now insufficient as the ability to bind it is adversely affected (Blass and Gibson, 1977). This was also shown to be familial with alcoholic fathers and naïve

sons being tested for TK activity and both found to be lower than that of non-alcoholics. This finding suggests that the genetic makeup of alcoholics or those who are at risk of becoming alcoholic (e.g., sons of alcoholics who are still alcohol naive) might cause them to be more affected by thiamine deficiency than non-alcoholic's (Mukherjee *et al.*, 1987).

1.7 Thesis aims

The aim of this thesis is to develop assays as well as test novel substrates for existing TK activity and to then rescreen these against the libraries produced. The libraries produced will be by one of three methods: EpPCR, mutator strain (XL1-Red) or saturation site directed mutagenesis (SSDM). The SSDM libraries will either focus mutations based on phylogenetic selection (Cochran et al., 2006) or with structurally focused mutations (in collaboration with E. Hibbert). The aim is to then screen these libraries with model reagents to determine their quality as well as any potential changes (Chapter 4), and then to screen novel substrates (Chapter 5) which have some existing activity with novel assays developed in Chapter 3. It is hoped that these studies will pave the way to utilising the potential promiscuity of TK, and for the investigation of the substrate walking concept using the enzyme transketolase (TK).

Chapter 2 – Materials and methods

Materials and methods

The protocols described in this chapter are standard practices and not experimental. Experimental techniques are described in subsequent Chapters.

2.1 Materials

Unless otherwise stated, all materials were purchased from Sigma-Aldrich Company Ltd. All primers were purchased from Stratagene, Netherlands. Water was purified to 15MΩ.cm^{-1} resistivity using an Elix 5 water purification system (Millipore Corp.). TOPO-blunt PCR and PCR-Script SK vectors were obtained from Invitrogen. Oligonucleotide PCR primers were from Qiagen. Taq Polymerase and dNTPs were from Roche. The mutator strain XL1-red and DNA restriction enzymes were from Stratagene.

2.2 Preparation of buffers, media, and reagents

Unless otherwise stated, all buffers, media, and reagents were stored at room temperature.

2.2.1 Luria Bertani (LB) medium

LB medium was prepared by dissolving 10g.L^{-1} tryptone, 10g.L^{-1} sodium chloride, and 5g. L^{-1} yeast extract in pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7. The medium was sterilised by autoclaving.

2.2.2 LB agar

LB agar was prepared by adding 20g.L^{-1} select agar to LB medium. The agar preparation was sterilised by autoclaving.

2.2.3 SOC medium

SOC medium was prepared by dissolving 20g.L^{-1}. tryptone, 5g.L^{-1}. yeast extract, 0.58g.L^{-1} sodium chloride, and 0.19g.L^{-1} potassium chloride in pure water. After autoclaving, the following filter-sterilised solutions were added: 2mM glucose, 1mM $MgCl_2$, and 1mM $MgSO_4$.

2.2.4 Ampicillin

Ampicillin was dissolved in pure water to a concentration of 150g.L^{-1}. Stocks were sterilised by filtration and stored at -20°C.

Ampicillin was used at a concentration of 150mg.L^{-1} to select bacteria carrying the plasmid pQR791. Preparations containing this concentration of ampicillin are labelled Amp^+.

2.2.5 HPA synthesis

The titration can be done up to 200ml on the autotitrator (3.54g of Bromopyruvate) the reaction takes approximately 2-3 hours and 18ml of the LiOH solution is added per 100ml of BrPA. The solution can turn slightly yellow at this stage and 1.5g of activated (coarse particle) charcoal can be used to decolourise the solution and then filtered. The reaction mixture is then adjusted pH 5 by addition of a few drops of acetic acid.

The solution was then reduced to 10-20% of the original volume used a high vacuum Bucho Rotavapor (dry ice/acetone) at 30°C and stored in a water bath in the fridge O/N. The next day the solution was frozen overnight and thawed on ice to keep solubility of HPA as low as possible. The thawed solution is then filtered and the crystals washed with cold methanol (stored at -20). A vacuum drier is then used to remove any remaining water from the methanol wash product should be a very fine and light white crystals (slightly grey if traces of charcoal remain).

2.2.6 0.3M buffer (pH 7.5)

2.2.6.1 Tris.HCl

0.3M tris(hydroxymethyl)aminomethane (Tris) buffer with a pH of 7.5 was prepared by dissolving 38.05g.L^{-1} Tris hydrochloride and 7.08g.L^{-1} Tris base into pure water.

2.2.6.2 Hepes

0.3M N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (Hepes) buffer with a pH of 7.5 was prepared by dissolving 84.99g.L^{-1} Hepes into pure water.

2.2.6.3 Mops

0.3M 4-Morpholinepropanesulfonic acid (Mops) buffer with a pH of 7.5 was prepared by dissolving 69.36g.L^{-1} Mops into pure water.

2.2.6.4 Glycine-Glycine

0.3M Glycine-Glycine (Gly-Gly) buffer with a pH of 7.5 was prepared by dissolving 39.63g.L^{-1} Gly-Gly into pure water.

2.2.7 Standard cofactor solution

Standard cofactor solution was prepared by dissolving 0.0915g magnesium chloride hexahydrate (M_r = 203.3) and 0.0576g TPP (M_r = 460.8) into the standard Ketone (e.g. HPA) solution 4.5ml of pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7.5. The ketone-cofactor solution was topped up to 5ml with pure water and stored at 4°C (for no more than 24 hours).

2.2.8 Standard substrate solution

Standard substrate solution was prepared by dissolving the desired ketone at 300 mM e.g. 0.165g β-HPA (M_r = 109.99), 0.0915g magnesium chloride hexahydrate (M_r = 203.3) and 0.0576g TPP (M_r = 460.8) in 5ml of pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7.5.

3ml of ketone substrate solution was topped up to 6ml by addition of 3ml (150 mM 1:1 v/v dilution) of chosen buffer solution and stored at 4°C (for no more than 24 hours).

In the cases of pyruvate, propionaldehyde, benzaldehyde and p-anisaldehyde solutions were prepared from commercial solutions. All substrates were prepared to a final concentration of 50 mM (ketone-cofactor solution 1/3 of total reaction volume final [substrate] 50 mM), except benzaldehyde, hydroxybenzaldehyde and anisaldehyde, were used at 10mM final (after addition of relevant buffer 1:1 ratio). Such a concentration is greater than the K_m of *Eco*TK for the given substrate in most cases. Thus the rate observed is a fixed condition and not a direct measurement of K_m.

Final concentrations in $g.L^{-1}$ for each substrate are as follows: bromopyruvate (9.2515 $g.L^{-1}$ (M_r = 185.03)); benzaldehyde (1.0612 $g.L^{-1}$ (M_r = 106.12 d = 1.044)); GA (6.005 $g.L^{-1}$ (M_r (dimeric) = 120.1)); hydroxybenzaldehyde (1.2212 $g.L^{-1}$ (M_r = 122.12)); β-HPA (5.4995 $g.L^{-1}$ (M_r = 109.99)); anisaldehyde (1.3615 $g.L^{-1}$ (M_r = 136.15 d = 1.119)); fluoropyruvate (6.4 $g.L^{-1}$ (Mr = 128)); heptadienal (5.5 $g.L^{-1}$ (Mr = 110 d = 0.881)); hexadienal (4.88 $g.L^{-1}$ (Mr = 96 d = 0.871)); mercaptopyruvate ((7.1055 $g.L^{-1}$ (Mr = 142.11)) with DTT (0.1542 $g.L^{-1}$ (Mr = 154.2)); pyruvate (5.502 $g.L^{-1}$ (M_r = 110.04 d = 1.267)); propionaldehyde (2.904 $g.L^{-1}$ (M_r = 58.08 d = 0.798))

2.3 Standard procedures

2.3.1 Streaked agar plates

A culture was streaked out on a Petri dish of LB agar (Amp^+ if appropriate) using a wire loop. The plate was incubated overnight at 37°C and stored at 4°C.

2.3.2 Overnight cultures

A single colony was picked from a plate into 5ml of LB medium (Amp^+ if appropriate) in a 50ml Falcon tube. The tube was incubated for 16 hours at

37°C with 220rpm agitation.

2.3.3 Shake flask cultures

1ml of an overnight culture was added to 49ml of LB medium (Amp^+ if appropriate) in a sterile 500ml shake flask. The shake flask was incubated for 16 hours at 37°C with 220rpm agitation.

2.3.4 Glycerol stocks

A 20% (v/v) glycerol stock was prepared by adding filter-sterilised 40% (v/v) glycerol to an overnight culture in a one to one volume ratio. Aliquots were stored at -80°C.

2.3.5 Sonication

An open 1.5ml Eppendorf tube containing 1ml of culture was packed into a small beaker with ice. The probe of an MSE Soniprep 150 (Sanyo Europe Ltd.) was placed in the culture and the following program of sonication was used: eight cycles of 15 seconds on-time and 15 seconds off-time. The amplitude of sonication was 8Å.

2.3.6 Preparation of plasmid DNA

Plasmid DNA was prepared from an overnight culture using a QIAprep Spin Miniprep (QIAGEN Ltd.). The concentration of DNA in the product was determined by measuring its absorbance at 260nm (A_{260}). One A_{260} unit corresponds to $50ng.ml^{-1}$ dsDNA at neutral pH when a path length of 1cm is used (QIAGEN Ltd., 2001). Plasmid DNA was stored at -20°C.

2.3.7 PCR purification

Single and/or double PCR reaction products were purified using QIAquick columns (QIAGEN Ltd) the elution volume can either be 50 or 30 µl depending on concentration of PCR product required. The concentration of DNA in the product was determined by measuring its absorbance at 260nm

(A_{260}). One A_{260} unit corresponds to 50ng.ml^{-1} dsDNA at neutral pH when a path length of 1cm is used (QIAGEN Ltd., 2001). Plasmid DNA was stored at -20°C.

2.3.8 Gel extraction of DNA

0.6% low melting point agarose gels run in TAE buffer were used for extraction of vector or PCR products where necessary. DNA products were then purified using QIAquick columns (QIAGEN Ltd) the elution volume can either be 50 or 30 µl depending on concentration of DNA required. The concentration of DNA in the product was determined by measuring its absorbance at 260nm (A_{260}) and a UNICAM UV/VIS UV2 spectrometer. One A_{260} unit corresponds to 50µg.ml^{-1} dsDNA at neutral pH when a path length of 1cm is used (QIAGEN Ltd., 2001). Plasmid DNA was stored at -20°C.

2.3.9 Mutagenesis using the mutator strain

XL1-Red mutator strain of E.coli (Stratagene, Netherlands) was transformed using PQR791 plasmid using heat shock (section 2.3.13) and cultured as per standard technique, but for 60-120 generations in order to introduce mutations.

2.3.10 PCR and EpPCR protocols

A mini-prep supplied by Qiagen, of JM107 PQR791 stock with a His6 tag (stock provided by Jean Aufcamp) was prepared and was found to have a content of ~161ng.ml^{-1}. After extracting the PQR791 plasmid (TK his tagged for ease of purification) from glycerol stocks (J.Aufcamp) the optimal conditions for error prone PCR had to be found. Work previously done by A.Correa was used as a starting point for reaction conditions. A range of both manganese and magnesium concentrations was tested and the optical density and 'quality' of the products were compared using gel electrophoresis and a spectrophotometer. The DNA was then used for EpPCR reactions using the following cycles:

94°C 5minutes, 94°C 30seconds (30 cycles), 55°C 60 seconds (30 cycles), 72°C 60 seconds (30 cycles), 72°C for 5 minutes. The reactions are run in a Techne techgene PCR machine with the initial conditions as follows:

Component	Volume (µl)
10X taq buffer (-Mg)	10 µl
PQR711(-15)(His6)	0.3µl (50ng)
Tk -15 Mut F	1µl
Tk CTERM	1µl
dNTP mix	8µl (0.8mM final)
$MgCl_2$	4.0,4.4 and 4.8µl (1.0,1.1,1.2mM)
$MnCl_2$	10µl (1mM stock 0.1mM final)
Taq DNA polymerase	1µl (2.5 units)

Table 2.1 EpPCR reaction setup

A 0.6% agarose gel was then loaded with 1 µl of molecular weight markers (0.5-12kbs), and 3.6 µl of each sample (3 µl sample with 0.6 µl of sample buffer mixed on parafilm prior to loading).

The Experiment was then repeated with the volume of template increased to 0.6 µl and the volume of dNTP's reduced to 2 µl as excess was added previously. The samples were loaded onto a gel and imaged as previously. The next reaction was carried out without Manganese as shown below:

Components	Volumes added (µl)
10X Taq buffer without Mg2+	10µl
ddH20	81.8,81,80.2,79.4 and 78.6µl to samples to reach final volume of 100µl
Template	0.6µl
Primer 1 (CTERM)	1µl
Primer 2 (15 MUTF)	1µl
$MgCl_2$	2.6,3.4,4.4,5.2 and 6µl(0.6,0.8,1.0,1.2 and 1.4 mM)
dNTP's	2µl (0.8mM final)
Taq polymerase	1µl

Table 2.2 EpPCR reaction setup ($MgCl_2$ free)

The best conditions were selected from this reaction (1.0, 1.2 and 1.4 mM MgCl2) and run in duplicate with either 0.1 or 0.05 mM MnCl2 with a Manganese free sample of 1mM MgCl2 as a control. Subsequently the best conditions were found to be 1.0mM MgCl2 with 0.05mM MnCl2, 1.0mM MgCl2 with 0.1mM MnCl2.

2.3.11 Gel Extraction of Vector

The original method to mutate the TK gene tried was to perform EpPCR using the two primers Tk-15mut F and TKC-term. Then to do a double digest on a miniprep of the unmodified plasmid was performed using enzymes XbaII and BglII for 4 hours and then heat inactivated at 65°C and run on a 0.6% low melting point agarose. The gel the band corresponding to the vector (~3kb) was then gel extracted using a UV plate and a scalpel. However ligations and transformation efficiencies were very low as it was difficult to obtain the amount of vector required even using a Qiagen gel extraction kit even when doing multiple extractions.

2.3.12 Quick-change reactions

A standard QuickChange protocol was used to produce the libraries without much modification to the method required from primer to primer. Separate cycling of the sense and antisense primer strands, prior to a last combined QuickChange. Where the separate primer products are mixed seemed to have no significant improvement on the quality of the PCR product(s) observed. The factor which seemed to have the greatest impact on the quality of the reaction products was the concentration of the DNA template used. The PCRs for each Library site was set up in parallel with 3 replicates for each site being used: 5, 15 and 25 ng DNA.

Reactions were set up as follows:

10 µl of 10X *Pfu* buffer
1.5 µl of *Pfu Turbo*
2 µl of dNTP mixture (4mM)
1 µl of 5, 15 and 25ng/µl PQR791 stock solution
1 µl DMSO
82.5 µl ddH_2O
1 µl of 50mM forward and reverse NNS primer

After cycling the quality of the products was analysed using a 0.6% agarose.

2.3.13 Transformation by heat-shock

Only one strain of cells was routinely transformed in this way: *E. coli* XL1-Blue (supplied by Stratagene Ltd.).

The relevant competent cells were thawed on ice (XL1-Red or XL1-Blue) and an aliquot was transferred to a chilled 1.5ml Eppendorf tube. If necessary, 1 µl β-mercaptoethanol was added to increase efficiency and the cells were incubated on ice for 10 minutes. 1 µl of the relevant plasmid DNA was added to the cells and mixed gently. The transformation reaction was incubated on ice for 30 minutes. Heat-shock was performed in a 42 °C waterbath for a specified length of time (45seconds). The transformation reaction was then incubated on ice for a further 2 minutes. 0.5 ml of preheated

(42 °C) growth medium was added to the tube which was then incubated at 37 °C for 1 hour.

2.3.14 Transformation by Electroporation

The cell line used for transformation of library clones after growth O/N and DNA extraction from XL1-Blue cells were electrocompetent TOP10 cells (supplied by Invitrogen life technologies).

These cells must be thawed on ice; a microcuvette must also be pre-chilled with 5µl of chosen DNA ready for the cells to be transferred to. The aliquot of cells placed into the cuvette is 20-40µl (using a 0.1cm cuvette). With the 5µl of DNA present the total is made up to 100µl using deionised water to keep salt levels to a minimum to prevent arcing events. The contents of cuvette are mixed gently by shaking from side to side then placed on ice for 1 minute. Then any condensation is removed from sides of cuvette before electroporation (Biorad Genepulser). Immediately after pulsing the cells between 480-960 µl of SOC medium are added to cells and transferred to a 15ml falcon tube for 1hour incubation at 37oC 225rpm to enable antibiotic expression prior to plating on selective plates (amp+).

2.3.15 Measurement of absorbance and optical density

2.3.15.1 Cuvettes

Absorbance and optical density measurements were performed in a UV2 spectrophotometer (Unicam Ltd.). An ultra-micro quartz cuvette (Sigma-Aldrich Company Ltd.) was used when measuring absorbance in the ultraviolet region (200-400nm wavelength). Each sample was sufficiently diluted to register an absorbency (or an optical density) of ≤1AU (or ≤1ODU).

2.3.15.2 Microwell plates

Wells were scanned at 600nm wavelength in a FLUOstar Optima plate reader (BMG Labtechnologies GmbH.) to determine optical density.

2.3.16 HPLC system and methods

2.3.16.1 Determination of β-HPA and L-erythrulose concentrations by HPLC

This protocol was developed by Christine Ingram (Department of Biochemical Engineering, UCL).

2.3.16.2 Sample preparation

The reaction sample was diluted 1:1 by the addition of 0.2% (v/v) TFA for UV detection. The addition of acid quenched the reaction by dropping the pH to well below the lower limit for transketolase activity (Mitra *et al.*, 1998). The concentration of L-erythrulose in this quenched sample was determined by HPLC. The reaction sample was diluted 1:10 by addition of 0.1% (v/v) TFA for electrochemical detection (ECD).

2.3.16.3 HPLC system

The HPLC system consisted of an Endurance auto sampler (Spark Holland BV), a GP50 gradient pump (Dionex Corp.), an LC30 chromatography oven (Dionex Corp.), a PC10 pneumatic controller (Dionex Corp.), a ED40 electrochemical detector and an AD20 absorbance detector (Dionex Corp.). A chromatography workstation running PeakNet 5.1 (Dionex Corp.) was used to control the HPLC components and collect data from the relevant detector.

2.3.16.4 HPLC method

The mobile phase for the HPA-GA reactions was 0.1% (v/v) trifluoroacetic acid (TFA) and the flow rate was 0.6ml.min^{-1}. Separation of the components was achieved using a 300mm Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories). The temperature of the column was maintained at 60°C by the chromatography oven. The detector monitored the absorbance of the output stream from the column at 210nm wavelength (the range was 0.1AU).

The injection volume was 10µl for all samples.

2.3.16.5 Retention times and calibration curves

The retention times of β-HPA and L-erythrulose using this method are 8.32 and 11.32 minutes, respectively. Glycolaldehyde does not yield a peak of significant size at any concentration lower than 200mM. Standard curves were obtained for 0-25 mM β-HPA and L-erythrulose.

2.3.16.6 High-throughput HPA-GA HPLC method

This protocol was developed by Oliver Miller (Department of Biochemical Engineering, UCL).

The mobile phase was 0.1% (v/v) trifluoroacetic acid (TFA) and the flow rate was 1.5ml.min^{-1}. Separation of the components was achieved using a 50mm PL Hi-Plex H guard column (Polymer Laboratories Ltd.). The temperature of the column was maintained at 30°C by the chromatography oven. The detector monitored the absorbance of the output stream from the column at 210nm wavelength (the range was 0.1AU). The injection volume was 10µl for all samples.

2.3.16.7 Calibration curves

HPA and L-erythrulose passed through the column could be resolved: these compounds had retention times of 0.53 minutes and 0.66 minutes, respectively. Glycolaldehyde was found to have the same retention time as L-erythrulose, but its absorbance at 210nm wavelength was very low.

2.3.16.8 Determination of β-HPA, Propionaldehyde and PKD concentrations by HPLC

For the Propionaldehyde-HPA reaction analysis the mobile phase was an isocratic run with a flow rate of 0.5ml.min^{-1} of which 95% was 0.1% TFA and 5% 100% ACN solution. Separation of the components was achieved using an

ACE5 column. The temperature of the column was maintained at 30°C within the oven. The detector monitored electrochemical changes. PKD was chemically synthesised by Mark Smith (Dept. of chemistry).

2.3.16.9 Determination of Benzaldehyde and BKD concentrations by HPLC

An assay for the benzaldehyde reaction was established using an ACE C18 column, and also made use of the relatively strong UV absorbance of aromatic substrates/products. An example HPLC trace is shown in Figure 3.33. The flow rate was 0.6 ml/min for this method and the mobile phase was initially 85% 0.1% TFA and 15% 100% ACN with detection in the UV range at 225 nm. At 7 minutes a step change was made to 60% TFA and 40% ACN to help bring the substrates and product out of the column matrix. The initial mobile phase conditions (85% 0.1% TFA and 15% 100% ACN) were resumed at 15 minutes in preparation for the next assay/sample when the products/substrates of interest had been collected. The method is 20 minutes and would not be a feasible library screening method due to the impractical time that would be required to screen mutant library plates. However, it is likely to be useful for more accurate analysis of mutants of interest identified by high-throughput screening.

2.3.17 Thin layer chromatography

Thin-layer chromatography (TLC) is a chromatographic technique that is useful for separating organic compounds. Because of the simplicity and rapidity of TLC, it is often used to monitor the progress of organic reactions and to check the purity of products. Thin-layer chromatography consists of a stationary phase immobilized on a glass or plastic plate, and an organic solvent. The samples were loaded onto silica RP-TLC plates (PolyGram plastic sheets 0.2mm pre-coated with silica gel manufactured by Macherey-Nagel MN Germany) using a glass capillary tube. The sample spots were then dried using a heating plate or hot airgun (Black&Decker KX1600 1400W,

230V - 50HZ). The constituents of a biotransformation sample were made identifiable by simultaneously running standards with the unknown compounds (PKD and BKD chemically synthesised by Mark Smith Dept. of Chemistry). The bottom edge of the plate was then placed in a solvent reservoir containing ethyl acetate, which then moves up the plate by capillary action. When the solvent front reaches the upper edge of the stationary phase, the plate is removed from the solvent reservoir. Any excess solvent was then left to dry for a short time before placing the plate into staining reservoir. Phosphomobacetic acid (PMA) dip was used to stain the plate, which was then heated with the airgun to activate the stain and develop the plate.

2.3.18 UV absorbance screens

Reaction mixtures were set up with 200 µl sonicated and filtered cell lysate (JM107 pQR791), 200 µl of 150 mM HPA, 150 mM Tris buffer, pH 7.0, and cofactors, plus 200 µl of the relevant aldehyde substrate. Each aldehyde substrate was prepared to 10 mM final concentration, except heptadienal and hexadienal which had solubilities in buffer no higher than 3 mM (see section 2.2.8). A UV-visible spectrophotometer (Cecil Aquarius 7000 series) was used to perform wavelength scans on these substrates in the 250-600nm range. Wavelength scans were recorded one hour after the reactions were initiated with filtered cell lysate, and then further samples were taken after 24 hours and scans recorded to look for new peaks or measure any absorbance changes. Samples of 250 µl were taken and quenched with 250 µl of 0.2% (standard 1:1 UV assay dilution section 2.3.9) TFA solution prior to reading their absorbance in a quartz cuvette.

2.3.19 Agarose gel electrophoresis

A GNA-100 system (Amersham Biosciences Ltd.) was used for agarose gel electrophoresis of DNA. The amount of agarose used in a particular gel depended upon the desired linear range of DNA fragment separation: 0.7% (w/v) was used for 0.8–10.0kbp fragments (plasmid DNA) and 2.0% (w/v)

was used for 800–2000bp fragments (PCR products).

The appropriate amount of agarose was dissolved in 50ml of 1× TAE buffer (40mM Tris·acetate and 1mM EDTA in pure water) by heating in a microwave. 0.5mg.l^{-1} ethidium bromide was added to the gel to permit the visualisation of DNA by UV light. A comb was inserted at one end to form the sample wells. After the gel had set it was submerged in 1× TAE buffer in the gel tank and the comb was removed.

Samples were prepared for loading by adding 6× loading buffer. The wells of sample lanes were loaded with 3µl of sample. The wells of marker lanes were loaded with 1.5µl of Novagen 0.5–12.0kbp Perfect DNA Markers (EMD Biosciences Inc.) or Novagen 50–2000bp PCR Markers (EMD Biosciences Inc.). Electrophoresis was performed at 60V for 1 hour. The gel was visualised and photographed using a Gel Doc 2000 system (Bio-Rad Laboratories).

2.3.20 Polyacrylamide gel electrophoresis (DNA)

A Mini-Protean II system (Bio-Rad Laboratories) was used for polyacrylamide gel electrophoresis of DNA. Precast 4–20% gradient Tris·borate EDTA (TBE) gels (Bio-Rad Laboratories) were used for all analyses.

A precast gel was clipped into the gel cassette and submerged in 1× TBE buffer in the gel tank. Samples were prepared for loading by adding 6× loading buffer (Sigma-Aldrich Company Ltd.). The wells of sample lanes were loaded with 5µl of sample. The wells of marker lanes were loaded with 2.5µl of Novagen 50–2000bp PCR Markers (EMD Biosciences Inc.). Electrophoresis was performed at 80V for 2 hours. Following electrophoresis the DNA in the gel was stained by soaking in a solution of 0.5mg.l^{-1} ethidium bromide in 1× TBE. The gel was visualised and photographed using a Gel Doc 2000 system (Bio-Rad Laboratories).

2.3.21 SDS-PAGE

A Mini-Protean II system (Bio-Rad Laboratories) was used for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of proteins.

12.5% (w/v) acrylamide gels were used for all SDS-PAGE analyses.

2.3.21.1 Stock solutions

The acrylamide solution was 29.2% (w/v) acrylamide and 0.8% (w/v) N,N'-methylene bisacrylamide in water (Bio-Rad Laboratories). The separating gel buffer was 1.5M Tris buffer (pH 8.8). The stacking gel buffer was 0.5M Tris buffer (pH 6.8). The running buffer was 0.05M Tris·HCl, 0.38M glycine, and 0.1% (w/v) SDS in pure water, adjusted to pH 8.8. The staining solution was 0.05% (w/v) Coomassie Brilliant Blue, 50% (v/v) methanol, and 10% (v/v) acetic acid in pure water.

2.3.21.2 Gel casting

The gel cassette was assembled according to the manufacturer's instructions. 12.5% (w/v) separating and 6% (w/v) stacking gels were prepared according to Table 2.2.

The separating gel was poured first, overlaid with isopropanol to ensure a flat surface, and allowed to set. The solvent was carefully removed and the stacking gel was poured above the first gel. A comb was positioned in the stacking gel to form the sample wells. After polymerisation was complete, the comb was removed and the gel cassette was secured in the electrophoresis tank.

2.3.21.3 Sample preparation and running the gel

Samples were mixed with 2× Laemmli Sample Buffer (Bio-Rad Laboratories) and heated to 100°C for 2 minutes to denature the protein. The wells of sample lanes were loaded with 20μl of sample. The wells of marker lanes were loaded with 5μl of Precision Plus Protein Standards (Bio-Rad Laboratories). Electrophoresis was performed at 100V for 3 hours.

Component	Separating gel (ml)	Stacking gel (ml)
Acylamide solution	4.2	2.0
Stacking gel buffer	0.0	2.5
Separating gel buffer	2.5	0.0
10% (w/v) SDS solution	1.0	1.0
Pure water	2.3	4.5
10% (w/v) ammonium persulphate	0.1	0.1
TEMED	0.01	0.01

Table 2.3 Components of 12.5% (w/v) separating and 6% (w/v) stacking gel. Ammonium persulphate and TEMED (N,N,N',N'-tetramethylethyl-endiamine) were added to each gel at the last minute to initiate polymerisation.

2.3.21.4 Staining with Coomassie Brilliant Blue

Protein bands were visualised by staining with Coomassie Brilliant Blue.

The gel was placed in a plastic container, covered with 50ml of staining solution, and microwaved on full power for 3 minutes. The stain was poured away and the gel was destained by boiling for 10 minutes in 1 litre of pure water. The gel was photographed using a Gel Doc 2000 system (Bio-Rad Laboratories).

2.3.22 Protein Concentration assay

The dye reagent was prepared by diluting 1 part Dye Reagent Concentrate with 4 parts distilled deionised (DDI) water. It was then filtered through a Whatman #1 filter (or equivalent) to remove any particulates. This diluted reagent could then be used for approximately 2 weeks when kept at room temperature. Four dilutions of a protein standard, which is representative of the protein solution to be tested, were prepared. The linear range of the assay

for BSA is 0.2 to 0.9 mg/ml, whereas with IgG the linear range is 0.2 to 1.5 mg/ml. 100 µl of each standard and sample solution were combined in a clean, dry cuvette. In this case the standard used was BSA 0.2, 0.4, 0.6 and 0.8mg/ml. Protein solutions were assayed in duplicate and 5.0 ml of diluted dye reagent was added to each tube prior to vortexing. They were then incubated at room temperature for a minimum of 5 minutes. Absorbance increases over time; samples should incubate at room temperature for no more than 1 hour. When sufficient incubation time has passed the samples should be measured for absorbance at 595 nm.

The reading obtained from the protein assay can then be used to determine the total protein concentration within the sample using a BSA standard curve. Once the total protein content has been determined, the TK concentration can be found from its percentage weighting which was determined through the densitometry software.

2.3.23 Bioanalysis

Sample preparation was carried out as per Protein 200 plus assay kit instructions (Agilent technologies). 20 µl of each protein sample was combined with 10 µl of denaturing solution in a 0.5ml microcentrifuge tube. At the same time 30 µl of ladder solution was added to another 0.5 ml tube without denaturing solution. The tubes were then heated for 3 to 5 minutes at 95-100 °C using a heating block. The tubes were then vortexed for 15 seconds after 10 seconds cooling time and then 420 µl of deionised water was added to all tubes.

The Protein 200 assay chips were primed by loading 12µl gel dye to well 'G' in the top right hand corner of the protein 200 chips, use the syringe provided to pressurize the gel through the chip for 60 seconds, then remove any residual gel dye from well G. Now all the wells labelled G have 12µl of fresh gel dye added to them and 12µl of destain are loaded into well marked DS and 6µl of the diluted ladder solution are added to well with ladder symbol beside it. Now the chip is ready for loading and each sample (6µl per well) of interest loaded

alignment Editor (Hall 1999) was used.

2.4.3 Manipulation and imaging of protein structure

Protein structures were manipulated and visualised with Pymol v0.99 Pdb viewer (Delano 2006) which was also used to generate and render images produced.

2.4.4 Chemical reactions schemes

Chemical reaction schemes and all chemical compound drawings were done using ISISDraw 2.4 (MDL information systems 2001).

Chapter 3 – Substrate specificity

3.1 Introduction

3.1.1 Transketolase as a biocatalyst

In vivo transketolase (TK) catalyses the reversible transfer of a 1, 2-dihydroxyethyl ketol group from D-xylulose-5-phosphate to D-ribose-5-phosphate to yield D-seduheptulose-7-phosphate and D-glyceraldehyde-3-phosphate. In addition, it catalyses the reversible transfer of a 1,2-dihydroxyethyl group from D-xylulose-5-phosphate to D-erythose-4-phosphate to yield D-fructose-6-phosphate and D-glyceraldehyde-3-phosphate. In experimental and preparative condensations, β-HPA is frequently used to replace D-xylulose-5-phosphate as the donor substrate. The use of β-HPA renders the reaction effectively irreversible by virtue of the concomitant release of carbon dioxide as one of the two products (Datta & Racker, 1961).

The carbon-carbon bond formation that TK catalyses is both stereoselective, in that the enzyme has a preference for α-hydroxyaldehydes with (*R*)-configuration at C2, and stereospecific, in that the new chiral centre formed in the product has the (*S*)-configuration. Consequently, TK-mediated condensation of an α-substituted aldehyde with β-HPA produces enantiomerically pure chiral triols with D-threo stereochemistry (Kobori *et al.*, 1992). The best acceptor substrates for transketolase are α-hydroxyaldehydes with (*R*)-configuration, although α-oxo- and α-unsubstituted aldehydes are also accepted.

3.1.2 Previously studied aldehyde acceptors

Many examples of different aldehydes have been tested for their acceptance by TK. Table 3.2 summarises previously published data with relative rates compared to glyceraldehyde-3-phosphate (GAP) taken from five studies (Sprenger *et al.*, 1995, Dalmas and Demuynck., 1993, Kobori *et al.*, 1992, Demuynck *et al.*, 1991 and Bolte *et al.*, 1987).

Aldehydes

TK E.coli (using either HPA or X5P)

molecular formula	name	Conc. (mM)	*Vrel*	*Km* (M)	*Vmax* (U/mg)
	D,L-Glyceraldehyde-3-phosphate			2.1x10-3	100
	D-Erythrose-4-phosphate			90x10-6	>110
	D-Ribose-5-phosphate			1.4x10-3	50.4
	Glycolaldehyde			14x10-3	60
	D,L-Glyceraldehyde			10x10-3	25
	D-Erythrose			150x10-3	75
	D-Ribose			1.4x10-3	25
	Formaldehyde			31x10-3	12.5
	Acetaldehyde			1.2	20

Spinach TK (using HPA)

molecular formula	name	Conc. (mM)	*Vrel*	*Km* (M)	*Vmax* (U/mg)
	Glycolaldehyde	100	1		
	Acetaldehyde	100	0.12		
	Methoxyacetaldehyde	100	0.32		
	Methylthioacetaldehyde	100	0.35		
	Chloroacetaldehyde	50	0.23		
	Propionaldehyde	100	0.05		
	Methylthiopropionaldehyde	100	0.24		
	Methoxypropionaldehyde	100	0.31		

Yeast TK (with 20mM HPA)

molecular formula	name	Conc. (mM)	*Vrel*	*Km* (M)	*Vmax* (U/mg)
Non-Carbohydrate substrates R-GA					
H	Glycolaldehyde	20	1		
CH3		20	0.2		
CH3CH2		20	0.33		
CH3(CH2)2		20	0.22		
MeOCH2		20	0.27		
MeSCH2		20	0.33		
H2C=CH		20	0.56		
H2C=CHCH2		20	0.28		
(S)-H2C=CHCH0H(c)		20	0.36		
(R)-H2C=CHCHOH(d)		20	0.32		
(CH3)3C		20	0.11		
HO(CH2)2		20	<0.1		

Carbohydrate substrates			
	D-Glyceraldehyde-3-phosphate	20	0.44
	D-Erythrose-4-phosphate	20	0.33
	D-arabinose-5-phosphate	20	0.24
	D-Glucose-6-phosphate	20	0.09
	D-Glyceraldehyde	20	0.78
			<0.000
	L-Glyceraldehyde	20	1
	D,L-Glyceraldehyde	40	0.56
	D-Erythrose	20	0.56
	D-Glucose	20	0.04

Yeast TK (using HPA)

molecular formula (RCHO)	**name**	**Conc. (mM)**	***Vrel***	***Km* (M)**	***Vmax* (U/mg)**
H	Glycolaldehyde	100	1		
CH3	Acetaldehyde	150	0.25		
CH3OCH2	Methoxyacetaldehyde	100	0.32		
(CH3O)2-CH	Dimethoxyacetaldehyde	100	0.11		
(R,S) CH3-CHOH		60	0.44		
(R,S) CH2F-CHOH		60	0.47		
(R,S) CH3O-CH2-CHOH		80	0.2		
CH3-CO-	Pyruvaldehyde	100	0.19		
CH3-CH2-CH2-	Butyraldehyde	100	0.11		
(R,S)CH3-CHOH-CH2-		100	0.29		
(R,S)CH2OH-CHOH-CH2-		100	0.43		
(R) CH2OH-CHOH-CH2-		100	0.45		
(R,S) CH3-CH(OCH3)-CH2-	3-Methoxybutyraldehyde	100	0.3		
D-Erythrose		60	0.84		
L-Threose		60	0.39		
D-Ribose		150	0.3		
2-Deoxy-D-Ribose		100	0.16		
(CH3)2C=CH-	3-Methylcrotonaldeyde	100	0.11		
2,5-Dimethoxy-3-tetrahydrofuranyl		100	0.13		
Phenyl	Benzaldehyde	100	<0.1		
2-Hydroxy-phenyl		100	0.28		
2,3-Dihydroxyphenyl		100	0.32		
Benzoyl		150	0.11		
2-Pyrrolyl		100	0.21		
2-Furanyl		100	0.11		
2-Thiophenyl		100	0.32		
2-Pyridyl		150	0.13		
3-Pyridyl		150	0.13		

Yeast TK (using 1.25mM HPA)					
molecular formula	**name**	**Conc. (mM)**	***Vrel***	***Km* (M)**	***Vmax* (U/mg)**
	D-Glyceraldehyde-3-phosphate	5	1		
	D-Erythrose-4-phosphate	5	0.75		
	D-Glucose-6-phosphate	5	0.2		
	Glycolaldehyde	5	2.25		
	D-Glyceraldehyde	5	1.75		
	L-Glyceraldehyde	5	<0.01		
	D,L-Glyceraldehyde	5	1.25		
	D-Erythrose	5	1.25		
	D-Glucose	5	0.1		

Table 3.1 Kinetic parameters for previously published aldehydes with transketolases from Spinach, *E. coli* and Yeast (Sprenger *et al.*, 1995, Dalmas and Demuynck, 1993, Kobori *et al.*, 1992, Demuynck *et al.*, 1991 and Bolte *et al.*, 1987).

Interestingly, there are some differences between the various types of TK sourced from different organisms. For example, yeast TK is able to accept aromatic aldehyde substrates, whereas *E. coli* does not, and to date none of the types listed have been shown to have activity towards cyclic aldehydes. This demonstrates that the active site of TK is sufficiently exposed to accept large steric groups as substrates, at least in the case of aldehyde acceptors.

3.1.3 Previously studied TK inhibitors

3.1.3.1 Substrate binding analogues:

Several substrate binding analogues have been studied previously, such as arabinose-5-phosphate which occurs naturally in *E. coli* and other organisms and is accepted as a substrate for yeast TK, but not *E. coli* where it acts as a competitive inhibitor. The *Ki* for D-arabinose 5-phosphate was found to be around 6 mM when studied at a concentration of 2.5 mM (G.A. Sprenger *et*

al., 1995). The effect of a natural analogue of transketolase substrates on the catalytic activity of the TK enzyme has also been investigated. Para-hydroxyphenylpyruvate demonstrates a reversible competitive inhibition of transketolase with respect to substrate, though it was also able to displace the thiamine diphosphate cofactor from holotransketolase. It is claimed by the authors that the data suggests that p-hydroxyphenylpyruvate participates in the regulation of tyrosine biosynthesis by influencing the catalytic activity of transketolase (Solovjeva *et al.*, 1999).

3.1.4 Chapter aims

The areas that I aimed to explore (in Chapter 4) for extending substrate specificity beyond those currently accepted are: a) directed evolution library-type strategies; and b) the degree of change to the substrate that can be achieved using a single round of directed evolution. The latter knowledge could lead to the use of intermediate substrates as stepping stones towards a final target substrate in a process that is termed from this point onwards as 'substrate walking'. Therefore, this Chapter explores further the acceptance of alternative potentially useful ketols and aldehydes, to determine low levels of acceptance that would be useful to improve by directed evolution, and to identify non-accepted substrates that may form the basis of future 'substrate walking' concepts in directed evolution. This Chapter also describes some initial work that has identified one of the potential ketol donors as a TK inhibitor. This work was only partially explored as it was reaching beyond the intended scope of the thesis, and is currently being explored further elsewhere in a collaboration with the University of Exeter. Finally, during the studies detailed in this Chapter, a non-enzymatic mimetic of the transketolase catalysed reaction was serendipitously discovered. This mimetic reaction was characterised in collaboration with the Department of Chemistry at UCL. The initial characterisation is described in this Chapter to the extent of my involvement only, as the complete details also extend beyond the intended scope of this thesis.

3.2 Methods

3.2.1 Ketol donor activity assays

The six ketol donors in Figure 2 (HPA being the control substrate) were screened for activity using the HPLC assay, with TK containing sonicated lysates from the JM107 *E. coli* strain harbouring the plasmid pQR791 modified from pQR711 (French and Ward 1995). Reactions were performed for each ketol substrate at 50, 25 and 1 mM, added to buffer (50 mM Tris.HCl, pH7.0), 50 mM glycolaldehyde, 9 mM $MgCl_2$ and 2.4 mM TPP. The cell lysate sonicate was prepared and preincubated as per standard methods (section 2.3.5). Reactions were prepared as follows for a total volume 600 µl:

200 µl cell lysate sonicated
50 µl TPP/$MgCl_2$ stock (12X)
200 µl ketol/Tris.HCl buffer stock (3X)
100 µl ddH_2O
50 µl GA stock (12X)

Reactions were initiated by the addition of GA and after 1 and 24 hours samples were taken quenched with 1:1 0.2% TFA solution as per standard methods (section 2.3.16.2) prior to standard HPLC analysis (section 2.3.16.4).

3.2.2 Aldehyde donor activity assays

Each substrate was prepared at 10 mM, except for 2, 4-heptadienal and 2, 4-hexadienal which had solubilities in water no higher than 1 mM, and also for nitro-benzaldehyde which had a solubility in water that was too low for

detection. Different buffer systems were tested for aldehyde substrate screens including Hepes, Mops, Tris.HCl and Gly-Gly all at 50 mM pH 7.0. HPA was added to a final concentration of 50 mM (section 2.2.8), cofactor to final concentrations of 9 mM $MgCl_2$ and 2.4 mM TPP, and 200 µl filtered cell lysate (JM107 pQR791), for a final reaction volume of 600 µl (section 2.2.7). As these were novel aldehydes with little or no previous data for activity, samples were taken over a 24-48 hour time period for analysis as below.

3.2.2.1 UV-visible absorbance assay development

Standard solutions of each aldehyde substrate at 10 mM were prepared in quartz cuvettes, except heptadienal and hexadienal which were no higher than 1 mM due to their poor solubilities in buffer. Wavelength scans were recorded with a UV-visible spectrophotometer (Cecil Aquarius 7000 series) in the range 250-600 nm. For reaction mixtures, scans were recorded at one hour after the reactions were initiated with filtered cell lysate, and again after 24 hours. In each case, 250µl samples were taken and quenched with 250µl of 0.2% TFA solution prior to measurement of absorbance over the range 250-600 nm, in a quartz cuvette. Absorbance data was captured by a laptop computer using the Datastream software.

3.2.2.2 HPLC assay development

3.2.2.2.1 C18 ACE, UV210/270 nm methods

Samples (10 µl) were injected onto a C18 ACE column using a Dionex HPLC system. The mobile phase (95% TFA and 5% ACN) was pumped through the column isocratically at a flow rate of 1.5 ml/min. Peaks were detected using

UV absorbance (AD20, Dionex) at both 210 nm and 270 nm, or an electrochemical detector (ED40, Dionex).

3.2.2.2.2 Derivative assay development for propionaldehyde TK biotransformation

Each substrate and product solution was made up to 50 mM, and a 0.2% solution of 2, 4-dinitrophenyhydrazine (2, 4-DNP) in 2 M HCl (aq) was also prepared. 10 µl of each of the three 50 mM solutions were diluted in 990 µl of 0.1% TFA in H_2O. To each of these was added 10µl of the 2, 4-DNP solution. A control of just 10 µl of 2,4-DNP in 990 µl of 0.1% TFA was also made up. The four solutions were then analysed separately by HPLC (using the propionaldehyde UV C-18 method section 3.2.2.2.1) and also then as a mixture.

3.2.2.3 TLC assay development

Samples were loaded onto silica RP-TLC plates (PolyGram plastic sheets 0.2mm pre-coated with silica gel manufactured by Macherey-Nagel MN Germany) using a glass capillary tube. The sample spots were then dried using a heating plate or hot airgun (Black&Decker KX1600 1400W, 230V - 50HZ). The constituents of a sample were made identifiable by simultaneously running substrate standards in parallel lanes, including the appropriate chemically synthesised products PKD and BKD. The bottom edge of the plate was then placed in a solvent reservoir containing ethyl acetate, which then moves up the plate by capillary action. When the solvent front reached the upper edge of the stationary phase, the plate was removed from the solvent reservoir. Any excess solvent was then left to dry for a short time before placing the plate into staining reservoir. Phosphomolybdic acid (PMA)

dip was used to stain the plate, which was then heated with the airgun to activate the stain and develop the plate.

3.2.2.4 Colourimetric assay development

Colourimetric assay development was work done in conjunction with M.E.B Smith (Dept Chemistry).

The test reaction buffer (90 µl, 50 mM Tris-HCL, HEPES, or Gly-Gly, pH 7.0), was pipetted into each well of a fresh 96-well plate (plate A). MP-carbonate resin (Biotage) (20 mg) was then added to each well using a Radleys resin loader. Individual biotransformation reaction solutions (10 µl) were then pipetted into each of the 96 wells of this plate (A) and left to stand for 3 hours. During this time any excess HPA is adsorbed, and therefore, quenched by the resin. A second fresh 96-well plate (plate B) was loaded with 50 µl of the test buffer (50 mM, and pH 7.0) in each well. A 50 µl sample of the mixture was transferred carefully, without transferring any resin beads, from the wells in plate A to the wells in plate B, and 20 µl 2, 3, 5-triphenyltetrazolium chloride (200 mg of 2,3,5-TPTC in 100 ml of MeOH) added to each of the 96-wells of plate B. Plate B was then transferred to a plate reader (BMG-labtech, Fluostar). Using the plate reader autoinjector, 10 µl 3M NaOH was added into each of the 96-wells of plate B, and the plate shaken for 10 seconds, then left to stand for 1 minute before an OD measurement taken for each of the wells at 485 nm (50 flashes per well).

3.2.3 Inhibition assays

A standard TK cell lysate reaction conditions were used. However, FlPA to final concentrations of 50, 25, 10, 5, 1, 0.1, 0.5 or 0.01 mM, was added to the cell lysate after incubation with the cofactors, and incubated for a further 20 minutes prior to addition of the substrates. HPA was added to 50 mM, whereas the GA concentration was increased to 100 mM, The reactions were

initiated with the GA addition and 50 µl samples were taken after 1 and 24 hours, then quenched in 50 µl 0.2% TFA prior to analysis. The samples were analysed using the standard high accuracy HPLC and UV monitoring method (section 2.3.16.1).

3.2.4 Inhibition profile modelling

The model used (by Dr Bing Chen, Dept Biochemical Engineering, UCL) for HPA/GA – TK reaction is for a Ping Pong Bi Bi mechanism as follows:

$$V = \frac{V_{max} AB}{K_b + K_a B + AB + \frac{K_a}{K_{iq}} BQ + \frac{K_a K_{ib}}{K_{iq}} Q}$$

A – HPA concentration, B – GA concentration, Q – erythrulose concentration

Where: K_a is the HPA equilibrium constant, K_b is the GA equilibrium constant, K_{ib} is the GA inhibition constant, and K_{iq} is the erythrulose inhibition constant.

If FlPA is competitive with HPA the equation becomes:

$$V = \frac{V_{max} AB}{K_b + K_a B(1 + \frac{F}{K_{if}}) + AB + \frac{K_a}{K_{iq}} BQ + \frac{K_a K_{ib}}{K_{iq}} Q}$$

If FlPA is competitive with HPA and GA the equation becomes:

$$V = \frac{V_{max} AB}{K_b(1 + \frac{F}{K_{if}}) + K_a B(1 + \frac{F}{K_{if}}) + AB + \frac{K_a}{K_{iq}} BQ + \frac{K_a K_{ib}}{K_{iq}} Q}$$

Where K_{if} is the FlPA inhibition constant.

3.3 Results

3.3.1 Choice of novel ketol donors

Five previously uncharacterised ketol donors, shown in Figure 3.1, were selected to be screened for acceptance by TK, with glycolaldehyde (GA) as the acceptor aldehyde. HPA was also assessed for comparison.

Substrates:

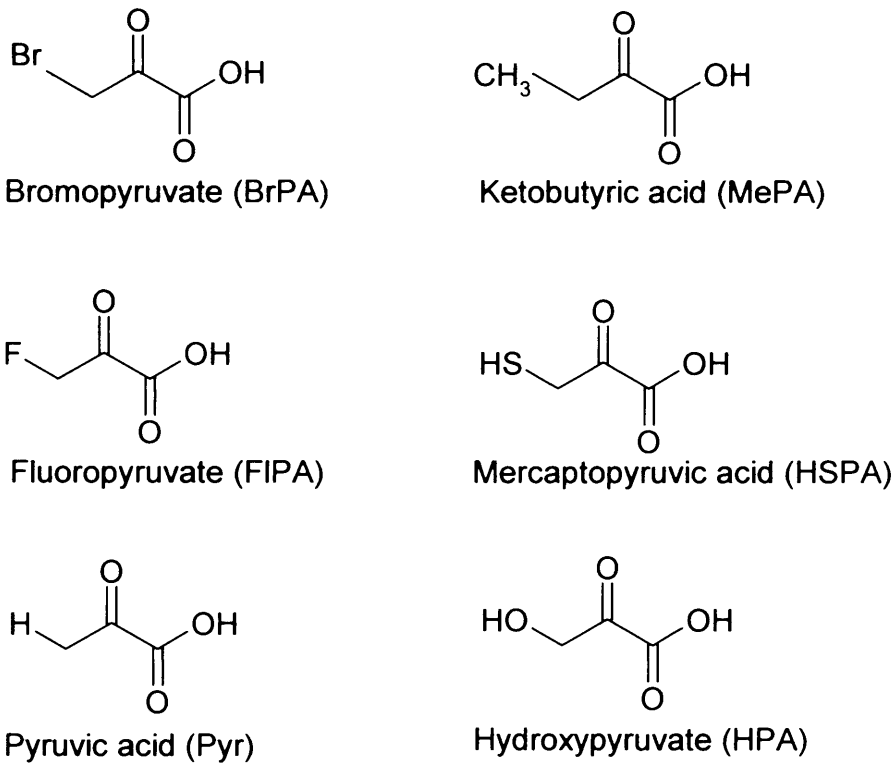


Figure 3.1 shows novel ketol donors chosen for testing for acceptance by TK, with GA as the acceptor aldehyde.

These ketol donors were chosen for analysis due to their potential synthetic usefulness, and also as they represent a chemical reactivity series that would potentially fall on a Hammett plot for the TK-catalysed reaction. Hammett plots demonstrate a linear free energy relationship (LFER) which is frequently used experimentally to study chemical reaction mechanisms. Hammett-plots

demonstrate a linear relationship between the logarithm of rate (or equilibrium) constants and σ-values which represent a series of substituent changes to one compound in a chemical reaction. Initial σ-values are obtained from the ratio of equilibrium constants for the substituted compound against a reference compound, in a representative reaction. σ-values for the ketol donors in Figure 3.1 were taken, by analogy, from acetate ester hydrolysis data as shown in Table 3.3 The electronic parameter, σ, was developed by Hammett. Hammett related the electronic effect of a substituent to the difference between log K_a of the substituted and unsubstituted benzoic. By definition, the σ value of hydrogen is 0. Since the logarithm of an equilibrium constant is proportional to the free energy change of the reaction (ΔG), σ values measure the free energy change caused by a particular substituent relative to hydrogen. Negative σ values indicate chemical groups that donate electrons favouring the neutral species and decreasing K, and positive σ values are chemical groups that withdraw electrons from the ring favouring the anions thereby increasing K

Chemical Group (RCOOH)	σ	Equivalent ketol donor abbreviation
$ClCH_2$	1.1	ClPA
$FlCH_2$	1	FlPA
$HOCH_2$	0.555	HPA
CH_3	0	Pyr
$BrCH_2$	-0.1	BrPA
$HSCH_2$	N/A	HSPA

Table 3.2 Potential ketol donor substrates for TK, and their sigma values.
Data sourced from Table 7-13 pp222 J.E Leffer & E. Grunwoald (reference 16).

Pyr>MePA>HSPA>OHPA>BrPA>ClPA>FlPA

$\xrightarrow[\sigma]{}$

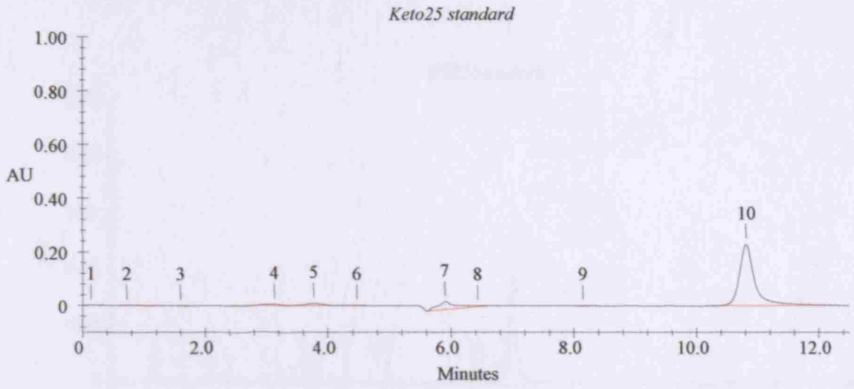
This series from left to right has an increasingly electronegative substituent group. This is expected to primarily affect their inherent chemical reactivity, as opposed to binding affinity to TK, and as a result is expected to affect k_{cat}. It should be noted that as HSPA data was not available, I have estimated its position in the series.

