REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author MILLER D.J.

Declaration

I hereby declare that my thesis entitled "Directed Evolution of Transketolase, a Carbon-Carbon Bond-Forming Enzyme" has not been submitted for a degree, diploma, or other qualification at any other university.

I am the sole author of this thesis and all data presented here is the result of my own work.

Oliver J. Miller

Directed Evolution of Transketolase, a Carbon-Carbon Bond-Forming Enzyme

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

Oliver J. Miller

Department of Biochemical Engineering
University College London

2005

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

Watching Doctor Gonzo leave

Raoul Duke: There he goes. One of God's own prototypes. Some kind of high-powered mutant never even considered for mass production. Too weird to live and too rare to die.

"Fear and Loathing in Las Vegas" (1998)

Acknowledgments

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

I am also grateful to all of the students that inhabit Foster Court: they made the workplace lively and extremely enjoyable. The HPLC tutelage generously provided to me by Christine Ingram is gratefully acknowledged. Thanks also to Richard Dunn of the School of Crystallography, Birkbeck College, for help with writing the Java program that made the consolidation of screening data an almost pleasant experience. Omar Qazi of the Centre for Molecular Microbiology and Infection, Imperial College, provided useful (and occasionally bizarre) insights that aided the development of this project.

I would also like to thank my parents Terry and Jane, my brother Toby, and my partner-in-crime Fay Newman for the support and encouragement that they given me over the past three years.

This project was financed by the UK Biotechnology and Biological Sciences Research Council (BBSRC).

Abstract

The enzyme transketolase (Enzyme Commission number: 2.2.1.1) has significant potential as a biocatalyst in the production of pharmaceuticals and fine chemicals. The enzyme catalyses the irreversible transfer of a C_2 (1,2-dihydroxyethyl) moiety from β-hydroxypyruvate (β-HPA) to a wide range of acceptor substrates in a stereospecific reaction. However, commercial application of transketolase is currently restricted by the limited availability and expense of β-HPA. This project describes efforts to generate and identify variants of *E. coli* transketolase that are capable of accepting pyruvate: a related, but much cheaper compound. The variants were prepared by a novel directed evolution technique, "focused" error-prone PCR (fepPCR), and then screened for the desired activity: pyruvate (donor) and glycolaldehyde (acceptor) to (*S*)-3,4-dihydroxybutan-2-one (and carbon dioxide).

The high-throughput screen consisted of the following steps: (1) transformation of a plasmid library into *E. coli* XL10-Gold competent cells; (2) culture of individual colonies in 384-well plates; (3) lysis of the cultures; (4) incubation of the lysates with cofactors and the target substrates; and finally (5) high-throughput HPLC analysis measuring donor substrate (pyruvate) depletion. The first four steps were optimised to ensure the highest possible concentration of holotransketolase in each screening reaction. The HPLC method utilised a 50mm guard column as the separation matrix and was capable of processing one sample every 1.2 minutes.

Several libraries of transketolase variants were generated using a novel mutagenesis technique: "focused" error-prone PCR (fepPCR). FepPCR uses knowledge of an enzyme as a map for targeting mutation to the most beneficial regions of the gene - in this case, three stretches of residues in the active site of *E. coli* transketolase (Ser24-His26, Gly99-Pro101, and Asp469-His473). Primers were designed to flank these small target sites (9-15bp) and they were PCR-amplified from the *tkt* gene under conditions that drastically lowered the fidelity of *Taq* DNA polymerase (0.14 misincorporations per nucleotide). The three mutated fepPCR products (50-54bp) were then cloned into the *tkt* vector pQR711 singly and in combination to create four distinct libraries. The PCR-based cloning techniques QuikChange Site-Directed Mutagenesis (Stratagene Ltd.) and QuikChange Multi Site-Directed Mutagenesis (Stratagene Ltd.) were both found to be effective for this step.

The four libraries were screened for the target activity and the following levels of coverage were achieved: ~100% of all possible single point mutations in the three targets sites and ~22% of all possible combinations of double point mutations in the target sites. Careful analysis of the HPLC chromatograms failed to identify any variants with the desired activity. It is proposed that larger libraries may yield positive results.

Abbreviations

σ_M	standard error of the mean
A_n	absorbance at *n*nm
Amp^+	denotes the inclusion of 150mg.l^{-1} ampicillin
amp^R	ampicillin resistance gene
AU	absorbance unit
bp	base-pair
BSA	bovine serum albumin
CLERY	combinatorial libraries enhanced by recombination in yeast
DNA	deoxyribonucleic acid
DNAP	DNA polymerase
dsDNA	double-stranded DNA
DXP	1-deoxy-D-xylulose-5-phosphate
EC	Enzyme Commission
EDTA	ethylenediaminetetraacetic acid
epPCR	error-prone PCR
FDA	Federal Drugs Agency
fepPCR	focused error-prone PCR
β-HPA	β-hydroxypyruvate
HPLC	high performance liquid chromatography
IGPS	indole-3-glycerol-phosphate synthase
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
KDPG	2-keto-3-deoxy-6-phosphoguconate
K_m	Michaelis constant
LB	Luria Bertani
MEGAWHOP	megaprimer PCR of whole plasmid
mpv	mutations per variant
M_r	relative molecular mass
MSDM	"multi" site-directed mutagenesis
MTEH	5-(2-methylthioethyl)-hydantoin
NADH	nicotinamide adenine dinucleotide, reduced form
OD_n	optical density at *n*nm
ODU	optical density unit
PCR	polymerase chain reaction
PDA	protein data automation
PDB	RCSB Protein Data Bank
pQR711A^μ	a library of pQR711 plasmids with mutations in target A
pQR711B^μ	as above with mutations in target B
pQR711C^μ	as above with mutations in target C

pQR711ABC$^{\mu}$	as above with mutations in targets A, B, and C
pQR711tkt$^{\mu}$	a library of pQR711 plasmids with mutations in the *tkt* gene
PRAI	phosphoribosylanthranilate isomerase
RACHITT	random chimeragenesis on transient templates
rpm	revolutions per minute
RPR	random-priming recombination
scFv	antibody single-chain Fv fragment
SCOPE	structure-based combinatorial protein engineering
SDM	site-directed mutagenesis
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SHIPREC	sequence homology-independent protein recombination
StEP	staggered extension process
TAE	Tris·acetate EDTA
TBE	Tris·borate EDTA
TEMED	N,N,N',N'-tetramethylethylendiamine
TFA	trifluoroacetic acid
ThDP	thiamine diphosphate
tkt	*Escherichia coli* transketolase gene
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane
UCL	University College London
UV	ultraviolet
V_{max}	maximal velocity

Note that an *E. coli* XL10-Gold pQR711ABC$^{\mu}$ colony with the label pQR711ABC$^{\mu}$-M5H24 would have been generated by QuikChange™ MSDM ("M") and stored in well H24 of pQR711ABC$^{\mu}$ master plate 5. Similarly, an *E. coli* XL10-Gold pQR711C$^{\mu}$ colony with the label pQR711C$^{\mu}$-S1J16 would have been generated by QuikChange™ SDM ("S") and stored in well J16 of pQR711C$^{\mu}$ master plate 1.

Table of contents

List of figures

List of tables

Chapter 1 - Introduction

1.1 Transketolase

Transketolase (EC 2.2.1.1) was first purified 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).

The ubiquitous transketolase is a thiamine diphosphate (ThDP) dependent enzyme that catalyses the transfer of a C_2 moiety (a 1,2-dihydroxyethyl group) between a ketose sugar and an aldose sugar (Figure 1.1). *In vitro*, transketolase is 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 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. In plants and

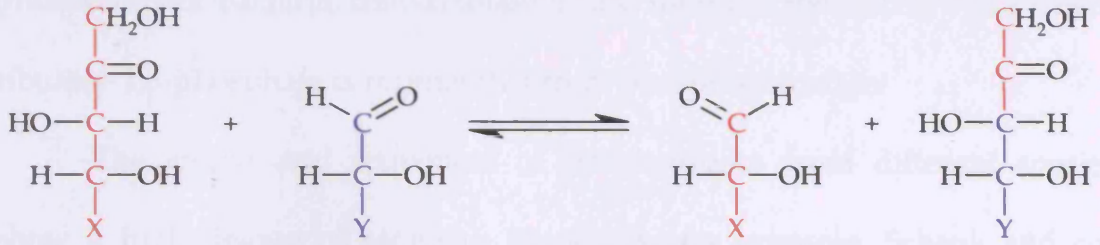


Figure 1.1. Scheme of the transketolase reaction shown in Fischer projection. A 1,2-dihydroxyethyl group is transferred between a ketose (red carbon skeleton) and an aldose (blue carbon skeleton). X and Y are variable (R groups).

```
Sc   1    MTQFTDIDKLAVSTIRILAVDTVSKANSGHPGAPLGMAPAAHVLWSQMRMN-PTNPDWIN  59
Ec   1    ----MSSRKELANAIRALSMDAVQKAKSGHPGAPMGMADIAEVLWRDFLKHNPQNPSWAD  56

Sc   60   RDRFVLSNGHAVALLYSMLHLTGYDLSIEDLKQFRQLGSRTPGHPE-FELP-GVEVTTGP  117
Ec   57   RDRFVLSNGHGSMLIYSLLHLTGYDLPMEELKNFRQLHSKTPGHPESGVTPLGVETTTGP  116

Sc   118  LGQGISNAVGMAMAQANLAATYNKPGFTLSDNYTYVFLGDGCLQEGISSEASSLAGHLKL  177
Ec   117  LGQGIANAVGMAIAEKTLAAQFNRPGHDIVDHYTYAFMGDGCMMEGISHEVCSLAGTLKL  176

Sc   178  GNLIAIYDDNKITIDGATSISFDEDVAKRYEAYGWEVLYVENGNEDLAGIAKAIAQAKLS  237
Ec   177  GKLIAFYDDNGISIDGHVEGWFTDDTAMRFEAYGWHVIRDIDGHDAASIKRAVEEARAVT  236

Sc   238  KDKPTLIKMTTTIGYGSLHAGSHSVHGAPLKADDVKQLKSKFGFNPDKSFVVPQEVYDHY  297
Ec   237  DKPSLLMCKTIIGFGSPNKAGTHDSHGAPLGDAEIALTREQLGW-KYAPFEIPSEIYAQW  295

Sc   298  QKTILKPGVEANNKWNKLFSEYQKKFPELGAELARRLSGQLPANWESKLPTYTAKDSA--  355
Ec   296  DAKEA--GQAKESAWNEKFAAYAKAYPQEAAEFTRRMKGEMPSDFDAKAKEFIAKLQANP  353

Sc   355  --VATRKLSETVLEDVYNQLPELIGGSADLTPSNLTRWKEALDFQPPSSGSGNYSGRYIR  413
Ec   354  AKIASRKASQNAIEAFGPLLPEFLGGSADLAPSNLTLWSGSK--AINEDA----AGNYIH  407

Sc   414  YGIREHAMGAIMNGISAFGANYKPYGGTFLNFVSYAAGAVRLSALSGHPVIWVATHDSIG  473
Ec   408  YGVREFGMTAIANGISLHGG-FLPYTSTFLMFVEYARNAVRMAALMKQRQVMVYTHDSIG  466

Sc   474  VGEDGPTHQPIETLAHFRSLPNIQVWRPADGNEVSAAYKNSLESKHTPSIIALSRQNLPQ  533
Ec   467  LGEDGPTHQPVEQVASLRVTPNMSTWRPCDQVESAVAWKYGVERQDGPTALILSRQNLAQ  526

Sc   534  LEGSS--IESASKGGYVLQDVAN-PDIILVATGSEVSLSVEAAKTLAAKNIKARVVSLPD  590
Ec   527  QERTEEQLANIARGGYVLKDCAGQPELIFIATGSEVELAVAAYEKLTAEGVKARVVSMSS  586

Sc   591  FFTFDKQPLEYRLSVLPDNVPIMSVE-VLATTCWGKYAH---QSFGIDRFGASGKAPEVF  646
Ec   587  TDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWYKYVGLNGAIVGMTTFGESAPAELLF  646

Sc   647  KFFGFTPEGVAERAQKTIAFYKGDKLISPLKKAF  680
Ec   647  EEFGFTVDNVVAKAKELL                  664
```

Figure 1.2. Amino acid sequences of *S. cerevisiae* ("Sc") and *E. coli* ("Ec") transketolases aligned using ClustalW. The 302 identical residues are shaded red and the 98 similar residues (determined by the BLOSUM62 matrix: Henikoff and Henikoff, 1992) are shaded blue. (Adapted from Schenk *et al.*, 1997.)

photosynthetic bacteria, transketolase is also found in the Calvin cycle where ribulose-1,5-phosphate is regenerated from phosphoglycerate.

The amino acid sequences of transketolases from different species show a high degree of sequence homology; for example, Schenk and co-workers (1997) aligned 22 sequences, 519–710 residues in length, and found that 50 residues were totally invariant. Figure 1.2 shows an alignment of the *S. cerevisiae* and *E. coli* transketolases that illustrates the high degree of homology. The majority of transketolases that have been studied are homodimeric with subunit molecular weights ranging from 67 to 75 kDa.

The enzyme from *S. cerevisiae* was the first to yield to a crystallographic structural analysis. The crystal structure, initially determined at 2.5 Å resolution (Lindqvist *et al.*, 1992) and subsequently extended to 2.0 Å resolution (Nikkola *et al.*, 1994), revealed the first insights into the catalytic mechanism of transketolase. *E. coli* transketolase has also been crystallised (Littlechild *et al.*, 1995) and the crystal structure has been solved at 1.9 Å resolution (RCSB Protein Data Bank structure file 1QGD) (Figure 1.3).

Homologous expression systems for transketolase have been established in *S. cerevisiae* (Sundström *et al.*, 1993) and *E. coli* (Draths and Frost, 1990; French and Ward, 1995). Bacterial heterologous expression systems have been created for human (Singleton *et al.*, 1996) and maize (Gerhardt *et al.*, 2003) transketolases.

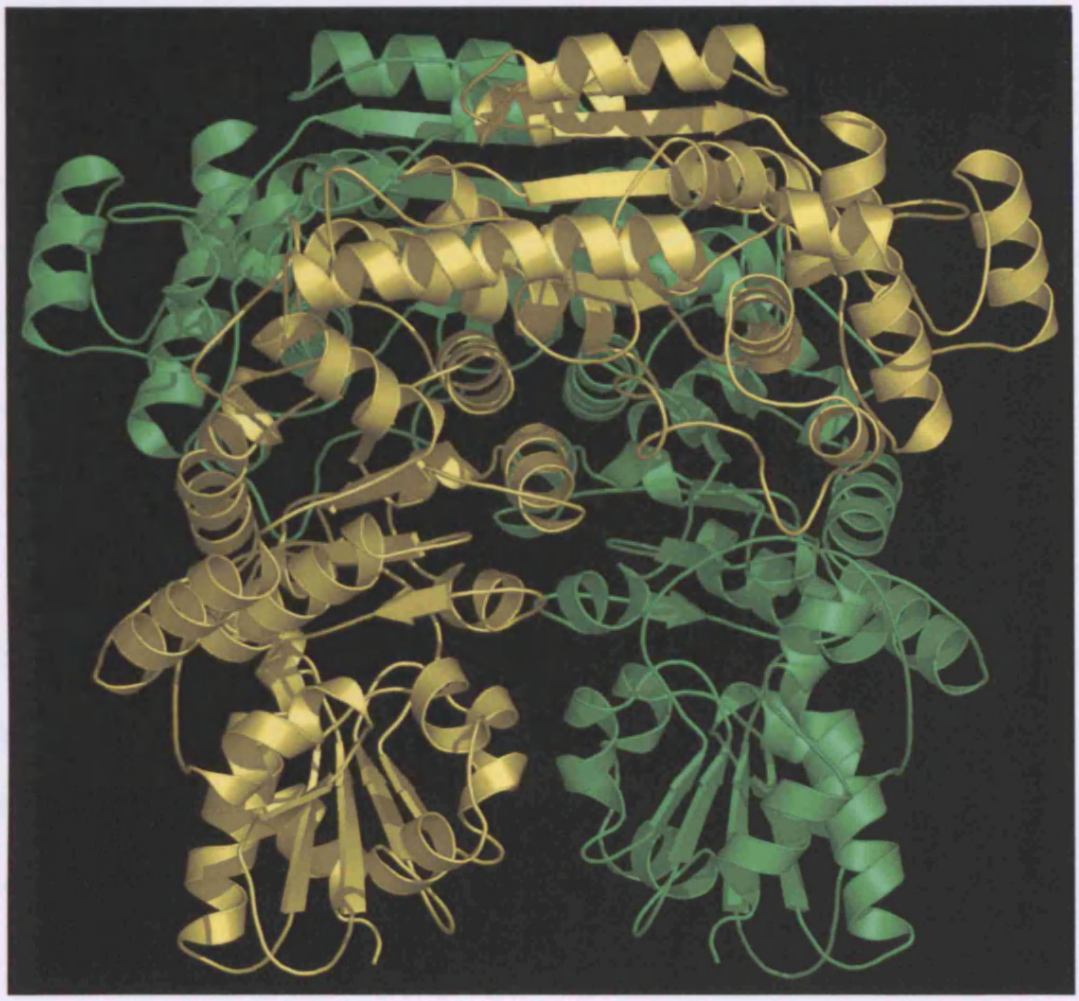


Figure 1.3. Ribbon diagram of the *E. coli* transketolase homodimer. One monomer is coloured yellow and the other is coloured green. (Created from the PDB structure file 1QGD.)

1.1.1 Structure

The following discussion is based on articles published by Schneider and Lindqvist (Schneider and Lindqvist, 1998 and references therein) and concerns the structure of transketolase from *S. cerevisiae*. The PDB structure file of *E. coli* transketolase (1QGD) permitted the counterparts of important residues to be identified.

Each subunit of transketolase comprises three domains (Figure 1.4). The amino-terminal, or PP-domain, consists of a five-stranded parallel β-sheet with α-helices on both sides. This domain is 320 amino acids in *S. cerevisiae*, which corresponds to a 315 amino acid stretch in *E. coli*. The second domain, the Pyr-domain, is a six-stranded parallel β-sheet, sandwiched between α-helices. Residues 323–538 constitute this domain in *S. cerevisiae* and residues 318–529 in *E. coli*. Both domains are involved in binding ThDP and their topology has been denoted the "ThDP binding fold" (Müller *et al.*, 1993). The carboxy-terminal domain contains a mixed β-sheet with four parallel and one antiparallel strand. Residues 539–680 comprise this domain in *S. cerevisiae* and residues 530–663 in *E. coli*. The C-domain does not contribute any amino acids to the active site and its function remains unclear.

The transketolase dimer is formed through tight interactions between the PP- and Pyr-domains (Figure 1.4). The C-domain makes few contacts with the other subunit.

Experiments with transketolase from *S. cerevisiae* (Heinrich *et al.*, 1972) and *E. coli* (Sprenger *et al.*, 1995) have demonstrated the necessity of divalent

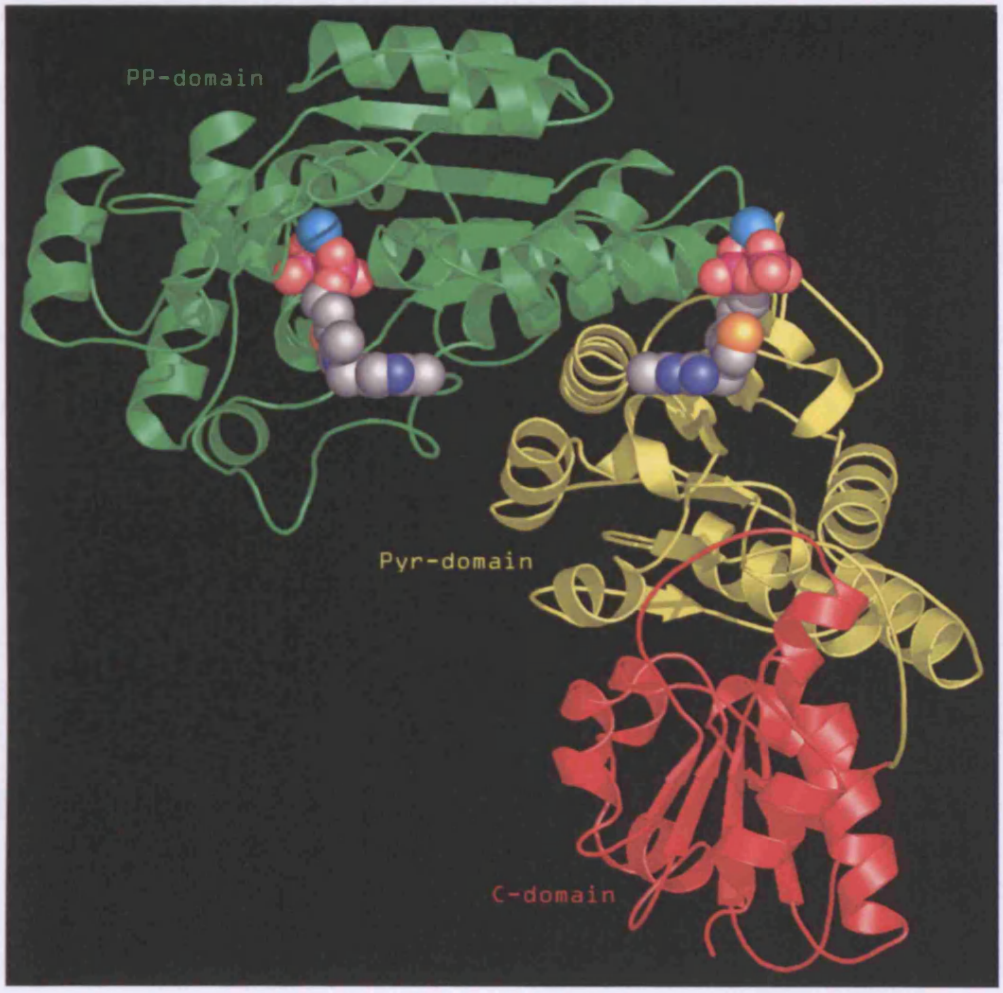


Figure 1.4. Ribbon diagram of a single subunit of *E. coli* transketolase. Two ThDP molecules, which bind to the PP- and Pyr-domains, are shown as CPK space-filling models. The divalent metal ion required for catalytic activity, in this case Ca^{2+}, is highlighted in cyan. (Created from the PDB structure file 1QGD.)

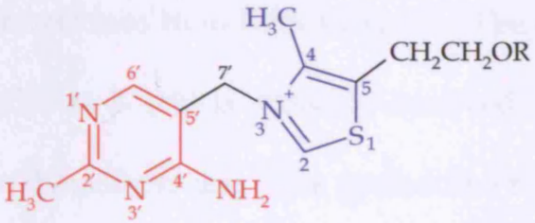


Figure 1.5. Structure of thiamine and ThDP. The methylaminopyrimidine ring (coloured red) is linked via a methylene bridge to the methylthiazolium ring (coloured blue). R is a hydrogen atom in thiamine and a diphosphate moiety ($P_2O_6^{3-}$) in ThDP.

metal ions for the formation of the active holoenzyme. A cation binds to each subunit through the main chain oxygen of Ile189 and by the side chains of Asp157 and Asn187 (Ile189, Asp155, and Asn185 in *E. coli*). The reconstitution rates of holoenzyme from apoenzyme vary depending on the nature of the cation, but final catalytic activity is not affected. For *S. cerevisae* transketolase the rates descend in the order $Ca^{2+} > Mn^{2+} > Co^{2+} > Mg^{2+} > Ni^{2+}$ (Heinrich *et al.*, 1972) and for the *E. coli* enzyme they descend in the order $Mn^{2+} > (Mg^{2+}$ and $Ca^{2+}) > (Co^{2+}$ and $Zn^{2+}) > Ni^{2+}$ (Sprenger *et al.*, 1995).

ThDP binds to transketolase through several interactions (Figures 1.4 and 1.5). Its diphosphate group is bound at the switchpoint between neighbouring strands 1 and 3 at the amino-terminus of a small α-helix in the PP-domain of one subunit. The diphosphate group interacts with the protein in two ways. Firstly, there are direct interactions through hydrogen bonds to His69, His263, and Gly158 (His66, His261, and Gly156 in *E. coli*). Secondly, there are indirect interactions through the bound divalent metal ion. The thiazolium ring of ThDP sits in the cleft between the PP-domain of one subunit and the Pyr-domain of the second subunit and interacts hydrophobically with residues from both subunits. The conserved aspartate residue Asp382 (Asp381 in *E. coli*) is probably involved in compensating the positive charge of the thiazolium ring. The pyrimidine ring of ThDP binds in a hydrophobic pocket, formed by residues from the Pyr-domain of the second subunit. The pocket is made up of a number of aromatic side chains belonging to Phe442, Phe445, and Tyr448 (Phe434, Phe437, and Tyr440 in *E.*

coli). A hydrogen bond between the N1′ atom of the pyrimidine ring and the side chain of a conserved glutamate residue (Glu418 in *S. cerevisiae* and Glu411 in *E. coli*) is significant since it is observed in all ThDP-dependent enzymes.

An interesting feature of the enzyme is a large solvent-filled channel which spans the region between the two ThDP molecules. This channel contains several conserved glutamate residues: Glu162, Glu167, and Glu418 (Glu160, Glu165, and Glu411 in *E. coli*). A hydrogen bonding network through these residues connects the N1′ atom of the pyrimidine ring of one ThDP to the corresponding atom of the ThDP molecule in the other subunit. The purpose of this conserved network is not clear, but it could be to shuffle protons between the two active sites. However, no cooperative behaviour between the active sites during catalysis has yet been observed.

1.1.2 Catalysis

1.1.2.1 Substrate binding and recognition

A crystallographic study of a complex of *S. cerevisiae* transketolase with the acceptor substrate D-erythrose-4-phosphate identified several residues in the substrate channel that may be involved in substrate binding (Nilsson *et al.*, 1997). Table 1.1 lists these residues and their proposed roles.

The structure of the active site suggests that the invariant active site residues His69 and His103 (His66 and His100 in *E. coli*) bind the C1-hydroxyl group of the donor substrate and stabilise the reaction intermediate (Wikner

Residue number		Residue identity	Function
S. cerevisiae	*E. coli*		
30	26	His	Catalysis and stereospecificity
69	66	His	Substrate recognition and binding
103	100	His	Substrate recognition and binding
263	261	His	Catalysis and stereospecificity
359	358	Arg	Phosphate binding
386	385	Ser	Phosphate binding
418	411	Glu	Catalysis
469	461	His	Phosphate binding
477	469	Asp	Stereospecificity
481	473	His	Transition state stabilisation
528	520	Arg	Phosphate binding

Table 1.1. Residues that are proposed to be involved in substrate binding and/or catalysis in transketolase.

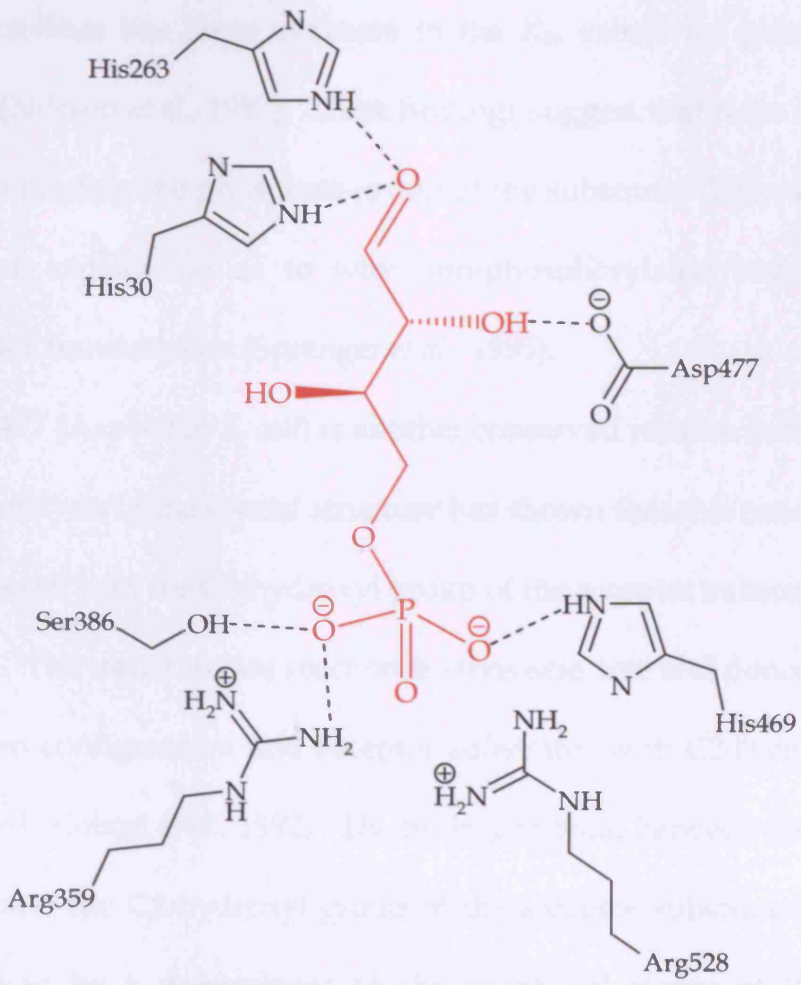


Figure 1.6. Active site of *S. cerevisiae* transketolase with D-erythrose 4-phosphate (coloured red). (Adapted from Hecquet *et al.*, 2001.)

et al., 1995; Wikner *et al.*, 1997). β-hydroxypyruvate (β-HPA) is a donor substrate for transketolase and it has been widely used as the donor in biotransformations (refer to Section 1.1.3). However, pyruvate is not a substrate for transketolase and the recognition of the C1-hydroxyl group of the donor by the side chains of these two histidine residues may be the molecular basis for this discrimination in substrate binding.

Residues Arg359, Arg528, and His469 (Arg358, Arg520, and His469 in *E. coli*) are close to the phosphate group of the substrate. Replacement of any of these residues with alanine yields mutant enzymes with considerable residual activities but large increases in the K_m values for phosphorylated substrates (Nilsson *et al.*, 1997). These findings suggest that these residues are involved in binding the phosphate group of the substrate. They also provide a molecular explanation as to why non-phosphorylated sugars are bad substrates for transketolase (Sprenger *et al.*, 1995).

Asp477 (Asp469 in *E. coli*) is another conserved residue in the substrate channel. Analysis of the crystal structure has shown that this residue forms a hydrogen bond with the C2-hydroxyl group of the acceptor substrate (Nilsson *et al.*, 1997). The transketolase reaction is stereoselective and donor substrates with D-threo configuration and acceptor substrates with C2-D configuration are preferred (Kobori *et al.*, 1992). The hydrogen bond between the side chain of Asp477 and the C2-hydroxyl group of the acceptor substrate is therefore very likely to be a determinant of the enantioselectivity of the enzyme (Nilsson *et al.*, 1998).

1.1.2.2 Catalytic mechanism

The first half of the reaction is the cleavage of the donor substrate and formation of the first product, an aldose, and a covalent intermediate, the 2-α-carbanion of 1,2-dihydroxyethyl ThDP. In the second half of the reaction, a nucleophilic attack by the 2-α-carbanion on the acceptor substrate occurs and the second product, a ketose with its carbon skeleton extended by two carbon atoms, is formed. This reaction mechanism, proposed by Schneider and Lindqvist (1993), is illustrated in Figure 1.7.

Glu418 (Glu411 in *E. coli*) is sufficiently close to form a hydrogen bond with the N1′ atom of the pyrimidine ring of ThDP. It is proposed that this hydrogen bond promotes the generation of a resonance form of ThDP with a positively-charged imino-group at the 4′-position of the pyrimidine ring. The side chain of His481 (His473 in *E. coli*) is close to this charged imino group and could conceivably abstract a proton. This would increase the pK_a of the group sufficiently that it would, in turn, remove a proton from C2 of the thiazolium ring. Once formed, the C2-carbanion of the cofactor is able to attack the carbonyl carbon of the donor substrate. For covalent bond formation between these two atoms to occur, stabilisation of the negative charge developing at the carbonyl oxygen is required. His481 (His473 in *E. coli*) and the charged 4′-imino group of ThDP are the most likely candidates for this role. Subsequently, the abstraction of a proton from the C3-hydroxyl group of the substrate, possibly mediated by the concerted action of His30 and His263 (His26 and His261 in *E. coli*), leads to cleavage of the addition

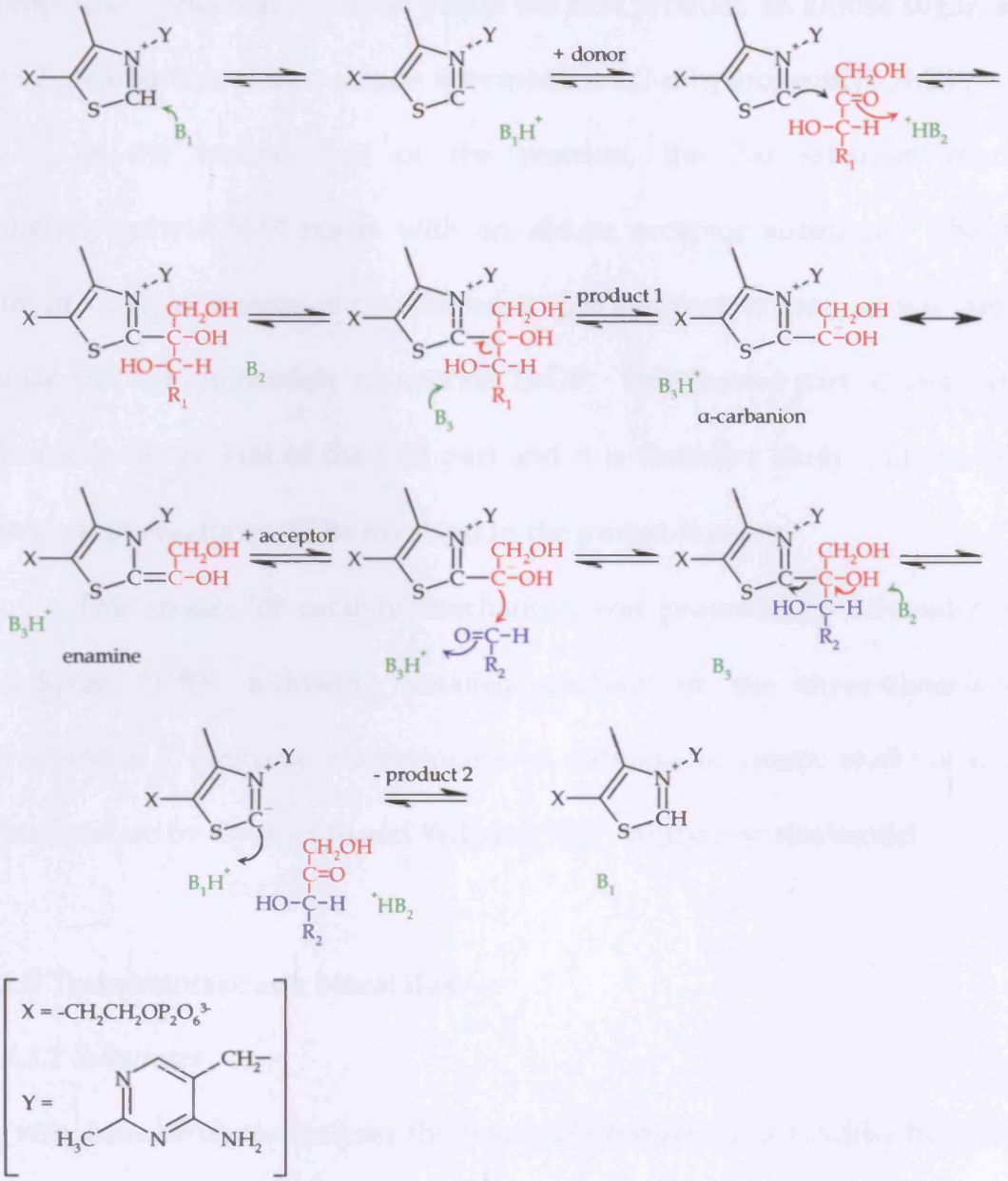


Figure 1.7. Reaction mechanism of transketolase. The donor substrate (a ketose) and the acceptor substrate (an aldose) are coloured red and blue respectively. B_1 is most likely to be the 4'-imino group of the ThDP pyrimidine ring. This group is also a candidate for B_2; His481 (His473 in *E. coli*) is another possibility. His30 and His263 (His26 and His261 in *E. coli*) are the two candidates for B_3. (Adapted from Schneider and Lindqvist, 1998.)

compound. This manipulation yields the first product, an aldose sugar, and the 2-α-carbanion of the reaction intermediate 1,2-dihydroxyethyl-ThDP.

In the second half of the reaction, the 2-α-carbanion of 1,2-dihydroxyethyl-ThDP reacts with an aldose acceptor substrate. The 1,2-dihydroxyethyl moiety is transferred to the acceptor to form a new ketose sugar and simultaneously regenerate ThDP. This second part of catalysis is essentially a reversal of the first part and it is therefore likely that the same amino acid residues will be involved in the proton transfer.

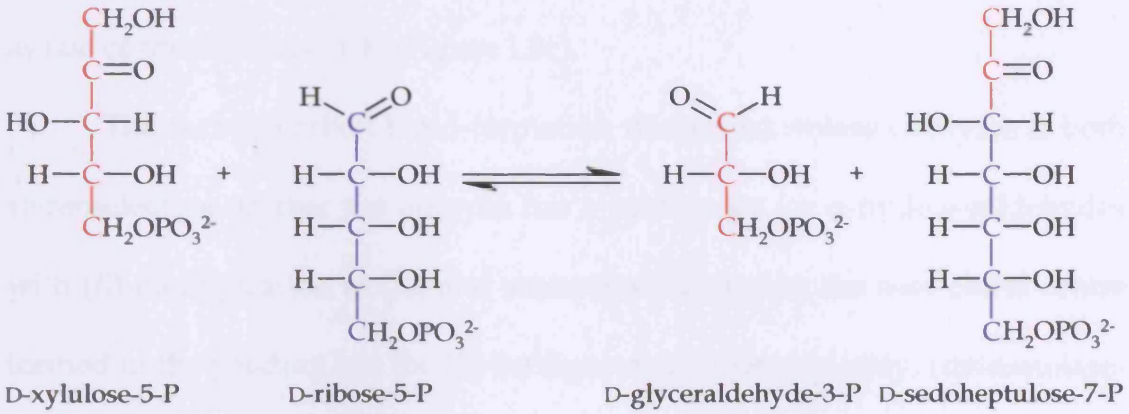
This model for catalytic mechanism was proposed by Schneider and Lindqvist (1993) following detailed analysis of the three-dimensional structure of *S. cerevisiae* transketolase. A subsequent kinetic study of *E. coli* transketolase by Gyamerah and Willetts (1997) supported this model.