3.3.2 Assay development for activity towards novel ketol donors

HPLC assays needed to be developed for these novel substrates in order to measure their activities with TK. Initially HPLC conditions were identified which allowed good separation of the desired ketol donor standards.

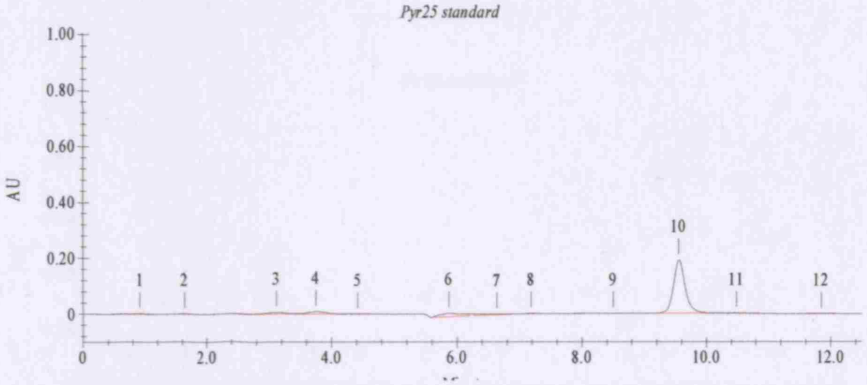
3.3.3 Ketol donor standards

Each ketol donor was initially resolved by HPLC to identify their retention times, as shown in Figures 3.2-3.7. Single peaks were resolved for all ketol donors except bromopyruvic acid (BrPA) and mercaptopyruvic acid (HSPA). Peaks were observed at 11 minutes for ketobutyric acid (MePA), 9.5 minutes for pyruvic acid (PA), 8 minutes for hydroxypyruvic acid (HPA), and 6.5 minutes for fluoropyruvic acid (FlPA). Bromopyruvic acid, however, showed a peak at 5.6 minutes for BrPA, but also a small peak at 8 minutes, presumably HPA formed by hydrolysis, and an unknown peak at 9 minutes, possibly pyruvic acid formed by elimination of bromide. A clear peak for mercaptopyruvic acid, either in the absence or presence of DTT, was difficult to determine.



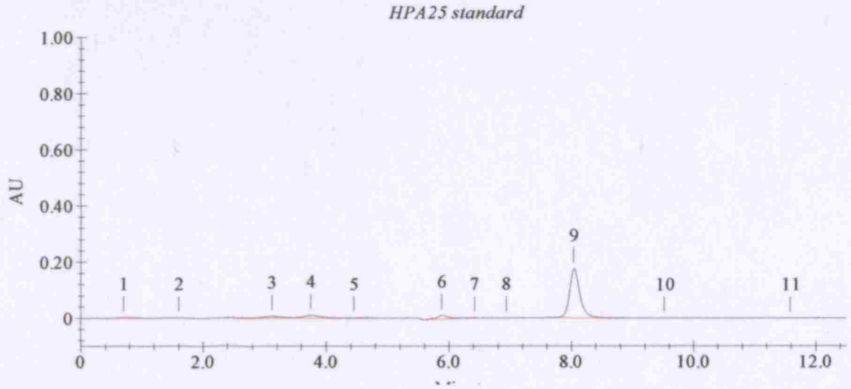
t (mins)

Figure 3.2 25 mM MePA in 50 mM Tris.HCl, pH7.0



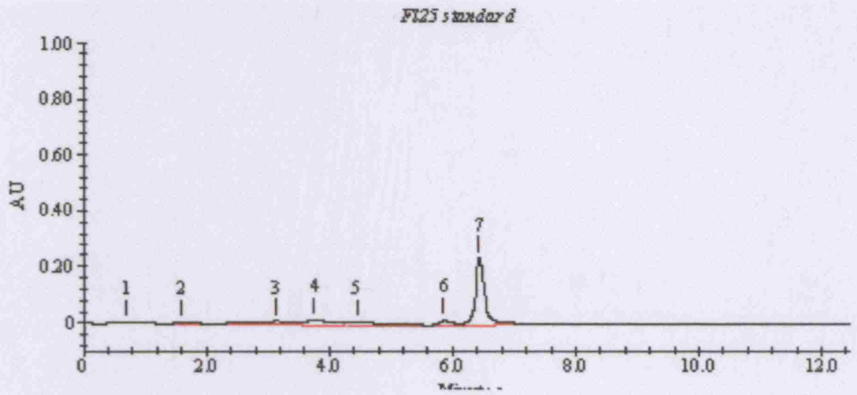
t (mins)

Figure 3.3 25 mM Pyr in 50 mM Tris.HCl, pH7.0



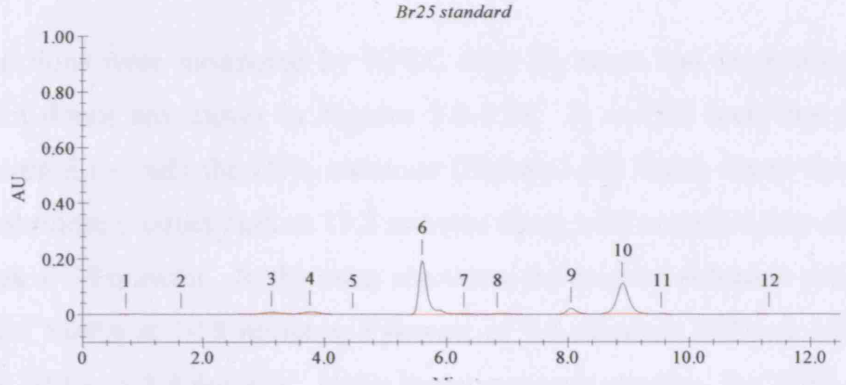
t (mins)

Figure 3.4 25 mM HPA in 50 mM Tris.HCl, pH7.0



t (mins)

Figure 3.5 25 mM FlPA in 50 mM Tris.HCl, pH7.0



t (mins)

Figure 3.6 25 mM BrPA acid in 50 mM Tris.HCl, pH7.0

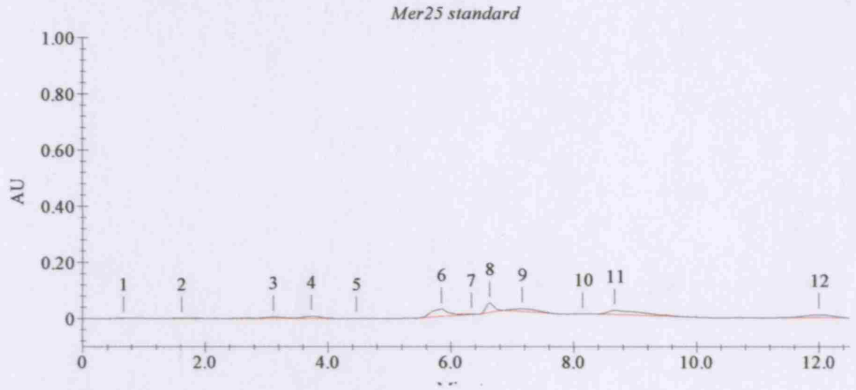


Figure 3.7 25 mM HSPA, in 1 mM DTT, 50 mM Tris.HCl, pH7.0

3.3.3.1 Activity of E.coli TK towards novel ketol donors

Reactions were monitored by HPLC after 24 hours and the results for each ketol donor are shown in Figures 3.8-3.14. It can be seen that a reaction occurred for only the HPA substrate (Figure 3.10) which forms the expected erythrulose product peak at 11.2 minutes along with complete loss of the HPA peak at ~8 minutes. In the other reactions, the original substrate peaks remain with: MePA at 10.8 minutes; Pyruvate at 9.8 minutes; FlPA at 8.2 minutes; and HSPA at 8.8 minutes. In the bromopyruvate reaction, the BrPA peak was still observed at 5.6 minutes along with a small erythrulose-like product peak at 11.2 minutes.

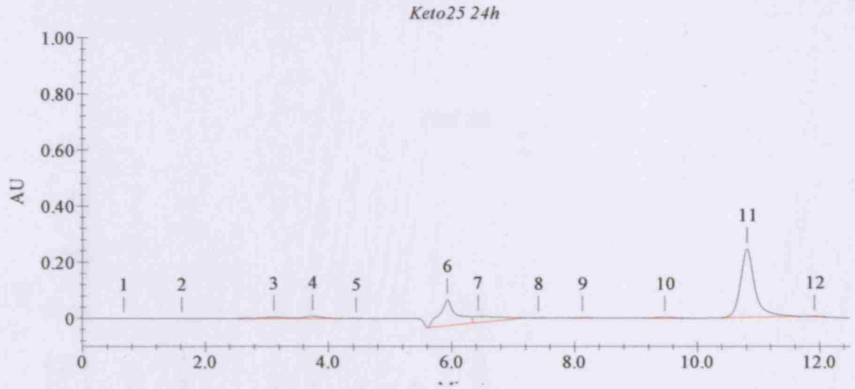


Figure 3.8 25 mM MePA, 50 mM GA in 50 mM Tris.HCl, pH7.0 after 24 hours reaction with 200 µl of crude TK lysate.

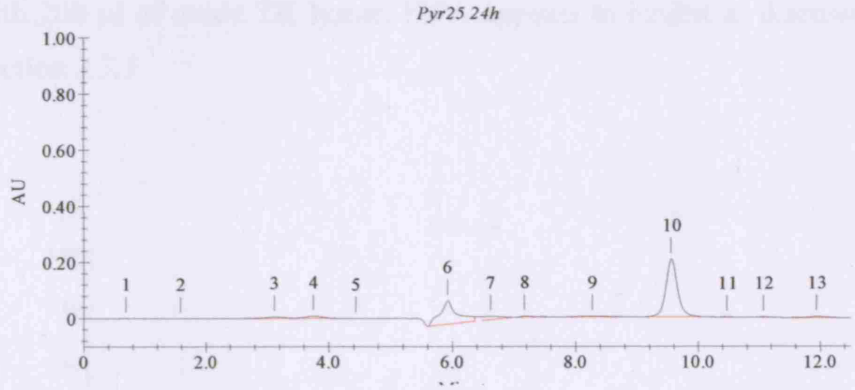


Figure 3.9 Pyr 25 mM, GA 50 mM pH7 50 mM Tris.HCl 24 hour reaction with 200 µl of crude TK lysate.

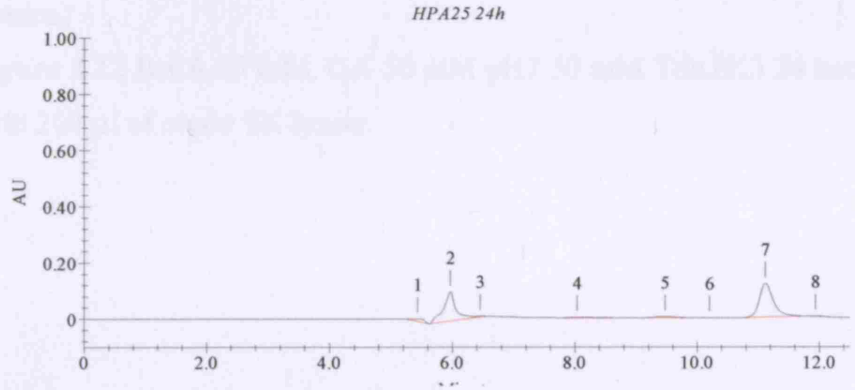


Figure 3.10 HPA 25 mM, GA 50 mM pH7 50 mM Tris.HCl 24 hour reaction with 200 µl of crude TK lysate.

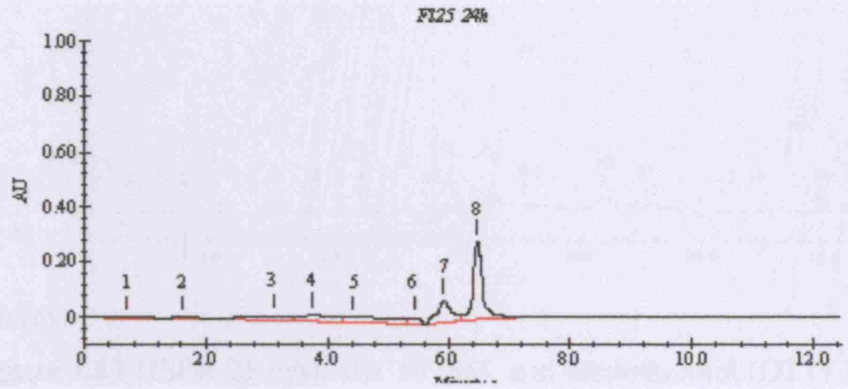


Figure 3.11 FlPA 25 mM, GA 50 mM pH7 50 mM Tris.HCl 24 hour reactions with 200 µl of crude TK lysate. FlPA appears to inhibit as discussed later in Section *3.3.5*

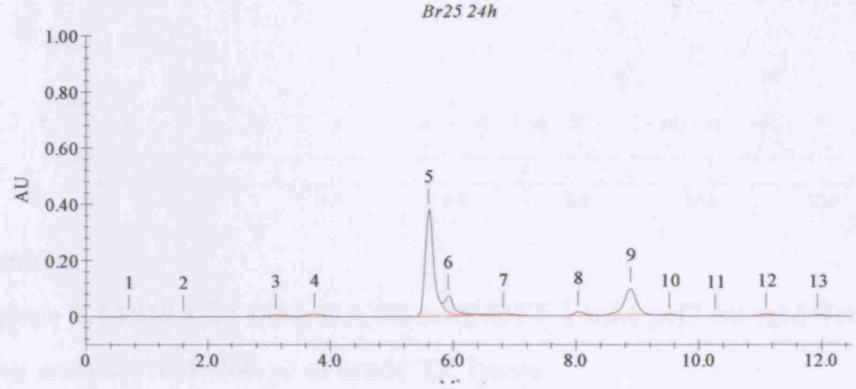


Figure 3.12 BrPA 25 mM, GA 50 mM pH7 50 mM Tris.HCl 24 hour reaction with 200 µl of crude TK lysate.

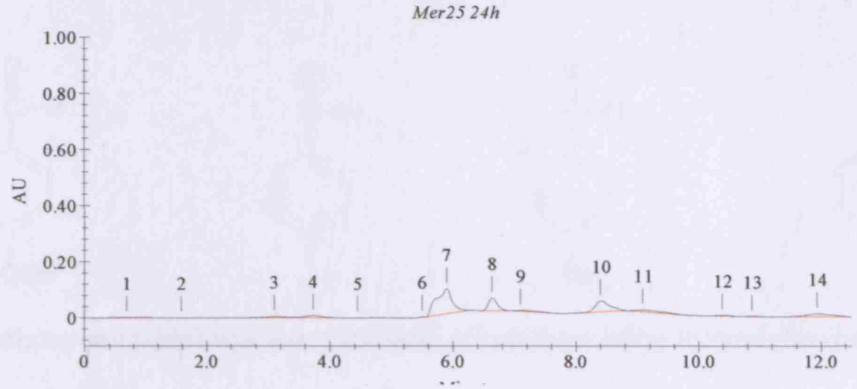


Figure 3.13 HSPA 25 mM, GA 50 mM, and dithiothreotol (DTT) 1 mM pH7 50 mM Tris.HCl 24 hour reaction with 200 µl of crude TK lysate.

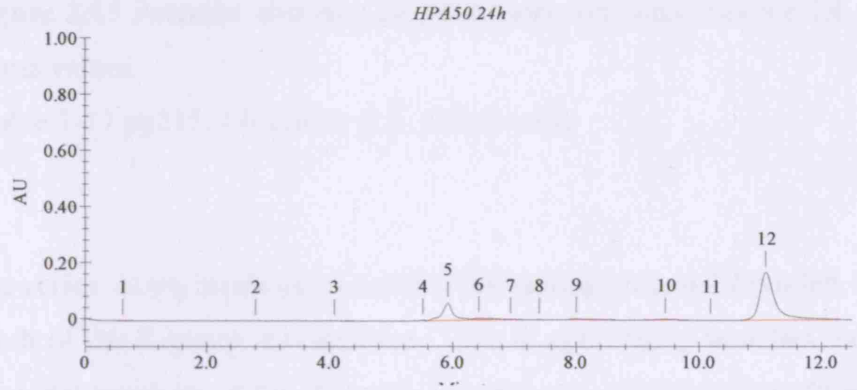
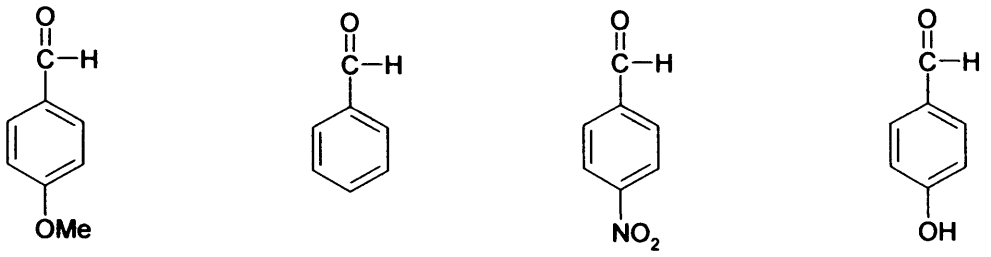


Figure 3.14 HPA 25 mM, GA 50 mM, DTT 1 mM pH7 50 mM Tris-HCl 24 hour reaction with 200 µl of crude TK lysate.

3.3.4 Choice of novel aldehyde acceptors

A number of previously unexplored aldehyde acceptors were also explored (Figure 3.15). These were picked as a potential Hammett series, as for the ketol donors, and also to investigate their varying properties including: a) solubility; b) reactivity; c) binding specificity



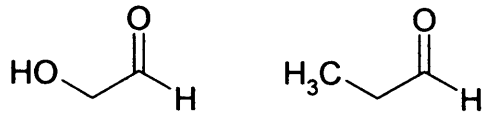
Methoxybenzaldehyde benzaldehyde nitrobenzaldehye hydroxybenzaldehyde

-0.12 0 0.82 0.98

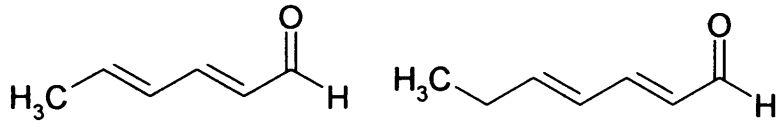
Hammett sigma values (σ)

Figure 3.15 Potential aromatic aldehyde acceptor substrates for TK, and their sigma values.
(Table 7-11 pp215, J.E Leffer & E. Grunzoald)

The series shows increased electron withdrawing potential from left to right as result of the R-group substitutions. This is again likely to affect the inherent chemical reactivity of the aldehyde acceptor, as well as binding affinity, and as a result is expected to affect k_{cat}, as well as K_m.



Glycoaldehyde Propionaldehyde



2, 4 -Hexadienal 2, 4-Heptadienal

Figure 3.16 Potential non-aromatic aldehyde acceptor substrates for transketolase.

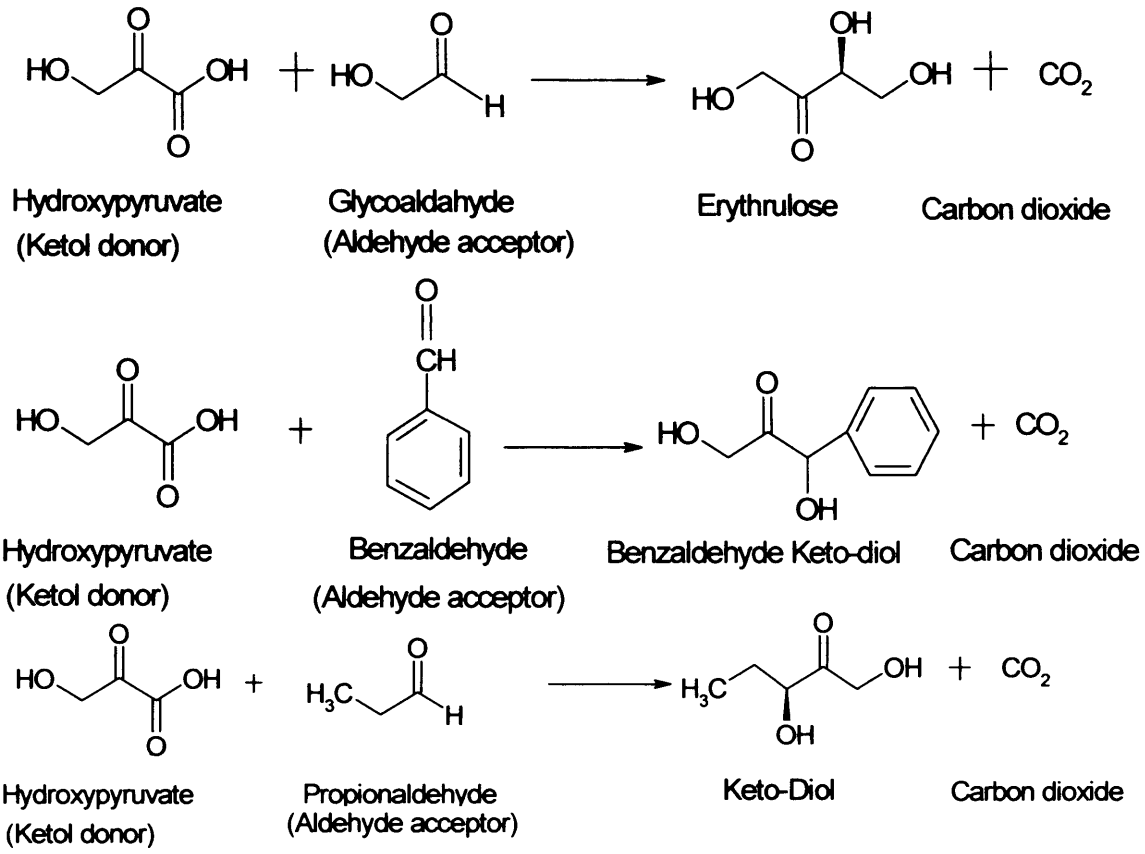


Figure 3.17 Reaction schemes for TK conversions of HPA with GA, PAL and BAL.

3.3.4 Assay development for activity towards novel aldehyde acceptors

3.3.4.1 UV-visible absorbance assay development

The initial objective was to identify potential spectroscopic changes for a high-throughput colourimetric assay. The spectra shown in Figures 3.18-3.22 for the potential aldehyde acceptor substrates are also useful for optimising HPLC detection wavelengths. A UV-visible spectrophotometer (Cecil Aquarius 7000 series) was used to perform wavelength scans on these substrates in the 250-600nm range. The solubility of each substrate to be tested was investigated initially. Also, as there was no previous inhibition data for each of these

substrates a wide range of concentrations at which reactions could be monitored was desirable so that conditions with no inhibition were more likely to be observed. However, the concentrations were also required to fall within the dynamic detection range for the spectrophotometer.

Reaction mixtures were set up (as described in Section 2.3.18) with 200 µl filtered cell lysate (JM107 pQR791), 200 µl of 150 mM HPA, 150 mM Tris buffer, pH 7.0, and cofactors, plus 200 µl of the relevant aldehyde substrate. Each aldehyde substrate was prepared to 10 mM final concentration, except heptadienal and hexadienal which had solubilities in buffer no higher than 3 mM, benzaldehyde and nitro-benzaldehyde for which the solubility in buffer was too low for detection. Wavelength scans were recorded one hour after the reactions were initiated with filtered cell lysate, and then further samples were taken after 24 hours and scans recorded to look for new peaks or measure any absorbance changes. Samples of 250µl were taken and quenched with 250µl of 0.2% (standard 1:1 UV assay dilution section 2.3.16.2) TFA solution prior to reading their absorbance in a quartz cuvette. Example spectra of each aldehyde monitored are shown in Figures 3.18-3.23 for initial substrate standards in buffer. Spectra recorded after reaction of each of these substrates in the presence of TK lysate and HPA, showed no noticeable change in the absorbance profiles and no altered absorbance readings even after 24 hours of incubation. Initial scans of benzaldehyde provided no significant UV absorbance traces. This was possibly due to benzaldehyde being to low in solubility and concentration in solution; and it formed an emulsion when efforts were made to produce solution at higher concentrations although addition of cell lysate appeared to increase solubility detection was still problematic. As a result of these problems UV absorbance assays of benzaldehyde were sidelined in favour of HPLC and TLC.

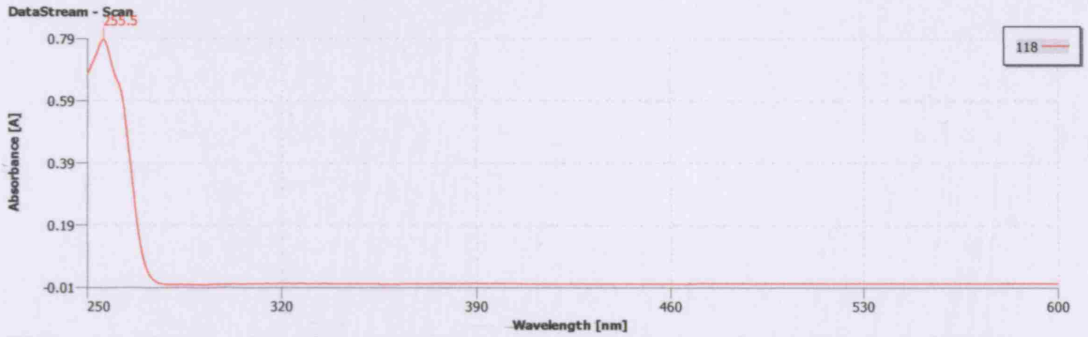


Figure 3.18 Wavelength scan of 4-methoxybenzaldehyde in 50mM Tris.HCl, pH 7.0

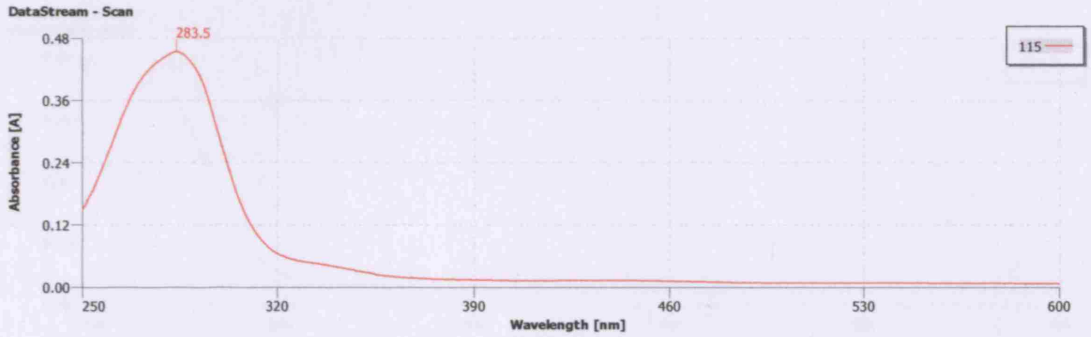


Figure 3.19 Wavelength scan of 4-hydroxybenzaldehyde in 50mM Tris.HCl, pH 7.0

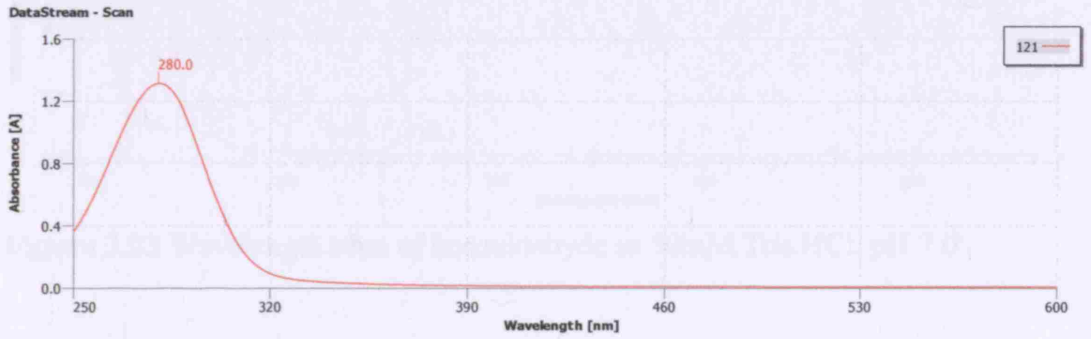


Figure 3.20 Wavelength scan of 2, 4-heptadienal in 50mM Tris.HCl, pH 7.0

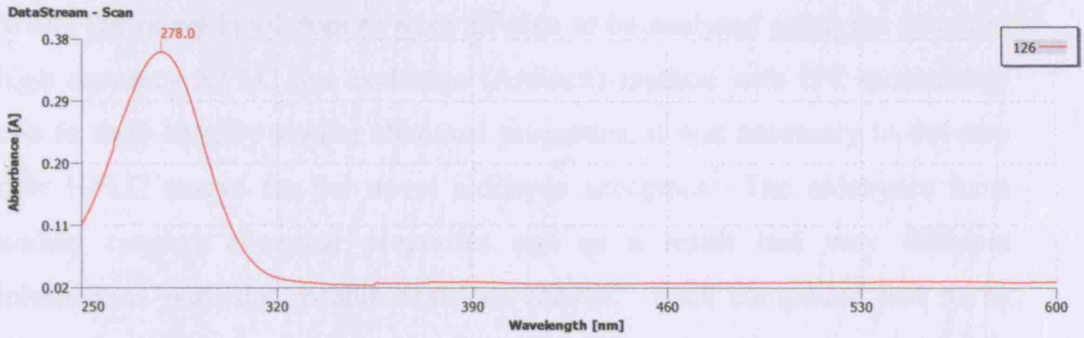


Figure 3.21 Wavelength scan of 2, 4-hexadienal in 50mM Tris.HCl, pH 7.0

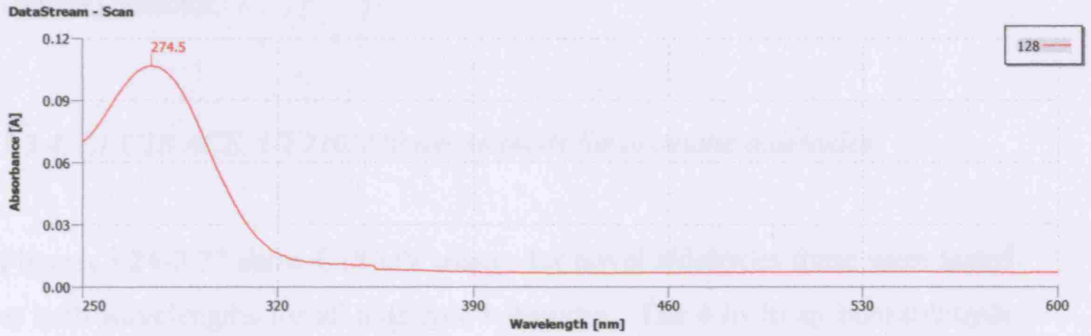


Figure 3.22 Wavelength scan of propionaldehyde in 50mM Tris.HCl, pH 7.0

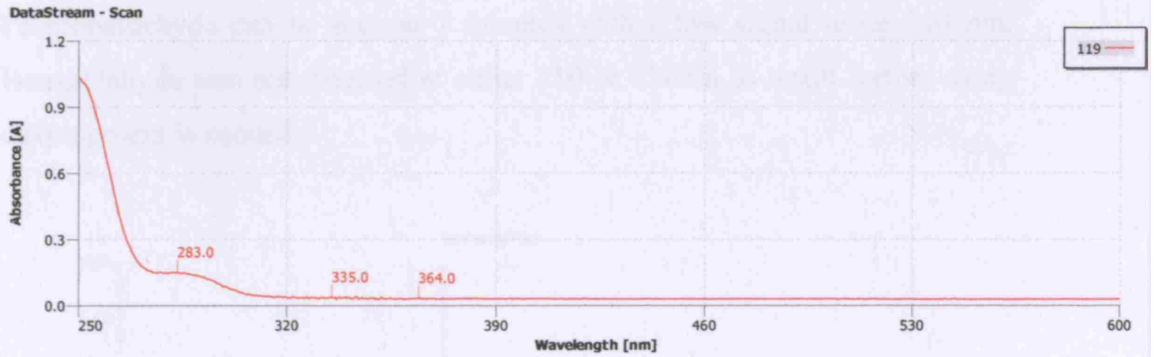


Figure 3.23 Wavelength scan of benzaldehyde in 50mM Tris.HCl, pH 7.0

3.3.4.2 HPLC assay development for novel aldehydes

While the novel ketol donors were all able to be analysed using the standard high accuracy HPLC ion exchange (Aminex) method with UV monitoring, due to their broadly similar chemical properties, it was necessary to develop new HPLC assays for the novel aldehyde acceptors. The aldehydes have widely ranging chemical properties and as a result had very different interactions with the column matrices chosen. Each compound had to be tested for retention times and column binding, as well as the method of visualising them. Visualisation by various methods was assessed, including: a UV-absorbance (AD20, Dionex) detector, both with and without derivatisation; and a standard triple waveform electrochemical (ED40, Dionex) detector.

3.3.4.2.1 C18 ACE, UV210/270 nm methods for aromatic aldehydes

Figures 3.24-3.27 show C18 UV assays for novel aldehydes these were tested at both wavelengths for all aldehyde substrates. The 4-hydroxy benzaldehyde peak can be seen at 17 minutes using the 210 nm assay. The 4-methoxy benzaldehyde peak can also be seen at 2.2 minutes using the 210 nm assay. Propionaldehyde can be seen at 7 minutes with a low signal using 210 nm. Benzaldehyde was not detected at either 210 or 270nm as result further assay development is needed.

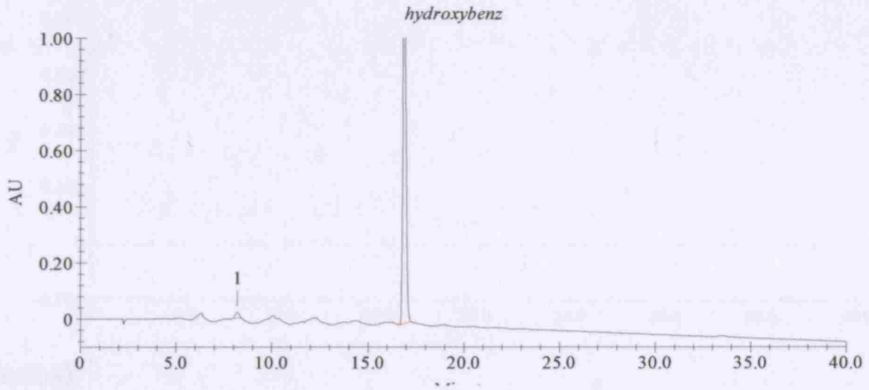


Figure 3.24 4-hydroxy benzaldehyde using a C18 column and absorbance at 210nm.

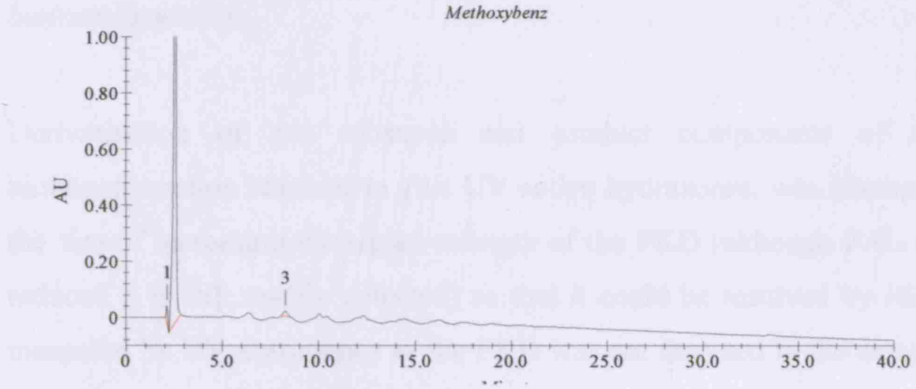


Figure 3.25 4-methoxy benzaldehyde using a C18 column and absorbance at 210nm.

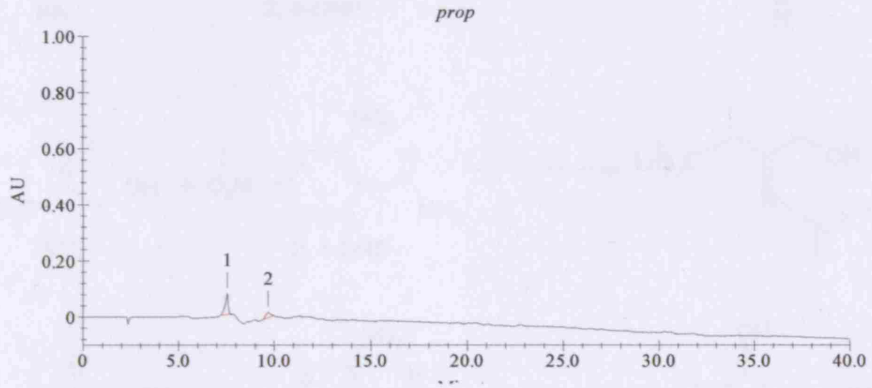


Figure 3.26 propionaldehyde using a C18 column and absorbance at 210nm.

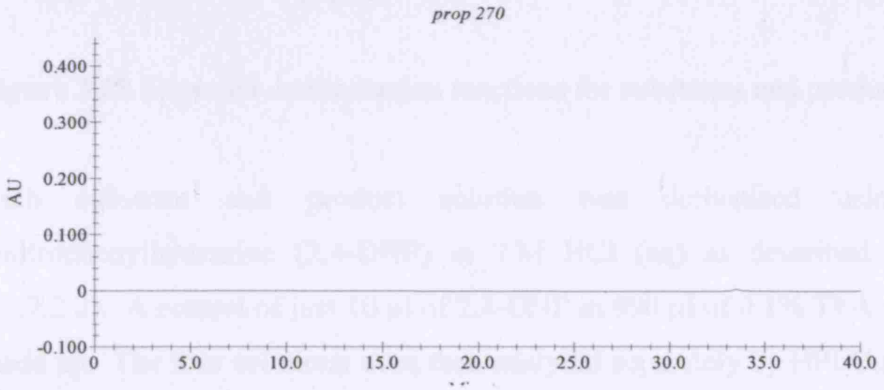


Figure 3.27 propionaldehyde using a C18 column and absorbance at 270nm.

3.3.4.2.2 Derivative assay development for propionaldehyde TK biotransformation

Derivatisation of the substrate and product components of the TK biotransformation reaction to give UV active hydrazones, was attempted with the aim of increasing the signal strength of the PKD (although PAL signal is reduced it is still readily detected) so that it could be resolved by HPLC and measured by UV absorbance as the PKD was not detected in the assays above only PAL was visualised.

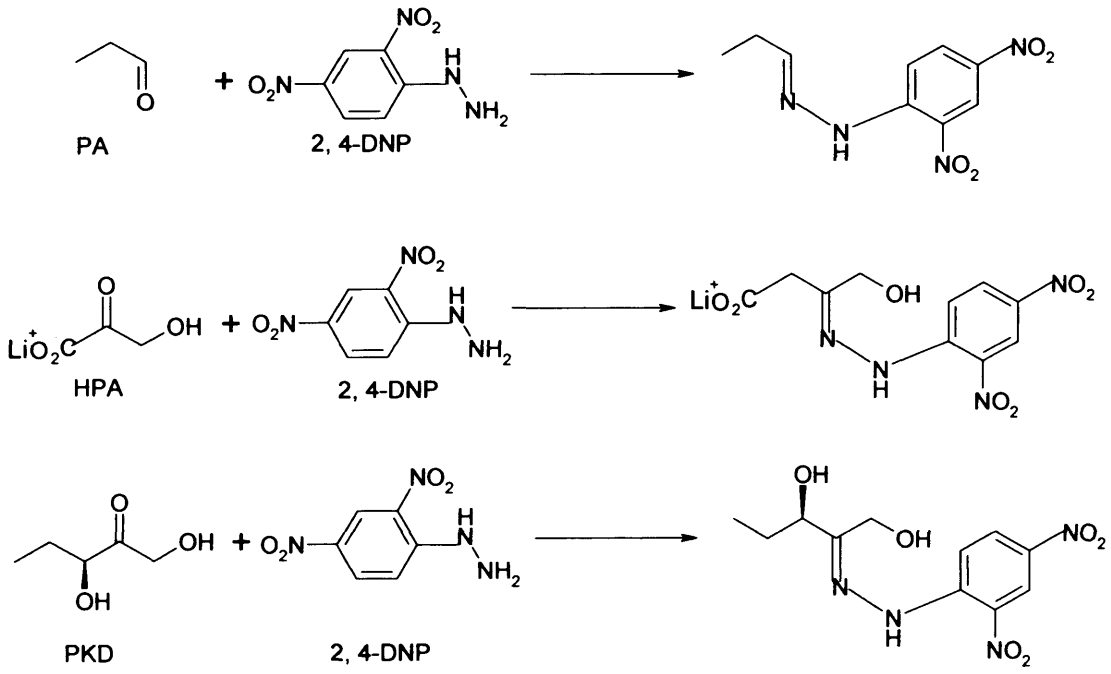


Figure 3.28 Expected derivatisation reactions for substrates and products.

Each substrate and product solution was derivatised using 2,4-dinitrophenylhydrazine (2,4-DNP) in 2 M HCl (aq) as described (section 3.2.2.2.2). A control of just 10 µl of 2,4-DNP in 990 µl of 0.1% TFA was also made up. The four solutions were then analysed separately by HPLC and then also as a mixture.

Figures 3.29-3.33 show derivatised samples for PA, HPA and PKD and their respective retention times. 2,4-DNP produces a peak at 9.6 minutes,

derivatised HPA has a peak at 9.8 minutes, PA has a peak at 17.2 minutes and PKD has a peak at 12 minutes. However mixing the separate solutions and performing a biotransformation appears to result in the loss of the PKD peak at 12 minutes.

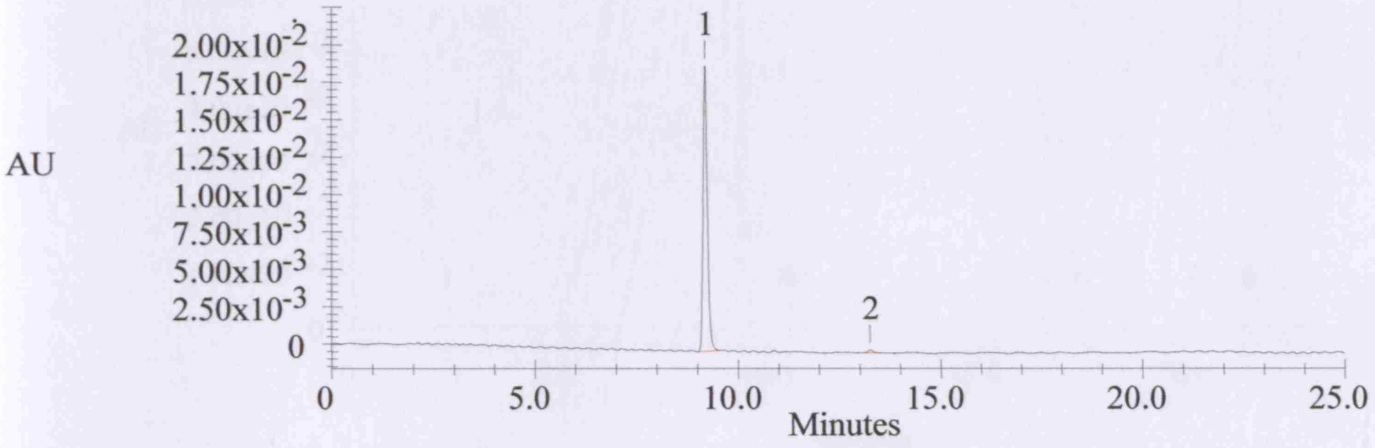


Figure 3.29 Solution 1: Reagent control with 2, 4-DNP only.

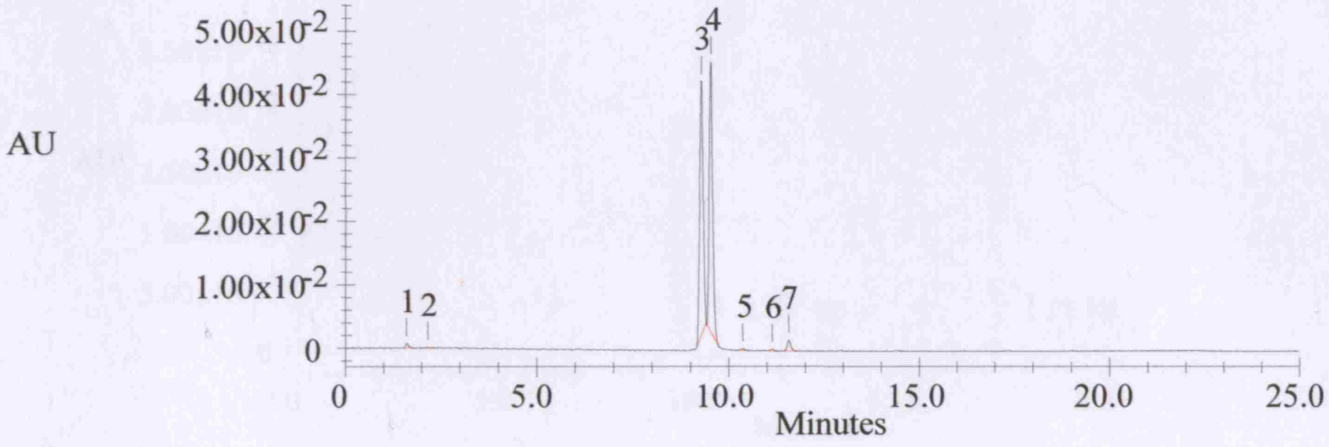


Figure 3.30 Solution 2: HPA and 2, 4-DNP.

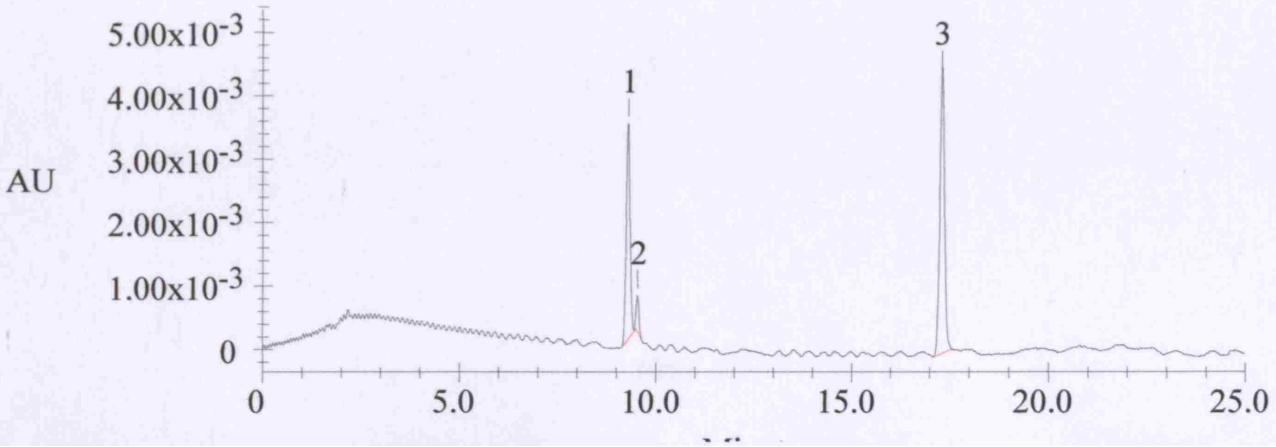


Figure 3.31 Solution 3: Propionaldehyde and 2, 4-DNP.

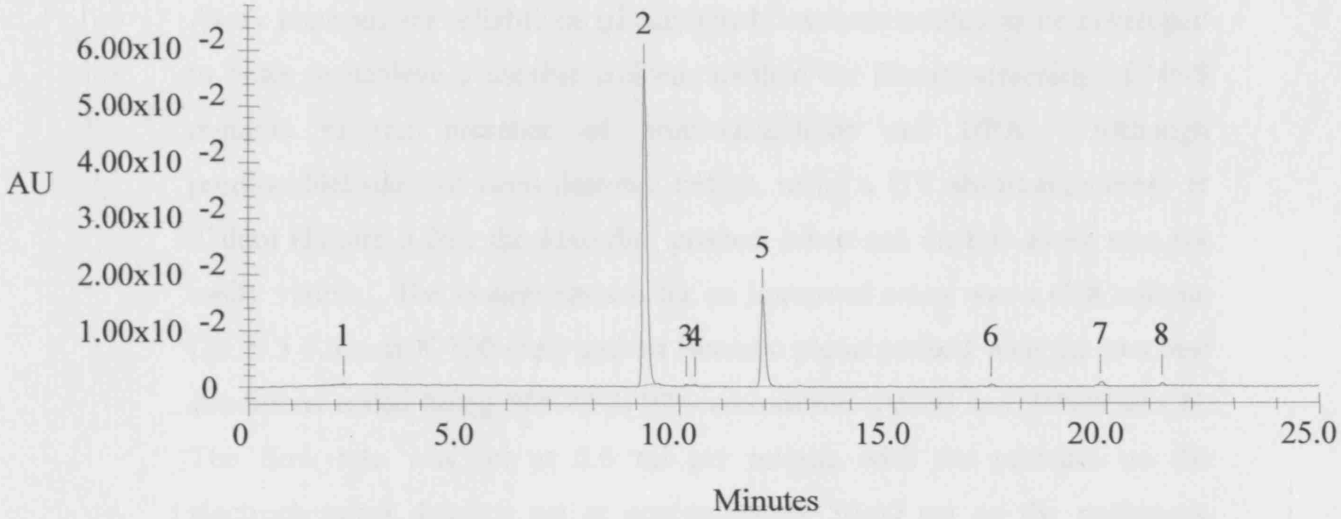


Figure 3.32 Solution 4: PKD and 2, 4-DNP.

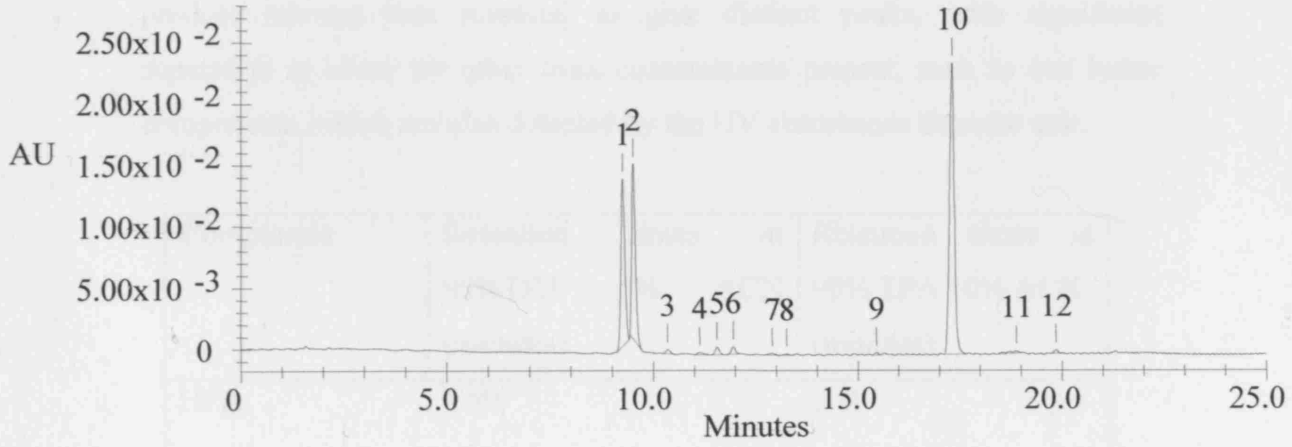


Figure 3.33 Mixture of solutions 2, 3 and 4.

3.3.4.2.3 ACE column (C18) electrochemical detection assay method for propionaldehyde

Assay methods for reliable propionaldehyde analysis needed to be developed in order to achieve a useable analysis method for library screening of NNS mutants in the presence of propionaldehyde and HPA. Although propionaldehyde had been detected before, using a UV absorbance assay at 210nm (Figure 3.26), the keto-diol product when run on this assay was not easily visible. The system chosen for an improved assay was a C18 column (ACE 5 4.6 mm X 150 mm) and an isocratic phase method with the two best conditions tested being 90%TFA/10% acetonitrile (ACN) and 95%/5%ACN. The flow rate was set at 0.6 ml per minute with the pressure on the electrochemical detector set at approximately 30-60 psi on the pneumatic controller, and with an initial method length of 30 minutes.

The percentage of ACN and TFA used will alter the retention times, and the aim was to develop a method where an HPA, propionaldehyde and keto-diol product mixture was resolved to give distinct peaks, with significant separation to allow for other trace contaminants present, such as cell lysate components, which are also detected by the UV absorbance detector unit.

Component	Retention times at 95%TFA 5% ACN (minutes)	Retention times at 90% TFA 10% ACN (minutes)
HPA	4.50	3.80
Keto-Diol	5.70	5.46
Propionaldehyde	3.40	3.62

Table 3.3 Changed retention times of propionaldehyde reaction substrates and products resulting from the alterations in mobile phase as a result of increasing ACN concentration.

The best assay method in terms of maximum separation of the individual components was 95%/5% TFA/ACN with the last peak being observed before 8 minutes. As such this assay could be used to screen a 96-well plate in approximately 13 hours, which is significantly faster than the ion-exchange Aminex column method, which takes over 24hours in total. Figures 3.34 and 35 show HPA-PAL and PKD calibration curves respectively.

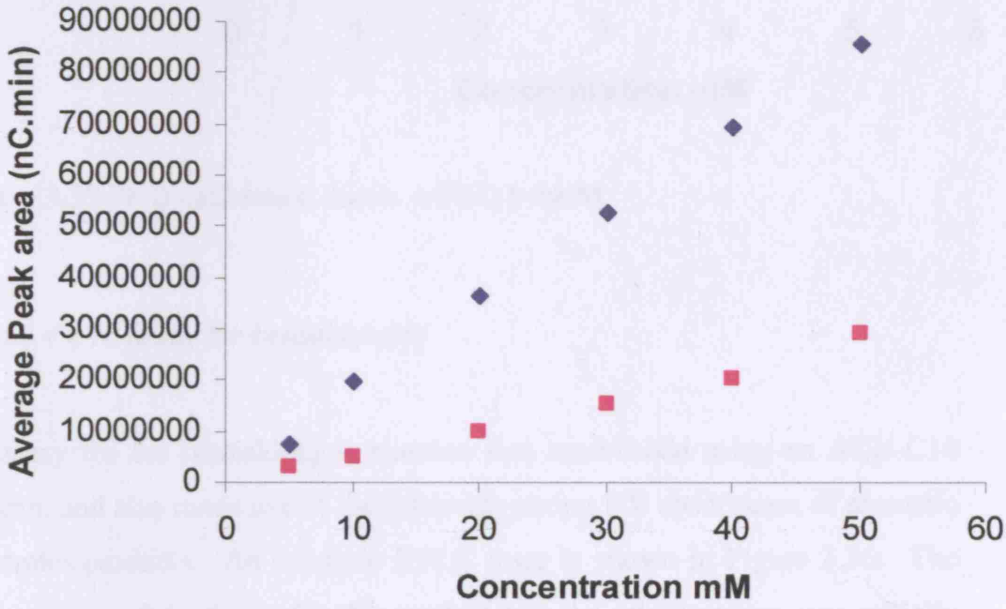


Figure 3.34 HPA and PAL calibration curve. PAL ■ HPA◆

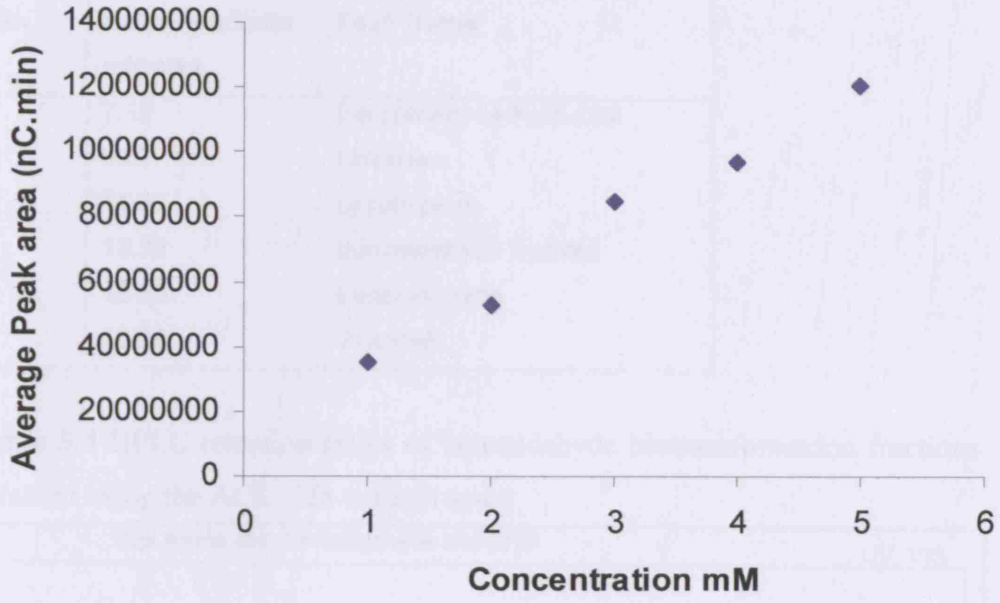


Figure 3.35 PKD calibration curve. ◆ PKD **1-5mM**

3.3.4.2.4 C18 assay for benzaldehyde

An assay for the benzaldehyde reaction was established using an ACE C18 column, and also made use of the relatively strong UV absorbance of aromatic substrates/products. An example HPLC trace is shown in Figure 3.36. The flow rate was 0.6 ml/min for this method and the mobile phase was initially 85% 0.1% TFA and 15% 100% ACN with detection in the UV range at 225 nm. At 7 minutes a step change was made to 60% TFA and 40% ACN to help bring the substrates and product out of the column matrix. The initial mobile phase conditions (85% 0.1% TFA and 15% 100% ACN) were resumed at 15 minutes in preparation for the next assay/sample when the products/substrates of interest had been collected. The method is 20 minutes and would not be a feasible library screening method due to the impractical time that would be required to screen mutant library plates. However, it is likely to be useful for more accurate analysis of mutants of interest identified by high-throughput screening.

Peak No.	Retention Time minutes	Peak Name
1	7.12	Benzaldehyde Keto-Diol
2	9.67	Unknown
3	11.57	Lysate peak
4	13.30	Benzaldehyde hydrate
5	15.20	Benzaldehyde
6	28.08	Unknown

Table 3.5 HPLC retention times of benzaldehyde biotransformation fractions obtained using the ACE C18 column assay.

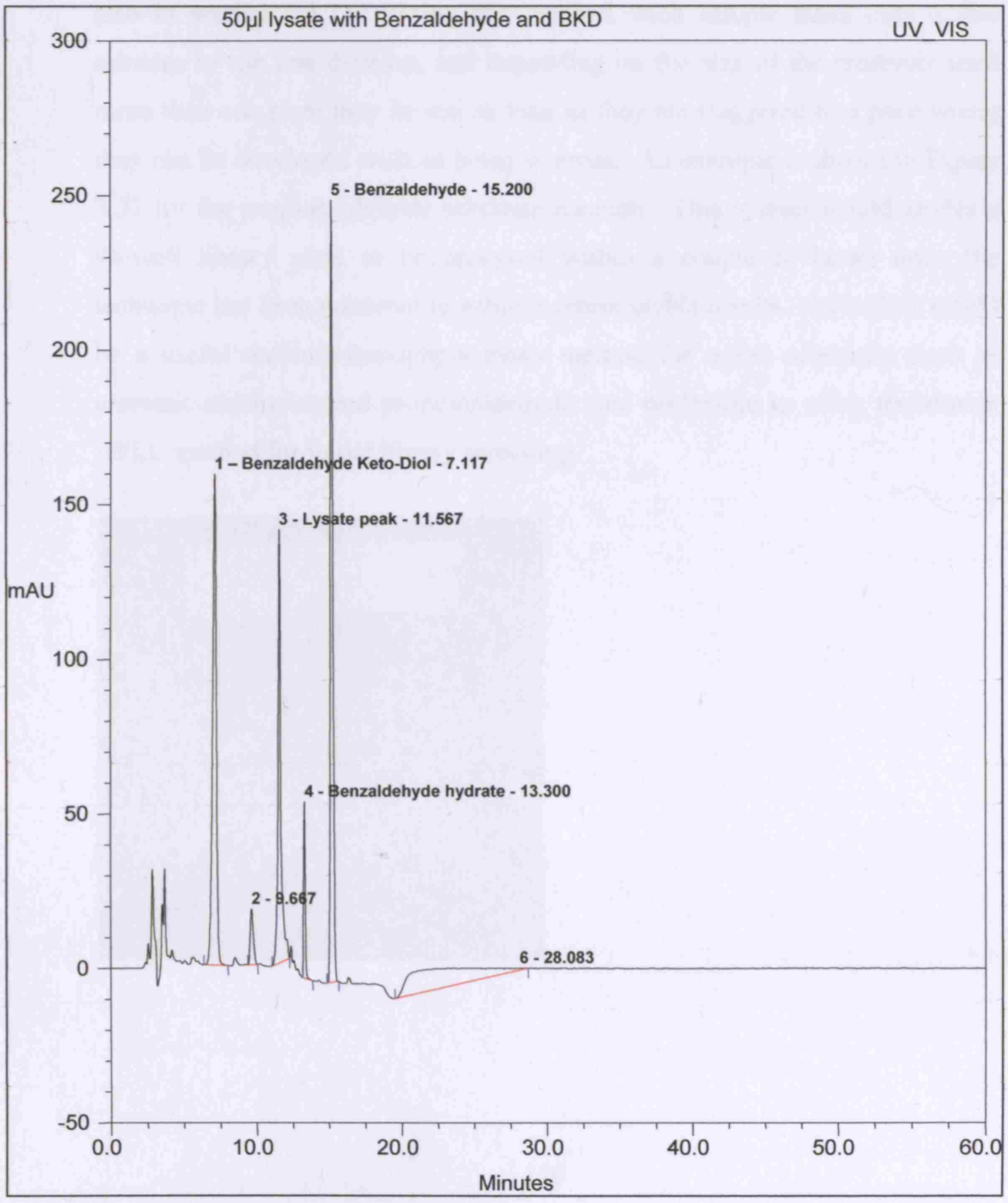


Figure 3.36 Example of the HPLC assay for the benzaldehyde biotransformation using an ACE C18 column. Peaks identified by comparison of retention times of standard solutions.

3.3.4.6 TLC assay development for novel aldehydes

Thin-layer chromatography (TLC) (section 3.2.2.4) was used as an alternative chromatographic technique for separating and analysing reaction substrates and products at medium throughput. Biotransformation reaction sample spots were added to TLC plates using glass capillaries, and developed as described, also in section 2.3.17. Using this method, each sample takes only a few minutes to run and develop, and depending on the size of the reservoir used more than one plate may be run as long as they are staggered to a pace where they can be developed without being overrun. An example is shown in Figure 3.37 for the propionaldehyde substrate reaction. This system would enable a 96-well library plate to be analysed within a couple of hours once the technique has been mastered to achieve reproducible results. As such it would be a useful medium-throughput assay method for novel substrates such as aromatic aldehydes and propionaldehyde, and preferable to using the slower HPLC method for initial library screening.

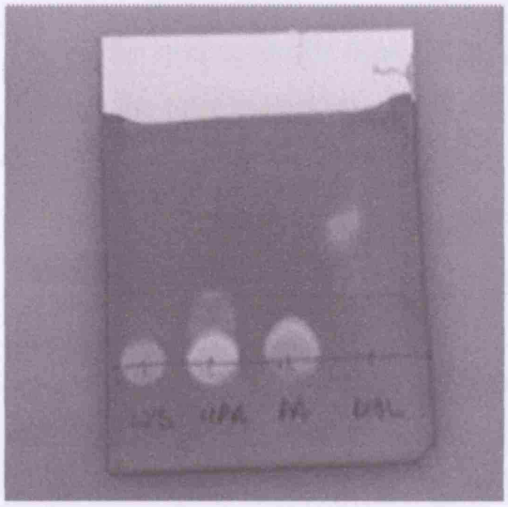
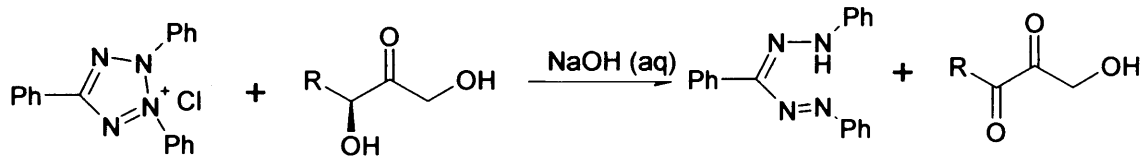


Figure 3.37 TLC assay establishing substrate and product signals for the PAL reaction: lysed cell control spot (LYS); 50 mM HPA substrate (HPA); 50 mM PAL (PA); ~10 mM keto-diol product (DIOL).

3.3.4.7 Colourimetric assay development

The colourimetric assay was developed in collaboration with Dr Mark Smith (Dept Chemistry, UCL). During development the assay had to be tested for any false positive interaction with all sample components, buffer and substrates. The assay, based on the reaction shown below, was initially developed to screen for propionaldehyde bioconversion, but if successful it could be used for screening of other novel substrates. A key requirement was to quench any excess HPA prior to the addition of colour reagents. The quaternary amine ion exchange resin used was shown to quench HPA efficiently.



The buffer system results from the next section (Section 3.3.4.8) were used to develop the following assay which appeared to be viable in either Tris.HCl or Gly-Gly buffer system. The assay described in detail in Section 3.2.2.4, was initially developed on the basis of the propionaldehyde bioconversion, and cannot be used for α-hydroxyaldehydes, such as glycolaldehyde.

An example of the colourimetric assay as it appears in the microtitre plate wells for the keto-diol product (PKD), formed from HPA and propionaldehyde, is shown in Figure 3.38A. The corresponding calibration curve, as measured by absorbance at 485 nm in a platereader, is shown in Figure 3.38B.

A

B

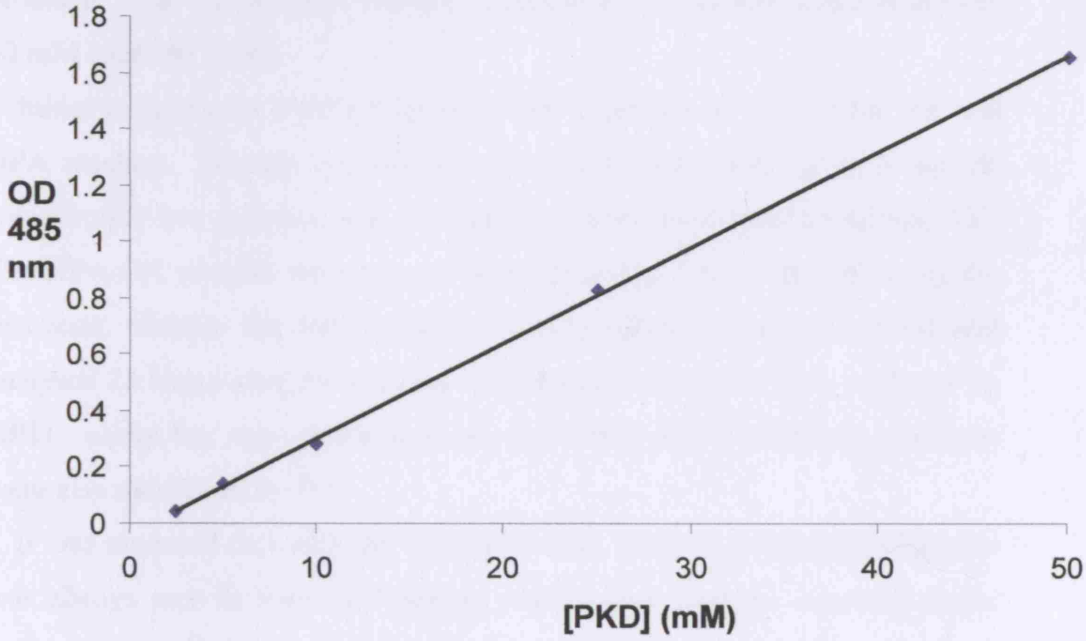


Figure 3.38A&B. Colourimetric assay calibration by absorbance at 485nm, using 50 mM Tris.HCl and pH7.0. (**A**) Photographic image of the colourimetric signal at different concentrations of PKD product. (**B**) Calibration curve for PKD as measured at 485 nm using a plate reader.

3.3.4.8 Exploration of reaction buffers

During assay development for the novel aldehyde substrates it was necessary to evaluate the effect of different buffer systems on TK activity and peak position, this was important not only for colourimetric assay development, but also for other potential screening techniques such as HPLC and TLC assays. Also, the buffers needed to be favourable for substrate and product stability and to prevent possible by products or side reactions. The four buffers Tris.HCl, Hepes, Gly-Gly and Mops, were investigated. These all have buffering potential (*pKa*) optima at around pH7-8, but their effects on the control sample, and also the novel substrates, reactions and products needed to be determined to identify any potentially unwanted side reactions or by-products. The buffers were assessed independently at a final concentration of 50 mM (Section 2.2.6).

Initial experiments were set up using HPA and GA as well as the PA and HPA reaction. Enzyme-free control experiments were also set up alongside some buffer-free samples, and also samples with lysed cells containing TK. The HPA-GA samples were quenched and analysed 1 hour after initiating the reactions, whereas the HPA-propionaldehyde reactions were quenched and analysed 24 hours after the reaction was initiated. Samples were analysed by HPLC using the appropriate method, and HPA-propionaldehyde reactions were also monitored by TLC.

It was observed that with the Tris.HCl buffer systems, the propionaldehyde was always seen to form two distinct peaks rather than the expected single peak. It was initially thought that the propionaldehyde may be interacting with this buffer and/or causing it to break down or react in some way. It was also possible that the propionaldehyde was reacting with the primary amines in the Tris.HCl buffer. The second observation was that controls of Mops and Hepes both appeared to have either erythrulose or PKD formation in the enzyme-free wells this was observed by TLC, HPLC and colorimetric assay turning red. Initially the assumption taken was that some contamination of controls had occurred and as a result the experiment was repeated. However, the unexpected products were seen once more despite taking even more care to prevent contamination of control wells. The next experiment set up was a simple screen of all four buffers against HPA and GA each at 50 mM, and

with the buffers at pH7.0. The reactions were initiated with a sample being taken at 1 hour and 24 hours after initiating the experiment. The samples were then analysed using the Aminex column (Section 2.3.16.5) and as previously observed, reactions for both Hepes and Mops containing samples appeared to have proceeded in the absence of TK, as shown below for the HPA-GA reaction (Figures 3.39-3.42).

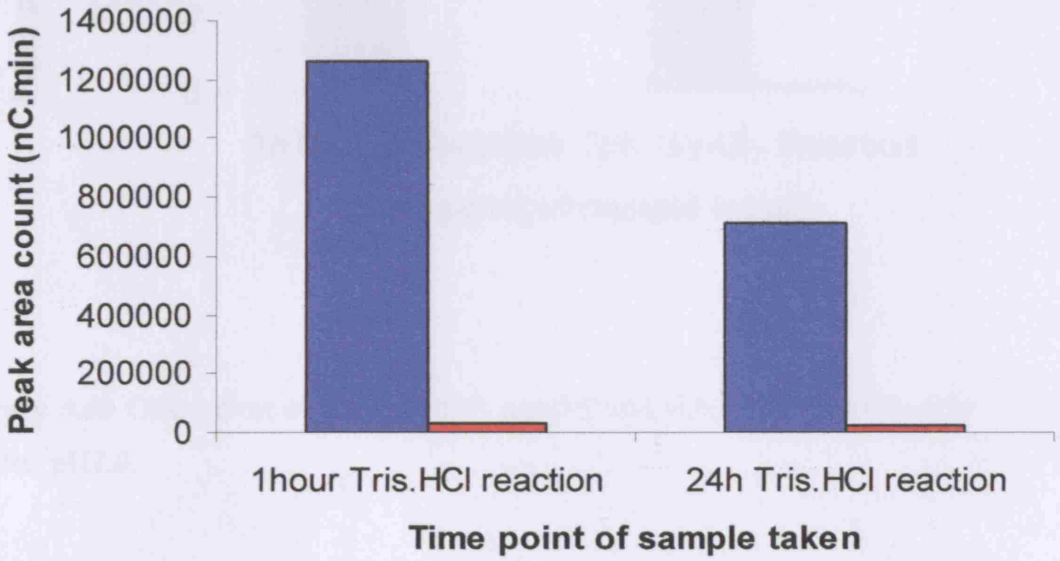


Figure 3.39 Conversion of 50 mM HPA and 50 mM GA in 50 mM Tris.HCl, pH7.0. Note that apparent loss of HPA in absence of erythrulose production is a likely result of poor loop-fill on injection of sample by HPLC.

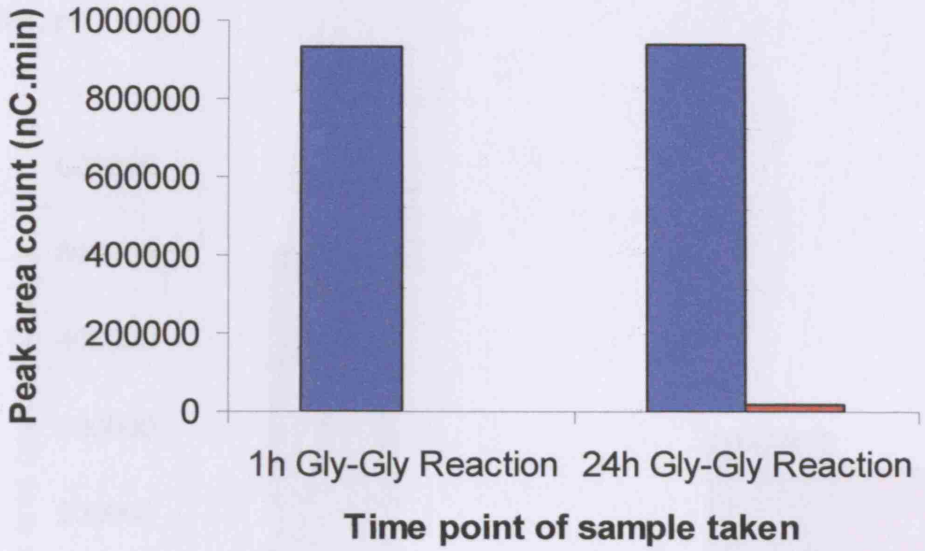


Figure 3.40 Conversion of 50 mM HPA and 50 mM GA in 50 mM Gly-Gly buffer, pH7.0.

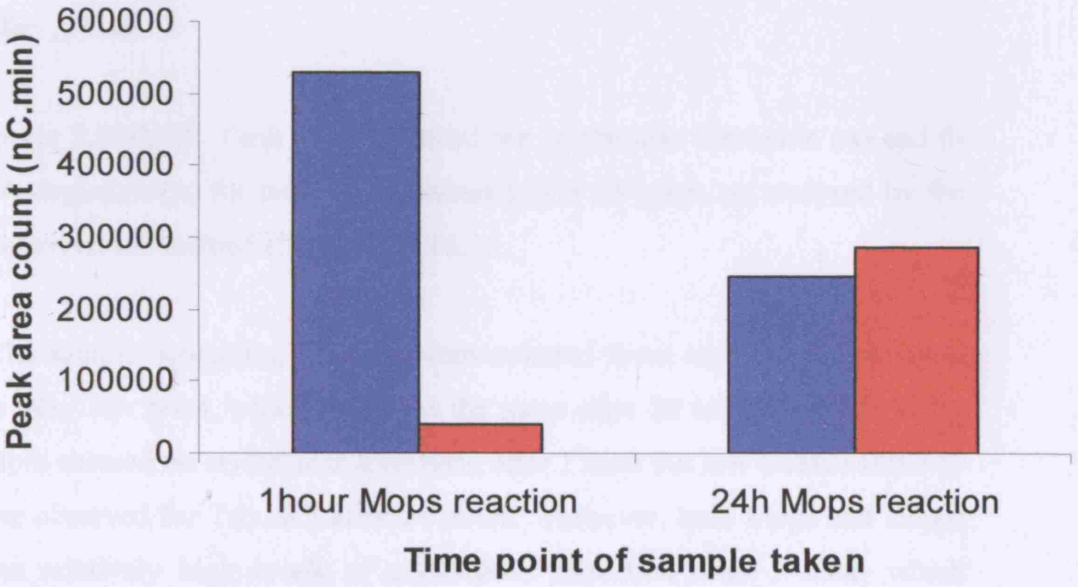


Figure 3.41 Conversion of 50 mM HPA and 50 mM GA in 50 mM Mops buffer, pH7.0.

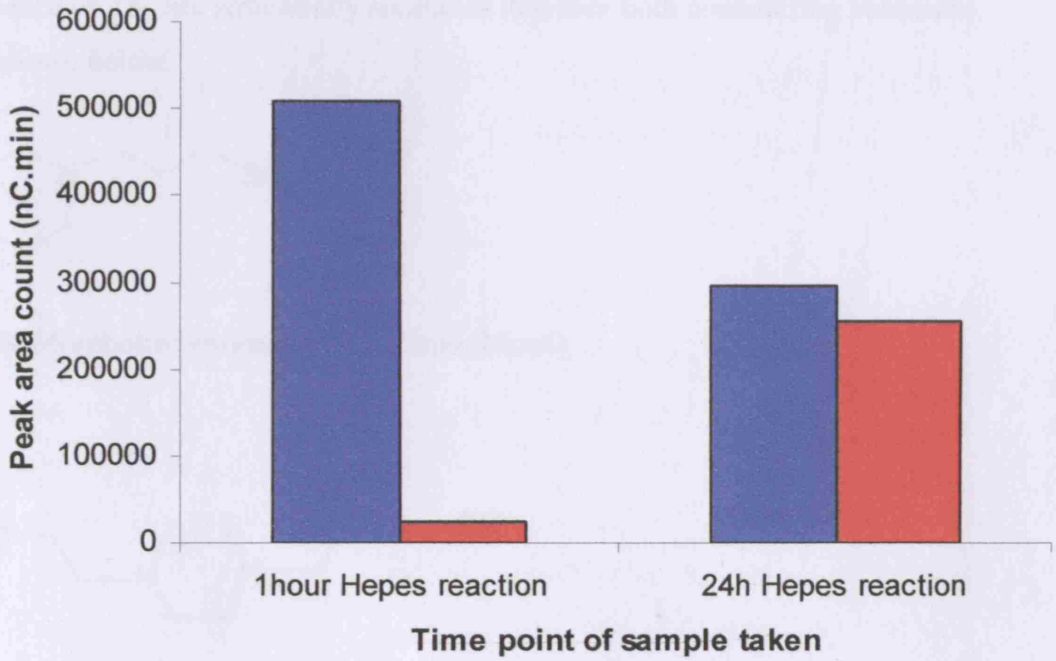


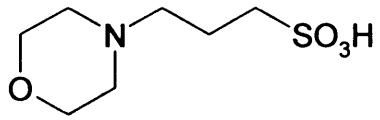
Figure 3.42 Conversion of 50 mM HPA and 50 mM GA in 50 mM Hepes buffer, pH7.0.

Figures 3.39-3.42. Peak areas obtained for erythrulose formation (■) and β-HPA depletion (■) for reactions monitored over 24 hours, as analysed by the standard HPLC method (Section 2.3.16.3).

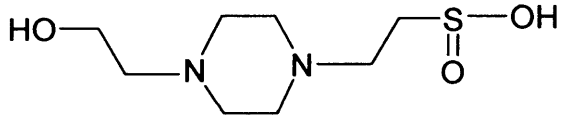
The sample containing Tris-HCl demonstrated some erythrulose formation at a very low level, which remained the same after 24 hours. The Gly-Gly sample showed no erythrulose formation after 1 hour but low levels similar to those observed for Tris.HCl after 24 hours. However, both Mops and Hepes show relatively high levels of erythrulose formation after 1 hour, which increases with time to 24 hours, with the Mops sample demonstrating the highest consumption of HPA and the highest level of erythrulose formation.

3.3.4.9 Structural and mechanistic analysis of buffers that act as a mimetic of TK

The buffers observed to catalyse the reaction between HPA and GA in the absence of TK, are structurally similar in that they both contain ring structures as shown below.



3-(N-Morpholino)propanesulfonic acid (Mops)



4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (Hepes)
(also N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid))

(Further analysis, not detailed here, was carried out by Dr Mark Smith in the Department of Chemistry at UCL.)

3.3.4.10 Activity of E.coli TK towards novel aldehyde acceptors

No initial activity towards the novel aldehyde acceptors was observed. Propionaldehyde exhibited activity with the wild type transketolase enzyme, but this had already been observed and documented in previous data (Tables 3.2). It is possible that some of the novel aldehydes may show activity when screened using the libraries constructed as described in Chapters 4 and 5

3.3.5 Inhibition of E.coli TK by fluoropyruvate

3.3.5.1 Fluoropyruvate and hydroxypyruvate inhibition experiments

Fluoropyruvate was initially thought to be one of the more promising candidates as a ketol donor potentially accepted by TK due to the similarities in size/charge between the R-groups of HPA and FlPA. However, initial experiments didn't show FlPA depletion or the formation of a product peak on HPLC analysis. As a results a series of experiments were performed to see if it might be binding but inhibiting, or remaining bound to the enzyme. These were performed as with the HPA ketol donor assays, except that FlPA was also added to final concentrations of 50, 25, 10, 5, 1, 0.1, 0.5 and 0.01 mM

In an initial protocol the ketol donor and inhibitor were added together with the glycoaldehyde acceptor to initiate the reaction after previous cofactor incubation of the enzyme. The ketol donor, inhibitor, and glycolaldehyde acceptor were each added to 50 mM. However, as the mode and kinetics of the inhibition were unknown, the initial conditions were modified to take this into account. It was possible that the mode of inhibition was irreversible or had slow kinetics. If this was the case, then by allowing incubation of FlPA prior to addition of HPA the observation of any existing inhibition was more likely. As a result the FlPA was allowed 20 minutes to incubate with the enzyme after the initial cofactor incubation. Also, as there were two ketols in this system the concentration of GA was increased to 100 mM, to ensure that the availability of aldehyde would not be a limiting factor. The addition of up to 100 mM GA is known to have no significant toxic or inhibitory effects. As in standard reactions, the cells were incubated for 20 minutes with co-factors, after sonication and prior to use. The FlPA was then added and left a further 20 minutes to allow FlPA access to enzyme prior to addition of HPA and then GA. The resulting effect on erythrulose formation is shown in Figure 3.43.

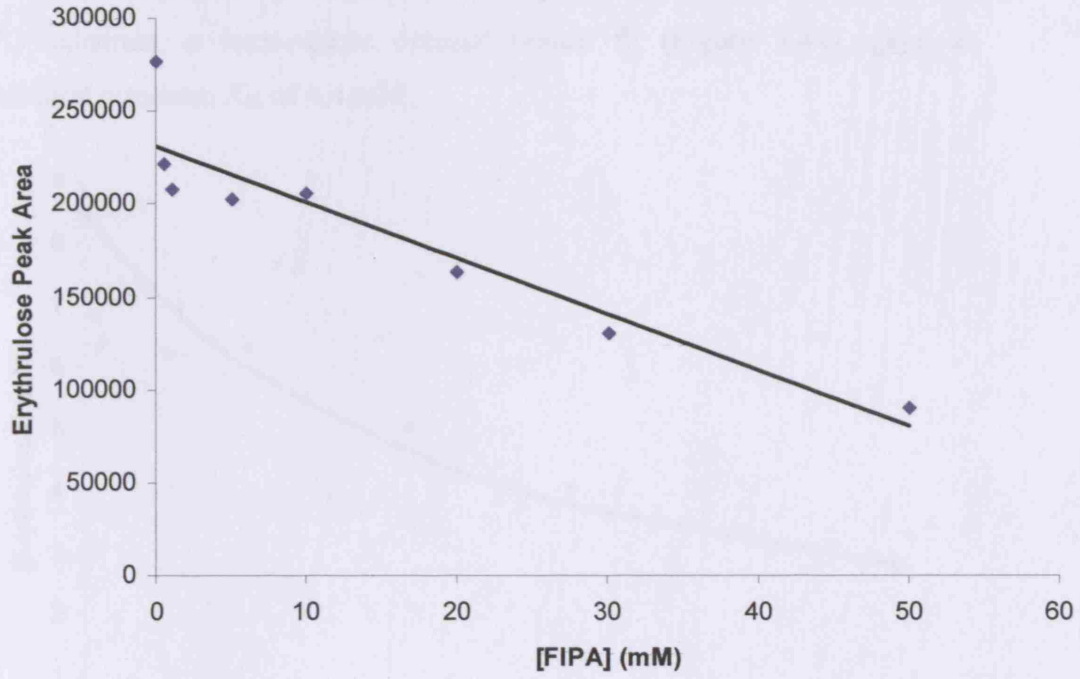


Figure 3.43 Erythrulose peak areas formed after 24 hours reaction, in the presence of increasing concentrations (0.5-50 mM) of FlPA, for the reaction of 50 mM HPA and 100 mM GA in 50 mM Tris buffer, pH7.0.

The addition of increasing concentrations of FlPA results in the decreasing formation of erythrulose, indicating inhibition. However, complete inhibition is not reached at any concentration of FlPA studied.

3.3.5.2 Modelling of E.coli TK inhibition by fluoropyruvate

While the experiments carried out to date do not enable the determination of the type of inhibition (e.g. competitive, non-competitive, and un-competitive) that is caused by fluoropyruvate, it was thought useful to determine the approximate inhibition constant where competitive inhibition is assumed. The modelling of the data to obtain the possible inhibition constants was carried out by Dr Bing Chen.

When assuming competitive inhibition by FlPA, with only the HPA (not GA) substrate, a least-square optimal search fit (Figure 3.44), gave an inhibition constant, K_{if}, of 4.4 mM.

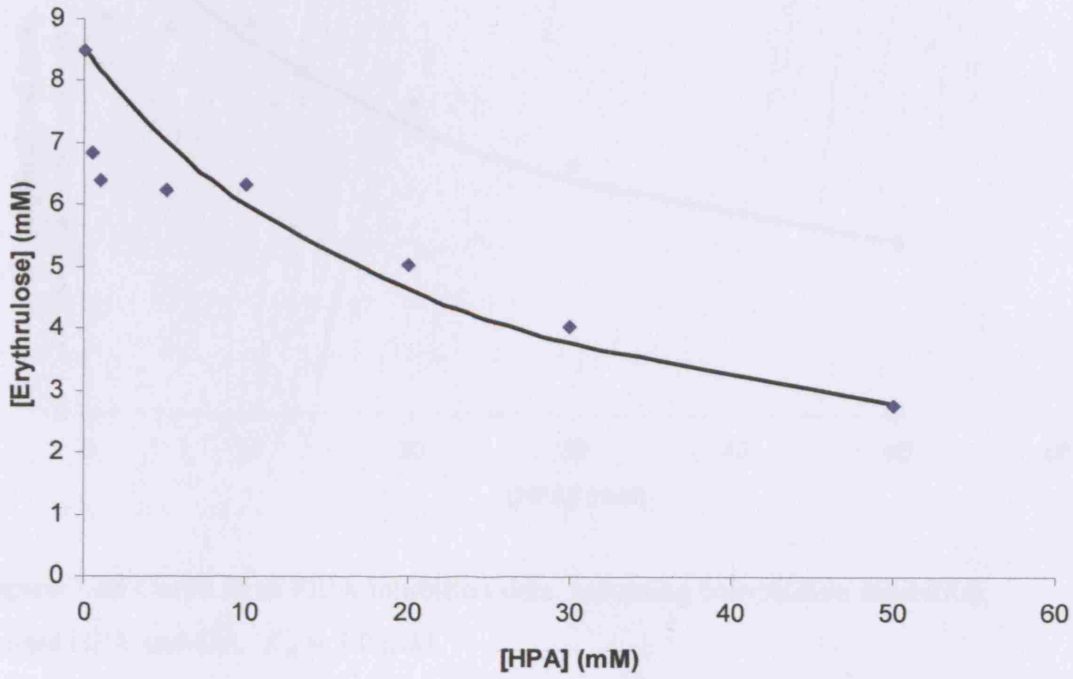


Figure 3.44 Curve fit to FlPA inhibition data, assuming competitive inhibition against HPA only. $K_{if} = 4.4$ mM

When assuming competitive inhibition by FlPA, with both the HPA and GA substrates, a least-square optimal search fit (Figure 3.43), gave an inhibition constant, K_{if}, of 7.0 mM.

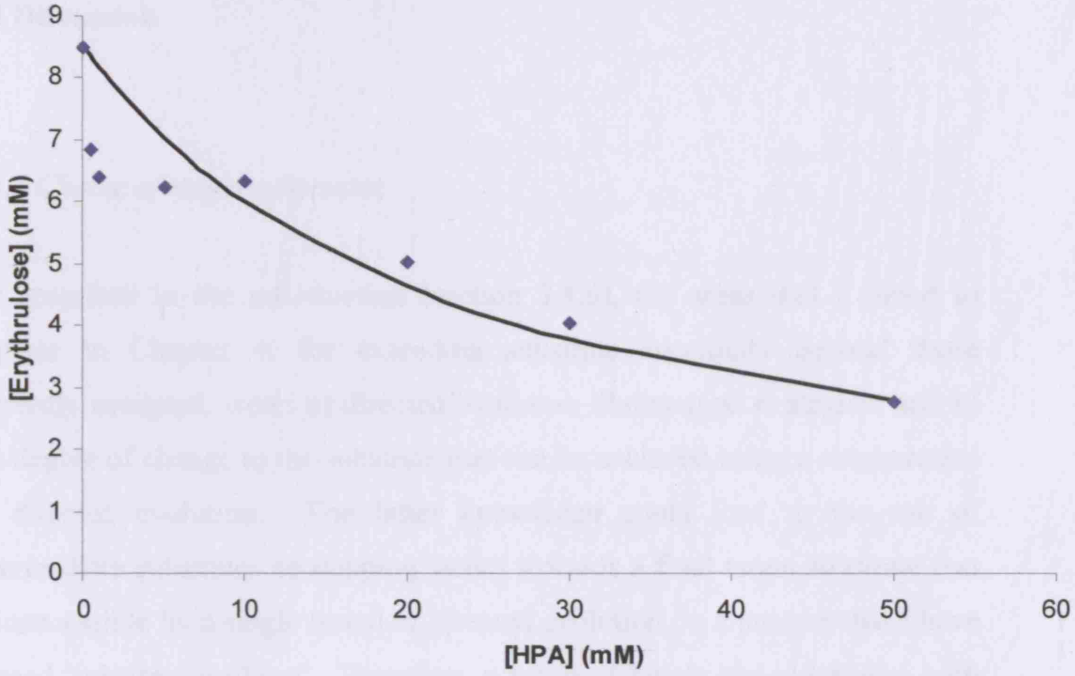


Figure 3.45 Curve fit to FlPA inhibition data, assuming competitive inhibition against HPA and GA. K_{if} = 7.0 mM.

3.4 Discussion

3.4.1 Choice of target substrates

As described in the introduction (section 3.1.6), the areas that I aimed to explore in Chapter 4, for extending substrate specificity beyond those currently accepted, were: a) directed evolution library-type strategies; and b) the degree of change to the substrate that can be achieved using a single round of directed evolution. The latter knowledge could lead to the use of intermediate substrates as stepping stones towards a final target substrate that is inaccessible by a single round of directed evolution, in a process that I have termed 'substrate walking'. Therefore, a range of ketols and aldehydes, with potential synthetic use, were chosen such that they formed a sequence of iterative chemical changes. To define these iterative changes, the substrates were chosen such that they formed a so-called Hammett series, in that they could potentially be plotted in a linear free-energy relationship. The series are shown in Figures 3.1, 3.15, and 3.16. Hammett-plots demonstrate a linear relationship between the logarithm of rate (or equilibrium) constants and σ-values which represent a series of substituent changes to one compound in a chemical reaction. Initial σ-values are obtained from the ratio of equilibrium constants for the substituted compound against a reference compound, in a representative reaction. σ-values for the ketol donors in Figure 3.1 were taken for example, by analogy to acetate ester hydrolysis data as shown in Table 3.3 (section 3.3.1). Data for the aldehyde acceptors shown in Figure 3.16 were not available but the substrates were expected to potentially form a similar series. As many of the substrates had not previously been assessed with TK from *E.coli*, their activity with wild type had to be assessed prior to screening against mutant TK libraries. To achieve this, assays needed to be developed for each substrate.

3.4.2 Assay development

Screening methods and assays were developed for most of the target novel ketols (Figure 3.1) and aldehydes (Figures 3.15 & 3.16), with HPA and glycolaldehyde being used as the standard ketol or aldehyde partner substrate in each case. HPA and glycolaldehyde have been widely studied previously.

3.4.2.1 Ketol donor assays

Assay development for the novel ketols was relatively simple in that they all had physicochemical properties broadly similar to HPA and as a result could be assayed using the same Aminex HPLC column (Biorad), often with only an extended run time being necessary due to increased retention times. Care had to be taken with the thiol containing mercaptopyruvic acid, in that DTT was required as a reducing agent to prevent disulphide bond formation. Care also had to be taken with bromopyruvate as it is readily hydrolysed to form HPA at pH greater than 8.5, and so pH adjustment prior to use had to be carefully controlled.

Each ketol donor was resolved by HPLC to identify their retention times, as shown in Figures 3.2-3.7 (section 3.3.3). Single peaks were resolved for all ketol donors except bromopyruvic acid (BrPA) and mercaptopyruvic acid (HSPA), and were clearly distinguishable from the peak of HPA. Bromopyruvic acid, however, showed a more complex HPLC trace with a peak at 5.6 minutes for BrPA, a small peak at 8 minutes, and an unknown peak at 9 minutes. As BrPA is known to hydrolyse slowly to form HPA, a method in fact used to synthesis HPA (section 2.2.5), the peak at 8 minutes was readily explained. The peak at 9 minutes correlated with pyruvic acid, and it is possible that this product is formed by elimination of bromide in some way, or is a byproduct of the synthesis of BrPA. Most problematic was mercaptopyruvic acid for which no clear peak was resolved either in the absence or presence of DTT.

3.4.2.2 Aldehyde acceptor assays

As many of the aldehyde acceptors were expected to strongly absorb light at UV-wavelengths, due to their conjugated double bond or aromatic natures (Figures 3.15 & 3.16), initial assay development was attempted using absorbance spectra. Such an assay would be rapid, quantitative and amenable to high-throughput screening if successful. Analysis of their spectra was also deemed useful for optimising the detection of such compounds using a UV absorbance detector for HPLC. Most of the aldehyde substrates had relatively low solubilities of less than 10mM, or occasionally less than 3 mM. In particular, para-nitrobenzaldehyde had a solubility in buffer that was too low for detection by UV-absorbance. Heptadienal and hexadienal were also very low in solubility and often formed emulsions rather than solutions at high concentrations, and also had a frequent tendency to degrade certain types of 'soft' plastics, including eppendorfs and microtitre plates (benzaldehyde also degrades some plastics). As a result these substrates were used in glassware rather than plastic whenever possible. The majority of aldehyde substrates were readily detectable by UV-absorbance as shown in Figures 3.18 to 3.23 (excluding 4-nitrobenzaldehye). However, the spectra of samples taken during the reaction of substrates with HPA and TK from *E.coli*, showed no novel absorbance peaks which might have corresponded to a product, or measurable changes to the initial spectra, even after 24 hours of incubation. Attempts were also made during assay development to exploit potential absorbance at weak symmetry-forbidden bands that aldehydes and ketones can show in the 275-295nm range as well as potential solvent effects when considering different phases to 100% TFA (Pages 17-20, Dudley. H. Williams & Ian Fleming), but these were similarly unsuccessful.

High accuracy HPLC methods were developed for each of the aldehydes after the unsuccessful attempts to develop a rapid UV-absorbance screen. HPLC assay development for aldehyde substrates required the use of either UV detection (AD20) or standard electrochemical triple-wave detection (ED40), and a C18 column, to resolve many of the aromatic compounds, as

well as the propionaldehyde keto-diol product, as these were found to be impossible to resolve using an ion-exchange column.

As can be seen in Figures 3.24-3.27 (section 3.3.4.2.1) peaks were readily resolved using a C-18 column and UV-absorbance detection for 4-hydroxy benzaldehyde, 4-methoxy benzaldehyde, and propionaldehyde. No single peaks were resolved for heptadienal or hexadienal and their HPLC traces (data not shown), indicated that they had begun to degrade the HPLC column matrix or interact with other parts of the HPLC system, as a broad range of unresolved peaks were observed. The multiple peaks observed were possibly due to compound interaction with any number of surfaces or plastics within the HPLC column or sample vials, although glass was used whenever possible with these compounds. As a result of the difficulties with these compounds and potential damage they could cause to the HPLC system or column matrix, further assay development was not carried out for these two compounds.

The observed peak intensity for propionaldehyde using the UV-absorbance detector was low at both 210 nm and 270 nm. Consequently, derivitisation was attempted as a means to allow UV-absorbance detection of both the propionaldehyde substrate and the expected keto-diol product. Individual components of the TK catalysed reaction gave clearly resolved peaks after derivatisation with 2,4-DNP (Figures 3.29-3.32, section 3.3.4.2.2). However, on mixing the three solutions containing the expected keto-diol product (PKD), and the substrates HPA and PA, with 2,4-DNP, there was an apparent signal loss for the keto-diol and an increase in signal for the aldehyde (Figure 3.33), possibly as a result of the interaction of these derivatized compounds. Despite attempts to resolve this issue it was not possible to detect the keto-diol in a mixture or a biotransformation reaction, and as a result this method was not feasible to carry forward. An alternative and successful solution to the low UV-absorbance level of propionaldehyde and the expected keto-diol product was the use of an electrochemical detection method which gave clearly visible and well resolved peaks.

Benzaldehyde assay development had been a bit more problematic than the other aromatic substrates in that it did not have the same solubility or

similar UV absorbance. As a result of this many wavelengths were tested until the ACE 225nm C-18 screen was finally developed (Figure 3.36). At best this assay is 16 minutes to also observe the substrate peak at 15.20 minutes the benzaldehyde ketol-diol is observed at 7.20 minutes.

As the HPLC method was still relatively slow for high-throughput screening work, with both the keto-diol product PKD resolved after 6 minutes (Table 3.4) and BKD resolved after 7 minutes (Table 3.6), thin-layer chromatography (TLC) was explored as a potential rapid method for pre-screening enzyme libraries before quantification using HPLC. It was investigated as a possible high throughput screen method for several novel substrates as the screening time for potential libraries with many of their HPLC developed assays would be impractical. The results for both benzaldehyde and propionaldehyde (Figure 3.37) were promising when comparing chemically synthesised products and biotransformation reactions with propionaldehyde, for which there exists a low level of activity with *E.coli* TK. Because of the simplicity and rapidity of TLC, it is often used to monitor the progress of organic reactions and to check the purity of products. TLC could potentially save a lot of screening time as the method is faster than the aldehyde HPLC methods developed. Although it is not as quantitative as HPLC analysis, it is capable of determining a 'yes or no' answer in terms of 'hit or miss' from a library screen, and could be used to greatly reduce the number of mutants selected for further analysis by HPLC.

As a result of the successful high-throughput assay development using TLC as a pre-screen and HPLC for subsequent quantitation, both benzaldehyde and propionaldehyde were chosen for further development of reaction conditions and enzyme evolution, alongside the better characterised glycolaldehyde. It should be noted that while the colorimetric assay was being developed, I had already begun to use the TLC method in the interests of time and progress with the library screening.

3.4.3 Effect of buffers and observation of the biomimetic reaction

During the development process for colorimetric assays, the effects of various buffer systems were investigated. It was noticed some control wells, containing either no cells or TK enzyme, were showing either the erythrulose or the propionaldehyde keto-diol (PKD) product formation by HPLC, in the presence of certain buffer compounds. These all have buffering optima (pKa) around pH 7-8, but their effects on novel substrates, reactions and products needed to be determined in more detail. As a result further analysis was made to ascertain whether this had arisen due to novel chemical process or contamination of controls due to, for example, poor sterile technique.

Upon further investigation it was found that two of the buffers selected, Mops and Hepes, appeared to have 'catalytic' potential (Figures 3.39-3.42, section 3.3.4.8). As a result these two buffers were not suitable for use in novel aldehyde screens and further investigation of the apparent TK biomimetic phenomenon was taken up by Mark Smith in the Department of Chemistry at UCL. Of the two remaining buffers, Tris.HCl and Gly-Gly, the latter was chosen due to fears of possible interactions of propionaldehyde with the amine of Tris.HCl to form a Schiff's base. The second deciding factor in buffer choice was the possible need for mass spectrometry to confirm the products formed, and a simple amino acid buffer like Glycine-Glycine would readily allow this, whereas the amine containing Tris.HCl would likely interfere with signal analysis.

Further study is ongoing (Dept Chemistry) into the ability of certain types of buffers to apparently mimic the activity of TK and facilitate chemical reaction in the absence of any enzyme. The buffers which have been found to be capable of these reactions both contained hetero-atomic ring structures which are likely to be involved in the conversions observed. Furthermore, it has been shown that TPP and Mg^{2+} are not required for this biomimetic reaction to occur, although the reaction will of course be slower than the enzymatic route, and will also yield a racemic mixture. In summary, a one-pot tertiary-amine-mediated carbon–carbon bond-forming reaction in water has been observed that is chemo-selective for hydroxypyruvate as a donor group. Interestingly, the observed donor substrate tolerance to date is analogous to that observed with transketolase (Smith *et al.*, 2006). However, the products

are a racemic mixture and so enzymatic conversion still has advantages if optically pure products are required. The biomimetic reaction did, however, provide a very convenient route for the synthesis of the keto-diol product standards that were required.

3.4.4 Initial activity screens

The ketol donors, other than HPA, all failed to show activity in initial screens with *E.coli* cell lysates expressing transketolase (TK), using the assays developed above (Figures 3.8-3.14, section 3.3.3.1). Observations of small quantities of erythrulose-like product in the BrPA-GA reaction with TK led to the initial possibility that BrPA was reacting. However, this was thought more likely be the result of GA reacting with the small quantities of HPA that are formed from BrPA in water, especially at greater than pH 8.5 which may transiently occur when adjusting the pH of the substrate solution to 7.0 (see HPA synthesis section 2.2.5). The lack of previously published data on non-natural ketol donors, except for HPA or sugars that are similar to the naturally accepted ones, suggests that the hydroxyl binding pocket in TK is unable to readily accept novel R-groups in the C-1 position of the ketol donor. Differences in size, charge and hydrogen bonding to active-site residues are most likely preventing the substrates tested here from binding to the TK active-site in a productive conformation, if at all. However, it was expected that fluoropyruvate would be able to bind as the fluorine atom is capable of hydrogen bonding and is small enough to fit into the cavity usually occupied by the hydroxyl group of HPA. For this reason, the possibility that FlPA binds but causes inhibition was explored. As indicated by the Hammett series data in Table 3.3, the high electronegativity of the fluorine atom may deactivate the ketone carbonyl group of the FlPA substrate such that it is no longer chemically activated by TK. Indeed, as discussed in Section 3.4.5, fluoropyruvate was found to have some inhibitory properties when incubated with TK prior to reactions with HPA. However, this needs further work to determine an accurate *Ki* and also the mode of inhibition, which was

considered beyond the scope of this thesis. The other ketol substrates tested did not appear to react with TK at all. This may be due to the greater sizes of their R-groups when compared to the hydroxyl of HPA, such as -SH, -Br, or -CH_3, that prevent entry into, interaction with or correct orientation by the hydroxyl binding pocket, or as the result of the loss of hydrogen bonding as is expected with pyruvate.

No initial activity towards any of the novel aldehyde acceptors was observed, except minor activity of propionaldehyde with the wild-type transketolase enzyme, which had already been observed and documented in previously (Table 3.2). Some of the aldehydes were not able to be tested, however. For example, para-nitrobenzaldehyde was too insoluble for detection, even in the presence of up to 2% DMSO (in an effort to increase solubility). As mentioned previously, heptadienal and hexadienal were also both problematic in assay development as their solubility in buffer was very low. Furthermore, these two compounds were observed to degrade certain types of plastics (soft pipette tips, some microtitre plates and Eppendorf tubes), and their HPLC traces seemed to indicate they had begun to degrade the HPLC column matrix or interacted with other system parts. The multiple signal spiking observed was possibly due to compound interaction with any number of surfaces or plastics within HPLC column or sample vials, although glass was used whenever possible with these compounds. As a result of the difficulties with these compounds and potential damage they could cause to HPLC system or column matrix further assay development was not carried out for these two compounds.

It was hoped that one or more of the initial substrates from both the ketol and aldehyde side of reaction would show some initial activity for exploration of our Hammet plot series, and also to test the potential for substrate walking as a means to expand the TK enzyme substrate specificity with the libraries produced. Unfortunately too many other factors came into play as described. The substrates ultimately chosen for screening of libraries represent two extremes of a series for the aldehyde acceptor side of the reaction. Whereas the aldehydes had displayed some initial activity, this was not the case for ketol donors and so the latter were not explored further. As

mentioned propionaldehyde has been shown to have an activity 5% relative to that of glycolaldehyde, using wild-type TK from *E.coli*. It was possible, however, that some of the novel aldehydes may show some activity when screened using the libraries constructed, though it was considered much less likely that the ketol substrates would as there was no evidence of initial activity detected on any other substrate than HPA, to serve as a starting point for improvements by directed evolution.

3.4.5 Inhibition by FlPA

While the experiments carried out to date do not enable the determination of the type of inhibition (e.g. competitive, non-competitive, and un-competitive) that is caused by fluoropyruvate, it was thought useful to determine the approximate inhibition constant when competitive inhibition is assumed. This information would provide valuable data for crystallisation studies to determine the structure of TK complexed to FlPA, being carried out in collaboration with Prof. Littlechild at Exeter University.

Modelling of the data to obtain the possible inhibition constants was carried out by Dr Bing Chen in the Department of Biochemical Engineering at UCL. Competition between FlPA and only HPA, and also between FlPA and both HPA and GA were considered. Inhibition constants (K_{if}), were predicted to be 4.4 mM for competition with HPA only (Figure 4.42), and 7.0 mM for competition with both GA and HPA (Figure 4.43).