1.1.3 Transketolase as a biocatalyst

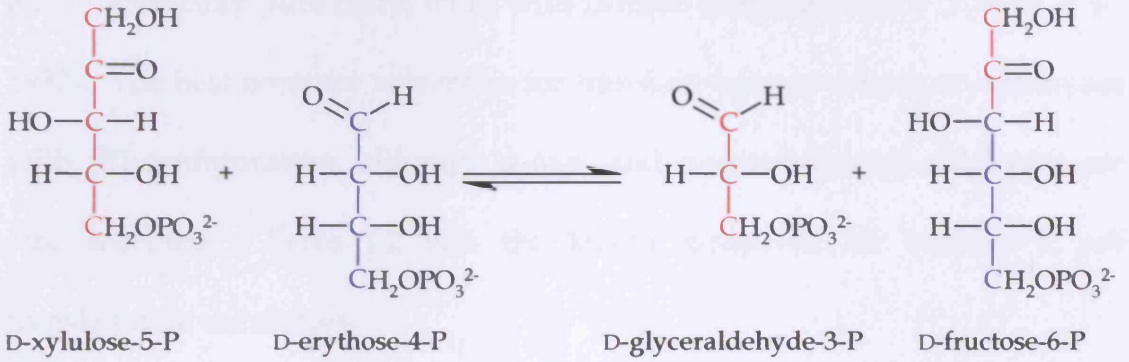
1.1.3.1 Substrates

In vivo, transketolase catalyses the reversible transfer of a 1,2-dihydroxyethyl group from D-xylulose-5-phosphate to D-ribose-5-phosphate to yield D-sedulose-7-phosphate and D-glyceraldehyde-3-phosphate (Figure 1.8a). 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 (Figure 1.8b). In experimental and preparative condensations β-HPA frequently replaces D-xylulose-5-phopshate as the donor substrate. The use of β-HPA renders the reaction

(a)



(b)



(c)

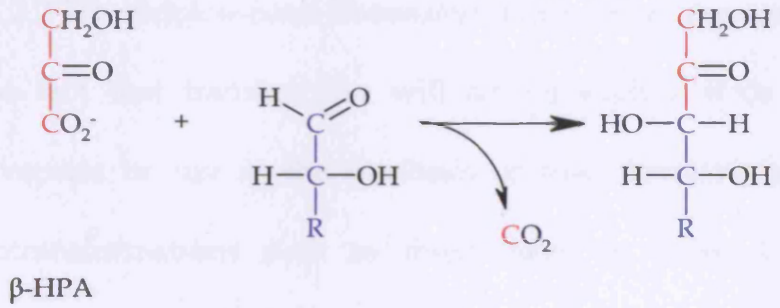


Figure 1.8. *In vivo* and *in vitro* reactions of transketolase shown in Fischer projection. Reactions (a) and (b) occur *in vivo* and are reversible. β-HPA is used as the donor substrate in biotransformations to make the reaction irreversible. If the R group is a hydrogen atom, then the reaction is β-HPA and glycolaldehyde to L-erythrulose (and carbon dioxide).

effectively irreversible by virtue of the concomitant release of carbon dioxide as one of the two products (Figure 1.8c).

The carbon-carbon bond-formation that transketolase 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, transketolase-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. Table 1.2 lists the kinetic constants for various *E. coli* transketolase substrates.

1.1.3.2 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 such as those listed in Table 1.3 have several key advantages over traditional chemistry-based routes to the same compounds. These advantages are summarised in Table 1.4.

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

Substrate	K_m (M)	V_{max} (U.mg^{-1})
Acceptor components		
β-Hydroxyaldehyde phosphates		
D,L-Glyceraldehyde-3-phosphate	2.1×10^{-3}	100
D-Erythrose-4-phosphate	90.0×10^{-6}	≥110
D-Ribose-5-phosphate	1.4×10^{-3}	50.4
β-Hydroxyaldehydes		
Glycolaldehyde	14.0×10^{-3}	60
D,L-Glyceraldehyde	10.0×10^{-3}	25
D-Erythrose	150.0×10^{-3}	75
D-Ribose	1.4	25
Aldehydes		
Formaldehyde	31.0×10^{-3}	12.5
Acetaldehyde	1.2	20
Donor components		
β-HPA	18.0×10^{-3}	60
Xylulose-5-phosphate	160.0×10^{-6}	≥110
Fructose-6-phosphate	1.1×10^{-3}	62
Sedoheptulose-7-phosphate	4.0×10^{-3}	6

Table 1.2. Kinetic constants for *E. coli* transketolase. The K_m values for D-ribose, D-ribose-5-phosphate, and sedoheptulose-7-phosphate are apparent constants, because only a small part of each compound exists in an open form, which can serve as a transketolase substrate. (Reproduced from Sprenger *et al.*, 1995.)

Substrate		Product	Reference
Donor	Acceptor		
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 1.3. A selection of fine chemicals and drug precursors that have been synthesised using transketolase.

Advantages
Mild operating conditions
Can be highly chemo-, regio-, and stereoselective
Potential to perform complex reactions with no equivalent in traditional chemistry
Potential for low environmental impact
Aqueous reaction medium reduces need for organic solvents
Potential for fewer reaction by-products
Potential avoidance of protection and de-protection steps
Disadvantages
Biocatalyst may show poor operational stability
Low volumetric productivities are common
Biocatalyst production may be expensive
Absolute substrate specificity may make the process inflexible
Bulk removal of water may be a problem in downsteam processing

Table 1.4. Summary of the advantages and disadvantages of biotransformations when compared to traditional chemical processes. (Adapted from Cheetham, 1994.)

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.

1.1.3.3 Overexpression of transketolase

Early work on transketolase-based biotransformations used transketolase protein extracted from two sources: *S. cerevisiae* and spinach. The yields from these sources were relatively poor, so high-yielding recombinant expression systems were developed. Initially, Draths and Frost (1990) cloned the *E. coli* transketolase gene (*tkt*, 1991bp) on a 5kb fragment into the low copy number

vector pBR325. French and Ward (1995) subsequently improved expression levels by subcloning the *tkt* gene from this construct into a variety of high copy-number expression vectors. Host cells containing these new constructs expressed sufficient recombinant *E. coli* transketolase to permit large-scale biotransformations. One of these strains, *E. coli* JM107 pQR711 (Figure 1.9), was grown to 20g dry cell weight per litre, producing 4kg of enzyme from a 1000 litre glycerol-fed fermentation (Hobbs *et al.*, 1996).

1.1.3.4 Reaction conditions

The principal transketolase-based biotransformation studied by researchers at UCL is the synthesis of L-erythrulose from β-HPA and glycolaldehyde (Hobbs *et al.*, 1993; Hobbs *et al.*, 1996; Mitra and Woodley, 1996; Woodley *et al.*, 1996; Chauhan *et al.*, 1997; Mitra *et al.*, 1998; Brocklebank *et al.*, 1999). Standard reaction conditions are used whenever possible to permit the comparison of data generated by different researchers (Table 1.5). The biocatalyst itself is usually derived from *E. coli* JM107 pQR711.

The concentrations of cofactors used in transketolase-based biotransformations vary from laboratory to laboratory. Generally speaking though, these concentrations are always high enough to ensure an excess of both cofactors in the reaction. The cofactors are incubated with the enzyme for a period of time prior to the addition of substrates to permit reconstitution of the holoenzyme. This period of time, "preincubation", varies between 3 and 30 minutes, depending on the laboratory.

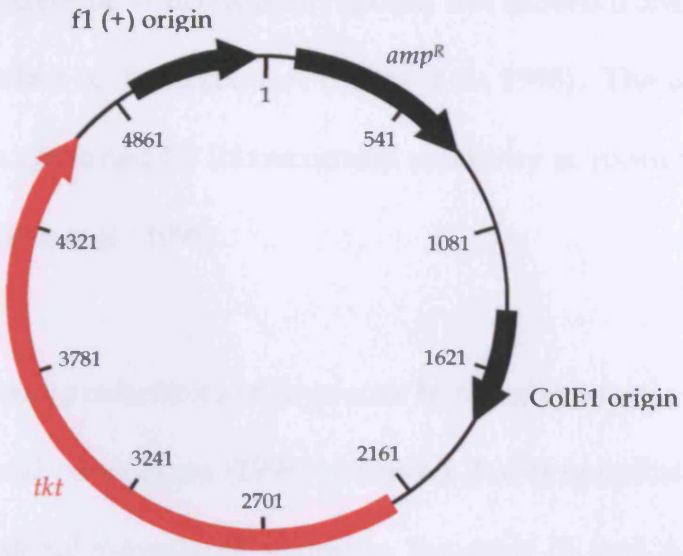


Figure 1.9. Map of the *E. coli* transketolase overexpression vector pQR711 (French and Ward, 1995). The orientation of each gene is represented by the direction of its arrow. amp^R is the ampicillin resistance gene and *tkt* is the transketolase gene.

Variable	Optimum	Literature	UCL standard
pH	7.0–7.5		7.5
Preincubation period		3–30min	15min
Temperature	20–40°C		25°C
Cofactor concentrations			
Mg^{2+}		0.9–10mM	9mM
ThDP		0.2–2.5mM	2.5mM
Substrate concentrations			
β-HPA (donor)	<0.6M		100mM
Glycolaldehyde (acceptor)	<0.5M		100mM

Table 1.5. Reaction conditions for *E. coli* transketolase-based bioconversion of β-HPA and glycolaldehyde to L-erythrulose (and carbon dioxide). "Optimum" ranges are for maximum transketolase activity (Sprenger *et al.*, 1995; Mitra *et al.*, 1998). "Literature" ranges are based on the conditions used by various laboratories around the world. "UCL standard" values are the conditions typically used by researchers at UCL.

Glycolaldehyde concentration should not exceed 0.5M because it has a destabilising effect on transketolase (Mitra *et al.*, 1998). The upper limit for β-HPA (0.6M) is governed by its maximum solubility at room temperature and neutral pH (Mitra *et al.*, 1998).

1.1.3.5 Maximising productivity in large-scale biotransformations

Brocklebank and co-workers (1998) observed that immobilising transketolase on the commercially-available supports Eupergit C and Amberlite XAD-7 increased the enzyme's half-life during biotransformations by 80–100-fold. Immobilisation was found to stabilise the enzyme against the denaturing effects of the aldehyde acceptor substrate (glycolaldehyde).

In an effort to reduce product inhibition, Chauhan and co-workers (1997) explored the use of *in situ* product removal (ISPR). Immobilised boronate resins were found to effectively sequester product as it formed in the reactor, thus preventing the build-up of inhibiting concentrations. The authors note that this method could be used to recover any transketolase product, providing that it is derived from an α-hydroxylated aldehyde acceptor substrate.

Woodley and co-workers (1996) suggested that substrate inhibition and product inhibition could be tackled simultaneously by prudent reactor selection. A fed-batch stirred tank reactor is recommended for transketolase-based biotransformations as it permits the maintenance of low aldehyde acceptor concentration (by feeding) and low product concentration (by ISPR).

1.2 Directed evolution

The excellent advantages that biotransformations have over traditional chemistry have already been discussed (refer to Section 1.1.3.2). However, the application of wild-type enzymes as synthetic agents is limited by factors such as substrate specificity, instability, and low activity in non-aqueous solvents (Cheetham, 1994). Being a product of millions of years of evolution, enzymes did not evolve to be efficient in every situation a process chemist might like.

A way around this problem is "rational protein engineering" by site-directed mutagenesis. Specific amino acids are substituted for different ones at defined positions in the peptide chain of an enzyme with the aim of improving activity or selectivity. Successful implementations of this approach have, however, been few and far between (for examples see: Cedrone *et al.*, 2000; Hult and Berglund, 2003). Attempts have, perhaps, relied too heavily on static protein structure models and transition-state theory (Dalby, 2003). Ignoring the contributions that enzyme dynamics and quantum tunnelling make to catalysis may make the design process easier, but it will almost certainly fail to deliver many positive results (Sutcliffe and Scrutton, 2000; Sutcliffe and Scrutton, 2002).

Until enzyme mechanics are better understood, an alternative approach presents itself for the modification of biocatalysts. "Directed evolution" contrasts with rational protein design in that it does not require a complete, or indeed any, understanding of how an enzyme works (Shao and

Arnold, 1996). It is an algorithm that mimics natural evolution with successive cycles of mutation and selection. A set of mutant genes, a "library", is created by either random mutagenesis of the gene encoding the enzyme or recombination of gene fragments. The mutant genes are then inserted into an appropriate microorganism and the mutant enzymes, "variants", are expressed individually. The optimal mutant is isolated and used as the basis for another cycle of mutagenesis and selection. Further cycles create an evolutionary pressure, leading to the formation and identification of superior biocatalysts at every generation. A typical directed evolution experiment is shown in Figure 1.10.

1.2.1 Methods of mutagenesis

There are two general approaches to creating libraries for directed evolution: (a) non-recombinative mutagenesis methods, such as error-prone PCR (epPCR) (Zhou *et al.*, 1991); and (b) recombinative mutagenesis methods, such as DNA-shuffling (Stemmer, 1994). In non-recombinative mutagenesis, diversity is generated by random point mutation. The majority of variants are typically less fit than the parental gene. The top variant is identified by screening or selection and used as the basis for a new cycle of mutation and selection. In recombinative mutagenesis, full-length chimerical genes are generated from a single gene or a family of genes and the best variants are pooled and recombined.

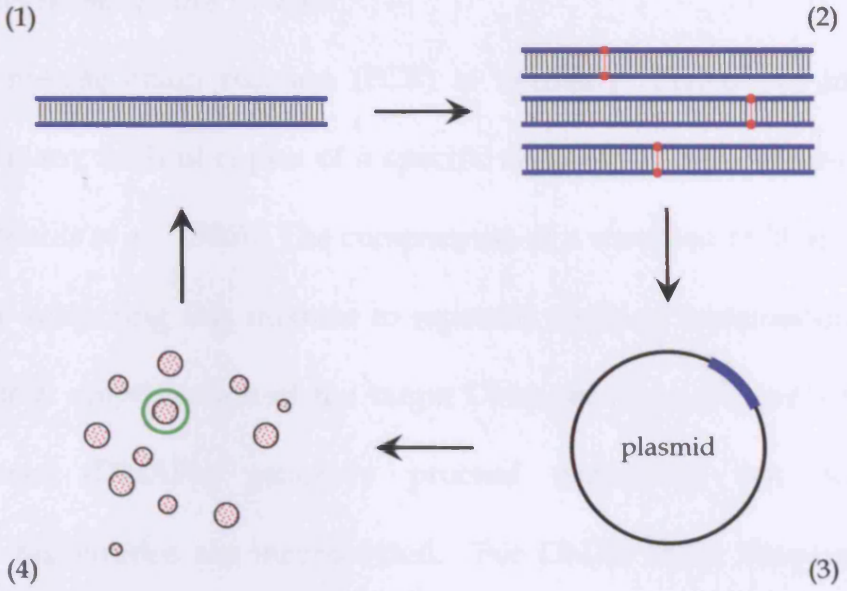


Figure 1.10. Typical steps in the directed evolution of an enzyme. (1) A gene is selected. (2) Mutation: a library of mutant genes is created (mutations are shown as red dots). (3) The gene library is inserted into an expression vector (coloured blue) to express them. (4) Selection: the colonies are screened for improvement in a particular enzyme characteristic and the optimal variant (circled in green) is isolated to repeat the entire process.

Reagent	Volume	Final conc.
ddH_2O	variable	
10× PCR buffer pH 8.3	10µl	
Tris-HCl		10mM
$MgCl_2$		1.5mM
KCl		50mM
dNTPs	2µl	
dATP		0.2mM
dCTP		0.2mM
dGTP		0.2mM
dTTP		0.2mM
Primer 1	0.2–2µl	0.1–1.0µM
Primer 2	0.2–2µl	0.1–1.0µM
Template DNA	variable	<1µg
Taq DNAP	0.5µl	2.5U
Total	100µl	

Table 1.6. The components of a typical PCR. The stock solution of dNTPs is usually a mixture of 10mM dATP, 10mM dCTP, 10mM dGTP, and 10mM dTTP. Primers are usually stored at 50µM concentration. (Adapted from Roche Diagnostics Ltd., 2000.)

1.2.1.1 Non-recombinative methods

The polymerase chain reaction (PCR) is normally carried out in order to generate many faithful copies of a specific segment of deoxyribonucleic acid (DNA) (Mullis *et al.*, 1986). The components of a standard PCR are shown in Table 1.6; subjecting this mixture to repeated cycles of temperature changes brings about amplification of the target DNA sequence (Figure 1.11). DNA polymerases (DNAPs) generally proceed accurately, but occasionally incorrect nucleotides are incorporated. For DNAP from *Thermus aquaticus* (*Taq* DNAP), the frequency of misincorporations is 1.1×10^{-4} per nucleotide (Tindall and Kunkel, 1988). This very small error rate can be increased for epPCR by: (a) increasing the concentration of Mg^{2+}; (b) adding Mn^{2+}; (c) increasing the concentrations of dCTP and dTTP; or (d) increasing the amount of *Taq* DNAP (Cadwell *et al.*, 1995). Following epPCR the mutant genes are cloned into a suitable vector by restriction and ligation.