3.5 Conclusions

The lack of previously published data on non-natural ketol donors, except for HPA or sugars similar to the naturally accepted ones, suggests that the hydroxyl binding pocket in TK is unable to readily accept novel R-groups in the C-1 position of the ketol donor. Differences in the size, charge and hydrogen bonding to active-site residues are most likely preventing the ketol

donor substrates tested here from reacting, except for fluoropyruvate which does at least bind. The fluorine group in FlPA is very electronegative and it may be that the resulting displacement depolarisation of the charge across the carbonyl bond of the ketone group, prevents susceptibility to nucleophilic attack by the carbanion of the activated TPP cofactor in TK, after binding to the TK active-site. The other substrates tested did not appear to react with TK at all. This may be due to their greater changes in size of the R-group compared to –OH, such as –SH, -Br, or $-CH_3$ that prevent entry into, interaction with or correct orientation by the hydroxyl binding pocket, or as the result of weak or loss of hydrogen bonding as expected with pyruvate. The concentration range at which FlPA inhibits at also needs to be more closely examined >50mM with pre-incubation prior to the addition of not only GA, but also HPA. The presence of HPA would be problematic as not only will it be present at similar concentrations, but also we know that dihydroxyethylthiamine pyrophosphate (DHETTP) intermediates can form, and without GA available the HPA would remain bound such that FlPA would not have access to the active site of these enzymes to facilitate inhibition (Meshalkina *et al.*, 1995)

Assay development has shown that time needs to be taken to tailor/produce suitable assays for the library to be developed. As can be seen in many cases it is not enough to just determine the optimum absorbance for the substrates and products as their chemical properties often differ widely making resolution under some systems impossible. Often the new products are retained for such longer periods that the substrates or products are not visible under the same conditions requiring a change of detection systems. Interactions can also be observed so despite visualising each component separately mixtures and biotransformations do not always appear the same as single traces (derivitisation figures 3.30-3.33). This can either render the assay useless due to inability to detect fractions or as a result of the extended retention time of one of the fraction making the method time much longer and as a result unrealistic for use in library screening. As a result library size and screening methods must be closely linked as a huge library of mutants is useless if it is not able to be screened in a realistic time.

As shown by a number of groups previously, the range of aldehyde donors accepted is fairly broad even if many have low activity. It may be that with rounds of mutation and directed evolution, greater activity with more non-natural ketol donors can be found. However, it appears more likely that the spectrum of aldehydes accepted can be broadened more readily and probably to greater effect. Also the initial results observed for propionaldehyde and previous mutations on TK towards aromatic substrate acceptance (Dr. J.Ward private communication) suggest an aldehyde substrate series for substrate walking, and as determined by a Hammet series could potentially yield interesting mutants with novel activities. The overall results gained from the assay development and substrates testing enabled certain substrates (propionaldehyde and benzaldehyde) to be carried forward for further investigation with the libraries produced in Chapters 4 and 5.

3.6 Future and Ongoing Work

Future work that is beyond the scope of this project, but nevertheless of interest would be determining the mode of inhibition of FlPA as well as a more accurate measurement of its *K*i. Although this would by no means ever be considered a therapeutic inhibitor due to potential inhibition of other enzymes, the close links between TK and cell survival, and in more recent studies as a target for cancerous growth control, studies using FlPA have the potential to elucidate our understanding of the ketol substrate binding. Previous crystal structures have been produced with HPA bound to TK in its intermediate state (DHETPP). However, with the structural similarity between HPA and FlPA and the non reactivity of FlPA, it is possible that they bind in a similar manner, but with FlPA not reaching the same intermediate state. If this was shown to be the case, the residues and method of binding/docking by the ketol substrate would be significantly better understood and determined. As a result cell paste was submitted to J. Littlechild (Exeter University) where possible crystallography of TK-FlPA may be performed.

Further study is ongoing in the Department of Chemistry, into the ability of certain types of buffers to apparently mimic the activity of TK and facilitate chemical reaction in the absence of any enzyme. Furthermore, the TPP and Mg^{2+} cofactors are not required for this to occur, although the reaction will of course be slower than enzymatic formation and will yield a racemic mixture. In summary, a one-pot tertiary-amine-mediated carbon–carbon bond-forming reaction in water has been observed that is chemo-selective for hydroxypyruvate as a donor group. Interestingly, the observed donor substrate tolerance to date is analogous to that observed with transketolase.

Chapter 4 – Library construction and tolerance to mutation

4.1 Introduction

Enzymes are desirable for use in chemical synthesis due to the numerous potential advantages they offer, such as rate acceleration and high regio- and stereo-specificity which conventional chemical catalysts are unable to provide. However, they are not always suited to the harsh conditions or substrates that chemists wish to operate under. This has led to the use of protein engineering techniques such as directed evolution, to broaden the specificity, enhance the stability, or even alter the function of enzymes in some cases, and overcome some of the process related problems while maintaining the advantageous aspects of enzyme function. Directed evolution thus has great potential to improve existing chemical reactions in the ways described and also potentially reduce the presence of harmful by-products often associated with traditional chemical synthesis.

The areas that I aimed to explore (in Chapter 4) for extending substrate specificity beyond those currently accepted are: a) directed evolution library-type strategies; and b) the degree of change to the substrate that can be achieved using a single round of directed evolution. The latter knowledge could lead to the use of intermediate substrates as stepping stones towards a final target substrate in a process that is termed from this point onwards as 'substrate walking'. Therefore, this Chapter explores the potential library production techniques which can be employed to produce a mutant library of TK for screening with the assays developed in Chapter 3 and any potential changes or improvements in the model reaction in this first round of directed evolution. The quality of the libraries in terms of expression levels and DNA sequence diversity will also be investigated before subjecting the libraries produced to novel substrate screens in Chapter 5.

4.2 Materials & Methods

4.2.1 Library construction

4.2.1.1 Mutator strains

XL1-Red Mutator strain (Stratagene, Netherlands) of *E.coli* was transformed using PQR791 plasmid using standard heat shock (section 2.3.13) transformation and cultured for 8 hour cycles (5-6 generations) in LB medium with ampicillin, and in some cases enriched with 1mM $MgCl_2$ and glucose. This was repeated for 60-120 generations in order to introduce mutation(s). Plasmid DNA was extracted using standard miniprep protocols (section 2.3.6) where possible.

4.2.1.2 Error-prone PCR (EpPCR)

Error-prone PCR reactions (Roche Products Limited, Welwyn Garden City, UK)
for the TK gene fragment were set up as follows:

10 µl 10x Taq buffer
2 µl plasmid template DNA (at 25ng/µl)
2 µl primer 'Tk-15mut F'
2 µl primer 'TKC-term'
8 µl 10mM dNTP mix for 0.2mM final concentration of each nucleotide
4.0-4.8 µl $MgCl_2$ for final concentrations between 1.0 and 1.2mM
10 µl $MnCl_2$ for final concentration of 0.1 mM
2 µl Taq polymerase
60-59.2 µl ddH_2O for final volume of 100 µl

Primer 'TK-15mut F' at the N-terminal end of the TK gene:
5'-CCCTTCATCATCAGATCTGGAGTCA-3'

Primer 'TKC-term' at the C-terminal end of the TK gene:

5'-TACCGCCGGACGGTACCTCTAGACGCCG-3'
Primers ordered from Stratagene, Netherlands.

PCR was performed with the following thermocycling protocol:

94 °C 5minutes (1 cycle) Initial activation step
94 °C 30seconds (30 cycles) Denaturation
55 °C 60 seconds (30 cycles) Annealing
72 °C 60 seconds (30 cycles) Extension
72 °C for 5 minutes Final extension

4.2.1.3 Ligation of gene products into Capture Vectors

4.2.1.3.1 TOPO plasmid capture vector

Ligations were performed with a ration of 100 fmol of EpPCR fragment insert to 1 fmol TOPO vector, and then left for two hours at room temperature according to the manufacturer's instructions (Invitrogen, Paisley, UK). After ligation was complete 2 µl of each was added to a cold 0.1 cm^3 electroporation cuvette with 100 µl of electro-competent TOP10 cells, and these were then pulsed with the EC1 setting (section 2.3.14). 500 µl of SOC media (section 2.2.3) was added and cells were then grown for 1 hour at 37 °C and 10 µl plated out on 300 µg/µl Ampicillin with IPTG/XGAL plates.

4.2.1.3.2 PCR Script SK+ plasmid capture vector (Invitrogen, Paisley, UK)

Two approaches were used for 4 hours, heat inactivated at 65 °C and resolved on a 0.6% low melting point agarose gel. The band corresponding to the vector (~3kb) was then gel extracted (section 2.3.8) using a UV plate and a scalpel and used in the ligations with double digested EpPCR products. In the second method, the original commercially available capture vector was used for ligations with the EpPCR products. For this vector it is necessary to polish

the EpPCR products using a cloned Pfu fragment, to improve the ligations. For the polishing reaction, 10 µl of the purified PCR product, 1 µl of 10 mM dNTP mix (2.5 mM each), 1.3 µl of 10x polishing buffer and 1 µl of cloned *Pfu* DNA polymerase (0.5 U) were combined and left for 30 minutes at 72 °C. This polished product was then added directly to the ligation reaction, 400 ng of polished PCR product are required per 10 ng insert (40:1 ratio for a 2.5kb insert) the reaction is then left for one hour at room temperature in the presence of 10 mM rATP, *Srf* I restriction enzyme (5 U/µl) and T4 DNA ligase (4 U/µl), before transformation.

4.2.1.4 Saturation site-directed mutagenesis

Saturation site-directed mutagenesis was performed using the Quikchange kit (Stratagene, Netherlands). PCR reactions for each library site were set up in parallel with three different concentrations of DNA template for each site (5, 15 and 25ng DNA) as follows:

10 µl of 10X *Pfu* buffer
1 µl of 50mM forward and reverse NNS primer
1 µl of 5, 15 and 25ng/µl PQR791 stock solution
2 µl of dNTP mixture (4mM)
1 µl DMSO
82.5 µl ddH_2O to a final volume of 98.5 µl
1.5 µl of *Pfu Turbo* added after the hot-start protocol

The thermocycle protocol consisted of:

95 °C 5minutes (1 cycle) Activation
95 °C 30seconds (16 cycles) Denaturation
55 °C 60 seconds (16 cycles) Annealing
58 °C 5minutes (1min/kb) (16 cycles) Extension
72 °C for 5 minutes. Final Extension

After thermocycling, 1 µl of Dpn1 (Stratagene) restriction enzyme was added to each PCR reaction, and incubated at 60 °C in a water bath for one hour to remove any excess parental DNA template from the reactions. The DpnI is specific for methylated DNA obtained from *E.coli* and does not digest the DNA synthesised by PCR. The quality of the products were then analysed using electrophoresis on a 0.6% agarose gel.

4.2.1.5 Definition of library target sites

The cycle-time for measurement of samples by HPLC set a practical limit of twenty target sites for saturation mutagenesis. Statistically it can be shown that a minimum of 61 colonies would need to be picked to obtain a 95% probability of including all twenty different residues in each library. Accounting for potential bias it was decided to aim for screening of 90 colonies from each saturation mutagenesis site which would require 20x 96-well plates in total.

The overall aim was to define amino-acid sites that were most likely to elicit the desired changes in substrate specificity. For the twenty sites, ten were determined by structural analysis, and ten were determined by phylogenetic considerations.

4.2.1.5.1 Structurally determined primary-shell library

The ten structurally defined sites were chosen by Ed Hibbert (UCL) as being within 4 Å of the erythrose-4-phosphate substrate bound to the active site in the crystal structure for yeast TK (Nilsson *et al.*, 1997. 1NGS.PDB). The ten sites selected were H26, H100, I189, H261, G262, R358, S385, H461, D469 and R520. The primers used for saturation mutagenesis are shown below, and were obtained from Stratagene (Netherlands).

NNN 26 5'-AAAGCCAAATCCGGTNNNCCGGGGGCCCCTATGG-3',
NNN 100 5'-CACTCTAAAACTCCGGGTNNNCCGGAAGTGGGTTACAC-3',
NNN189
5'-ACGATGACAACGGTATTTCTNNNGATGGTCACGTTGAAGGCTGG-3',

NNN 261 5'-TACCCACGACTCCNNNGGTGCGCCGCTGGG-3',
NNN 262 5'-CACGACTCCCACNNNGCGCCGCTGGGCG-3',
NNN 358 5'-CGGCGAAAATCGCCAGCNNNAAAGCGTCTCAGAATGCT-3',
NNN 385 5'-GCTGACCTGGCGCCGNNNAACCTGACCCTGTGGT-3',
NNN 461 5'-GGTGATGGTTTACACCNNNGACTCCATCGGTCTGGG-3',
NNN 469 5'-TCGGTCTGGGCGAANNNGGGCCGACTCACCAG-3',
NNN 520 5'-GCACTGATCCTCTCCNNNCAGAACCTGGCGCAGC-3',

After centrifugation, all primers were resuspended in ddH_2O to a final concentration of 100 mM, based on yields stated in the DNA product analysis form. Each solution was diluted 40,000-fold in MilliQ water to give 25 ng/µl, prior to use in PCR.

4.2.1.5.2 Bioinformatically determined second-shell library

The ten phylogenetically defined sites were chosen by Sean Costelloe (UCL), from those sites within 10 Å of the TPP cofactor within the TK active-site that were also shown by multiple sequence alignments and phylogenetic reconstruction to vary most during the evolution of TK (SC personal communication). In the phylogenetic analysis 52 TK sequences representing bacteria, fungi and a small number of plants, as well as one trypanosome were chosen from sequences identified in using a BLAST search with *E.coli* TKTA as the input sequence. Sequences designated as 'putative' were eliminated as they are not proven to be transketolases. A phylogenetic tree was constructed by Sean Costelloe (unpublished) using the Phylip and PAML sequence analysis tools. The full sequences at the nodes of the tree for the Sce and Eco lineages were determined back to their common ancestor. The 52 residues comprising the active-site of TK were identified as those within 10 Å of the TPP cofactor. Of these 52 residues, only ten were found to vary throughout the evolutionary history of TK. These 10 residues formed the basis for SSDM in the Bioinformatically determined, (so-called phylogenetic) library. The ten phylogenetic sites therefore identified as having greatest likelihood to elicit substrate specificity changes were K23, A29, N64, M159, S188, D259, A383,

P384, V409 and L466. The primers used for saturation mutagenesis are shown below, and were obtained from Stratagene (Netherlands).

NNS 23 5'-CGCAGTACAGAAAGCCNNSTCCGGTCACCCGGG-3',
NNS 29 5'-GGTCACCCGGGGNNSCCTATGGGTATGGC-3',
NNS 64 5'-TCGTGCTGTCCNNSGGCCACGGCTC-3',
NNS 159 5'-GACGGCTGCATGNNSGAAGGCATCTCC-3',
NNS 188 5'-CTACGATGACAACGGTATTNNSATCGATGGTCACGTTGAAG-3',
NNS 259 5'-CGGTACCCACNNSTCCCACGGTG-3',
NNS 383 5'-GTTCTGCTGACCTGNNSCCGTCTAACCTGAC-3',
NNS 384 5'-CTGACCTGGCGNNSTCTAACCTGACCC-3',
NNS 409 5'-CTACATCCACTACGGTNNSCGCGAGTTCGGTAT-3',
NNS 466 5'-CGACTCCATCGGTNNSGGCGAAGACGG-3'

After centrifugation, all primers were resuspended in ddH_2O to a final concentration of 100 mM, based on yields stated in the DNA product analysis form. Each solution was diluted 40-fold in MilliQ water to give 25 ng/µl, prior to use in PCR.

4.2.1.6 Transformation of competent cells

Following QuickChange reactions and Dpn1 digestion, the PCR products (1 µl) were transformed into XL1-Blue cells (100 µl) by standard heat shock methods (section 2.3.13). The cells were then grown as streaked plates (section 2.3.1) on LB-Agar (Amp) plates. The resulting colonies were harvested together in 2-5 ml LB and the library plasmid mixture obtained by DNA miniprep. 1 µl of the resulting DNA mixture was then electroporated into 100 µl of TOP10 electrocompetent cells using a 0.1 cm^3 cold cuvette and electroporation setting EC1 (section 2.3.14).

4.2.1.7 Master library preparation and storage

A Qpix2 robot (Genetix Ltd.) was used to pick 90 colonies of each library (on average there were some 150-200 colonies per plate) in *E.coli* TOP10 into

90 wells of a single 96 deep-square well microtitre plate containing 500 µl of LB/Amp. Three wells were inoculated with *E.coli* TOP10 pQR791 (wild-type TK) to provide controls. Each plate was incubated at 37 °C and agitated at 1600 rpm for 16 hours using a Variomag plate mixer. Six identical reaction plates were then made by transferring 60 µl of the cell culture into 96-well microtitre plates (maintaining the same well formatting) and the remaining cell culture was used to produce two library master plates with 20% (v/v) glycerol by combining 60 µl of cell culture with 60 µl of 40% sterile-filtered glycerol solution and mixing thoroughly. The two library master plates and reaction plates were stored at -80 °C. The master plates were labelled by their library-site names e.g 'NNS23 glycerol' and designated as 'M' (master) or 'R' (replicate). All M plates were stored in a locked freezer and R plates kept with the corresponding reaction plates to enable replication of the reactions plates as necessary, while maintaining an untouched glycerol master stock plate (M).

4.2.1.8 Physicochemical analysis of library sites

4.2.1.8.1 B-factors

B-factors for each TK residue targeted in the libraries were obtained from the pdb file (1QGD.pdb) for the structure of TK from *E.coli* determined and obtained from Prof. Jennifer Littlechild, Exeter University. B-factors for each residue were calculated as the average of all atoms in the residue, and then averaged over both chains in the homodimeric structure. This was calculated for both 'whole residues', and for 'side-chain only'.

4.2.1.8.2 Solvent accessible surface area

Solvent accessible surface areas for every TK residue were calculated from the pdb file (1QGD.pdb) for the structure of TK from *E.coli* determined and obtained from Prof. Jennifer Littlechild, Exeter University. This was done using the program Discovery Studios Modelling 1.1.61.0 (Accelrys 2003) which calculates the solvent accessible surface of protein residues employing a probe. A probe radius of 1.4 Å was used to determine the degree of accessibility equivalent to a water molecule. The probe acts as an "eraser", moving from the outside of the protein inward, it removes free grid points until the openings are too small for it to move forward. Residues neighbouring the erased grid points are considered exposed, while the others are considered buried. The smaller the probe radius, the more exposed residues are defined. In this way, the total Surface Accessible Surface (SAS) for the whole protein, and then the total SAS, sidechain SAS, and percent SAS for each residue, were obtained.

4.2.1.8.3 Sequence entropy

A sequence alignment for 52 bacterial and yeast TK sequences was obtained (section 2.4.1) using the BioEdit software package (Hall 1999). This software was then used to calculate the sequence entropy ($H(x)$) using the bit measurement method as defined by Equation 4.1:

$$H(l) = -\Sigma f(b,l)\log(base2)f(b,l) \qquad Eq\ 4.1$$

where $H(l)$ is the uncertainty or entropy at position l, b represents a residue (out of the allowed choices for the sequence in question), and f(b,l) is the frequency at which residue b is found at position l. The information content of a position l, then, is defined as a decrease in uncertainty or entropy at that position.

4.2.2 Library validation and screening

4.2.2.1 DNA sequencing

The target regions of the TK gene in the plasmids extracted from at least ten individual colonies from each residue library were sequenced to assess the diversity of the libraries. DNA was extracted using Qiagen minipreps and standard protocols described in the manual. DNA cycle sequencing (section 2.3.24) was performed using the sequencing services at WIBR (Wolfson Institute for Biomedical Research, UCL). All DNA samples were given to the service at a concentration of 100 fmoles in 6 µl.

4.2.2.2 Statistical analysis of library diversity

The extent of mutagenic bias for each library was determined by comparing the frequency of occurrence of each codon observed by DNA sequencing of random clones, to the probability of obtaining each codon at that frequency. The probability (P) of obtaining no examples of a codon (frequency, F=0), from a random sample of X clones, is given by Equation 4.2:

$$P(F=0) = [(N-1)/N]^X \qquad Eq\ 4.2$$

where N is the total number of possible codon variants (64 for NNN, and 32 for NNS libraries). The probability of obtaining F examples of a codon, for $0<F\leq X$, is given by Equation 4.3:

$$P(F>0) = {}^{X}C_F\,[(N-1)/N]^{(X-F)}\,[1/N]^F \qquad Eq\ 4.3$$

The occurrence of a codon at a frequency that is less than 1% probable in an unbiased library is considered to indicate a bias towards the occurrence of that codon.

4.2.2.3 Screening on the model reaction

Every microwell reaction plate was screened using the high-throughput procedure described in (section 2.3.16.6 see also Figure 4.27). Plates were

thawed from -80 °C resulting in cell lysis. Substrates and cofactors were prepared at standard concentrations (sections 2.2.8 and 2.2.7 respectively) with 50 mM Tris-HCl buffer at pH7 (section 2.2.6.1). HPLC chromatograms were collected for all of the screening reactions following one hour of incubation at 25 °C.

4.2.2.4 Screening data analysis

Screening data was batch exported using Dionex Peaknet 5.1 software allowing whole HPLC runs to be extracted into Excel. The peak area data was then used to convert the quantity of substrate and product present at the end of the reaction to a level of conversion, using standard curves (Section 2.3.16.7 see also Figure 4.41). The calculation of conversion rather than using an absolute concentration of product allowed for any experimental error in the initial substrate concentration to be eliminated. This then removed any well-to-well variation due to pipetting of substrates. The inclusion of wild-type colonies and blanks in three wells each allowed plate-to-plate variations to be monitored.

4.2.3 High-accuracy characterisation of selected mutants

Individual mutants of interest identified from library microwell fermentation plates, based on high-throughput analysis, were screened using the procedure described in (section 2.3.16.1). HPLC chromatograms were collected for all of the screening reactions following up to one hundred hours of incubation at 25°C.

4.2.3.1 Model reaction conditions

All model reactions were performed at 50 mM final concentrations of HPA and GA using standard substrate solutions (section 2.2.8) and Tris-HCl buffer at pH7 (2.2.6.1) with standard cofactor solutions (section 2.2.7). Reactions were performed at 25 °C in 96-well microtitre plates.

4.2.3.2 Determination of protein concentration by bioanalysis

Samples were analysed using standard Protein 200 plus assay instructions. After priming and preparation of an Agilent chip (section 2.3.22), 6 µl of the protein samples of interest were loaded in triplicates or more onto the 10-lane chip. The Agilent 2100 bioanalyser, then uses capilliary electrophoresis to separate the proteins, and analyses the quantity of a protein-binding fluorescent dye present in each peak, all within 30 minutes. The data was then used in conjunction with a standard curve obtained with commercial TK samples, to ascertain the TK protein concentration of the samples analysed.

4.2.3.3 Determination of protein concentration by densitometry

Protein samples used in reactions were sonicated and 10 µl of each sample was electrophoresed using a precast 7.5% SDS-Page gel (Bio-Rad) in duplicate. Standard TK protein (5 µl) was added to one of the gel wells. The gel was then stained and destained with Coomassie blue (section 2.3.21.4) and then imaged using Imagesoft software which enabled the densitometry calculation (section 2.3.21.5) from the number of bands per lane on the gel, and their relative intensity. The results of a standard curve obtained using an independent Bio-Rad protein concentration assay enables the weighting given to each band to be converted to a concentration of each protein band present.

A BSA standard curve was obtained by first measuring the protein concentration with a Bradford assay. The dye reagent was prepared by diluting one part Dye Reagent Concentrate with four parts distilled deionised (DDI) water. It was then filtered through a Whatman #1 filter to remove any particulates (section 2.3.21.6). Five dilutions of the BSA protein standard were prepared, which are representative of the concentration range of the transketolase protein solution to be tested. The linear range of the assay for BSA is 0.2 to 0.9 mg/ml, whereas with IgG the linear range is 0.2 to 1.5 mg/ml. 100 µl of each standard is pipetted into a clean, dry cuvette. The standard BSA concentrations used were 0.2, 0.4, 0.6 and 0.8 mg/ml. Protein

solutions were assayed in duplicate and 5.0 ml of diluted dye reagent was added to each tube prior to vortexing. They were then incubated at room temperature for a minimum of 5 minutes. Absorbance increases over time so samples were incubated at room temperature for no more than 1 hour. After a fixed incubation period the samples were measured for absorbance at 595 nm. The reading obtained from the Bradford protein assay was then used to determine the total protein concentration within densitometry samples where the BSA standard curve was also used. Once the total protein content has been determined, the TK concentration can be found from its percentage weighting which was determined through the densitometry software.

4.3 Results

4.3.1 Mutator strains

XL1-red mutator strain cells proved to be problematic to grow, even when using media enriched with magnesium chloride and glucose. Cell viability was usually not maintained beyond 60 generations, and miniprep DNA obtained from these cells contained no plasmid, appearing only as a 'smear' of DNA by agarose gel electrophoresis. Miniprep plasmid DNA from earlier generations was transformed into XL1-blue cells and plasmid DNA from individual colonies were sequenced. However, no evidence of mutagenesis was found and the method was abandoned.

4.3.2 Error-prone PCR (EpPCR)

EpPCR was performed to mutate the entire TK gene using the two primers TK-15mutF and TKC-term, using the conditions and PCR protocol in section 4.2.1.2. The best conditions for PCR were found to be: 1.0 mM $MgCl_2$ with 0.05 mM $MnCl_2$; and 1.0 mM $MgCl_2$ with 0.1 mM $MnCl_2$ (Figure 4.1). Both these conditions gave good clear product bands when run on 0.6% agarose gels for PCR product analysis although the 0.1mM manganese products appear the best in terms of quantity/quality of signal.

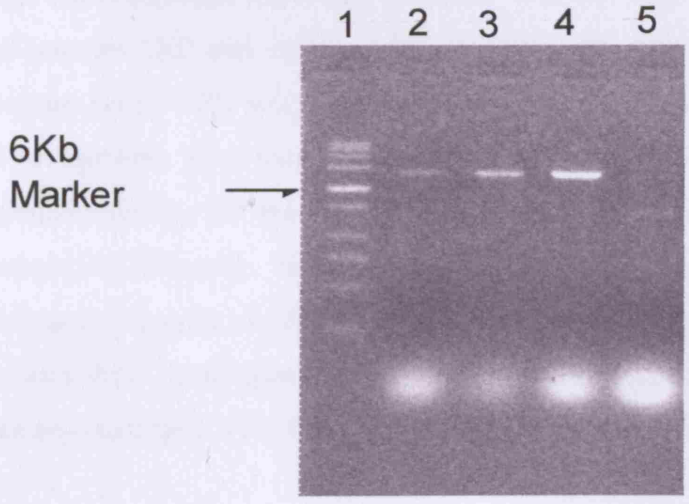


Figure 4.1 Error-prone PCR products for the TK plasmid. Lane 1 shows MW markers, lane 2 shows 1 mM $MgCl_2$ with 0.025 mM $MnCl_2$, lane 3 shows 1.0 mM $MgCl_2$ with 0.05 mM $MnCl_2$, lane 4 shows 1.0 mM $MgCl_2$ with 0.1 mM $MnCl_2$ and lane 5 shows 1.0 mM $MgCl_2$ with 0.2 mM $MnCl_2$.

4.3.3 Ligation of EpPCR products to capture vectors

Initially an attempt was made to modify the whole TK gene using EpPCR and subsequently ligate the gene back into the same vector (gel extracted). A double digest on a miniprep of the unmodified plasmid was performed using enzymes XbaII and BglII for 4 hours and then heat inactivated at 65°C and run on a 0.6% low melting point agarose gel. The band corresponding to the vector (~3kb) was then excised from the gel using a UV plate and a scalpel and the DNA gel extracted (section 2.3.8). However, the ligation and transformation efficiencies achieved were low and the amount of vector recovered after gel extraction (Qiagen), even when repeated and products pooled, was not sufficient to perform a good efficiency for ligation.

As a result of the difficulties recovering the original vector, two vectors were ordered: TOPO and PCR script SK+. The PCR products for both EpPCR conditions were ligated into the TOPO vector first due to its apparent high efficiency as the result of TOPOisomerase being present in the ligation and increasing the number of inserts which get ligated into the vector. The final EpPCR conditions were kept the same with the only difference being the use of primers TKC and TKN in order to include the TK constitutive promoter for capture into TOPO with the aim of making it into an expression vector for TK. The ligations were very successful with 9 out of 10 random transformants picked showing the correct gene size when a double digest was performed to extract the TK gene. However this system is not designed to be an expression vector and despite cloning in the TK promoter, numerous SDS page gels of the clones from heat shock transformations (section 2.3.13) and some HPLC assays (section 2.3.16) failed to show over-expression of TK.

As TOPO was not very successful in expression it was decided to go back to the methods used to construct the original pQR711 vector with the PCR script SK+ capture vector. It has been shown by French and Ward (1995) that over expression is readily achievable in this system and gives similar expression levels which ever orientation the gene ligates into the vector with. The primers used for this PCR were TKC-term and TKN and the conditions were same as for the previous EpPCR. Again however the number of colonies seen after transformation was disappointing with less than 50 colonies on the plate.

Various library construction techniques were tested as mentioned with varying degrees of success, but it was still difficult to produce a large number of transformants. As a result, I began to focus more on site-saturation mutagenesis (SSDM) which provided a simpler, ligation-free method. The SSDM method was considered most appropriate for targeting mutagenesis towards sites that directly affect substrate specificity and would enable me to test the hypothesis that phylogenetic variations in nature and/or structure models provided useful means of defining target residues.

4.3.4 Focused Mutagenesis

4.3.4.1 Definition of libraries

Two sets of ten active-site libraries were constructed in parallel, in which the different amino-acid target sites chosen for modification were determined by two different approaches. The first set was derived from comparison of the conservation and variation during evolution observed in 54 different bacterial and yeast TK sequences. Residues were chosen within 10 Å of the erythrose-4-phosphate substrate when bound to the TK active-site, that were also found to vary during evolution (Sean Costelloe, personal communication). The second approach determined ten sites that were simply within 4 Å (average H-bond distance) of the erythrose-4-phosphate substrate when bound to the TK active-site (Ed Hibbert, personal communication).

These different approaches were chosen to reflect alternative potential starting points, and to determine whether structurally conserved residues

should be included. The aim was also to determine whether residues with direct contact to substrates are more or less relevant than so called second shell residues that do not directly interact with substrates and products.

The first potential starting point is the case where no or very little structural information is available. For example, lack of a structure, a homology model, or a structure without substrate bound, makes the use of sequence conservation analysis potentially more useful. By contrast, the availability of a structure with substrate/product bound to the active site provides potentially rich information regarding the residues which actually interact with the substrates/products. It is of course possible that residues not directly interacting with the substrate are equally important. These residues would be more difficult to identify by structure alone and so perhaps a combination of the two approaches may be more useful.

The locations of 'first-shell' residues identified using only the structural proximity to substrate are shown in Figure 4.2. For comparison, the locations of the 'second-shell' residues identified using the phylogenetic analysis are also shown in Figure 4.2, and the evolutionary pathway determined by Sean Costelloe (personal communication) is shown in Table 4.1.

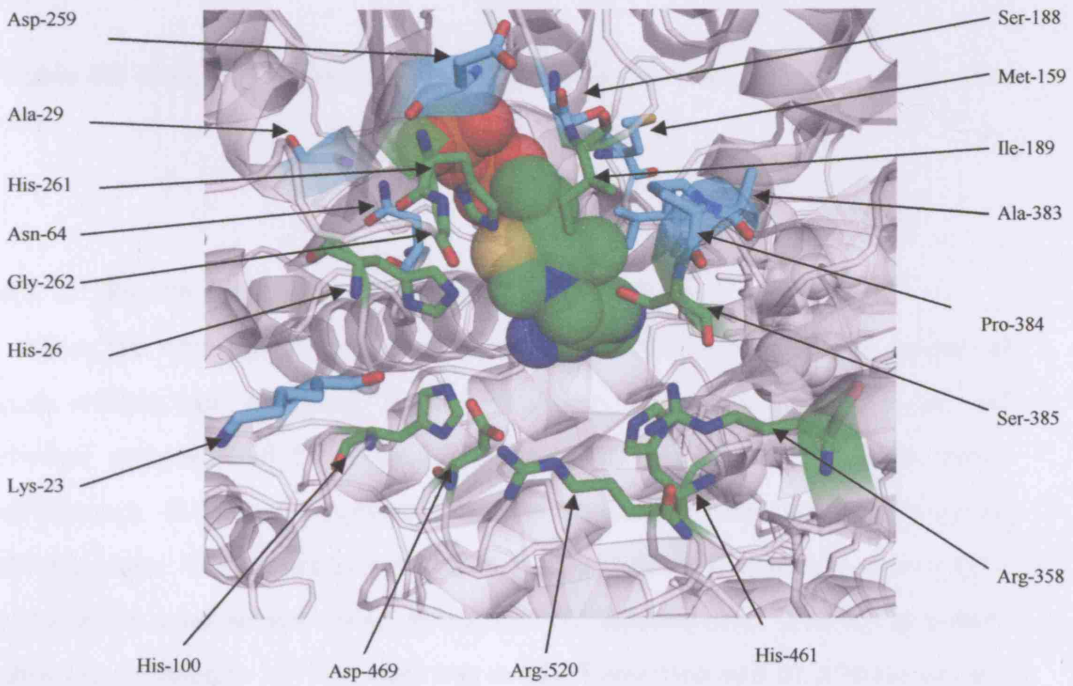
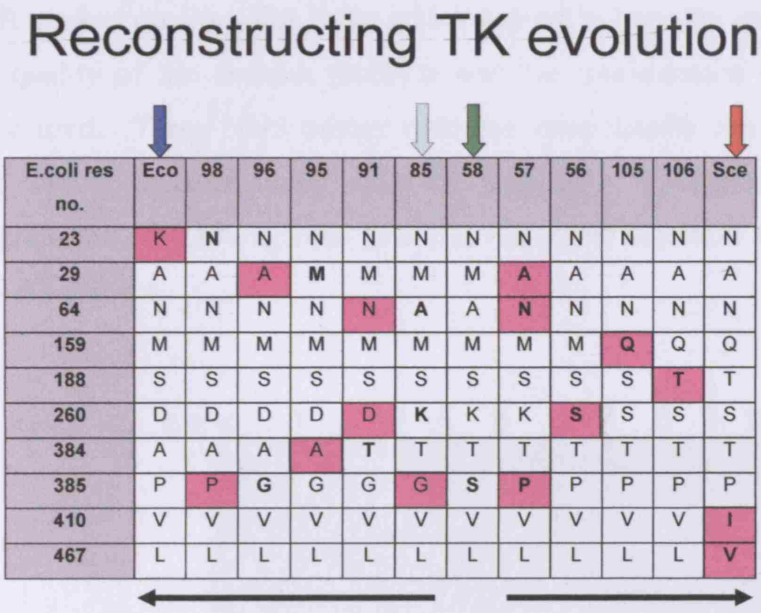


Figure 4.2. Location of the ten structurally defined sites (green), and the phylogenetically defined sites (cyan), relative to the TPP-cofactor in the *E.coli* TK active-site (from 1QGD.pdb structure). Structurally defined 'first-shell' sites are His26, His100, Ile189, His261, Gly262, Arg 358, Ser385, His461, Asp469, Arg520. Phylogenetically defined 'second-shell' sites are K23, A29, N64, M159, S188, D259, A383, P384, V409 and L466.



E.coli res no.	Eco	98	96	95	91	85	58	57	56	105	106	Sce
23	K	**N**	N	N	N	N	N	N	N	N	N	N
29	A	A	**A**	**M**	M	M	M	**A**	A	A	A	A
64	N	N	N	N	**N**	**A**	A	**N**	N	N	N	N
159	M	M	M	M	M	M	M	M	M	**Q**	Q	Q
188	S	S	S	S	S	S	S	S	S	S	**T**	T
260	D	D	D	D	**D**	**K**	K	K	**S**	S	S	S
384	A	A	A	**A**	**T**	T	T	T	T	T	T	T
385	P	**P**	**G**	G	G	**G**	**S**	**P**	P	P	P	P
410	V	V	V	V	V	V	V	V	V	V	V	**I**
467	L	L	L	L	L	L	L	L	L	L	L	**V**

Table 4.1 Phylogenetic basis for selection of library sites

4.3.4.2 Site-Directed Saturation Mutagenesis (SSDM)

Once the sites had been chosen, primers to facilitate the changes desired at each residue were designed. A complementary pair of primers, (forward and reverse complement) for each set of ten sites were made, but with some differences. For the phylogenetic sites the degenerate NNS codon (where N is an equimolar AGCT mixture, and S is an equimolar GC mixture) was used to provide the randomised codon at each of the desired sites. For the structural sites the degenerate NNN codon was used. These two sets of primers encode significant differences at the amino-acid level, but as a result of the wider

selection of codons accessible with NNN primers, more stop codons are intrinsically possible. The template DNA used for library construction was pQR791 which includes a His6-tagged N-terminal fusion to TK for ease of purification of interesting mutants (J. Aucamp unpublished).

A standard quick-change protocol (section 2.3.12) was used to produce the libraries with minor modifications to the method required from primer to primer. Separate PCR thermo-cycling of the sense and antisense primer strands, prior to a final PCR which combined these products did not improve the PCR product quality. The factor which seemed to have the greatest impact on the quality of the reaction products was the concentration of the DNA template used. Three NNS primer reactions were usually run in triplicate (varied DNA concentrations) and in parallel. Examples of these, electrophoresed on 0.8% agarose gels after the *DpnI* digestion are shown in Figures 4.3 and 4.4.

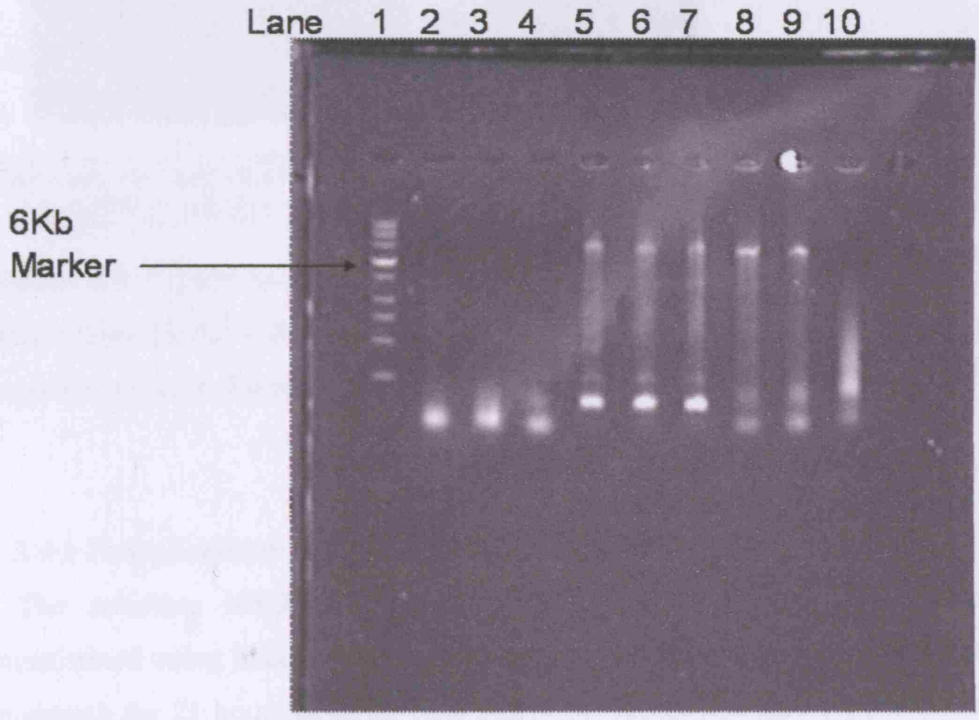


Figure 4.3: Examples of *Dpn*I digested products of SSDM Quikchange PCR, with 5 µl sample per lane. The brightest MW marker (labelled) is 6kb so products just above this marker are of the expected size. Lane 1: MW

markers, lane 2: NNS466 5ng pQR791 template, lane 3: NNS466 10ng, lane 4: NNS466 25ng, lane 5: NNS23 5ng, lane 6: NNS23 10ng, lane 7: NNS23 25ng, lane 8: NNS64 5ng, lane 9: NNS64 10ng and lane 10: NNS64 25ng template.

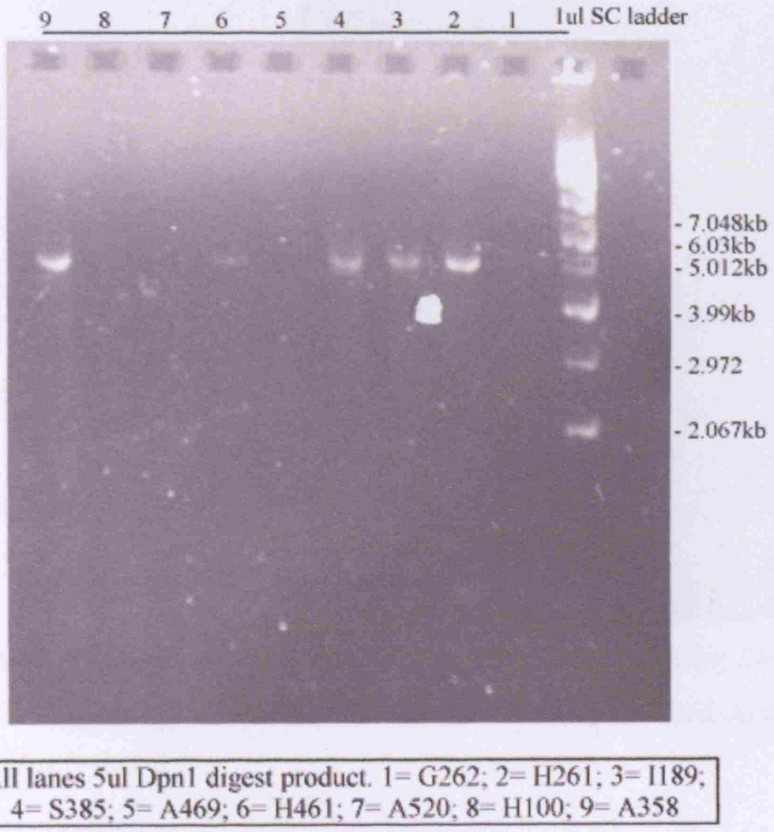


Figure 4.4: Examples of *Dpn*I digested products of nine different SSDM Quikchange PCR, with 5 µl sample per lane. The target residue for each sample is noted in the box legend.

4.3.4.3 Transformation and master library preparation

The resulting NNS/NNN QuickChange™ products from above were transformed using heat shock (section 2.3.13) into *E.coli* Xl1-blue cells prior to growth for 21 hours at 37 °C on a Petri dish of nutrient agar supplemented with Ampicillin. At this stage some randomly picked colonies were used to inoculate 5-10 ml of LB/Amp and incubated overnight (section 2.3.2). Overnight cultures were used to perform SDS-PAGE analysis (section 2.3.21) to make sure that TK is being constitutively expressed as seen in Figure 4.5.

DNA was also obtained by miniprep (section 2.3.6) for a few randomly picked colonies and sent for DNA sequencing (section 2.3.24) to analyse the success rate of the mutagenesis.

Figure 4.5: SDS-PAGE of randomly picked colonies from the NNS23 library in XL1-blue. TK runs close to the 75kDa marker indicated on the gel.

The remaining colonies were then harvested together and their library plasmid DNA was extracted and transformed into the *E.coli* expression strain TOP10 (Stratagene Ltd) for the phylogenetically defined sites, and into *E.coli* XL10 (Stratagene Ltd) for the structurally defined sites. The switch to XL10 was undertaken after it was found that the plasmids in the TOP10 strain were not stably maintained (see section 4.3.4.4 below). Colonies were then picked after overnight growth on LB/Amp agar plates using a Qpix colony picker (Genetix, UK) to inoculate each well of a 96-deep-well plate, with 90 wells containing variants and 6 control wells comprising 3 blank wells and 3 'WT' unmodified PQR791 TOP10 clones (A6 WT, A7 blank, E6 WT, E7 blank, H11 blank and H12 WT). Incubation of plates at 37 °C, agitated at a rate of 1600 rpm for 16 hours using a Varimag plate mixer provided uniform cell growth across each plate. The quality and diversity within each library was

determined by sequencing (section 2.3.23) of more randomly chosen variants, and by initial screening against the TK model reaction between HPA and GA.

4.3.4.4 Initial Screening on the model reaction

Initial screening with the model (HPA + GA) reaction was carried out on the libraries to ascertain the quality of the variants produced by SSDM, and to determine which variants are of interest for more in-depth analysis, for example those with improved activity. Reactions and HPLC analysis was carried out as described in section 4.2.2.3, allowing a 96-well plate to be analysed in a few hours. The results of each library plate screen using this method are shown in Figures 4.6-4.14 for the phylogenetically defined libraries expressed in TOP10. The three spiked wild-type colonies are shown as red data points for comparison. Data is represented as conversion, which ranges from 0 to a complete conversion of 1. The data points have been ordered into ascending order as a resulty are not presented in the order found on a plate format.

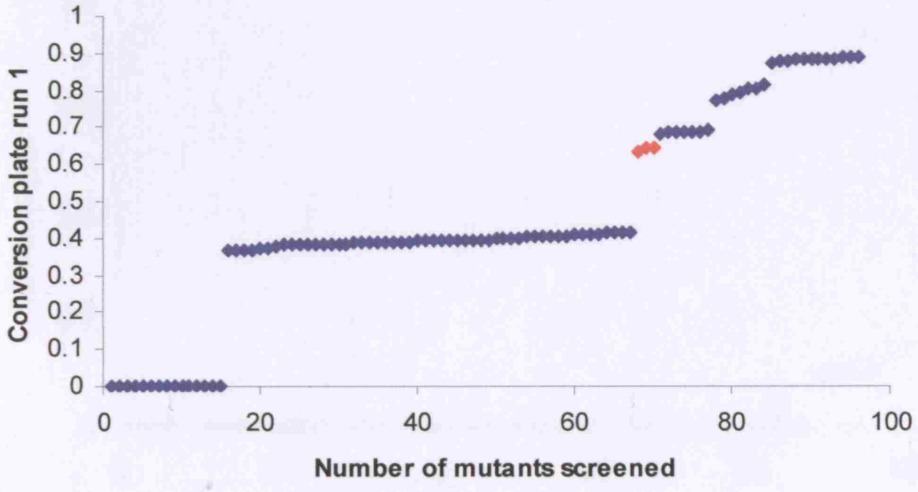


Figure 4.6: Conversion after one hour for the K23X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

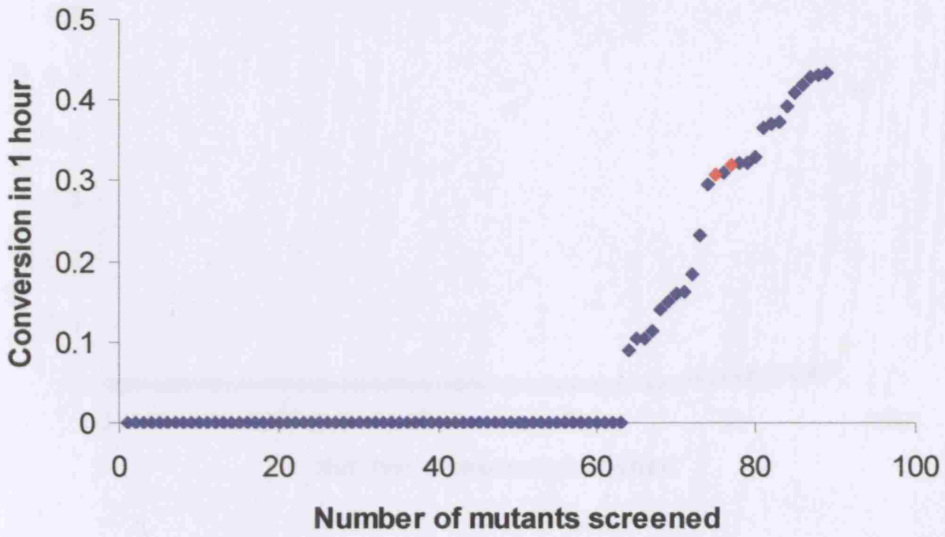


Figure 4.7: Conversion after one hour for the A29X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

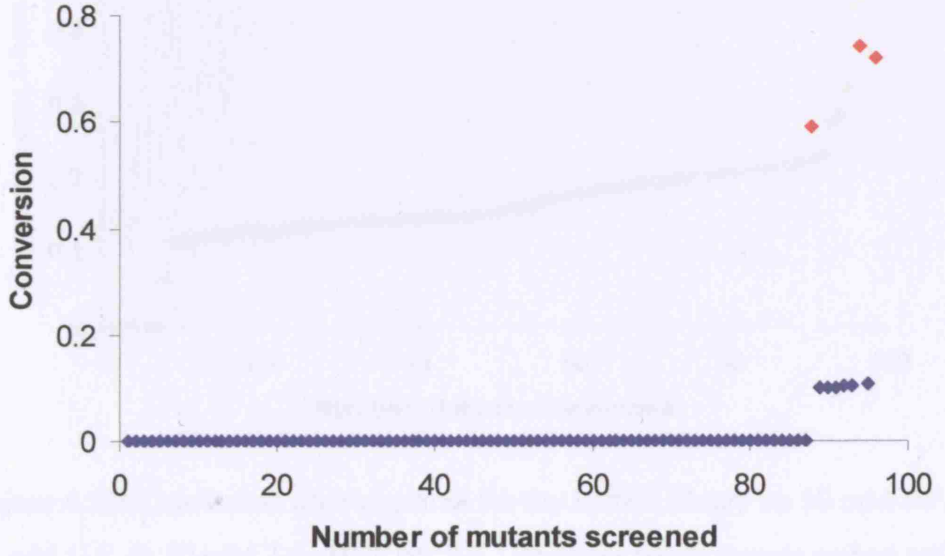


Figure 4.8: Conversion after one hour for the N64X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

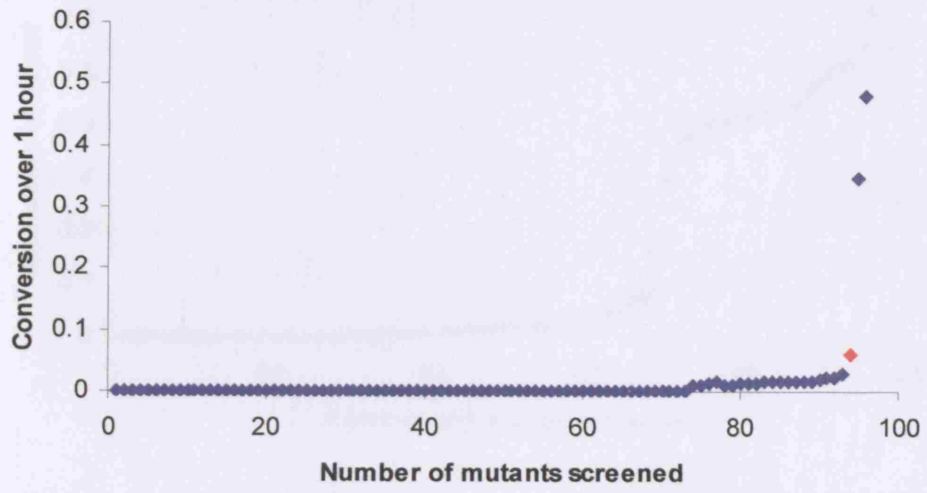


Figure 4.9: Conversion after one hour for the M159X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

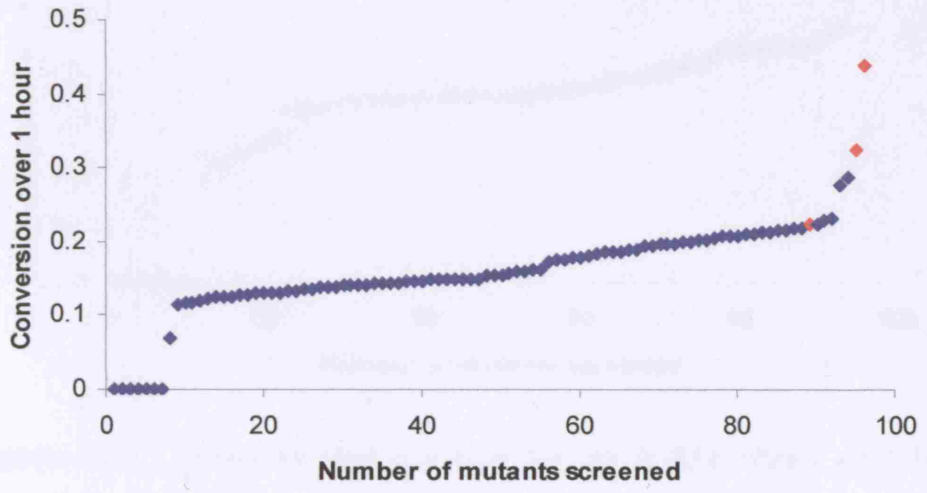


Figure 4.10: Conversion after one hour for the S188X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

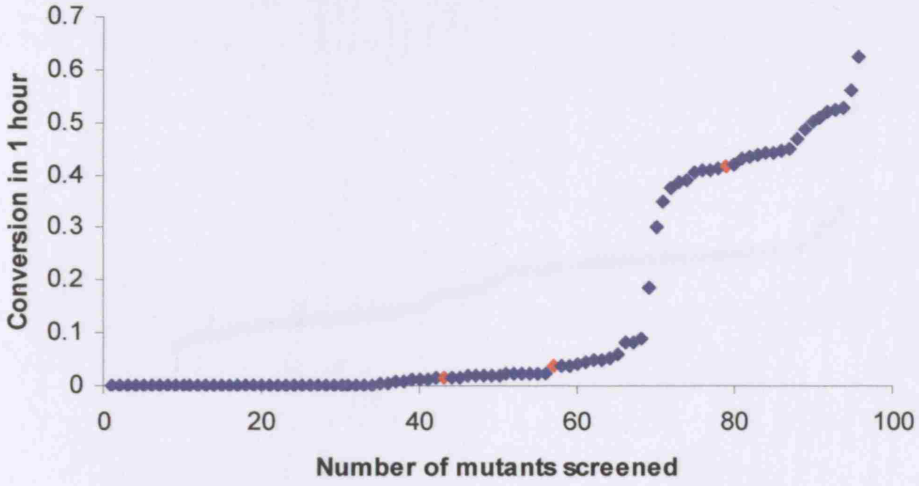


Figure 4.11: Conversion after one hour for the D259X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

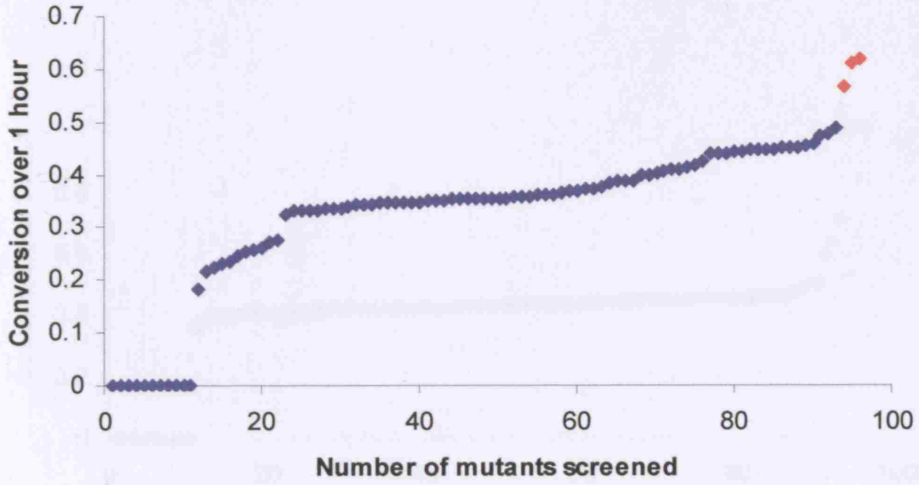


Figure 4.12: Conversion after one hour for the A383X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

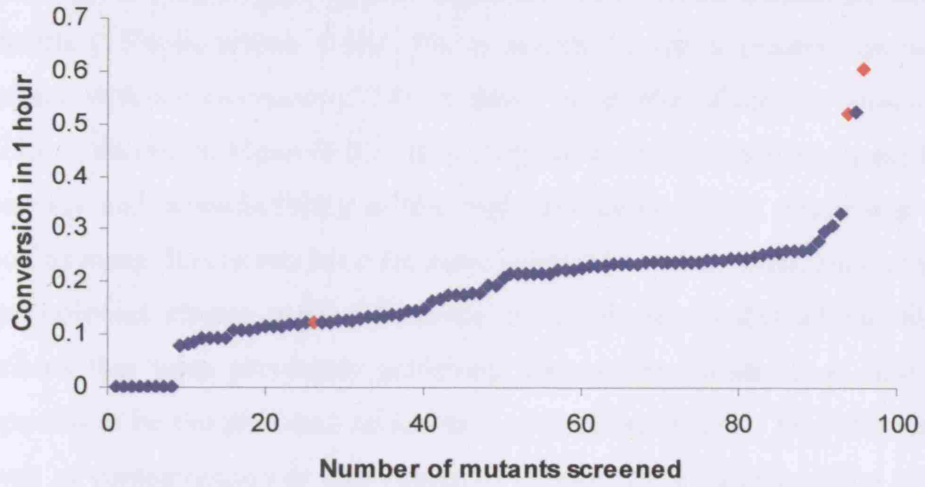


Figure 4.13: Conversion after one hour for the P384X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

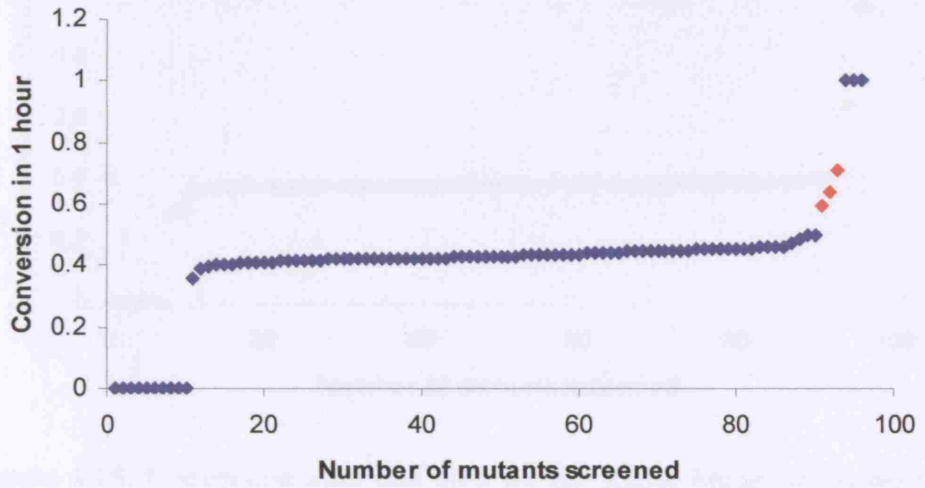


Figure 4.14: Conversion after one hour for the V409X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

Two of the libraries, K23X and V409X appear to have clustered data points at a conversion of around 0.4 (40%). To analyse this further, the K23X plate was recultured from the master plate and rescreened (Figure 4.15). After these

further generations of growth, there appeared to be a lower number of inactive mutants (15% in screen 1 and 7% in screen 2), but a greater number of mutants with a conversion of 0.4. A direct parity plot of the two screens for K23X is shown in Figure 4.16. It is clear from this plot that the underlying accuracy and reproducibility of the high-throughput HPLC assay was very good as many data points have the same activity in both screens, and the wild-type colonies cluster well. However, a significant number of the library variants that were previously achieving conversions greater than wild-type appeared to be unstable and shifted to a conversion of 0.4. This may be the result of contamination or enrichment of cultures by a faster growing variant which becomes more apparent after further generations of growth. It is possible that the V409X library suffered the same problem. Further analysis of the library quality was undertaken by DNA sequencing of random variants.

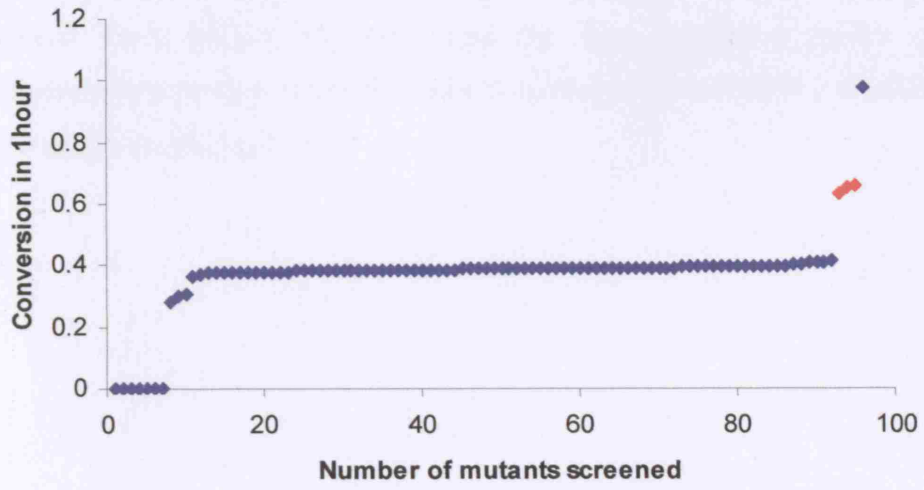


Figure 4.15: Conversion after one hour for the K23X library (regrown from master plate) on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

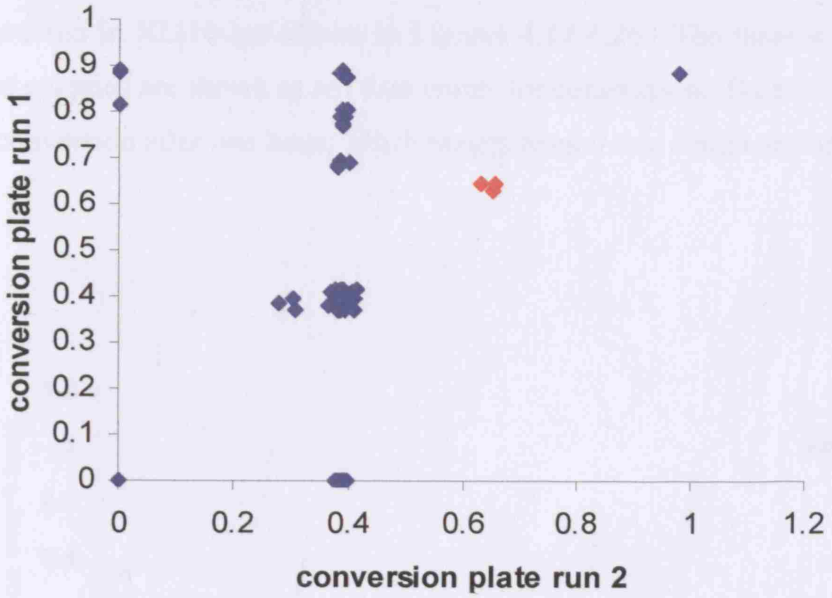


Figure 4.16: Comparison of conversions obtained from the first generation reaction plate for K23X, with those obtained from a second generation plate cultured for a further 16 hours from the first generation master plate. Conversions are measured for the model reaction of 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0.

The results of each library plate screen for the structurally defined libraries expressed in XL-10 are shown in Figures 4.17-4.26. The three spiked wild-type colonies are shown as red data points for comparison. Data is represented as conversion after one hour, which ranges from 0 to a complete conversion of 1.

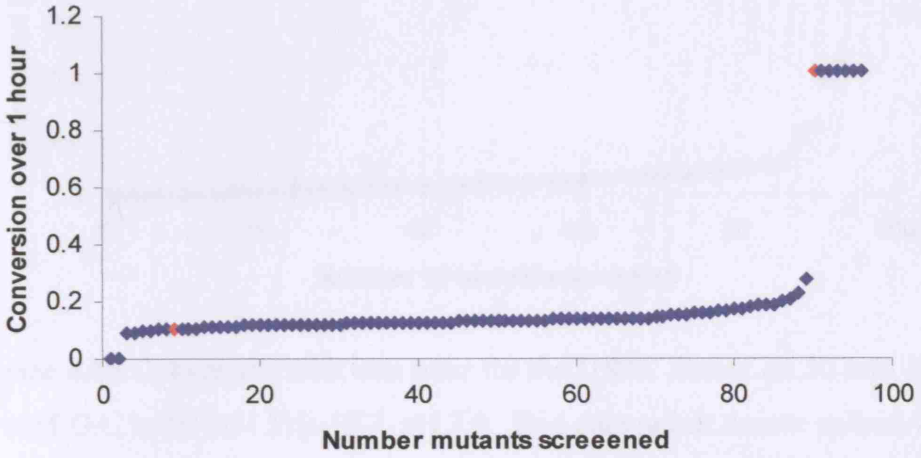


Figure 4.17 Conversion after one hour for the H26X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

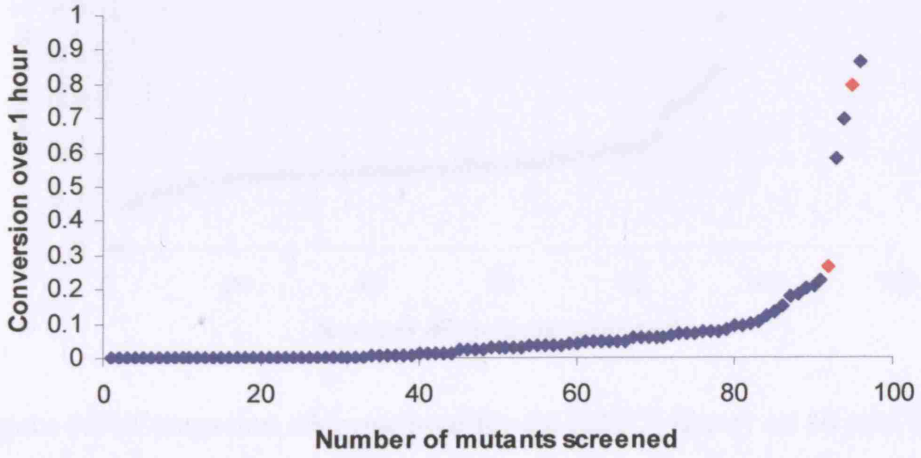


Figure 4.18 Conversion after one hour for the H100X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

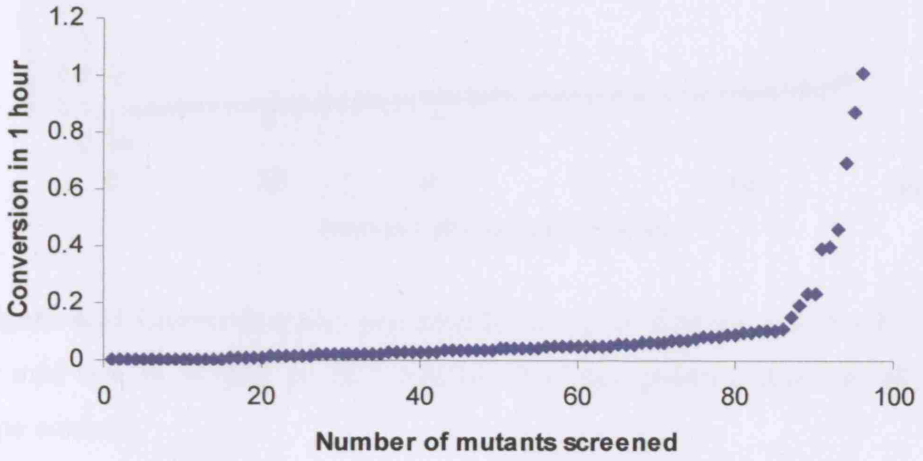


Figure 4.19 Conversion after one hour for the I189X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

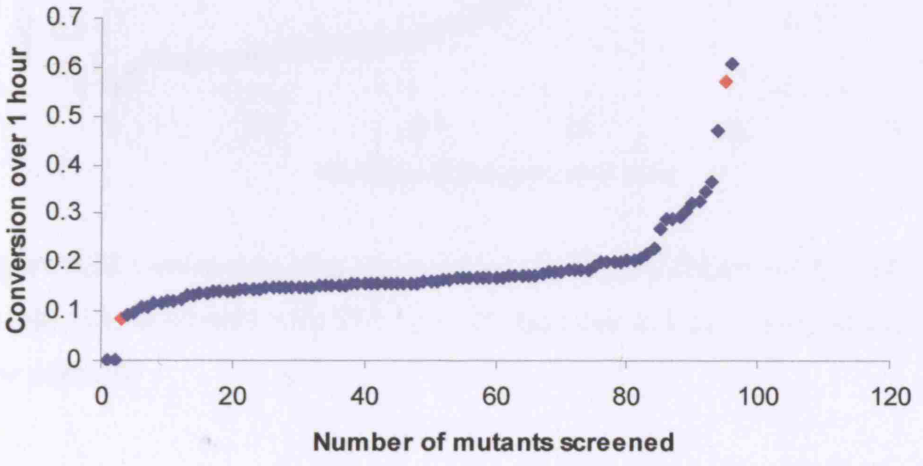


Figure 4.20 Conversion after one hour for the H261X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

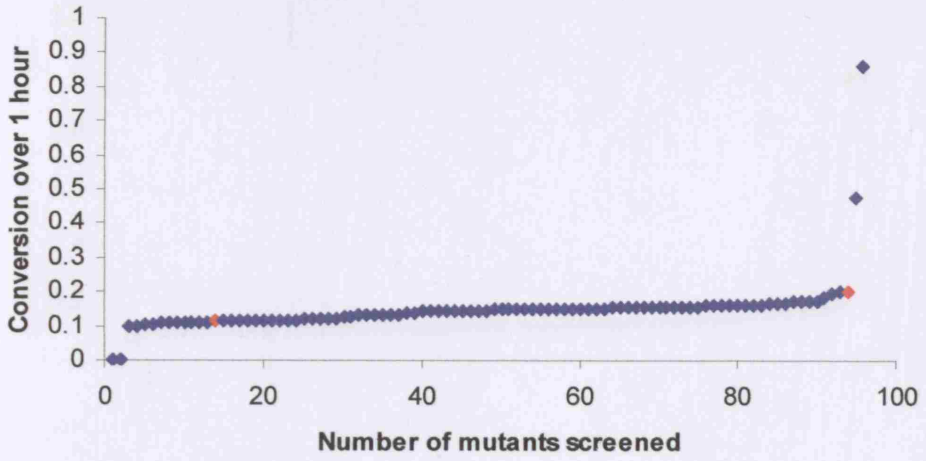


Figure 4.21 Conversion after one hour for the G262X library on 50 mM HPA, 50 mM GA, in 50 mM Tri-HCl, pH 7.0. Red data points denote spiked wild-type controls.

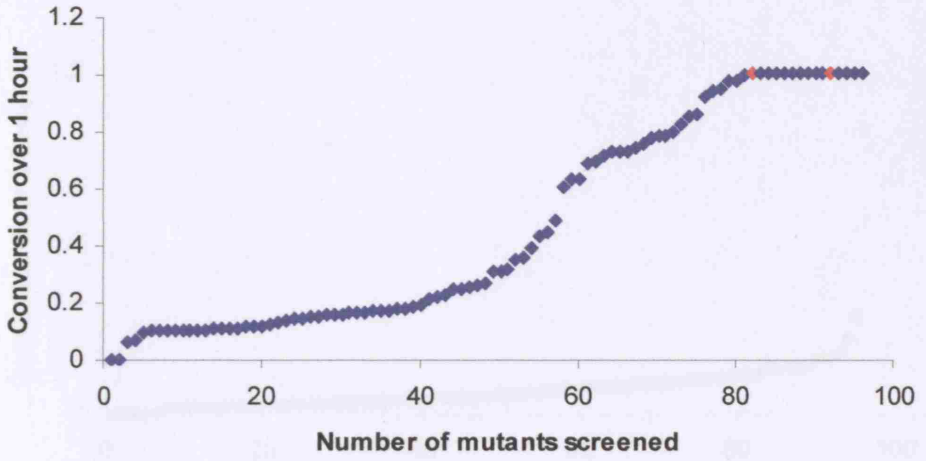


Figure 4.22 Conversion after one hour for the R358X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

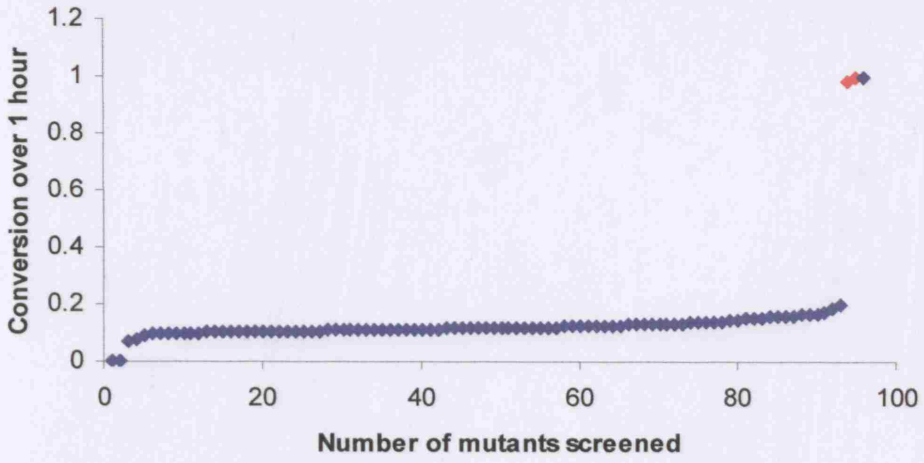


Figure 4.23 Conversion after one hour for the S385X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

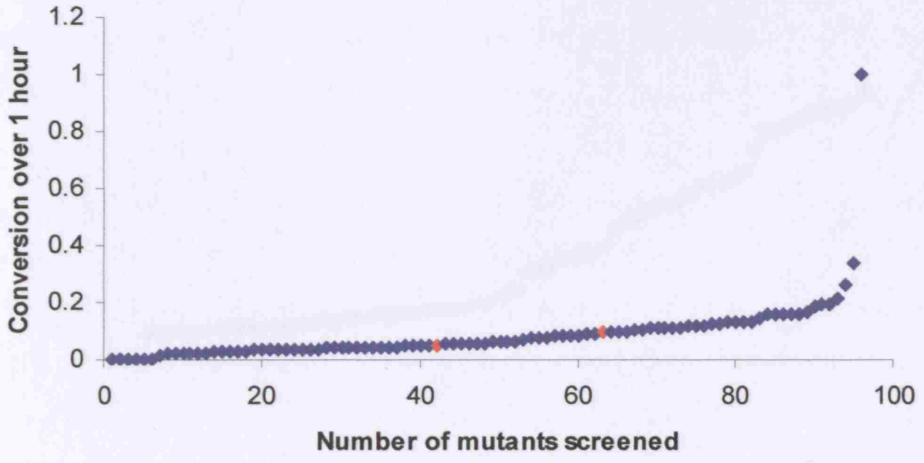


Figure 4.24 Conversion after one hour for the H461X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

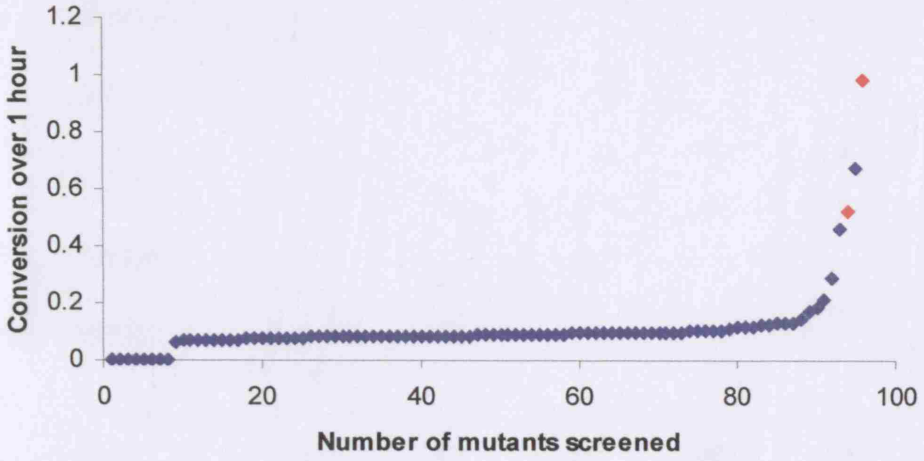


Figure 4.25 Conversion after one hour for the D469X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

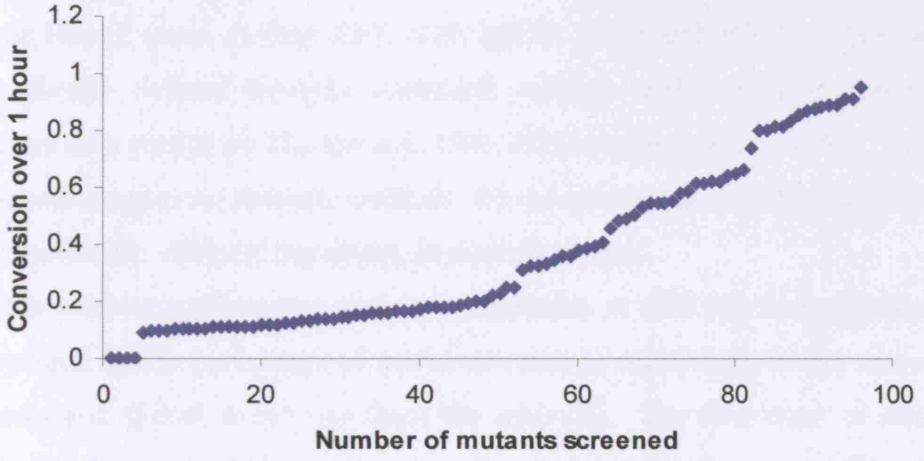


Figure 4.26 Conversion after one hour for the R520X library on 50 mM HPA, 50 mM GA, in 50 mM Tris-HCl, pH 7.0. Red data points denote spiked wild-type controls.

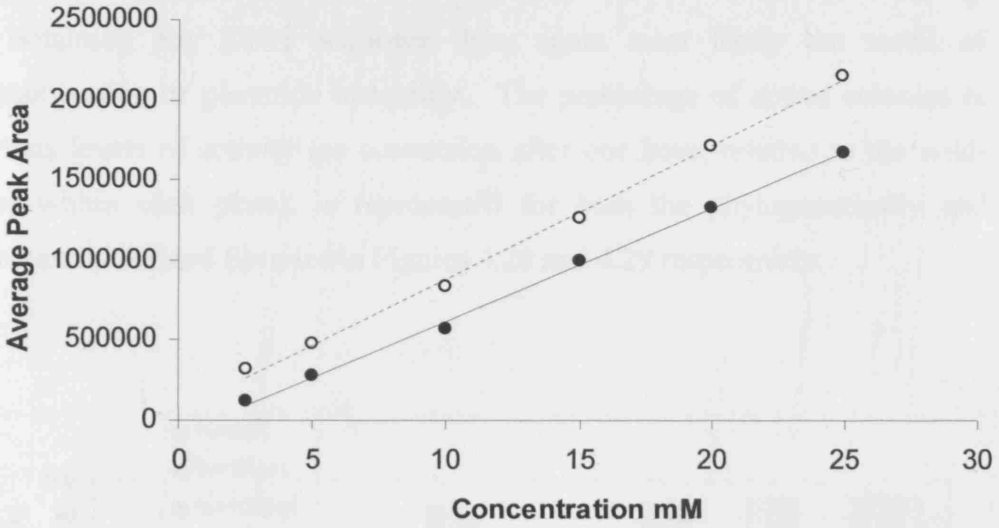


Figure 4.27 High-throughput HPLC calibration curves for L-erythrulose (●) and β-HPA (○). Above 25mM the relationships become more curved.

Five of the phylogenetically defined libraries contained variants with activities potentially greater than the wild-type TK obtained from within the same library plate, namely K23, A29, M159, D259 and V409. Three of the structurally defined libraries contained variants with activities potentially greater than wild-type TK, namely I189, R358 and potentially H461. Further analysis using more accurate methods, for the interesting variants was required to confirm the effect of mutations, as described later.

The tolerance of enzyme activity to mutation at each site is approximately measured by the percentage of active colonies in each library (after removing blanks and spiked wild-types from the analysis). The frequency of inactive colonies (below the limit of detection) for phylogenetically defined libraries is 3% for K23X, 72% for A29X, 93% for N64X, 77% for M159X, 8% for S188X, 36% for D259X, 6% for A383X, 4% for P384X, and 5% for V409X. The frequency of inactive colonies for structurally defined libraries is 1% for H26X, 42% for H100X, 18% for I189X, 1% for H261X, 3% for G262X, 14% for R358X, 73% for S385X, 9% for H461X, 0% for D469X, and 4% for R520X. The data for K23X, and G262X are excluded from all subsequent analyses. The K23X library contained poor sequence diversity combined with a high level of apparently active K23Z mutants suggesting contamination.

The G262X library had very poor activities overall, combined with a difficulty in obtaining any DNA sequence data, again most likely the result of contamination or plasmids instability. The percentage of active colonies at various levels of activity (as conversion after one hour, relative to the wild-type within each plate), is represented for both the phylogenetically and structurally defined libraries in Figures 4.28 and 4.29 respectively.

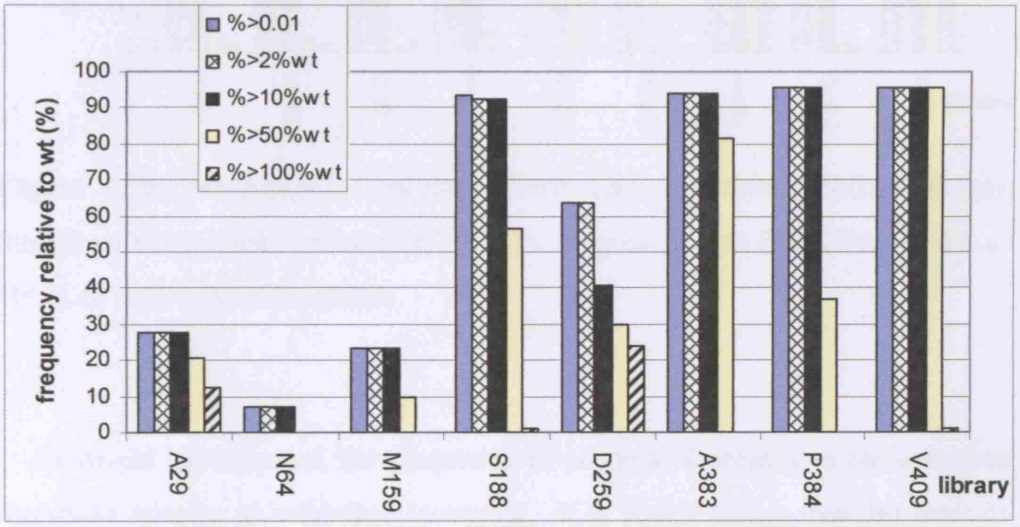


Figure 4.28: Percentage of colonies within each phylogenetically defined library that show measurable conversion (>0.01), or greater than 2%, 10%, 50% and 100% of wild-type conversion.

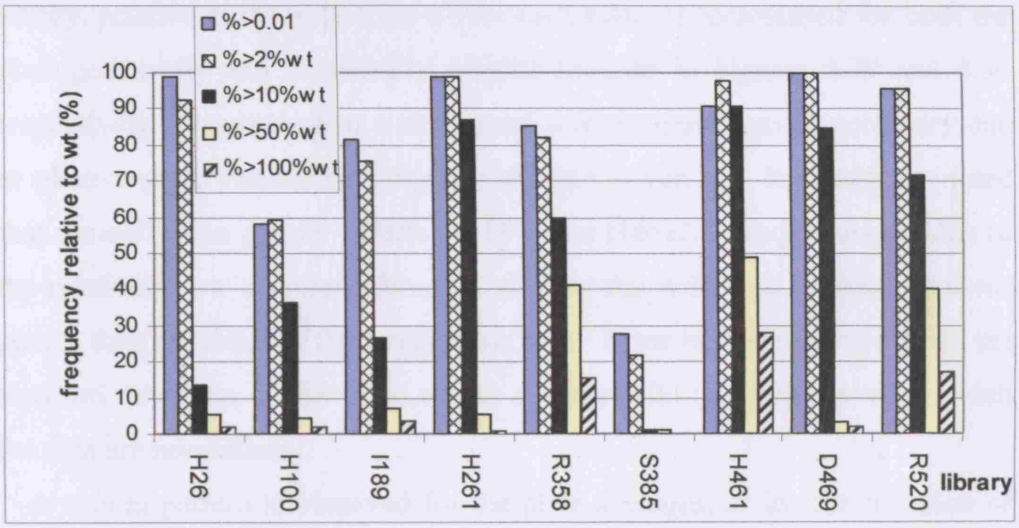


Figure 4.29: Percentage of colonies within each structurally defined library that show measurable conversion (>0.01), or greater than 2%, 10%, 50% and 100% of wild-type conversion.

As would be expected, the frequency of mutants decreases as the selection threshold relative to wild-type increases. It is worth noting that the limit of detection (fractional conversion of 0.01) is typically at around 2% of the average wild-type conversion. The structurally defined libraries appear to retain high percentages of active clones, at low conversion thresholds, in more cases (all but S385X) than for the phylogenetically defined libraries (only S188X, D259X, A383X, P384X and V409X). Conversely however, the percentage of active clones in each structurally defined library appears to drop off much more noticeably than for the phylogenetically defined libraries, at the higher thresholds of activity relative to wild type. In the structurally defined libraries, the H461X, R358X and R520X retain the greatest number of active clones at higher selection thresholds. The case of H461X may be skewed by a low average conversion for wild-type found in that library, as discussed below. In the phylogenetically defined libraries, A29X, S188X, D259X, A383X, P384X, and V409X retain the greatest number of active clones at higher (>50% relative to wild type) selection thresholds.

The average activity (as conversion after one hour) for the entire of each library, relative to the wild-type within each plate, is represented for both the phylogenetically and structurally defined libraries in Figures 4.29 and 4.30 respectively. Normalisation with internal wild-type averages is necessary due to plate-to-plate variation of high-throughput screening. It should be noted that normalisation greatly affects M159X and H461X library averages due to the relatively low averages observed also for the wild-type colonies in these plates, thus resulting in the larger error bars. Error bars are derived from the standard deviation of the mean of the average wild-type activity with which the data are normalised.

A similar pattern is observed for the plate averages, as for the retention of activities at high selection thresholds. For the structurally defined libraries, H461X, R358X and R520X have the highest overall averages relative to wild-type. It is worth noting here that these three residues interact with the phosphate groups of the natural substrates for transketolase, and that the phosphorylated substrates are not used in this study. For the phylogenetically defined libraries, A29X, M159, S188X, D259X, A383X, P384X and V409X have the highest overall averages relative to wild-type.

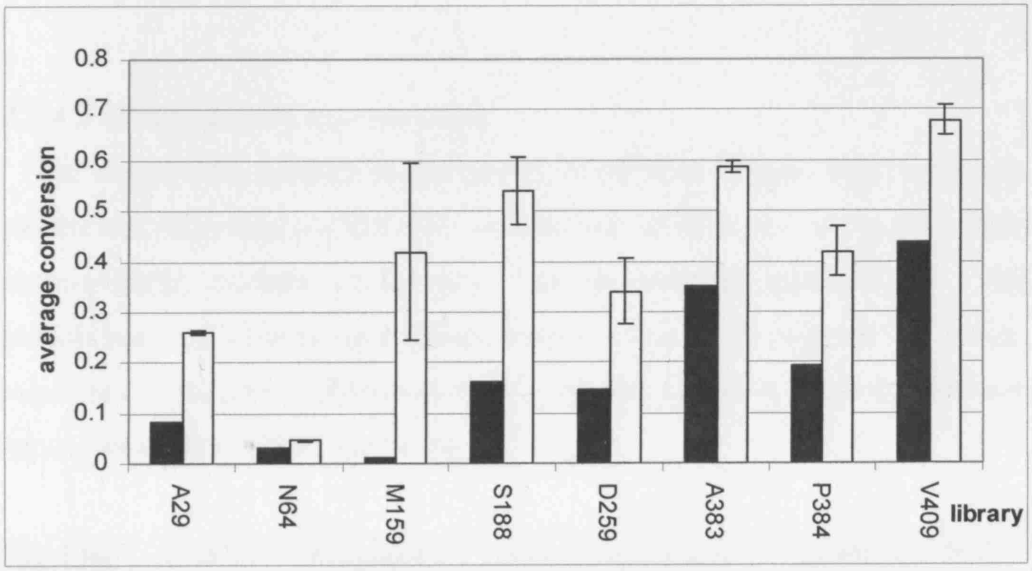


Figure 4.30: Average conversions (filled boxes), and average conversions relative to internal wild-type average (open boxes), within each phylogenetically defined library.

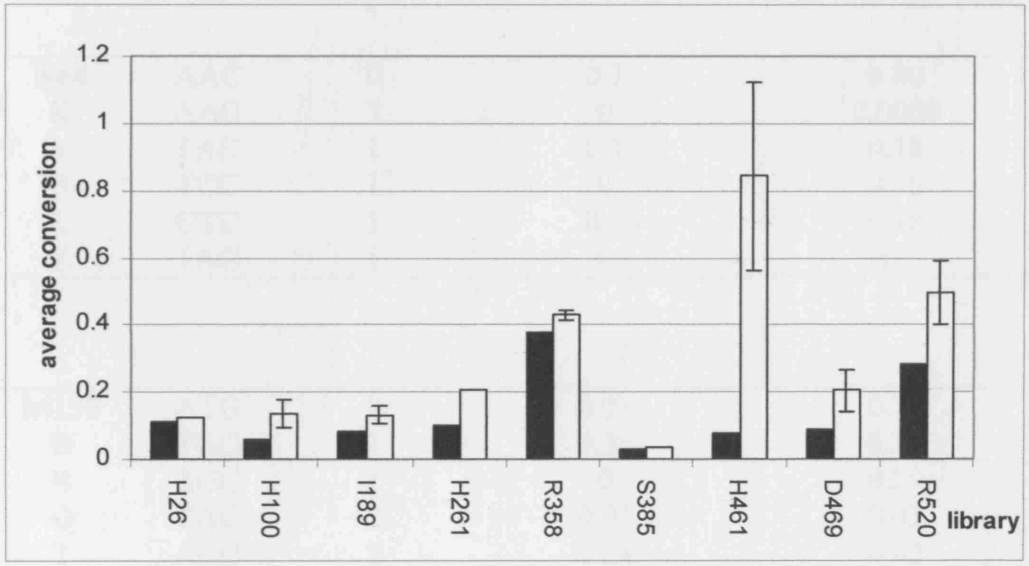


Figure 4.31: Average conversions (filled boxes), and average conversions relative to internal wild-type average (open boxes), within each structurally defined library.

4.3.4.5 Library quality determination

The amino-acids present at the target-site of each library, from randomly picked and sequenced variants are summarised in Table 4.2, along with their corresponding average conversion, and the codons represented. The probabilities of observing a given codon for a given number of times, calculated as described in Section 4.2.2.2, are also shown in Table 4.2, and are dependent on the sample size taken.

Residue	Codons	Frequency	average conversion	probability P(F)
K23	AAA	0	0.65	0.80
L	CTG	4	0.54	**$3x10^{-5}$**

Z	TAG	3	0.38	0.0009

A29	GCC	0	0.31	0.68
D	GAC	3	0.35	0.005
E	GAG	6	0.4	**$7x10^{-7}$**
Q	CAG	1	0.43	0.26
T	ACC	2	0.3	0.05

N64	AAC	0	0.7	0.80
K	AAG	3	0	0.0009
Y	TAC	1	0.1	0.18
S	TCC	1	0	0.18
L	CTC	1	0.1	0.18
Z	TAG	1	0	0.18

M159	ATG	0	0.03	0.78
W	TGG	1	0.35	0.20
S	AGC	1	0	0.20
Q	CAG	2	0.02	0.02
T	ACG	2	0.24	0.02
K	AAG	1	0.03	0.20
Z	TAG	1	0.02	0.20

S188	TCT,TCG	1,1	0.33	0.71
T	ACC	1	0	0.25
R	CGG	1	0.23	0.25
Q	CAG	1	0.28	0.25
Z	TAG	2	0.17	0.04
P	CCG,CCC	1,2	0.21	0.25, 0.04
Y	TAC	1	0.18	0.25

D259	GAC	1	0.40	0.37
Y	TAC	8	0.36	**$1x10^{-5}$**
S	AGC	1	0.08	0.37
G	GGC	15	0.45	**$8x10^{-14}$**
P	CCC	2	0.02	0.21
V	GTC	1	0	0.37
A	GCC,GCT	4,1	0.38	0.02, 0.37
L	CTC	1	0	0.37
N	AAC	1	0	0.37
Z	TAG	1	0.02	0.37

A383	GCG	0	0.60	N/A[a]

P384	CCG	2	0.46	0.009
N	AAC	1	0.26	0.14
R	AGG	1	0.26	0.14
C	TGC	1	0.3	0.14

V409	GTT	0	0.65	N/A[a]

H26	CAC	4	0.92	$2x10^{-5}$
A	GCG	1	0.23	0.15
K	AAG	1	0.92	0.15
T	ACA,ACC	1,1	0.18	0.15, 0.15
Y	TAC	1	0.92	0.15
V	GTC	1	0.92	0.15
Z	TGA	1	0.1	0.15

H100	CAT	2	0.43909	0.003
V	GTT	1	0.18	0.15
A	GCC	1	0.23	0.15
I	ATC	1	0.7	0.15
M	ATG	1	0.86	0.15

I189	ATC,ATA	3,2	0.65	0.0004, 0.01
L	CTT,TTG	1,1	0.38	0.14, 0.14
A	GCA	1	0.01	0.14
M	ATG	1	0.39	0.14
Y	TAC	1	0.19	0.14

H261	CAC	0	0.49	
T	ACA	1	0.12	[a]N/A

R358	CGT	4	0.88	$1x10^{-5}$
I	ATC	1	0.91	0.14
Q	CAA	2	0.79	0.01
P	CCA,CCT	1,1	0.62	0.14, 0.14
G	GGG	1	0.59	0.14

S385	TCT	0	0.88	0.95
L	CTT	1	0.038	0.05
P	CCA	1	0.87	0.05
W	TGG	1	0.88	0.05

H461	CAC	3	0.09	0.002
Q	CAA	1	0.98	0.22
S	AGT	6	0.09	**$2x10^{-7}$**

Y	TAC	1	0.1	0.22
P	CCG	2	0.08	0.03
W	TGG	4	0.2	0.0001
G	GGT	1	0.14	0.22

D469	GAC	4	0.23	$1x10^{-5}$
Y	TAT	1	0.12	0.14
S	TCA	1	0.06	0.14
A	GCA	1	0.03	0.14
L	TTA	1	0.39	0.14
T	ACT	1	0.15	0.14
Z	TAG	1	0.1	0.14

R520	CGT	0	0.57	0.74
S	AGT	3	0.52	0.003
I	ATC,ATT	1,3	0.64	0.22, 0.003
L	CTA	1	0.71	0.22
G	GGC	2	0.8	0.03
C	TGT	1	0.11	0.22
V	GTC	1	0.8	0.22
P	CCG,CCT	3,1	0.56	0.003, 0.22
Z	TAA	1	0.82	0.22
Y	TAT,TAC	1,1	0.38	0.22, 0.22

Table 4.2 Sequence data from randomly picked colonies within each library (excluding known spiked wild-type colonies), their average conversions after one hour of reaction obtained from high-throughput screening, and the probability of observing each codon at frequency F. Cases of bias are highlighted in bold. [a]Insufficient DNA sequences obtained.

The probability data in Table 4.2 indicate a slight bias towards one non-wild-type DNA sequence for K23X, A29X and H461X, whereas D259 shows considerable bias towards D259G sequences.

The natural sequence variations observed at each site in the 52 predominantly bacterial and yeast TK sequences analysed, are also shown for comparison in Tables 4.3 & 4.4.

Residue Number	Percentage active mutants	Natural Amino Acid Distribution (Percentage Frequency)
23	97%	N 56%, K 18%, G 15%, Q 5%, S 3%, , H 1.5%, R 1.5%
29	29%	A 26.5%, M 25%, L 23.5%, S 14%, T 5%, V 3%, G 1.5%, I 1.5%
64	7%	A 53.5%, N 30%, K 6.5%, V 5%, G 3.3%, H 1.7%
159	23%	M 66.7%, Q 16.6%, S 10%, F3.3%, A 1.7%, T 1.7%
188	93%	S 65%, T 10%, Q 8.3%, G 6.6%, C 5%, N 1.7%, I 1.7%, V 1.7%
259	64%	K 25%, A 18.7%, D 16.3%, G 13.3%, S 10%, E 8.3%, L 3.3%, F 1.7%, R 1.7%, V 1.7%
383	94%	A 43.3%, T 26.7%, S 15%, K 8.3%, G 5% R 1.7%
384	96%	G 30%, P 26.7%, S 13.%, N 6.7%, R 3.7%, H 5%, A 3.3%, C 3.3%, T 3.3%, L 3.3%, I 1.7%
409	95%	V 65.%, I 31.7%, L 3.3%
466	N/A	L 53.7%, V 40.7%, Q 3.7%, H 1.8%

Table 4.3 Phylogenetic library activity screens and natural amino acid distribution in the initial phylogenetic sequence alignment.

Residue Number	Percentage active mutants	Natural Amino Acid Distribution (Percentage Frequency)
26	99%	100% H
100	58%	95% H, 5% F
189	82%	63.3% I, 26.7% L, 6.7% Q, 1.7% C, 1.7% M
261	99%	95% H, 5% Y
262	97%	91.6% G, 5% V, 1.7% A, 1.7% W
358	86%	93.3% R, 3.3% P, 1.7% A, 1.7% T
385	27%	93.3% S, 5% G, 1.7% P
461	91%	95% H, 5% Q
469	100%	98.3% D, 1% N
520	96%	95% R, 3.3% N, 1.7% K

Table 4.4 Structural library activity screens and natural amino acid distribution in the initial phylogenetic sequence alignment.

4.3.5 Analysis of B-factors, solvent surface accessible surface area (SAS) and sequence entropy (H(x))

The B-factors obtained from the *E.coli* TK crystal structure file (1QGD.pdb) are summarised in Table 4.5, as an average for each library residue. The values assigned to each atom are also represented by colour for each library residue shown in the active-site structures of Figure 4.31.

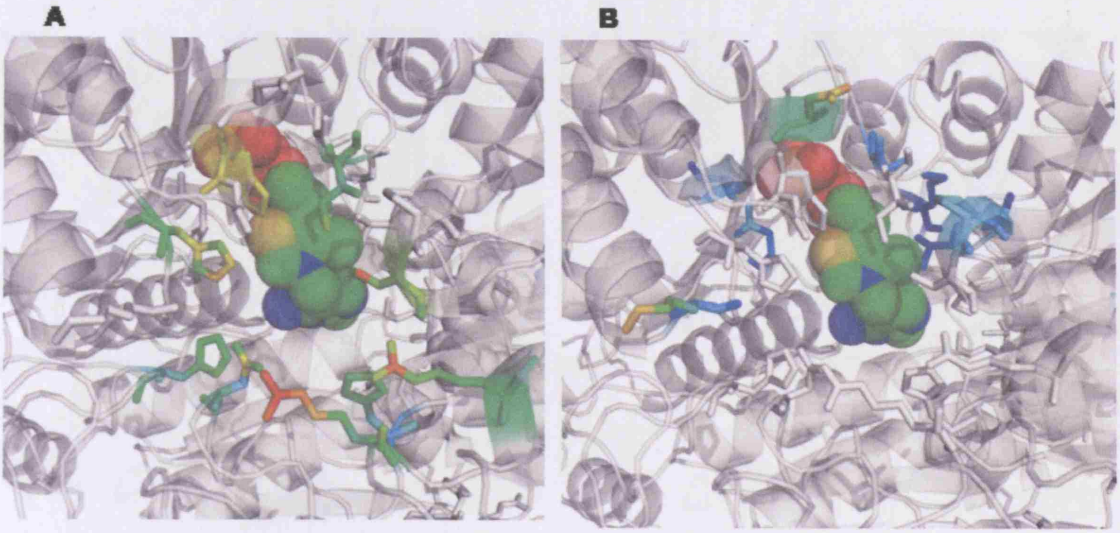


Figure 4.32 A&B Residues coloured by B-factor unit values. **A** structurally defined library residues highlighted. **B** phylogenetically defined library residues highlighted.

(1 BFU = 79Å^2)

The solvent accessible surface areas (SAS) obtained from the *E.coli* TK crystal structure file (1QGD.pdb) are summarised in Table 4.5, as an average for the side-chain of each library residue. The SAS values assigned to each residue are also represented by colour in the active-site structures of Figure 4.32.

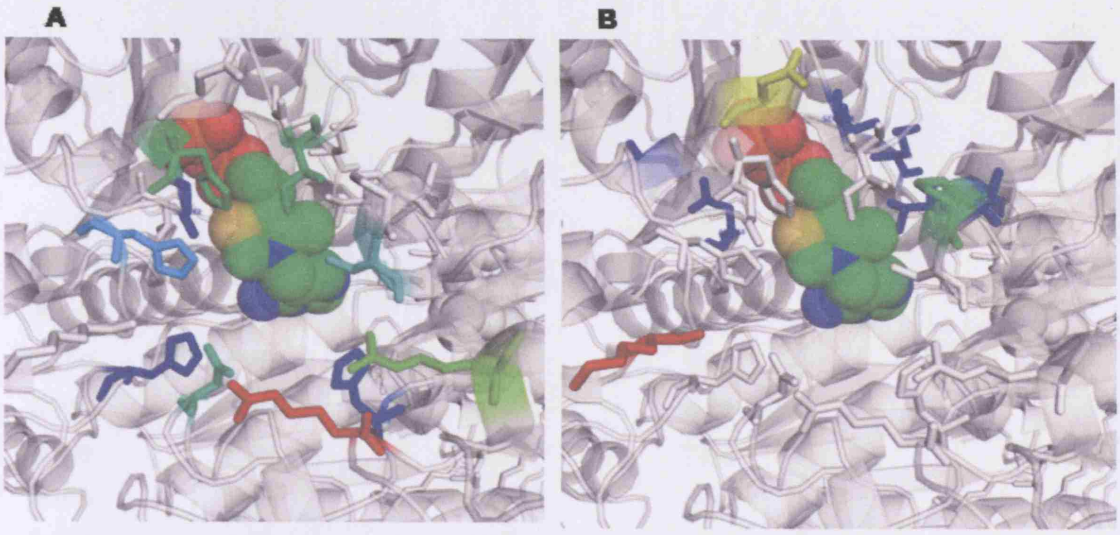


Figure 4.33 A&B Residues coloured by solvent accessible surface (SAS) for the side-chain according to the key below. **A** structurally defined library residues highlighted. **B** phylogenetically defined library residues highlighted.

The solvent accessible surface (SAS) areas observed for the library residues are on average less than 30 $Å^2$, indicating that the majority are not very solvent exposed. Five of the ten structurally defined residues have at least 30 $Å^2$ exposed to solvent, with R520 being particularly solvent exposed (97 $Å^2$) and H26, H100, G262 and H461 having less than 21 $Å^2$ solvent exposure (dark blue). The phylogenetic residues tend to be more varied and extreme with the highly exposed K23 (130 $Å^2$), D259 (74 $Å^2$) and P384 (46 $Å^2$), and the highly buried M159 (0 $Å^2$), A383 (3 $Å^2$), N64 (5 $Å^2$), S188 (5 $Å^2$) and A29 (6 $Å^2$). The phylogenetic residues are more clustered in terms of exposure, with few residues having intermediate solvent exposure (light blue to green). This is likely due to their second shell placement.

Residue	Sequence entropy H(x)	B-factor units (1 BFU = 79Å^2)	SAS (Å^2)
K23	1.32	38.7	129.3
A29	1.71	14.8	5.98
N64	1.21	15.2	4.68
M159	1.05	9.1	0.00
S188	1.25	14.4	4.91
D259	1.98	38.9	74.0
A383	1.42	15.0	3.34
P384	1.97	19.9	48.9
V409	0.76	9.9	8.55
H26	0	18.3	21.1
H100	0.20	13.7	5.73
I189	0.96	15.0	34.3
H261	0.20	21.9	35.9
G262	0.37	22.7	2.96
S385	0.28	18.3	28.8
H461	0.20	11.4	12.5
D469	0.085	14.8	31.2
R520	0.23	20.1	97.1

Table 4.5 Summary of sequence entropy H(x), B-factor units (from 1QGD.pdb), and solvent accessible surface (SAS) data for residues in each library.

A plot of the B-factors versus solvent accessible surface areas for the library residues, is shown in Figure 4.34. A reasonable correlation (R^2 = 0.68) is obtained for the B-factor averaged over the side-chains, versus the solvent accessibility of the side-chains. A plot of natural sequence entropy, H(x), for each library residue, versus average B-factors and solvent accessible surface area (SAS) for their side-chains, is shown in Figure 4.35. A poor correlation is obtained in both cases.

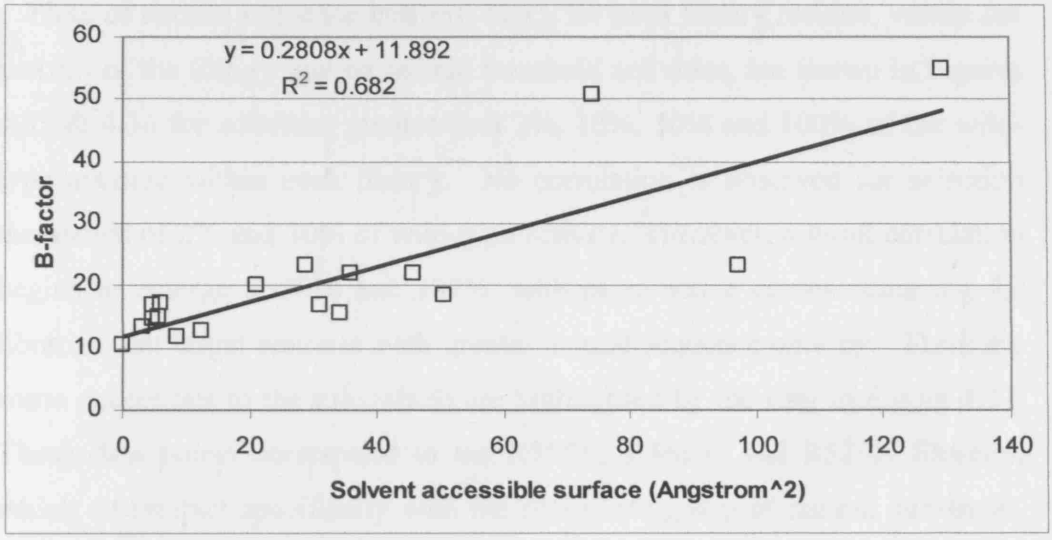


Figure 4.34 Correlation of B-factor unit averages (for total residue and side-chain only, 1 BFU = 79Å^2) with solvent accessible surface area (SAS) for the side-chains of each library residue.