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^{-7} per nucleotide 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 5×10^{-4} mutations

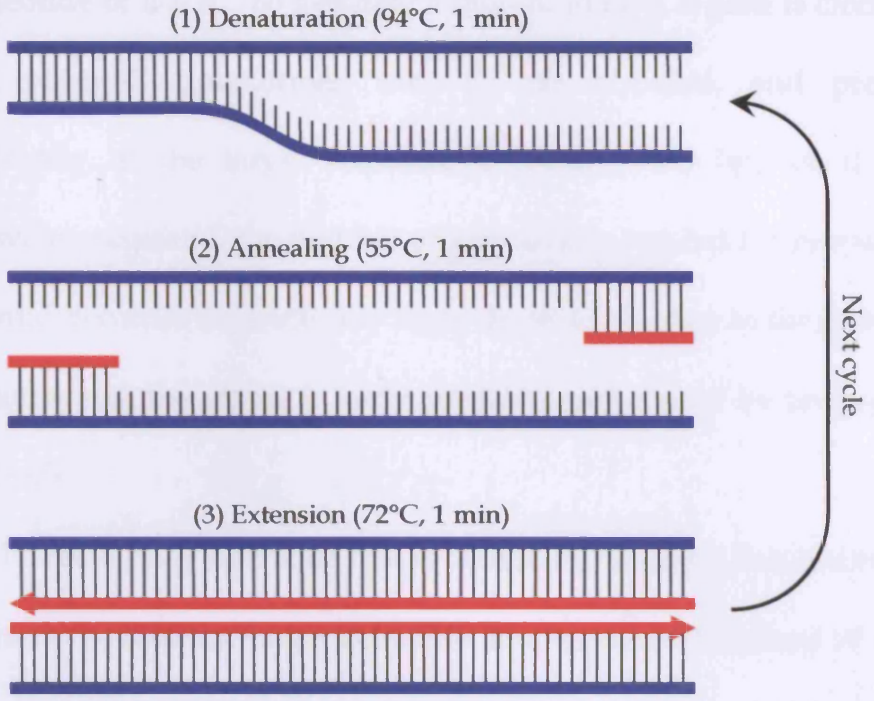


Figure 1.11. The mechanism of PCR. (1) Denaturation (94°C): double-stranded DNA melts open to form single-stranded DNA. (2) Annealing (55°C): a reduction in temperature allows stable bonds to form between the oligonulcleotide primers (coloured red) and complementary bases of the template DNA (coloured blue). (3) Extension (72°C): DNAP extends the primers according to the sequence of the template DNA. The entire cycle is then repeated by denaturing the two copies of the target sequence and starting again. In this way the number of copies of the target sequence grows exponentially. (Mullis *et al.*, 1986.)

per nucleotide 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.

The remaining technique that is commonly used for non-recombinative mutagenesis is random oligonucleotide mutagenesis (Oliphant *et al.*, 1986; Sneeden and Loeb, 2003). Oligonucleotides that encode a specific region of a plasmid-borne gene are synthesised with certain bases randomised either fully (A/C/G/T) or partially (e.g. C/G). These oligonucleotides are amplified by minimal cycles of PCR to generate double-stranded fragments which are then cloned into plasmids by restriction and ligation. The libraries generated by this approach can be very large: for example, randomising just four residues in a protein results in a library size of 1.6×10^5 unique protein sequences.[1] Such a library would require a large amount of time to screen.

1.2.1.2 Recombinative methods

One of the oldest hypotheses for the advantage of sex and recombination is that it allows beneficial mutations, initially carried by different individuals, to be combined together into the same individual (Fisher, 1930). In other words,

[1] Every possible amino acid is accessed at each of the four positions so there are $20^4 = 1.6 \times 10^5$ unique protein sequences.

mutations that arise on a poor genetic background can recombine onto a better genetic background. Recombinative mutagenesis methods were developed in an effort to recreate this phenomenon *in vitro*.

Stemmer published the first account of *in vitro* recombination in 1994 (Stemmer, 1994). In DNA-shuffling one or more closely-related genes are digested with DNase I to yield double-stranded oligonucleotide fragments of 10-50bp, which are purified and used in a PCR-like reaction. Full-length genes are reconstructed from the fragments over the course of the reaction. Recombination occurs by template-switching: fragments from one copy of a gene prime on another. The combination of this technique with epPCR or bacterial mutator strains to introduce point mutations allows the simultaneous permutation of both single mutations and large blocks of DNA sequences. Mutants with improved properties can be reshuffled against each other or against the wild-type gene which would correspond to the backcross technique of classical genetics. This process may eliminate neutral or deleterious mutations from the gene pool. It is also possible to perform DNA-shuffling with homologous genes from different species (family-shuffling) which results in sparse sampling of a large sequence space, thereby significantly accelerating the rate of functional enzyme improvement (Crameri *et al.*, 1998).

StEP is a simple and efficient variation of DNA-shuffling that is based on a staggered, PCR-like reaction (Zhao *et al.*, 1998). StEP employs thermal cycling with extremely abbreviated annealing/DNAP-catalysed extension

cycles. In each cycle the growing fragments anneal to different templates based on sequence complementarity and extend further. This is repeated until full-length chimerical genes are reassembled. RPR, CLERY, and RACHITT are other variations of the DNA-shuffling technique (Table 1.7).

To permit recombination of unrelated or distantly-related proteins, non-homologous recombination methods have also been developed (Table 1.7). SHIPREC and ITCHY both permit the creation of arrays of hybrid proteins from distantly-related genes. However, neither technique permits more than one crossover to occur per hybrid. This limitation was addressed by developing SCRATCHY, a combination of ITCHY and DNA-shuffling. SCOPE is a different technique to those already mentioned because it utilizes protein structure information to guide the generation of multiple-crossover hybrids from non-homologous genes.

1.2.2 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

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 1.7. Current techniques for generating libraries of mutants for directed evolution.

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 small regions of the gene to a higher level of random mutation (Figure 1.12). This approach may yield novel combinations of substitutions

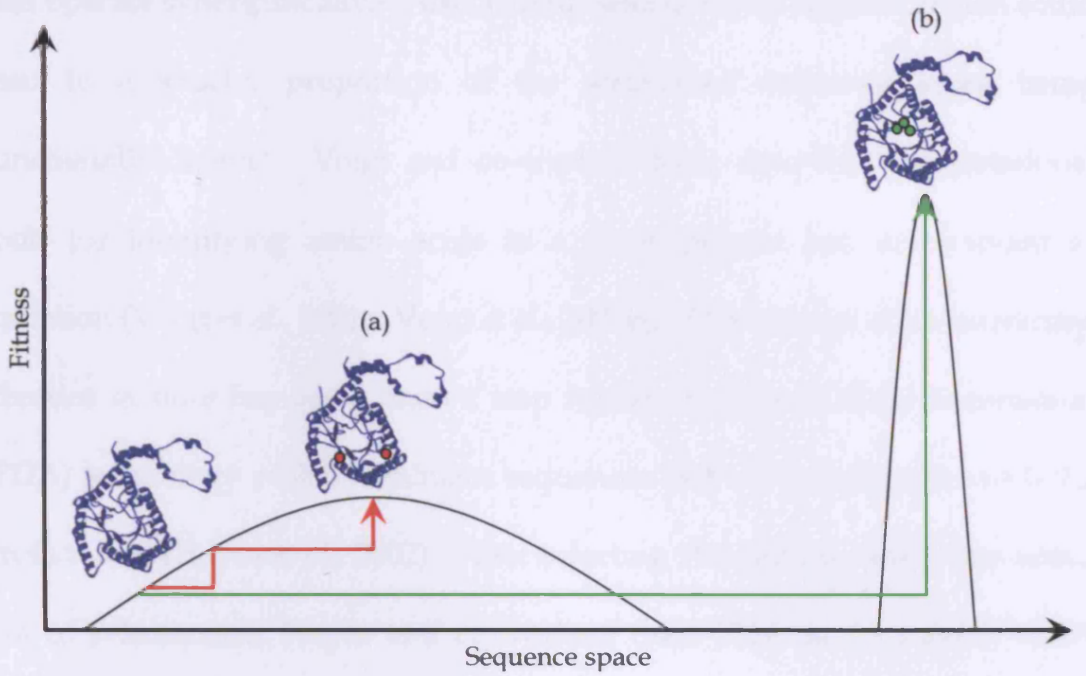


Figure 1.12. Simplified representation of evolutionary progressions by: (a) repeated single amino acid substitutions; and (b) intense mutation of small regions of the gene. It is possible for method (b) to yield highly successful mutants with substituted residues that work together. Note that, in reality, sequence space has many more dimensions.

that operate synergistically.[2] 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.

1.2.3 Library size and coverage

To demonstrate the calculations involved in assessing library size and coverage, an example utilising the *E. coli* transketolase gene and a non-recombinative mutagenesis method will be used.

1.2.3.1 Library size

The entire *tkt* gene is subjected to non-recombinative mutation at a frequency of *exactly* two base changes per variant. The total number of possible

[2] These combinations may not be identified by making single amino acid substitutions because the intermediates might not be favourable.

nucleotide sequences in this library can be calculated using the formula:

$$V = \binom{N}{x}(3^x)$$
$$= \frac{3^x N!}{x!(N-x)!}$$
$$= \frac{3^2 \times 1986!}{2!(1986-2)!}$$
$$= 1.77 \times 10^7$$

Where V = number of possible sequences, x = number of mutations, and N = length of the gene in base-pairs (excluding start and stop codons). There are $\binom{N}{x}$ different ways of choosing x bases to mutate and each of these may be mutated to any of the three other bases, giving the 3^x term (Patrick *et al.*, 2003).

The number of unique protein sequences encoded by this library is, however, less than the number of unique nucleotide sequences. Because the gene is long (1986bp), the probability that two nucleotide mutations would occur in the same codon is extremely low:

$$P = 2 \times \left(\frac{1}{1986}\right) \times \left(\frac{1}{1986-1}\right) \times \left(\frac{2}{3} \times 1986\right)$$
$$= 6.72 \times 10^{-4}$$

Thus, the mutations will not cooperate in the vast majority of cases, but will affect different residues. Codon degeneracy means that these separated mutations can grant access to only 5.98 new residues at each site (on average). This figure was calculated for the *tkt* gene by building a table of the possible residue changes for each codon (Table 1.8) and averaging the totals. As the number of residues accessible from each codon after a single base change was found to vary with the codon (5-7), the results were weighted for codon

Original residue			Identity of residue following a single point mutation at...												NOPE
Codon	CU *tkt*	Identity	First position				Second position				Third position				
			A	C	G	T	A	C	G	T	A	C	G	T	
AAA	4.82	K	K	Q	E	⊘	K	T	R	I	K	N	K	N	6
AAC	2.71	N	N	H	D	Y	N	T	S	I	K	N	K	N	7
AAG	0.45	K	K	Q	E	⊘	K	T	R	M	K	N	K	N	6
AAT	0.60	N	N	H	D	Y	N	T	S	I	K	N	K	N	7
ACA	0.15	T	T	P	A	S	K	T	R	I	T	T	T	T	6
ACC	3.46	T	T	P	A	S	N	T	S	I	T	T	T	T	5
ACG	0.30	T	T	P	A	S	K	T	R	M	T	T	T	T	6
ACT	1.05	T	T	P	A	S	N	T	S	I	T	T	T	T	5
AGA	0.00	R	R	R	G	⊘	K	T	R	I	R	S	R	S	5
AGC	0.45	S	S	R	G	C	N	T	S	I	R	S	R	S	6
AGG	0.00	R	R	R	G	W	K	T	R	M	R	S	R	S	6
AGT	0.00	S	S	R	G	C	N	T	S	I	R	S	R	S	6
ATA	0.00	I	I	L	V	L	K	T	R	I	I	I	M	I	6
ATC	3.01	I	I	L	V	F	N	T	S	I	I	I	M	I	7
ATG	3.31	M	M	L	V	L	K	T	R	M	I	I	M	I	6
ATT	1.81	I	I	L	V	F	N	T	S	I	I	I	M	I	7
CAA	0.30	Q	K	Q	E	⊘	Q	P	R	L	Q	H	Q	H	6
CAC	2.56	H	N	H	D	Y	H	P	R	L	Q	H	Q	H	7
CAG	3.16	Q	K	Q	E	⊘	Q	P	R	L	Q	H	Q	H	6
CAT	0.15	H	N	H	D	Y	H	P	R	L	Q	H	Q	H	7
CCA	0.00	P	T	P	A	S	Q	P	R	L	P	P	P	P	6
CCC	0.00	P	T	P	A	S	H	P	R	L	P	P	P	P	6
CCG	4.07	P	T	P	A	S	Q	P	R	L	P	P	P	P	6
CCT	0.30	P	T	P	A	S	H	P	R	L	P	P	P	P	6
CGA	0.15	R	R	R	G	⊘	Q	P	R	L	R	R	R	R	4
CGC	1.66	R	S	R	G	C	H	P	R	L	R	R	R	R	6
CGG	0.00	R	R	R	G	W	Q	P	R	L	R	R	R	R	5
CGT	2.41	R	S	R	G	C	H	P	R	L	R	R	R	R	6
CTA	0.00	L	I	L	V	L	Q	P	R	L	L	L	L	L	5
CTC	0.45	L	I	L	V	F	H	P	R	L	L	L	L	L	6
CTG	6.78	L	M	L	V	L	Q	P	R	L	L	L	L	L	5
CTT	0.15	L	I	L	V	F	H	P	R	L	L	L	L	L	6
GAA	5.87	E	K	Q	E	⊘	E	A	G	V	E	D	E	D	6
GAC	4.07	D	N	H	D	Y	D	A	G	V	E	D	E	D	7
GAG	1.36	E	K	Q	E	⊘	E	A	G	V	E	D	E	D	6
GAT	1.20	D	N	H	D	Y	D	A	G	V	E	D	E	D	7
GCA	2.71	A	T	P	A	S	E	A	G	V	A	A	A	A	6
GCC	2.11	A	T	P	A	S	D	A	G	V	A	A	A	A	6
GCG	4.22	A	T	P	A	S	E	A	G	V	A	A	A	A	6
GCT	4.37	A	T	P	A	S	D	A	G	V	A	A	A	A	6
GGA	0.00	G	R	R	G	⊘	E	A	G	V	G	G	G	G	4
GGC	2.86	G	S	R	G	C	D	A	G	V	G	G	G	G	6
GGG	0.30	G	R	R	G	W	E	A	G	V	G	G	G	G	5
GGT	5.57	G	S	R	G	C	D	A	G	V	G	G	G	G	6
GTA	0.90	V	I	L	V	L	E	A	G	V	V	V	V	V	5
GTC	0.90	V	I	L	V	F	D	A	G	V	V	V	V	V	6
GTG	1.51	V	M	L	V	L	E	A	G	V	V	V	V	V	5
GTT	2.41	V	I	L	V	F	D	A	G	V	V	V	V	V	6
TAA	0.15	⊘	K	Q	E	⊘	⊘	S	⊘	L	⊘	Y	⊘	Y	
TAC	2.71	Y	N	H	D	Y	Y	S	C	F	⊘	Y	⊘	Y	6
TAG	0.00	⊘	K	Q	E	⊘	⊘	S	W	L	⊘	Y	⊘	Y	
TAT	0.75	Y	N	H	D	Y	Y	S	C	F	⊘	Y	⊘	Y	6
TCA	0.30	S	T	P	A	S	⊘	S	⊘	L	S	S	S	S	4
TCC	2.56	S	T	P	A	S	Y	S	C	F	S	S	S	S	6
TCG	0.15	S	T	P	A	S	⊘	S	W	L	S	S	S	S	5
TCT	2.26	S	T	P	A	S	Y	S	C	F	S	S	S	S	6
TGA	0.00	⊘	R	R	G	⊘	⊘	S	⊘	L	⊘	C	W	C	
TGC	0.60	C	S	R	G	C	Y	S	C	F	⊘	C	W	C	6
TGG	1.66	W	R	R	G	W	⊘	S	W	L	⊘	C	W	C	5
TGT	0.15	C	S	R	G	C	Y	S	C	F	⊘	C	W	C	6
TTA	0.00	L	I	L	V	L	⊘	S	⊘	L	L	F	L	F	4
TTC	3.31	F	I	L	V	F	Y	S	C	F	L	F	L	F	6
TTG	0.15	L	M	L	V	L	⊘	S	W	L	L	F	L	F	5
TTT	0.60	F	I	L	V	F	Y	S	C	F	L	F	L	F	6

Table 1.8. (Previous page.) The possible identities of any given codon following a single point mutation at either the first, second, or third nucleotide position. The start (ATG) and stop (TAA) codons of the gene are not included. "CU *tkt*" is an abbreviation for "codon usage in the gene tkt". "NOPE" is an abbreviation for "number of possible exchanges" and is equal to the total number of new residues that can be accessed by a single point mutation in a particular codon. Changes leading to new residues are shaded green. Changes leading to or from stop (⊘) codons are shaded red.

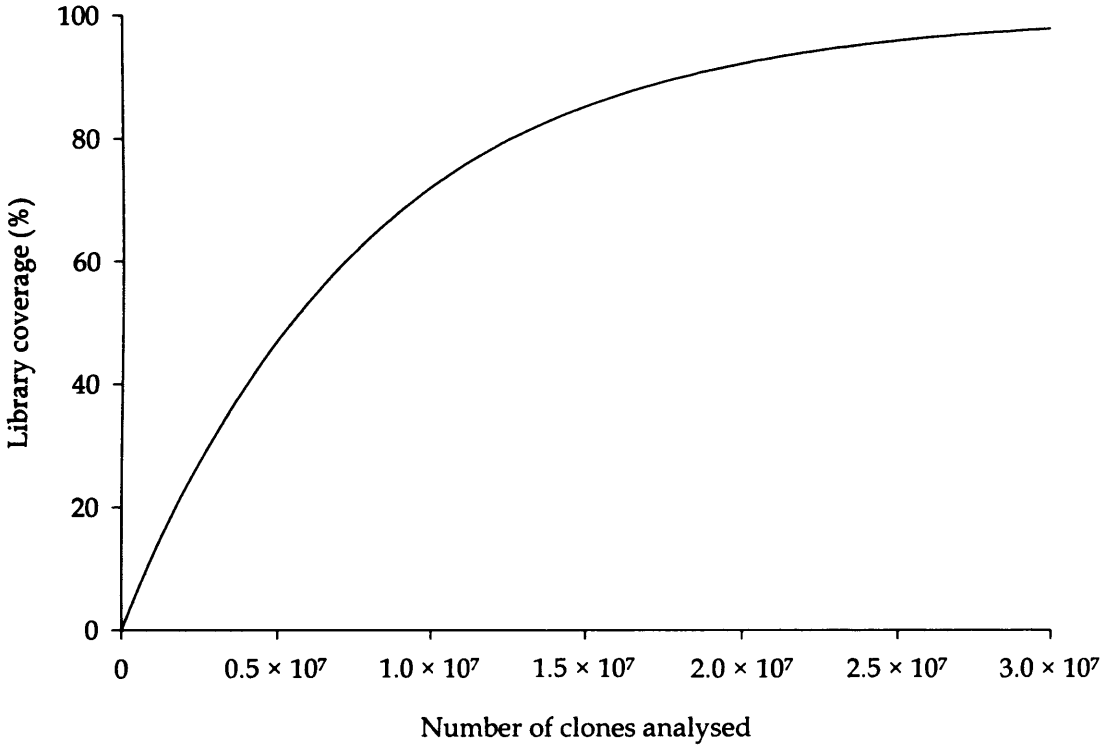


Figure 1.13. The relationship between the number of transformants analysed and percentage library coverage in the example scenario. Transformants are analysed at random from a library of 7.82×10^6 unique proteins. Percentage library coverage increases with the number of transformants analysed in an exponential rise to maximum (100%).

usage in the *tkt* gene before being averaged. The number of unique proteins encoded by the library of 1.77×10^7 nucleotide sequences is:

$$W = \binom{N/3}{x}(5.98^x)$$
$$= \frac{5.98^x (\frac{N}{3})!}{x!(N-x)!}$$
$$= \frac{5.98^2 \times 662!}{2!(662-2)!}$$
$$= 7.82 \times 10^6$$

Where W = number of possible sequences, x = number of mutations, and $N/3$ = length of the protein in residues. There are $\binom{N/3}{x}$ different ways of choosing x residues to mutate and each of these may be mutated to 5.98 other residues (on average), giving the 5.98^x term.

1.2.3.2 Library coverage

Screening or selecting 7.82×10^6 transformants would not, however, cover 100% of the library. The actual proportion would be:

$$C = 1 - e^{\left(\frac{-A}{L}\right)}$$
$$= 1 - e^{\left(\frac{-7.82\times10^6}{7.82\times10^6}\right)}$$
$$= 0.632$$
$$= 63.2\%$$

Where C = proportion of library examined, L = library size (number of unique proteins), and A = number of clones analysed. If 90% coverage of the library was required, then 2.53×10^7 transformants would need to be analysed. Although library coverage increases with the number of transformants analysed, it never actually reaches 100% (Figure 1.13). This is because

transformants are picked at random from the population: some protein sequences are analysed more than once and some are not analysed at all.

1.2.4 Screening and selection

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 (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 coexpressed 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 (3.1g.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 (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.

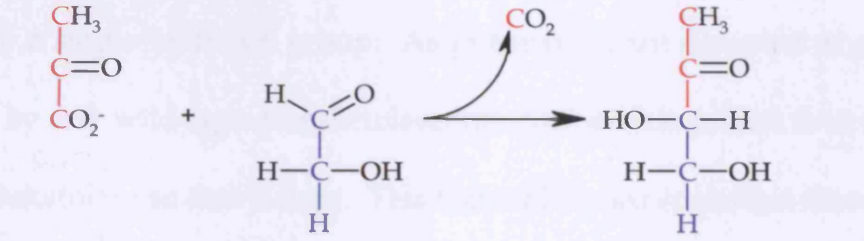
1.2.5 Successful applications of directed evolution

There are many excellent examples of biocatalysts being modified for specific applications. Characteristics of wild-type enzymes that have been improved include: activity, enantioselectivity, oxidative stability, solvent stability, substrate specificity, and thermostability (Table 1.9). These changes have all been brought about by favourable combinations of mutagenesis (refer to Section 1.2.1) and selection/screening (refer to Section 1.2.4).

With regard to carbon-carbon bond-forming enzymes, several have been successfully modified by directed evolution to accept unnatural substrates, including: the pyruvate-dependent aldolases *N*-acetylneuraminate lyase (Wada *et al.*, 2003) and 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase (Griffiths *et al.*, 2004); and the dihydroxyacetone phosphate-dependent aldolase tagatose-bisphosphate aldolase (Williams *et al.*, 2003). The *in vivo* selection strategy employed by Griffiths and co-workers involved a pyruvate kinase-deficient *E. coli* cell line. The auxotroph could be rescued by supplementing the growth medium with pyruvate or, in the case of the selection, retro-aldol cleavage of an unnatural 2-keto-4-hydroxybutyrate adduct by a suitable mutant of KDPG aldolase. 10^7 KDPG aldolase mutants were selected using this method and three active variants were identified.

Improvement			Mutagenesis technique	Reference
Type	Enzyme	Extent of change		
Activity	Phospho-triesterase	63-fold increase in k_{cat}	Saturation mutagenesis	Griffiths and Tawfik, 2003
Activity in an organic co-solvent	Subtilisin E	256-fold increase in k_{cat}/K_m in 60% aqueous dimethylformamide	epPCR	Chen and Arnold, 1993
Enantioselectivity	Hydantoinase	L-selective enzyme evolved from D-selective wild-type	epPCR	May *et al.*, 2000
Site specificity	Cre recombinase	Evolved to selectively recombine a novel recombination site	Saturation mutagenesis	Santoro and Schultz, 2002
Substrate specificity	β-Galactosidase	10–20-fold higher k_{cat}/K_m for fucose (not galactose) substrates	DNA-shuffling	Zhang *et al.*, 1997
Thermostability	*p*-Nitrobenzyl esterase	T_m increased 14°C without activity at lower temperatures being compromised	epPCR DNA-shuffling	Giver *et al.*, 1998

Table 1.9. Examples of enzyme characteristics that have been improved by directed evolution.



pyruvate glycolaldehyde (*S*)-3,4-dihydroxybutan-2-one

Figure 1.14. Scheme of the desired activity shown in Fischer projection. A 1-hydroxyethyl group is transferred from pyruvate to glycolaldehyde, forming (*S*)-3,4-dihydroxybutan-2-one and carbon dioxide.

1.3 The aims of this project

The ability of transketolase to transfer a C_2 moiety from a donor substrate to a wide range of acceptor substrates in a stereospecific manner advocates its use in the synthesis of fine chemicals and pharmaceuticals. β-HPA is frequently employed as the donor substrate because the formation of carbon dioxide as co-product renders the reaction irreversible. However, β-HPA is an extremely expensive compound (£113.60 per gram) so adoption of transketolase as an industrial biocatalyst is unlikely unless a new donor substrate can be found. Pyruvate would be a suitable replacement as it is inexpensive (£0.19 per gram) and differs from β-HPA chemically by the absence of a single hydroxyl group. As pyruvate is not accepted as a donor substrate by any wild-type transketolase, the goal of this project is to modify *E. coli* transketolase so that it does. This feat will be accomplished through the use of directed evolution and a high-throughput screen for transketolase activity. The aims of this project are, therefore:

Firstly, to generate a library of *E. coli* transketolase variants with mutations limited to stretches of residues likely to be involved in β-HPA/pyruvate discrimination. This "focused" approach has two advantages over "whole-gene" mutagenesis: (a) library size is dramatically reduced; and (b) much higher mutational loads can be applied, permitting the identification of synergistic combinations of mutations. It is believed that this approach may yield larger step enhancements than existing approaches. The development of a novel, focused mutagenesis technique - focused epPCR-

QuikChange - is discussed in Chapter 5. The implementation of this method in creating a transketolase library is discussed in Chapters 5 and 6. Chapter 4 discusses the initial design of the library, i.e. the selection of mutagenesis targets.

Secondly, to develop a high-throughput screen capable of identifying the desired activity (Figure 1.14) in a quantitative manner. The sensitivity of this screen will be maximised by doing the following: (a) selecting the best *E. coli* strain for hosting the transketolase library; (b) developing a method for culturing library members to high optical densities in a massively-parallel format; (c) developing an effective procedure for releasing transketolase protein from intact cells; and (d) optimising the cofactor concentrations in the reaction step so that the amount of holotransketolase is maximised. The synthesis of (*S*)-3,4-dihydroxybutan-2-one in each screening reaction will be monitored by a suitable high-throughput assay. The development of this screen is discussed in Chapter 3.

Thirdly, to screen the transketolase library for the desired reaction (Figure 1.14). A variant capable of catalysing the transfer of a 1-hydroxyethyl group from pyruvate to an acceptor substrate in a stereospecific manner would be extremely useful to the pharmaceutical and fine-chemical industries. The results of the screening operation are discussed in Chapter 6.

Chapter 2 - Materials and methods

The protocols described in this chapter are standard practices and not experimental. Experimental techniques are described in Chapters 3–6.

2.1 Materials

Unless otherwise stated, all materials were purchased from Sigma-Aldrich Company Ltd. Water was purified to 15MΩ.cm resistivity using an Elix 5 water purification system (Millipore Corp.).

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 NZY$^+$ medium

NZY$^+$ medium was prepared by dissolving 10g.l^{-1} NZ amine, 5g.l^{-1} yeast extract, and 5g.l^{-1} sodium chloride in pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7.5. After autoclaving, the following filter-sterilised solutions were added: 20mM glucose, 12.5mM $MgCl_2$, and 12.5mM $MgSO_4$.

2.2.4 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 $MgSO_4$.

2.2.5 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} (a 1000-fold dilution of the stock) to select bacteria carrying the plasmid pQR711. Preparations containing this concentration of ampicillin are labelled Amp$^+$.

2.2.6 0.5M Tris buffer (pH 7.5)

0.5M tris(hydroxymethyl)aminomethane (Tris) buffer with a pH of 7.5 was prepared by dissolving 63.5g.l^{-1} Tris hydrochloride and 11.8g.l^{-1} Tris base in 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 ThDP (M_r = 460.8) in 4.5ml of pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7.5. The 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 0.1375g β-HPA (M_r = 109.99), 0.0751g glycolaldehyde (M_r = 60.05), and 1.25ml of 0.5M Tris buffer (pH 7.5) in 3.25ml of pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7.5. The substrate solution was topped up to 5ml with pure water and stored at 4°C (for no more than 24 hours).

2.2.9 Cresol red substrate solution

Cresol red substrate solution was prepared by dissolving 0.0688g β-HPA, 0.0376g glycolaldehyde, and 25mg.l^{-1} cresol red in 4.5ml of pure water. A concentrated solution of sodium hydroxide was used to adjust the pH to 7.5. The substrate solution was topped up to 5ml with pure water and stored at 4°C (for no more than 24 hours).

2.3 Standard procedures

2.3.1 Streaked 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 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.

2.3.7 Transformation by heat-shock

Three strains of competent cells were routinely transformed: *E. coli* BL21-Gold(DE3), JM109, and XL10-Gold (all supplied by Stratagene Ltd.).

The relevant competent cells were thawed on ice and an aliquot (Table 2.1) was transferred to a chilled 1.5ml Eppendorf tube. If necessary, β-mercaptoethanol was added (Table 2.1) and the cells were incubated on ice for 10 minutes. 1µl 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 (Table 2.1). The transformation reaction was then incubated on ice for a further 2 minutes. 0.5ml of preheated (42°C) growth medium (Table 2.1) was added to the tube which was then incubated at 37°C for 1 hour with

E. coli strain	Aliquot (µl)	β-mercapto-ethanol	Heat-shock (seconds)	Growth medium
BL21-Gold(DE3)	50		20	SOC
JM109	50	0.4µl, 1.42M	45	SOC
XL10-Gold	45	2µl	30	NZY⁺

Table 2.1. Values for various parameters during the transformation of competent cells. The concentration of the β-mercaptoethanol solution provided with the competent XL10-Gold cells could not be discovered.

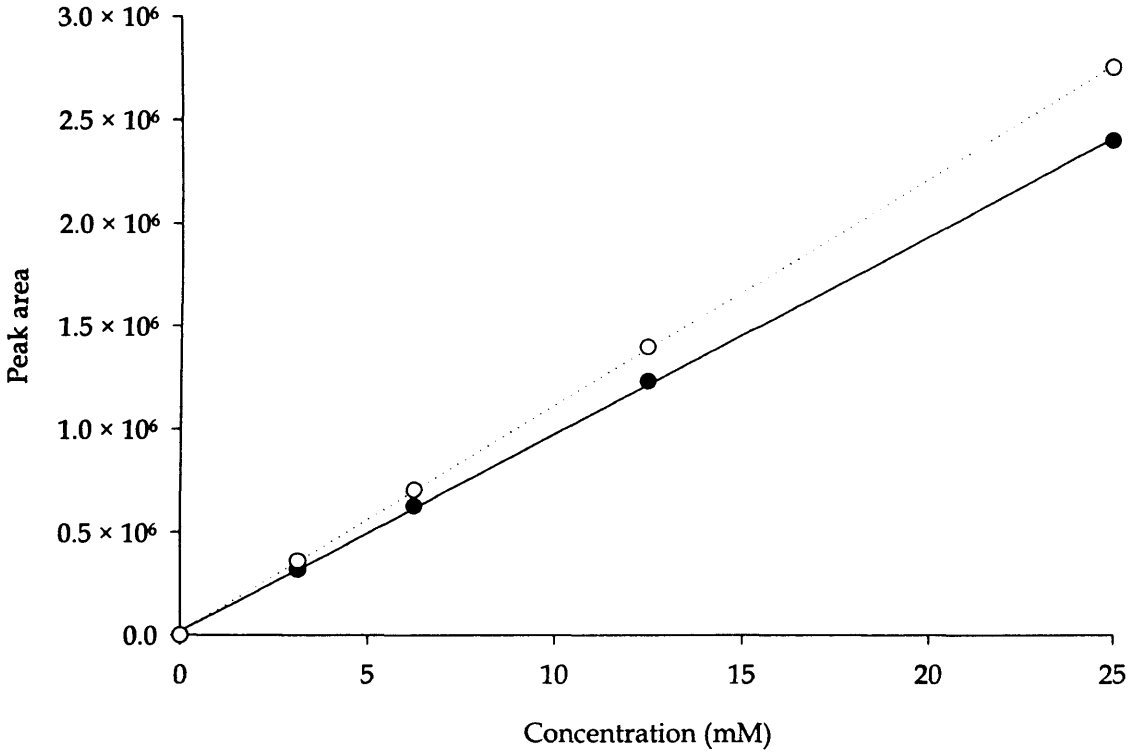


Figure 2.1. HPLC calibration curves for L-erythrulose and β-HPA. The relationships between concentration and detector response are approximately linear for L-erythrulose (•) and β-HPA (o) with R^2 values of 0.9997 and 0.9999, respectively. Above 25mM the relationships become more curved.

220rpm agitation.

2.3.8 Measurement of absorbance and optical density

2.3.8.1 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).

2.3.8.2 Microwell plates

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

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

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

2.3.9.1 Sample preparation

The reaction sample was diluted 1:4 by the addition of 0.2% (v/v) TFA. 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.

2.3.9.2 HPLC system

The HPLC system consisted of an Endurance autosampler (Spark Holland BV), a GP50 gradient pump (Dionex Corp.), an LC30 chromatography oven (Dionex Corp.), a PC10 pneumatic controller (Dionex Corp.), 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 detector.

2.3.9.3 HPLC method

The mobile phase was 0.1% (v/v) trifluoroacetic acid (TFA) and the flow rate was 0.8ml.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.

2.3.9.4 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.[3]

Figure 2.1 illustrates the calibration curves for β-HPA and L-erythrulose. In both cases the relationship between injection concentration

[3] Glycolaldehyde has a low extinction coefficient at 210nm wavelength.

and peak area is proportional up to 25mM.

2.3.10 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 (Sigma-Aldrich Company Ltd.). 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).

2.3.11 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).

2.3.12 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.

2.3.12.1 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.

2.3.12.2 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.

2.3.12.3 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 2.2. Components of 12.5% (w/v) separating and 6% (w/v) stacking gels. Ammonium persulphate and TEMED (N,N,N',N'-tetramethylethylendiamine) were added to each gel at the last minute to initiate polymerisation.

2.3.12.4 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).

2.3.13 DNA sequencing

Cycle sequencing (Sambrook *et al.*, 1989) was performed by the Sequencing Service of the Wolfson Institute for Biomedical Research (UCL). All DNA samples were given to the service at a concentration of 100fmoles in 6µl. Primer TKN (5'-GATCCAGAGATTTCTGA-3', *tkt* nucleotides -140 to -124) was used for sequencing targets A and B and primer TKC (5'-TATCTCCCTGCACGGTGGCTTCC-3', *tkt* nucleotides +1260 to +1282) was used for target C.

Chapter 3 - High-throughput screen

3.1 Introduction

It was envisaged that the screening phase of this project would involve four steps: (1) transformation of the pQR711 variants into a suitable host; (2) microwell culture of the colonies; (3) lysis of the cultures; and finally (4) incubation with the target substrates (pyruvate and glycolaldehyde). If any of the mutant transketolases were capable of synthesising (*S*)-3,4-dihydroxybutan-2-one (and carbon dioxide) from these substrates (Figure 1.14) they would be identified by a suitable assay.

The principal methods for measuring transketolase activity are a continuous multi-enzyme-linked spectrophotometric assay (Villafranca and Axelrod, 1971) and a discontinuous HPLC assay (Mitra and Woodley, 1996). The multi-enzyme-linked assay uses α-glycerophosphate dehydrogenase, triosephosphate isomerase, phosphoribose isomerase, and D-ribulose-5-phosphate-3-epimerase to link transketolase activity to NADH oxidation. By monitoring the rate of NADH depletion at 340nm wavelength (A_{340}), the transketolase activity can be elucidated. The spectrophotometric nature of this assay makes it a good choice for high-throughput screening: hundreds of transketolase variants could be screened simultaneously in a microwell plate. However, the assay relies on a specific chain of reactions that: (a) will not permit β-HPA or pyruvate to be used as the donor substrate; and (b) will not yield a true measure of transketolase activity if one of the downstream links becomes the rate-limiting step.

HPLC methods evaluate the concentration of product formed in a reaction at intervals, thus allowing enzyme activity to be elucidated. As the product is examined directly, HPLC-based determinations of transketolase activity do not suffer from the limitations of the multi-enzyme-linked assay. One such method, developed by Christine Ingram (Department of Biochemical Engineering, UCL), allows the bioconversion of β-HPA and glycolaldehyde to L-erythrulose (and carbon dioxide) to be followed.[4] This method has been useful in making precise measurements of transketolase activity (refer to Section 2.3.9), but it was too sluggish for high-throughput application (11 minutes per sample = 131 samples per day).

A recent publication describes an entirely new approach to determining transketolase activity: the use of a fluorogenic donor substrate (Sevestre *et al.*, 2003). When this compound (1) is cleaved by transketolase it yields a highly unstable compound (2) that rapidly undergoes β-elimination to form umbelliferone (Figure 3.1). Umbelliferone is a highly fluorescent compound whose formation could be easily measured in a high-throughput manner. This method would, therefore, be extremely useful for the directed evolution of new acceptor substrate-specificities in transketolase. However, as the goal of this project is to modify donor substrate specificity, this technique cannot be used.

As the existing methods for determining transketolase activity were not suitable for the screening phase of this project, a new one was required.

[4] This method was in turn developed from that of Mitra and Woodley (1996).

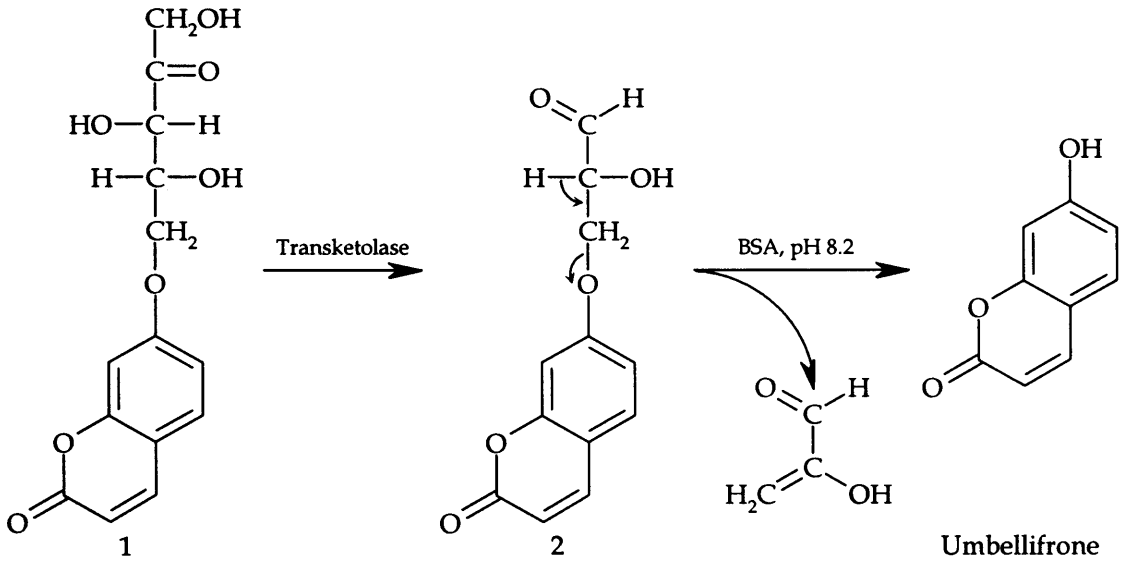


Figure 3.1. A fluorogenic assay for transketolase activity. Compound 1 is a donor substrate for the transketolase reaction. The transfer of a 1,2-dihydroxyethyl group from compound 1 to an acceptor substrate yields a ketose product and compound 2. Compound 2 undergoes rapid β-elimination to form the highly-fluorescent compound umbellifrone. "BSA" is an abbreviation for bovine serum albumin. (Adapted from Sevestre *et al.*, 2003.)