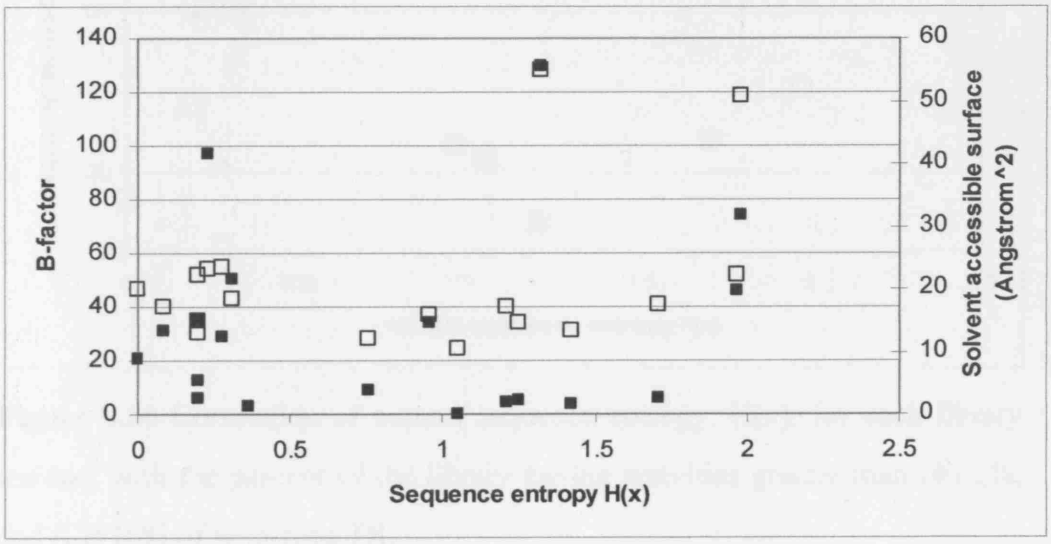


Figure 4.35 Correlation of natural sequence entropy, H(x), for each library residue, with B-factor units (1 BFU = 79Å^2) and solvent accessible surface area (SAS) for their side-chains.

Plots of natural sequence entropy, H(x), for each library residue, versus the percent of the library having certain threshold activities, are shown in Figures 4.35 & 4.36 for activities greater than 2%, 10%, 50% and 100% of the wild-type average within each library. No correlation is observed for selection thresholds of 2% and 10% of wild-type activity. However, a weak correlation begins to emerge at 50% and 100%, with more active clones occurring for libraries that target residues with greater natural sequence entropy. There are some exceptions to the rule which are highlighted by the ring in Figure 4.37. These data points correspond to the R358X, H461X, and R520X libraries, which all interact specifically with the phosphate group of natural substrates, and so these residues are no longer being selected for with the substrates used.

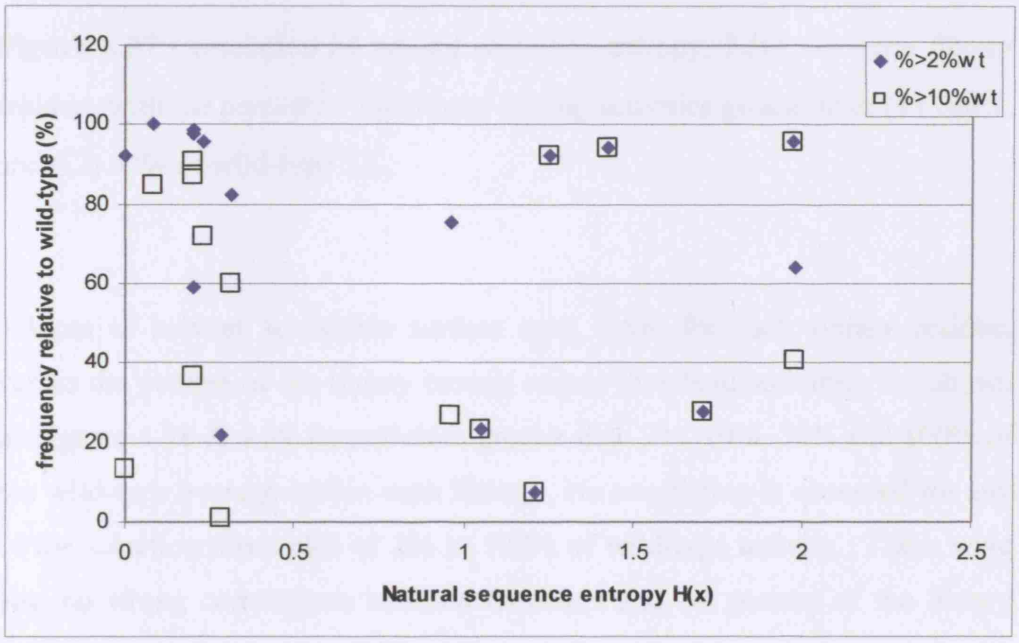


Figure 4.36 Correlation of natural sequence entropy, H(x), for each library residue, with the percent of the library having activities greater than (♦) 2%, and (□) 10% of wild-type TK.

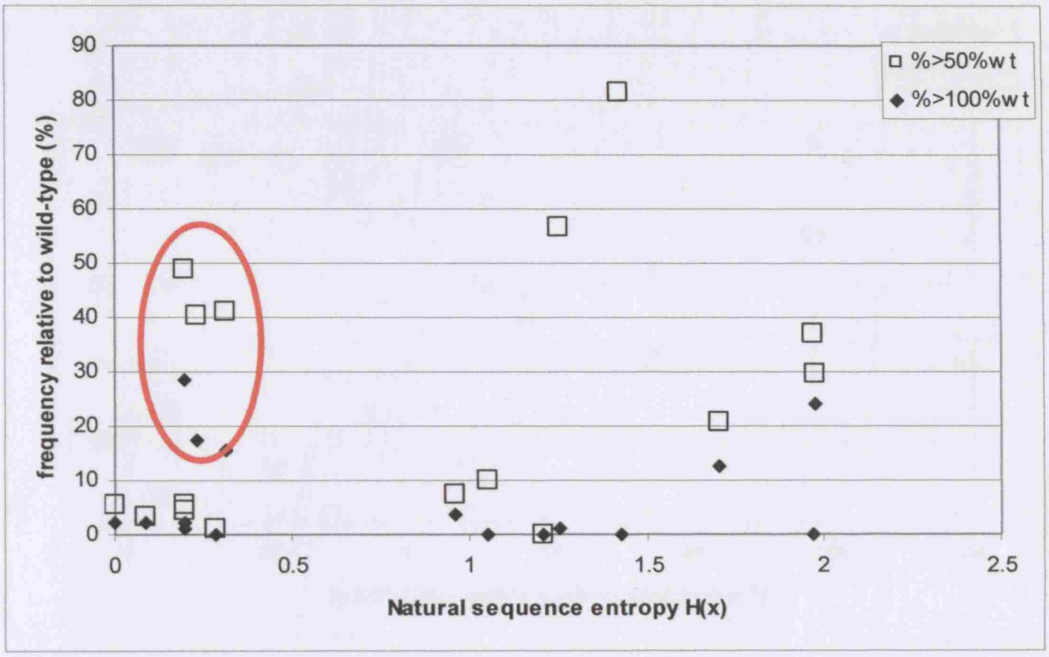


Figure 4.37 Correlation of natural sequence entropy, H(x), for each library residue, with the percent of the library having activities greater than (♦) 100%, and (□) 50% of wild-type TK.

Plots of solvent accessible surface area, SAS, for each library residue, versus the percent of the library having certain threshold activities, are shown in Figures 4.38 & 4.39 for activities greater than 2%, 10%, 50% and 100% of the wild-type average within each library. No correlation is observed for any of the selection thresholds of 2% to 100% of wild-type activity. There were also no strong correlations between B-factors and the percent of the library having certain threshold activities (data not shown).

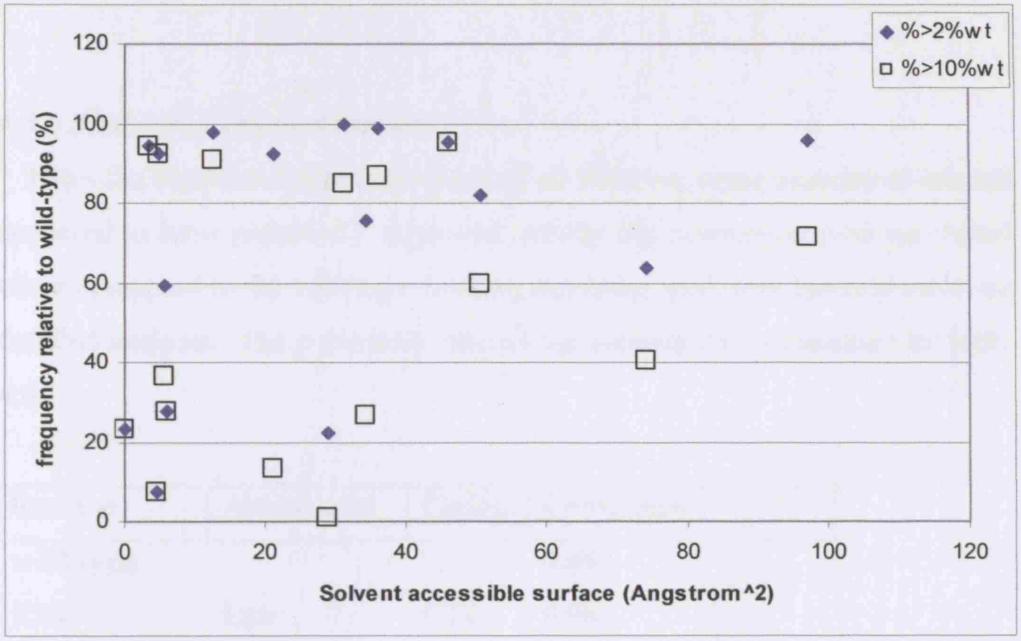


Figure 4.38 Correlation of solvent accessible surface area, for each library residue, with the percent of the library having activities greater than (♦) 2%, and (□) 10% of wild-type TK.

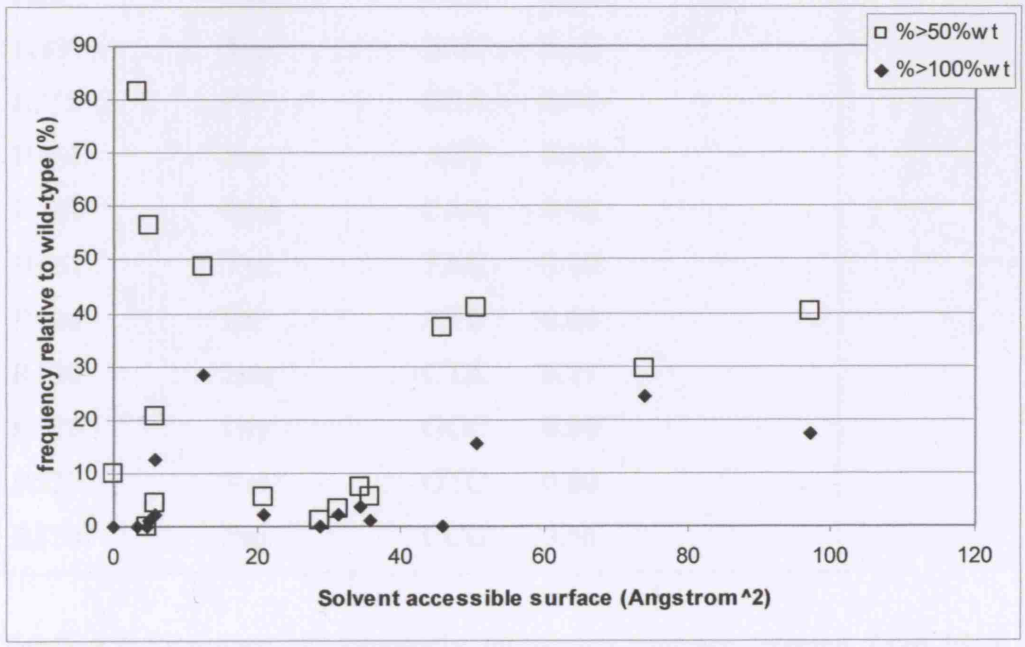


Figure 4.39 Correlation of solvent accessible surface area, for each library residue, with the percent of the library having activities greater than (♦) 100%, and (□) 50% of wild-type TK.

4.3.6 Analysis of selected mutants

From the high-throughput screening of all libraries, some mutants of interest appeared to have potentially improved activity (as conversion over one hour) when compared to the wild-type internal standards, and were isolated for more detailed analysis. The potentially interesting mutants are summarised in Table 4.6.

Residue	Amino acid	Codon	Conversion
wild type			**0.49**
K23	Leu	CTG	0.98
A29	Glu	GAG	0.48
A29	Asp	GAC	0.41
A29	Gln	CAG	0.47
M159	Trp	TGG	0.35
H26	Lys	AAG	0.92
H26	Tyr	TAC	0.92
I189	Tyr	TAC	0.19
R358	Pro	CCA	0.94
H461	Ser	AGT	0.10
H461	Gln	CAA	0.98
H461	Tyr	TAC	0.10
R520	Ile	ATC	0.64
R520	Leu	CTA	0.71
R520	Gly	GGC	0.80
R520	Val	GTC	0.80
R520	Pro	CCG	0.56

Table 4.6 Summary of potentially interesting mutants selected from high-throughput screening of structurally and phylogenetically defined libraries. Note these are from 1 plate screen and as a result nostandard deviation (SD) is available.

4.3.7 Characterisation of interesting mutants

A number of the mutants summarised in Table 4.6 were subjected to further analysis by repeating reactions and analysing the conversions to erythrulose at higher resolution, and also analysing the concentration of transketolase present so that specific activities could be compared.

4.3.7.1 Activity analysis on model reaction

The slower HPLC analysis method, described in Section 2.3.16.4, uses a longer column with a longer elution time and lower flow rate compared to the high-throughput screen (section 2.3.16.6). This allows for greater resolution of the compounds under investigation. Examples of the time courses obtained for conversion of HPA and GA to erythrulose, analysed at high resolution, by various mutants and wild-type transketolase, are shown in Figure 4.40.

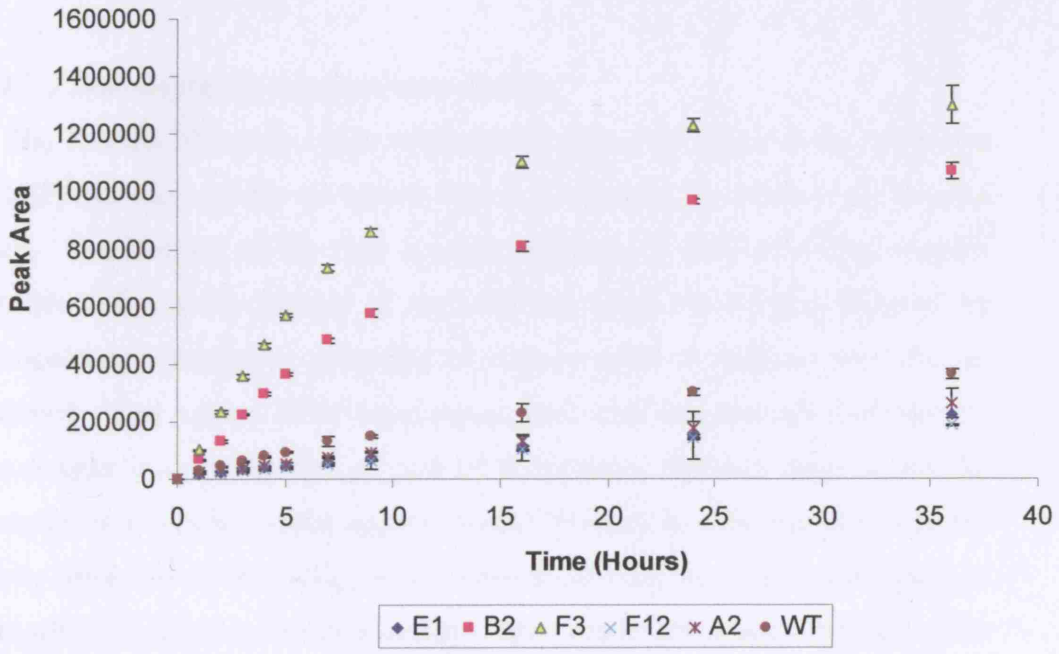


Figure 4.40 High resolution HPLC analysis of the reaction profile for HPA and GA catalysed by selected mutants and wild-type transketolase, over 40

hours using standard cofactor concentrations, 50 mM substrates and 50 mM Tris.HCl buffer at pH7.0.

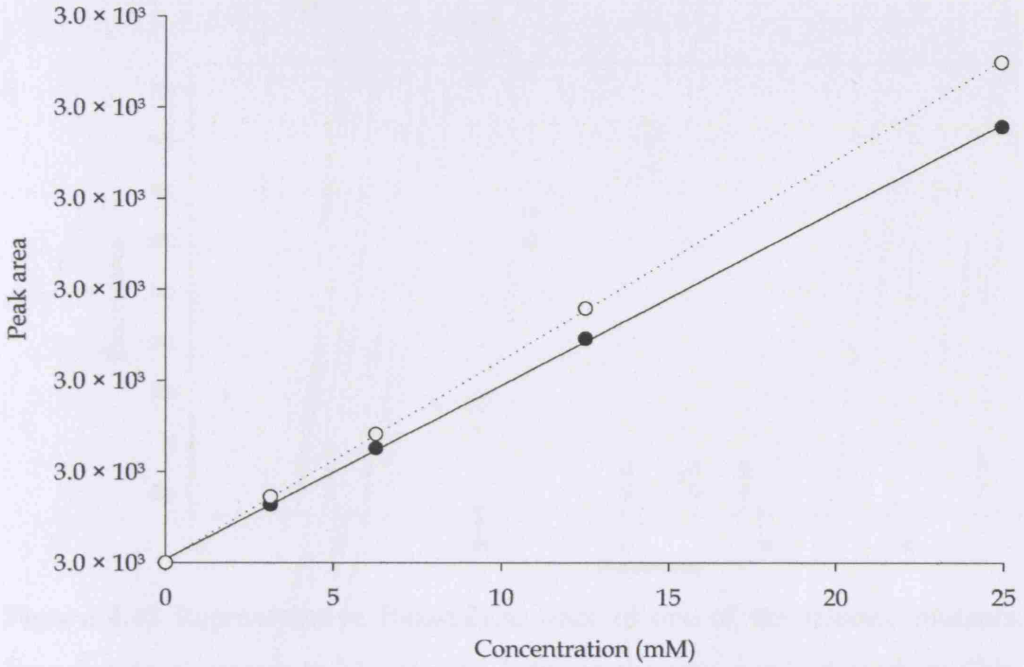


Figure 4.41 HPLC calibration curves for L-erythrulose and β-HPA. The relationships between concentration and detector response are approximately linear for L-erythrulose (●) and β-HPA (○). Above 25mM the relationships become more curved.

4.3.7.2 Bioanalysis for enzyme concentration

The Protein 200 assay chips were primed as per standard method (Section 2.3.22) and each sample of interest loaded in triplicate or more on the 10 lane chip. The loading of the chip is very important as there is a tiny channel which resides at the bottom of the well and must avoid being blocked by particulates (filtration or vortexing of samples prior to loading) and also air bubbles. The Agilent 2100 bioanalyser, then analyses through fluorescence the amount in concentration of each of the proteins and their sizes within the samples on each chip within approximately 30 minutes. However the chips are easily fouled by poor loading or particulates entering the chip channels and as a result it is necessary to often perform many replicates as some erratic results were seen. The data collected can then be used in conjunction with a standard curve to ascertain the protein concentration of the samples analysed. An

example can be seen below of a Bioanalysis trace (Figures 4.42) and a standard curve (Figure 4.43) from commercial TK (Fluka Ltd).

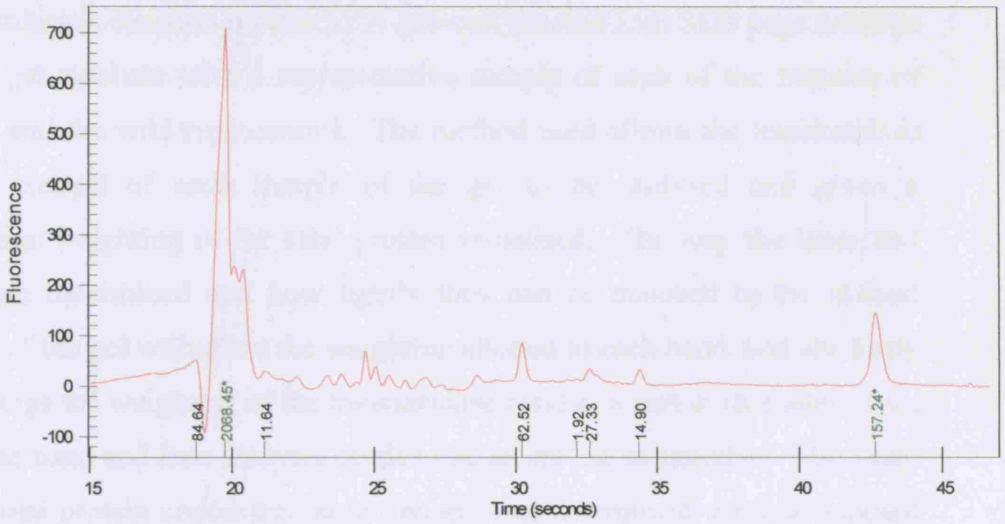


Figure 4.42 Representative Bioanalysis trace of one of the selected mutants. Transketolase appears at 31 seconds, between the two standard markers (20s, 43s).

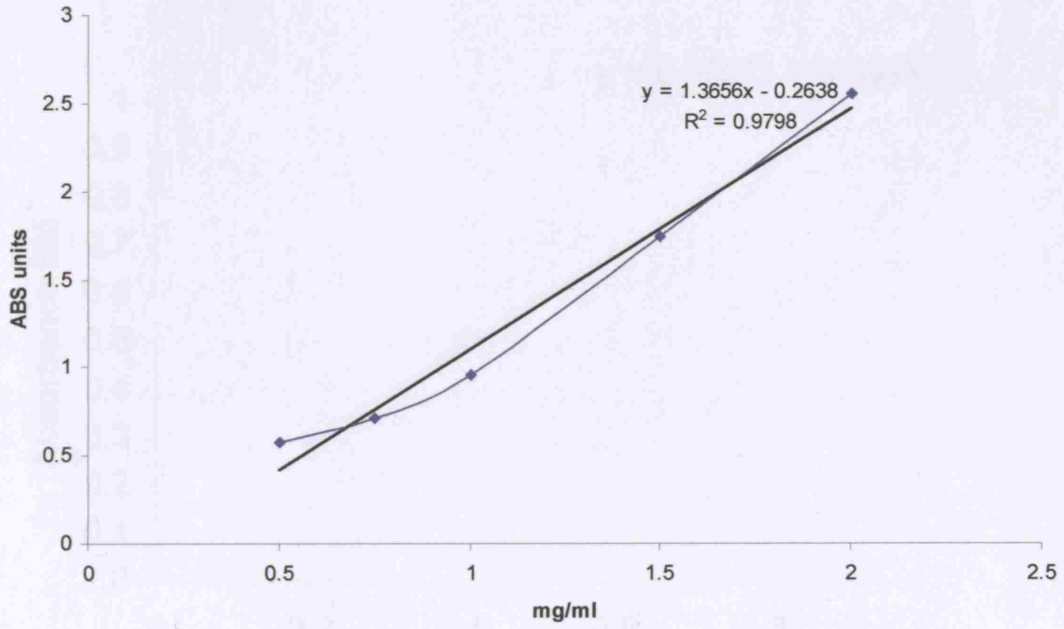


Figure 4.43 Standard protein concentration (mg/ml) curve obtained by Bioanalysis with commercially available TK (Fluka).

4.3.7.3 Densitometry for enzyme concentration

Densitometry was also performed to determine the concentration of transketolase in reaction mixtures. A pre-cast (Biorad Ltd) SDS page (section 2.3.21) gel was run with a representative sample of each of the mutants of interest and the wild-type control. The method used allows the transketolase protein content of each sample of the gel to be analysed and given a percentage weighting of the total protein visualised. The way the lanes and bands are determined and how tightly they can be matched to the stained regions of the gel will affect the weighting allotted to each band, and are likely also change the weighting of the transketolase band as a part of that total. As a result the band and lane analysis needs to be as precise as possible. The value of the total protein concentration on the gel was determined using a standard curve for bovine serum albumin (BSA), also determined by a Bradford assay as shown in Figure 4.44.

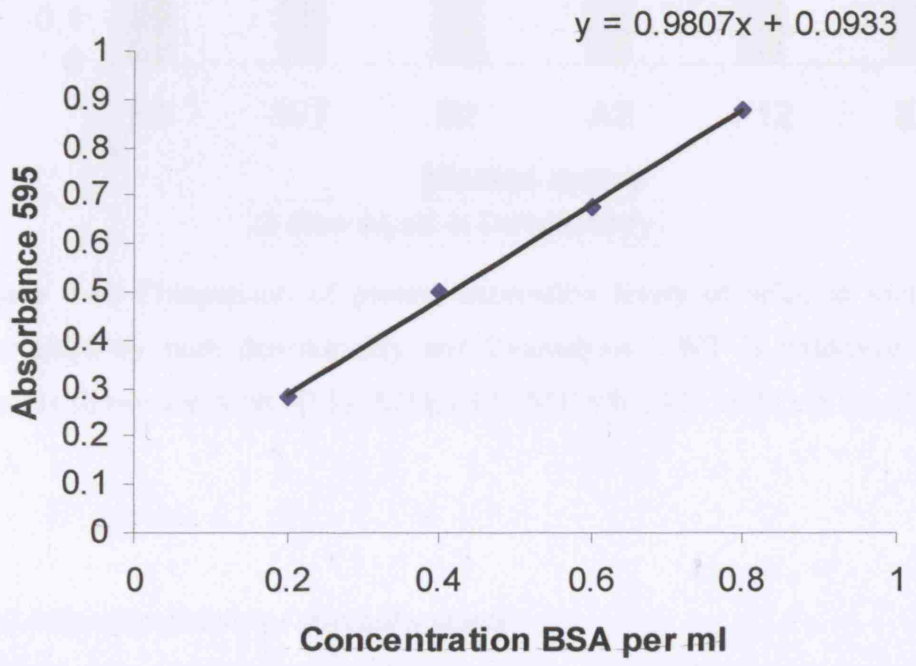


Figure 4.44 Standard curve absorbance BSA/LB solutions 0.2-08mg/ml A595

Biorad protein assay

4.3.7.4 Comparison of bioanalysis and densitometry

The two methods were used to analyse the expression of transketolase in reaction samples for each of the library mutants of interest. The data obtained from the two methods is compared in Figure 4.45. It can be seen that the two methods give protein concentrations that are in broad agreement. However, densitometry has the greater accuracy.

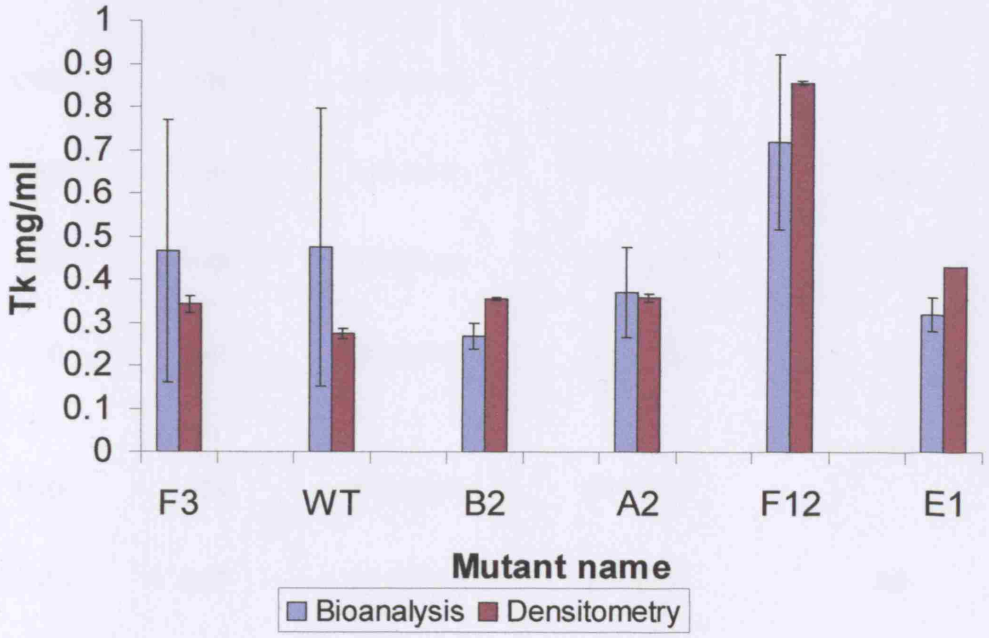


Figure 4.45 Comparison of protein expression levels of selected mutants determined by both densitometry and Bioanalysis. WT is wild-type TK. Mutants shown are A29D (F3), A29E (B2), M159W (A2), A29Q (F12), K23L (E1).

4.3.7.5 Specific activity of selected mutants

The specific activity of selected mutants was calculated from the conversion at one hour determined by HPLC, and the enzyme concentrations determined by both densitometry and bioanalysis. Table 4.7 summarises the results. Two

mutants, A29D and A29E show significant increases in the conversion of GA and HPA to erythrulose per hour, per mg of TK. Three mutants show significant increases in the expression level of TK within the cell, namely I189Y, R358P, and H461S. Such changes may or may not result directly from the changes to the targeted codon, or may be due to mutations elsewhere in the plasmid.

Sample name	Codon	Conversion by first screen	Conversion at high res (SD)	[TK] (mg/ml) (SD)	[E]/TK (M/mg/hour)
Wild-type	-	0.49	0.33 (0.01)	0.47 (0.3)	13.9
E1 (K23L)	CTG	0.98	0.20 (0.08)	0.32 (0.04)	14.5
B2 (A29E)	GAC	0.41	0.83 (0.02)	0.27 (0.03)	**61.5**
F3 (A29D)	GAG	0.48	0.93 (0.05)	0.47 (0.3)	**42.5**
F12 (A29Q)	CAG	0.47	0.19 (0.006)	0.72 (0.2)	5.7
A2 (M159W)	TGG	0.35	0.25 (0.05)	0.37 (0.1)	13.6
F4 (I189Y)	TAC	0.19	0.15 (0.03)	1.91 (0.3)	8.5
D5 (R358P)	CCA	0.94	0.67 (0.08)	6.81 (3.1)	3.0
F12 (H461S)	AGT	0.10	0.78 (0.07)	1.85 (0.3)	12.5

Table 4.7 Specific activity as erythrulose formed [E] per mg TK per hour, for selected mutants, including DNA sequencing, and reaction conversions for both initial high-throughput and high resolution screening. Expression levels of TK are determined from Bioanalysis. SD: standard deviation in brackets.

4.4 Discussion

4.4.1 Library construction

The initial library methods attempted were fairly unsuccessful as problems occurred with ligation methodology due to the need for high numbers of variants to produce each library condition (EpPCR conditions see section 4.2.2 Figure 4.1). However the efficiency of the ligations were never very high, probably due to large vector to insert size ratio the vectors tested were both approximately 3 kb and the transketolase gene insert is approximately 2.5 kb. The similarity of the sizes of vector to insert as well as possible low template yields from the EpPCR reactions and/or low vector concentration (gel extraction yields were often poor) resulted in very poor transformation efficiency, even with the use of electroporation. Attempts to use mutator strain XL1-Red were similarly unsuccessful with problems propagating the strains probably due to accumulated deleterious mutations in the hosts chromosomal DNA. Despite frequent minipreps and retransformations into fresh mutator strain host cells, the plasmid frequently seemed to be 'lost' or ejected from cell, despite these problems cells were taken to 60 and 100 generations and DNA sequenced (section 2.4.1) and found to have very low if any mutation rates, similar results were also observed by others using the strain (C .Ingram unpublished data).

The switch in strategy to the use of saturation site-directed mutagenesis (SSDM) avoided the problems associated with poor ligation efficiency or mutator strain host instability. In fact the optimisation of the QuickChange protocols and primer design were comparatively easy. The greatest factor found to impact on the yield of QuickChange products was found to be the plasmid DNA template concentration (Figure 4.7). Other factors such as primer concentration were also varied, but were found to have minimal effect.

4.4.2 Initial screening for erythrulose production

To be suitable for directed evolution experiments, a screening system should ideally be able to link the observed activity or phenotype, directly to the genotype. This can be easily achieved here as microwell plates containing libraries are replicated. After lysis of the cells and analysis of conversions obtained with the target substrates, the clones of interest identified from the screen can subsequently be further analysed in more detail kinetically, and also by DNA sequencing. As the cells in each well arise from a unique clone (assuming no mixed culture or contamination occurs), there is a biunivocal correlation between the enzyme activity and the DNA subsequently recovered. The initial screen by HPLC allowed the rapid analysis of a large number of mutants in a reasonable time. However, there is likely to be some trade off between the accuracy of the initial screening results and the speed with which they are attained. The fast HPLC method used gives a general indication of the viability of libraries as well as which mutants might be worthy of a further, more detailed, HPLC analysis. As can be seen in Figure 4.16, two plates generated independently from the same master stock and screened on different days using the same fast HPLC method show a number of significant differences, nevertheless overlayed by the majority of clones displaying reproducible conversions to erythrulose after one hour. The majority of the mutants run on this comparison seem to cluster in the same region as the internal wild-type standards. However, there are some differences such as the majority of the better performing mutants screened on plate 1 becoming more wild-type like on the second plate screen. Also two of the better performing mutants on plate 1 (showing 0.8 conversion) were in fact inactive on plate 2. These differences were possibly due to bacterial contamination or plasmid instability resulting in cells not expressing TK. It is also possible, though unlikely, that the slight differences seen are the results of mixed colonies being initially picked by the Qpix into some wells of the microplates. This would not always have happened as some libraries had different densities of cells on plates prior to picking.

4.4.3 Analysis of sequence diversity in libraries

It may be argued that any bias towards the fewest number of point mutations at the genetic level, occurring in a library, in some ways reflects the bias to mutability that occurs in nature to avoid disruptive changes to protein function. For example, the genetic codon table does not reflect the number of residue types evenly and single base changes tend to yield an amino acid with properties similar to the original. For my studies it was essential to analyse the extent of bias within the libraries made so that I could be confident that most amino-acids are represented in the number of colonies analysed, and also so that broader conclusions about tolerance to mutation could be drawn from the activity profiles of the libraries.

Bias may result in some codons not being represented in libraries, and these will not all be redundant mutations so some mutants of interest may be missed depending on the bias of each library position and number of base changes needed to effect that residue change. The bias inherent in each library will also have an effect on how tolerance to mutation is represented by the number of active mutants observed, as it will affect the frequency of occurrence of each amino acid in different ways, giving an uneven profile.

An analysis of the frequency distribution of codons observed in each library, obtained by DNA sequencing of random samples, is shown in Table 4.2. When compared to the probabilities, assuming an unbiased library, of obtaining a given codon at each frequency from the sample size studied, the DNA sequence data obtained generally show a low level of bias. Exceptions include a slight bias towards a single non-wild-type DNA sequence for K23X, A29X and H461X. However, for D259, a considerable bias towards D259G sequences occurs. The probability of observing D259G fifteen times in a sample size of thirty-six sequences is 8×10^{-14}. Caution must therefore be observed when drawing broad conclusions from the distribution of activity in the D259 library. In all other libraries where sufficient DNA sequences were analysed there does not appear to be significant bias towards single base substitutions, as two and three substitutions frequently occur.

4.4.4 Activity profiles obtained from high-throughput screening

The average conversions (where 0 is no reaction, and 1 is complete reaction), obtained relative to internal wild-type averages, are shown in Figures 4.29 & 4.30 for libraries respectively defined phylogenetically and structurally. The most noticeable effect in the structurally defined libraries is that H461X, R358X and R520X have significantly higher conversions averaged for the whole library. This reflects their higher tolerance to mutation than other structurally defined libraries. As will be discussed later, the conservation at the structurally defined libraries is generally at or close to 100% within bacterial and yeast TK sequences. The tolerance to mutation for all of these libraries was therefore expected to be low. However, H461X, R358X and R520X are all implicated in binding specificity for phosphorylated substrates (Nilsson *et al.*, 1997). Therefore, as the activities in this study have all been measured using non-phosphorylated substrates, the increased tolerance to mutation at these residues may result from removal of selection pressure to maintain phosphate binding function.

It is clear from comparison of Figures 4.30 and 4.31 that many residues in the phylogenetically defined libraries appear to be more tolerant to mutation than for the structurally defined libraries. The exceptions are positions A29 and N64, for which there is no immediate explanation using structural information alone.

4.4.5 Sequence dependent correlations (B-factors and solvent accessible surface

"B-factors" or "temperature factors" are among the worst determined parameters in protein structures. In principle, they describe the flexibility of the atom positions in the structure. In effect, however, the B-factors are normally underdetermined, and hence absorb all kinds of inaccuracies in the X-ray structure refinement, as well as being potentially influenced by protein-protein interactions within the crystal. The average B-factor for all atoms that

have an accessibility value of 0 is usually calculated. This average is normally between 20 and 40. Lower values can be obtained for frozen crystals, whereas high values are a possible indication of refinement problems. Since the B-factor is largely a parameter describing the motion of the protein atoms, and proteins are rather fixed entities that tend to rotate and translate as a whole, one expects the B-factors of atoms buried deep in the core to be lower than the ones at the surface of the protein. A less strictly adhered rule: one expects the B-factors of atoms that are close in space to be similar. Even less strict: one expects the B-factors of bonded atoms to be similar.

It is possible that the tolerance to mutation of a residue might in part relate to the B-factor at that residue, as B-factors relate to the flexibility of the residue which makes it difficult to resolve by x-ray diffraction. The average B-factors for the side-chains of the wild-type residues in the TK structure (1QGD.pdb) were analysed at each library position. Figure 4.32 shows the B-factors represented on a colour key as they are located in the TK structure, for both library types. B-factors are also summarised in Table 4.5. It is expected that B-factors would be greatest for residues that are more solvent exposed and for this reason the solvent accessible surface (SAS) for the same residues was estimated as summarised in Figure 4.33 and Table 4.5. As expected, the B-factors and SAS values correlate well as shown in Figure 4.34.

As can be seen in Figures 4.38 & 4.39, the solvent accessible surface does not correlate well with the observed levels of activity for each library. A similar result (data not shown) was observed for B-factors. This does not mean that neither measure can play a role in predicting sites that are tolerant to mutagenesis, but it does mean that neither measure is a strong factor on its own for determining highly tolerant sites.

4.4.6 Comparison of natural and evolved sequence diversity

The sequence entropy, i.e. degree of variation that occurs at each residue within a set of aligned natural sequences is expected to relate to the tolerance to mutation at that position, given a certain degree of selection pressure. When selection pressure in terms of substrate binding, catalysis or protein stability for example, is removed, it is expected that the observed sequence entropy will increase at residues important to that function. In a similar manner, the libraries produced at each residue position would potentially vary according to the selection pressure maintained and the importance of that residue in retaining the selected function.

Values of the sequence entropy for bacterial and yeast sequences were obtained using a sequence alignment in the Bioedit software suite and are summarised in Table 4.5. It can be seen in Figure 4.34 that sequence entropy does not correlate well to B-factors, or by inference of Figure 4.34 to solvent accessible surface. Plots of natural sequence entropy, H(x), for each library residue, versus the percent of the library having certain threshold activities, are shown in Figures 4.35 & 4.36 for activities greater than 2%, 10%, 50% and 100% of the wild-type average within each library. While no correlation is observed for screening thresholds of 2% and 10% of wild-type activity, a correlation begins to emerge at 50% and 100% levels of wild-type activity. In these cases, more active clones occur for libraries that target residues with greater natural sequence entropy. There are some exceptions to the rule which are highlighted by the ring in Figure 4.37. These data points correspond to the R358X, H461X, and R520X libraries, which all interact specifically with the phosphate group of natural substrates, and so these residues are no longer being selected for with the non-phosphorylated substrates used. These residues provide additional evidence that the application of stringent selection pressure leads to reduced tolerance to mutation in both natural sequences and directed evolution libraries. This observation has a direct impact on the decision whether or not to include highly conserved residues, such as those in the structurally defined library, in the choice of sites to target for directed evolution. It would seem that when attempting to evolve activities on substrates that differ from the natural ones, such as non-phosphorylated substrates for TK, then residues in contact with the altered regions of substrate

are still potentially a good target for mutagenesis. Overall though it is still not clear whether highly conserved sites may be hiding useful amino-acid variants that only become apparent when evolving an activity *in vitro*, i.e removing a selection pressure that only occurs in the cell.

The correlation of high natural sequence entropy with the greater tolerance to mutation is stronger for a selection threshold of 50% of wild-type activity than it is for 100%. This observation demonstrates that a high tolerance to mutation is not a guarantee for finding mutants with greater than wild-type activity in regions of high natural sequence entropy. However, focussing mutations to regions of high sequence entropy is more likely to produce beneficial mutations.

What is perhaps remarkable, is that when the screening threshold of activity is reduced to 10% of wild-type activity, nearly all sites, including those that are almost completely conserved in transketolases (H26, H100, H261, H461, D469), retain a high number of clones meeting the screening criterion (Figures 4.28 & 4.29). This implies that for transketolase at least, the drop in activity to 10% is sufficient to make the survival of the cell improbable. However, it also implies that the active-site of TK is remarkably tolerant to mutagenesis in real terms, where a 10% loss of activity is not significant relative to the overall level of activity in comparison to the uncatalysed reaction. Therefore, a reduced selection threshold in directed evolution experiments may be beneficial at the initial stages, for obtaining greater sequence diversity with reasonable activity, before recombining them to look for synergistic effects.

Another interesting observation is the difference between variants at a given position that are observed in Nature, versus those that appear in the libraries at low (10%) selection thresholds. For example, H26 is 100% conserved in bacterial and yeast sequences (Table 4.4), and yet can be mutated to Ala, Lys, Thr, Tyr and Val while retaining at least 10% activity. By contrast, A29 has a high degree of natural sequence variation (Table 4.3), including Ala, Met, Leu, Ser, Thr, Val, Gly and Ile, and yet the active sequences found in the A29X library with at least 10% activity are Ala, Asp, Glu, Gln, and Thr.

Furthermore, Ala29Asp and Ala29Glu mutations were found to have 4.4 and 3-fold greater specific activities than wild-type TK (Table 4.7), while Ala29Gln retains 41% activity. A similar effect is observed for K23L, I189Y, H461S and I189Y mutants which have wild-type-like levels of specific activity, and yet are not observed in the natural sequence variants. These surprising results may be in part the effect of utilising non-phosphorylated substrates in the directed evolution screen, or by using an *in vitro* selection process. However, it is more likely that the natural sequence variations are not necessarily beneficial as single mutants of the *E. coli* TK. The natural variants occur in the context of many other groups of simultaneous mutations, some of which may be necessary to observe wild-type activity.

These observations also demonstrate that creating mutants that contain only natural sequence variations is extremely limiting in terms of potential for enhanced activity. Using that approach, the mutants with greatest activity for erythrulose production (A29D and A29E) would not have been identified. This conclusion should perhaps be borne in mind when performing the artificial shuffling technique in which only natural variants are recombined during library construction (Ness *et al.*, 2002).

Rationalising any of the mutants observed is very difficult without structural analysis. A29 is in a very interesting location. It may have tweaked activity via the TPP without affecting substrate specificity. This is because A29 is in a location where its methyl sidechain is in direct contact (3.4 Å) with the terminal phosphate of the TPP. Such an interaction is not favourable and the switch to Asp or Glu could enable protonation of the TPP phosphate. Alternatively the new Asp and Glu mutants could form a hydrogen bond with the nearby N64 or H66 sidechains, where the latter residues already make a stabilising contact with the TPP phosphate. I189 and H461 are residues which are able to interact directly with the substrate and mutation of these residues may have helped improve binding and/or reaction rate by altering the orientation or binding of substrates. It is possible that H461S has switched from providing a positively charged histidine which normally interacts which

with the substrate, and the serine hydroxyl-group it is replaced with could potentially provide a better hydrogen-bonding interaction with the terminal hydroxyl groups of both substrates and the erythrulose product. R520 is also a residue made redundant by use of non-phosphorylated substrates. This residue normally extends into the active-site funnel and may to some extent hinder the access of substrates to their binding sites. Mutation to praline might be expected to remove much of the side-chain bulk and, therefore, create greater acess to the active site thus increasing the rate of reaction. Furthermore this residue occurs at the apex of a hairpin loop which protrudes into the active site. The restriced conformation of a proline residue could be easily accommodated into this loop without disrupting the overall hairpin structure.

Overall, the use of sequence entropy to identify sites for directed evolution is only partially useful as potential sites such as H461 may have been missed. However, the use of a combined approach as described here, involving residues with structural proximity, and then 'second shell' residues with high sequence entropy, may provide a more efficient route to choosing residues for targeted saturation mutagenesis.

4.5 Conclusions

Problems with initial library development methods led to the conclusion that mutator strains were unreliable for good production of libraries, this is because there is little or no control of mutation rate and where they will occur. The problems seen with loss of cell viability and plasmid quality give the impression that most mutation occurring is in undesired sites such as disruption of important cellular DNA and or also plasmid DNA e.g. ampicillin resistance gene or possibly origin replication for plasmid resulting in lack of ability to extract plasmid DNA. The other methods of library development were unsuccessful also I believe due to the lack of a defined mutation strategy. Random mutagenesis for such a large protein is likely to result in the need for

very large libraries but will also likely cause many inactivated or incurring numerous redundant mutations. The result of this is that library quality is affected and as a result the number of clones required to be picked to achieve satisfactory sequence coverage can rapidly become unrealistic. The other problem that occurred was the need to achieve good efficiency transformations with the use of a relatively large DNA fragment for ligation protocols. Most of the desirable commercial vectors are of similar size to the TK gene and ligations with like size fragments are not conducive to high efficiency. Clones were obtainable, but with such low efficiency as to become unrealistic as an effective library strategy. In my opinion, if ligation and random mutagenesis is carried out it is preferable to do it with smaller DNA fragments such as possibly a domain of interest in TK or by removal of, for example, the redundant C-terminal domain prior to modification (activity maybe observed in absence of C-terminus (S.Costelloe, unpublished)). Saturation site-directed mutagenesis is a powerful tool for directed evolution if used correctly and has many advantages in terms of reducing library size and reduced occurrence of redundant mutations if good target sites are identified by either phylogenetic or structural analysis or combination of both. The construction of just mutants which have been seen in nature (natural variation libraries) would however miss many of the mutants of interest produced and as a result I firmly believe that the production of saturation mutagenesis libraries with the hope of finding variants with a complete array of amino acids is vital to finding new and or improved activity (in comparison to artificial shuffling used by Ness *et al.*, 2002). There are however potential drawbacks to this method of mutagenesis in that multiple variables of the same codon can be produced resulting in the same final sequence of protein, but potentially causing other problems. It is possible that if a low usage codon is inserted into a library clone during QuickChange strategies, then the expressed level of TK may be limited by the low availability of the correct transfer RNA (tRNA). It is possible the clone produced could have improved activity, but this may not be detected in initial screens due to low expression levels of TK. This would not be so much of a problem in the detailed mutant analysis as expression level is analysed and a measure of activity can be assigned to each mutant via the amount of

erythrulose produced relative to the TK expression. However in the rapid screen analysis it is possible some of the 'poorer' performing mutants may have had expression level changes and as a result may have been overlooked. Library construction bias may also be a problem, as in principle any bias represents a liability to the quality of the libraries produced. The random variation and library bias that sometimes occurs, for example in the D259X library, calls into question the accuracy of the subsequent calculations for library coverage numbers. The levels of library coverage that were assumed to allow coverage for all variants (90 mutants picked per library) is also, therefore, only approximate. In future applications of SSDM through QuickChange protocols, the accuracy could be improved by sequencing more library members and sorting them before screening.

The results in Chapter 4 suggested that both structurally defined and phylogenetically defined sites are useful for the engineering of activity towards non-natural substrates. Furthermore, it was shown that both library types mostly yielded residues with a high tolerance to mutation in terms of enzyme inactivation, which is perhaps surprising for the fully conserved residues of the structurally defined library. Perhaps less surprising was the observation that residues no longer under selection pressure, such as the phosphate-binding R358 and R520, showed the greatest tolerance and range of acceptable mutations.

4.6 Future and Ongoing work

Work which is ongoing with these libraries are attempts to construct some double mutants from those of most interest in both libraries to investigate the impacts these may have as double mutants have been claimed to have much more impact on changes in activity and/or specificity than single ones. Future work, had time allowed, would have been to try to counteract the codon bias in the libraries to try and give a more even spread of the amino acids produced and as a result improve the quality of the libraries and potentially increase the mutants of interest found. If time allowed I would also investigate further the

difficulties in producing the 'missing' 466 library, possibly by attempting to produce a new PCR protocol or primer to get around the problem encountered initially.

Chapter 5 - Library screening with novel aldehydes

5.1 Introduction

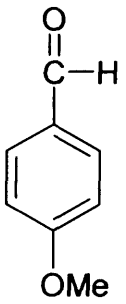
5.1.1 Prediction of sites: Tolerance (plasticity)

As discussed in Chapter 4, predictive methods for guiding the location of targeted mutagenesis, other than structural analysis, would be extremely valuable. The results in Chapter 4 suggested that both structurally defined and phylogenetically defined sites are useful for the engineering of activity towards natural substrates. Furthermore, it was shown that both library types mostly yielded residues with a high tolerance to mutation in terms of enzyme inactivation, which is perhaps surprising for the fully conserved residues of the structurally defined library. Perhaps less surprising was the observation that residues no longer under selection pressure, such as the phosphate-binding R358, H461 and R520, showed the greatest tolerance and range of acceptable mutations. As previously discussed in Chapter 4, analyses of directed evolution experiments have shown that mutated sites resulting in improvements, are also frequently those that are tolerant to mutation as determined by computational modelling of the protein thermostability (Voigt et al., 2001a;Voigt et al., 2001b). The aim of Chapter 5 was to extend the study beyond natural substrates and on towards non-natural substrates to determine whether there is a correlation between activity improvements and a preference for mutations occurring at residues from a particular library type or a correlation to the degree of tolerance to mutation as measured by the natural substrates.

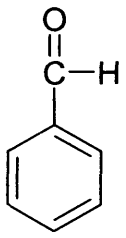
5.1.2 Substrate walking

The concept of substrate walking could be viewed as an application of promiscuous enzyme activity. The idea is to direct the specificity of an enzyme towards a novel substrate that is too different to be accepted even after

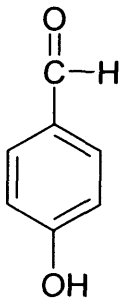
the first round of directed evolution. Therefore, a substrate that forms a structural stepping stone would be chosen that can exploit the existing enzyme promiscuity, and the enzyme engineered for improved activity towards this substrate. Directed evolution could be based on a study of evolutionarily related enzymes to identify residues which, when modified, might improve the desired activity. The final substrate specificity would then be achieved by engineering further the enzyme variant that accepts the so-called stepping stone substrate. The potential range of the substrates that might be introduced using such a method in conjunction with directed evolution needs further investigation. The aim of this Chapter is to begin to investigate the range of novel specificity that can be achieved in a single round of directed evolution, arising from the focused mutations based on phylogenetic and structural selection in chapter 4, and to utilise promiscuity for the investigation of substrate walking using the enzyme transketolase (TK). The substrates shown in Figure 5.1 had been selected for further testing based on those successful assay development results in Chapter 3.



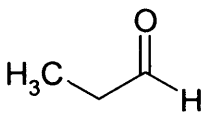
4-methoxybenzaldehyde



benzaldehyde



hydroxybenzaldehyde



propionaldehyde

Figure 5.1 Aldehyde acceptor substrates chosen for further study in Chapter 3

The areas that I aimed to explore (in Chapter 5) for extending substrate specificity beyond those currently accepted are: a) the scope of change achievable through one round of directed evolution library-type strategies (novel activities); and b) the degree of change to the substrate that can be achieved using a single round of directed evolution. The latter knowledge could lead to the use of intermediate substrates as stepping stones towards a final target substrate in a process that is termed from this point onwards as 'substrate walking'. Therefore, this Chapter explores the potential library production techniques which can be employed to produce a mutant library of TK for screening with the assays developed in Chapter 3 and any potential changes or improvements in the model reaction in this first round of directed evolution seen in Chapter 4. Assessment of the quality of the two libraries, and their initial screening have been performed in Chapter 4, and so it was then appropriate to begin to screen them with the more promising aldehyde substrates defined in Chapter 3. The first new substrate to be tested was propionaldehyde as this already exhibits a wild-type activity of ~5% relative to that of glycolaldehyde (GA). The most desirable assay for the screening of libraries with these aldehyde substrates was a colorimetric method being developed in collaboration with Mark Smith and Helen Hailes in the Department of Chemistry at UCL. However this was not sufficiently optimised for high-throughput screening use at the time when the libraries were developed, and as a result a rapid thin-layer chromatography (TLC) method was used in place of the HPLC screening method.