Two separate approaches were investigated: (a) a spectrophotometric assay based on a pH indicator and the pH increase that accompanies transketolase-mediated condensation (Mitra *et al.*, 1998); and (b) a fast HPLC protocol based on a short column. Approach (a) has been successfully employed by a number of groups (for example: May *et al.*, 2000). Although there are no examples of HPLC-based screens being used in directed evolution, they have been used in combinatorial chemistry (Ding *et al.*, 1999; Mikami *et al.*, 2001) and in the nutritional analysis of seeds from *Arabidopsis* mutants (Jander *et al.*, 2004).

The remainder of the chapter relates efforts to maximise the quantity of holotransketolase in each screening reaction. As the k_{cat} values of active transketolase variants might be low, it was reasoned that higher holoenzyme concentrations would aid their identification. The aspects of transketolase production and reconstitution that were examined were: (a) the choice of host strain; (b) the microwell culture process; (c) the lytic method; and (d) the concentrations of cofactors in the reaction step. As these aspects could not be optimised using the target reaction (Figure 1.14), the β-HPA and glycolaldehyde to L-erythrulose (and carbon dioxide) reaction was used as a model. Reaction conditions other than the concentrations of cofactors were maintained at departmental standards (refer to Section 1.1.3.4.), i.e. 15 minutes preincubation time, 100mM substrates, 25°C temperature, and pH 7.5.

In this chapter transketolase activity is expressed in $mkat.m^{-3}$ (millikatals per cubic meter). The "katal" is the Système International unit for

catalytic activity and is equal to one mole of product formed per second (Dybkaer, 2001).

3.2 Methods

3.2.1 Selection of the host strain

3.2.1.1 Transformation

E. coli JM107 and JM107 pQR711 were a gift from Dr. John Ward (Department of Biochemistry, UCL). Streaked plates (Section 2.3.1) and glycerol stocks (Section 2.3.4) were immediately prepared from both strains.

Plasmid DNA was extracted from an overnight culture (Section 2.3.2) of *E. coli* JM107 pQR711 by the standard procedure (Section 2.3.6). The resulting pQR711 DNA was transformed by heat-shock (Section 2.3.7) into competent cells of three strains: *E. coli* BL21-Gold(DE3), JM109, and XL10-Gold. 10µl aliquots of transformed cells were added to 90µl of LB medium and spread on LB Amp^+ agar plates. The plates were incubated overnight at 37°C.

Overnight cultures (Section 2.3.2) were grown from individual colonies of *E. coli* BL21-Gold(DE3) pQR711, JM109 pQR711, and XL10-Gold pQR711, and plasmid DNA was extracted (Section 2.3.6). 280ng of each DNA sample were added to 3µl of 10× NEBuffer 3 (New England Biolabs Inc.), 1µl of the restriction enzyme *Mlu* I (New England Biolabs Inc.), and sufficient pure water to make a total volume of 30µl. The samples were incubated at 37°C for one hour to permit digestion. 40ng of each digested sample were run on a

0.7% (w/v) agarose gel (Section 2.3.10).

3.2.1.2 Confirmation of transketolase overexpression

Overnight cultures (Section 2.3.2) were grown of *E. coli* BL21-Gold(DE3), JM107, JM109, XL10-Gold, and the corresponding pQR711 strains. The final OD_{600} of each culture was measured (Section 2.3.8.1) and an aliquot of each was diluted to 1.00ODU. These diluted samples were denatured and run on an SDS-PAGE gel (Section 2.3.12).

3.2.1.3 Reaction profiles

Shake flask cultures (Section 2.3.3) of *E. coli* BL21-Gold(DE3), JM107, JM109, XL10-Gold, and the corresponding pQR711 strains were prepared. Aliquots of these cultures were diluted with LB medium to OD_{600}s of 3.00ODU and lysed by sonication (Section 2.3.5). 0.5ml of each sonicate was added to 0.1ml of cofactor solution (Section 2.2.7) (LB medium replaced sonicate in a control sample). After a 15 minute preincubation period the reactions were initiated by adding 0.4ml of substrate solution (Section 2.2.8). The final concentrations in each reaction were: 100mM glycolaldehyde, 100mM β-HPA, 50mM Tris buffer (pH 7.5), 9mM $MgCl_2$, and 2.5mM ThDP. The reactions were allowed to proceed at 25°C for 90 minutes. At frequent time points samples of each reaction were quenched and analysed for L-erythrulose (Section 2.3.9).

3.2.1.4 Transketolase activity

Shake flask cultures (Section 2.3.3) of *E. coli* BL21-Gold(DE3), JM107, JM109, XL10-Gold, and the corresponding pQR711 strains, were prepared. All strains grew to similar final OD_{600}s (4–5ODU). Aliquots of the wild-type cultures were diluted with LB medium to OD_{600}s of 3.00ODU and lysed by sonication (Section 2.3.5). Aliquots of the pQR711 cultures were diluted to OD_{600}s of 0.50ODU and sonicated. 150µl of each sonicate was added to 30µl of cofactor solution (Section 2.2.7). Three replicate samples were prepared for each strain (LB medium replaced sonicate in three control samples). After a 15 minute preincubation period the reactions were initiated by adding 120µl of substrate solution (Section 2.2.8). The final concentrations in each reaction were: 100mM glycolaldehyde, 100mM β-HPA, 50mM Tris buffer (pH 7.5), 9mM $MgCl_2$, and 2.5mM ThDP. The *E. coli* JM107 pQR711, JM109 pQR711, and XL10-Gold pQR711 reactions were quenched after five minutes at 25°C (Section 2.3.9). The remaining reactions were quenched after a further 55 minutes at 25°C. The concentration of L-erythrulose in each quenched reaction was determined by HPLC (Section 2.3.9) and the corresponding transketolase activity value was calculated.

3.2.2 Microwell culture

Each well of three 96-well plates and three 384-well plates was filled with the appropriate volume of LB Amp^+ medium (Table 3.1). A QPix2 robot (Genetix Ltd.) was programmed to inoculate every well of the six plates with colonies

of *E. coli* XL10-Gold pQR711. Each plate was sealed by taking an identical (but empty) plate, inverting it, and taping it over the top. The sealed plates were incubated for 16 hours at various agitation speeds on a Variomag Teleshake unit (Camlab Ltd.) in a 37°C incubator (Table 3.1). A shake flask of *E. coli* XL10-Gold pQR711 was prepared in parallel (Section 2.3.3).

After incubation, the uniformity of cell growth across each plate was analysed (Section 2.3.8.2). The wells of each plate were pooled and the OD_{600} was measured (Section 2.3.8.1). The final OD_{600} of the shake flask culture was measured (Section 2.3.8.1).

3.2.3 Optimisation of cofactors

A shake flask culture (Section 2.3.3) of *E. coli* XL10-Gold pQR711 was prepared. An aliquot of this culture was diluted with LB medium to an OD_{600} of 0.50ODU and lysed by sonication (Section 2.3.5). A 96-well plate was set up with a grid of wells containing increasing concentrations of ThDP across one axis and increasing concentrations of $MgCl_2 \cdot 6H_2O$ down the other. The total volume of cofactors in each well was 30µl. 150µl of the sonicate was pipetted into each well. LB medium replaced sonicate in two control samples: one control had maximum cofactors and the other had none at all. After a 15 minute preincubation period the reactions were initiated by adding 120µl of substrate solution (Section 2.2.8). The final concentrations in each reaction were: 100mM glycolaldehyde, 100mM β-HPA, 50mM Tris buffer (pH 7.5), 0-32mM Mg^{2+}, and 0-8mM ThDP. The reactions were allowed to proceed at

Format	Medium volume (µl)	Agitation (rpm)
96-well	150	600
96-well	150	1000
96-well	150	1400
384-well	60	600
384-well	60	1000
384-well	60	1400

Table 3.1. Candidate microwell culture formats. "Medium volume" is volume of LB Amp$^+$ medium per well.

Treatment	Label
Sonicated by standard method (2.3.5)	Sonication
Incubated at 25°C with 10% (v/v) BugBuster	
for 30 minutes	10% BB 30min
for 1 hour	10% BB 60min
Incubated at 25°C with 20% (v/v) BugBuster	
for 30 minutes	20% BB 30min
for 1 hour	20% BB 60min
Cycles of freezing at -80°C and thawing at 25°C	
1 cycle	F/T × 1
2 cycles	F/T × 2
3 cycles	F/T × 3
No further treatment	Not treated

Table 3.2. Candidate lytic methods. BugBuster HT Protein Extraction Reagent is a proprietary lysing reagent sold by Novagen Inc.

25°C for ten minutes. Samples of each reaction were quenched and diluted 1:4 by the addition of 0.2% (v/v) TFA. The concentration of L-erythrulose in each quenched reaction was determined by HPLC (Section 2.3.9) and the corresponding transketolase activity value was calculated.

3.2.4 Lytic methods

A shake flask culture (Section 2.3.3) of *E. coli* XL10-Gold pQR711 was prepared. This culture was diluted with LB medium to an OD_{600} of 3.41ODU. 1ml aliquots were pipetted into 1.5ml Eppendorf tubes and lysed by different methods (Table 3.2). Three replicate samples were prepared using each lytic method. After treatment the samples were diluted 1:6.82 by the addition of LB medium. 150µl of each diluted sample was added to 30µl of cofactor solution (Section 2.2.7) (LB medium replaced lysate in three control samples). After a 15 minute preincubation period the reactions were initiated by adding 120µl of substrate solution (Section 2.2.8). The final concentrations in each reaction were: 100mM glycolaldehyde, 100mM β-HPA, 50mM Tris buffer (pH 7.5), 9mM Mg^{2+}, and 2.5mM ThDP. The reactions were allowed to proceed at 25°C for ten minutes. Samples of each reaction were quenched and diluted 1:4 by the addition of 0.2% (v/v) TFA. The concentration of L-erythrulose in each quenched reaction was determined by HPLC (Section 2.3.9) and the corresponding transketolase activity value was calculated.

3.2.5 High-throughput assays for transketolase activity

3.2.5.1 Spectrophotometric assay

A shake flask culture (Section 2.3.3) of *E. coli* JM107 pQR711 was prepared and its final OD_{600} was measured (Section 2.3.8.1). This culture was diluted with LB medium to an OD_{600} of 3.00ODU and lysed by sonication (Section 2.3.5). LB medium was added to aliquots of this sonicate to make three dilutions: 1:2, 1:4, and 1:8. 0.5ml of each dilution was added to 0.1ml of cofactor solution (Section 2.2.7) in a polystyrene cuvette (LB medium replaced sonicate in a control sample). After a ten minute preincubation period the reactions were initiated by adding 0.4ml of cresol red substrate solution (Section 2.2.9). The final concentrations in each reaction were: 50mM glycolaldehyde, 50mM β-HPA, 9mM $MgCl_2$, 2.5mM ThDP, and 10mg.l^{-1} cresol red. The A_{573} of each reaction was monitored in a UV2 spectrophotometer.

3.2.5.2 High-throughput HPLC method

3.2.5.2.1 HPLC system

The same HPLC system was used as described in Section 2.3.9.

3.2.5.2.2 HPLC method

The mobile phase was 0.1% (v/v) trifluoroacetic acid (TFA) and the flow rate was 2ml.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.

3.2.5.2.3 Retention times and calibration curves

Retention time data and calibration curves were collected for L-erythrulose, glycolaldehyde, β-HPA, and pyruvate.

3.3 Results

3.3.1 Selection of host strain

E. coli JM107 pQR711 yields transketolase at a concentration of 168mg per litre of shake flask culture, corresponding to 23% of total cell protein (French and Ward, 1995). However, *E. coli* JM107 was ruled out as a host for the mutant transketolase library because commercial competent cells were not available. Commercial competent cells are preferable for library construction because their transformation efficiencies are several orders of magnitude greater than those prepared in the laboratory. Several commercial strains of *E. coli* were investigated as potential hosts for the transketolase library: BL21-Gold(DE3), JM109, and XL10-Gold (Table 3.3).

3.3.1.1 Transformation

pQR711 DNA was extracted from *E. coli* JM107 pQR711 and transformed into the commercial strains by heat-shock. To confirm that the transformations had been successful, plasmid DNA was extracted from these new strains,

Strain	Genotype
Original host for pQR711	
JM107	e14$^-$(McrA$^-$) *endA1 gyrA96 thi-1 hsdR17* (r_K m_K^+) *supE44 relA* Δ(*lac-proAB*) [F′ *traD36 proAB lacI^{q}ZΔM15*]
Potential hosts for the library	
BL21-Gold(DE3)	F$^-$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte
JM109	e14$^-$(McrA$^-$) *recA1 endA1 gyrA96 thi-1 hsdR17* (r_K m_K^+) *supE44 relA* Δ(*lac-proAB*) [F′ *traD36 proAB lacI^{q}ZΔM15*]
XL10-Gold	Tet^r Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F′ *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r) Amy Cam^r]

Table 3.3. Gentotypes of the *E. coli* strains BL21-Gold(DE3), JM107, JM109, and XL10-Gold. Genes that are listed signify mutant alleles. Genes on the F′ episome, however, are wild-type unless indicated otherwise. (a) JM107 cells are endonuclease (*endA*) deficient, greatly improving the quality of miniprep DNA. The *hsdR* mutation prevents cleavage of cloned DNA by the *Eco*K endonuclease system. (b) BL21-Gold(DE3) is a general protein expression strain that lacks both the Lon protease and the OmpT protease, which can degrade proteins during purification. (c) JM109 cells are genetically identical to JM107 cells except for being recombination deficient (*recA*). This deficiency improves insert stability. (d) XL10-Gold cells are deficient in all known restriction systems [Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173*]. The strain is endonuclease deficient (*endA*) and recombination deficient (*recA*). The Hte phenotype increases the transformation efficiency of ligated and large supercoiled DNA.

linearised by the restriction enzyme *Mlu* I, and analysed by agarose gel electrophoresis (Section 3.2.1.1) (Figure 3.2). The digestion step was necessary to eliminate gel bands caused by circular DNA, plasmid dimers, and other molecular complexes. All three of the new strains were found to harbour the pQR711 plasmid (5.4kbp).

3.3.1.2 Confirmation of transketolase overexpression

Cultures of *E. coli* BL21-Gold(DE3), JM107, JM109, XL10-Gold, JM107 pQR711 and the three new pQR711 strains were analysed by SDS-PAGE (Section 3.2.1.2). Dark bands of ~70kDa were observed in the *E. coli* JM107 pQR711, JM109 pQR711, and XL10-Gold pQR711 lanes (Figure 3.3). A fainter band of the same molecular mass was observed in the *E. coli* B BL21-Gold(DE3) pQR711 lane. As 70kDa approximated the relative molecular mass of the *E. coli* transketolase monomer (72.036kDa) it was concluded that the enzyme was being overexpressed in all of the pQR711 strains. However, the expression of transketolase in the *E. coli* BL21-Gold(DE3) pQR711 strain appeared to be significantly lower than in the other strains.

3.3.1.3 Reaction profiles

Cultures of *E. coli* BL21-Gold(DE3), JM107, JM109, XL10-Gold, and the corresponding pQR711 strains were grown, diluted to the same final OD_{600}, and sonicated. Aliquots of these sonicates were incubated with the substrates β-HPA and glycolaldehyde (both 100mM) at 25°C and pH 7.5. Over the

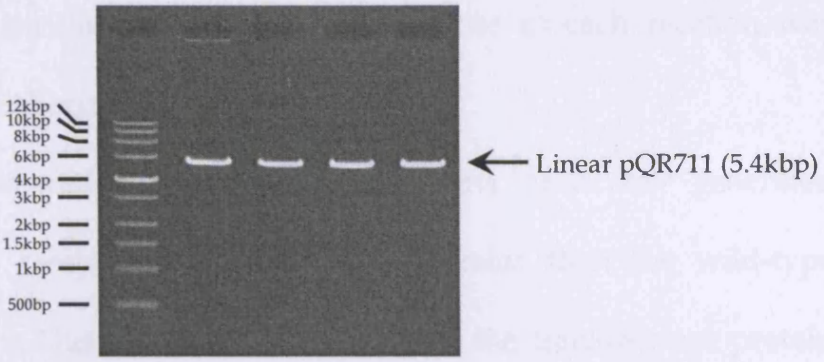


Figure 3.2. 0.7% agarose gel electrophoresis of linearised plasmid DNA extracts. Sample lanes are *E. coli* JM107 pQR711 ("J7p"), BL21-Gold(DE3) pQR711 ("Bp"), JM109 pQR711 ("J9p"), and XL10-Gold pQR711 ("Xp"). Lane "M" contains marker DNA (0.5–12kbp). The 5.4kbp band in all of the sample lanes corresponds to linear pQR711 DNA.

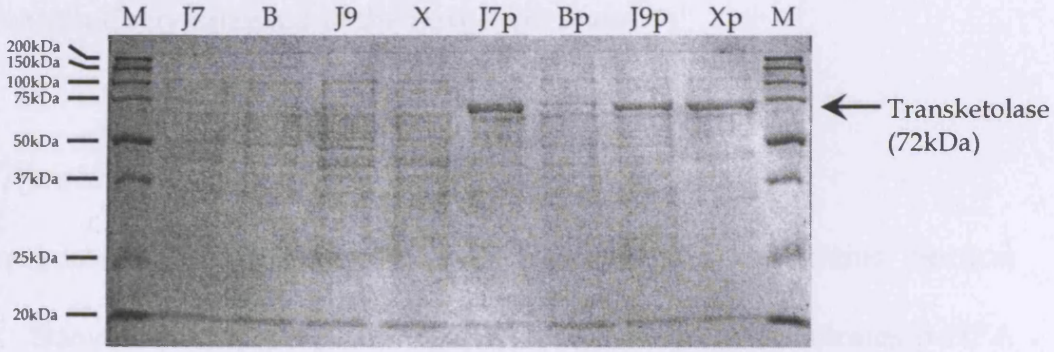


Figure 3.3. SDS-PAGE analysis of the new strains. Sample lanes are *E. coli* BL21-Gold(DE3) ("B"), JM107 ("J7"), JM109 ("J9"), XL10-Gold ("X"), BL21-Gold(DE3) pQR711 ("Bp"), JM107 pQR711 ("J7p"), JM109 pQR711 ("J9p"), and XL10-Gold pQR711 ("Xp"). "M" lanes contain protein standards (20–250kDa). The 70kDa band in the J7p, Bp, J9p, and Xp lanes corresponds to the transketolase monomer.

course of 90 minutes the synthesis of L-erythrulose in each reaction was monitored by HPLC (Section 3.2.1.3).

The data showed that the pQR711 strains of *E. coli* generated considerably more L-erythrulose from the substrates than the wild-type strains (Figure 3.4). This observation proved that the transketolase protein being overexpressed by the pQR711 strains was functional. The highest activity was observed in the *E. coli* XL10 pQR711 sonicate; the JM107 pQR711 and JM109 pQR711 sonicates had comparable activities, while the BL21-Gold(DE3) pQR711 sonicate had a substantially lower one. These activities were quantitatively assessed in the next experiment.

3.3.1.4 Transketolase activity

Initial reaction velocity data were collected for the eight strains (Section 3.2.1.4). Sonicates were prepared and incubated with the substrates β-HPA and glycolaldehyde (both 100mM) at 25°C and pH 7.5. The transketolase concentrations in the wild-type reactions were equal to 1.50ODU culture sonicates and the transketolase concentrations in the pQR711 reactions were equal to 0.25ODU culture sonicates. The *E. coli* JM107 pQR711, JM109 pQR711, and XL10-Gold pQR711 reactions were quenched after five minutes of incubation and the remaining reactions were stopped after one hour. The L-erythrulose concentration in each reaction was measured by HPLC and the corresponding transketolase activity was computed (Figure 3.5). Transketolase activity was calculated as "mkat.m^{-3} of a 0.25ODU culture

Strain	L-Erythrulose conc. (mM) at time (minutes)					
	0.5	10	20	30	60	90
Wild-type strains						
JM107	0.00	0.00	0.00	0.89	1.58	2.64
BL21-Gold(DE3)	0.00	0.00	0.83	1.26	2.95	4.76
JM109	0.00	0.00	0.71	0.96	1.71	2.52
XL10-Gold	0.00	0.00	0.78	1.42	2.68	4.47
pQR711 strains						
JM107 pQR711	1.68	33.98	62.55	89.59	96.72	99.04
BL21-Gold(DE3) pQR711	0.00	5.42	11.74	18.38	36.37	53.59
JM109 pQR711	1.60	32.27	74.19	97.50	99.95	97.77
XL10-Gold pQR711	3.27	92.70	96.74	97.86	98.80	98.65

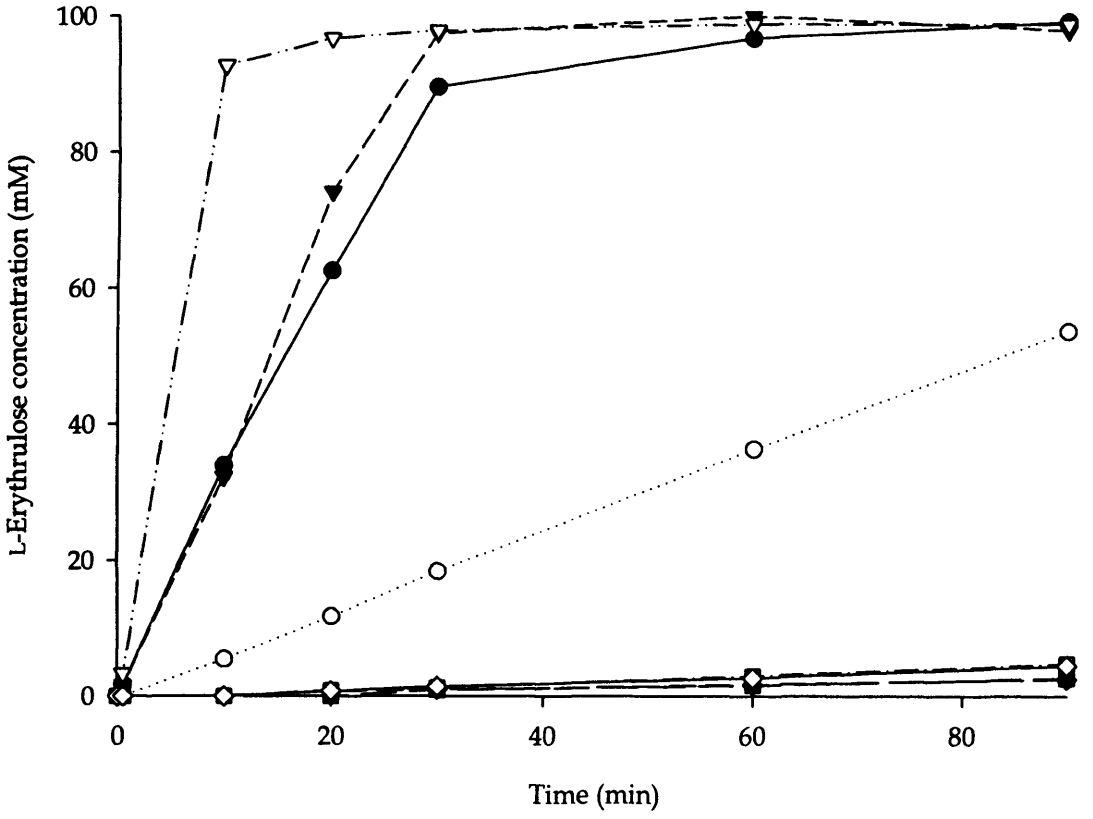


Figure 3.4. Bioconversion of β-HPA and glycolaldehyde to L-erythrulose (and carbon dioxide) by sonicates of the new strains. 100mM each of the two substrates were incubated with sonicates of *E. coli* JM107 pQR711 (●), BL21-Gold(DE3) pQR711 (○), JM109 pQR711 (▼), XL10-Gold pQR711 (▽), and the corresponding wild-type strains. The concentration of transketolase in each reaction was equal to that of a sonicated 1.50ODU culture. The control reaction did not produce detectable L-erythrulose.

Strain	Transketolase activity ($mkat.m^{-3}$)				
	A	B	C	Mean	σ_M
Wild-type strains					
JM107	0.089	0.086	0.085	0.087	0.0011
BL21-Gold(DE3)	0.106	0.129	0.132	0.122	0.0083
JM109	0.084	0.087	0.092	0.088	0.0025
XL10-Gold	0.139	0.137	0.134	0.137	0.0015
pQR711 strains					
JM107 pQR711	5.874	5.664	5.700	5.746	0.0649
BL21-Gold(DE3) pQR711	1.046	1.026	1.068	1.047	0.0123
JM109 pQR711	5.669	5.440	5.728	5.612	0.0879
XL10-Gold pQR711	12.967	12.744	13.416	13.042	0.1976

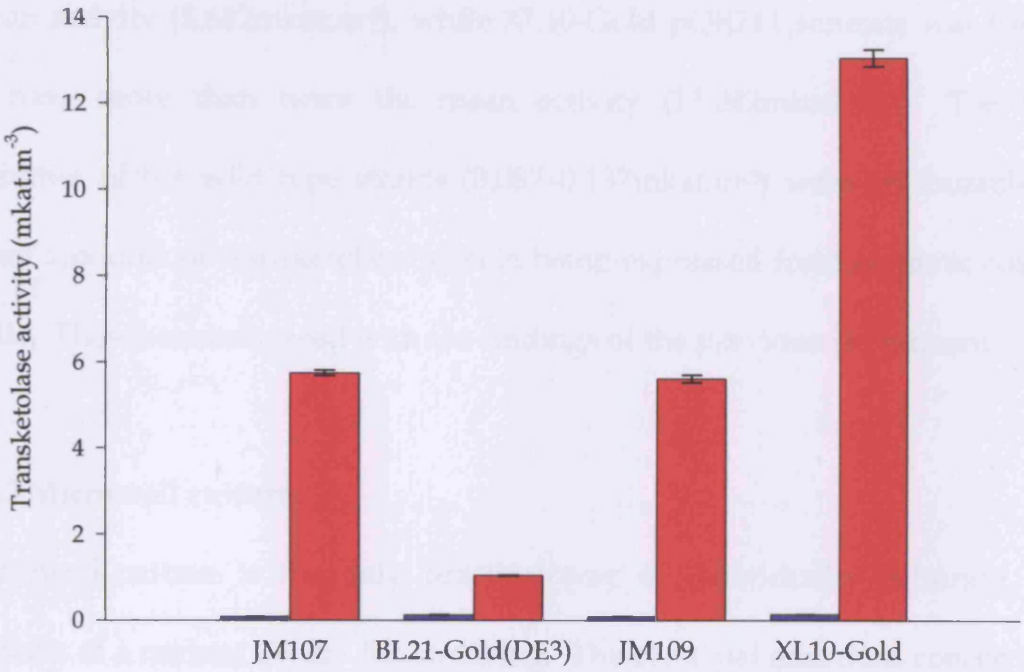


Figure 3.5. Transketolase activity in sonicates of the new strains. Cultures of *E. coli* JM107, BL21-Gold(DE3), JM109, XL10-Gold, and the corresponding pQR711 strains were sonicated and analysed for transketolase activity. The catalysed reaction was 100mM β-HPA and 100mM glycolaldehyde to L-erythrulose (and carbon dioxide). The activity values correspond to reactions containing the same concentration of transketolase as 0.25ODU culture lysates. Wild-type columns are on the left (coloured blue) and pQR711 columns are on the right (coloured red). The error bars correspond to $\pm\sigma_M$. The control reaction did not produce detectable L-erythrulose.

sonicate" for two reasons: (a) to adjust for the concentration of sonicate in the reaction (six times more dilute in the pQR711 reactions); and (b) to adjust for the length of the incubation period (six times shorter for the *E. coli* JM107 pQR711, JM109 pQR711, and XL10-Gold pQR711 reactions).

E. coli BL21-Gold(DE3) pQR711 sonicate was found to have a mean activity approximately six-fold lower (1.047mkat.m^{-3}) than JM107 pQR711 sonicate (5.746mkat.m^{-3}). *E. coli* JM109 pQR711 sonicate had a comparable mean activity (5.612mkat.m^{-3}), while XL10-Gold pQR711 sonicate was found to have more than twice the mean activity (13.042mkat.m^{-3}). The low activities of the wild-type strains (0.087–0.137mkat.m^{-3}) were attributable to small amounts of transketolase protein being expressed from genomic copies of *tkt*. These results agreed with the findings of the previous experiment.

3.3.2 Microwell culture

Microwell culture is the only practical way of individually culturing the variants of a mutant library for screening. The principal questions concerning the application of microwell culture in this project were: (a) could adequate cell densities be achieved?; (b) could cultures be grown to consistent densities across a plate?; and (c) was there any danger of external- or cross-contamination?

Microwell culture was studied in two plate formats: 96 round-well and 384 square-well (Section 3.2.2). The wells of three plates of each format were filled with LB Amp$^+$ medium and inoculated with individual colonies of *E.*

coli XL10-Gold pQR711. Each plate was sealed by taking an identical (but empty) plate, inverting it, and taping it over the top. This method of sealing prevented evaporation, but trapped sufficient air above each culture to permit aerobic growth. The sealed plates were incubated for 16 hours at various agitation speeds in a 37°C incubator.

The growth of bacteria across the plate was found to be uniform in all cases (a typical result is illustrated in Figures 3.6 and 3.7). Pooling the well cultures from each plate and measuring the OD_{600} in a cuvette revealed that the highest cell densities were obtained in 96-well plate cultures (Figure 3.8). Final OD_{600} increased with agitation speed from 4.14ODU at 600rpm, through 4.15ODU at 1000rpm, to 4.53ODU at 1400rpm. This latter value was close to that achieved in a shake flask (4.86ODU). The final OD_{600}s of the cultures in the 384-well plates were found to be lower: 1.48ODU at 600rpm, 1.91ODU at 1000rpm, and 3.41ODU at 1400rpm. Again, turbidity was found to increase with agitation speed.

Further experiments revealed that the risk of a microwell culture being contaminated by external sources or neighbouring wells was virtually nil.

3.3.3 Optimisation of cofactors

To determine the optimum cofactor concentrations for transketolase activity aliquots of an *E. coli* XL10-Gold pQR711 sonicate were incubated with the substrates β-HPA and glycolaldehyde (both 100mM), and variable concentrations of Mg^{2+} (0–32mM) and ThDP (0–8mM) at 25°C and pH 7.5

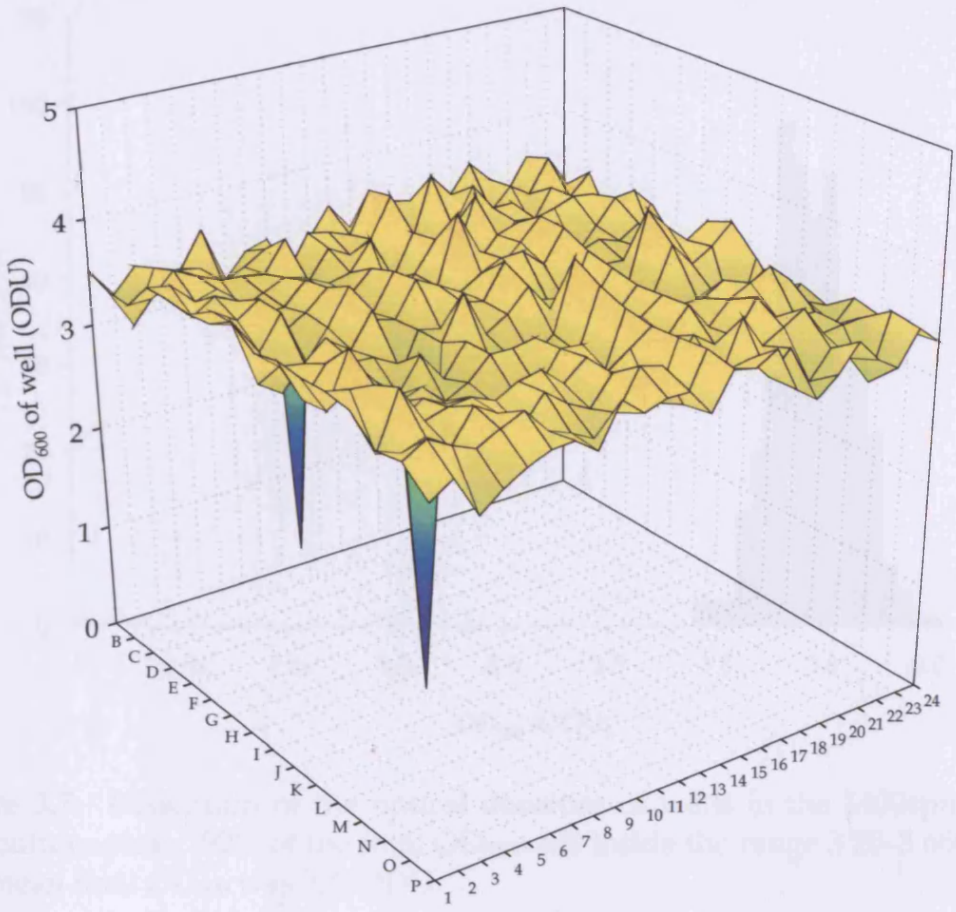


Figure 3.6. Uniform growth across the 1400rpm 384-well culture plate. There was no discernable pattern to the growth. Wells A10 and I9 failed to grow.

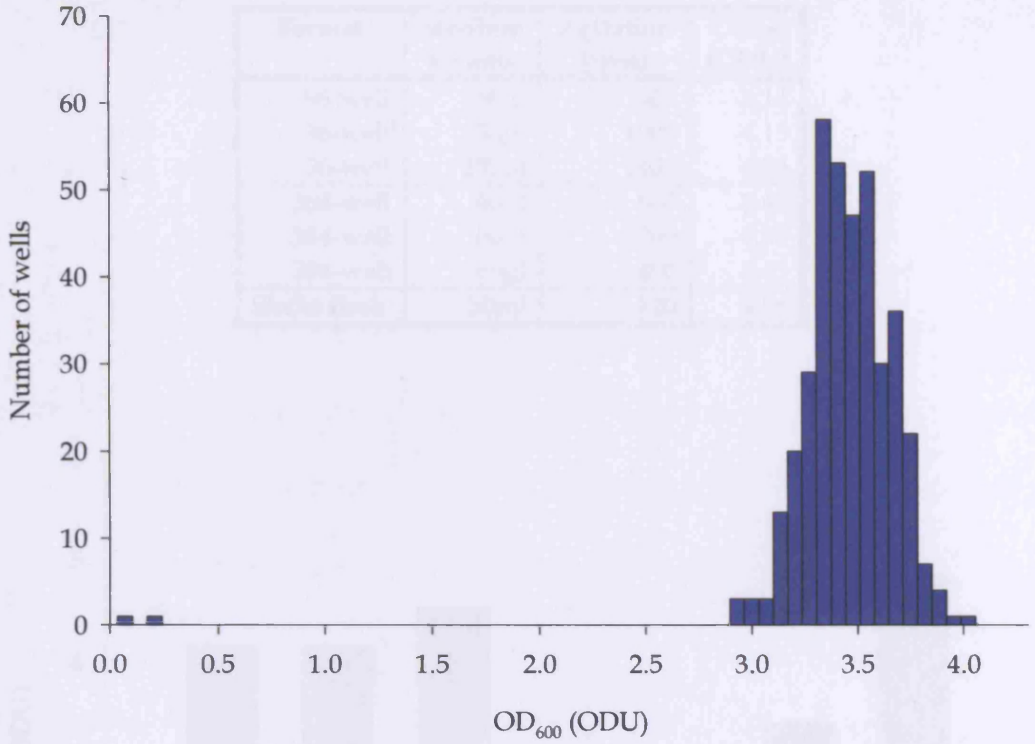


Figure 3.7. Histogram of the optical densities of wells in the 1400rpm 384-well culture plate. 90% of the final OD_{600}s fell inside the range 3.20–3.66ODU. The mean final OD_{600} was 3.41ODU.

Format	Medium volume	Agitation (rpm)	OD_{600} (ODU)
96-well	150µl	600	4.14
96-well	150µl	1000	4.15
96-well	150µl	1400	4.53
384-well	60µl	600	1.48
384-well	60µl	1000	1.91
384-well	60µl	1400	3.41
Shake flask	50ml	220	4.86

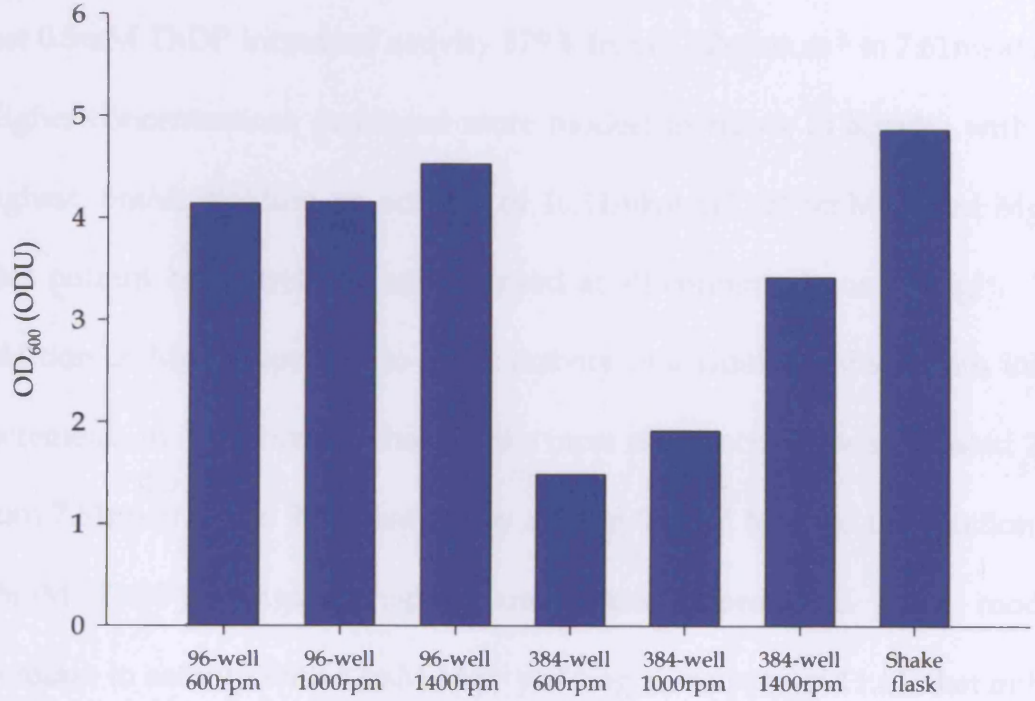


Figure 3.8. The effects of plate format and agitation speed on the OD_{600} of pooled microwell cultures.