5.2 Materials and methods

5.2.1 Keto-diol product standards

All keto-diol product standards were obtained from syntheses by Mark Smith (Department of Chemistry, UCL) as described in section 2.3.16.8

5.2.2 TK reactions in microplates

Reactions between TK variants in cell lysates from the libraries, between HPA and the appropriate aldehyde acceptor substrates were carried out in microtitre plates for 48 hours as described in Section 2.3.15.2. The aromatic aldehyde substrates were used at 10 mM rather than the usual 50 mM for GA and PAL, due to their lower solubility.

5.2.3 Thin-layer chromatography (TLC) screening

The TLC method for analysing the libraries was the same as described previously (section 2.3.17) with up to six different mutants being analysed on a single TLC plate, alongside a wild-type TK control reaction (plate wells A6, E6 or H12) and a keto-diol product standard as appropriate. This system allowed a whole plate to be screened in roughly 2-3 hours after preparing the plate meaning it was possible to sometimes screen two plates a day in comparison to the rapid HPLC analysis which takes approximately 4 hours per plate.

Propionaldehyde, benzaldehyde, hydroxybenzaldehyde and methoxybenzaldehyde were all tested for activity against the libraries described in Chapter 4, using the TLC methods. All aldehyde substrates were tested at 10 mM final concentration, except propionaldehyde which was tested at 50 mM. The reduced concentration was necessary to overcome problems with solubility. Each library site was tested systematically with each of the substrates prior to moving on to a new site for testing.

5.2.4 HPLC assays

Individual mutants of interest identified from library microwell fermentation plates, based on TLC screening analysis, were characterised in more detail using the HPLC procedure described in (section 2.3.16.8). HPLC chromatograms were collected for all of the screening reactions following up to sixty hours of incubation at 25 °C.

5.2.5 Determination of protein concentration by bioanalysis

Samples were analysed using standard Protein 200 plus assay instructions. After priming and preparation of an Agilent chip (section 2.3.22), 6 µl of the protein samples of interest were loaded in triplicates or more onto the 10-lane chip. The Agilent 2100 bioanalyser then uses capillary electrophoresis to separate the proteins, and analyses the quantity of a protein-binding fluorescent dye present in each peak, all within 30 minutes. The data was then used in conjunction with a standard curve obtained with commercial TK samples, to ascertain the TK protein concentration of the samples analysed.

5.3 Results

5.3.1 Propionaldehyde reaction screening by thin-layer chromatography (TLC)

A typical TLC screening plate for reactions between propionaldehyde (PAL) and hydroxypyruvate (HPA), catalysed by wild-type transketolase (TK) and various mutants, is shown in Figure 5.2. The product formed (1,3-dihydroxypentan-2-one), which I will name for simplicity as propionaldehyde keto-diol (PKD), appears as a distinct band as indicated.

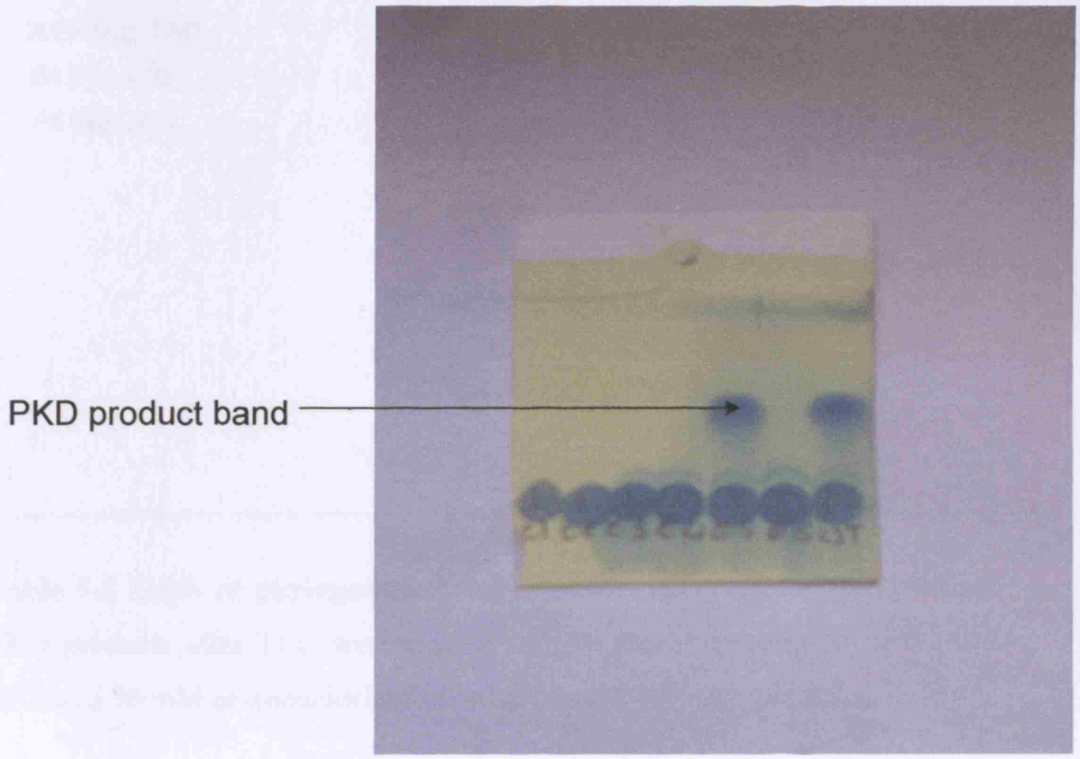


Figure 5.2 Example TLC plate after development, for the NNS159 library wells E1-E6. Wild-type TK is shown in the lane furthest to the right.

All libraries were initially screened in this way for the TK catalysed reaction of HPA and PAL. A summary of mutants from the phylogenetically determined library that were found to accept PAL at least as well as the wild-type TK is shown in Table 5.1.

Library (codon type / target residue)					
NNS A29	**NNS M159**	**NNS A188**	**NNS D259**		**NNS P384**
(Well position / amino acid / codon)					
B2 Asp GAC	E5 Tyr TGG	G6 Gln CAG	A2 Ser AGC	D7 Ala GCC	G7 n/d
D1 Glu GAG	A2 Tyr TGG		A3 Gly GGC	D10 Tyr TAC	
F3 Glu GAG	A8 Ser AGC		A10 Tyr TAC	D12 Gly GGC	
E12 Glu GAG	A9 Gln CAG		B2 Tyr TAC	E5 Gly GGC	
	A11 Stop TAG		B5 Gly GGC	E11 Gly GGC	
	F4 Thr ACG		B7 Tyr TAC	E12 Gly GGC	
	F5 Glu CAG		B11 Gly GGC	F8 Gly GGC	
			B12 Gly GGC	F11 Tyr TAC	
			C2 Tyr TAC	F12 Gly GGC	
			C6 Ala GCC	G7 Gly GGC	
			C12 Gly GGC	G11 Gly GGC	
			D4 Ala GCC	G12 Gly GGC	
			D5 n/d	H10 Ala GCC	
			D6 Gly GGC		

Table 5.1 Table of phylogenetic library mutants observed to have possible PKD products after TLC screening of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0.

A summary of mutants of the structurally determined library, found to accept PAL at least as well as wild-type TK is shown in Table 5.2.

Library (codon type / target residue)						
NNN H26	NNN H100	NNN I189	NNN R358	NNN H461	NNN D469	NNN R520
(Well position / codon / amino acid)						
B11 TAC Tyr	D6 GTT Val	B3 TTG leu	A4 ATC Ile	A1 CAA Gln	B6 TCA Ser	B1n/d
B12 GCG Ala	D12 n/d	D4 ATG Met	D9 n/d	F11 TAC Tyr	D6 n/d	C9 ATC Ile
D12 n/d	F6 GCC Ala	E4 ATA Ile	F9 n/d	F12 AGT Ser	H3 n/d	D11TGT Cys
E12 AAG Lys	F12 ATC Ile	F4 TAC Tyr		H12 AGT Ser	F11 ACT Thr	F7 TAA Stp
G5 ACA Thr		F10 n/d			F3 GAC Asp	F9 GGC Gly
G6 TGA Stp		H3 ATC Ile				F12 GTC Val
H7 GTA Val						G11 CCG Pro

Table 5.2 Table of structural library mutants observed to have possible PKD products after TLC screening of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0.

5.3.2 Aromatic aldehyde reaction screening by thin-layer chromatography (TLC)

No detectable activity was observed with any of the libraries when screened against 4-methoxybenzaldehyde or 4-hydroxybenzaldehyde, with HPA as the ketol donor, even when reactions were left for in excess of 60 hours in agitated microplates with TLC testing at regular intervals every 10-20 hours reaction time. Of the aromatic aldehydes selected for study, only benzaldehyde was observed to have potential product (BKD) formation bands by TLC after more than 60 hours agitation and reaction time. A typical TLC screening plate for reactions between benzaldehyde (BAL) and hydroxypyruvate (HPA), catalysed by various mutants, is shown in Figure 5.3. The product formed (1,3-dihydroxy-1-phenyl-propan-2-one), which I will call the benzaldehyde ketodiol (BKD) product for simplicity, appears as a distinct band as shown. The mutants summarised in Table 5.3 were observed to have possible/weak product banding after reaction with benzaldehyde.

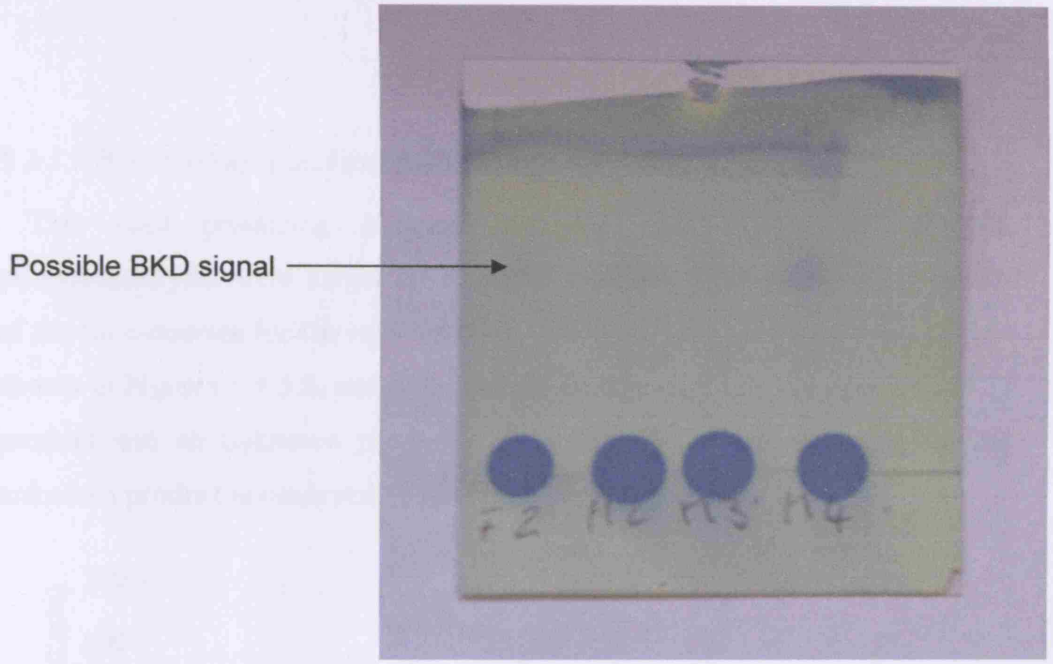


Figure 5.3 Selected NNS259 library mutants showing a potential product (BKD) signal after 48hours of the TK catalysed reaction between 50 mM HPA and 10 mM benzaldehyde (BAL), 50 mM Gly-Gly, pH 7.0. No BKD band is observed using wild-type TK under the same conditions.

Library (codon type / target residue)	
NNS188	NNS259
(Well position / amino acid / codon)	
D4 Thr ACC	A1 Tyr TAC
B8 n/d	A2 Ser AGC
	C3 Pro CCC
	C4 Val GTC
	F2 Stop TAG
	G3 Leu CTC
	G8 Ala GCC
	G10 Asn AAC
	H2 Asp GAC
	H3 Pro CCC

Table 5.3 Table of mutants observed to have possible BKD products after TLC screening of the TK catalysed reaction of 50 mM HPA and 10 mM benzaldehyde (BAL), 50 mM Gly-Gly, pH 7.0.

5.3.3 HPLC testing of mutants with promising activity towards PAL

The most promising mutants observed by TLC to be utilising propionaldehyde, were subjected to HPLC analysis. Examples of the results of the time-courses for the reactions with D259 mutants, and wild-type TK are shown in Figures 5.4-5.8, and highlight the presence of both the expected PKD product and an unknown product. A typical HPLC trace, for which the unknown product is observed, is shown in Figure 5.9.

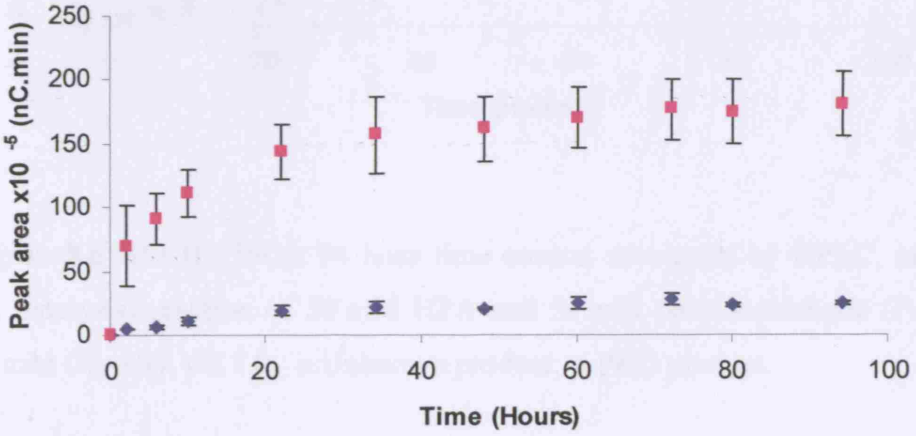


Figure 5.4 A2 (D259S): 94 hour time course, monitored by HPLC, of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0. ■ Unknown product; ◆ PKD product.

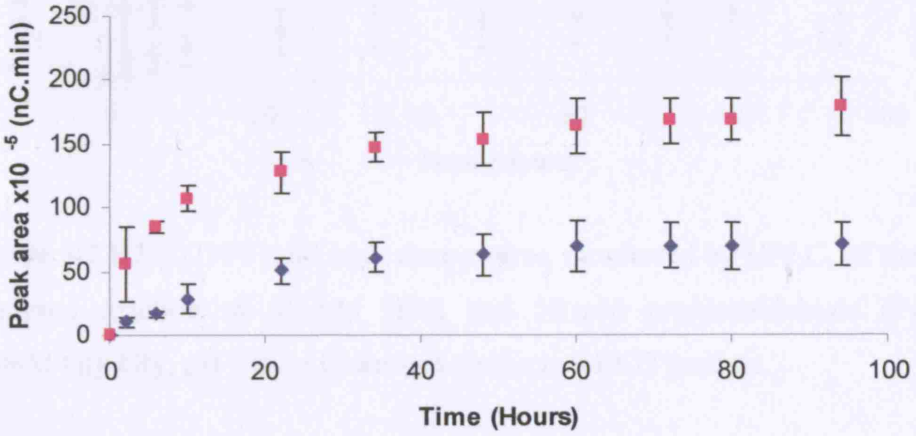


Figure 5.5 D7 (D259A): 94 hour time course, monitored by HPLC, of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0. ■ Unknown product; ◆ PKD product.

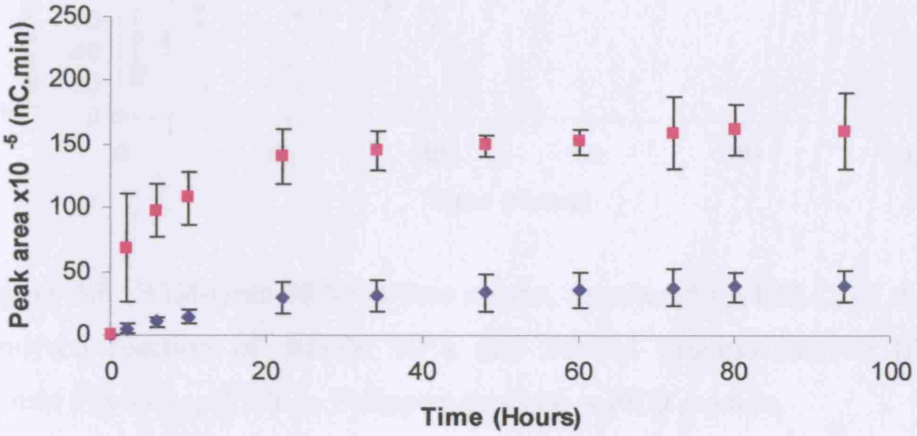


Figure 5.6 B12 (D259G): 94 hour time course, monitored by HPLC, of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0. ■ Unknown product; ◆ PKD product.

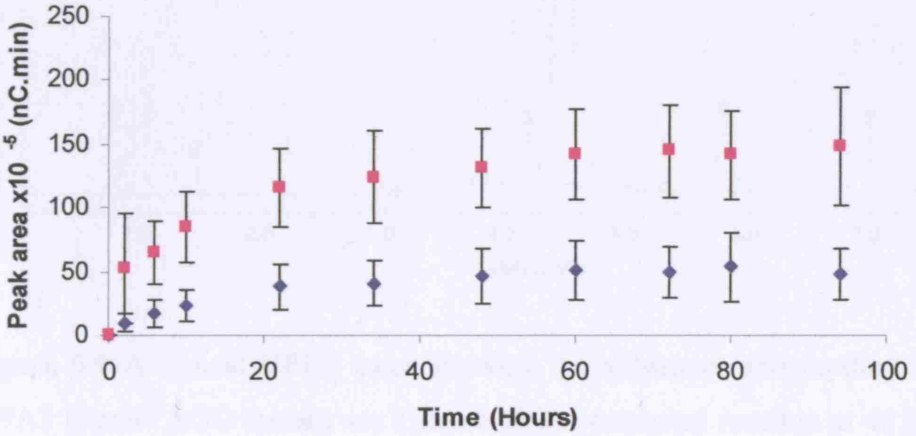


Figure 5.7 B7 (D259Y): 94 hour time course, monitored by HPLC, of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0. ■ Unknown product; ◆ PKD product.

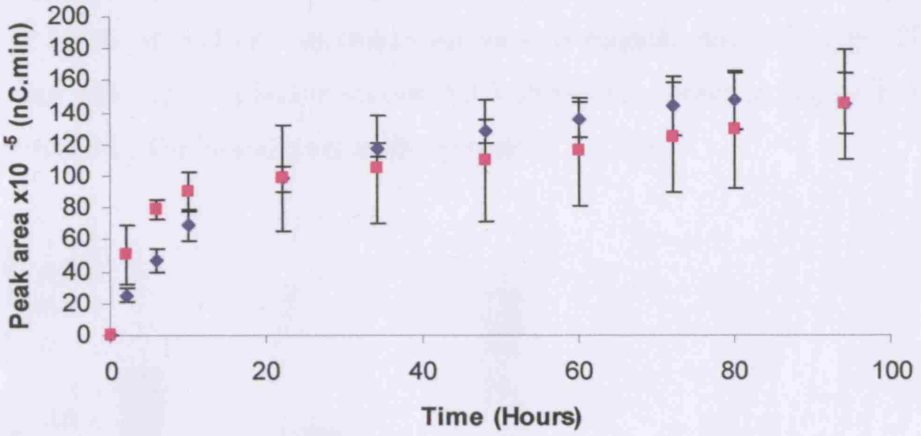


Figure 5.8 Wild-type: 94 hour time course, monitored by HPLC, of the TK catalysed reaction of 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0. ■ Unknown product; ◆ PKD product.

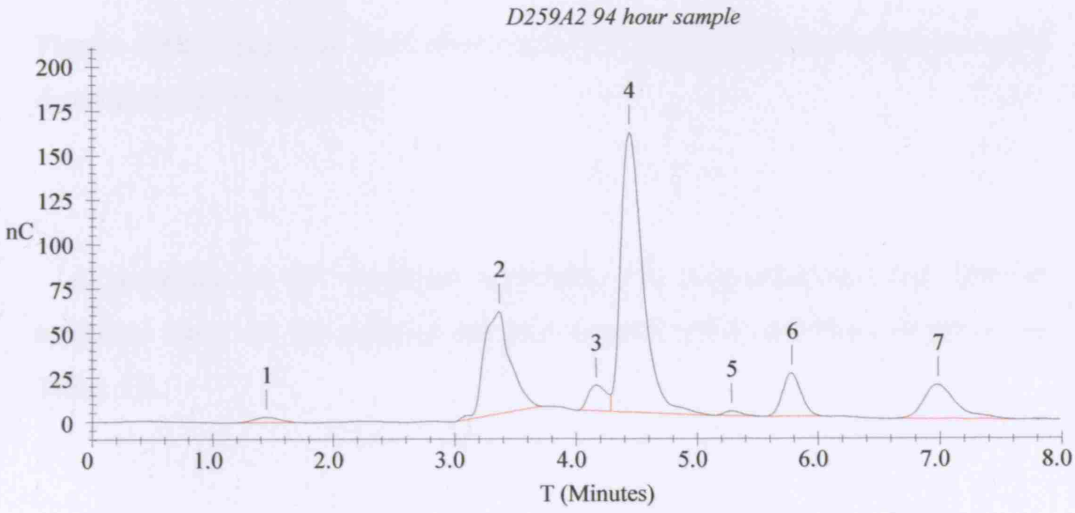


Figure 5.9 A typical HPLC trace showing the substrates and products of a 259A2 (Serine AGC mutant see Table 5.1) TK catalysed reaction at 48 hours between 50 mM HPA and 50 mM propionaldehyde (PAL), 50 mM Gly-Gly, pH 7.0. Peak 1 is due to a step change in the concentration of ACN in the mobile phase. Peak 2 is propionaldehyde (PAL), Peak 4 is HPA. Peak 6 is the PKD product, whereas Peak 7 is an unknown product.

5.3.4 Expression level analysis of selected HPA-PAL mutants by bioanalysis

The levels of protein expression for various mutants and wild-type TK in the reactions characterised in section 5.3.3 above, are shown in Figure 5.10, as determined by the bioanalysis method (section 2.3.22).

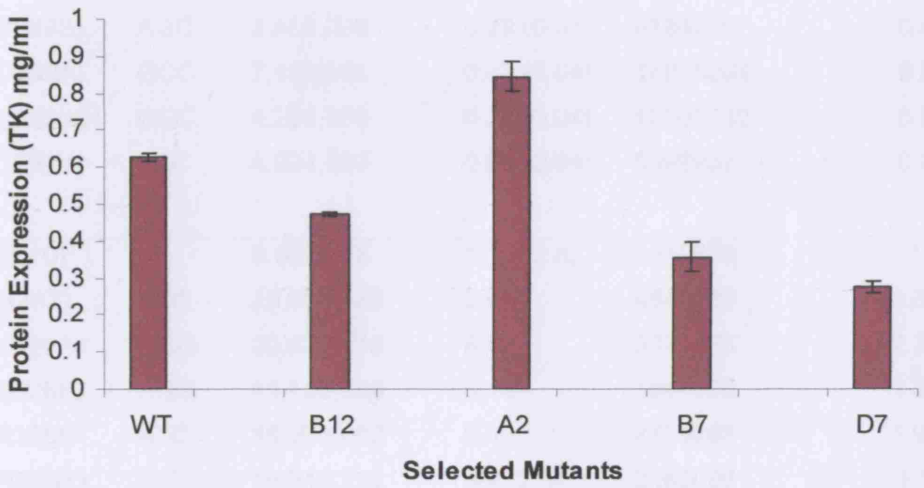


Figure 5.10 Expression level of selected propionaldehyde accepting mutants, determined by bioanalysis.

A summary of the observed activities, TK concentrations and specific activities observed for selected mutants towards HPA and PAL is shown in Table 5.4.

Sample	Codon	PKD peak area (AU.min)	[TK] (mg/ml)	Yield PKD/TK (AU.min/mg)	Yield/TK relative to wt
WT (Top10)	-	14,626,010	0.63 (0.01)[a]	23266373	1
A2 (D259S)	AGC	2,454,703	0.28 (0.01)	8834220	0.4
D7 (D259A)	GCC	7,108,658	0.47 (0.04)	14976284	0.6
B12 (D259G)	GGC	4,296,768	0.36 (0.04)	12005230	0.5
B7 (D259Y)	TAC	4,824,293	0.85 (0.04)	5689937	0.2
WT (XL10)	-	8,869,278	6.12 (2.8)	1449696	1
G5 (H26T)	ACA	29,663,436	6.12 [b]	4848529	3.3 [c]
B12 (H26A)	GCG	20,623,715	6.12 [b]	3370975	2.3 [c]
E12 (H26K)	AGG	11,100,089	6.12 [b]	1814325	1.3 [c]
A4 (R358I)	ATC	16,973,032	6.12 [b]	2774265	1.9 [c]
D9 (R358X)	N/D	19,210,732	8.9 (2.3)	2163099	1.5
H12 (H461S)	AGT	5,457,127	2.0 (0.4)	2721375	1.9
WT (XL10)	-	14,462,565	9.9 (1.1)	1468020	1
F11 (D469T)	ACT	25,062,568	3.5 (0.3)	7200963	5.0
B6 (D469S)	TCA	24,204,904	3.3 (0.3)	7371644	5.1

Table 5.4 Selected mutants of interest from both libraries after HPA-PAL TLC and HPLC analysis. Samples from phylogenetic library (NNS mutants in Top10 cells) and structural library (NNN mutants in XL10 cells prepared by E. Hibbert) are separated by the dotted line above. [a] standard deviation in brackets. Each row of wild-type data represents the wild-type measurement taken on the same day as for the sample rows below. Wild-type expression levels from the two cell types used are considerably different. PKD peak areas are also a factor of 16-fold higher for the Top10 cell measurements than the XL10 measurements due to an altered protocol for the analysis by HPLC. [b] [TK] not determined so wild-type value assumed. [c] denotes values assumed wild-type level for [TK].

As can be seen in Table 5.4, several mutants, notably H26T, H26A, R358I, H461S, and especially D469T and D469S demonstrate increased activity towards propionaldehyde when compared to wild-type TK.

5.3.5 Structural modelling selected HPA-PAL mutants (energy minimised)

Minimised structural models of selected mutants were generated as protein databank (PDB) files by John Strafford (Dept. Biochemical Engineering, UCL) using the wild-type *E.coli* TK structure file 1QGD as a template then edited and minimised in Discovery studio (Accelrys) using the CharmM potential energy function. Figures 5.11 and 5.12 show the mutants D259F, and D259Y respectively, and highlight their proximity to the TPP cofactor also shown as spheres. While the mutant D259F was not observed, the model helps to discern any structural effects that may occur due to the addition of the hydroxyl group in D259Y.

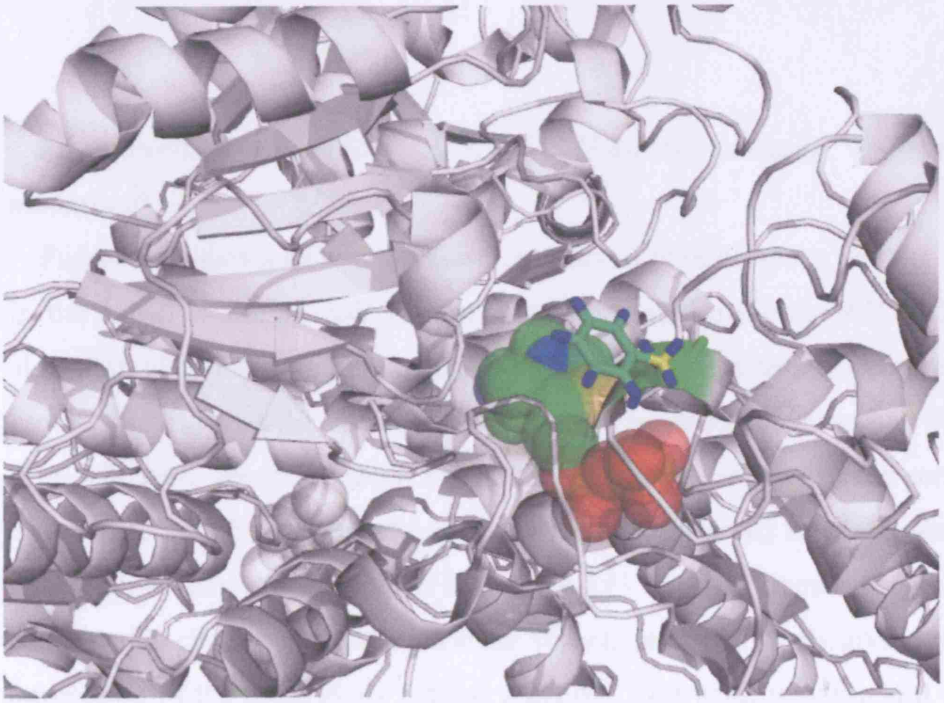


Figure 5.11 D259F showing orientation of modelled novel library residue.

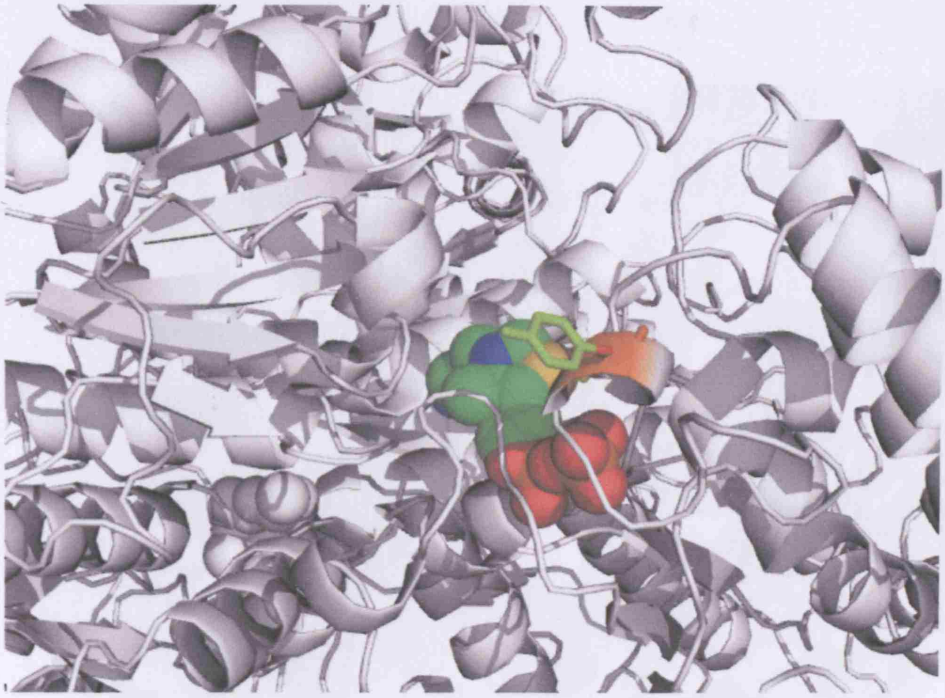


Figure 5.12 D259Y showing orientation of modelled novel library residue.

5.3.6 Structural modelling selected HPA-PAL mutants (non-energy-minimised)

Figure 5.13 shows an energy minimised model of erythrulose product bound in the active site of *E. coli* TK, and was calculated using the AUTODOCK algorithm (J. Strafford unpublished). Figure 5.14 is the same structure as in Figure 5.13 but with the PKD product bound. The structural images in Figures 5.15-5.22 were generated using Pymol and the mutagenesis wizard it contains, to generate possible structural models of mutants observed from the libraries. Unlike minimised models shown in Figures 5.11 and 5.12, these models select the best side-chain rotamer from a small subset, and so will only give a general impression of the most likely residue position. However, they are still useful for gaining some insight into the effect of the substitution and its impact on neighbouring residues and their collective substrate binding.

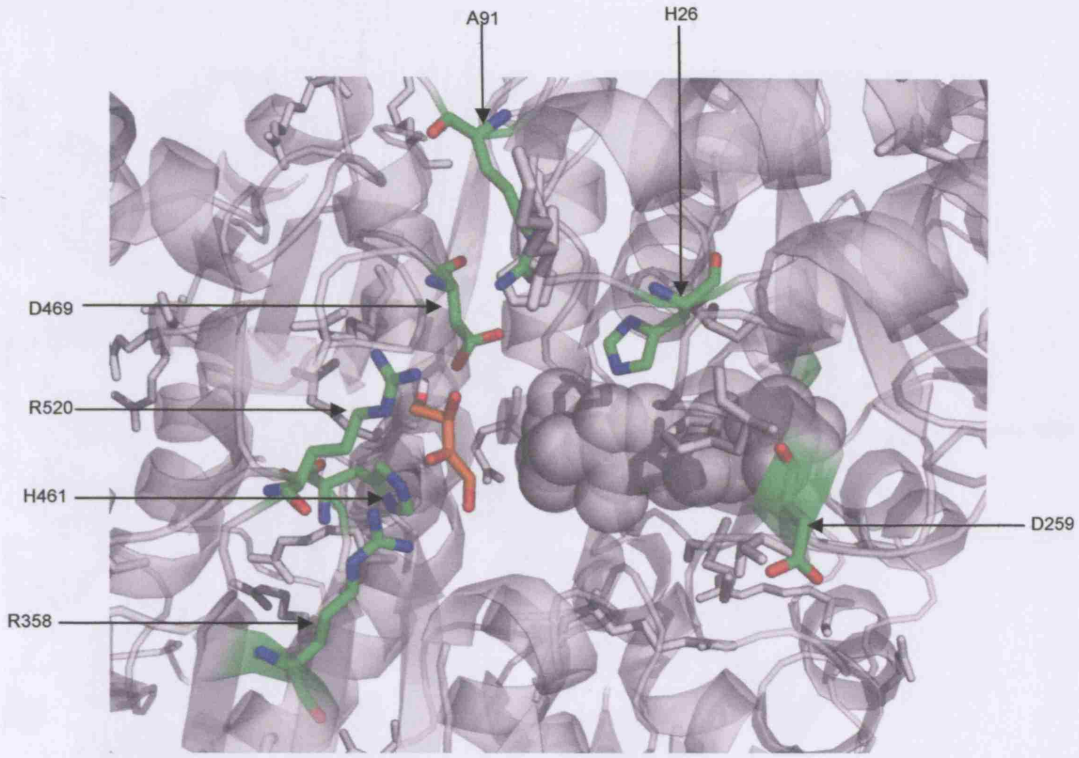


Figure 5.13 Active-site of *E. coli* TK with library residues of interest and their closest neighbours highlighted in green. The erythrulose product is highlighted in orange. TPP cofactor is in grey spheres.

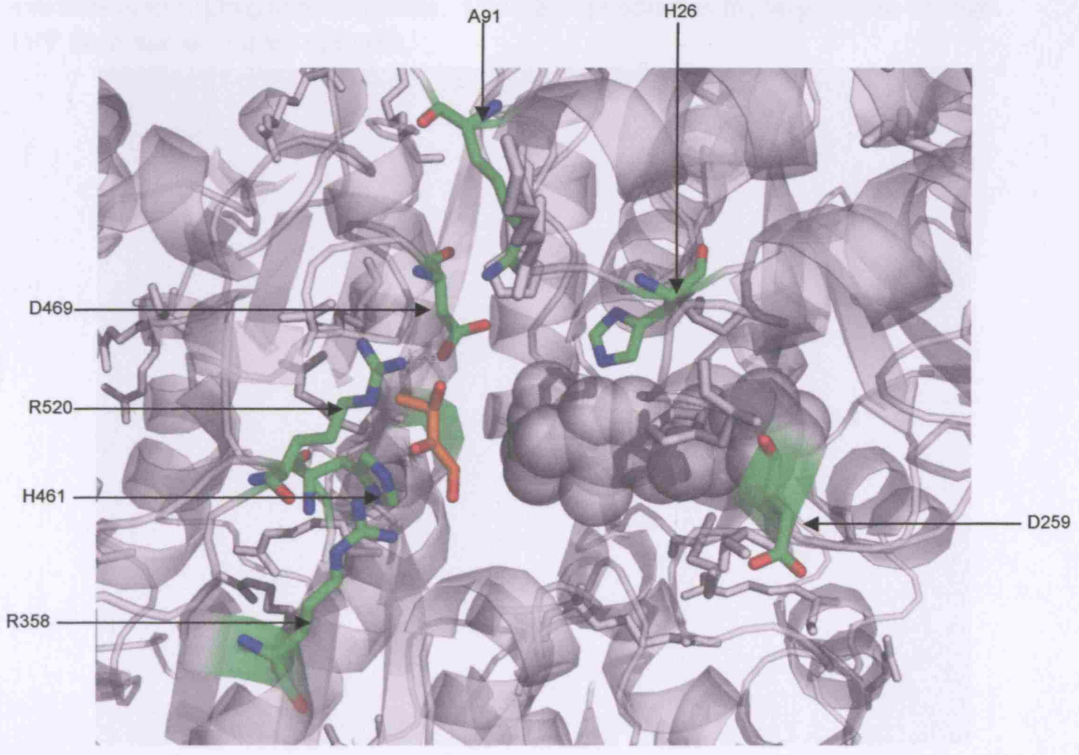


Figure 5.14 Active-site of *E. coli* TK with library residues of interest and their closest neighbours highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

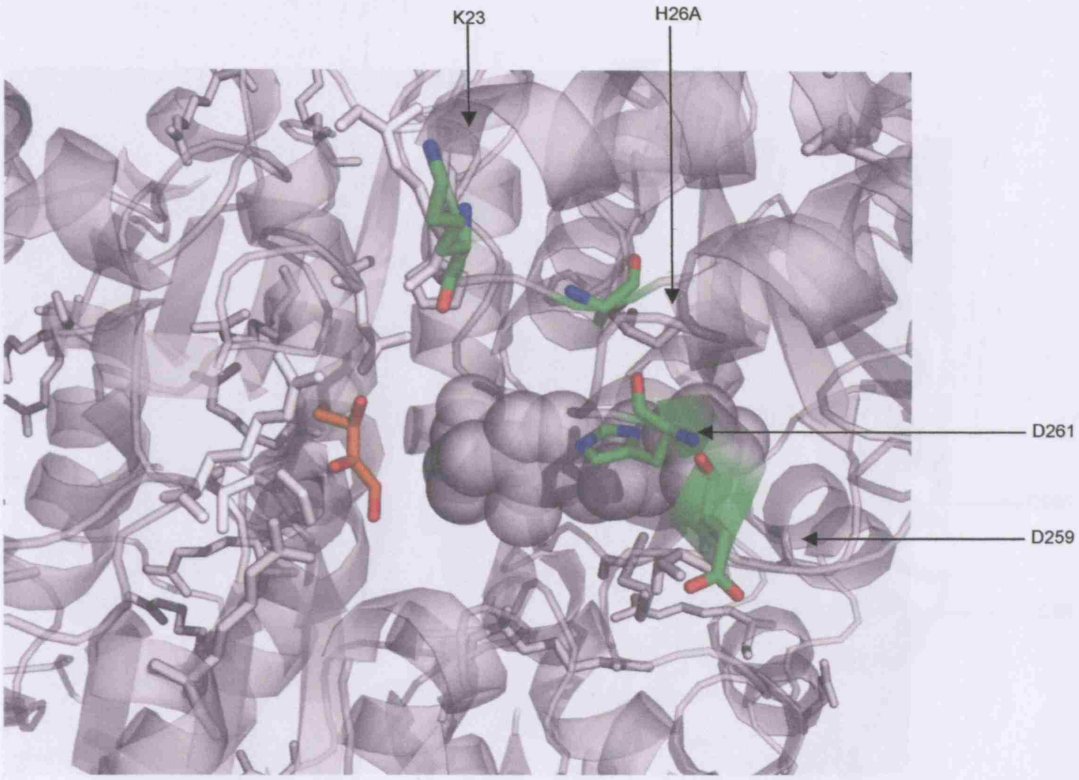


Figure 5.15 Active-site of *E. coli* TK H26A mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

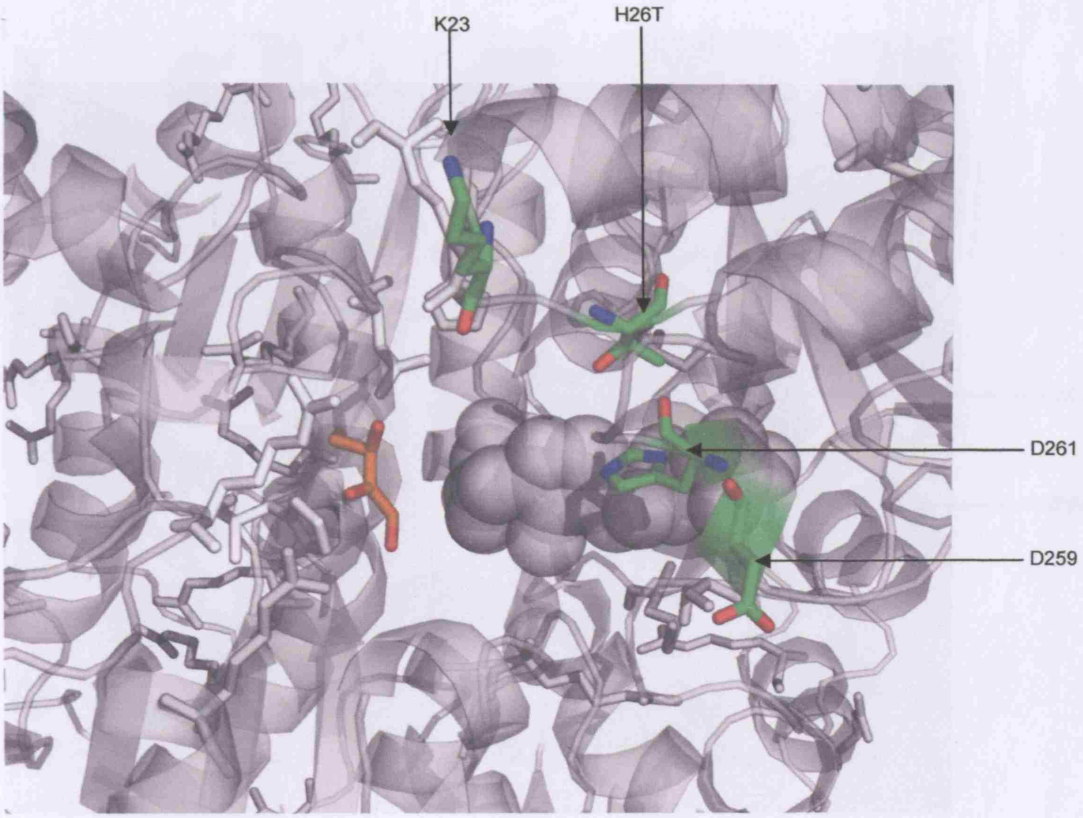


Figure 5.16 Active-site of *E. coli* TK H26T mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

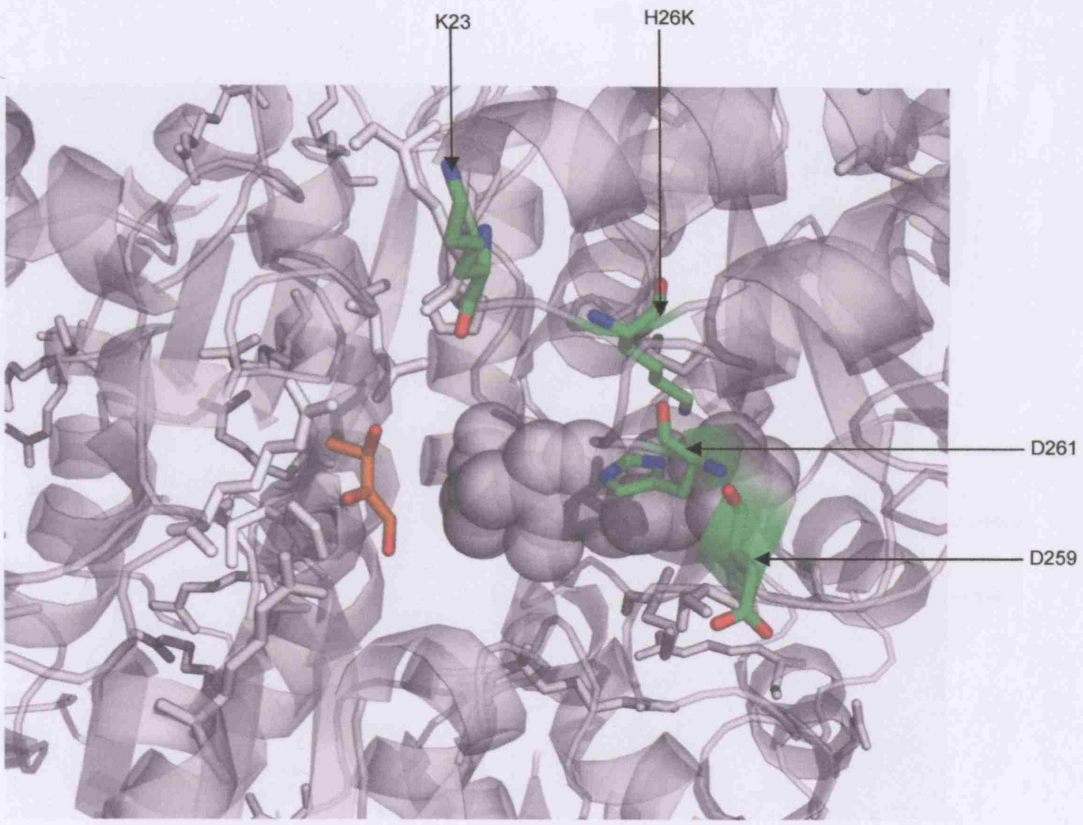


Figure 5.17 Active-site of *E. coli* TK H26K mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

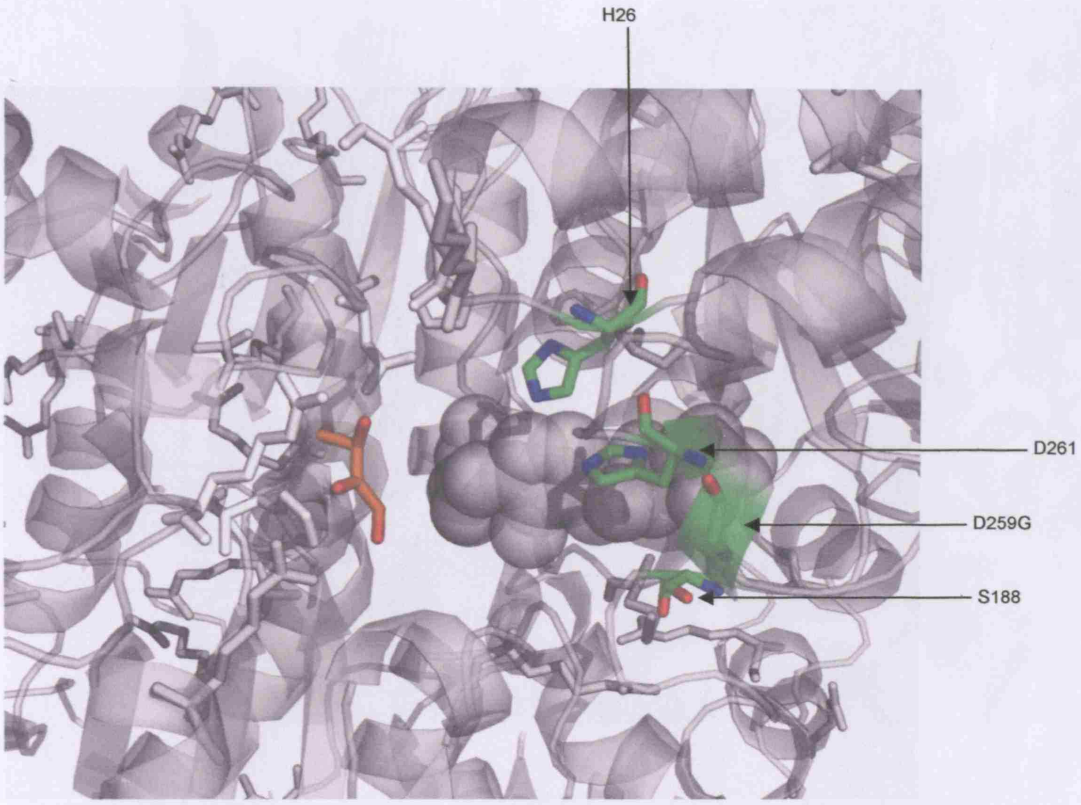


Figure 5.18 Active-site of *E. coli* TK D259G mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

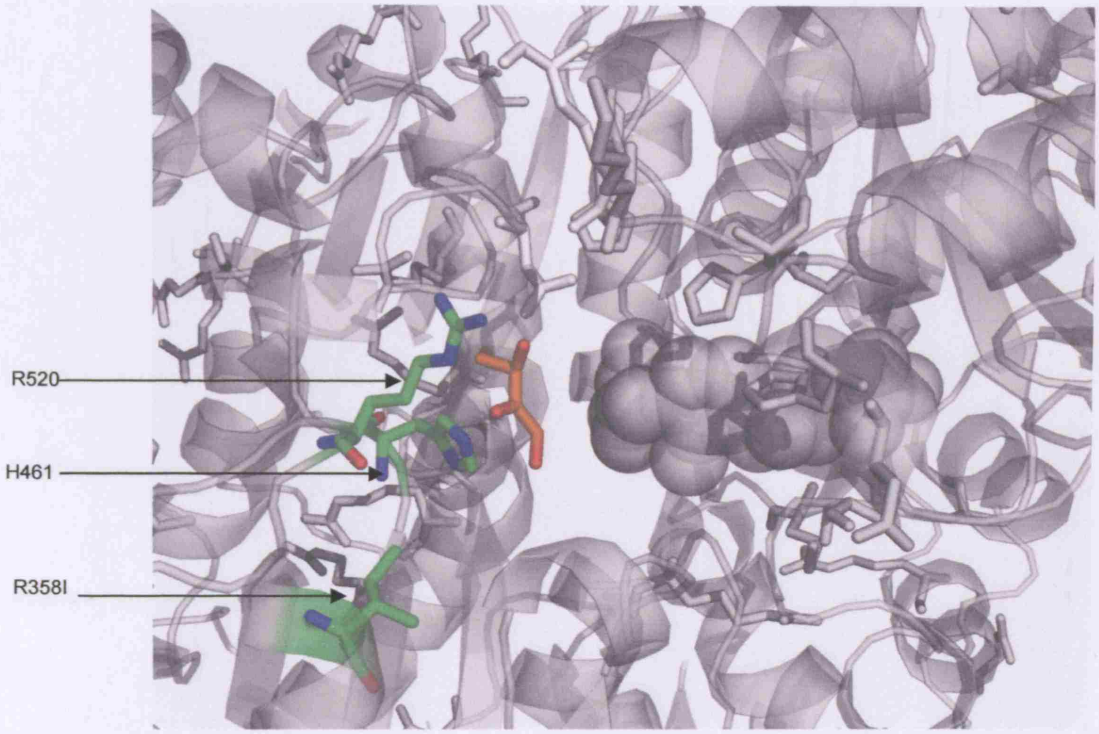


Figure 5.19 Active-site of *E. coli* TK R358I mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

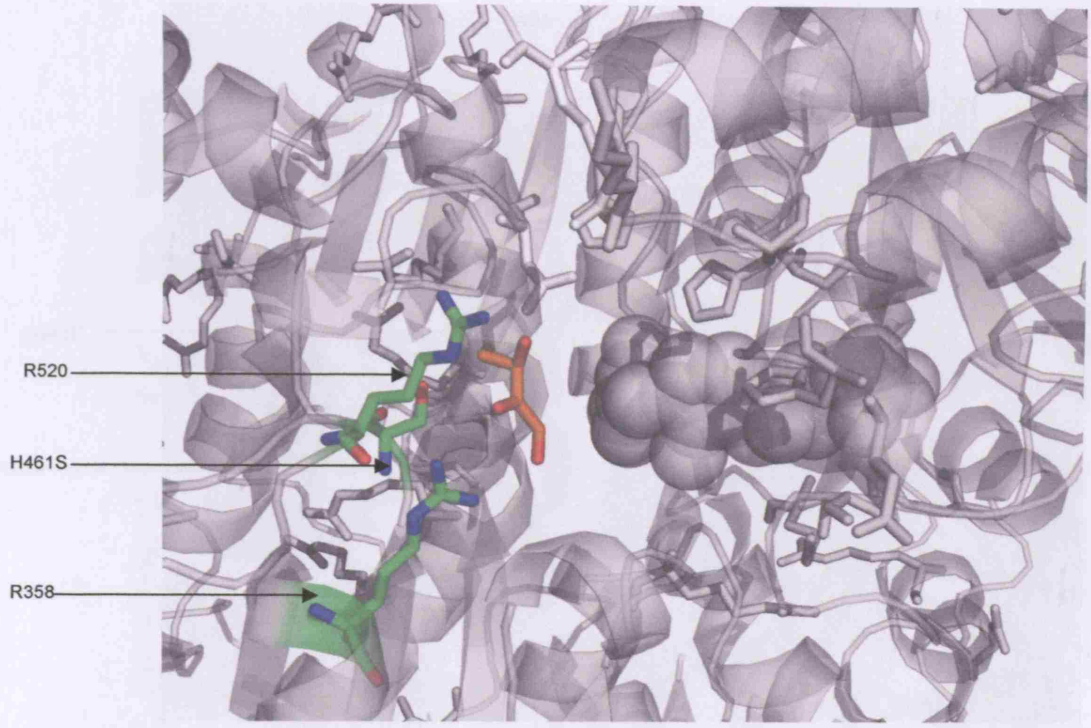


Figure 5.20 Active-site of *E. coli* TK H461S mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

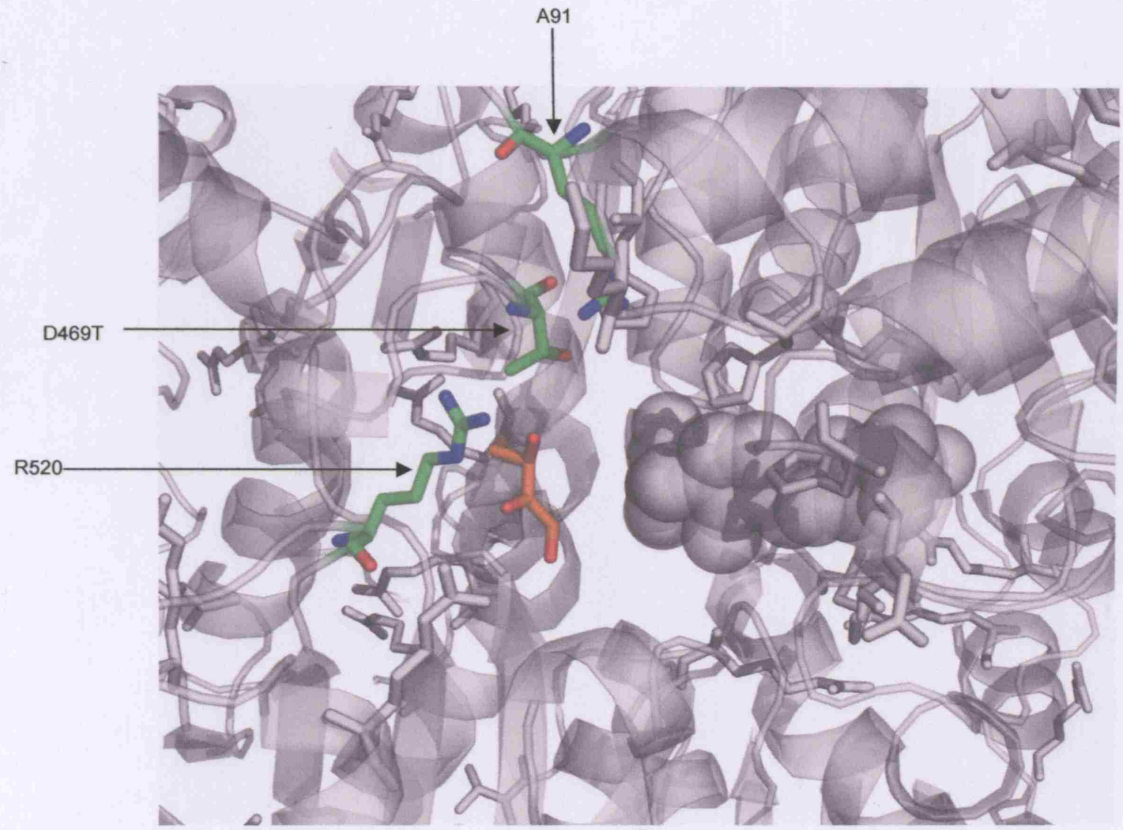


Figure 5.21 Active-site of *E. coli* TK D469T mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

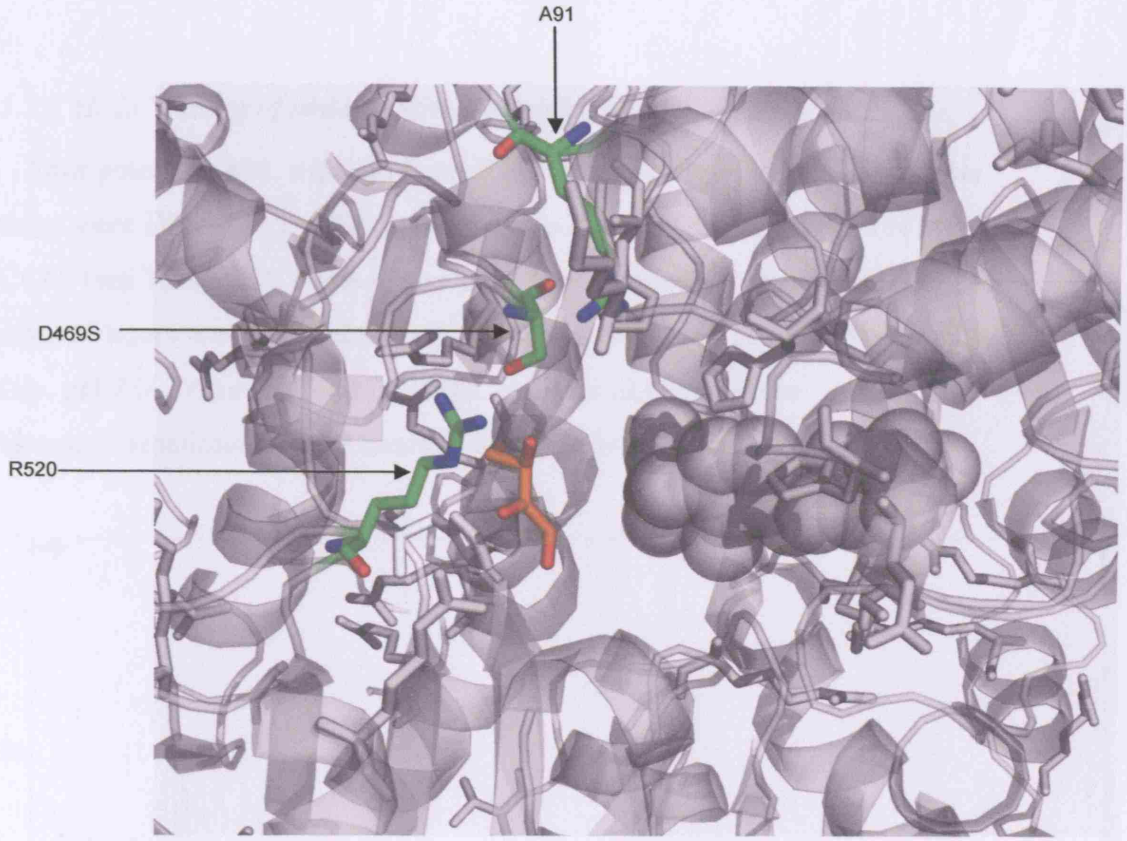


Figure 5.22 Active-site of *E. coli* TK D469S mutant with neighbouring residues also highlighted in green. The PKD product is highlighted in orange. TPP cofactor is in grey spheres.

5.3.7 HPLC testing of mutants with promising activity towards BAL

Four potential BAL mutants from D259 library were selected for analysis these were D259A1 (Tyr TAC), A2 (Ser AGC), H2 (Asp GAC) and H3 (Pro CCC) (see Table 5.3). None of these mutants showed any detectable BKD after 48 hours reaction with 50 mM HPA, benzaldehyde (BAL), 10 mM Gly-Gly, pH 7.0. An example of a HPLC analysis of one of these biotransformations after 48 hours is shown below (Figure 5.23)

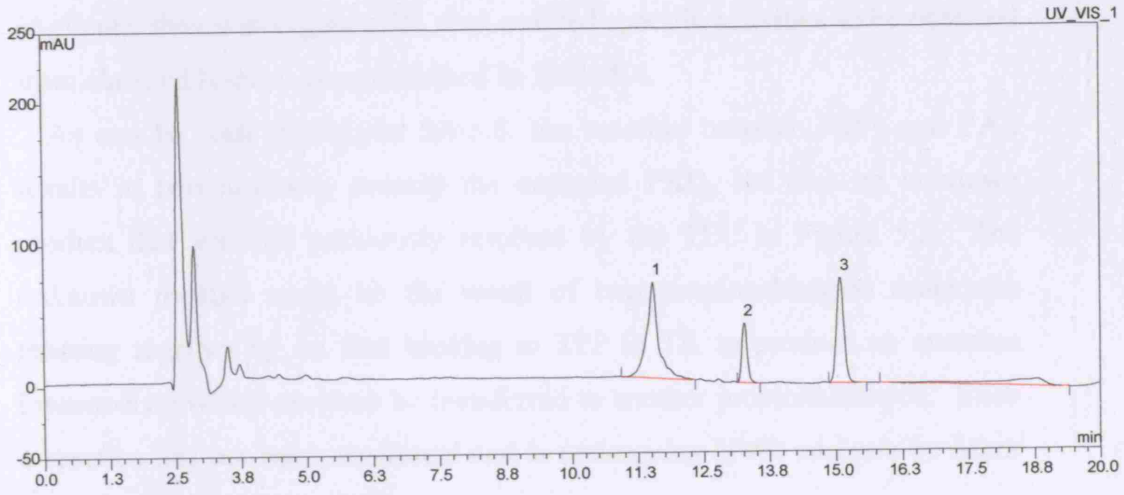


Figure 5.23 A typical HPLC trace showing the substrates and products of a D259G8 (Alanine mutant see Table 5.3) TK catalysed reaction at 48 hours between 50 mM HPA and 10 mM benzaldehyde (BAL), 50 mM Tris.HCl, pH 7.0. Peak 1 is due to a step change in the concentration of ACN in the mobile phase. Peak 2 is benzaldehyde hydrate and peak 3 is benzaldehyde. No discernable BKD peak is observed.

5.4 Discussion

5.4.1 Mutants with altered activity towards propionaldehyde substrate

Screening of the libraries by TLC, against propionaldehyde and HPA identified several mutants that were potentially more active than the wild-type transketolase (Table 5.1). Some of these were then analysed in greater detail by HPLC to monitor the time-course of the reaction, the quality and yield of the product. Measurement of the TK protein concentration by bioanalysis (Agilent), shown in Figure 5.10, also enabled specific activities to be obtained from clarified lysates, as summarised in Table 5.4.

As can be seen in Figures 5.4-5.8, the reaction between HPA and PAL results in two products, namely the expected PKD, but also an unknown product that was not previously resolved by the TLC in Figure 5.2. The unknown product could be the result of two propionaldehyde molecules reacting together by on first binding to TPP in TK to produce an enamine intermediate which can then be transferred to another propionaldehyde. Such a reaction has not been confirmed and is undergoing NMR analysis by Mark Smith (Dept Chemistry, UCL).

As can be seen in Table 5.4, several mutants, notably H26T, H26A, R358I, H461S, and especially D469T and D469S demonstrate increased activity towards propionaldehyde when compared to wild-type TK.

5.4.2 Structure modelling of selected HPA-PAL mutants

Rationalising the mutations observed with their altered activities is not straightforward without detailed kinetic data or new protein structures. However, the mutations are modelled and compared to wild-type TK in Figures 5.11-5.22.

One naturally occurring amino acid, Phenylalanine was not observed (G, S and A were observed) as one of the residues to occur at D259, but a non-naturally occurring residue was namely tyrosine (D259Y). These substitutions were unexpected as phenylalanine is seen in only 1.7 %, and tyrosine in 0 % of

natural variants (Table 4.3). They are also chemically very different to the aspartate that naturally occurs in 16.3% of TK sequences. Aspartic acid is a polar acidic amino acid (negatively charged at physiological pH), whereas phenylalanine is a hydrophobic, aromatic, and bulky amino acid, and tyrosine is similar to phenylalanine but with an additional polar hydroxyl group. It was thought that these residues may sterically block the active site of TK and so modelling was performed by J. Strafford for both of these substitutions to determine their possible orientation. As can be seen from Figures 5.10 and 5.11 neither of these residues appears to significantly obscure the active site or accessibility to the TPP cofactor.

Most of the mutations observed increase the hydrophobicity of the active site, in keeping with the increased hydrophobicity of the propionaldehyde substrate and PKD product, compared to the glycolaldehyde substrate and erythrulose product. Structural analysis reveals a clear link only for the H461S, D469T and D469S mutants. From comparison of Figure 5.20 and Figure 5.14 showing H461S and wild-type TK bound to PKD respectively, it can be seen that the H461S mutation increases the hydrophobicity in a region close to the PKD product. Even more striking are the modelled structures for D469T and D469S in Figures 5.21 and 5.22. In wild-type TK, D469 presents a carbonyl group as part of a carboxylate that can hydrogen bond to erythrulose. The mutation D469T removes this interaction, replacing the carbonyl group with a methyl which can subsequently interact by Van der Waals forces with the methyl group of the PKD product, in an equivalent position to the hydroxyl group in erythrulose. D469S is similar except that the mutation removes the carbonyl group without adding a methyl group. Instead, the PKD product could potentially interact with the hydrophobic methylene (CH2) in the serine side-chain, depending on the preferred rotamer position of the serine hydroxyl group.

5.4.3 Relationship of mutants observed to promiscuity concept

Aharonie and colleagues have observed that it is possible to find evolutionary pathways in which promiscuous activities increased while the

native activity remained virtually unchanged (Aharonie *et al.*, 2005). The mutations in these pathways are primarily found to occur in flexible regions of the protein scaffold rather than in the catalytic or substrate binding residues (Fernandez *et al.*, 2005). However I have observed a number of TK mutants which have exhibited propionaldehyde activity, and yet did not exhibit significant activity towards the model (HPA-GA) reaction in Chapter 4. For example, D259S and D469S both display poor activity with HPA and GA, and yet produce 40% and 500% respectively, of the yield of PKD obtained by wild-type after 94 hours.

Most of the best performing mutants for the HPA-PAL reaction have retained 'wild-type-like' activity or better (see Table 5.4), with all improved mutants being found in the structural libraries. The six most significant mutants, H26T (G5), H26A (B12), R358I (A4), H461S (H12), D469T (F11), and D469S (B6) (Table 5.4), are all from the structural libraries. These results slightly conflicts with the findings of Aharonie *et al.*, in that these residues are most likely to be involved in substrate binding and yet they seem to have allowed for improvement of TK promiscuous activities. These four mutants are also of interest because none of them are variants observed in natural sequences at these sites (Table 4.4). In natural sequences, H26 is 100% conserved, yet H26 can be mutated to T, A or K with improved activity for PKD production. Similarly for R358I, the natural sequence variation is R, P, A or T. Compared to H461S the natural variation observed is H or Q, and compared to D469T or D469S, the natural sequence variation is D or N.

The observed non-naturally occurring structural mutants perhaps suggest that HPA-PAL activity may be a promiscuous activity and that improvement of this activity requires non-natural variations, some of which, but not all, retain activity towards the natural-like HPA-GA reaction. A striking result is that in Chapter 4, mutants that improved the HPA-GA (natural-like) reaction, came from both the phylogenetically and structurally defined libraries, whereas mutants that improved the non-natural HPA-PAL reaction all came from the structurally defined libraries. This indicates that to obtain novel non-natural reactions it is useful to target residues that are highly conserved in natural sequences.

Four mutants with significant yet less than wild-type activity were obtained from the phylogenetic library, with activities descending in the order D259A, D259S, D259G and then D259Y. Three of these, D259A, D259S, and D259G are residues observed in the natural sequence variation, whereas D259Y is novel, yet similar to the D259F observed in 1.7% of natural sequence variants (Table 4.3).

It is possible that the unknown product observed for the D259 library mutants have in some way detracted from the yield of PKD obtained, especially if these mutations enhance an activity similar to that observed in benzaldehyde lyase. This reaction facilitates the condensation of two benzaldehyde molecules and may be similar for the mutants in producing a PAL-PAL product. However this would need further investigation by NMR and mass spectrometry to determine the identity of the unknown product.

5.4.4 Mutants with altered activity towards benzaldehyde substrate

Promising mutants that appeared to form a new product with HPA and benzaldehyde were observed by TLC, although only very low levels of a possible BKD product were present (Figure 5.3). However upon further testing by HPLC, no BKD-like product was observed (Figure 5.23). The low levels of possible TLC signal as well as an inability to detect activity in the HPLC analysis led to the inability to further analyse these potential mutants. The lack of readily available BKD product standards (due to synthesis difficulties for Dept. chemistry) made identification of possible TLC products even harder to verify by TLC or HPLC, and also the lack of any activity using wild-type TK made a comparative analysis impossible. Overall it appears that the directed evolution strategy taken is not capable of finding any mutants with the desired activity towards benzaldehyde.

5.5 Conclusions

The discovery of mutants from structural library sites, with significant improvements in activity towards propionaldehyde, were interesting in that they are thought to be likely to influence substrate binding or the active-site conformation. Beneficial mutations at such highly conserved and directly substrate-interacting residues was not previously thought typical (Fernandez *et al.*, 2005) and consequently I was expecting the more significant improvements to come from the phylogenetic library residues which fit the typical criteria more in terms of being naturally variant and only indirectly interacting with substrates. Another point of interest was the observation that some mutants displayed activity towards propionaldehyde despite poor activity towards the natural analogue glycolaldehyde substrate. It is typically thought that wild-type activity is retained until sufficient improvements in the novel activity occur. However this was not always the case when comparing mutants active towards glycolaldehyde and propionaldehyde, as some mutations appeared to be able to significantly switch the preference between the two substrates.

Although I do not believe any of these mutants have been changed to a new class of enzyme, it is possible that other novel activities have arisen along with the loss of HPA-GA activity in some cases, which could be the first step in the process of altering the reaction catalysed. Some variants were observed which retained activity towards glycolaldehyde, but showed no discernible improvement in HPA-PAL activity, for example D259S, D259A, D259G and D259Y. It is possible that these mutants are retaining HPA-GA activity before being able to alter substrate specificity in line with the model of promiscuity (Aharonie *et al.*, 2005). The natural variation observed at D259 is K, A, D, G, S, E, L, F, R and V (one of the most variable sites) but not Y. Further mutation of these variants with mutants observed at other residues of interest may result in greatly improved activities, even where neither single mutant exhibits a significant improvement in activity at the first round of mutation. D259G was the most common of mutations for PAL-HPA mutants, with 12

out of 26 active mutants found in this library having this mutation. Although the D259 activities were not found to improve upon existing activity, these may be good targets for finding synergistic mutations.

5.6 Future and Ongoing work

The mutants observed in the D259X library, which is also the most variable site in terms of natural sequences, may have the potential to influence substrate specificity. It would, therefore, be worth further investigation of these mutations in a round of double mutant production, to look for potential synergistic coupling and further improvements in substrate specificity.

It is essential that the enantioselectivity of TK is retained in mutants with improved activity or altered substrate specificity. Therefore, the enantiomeric excess of products needs to be investigated. Mark Smith and Kirsty Smithies (Dept of Chemistry, UCL) have been recently developing a chiral GC analysis of HPA-PA biotransformation products. The mutants analysed to date with this method have been D469T, D469A as well as the wild-type TK. Preliminary data suggests enantiomeric excesses of PKD product of 58% for wild type, 64% for D469T, and 33% for D469A, although which enantiomer is in excess is still to be determined.

Now that it has been observed that a single round of directed evolution by the methods used, can improve activity towards propionaldehyde but not achieve activity towards benzaldehyde, then a series of stepping-stone substrates between these two will need to be constructed to test the 'substrate-walking' concept. An example series is suggested in Figure 5.24.

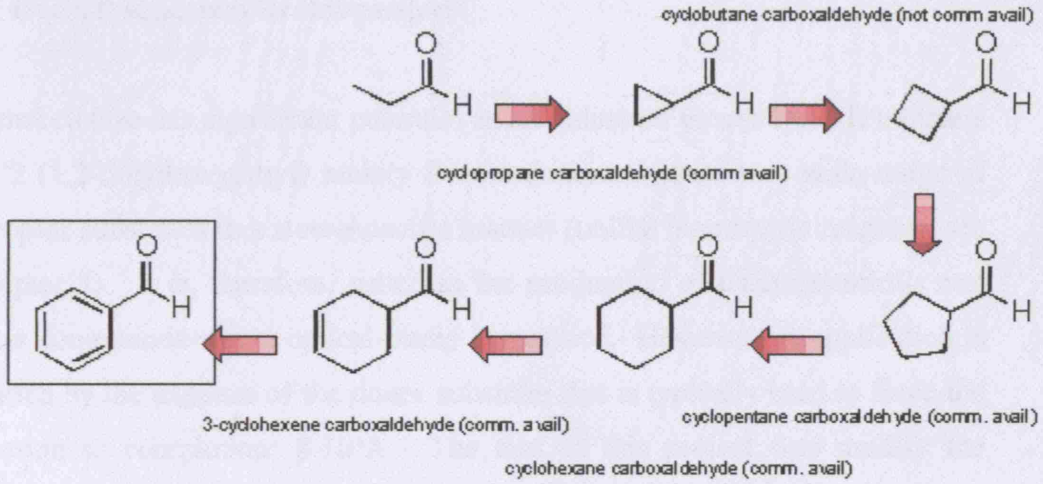


Figure 5.24 Potential series of substrates for testing the substrate-walking concept.

Chapter 6 – General discussion

6.1 Overall summary of this project

Transketolase has significant potential as an industrial biocatalyst. It transfers a C2 (1,2-dihydroxyethyl) moiety from a donor substrate to a wide range of acceptor substrates in a stereospecific manner (unlike biomimetic reagents, see Chapter 3). It is, therefore, suited to the production of pharmaceuticals and other compounds where optical-purity is required. However, its application is limited by the expense of the donor substrate that is typically used to force the reaction to completion: β-HPA. The aim of this project was modify the substrate-specificity of *E. coli* transketolase so that it would accept alternative ketols as donor substrates. The other ketols tested were vastly cheaper than β-HPA but were chemically similar and would still render the reaction irreversible as they all liberate CO_2 as a byproduct. However no initial activity was determined for any of the novel ketol donors selected for testing with unmodified TK, as a result the focus was changed to finding novel aldehyde acceptors preferably with some existing level of activity with TK (e.g. propionaldehyde 5% of that of GA). HPA was also synthesised (chapter 2) rather than commercially bought to reduce the cost of the ketol donor since its replacement was not feasible.

E. coli transketolase required some modifications for it to be able to catalyse the desired reaction with improved reaction rates. As mentioned (chapter 4) two library approaches were tried in parallel phylogenetic (NNS primers in TOP10 cells) and structural library (NNN primers in XL10 produced and tested by E. Hibbert).

The library types each had ten sites of residues chosen for SSDM were K23, A29, N64, M159, S188, D259, A383, P384, V409 and L466 for the library based on phylogenetic studies and H26, H100, I189, H261, G262, R358, S385, H461, D469 and R520, based on the structural analysis.

A variety of high-throughput screens for the desired activities were

developed prior to and during library construction. The screens has the following general steps: (1) transformation of the plasmid library into *E. coli* TOP10 or XL10-Gold competent cells; (2) microwell fermentation of individual colonies; (3) lysis of the cultures; (4) incubation of the lysates with cofactors and the target substrates (HPA and GA/BAL/PAL etc); and finally (5) high-throughput (as much as is possible depending on the compound screened) HPLC analysis measuring donor substrate (HPA) depletion and product formation.

The compositions of the libraries were determined by Sambrook cycle sequencing clones of interest. Expression level analysis was also carried out on library clones, both bioanalysis and densitometry was tested as means to obtain accurate measure of TK concentration present (chapter 4). Bioanalysis was the method of choice to be taken forward for subsequent expression level analysis carried out (Chapter 5) due to the ease of use and reproducibility of this system.

6.2 Overall conclusions

The following conclusions were reached during the development of the high-throughput screens for the desired activities:

- Of the *E. coli* strains tested, XL10-Gold and TOP10 were the best hosts for the pQR791 plasmid; the libraries were constructed using both strains as a result. However for long term culturing of library plates it was found that XL10 was more stable cell line and less likely to plasmid copy number loss or changes in expression levels.
- During assay development it was noticed that two buffers facilitated the conversion of TK substrates in absence of the enzyme (Hepes and Mops buffers) this was termed biomimetic reaction (chapter 3) and as a result these buffers are not viable for use in novel screens, Tris-HCl or Gly-Gly buffers must be employed.
- During initial ketol donor assay it was also observed that FlPA was able to act as an inhibitor to HPA-GA reaction occurring (chapter 4),

this needs further experimental analysis, but work in ongoing to develop a possible crystal structure (J. Littlechild University of Exeter) with FlPA bound which might elucidate residues involved in binding of ketol donors.

- A variety of accurate, high-throughput HPLC assays for the desired activities were developed. Using either the 50mm ion exchange guard column or ACE5 short C18 column depending on the type of aldehyde to be screened.
- The discovery of a biomimetic reaction was not only of interest as a technique to synthesis some chemicals in a one step process (albeit a racemic mixture), but also demonstrated the need for thorough assay development when working with novel substrates.
- The use of TLC as a rapid screening method for hit or miss answers in terms of yes or no activity in mutants, allowing selection of a subset for the more accurate slower HPLC analysis methods allowed faster analysis of library plates than if all required HPLC analysis for initial screening results (assay times then being the limiting factor).

The following conclusions were reached during the development SSDM and the construction of transketolase libraries:

- The residues most likely to perform the role of enhancing β-HPA/GA discrimination in *E. coli* transketolase are A29 (E&D), I189 (Y) and 461 (S). Mutants obtained from these residues SSDM library displayed increased activity and reaction rates in presence of HPA-GA than unmodified pQR791 TK.
- The residues most likely to positively affect aldehyde donor substrate specificity are H26 A, T and K, R358 I, H461 S and D469 T and S. These appeared to have improved activity and reaction rates towards HPA-PAL compared to unmodified TK.
- The residues which may have caused substrate switch in that these mutants were observed to have potential HPA-PAL or HPA-BAL activity despite no initial sign of HPA-GA activity remaining on initial screens are M159, D259 and P384 (Table 5.4).

- Residues no longer under selection pressure, such as the phosphate-binding R358 and R520, showed the greatest tolerance and range of acceptable mutations.

The following conclusions were reached during the screening step:

- The first phase did reveal some of the desired activity in the libraries screened, improvements were detected in phylogenetic libraries screened for the HPA-GA reaction. This suggests that a single point mutation in any of the target libraries can yield useful mutations however in this case the greatest impact appears to be from mutants in the phylogenetic library.
- The second phase of screening with novels substrates all of which excluding propionaldehyde exhibited no initial activity in presence of unmodified TK identified variants with improved activity towards HPA-PAL in structural libraries. Changes also occurred in some of the phylogenetic library mutant reaction rates although none showed a marked improvement in comparison to wt controls. It is clear that single point mutation in any of the target sites was capable of yielding the desired activity however it seems the greatest improvements in this case were found in the structural library.

The failure of this project to identify definitively (TLC detection, but no HPLC detection possibly due to dilution constraints to load samples to UV detector 1:1 v/v) a single variant of transketolase with a novel desired activity introduced was extremely disappointing. It is possible that the limited sensitivity of the HPLC assay meant that variants with very low levels of the desired activity were screened but not identified (although the detection of possible BKD product at TLC should have allowed detection by HPLC, if the BKD product observed was real and not an artefact). A possible explanation for this lack of variants with detectible novel activities is that the project goal was too ambitious for the libraries that were screened. The project goal was ambitious for two reasons: (a) too many significant modifications to the

enzyme functionality were required; and (b) the enzyme did not exhibit the desired activity to any extent to begin with. It is also possible that since the residues targeted were those that varied among TKs during evolution and no TK has been shown to use BAL, perhaps a better approach would be to target highly conserved residues that may be preventing use of substrates, such as BAL. Targets could be suggested by comparison of TK with evolutionary related enzymes that can use BAL (if there are any). This may be also be the limit of change that is achievable in one round of evolution and further rounds of SSDM employing double and synergistic mutant pairs from both libraries maybe what is required to achieve the next step in substrate walking to a substrate with no initial detectable activity.

6.3 Future work

Possible extensions to this project that might enable the achievement of the project goal are:

- Further testing of libraries with other novel aldehyde substrates forming new reactivity series such as cyclic aldehydes, to further establish the limits of change that one round of evolution can facilitate in substrate specificity.
- Increase the number of target sites or vary other sites in mutants of interest already found in an effort to find a second mutation which may allow a greater switches in specificity.
- Analyse the structural and phylogenetic libraries and attempt to identify possible sites which mutant coupling could be employed and its impacts explored on changes in substrate specificity.
- Establish the mode of mechanism of inhibition of FlPA as well the K_i and the production of a crystal structure of FlPA bound to *E.coli* TK.
- Employ colourimetric screening as a more quantitatively comparable method (in comparison to TLC) to establish levels of conversion of novel substrates between library sites.

- Attempt to improve the quality of libraries (reduced redundancy of mutation) as well as the quality of screens employed to identify an possible variants of interest in the most efficient way possible.
- Produce all further libraries in single host strain XL10-Gold to avoid problems of differing expression levels as well as changes in plasmid copy number encountered with other library production strains.
- GC chiral column analysis and mass spec analysis of HPA-PAL mutant products.

Reference List

Abecassis V., Pompon D., Truan G. 2000. High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *PNAS 28*:e88.

Aharoni A., Gaidukov L., Khersonsky O., Mc Gould S., Roodvelt C., Tawfik D. 2005. The 'evolvability' of promiscuous protein functions. *Nature Genetics 37*:73-76.

Amin N., Liu A.D., Ramer S., Aehle W., Meijer D., Metin M., Wong S., Gualfetti P., Schellenberger V. 2004. Construction of stabilized proteins by combinatorial consensus mutagenesis. *PEDS 17*:787-793.

Arnold F. 1998. Design by directed evolution. *Acc Chem Res 31*:125-131.

Ashworth J., Havranek J., Duarte C., Sussman D., Monnat R., Stoddard B., Baker D. 2006. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*656-659.

Bernath K., Hai M., Mastrobattista E., Griffiths A., Magdassi S., Tawfik D. 4 A.D. In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Analytical Biochemistry 325*:151-157.

Bessler C., Schmitt J., Maurer K., Schmid R. 2003. Directed evolution of a bacterial -amylase: Toward enhanced pH-performance and higher specific activity. *Protein Science 12*:2141-2149.

Blass J., Gibson G. 1979. Genetic factors in Wernicke-Korsakoff syndrome. *Alcohol Clin Exp Res 3*:126-134.

Blass J., Piacentini S., Boldizsar E., Baker A. 1982. Kinetic studies of mouse brain transketolase. *J Neurochem 39*:729-733.

Bloom J., Labthavikul S., Otey C., Arnold F. 2006. Protein stability promotes evolvability. *PNAS 103*:5869-5874.

Bogarad L., Deem M. 1999. A hierarchical approach to protein molecular evolution. *Proc Natl Acad Sci USA 96*:2591-2595.

Bongs J., Hahn D., Schorken U., Sprenger G.A., Kragl U., Wandrey C.. 1997. Continuous production of erythrulose using transketolase in a membrane reactor. *Biotechnology Letters 19*:213-216.

Bornscheuer U., Pohl M. 2001. Improved biocatalysts by directed evolution and rational protein design. *Current Opinion in Chemical Biology 5*:137-143.

Bornscheuer U., Kazlauskas R. 2004. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angewandte Chemie International Edition 43*:6032-6040.

Brandon.C., Tooze.J. 1991. *Introduction to protein structure*. Second Edition Published by Garland.

Burton S., Cowan D., Woodley J.M. 2002. The search for the ideal biocatalyst. . *Nat Biotechnol 20*:37-45.

Carter P, Wells J. 1987. Engineering enzyme specificity by "substrate-assisted catalysis". *Science 237*:394-399.

Chauhan R., Woodley J., Powell L.. 1996. *In situ* product removalfrom *E. coli* transketolase-catalyzed biotransformations. *NY Acad Sci 199*:545-554.

Cochran J., Kim Y., Olsen M., Bhandari R., Wittrup K. 2004. Domain-level antibody epitope mapping through yeast surface display of epidermal growth factor receptor fragments. *Journal of Immunological Methods 287*:147-158.

Cochran J., Kim Y., Lippow S., Rao B., Wittrup K. 2006. Improved mutants from directed evolution are biased to orthologous substitutions. *PEDS 19*:245-253.

Coco W., Levinson W., Crist M., Hektor H., Darzins A., Plenkos P., Squires C., Monticello D. 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnology 19*:354-359.

Coco W. 2003. RACHITT: Gene family shuffling by Random Chimeragenesis on Transient Templates. *Methods Mol Biol 231*:111-127.

Comin-Anduix B., Boren J., Martinez S., Moro C., Centelles J.J., Trebukhina R., Petushok N., Lee W.N., Boros L.G., Cascante M. 2001. The effect of thiamine supplementation on tumour proliferation. A metabolic control analysis study. *Eur J Biochem 268*:4177-4182.

Comin-Anduix B., Boros L.G., Marin S., Boren J., Callol-Massot C., Centelles J.J., Torres J.L., Agell N., Bassilian S., Cascante M. 2002. Fermented wheat germ extract inhibits glycolysis/pentose cycle enzymes and induces apoptosis through poly(ADP-ribose) polymerase activation in Jurkat T-cell leukemia tumor cells. *J Biol Chem 277*:46408-46414.

Copley S. 2003. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Current Opinion in Chemical Biology 7*:265-272.

Corey D., Willet W., Coombs G., Craik C. 1995. Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry 34*:11521-11527.

Dalby P, 2003. Optimising enzyme function by directed evolution. *Current Opinion in Structural Biology 13*:1-6.

Daniels D., Chen H., Hicke B., Swiderek K., Gold L. 2003. A tenascin-C aptamer identified by tumour cell SELEX: Systematic evolution of ligands by exponential enrichment. *PNAS 100*:15416-15421.

Datta A.G., Racker E. 1961. Mechanism of Action of Transketolase properties of the crystalline yeast enzyme. *J Biol Chem 236*:617-623.

De La Haba G., Leder I., Racker E. 1955. Crystalline Transketolase from bakers' yeast: Isolation and properties. *J Biol Chem 214*:409-426.

de Raadt A., Griengl H. 2002. The use of substrate engineering in biohydroxylation. *Current Opinion in Biotechnology 13*:537-542.

Demuynck C., Bolte J., Hecquet L., Samaki H.. 1990. Enzymes as reagents in organic chemistry: transketolase-catalysed synthesis of D-[1,2-^{13C2}]xylulose. *Carbohyd Res 206*:79-85.

Demuynck C., Bolte J., Hecquet L., Dalmas V. 1991. Enzyme-catalyzed synthesis of carbohydrates: synthetic potential of transketolase. *Tetrahedron Lett 32*:5085-5088.

Drummond D.A., Iverson B.L., Georgiou G., Arnold F.H. 2005. Why higherror-rate random mutagenesis libraries are enriched in functional and improved proteins. *J Mol Biol.* 806-816.

Du M.X., Sim J., Fang L, Yin Z., Koh S., Stratton J., Pons J., Wang J., Carte B. 2004. Identification of Novel Small-Molecule Inhibitors for Human Transketolase by High-Throughput Screening with Fluorescent Intensity (FLINT) Assay. *J Biomol Screen 9*:427-433.

Effenberger F., Null V. 1992. Enzyme-catalyzed reactions a new, efficient synthesis of fagomine. *Liebigs Ann Chem 11*:1211-1212.

Eijsink V., Gaseidnes S., Borchert T., van den Burg B. 2005. Directed evolution of enzyme stability. *Biomolecular Engineering 22*:21-30.

Farber G., Petsko G. 1990. The evolution of a/b barrel enzymes. *Trends Biochem Sci 15*:228-234.

FDA. Policy statement for the development of new stereoisomeric drugs. 1992. http://www.fda.gov.cder.guidance.stereo.htm.

Fernandez A., Tawfik D., Berkhout B., Sanders R., Kloczkowski A., Sen T., Jernigan B. 2005. Proteing promiscuity: Drug resistance and native functions-HIV-1 Case. *Journal of Biomelecular Structure & Dynamics 22*:615-624.

Flores H., Ellington A. 2005. A modified consensus approach to mutagenesis inverts the cofactor specificity of Bacillus stearothermophilus lactate dehydrogenase. *PEDS 18*:369-377.

Freeman A., Cohen-Hadar N., Abramov S., Modai-Hod R., Dror Y., Georgiou G. 2004. Screening of Large Protein Libraries by the "Cell Immobilized on Adsorbed Bead" Approach. *Biotechnol Bioeng 86*:196-200.

French C., Ward J.M. 1995. Improved production and stability of E.coli recombinants expressing transketolase for large scale biotransformation. *Biotech Letts 17*:247-252.

Glieder A., Farinas E.T., Arnold F.H. 2002. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat Biotechnol 20*:1135-1139.

Gonzalez-Blasco., Sanz-Aparicio J., Gonzalez B., Hermoso J., Polaina J. 2000. Directed Evolution of b-Glucosidase A from *Paenibacillus polymyxa* to Thermal Resistance. *The Journal of Biological Chemistry 275*:13706-13712.

Goud G., Artsaenko O., Bols M., Sierks M. 2001. Specific glycosidase activity isolated from a random phage display antibody library. *Biotechnol Prog*197-202.

Gould S., Tawfik D. 2005. Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry 44*:5444-5452.

Greener A., Callahan M., Jerpseth B. 1997. An efficient random mutagenesis technique using an *E. coli* mutator strain. *Molecular Biotechnology 7*:189-195.

Griffiths A., Tawfik D. 2003. Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization. *EMBO J 22*:24-35.

Guerard C., Alphand V., Archelas A., Demuynck C., Hecquet L., Furstoss R., Bolte J. 1999. Transketolase-mediated synthesis of 4-deoxy-D-fructose 6-phosphate by epoxide hydrolase-catalysed resolution of 1,1-diethoxy-3,4-epoxybutane. *Eur J Org Chem 3399*-3402.

Guo K., Wendel H., Lutz S., Ziemer G., Scheule A. 2005. Aptamer-based capture molecules as a novel coating strategy to promote cell adhesion. *J Cell Mol Med 9*:731-736.

Hayes R., Bentzien J., Ary M., Hwang M., Jacinto J., Vielmetter J., Kundu A., Dahiyat B. 2002. Combining computational and experimental screening for rapid optimization of protein properties. *Proc Natl Acad Sci USA*15926-15931.

Hecquet L., Bolte J., Demuynck C. 1996. Enzymatic synthesis of "natural-labeled" 6-deoxy-L-sorbose, precursor of an important food flavour. *Tetrahedron 52*: 8223-8232.

Heinrich P.C., Wiss O. 1971. Transketolase from human erythrocyte: purification and properties. *Helv Chim Acta 54*:2658-2668.

Hobbs G., Mitra R., Chauhan R., Woodley J., Lilly M.. 1996. Enzyme-catalysed carbon–carbon bond formation: large-scale production of *Escherichia coli* transketolase. *J Biotechnol 45*:173-179

Horecker B.L., Smyrniotis P.Z., Hurwitz J. 1956. The role of xylulose 5-phosphate in the transketolase reaction. *J Biol Chem 223*:1009-1019.

Humphrey A.J., Parsons S.F., Smith M.E.B., Turner N.J. 2000. Synthesis of a novel N-hydroxypyrrolidine using enzyme catalysed asymmetric carbon-carbon bond synthesis. *Tetrahedron Letters 41*:4481-4485.

Jensen R. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol 30*:409-425

Joerger A., Mayer S., Fersht A. 2003. Mimicking natural evolution *in vitro*: An *N*-ac etylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity. *PNAS 100,* 5694-5699.

Johannes.F., Dressler.D., Wilde.U., Schubert.P. 2005. Mutations in the Transketolase-like Gene TKTL1: Clinical Implications for Neurodegenerative Diseases, Diabetes and Cancer. *Clin Lab:* 257-273.

Kauffman S. 1993. The origins of order. *Oxford University Press (Oxford, UK)*

Kiely M., Tan E., Wood T. 1969. The purification of transketolase from *Candida utilis*. *Can J Biochem 47*:455-460.

Kjosen B., Seim S. 1977. The transketolase assay of thiamine in some diseases. *American Journal of Clinical Nutrition 30*:1591-1596.

Kobori Y., Myles D., Whitesides G. 1992. Substrate specificity and carbohydrate synthesis using Transketolase. *J Org Chem 57*:5899-5907.

Langbein S., Zerilli M., Zur H.A., Staiger W., Rensch-Boschert K., Lukan N., Popa J., Ternullo M.P., Steidler A., Weiss C., Grobholz R., Willeke F., Alken P., Stassi G., Schubert P., Coy J.F. 2006. Expression of transketolase TKTL1 predicts colon and urothelial cancer patient survival: Warburg effect reinterpreted. *Br J Cancer 94*:578-585.

Lee S., Kirschning A., Muller M., Way C., Floss H.G. 1999. Enzymatic synthesis of [7-$^{14C, 7-3H}$]- and [1-13C]sedoheptulose 7-phosphate and [1-13C]*ido*-heptulose 7-phosphate. *J Mol Cat B: Enzymatic 6*: 369-377.

Leffer.J.E. & Grunwald.E. 1963. *Rates and Equilibria of Organic Reactions, First Edition Dover publications.*

Lehmann M., Kostrewa D., Wyss M., Brugger R., D'Arcy A., Pasamontes L., van Loon A. 2000. From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *PEDS 13*:49-57.

Lehmann M., Loch C., Middendorf A., Studer D., Lassen S., Pasamontes L., van Loon A., Wyss M. 2002. The consensus concept for thermostability engineering of proteins: further proof of concept. *PEDS 15*:403-411.

Loverix S., Geerlings P., McNaughton M., Augustyns K., Vandemeulebroucke A., Stevaert J., Versees W. 2005. Substrate-assisted leaving group activation in enzyme-catalyzed N-glycosidic bond cleavage. *J Biol Chem 280*:14799-14802.

Lutz S., Ostemeier M., Benkovic S. 2001. Rapid generation of incremental truncation libraries for protein engineering using á-phosphothioate nucleotides. *Nucleic Acids Research 29*:16e.

Lutz S., Ostemeier M., Moore G., Maranas C., Benkovic S. 2001. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc Natl Acad Sci USA 98*:11248-11253.

Lutz S., Patrick W. 2004. Novel methods for directed evolution of enzymes: quality, not quantity. *Current Opinion in Biotechnology 15*:291-297.

Lye G., Dalby P., Woodley J. 2002. Better biocatalytic processes faster: new tools for the implementation of biocatalysis in organic synthesis. Organic . *Proc Res Dev 6*:440.

Lye G., Ayazi-Shamlou P., Baganz F., Dalby P., Woodley J. 2003. Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends Biotechnol 21*:29-37.

Martineau P., Jones P., Winter G. 1998. Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J Mol Biol 280*:117-127.

May O., Nguyen P., Arnold F. 2000. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nat Biotechnology 18*:317-320.

Merz A., Yee M.C., Szadkowski H, Pappenberger G., Crameri A., Stemmer W.P., Yanofsky C., Kirschner K. 2000. Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry 39*:880-889.

Meshalkina L., Neef H., Tjaglo M., Schellenberger A., Kochetov G. 1995. The presence of a hydroxly group at the C-1 atom of the transketolase substrate molecule is necessary for the enzyme to perform the transferase reaction. *FEBS Lett*220-222.

Meyer A., Schmid A., Held M., Westphal A.H., Rothlisberger M., Kohler H.P.E., van Berkel W.J.H., Witholt B. 2002. Changing the substrate reactivity of 2-hydroxybiphenyl 3- monooxygenase from Pseudomonas azelaica HBP1 by directed evolution. *J Biol Chem 277*:5575-5582.

Miller J. 1992. A short course in bacterial genetics. *Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York, USA).*

Miller S., Orgel L. 1974. The origins of life on earth. *Prentice-Hall (Englewood Cliffs, New Jersey, USA).*

Miyazaki K., Arnold FH. 1999. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol 49*:716-720.

Mocali A., Paoletti F. 1989. Transketolase from human leukocytes. Isolation, properties and induction of polyclonal antibodies . *Eur J Biochem 180*:213-219.

Morley KL., Kazlauskas RJ. 2005. Improving enzyme properties: when are closer mutations better? *Trends Biotechnol 23*:231-237.

Mukherjee A., Svoronos S., Ghazanfari A., Martin P., Fisher A., Roecklein B., Rodbard D., Stanton R., Behar D., Berg C., Manjunath R. 1987. Transketolase abnormality in cultured fibroblasts from familial chronic alcoholic men and their male offspring. *J Clin Invest 79* :1039-1043.

Myles D.C., Andrulis P.J., Whitesides G.M.. 1991. A Transketolase-Based Synthesis of (+)-Exo-Brevicomin. *Tetrahedron Lett 32* :4835-4838.

Ness J., Kim S., Gottman A., Pak R., Krebber A., Borchert T., Govindarahan S., Mundorff E., Minshull J. 2002. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat Biotechnology 20*:1251-1255.

Nilsson U., Meshalkina L., Lindqvist Y., Schneider G. 1997. Examination of substrate binding in Thiamin Diphosphate-dependent Transketolase by protein crystallography and site-directed mutagenesis. *The Journal of Biological Chemistry 272*:1864-1869.

O'Maille P., Bakhtina M., Tsai M. 2002. Structure-based combinatorial protein engineering (SCOPE). *J Mol Biol 321*:677-691.

Orengo C., Thornton J. 2005. Protein families and their evolution-a structural perspective. *Annual Review of Biochemistry 74*:867-900.

Parikh M., Matsumura I. 2005. Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of b-fucosidase from b-galactosidase. *J Mol Biol 352*:621-628.

Park S., Morley K., Horsman G., Holmquist M., Hult K., Kazlauskas R. 2005. Focusing Mutations into the *P. fluorescens* Esterase Binding Site Increases Enantioselectivity More Effectively than Distant Mutations. *Chemistry & Biology 12*:45-54.

Peimbert M., Segovia L. 2003. Evolutionary engineering of a ß-Lactamase activity on a D-Ala D-Ala transpeptidase fold. *PEDS 16*:27-35.

Racker E. 1961. Transketolase. *The Enzymes, Academic Press (New York, USA).*

Rader C., Barbas C.I. 1997. Phage display of combinatorial antibody libraries. *Current Opinion in Biotechnology 8*:503-508.

Rais B., Comin B., Puigjaner J., Brandes J.L., Creppy E., Saboureau D., Ennamany R., Lee W.N., Boros L.G., Cascante M. 1999. Oxythiamine and dehydroepiandrosterone induce a G1 phase cycle arrest in Ehrlich's tumor cells through inhibition of the pentose cycle. *FEBS Lett 456*:113-118.

Rao B., Driver I., Lauffenburger D., Wittrup K. 2004. Interleukin 2 (IL-2) Variants Engineered for Increased IL-2 Receptor -Subunit Affinity Exhibit Increased Potency Arising from a Cell Surface Ligand Reservoir Effect. *Mol Pharmacology 66*:864-869.

Rao B., Driver I., Lauffenburger D., Wittrup K. 2005. High-Affinity CD25-Binding IL-2 Mutants Potently Stimulate Persistent T Cell Growth. *Biochemistry 44*:10696-10701.

Reetz M.T., Jaeger K.E. 1999. Superior biocatalysts by directed evolution. *Top Curr Chem 200*:31-57.

Reetz M., Bocola M., Carballeira J., Zha D., Vogel A. 2005. Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angewandte Chemie International Edition 44*:4192-4196.

Santoro S., Schultz P. 2002. Directed evolution of the site specificity of Cre recombinase. *PNAS 99*:4185-4190.

Sato K., Simon M., Levin A., Shokat K., Weiss G. 2004. Dissecting the engrailed homeodomain-DNA interaction by phage-displayed shotgun scanning. *Chemistry & Biology 11*:1017-1023.

Schenk G., Duggleby R., Nixon P. 1998. Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *Int J Biochem Cell Biol 30*:1297-1318.

Schmitzer A., Lepine F., Pelletier J. 2004. Combinatorial exploration of the catalytic site of a drug- resistant dihydrofolate reductase: creating alternative functional configurations. *PEDS 17*:809-819.

Seyden-Penne J. 1995. Chiral auxiliaries and ligands in asymmetric synthesis. *Wiley-Interscience.*

Shao Z., Zhao H., Giver L., Arnold F. 1997. Random-priming *in vitro* recombination: an effective tool for directed evolution. *Nucleic Acids Research 26*:681-683.

Shao Z., Zhao H., Giver L., Arnold F. 1998. Random-priming *in vitro* recombination: an effective tool for directed evolution. *Nucleic Acids Research 26*:681-683.

Sheari.J. The multi-enzymatic synthesis of xylulose 5-phosphate. Thesis 2006. University Of London. 2006.

Shinkai A., Patel P.H., Loeb L.A. 2001. The Conserved Active Site Motif A of *Escherichia coli* DNA Polymerase I Is Highly Mutable. *J Biol Chem 276*:18836-18842.

Sieber V., Martinez C., Arnold F. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat Biotechnology 19*:456-460.

Simpson F. 1960. Preparation and properties of transketolase from pork liver. *Can J Biochem Physiol 38*:115-124.

Smith.M.E.B, Smithies.K., Senussi.T., Dalby.P.A., Hailes.H.C. 2006. The First Mimetic of the Transketolase Reaction. *European Journal of Organic Chemistry 5*:1121-1123.

Sneeden J., Loeb L. 2003. Random oligonucleotide mutagenesis. *Methods Mol Biol 231*:65-73.

Solovjeva O.N., Kochetov G.A. 1999. Inhibition of transketolase by p-hydroxyphenylpyruvate. *FEBS Lett 462*:246-248.

Sprenger G.A. 1991. Cloning and preliminary characterization of the transketolase gene from *Escherichia coli* K12. *In: Bisswanger H , Ullrich J (eds) Biochemistry and physiology of thiamin diphosphate enzymes, VCH (Weinheim, Germany)*322-326.

Sprenger G.A., Schorken U., Sprenger G., Sahm H. 1995. Transketolase A of Escherichia coli K12. Purification and properties of the enzyme from recombinant strains. *Eur J Biochem 230*:525-532.

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

Sterzel R., Semar M., Lonergan E., Treser G., Lange K. 1971. Relationship of nervous tissue transketolase to the neuropathy in chronic uremia. *J Clin Invest 50*:2295-2304.

Stinson S. 2000. Chiral drugs. *Chem Eng News 78*:57-78.

Tawfik D., Griffiths A. 1998. "Man-made cell-like compartments for molecular evolution.". *Nat Biotechnology 16*:652-656.

Ting A., Kain K., Kleme R., Tsien R. 2001. Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. *PNAS 98*:15003-15008.

Trost B., Fleming I. 1991. Comprehensive organic synthesis. *Pergamon Press (New York, USA).*

Tuerk C., Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage-T4 DNA-polymerase. *Science 249*:505-510.

Udit A., Silberg J., Sieber V. 2003. Sequence homology-independent protein recombination (SHIPREC). *Methods Mol Biol 231*:163.

Voigt C., Mayo S., Arnold F., Wang Z. 2001. Computationally focusing the directed evolution of proteins. *Journal of Cellular Biochemistry Suplement 37*:58-63.

Voigt C., Mayo S., Arnold F.H., Wang Z. 2001. Computational method to reduce the search space for directed protein evolution. *PNAS 98*:3778-3783.

Wang C., Liao J. 2001. Alteration of product specificity of *Rhodobacter sphaeroides* phytoene desaturase by directed evolution. *J BiolChem 276*:41161-41164.

Weinreich D., Delaney N., DePristo M., Hartl D. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science 312*:111-114.

Winker C, Meshalkina L, Nilsson U, Backstrom S, Lindqvist Y and Schneider G. 1995. His103 is required for substrate recognition and catalysis. Eur. J. Biochem. 233, 750-755.

Williams.H. & Fleming.I. 1995. *Spectroscopic methods in organic chemistry* Fourth edition published by McGraw-hill education.

Zaccolo M., Williams D., Brown D., Gheradi E. 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleotide analogues. *J Mol Biol 255*:589-603.

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

Zhao H., Giver L., Shao Z., Affholter J, Arnold F. 1998. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat Biotechnology 16*:258-261.

Zhao H., Arnold F.H. 1997. Combinatorial protein design: strategies for screening protein libraries. *Curr Opin Struct Biol 7*:480-485.

Zhou Y., Zhang X., Ebright R. 1991. Random mutagenesis of gene-sized DNA molecules by use of PCR with *Taq* DNA polymerase. *Nucleic Acids Research 19:*6052.

Zimmerman F.T., Schneider A., Schorken U., Sprenger G.A., Fessner W.D. 1999. Efficient multi-enzymatic synthesis of D-xylulose 5-phosphate. *Tetrahedron: Asymmetry 10*:1643-1646.