(Section 3.2.3). The concentration of transketolase in each reaction was equal to a 0.25ODU culture sonicate. The reactions were quenched after ten minutes and analysed for L-erythrulose. Transketolase activity values were then calculated.

The results showed that transketolase activity was dramatically enhanced by the addition of ThDP (Figure 3.9). At 0mM added Mg^{2+}, adding just 0.5mM ThDP increased activity 579% from 1.12mkat.m^{-3} to 7.61mkat.m^{-3}. Higher concentrations produced more modest increases in activity with the highest, 8mM, yielding an activity of 10.11mkat.m^{-3} (at 0mM added Mg^{2+}). This pattern of behaviour was observed at all concentrations of Mg^{2+}. The addition of Mg^{2+} appeared to affect activity in a similar fashion with initial increments in concentration having the most effect: activity was boosted 24% from 7.61mkat.m^{-3} to 9.44mkat.m^{-3} by adding 0.5mM Mg^{2+} to the reaction (at 0.5mM ThDP). Again, higher concentrations produced more modest increases in activity with 32mM Mg^{2+} yielding an activity of 11.43mkat.m^{-3} (at 0.5mM added ThDP). This pattern of behaviour was observed at all concentrations of ThDP except at 0mM (added).

The highest transketolase activity was observed when 8mM ThDP and 32mM Mg^{2+} were added (12.91mkat.m^{-3}).

3.3.4 Lytic methods

The following methods were compared for their effectiveness at releasing active transketolase from *E. coli* XL10-Gold pQR711: sonication, a proprietary

ThDP (mM)	Transketolase activity (mkat.m^{-3}) at Mg^{2+} conc. (mM)							
	0.0	0.5	1.0	2.0	4.0	8.0	16.0	32.0
0.0	1.12	1.13	1.13	1.13	1.14	1.18	1.18	1.17
0.5	7.61	9.44	10.19	10.80	11.09	11.30	11.45	11.43
1.0	8.92	10.24	10.83	11.40	11.73	11.94	11.93	11.90
2.0	9.72	10.62	11.30	11.70	12.00	12.18	12.37	12.40
4.0	10.15	10.86	11.40	11.94	12.25	12.50	12.79	12.85
8.0	10.11	10.79	11.32	11.89	12.31	12.65	12.88	12.91

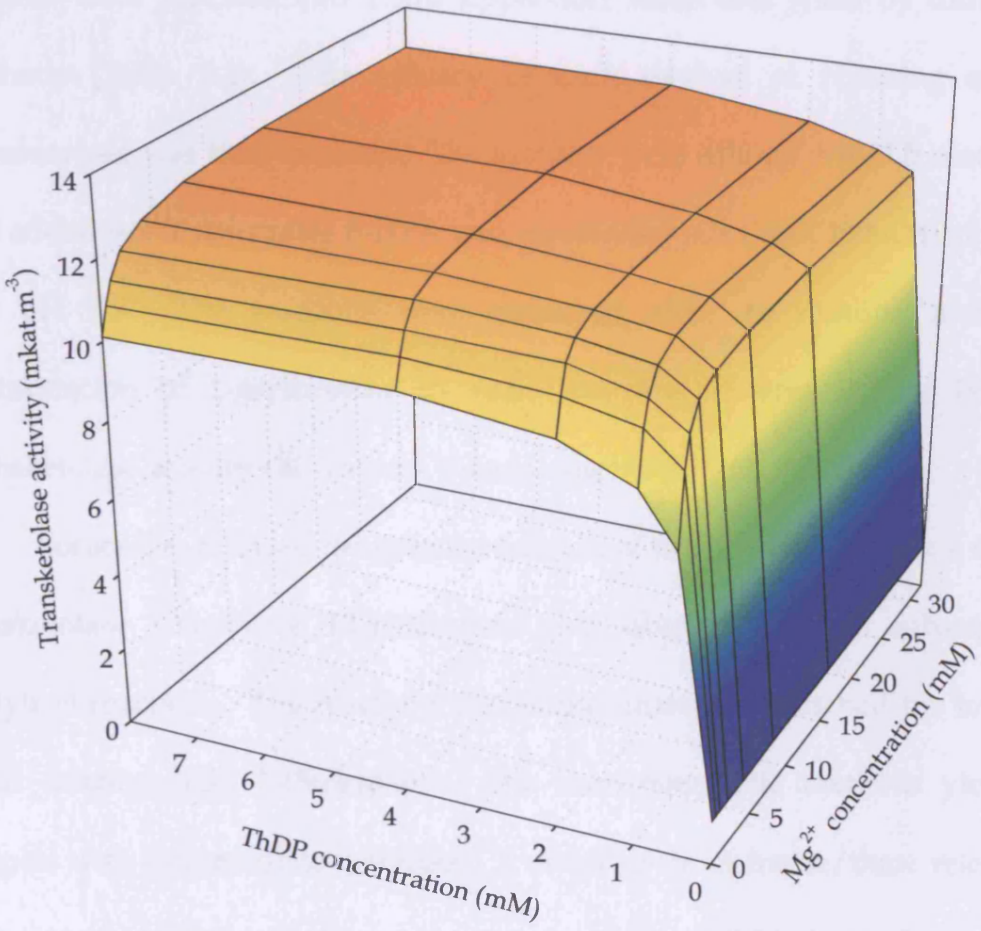


Figure 3.9. The effects of cofactors on transketolase activity. A culture of *E. coli* XL10 pQR711 was sonicated and analysed for transketolase activity at different concentrations of ThDP and Mg^{2+}. The catalysed reaction was 100mM β-HPA and 100mM glycolaldehyde to L-erythrulose (and carbon dioxide). The concentration of free transketolase in each reaction was equal to that of a 0.25ODU culture sonicate. The control reactions did not produce detectable L-erythrulose.

lysing reagent, and repeated cycles of freezing and thawing (Section 3.2.4).

A shake flask culture of *E. coli* XL10-Gold pQR711 was prepared and diluted with LB medium to an OD_{600} of 3.41ODU: equal to the mean final optical density previously observed for a microwell culture grown in a 384-well plate (incubated at 37°C and 1400rpm for 16 hours) (Figure 3.8). 1ml aliquots were pipetted into 1.5ml Eppendorf tubes and lysed by different methods (Table 3.2). The efficacy of each method at releasing active transketolase was then assessed. The samples were diluted with LB medium and added to the substrates β-HPA and glycolaldehyde (both 100mM) at 25°C and pH 7.5. The reactions were quenched after ten minutes and the concentration of L-erythrulose in each one was determined by HPLC. Transketolase activity values were then calculated (Figure 3.10).

Sonication released the greatest amount of active transketolase: a mean transketolase activity of 12.40mkat.m^{-3} was observed in the subsequent analytical reactions. The reactions containing untreated cells had the lowest mean activity: only 7.49mkat.m^{-3}. The remaining lytic methods yielded samples with intermediate activities. A single cycle of freeze/thaw released sufficient transketolase to boost the mean activity to 8.54mkat.m^{-3}. Further cycles increased this activity to 9.45mkat.m^{-3} (two cycles) and 9.41mkat.m^{-3} (three cycles). A mean activity of 8.13mkat.m^{-3} was observed with 10% (v/v) BugBuster Reagent after 30 minutes, rising to 8.95mkat.m^{-3} after one hour. This pattern was also observed with a BugBuster Reagent concentration of 20% (v/v): a mean activity of 8.49mkat.m^{-3} was recorded after 30 minutes,

Method	Transketolase activity (mkat.m^{-3})				
	A	B	C	Mean	σ_M
Sonication	11.98	12.82	12.40	12.40	0.242
10% BB 30min	8.26	7.90	8.22	8.13	0.115
20% BB 30min	8.79	8.41	8.26	8.49	0.158
10% BB 60min	8.39	9.21	9.25	8.95	0.280
20% BB 60min	9.10	8.64	9.33	9.02	0.203
F/T × 1	8.59	8.80	8.24	8.54	0.164
F/T × 2	9.59	9.54	9.22	9.45	0.117
F/T × 3	9.70	9.15	9.37	9.41	0.160
Not treated	7.71	7.48	7.28	7.49	0.123

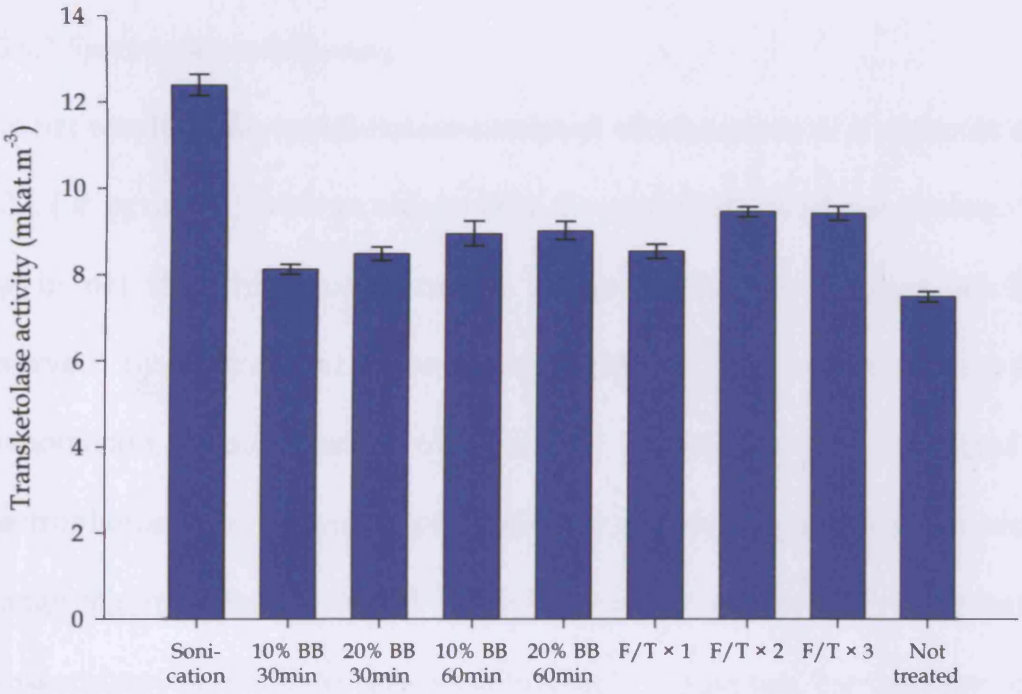


Figure 3.10. Comparison of lytic methods. Aliquots of a 3.41ODU *E. coli* XL10-Gold pQR711 culture were lysed by different methods, diluted, and analysed for transketolase activity (overall dilution 1:13.64). The catalysed reaction was 100mM β-HPA and 100mM glycolaldehyde to L-erythrulose (and carbon dioxide). "*n*% BB" is an abbreviation for "*n*% (v/v) BugBuster HT Protein Extraction Reagent". "F/T × *n*" is an abbreviation for "*n* cycles of freezing at -80°C and thawing at 25°C". The error bars correspond to $\pm\sigma_M$. The control reaction did not produce detectable L-erythrulose.

rising to 9.02mkat.m^{-3} after one hour.

3.3.5 High-throughput assays for transketolase activity

Two alternative approaches to quantifying transketolase activity in a high-throughput manner were investigated: (a) a spectrophotometric assay (Section 3.2.5.1); and (b) a fast HPLC assay (Section 3.2.5.2).

3.3.5.1 Spectrophotometric assay

The net result of the transketolase-catalysed condensation of a molecule of β-HPA (or pyruvate) with an aldehyde is the consumption of one proton. The rise in pH that this process causes in the absence of a buffer has been observed by Mitra and co-workers (1998). It was hoped that this phenomenon would permit transketolase activity to be ascertained by spectrophotometric means: a pH indicator present in the reaction would change colour as the pH shifted. Such an assay would allow several hundred transketolase variants to be simultaneously screened for activity in a transparent 384-well plate. The pH indicator chosen to investigate the potential of this approach was cresol red. The A_{573} of this compound increases with alkalinity from pH 7.5 (it visibly changes in colour from orange to purple).

Aliquots of an *E. coli* JM107 pQR711 sonicate were diluted with LB medium to form a set of three strengths. These diluted sonicates were incubated with 50mM β-HPA, 50mM glycolaldehyde, and 10mg.l^{-1} cresol red

at 25°C and pH 7.5 in the absence of a buffer. The concentration of transketolase in each reaction was equal to either a 0.1875ODU, 0.375ODU, or 0.75ODU culture sonicate. The pH of each reaction was monitored by A_{573} (Figure 3.11).

A_{573} was found to increase in all of the reactions over 60 minutes - including the control. The relationship between time and absorbance was most linear in the first 15 minutes of each reaction so lines of best fit were fitted to these data. The rate of change in A_{573} of the control reaction over this period was calculated to be 1.69mAU.min^{-1}: the baseline rate of change. Deducting this value from the sonicate reactions yielded the following rates of change: 2.33mAU.min^{-1} for the 0.1875ODU reaction, 3.06mAU.min^{-1} for the 0.375ODU reaction, and 4.26mAU.min^{-1} for the 0.75ODU reaction. The observed correlation between rate of change in A_{573} and concentration of transketolase suggested that a spectrophotometric assay for transketolase activity might be possible. However, further experiments with this assay proved it to be highly erratic: results could not be reproduced day to day. In fact, A_{573} was occasionally observed to decrease over time!

3.3.5.2 High-throughput HPLC method

The HPLC method used elsewhere in this thesis to determine transketolase activity (Section 2.3.9) is a slow procedure (≤131 samples per day). To develop a faster method for screening the 300mm analytical column was replaced with a 50mm guard column containing a similar stationary phase

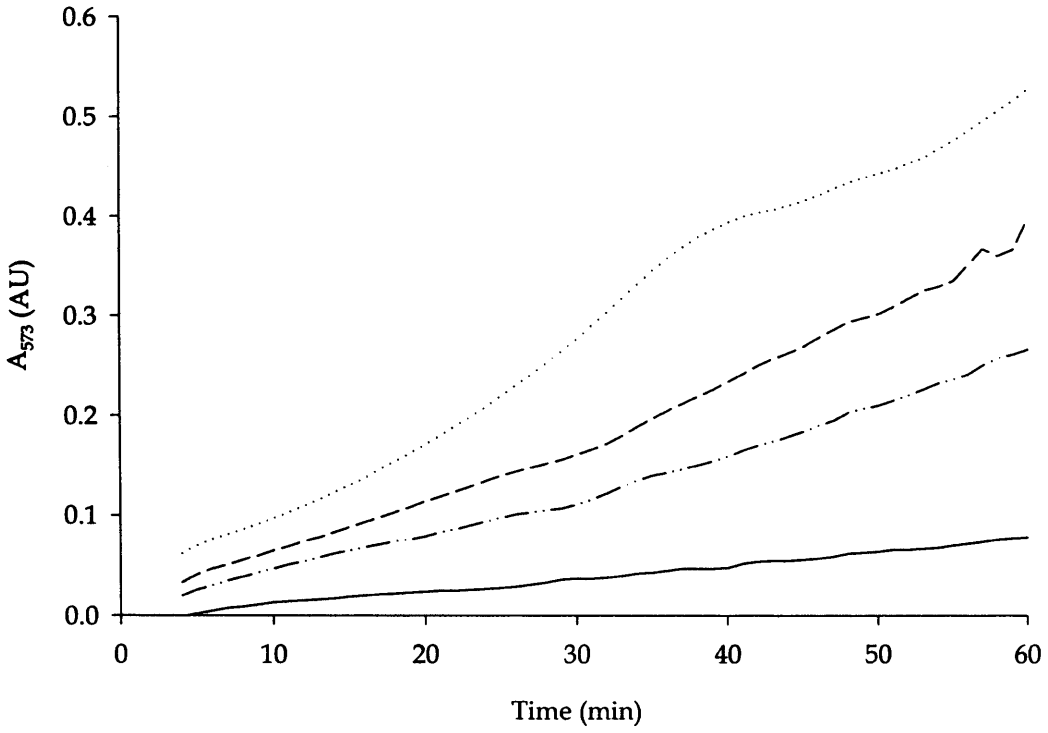


Figure 3.11. High-throughput spectrophotometric assay for transketolase activity. Dilutions of a 3.00ODU *E. coli* JM107 pQR711 sonicate were incubated with 50mM β-HPA and 50mM glycolaldehyde in the absence of a buffer. The pH of each reaction was monitored by measuring the A_{573} of an incorporated pH indicator (cresol red). The data illustrated in the graph correspond to the control reaction (solid) and reactions with the following overall dilutions of sonicate: 1:4 (dotted), 1:8 (dashed), and 1:16 (dash-dot-dot).

(PL Hi-Plex H guard column, Polymer Laboratories Ltd.). The intended role of this column is to protect a longer analytical column from degradation by acting as a trap for highly-retained sample components and particulate matter. It was discovered, however, that it was capable of a modest degree of separation itself.

Mixtures of β-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 (Figure 3.12). Glycolaldehyde was found to have the same retention time as L-erythrulose, but its absorbance at 210nm wavelength was very low. Pyruvate yielded a strong peak at 0.59 minutes: 0.06 minutes later than β-HPA (Figure 3.13). (*S*)-3,4-dihyroxybutan-2-one could not be obtained from any supplier so its retention time could not be determined.[5] It was calculated that the fast analysis time of this HPLC method (1.2 minutes) would permit ≤1200 mutants a day to be screened: 817% faster than the 11 minute protocol.

Calibration curves for L-erythrulose, β-HPA, and pyruvate were found to be relatively linear up to 25mM concentration (Figure 3.14). The ordinate data in these curves is "detector response", not "peak area" as preferred. The PeakNet 5.1 software could not provide reliable peak height or area data for the wide peaks so a simple program was written in Java (Sun Microsystems Inc.) to find these peaks and output the maximum detector response in each case.

[5] It was estimated to be around 0.72 minutes: 0.06 minutes later than L-erythrulose.

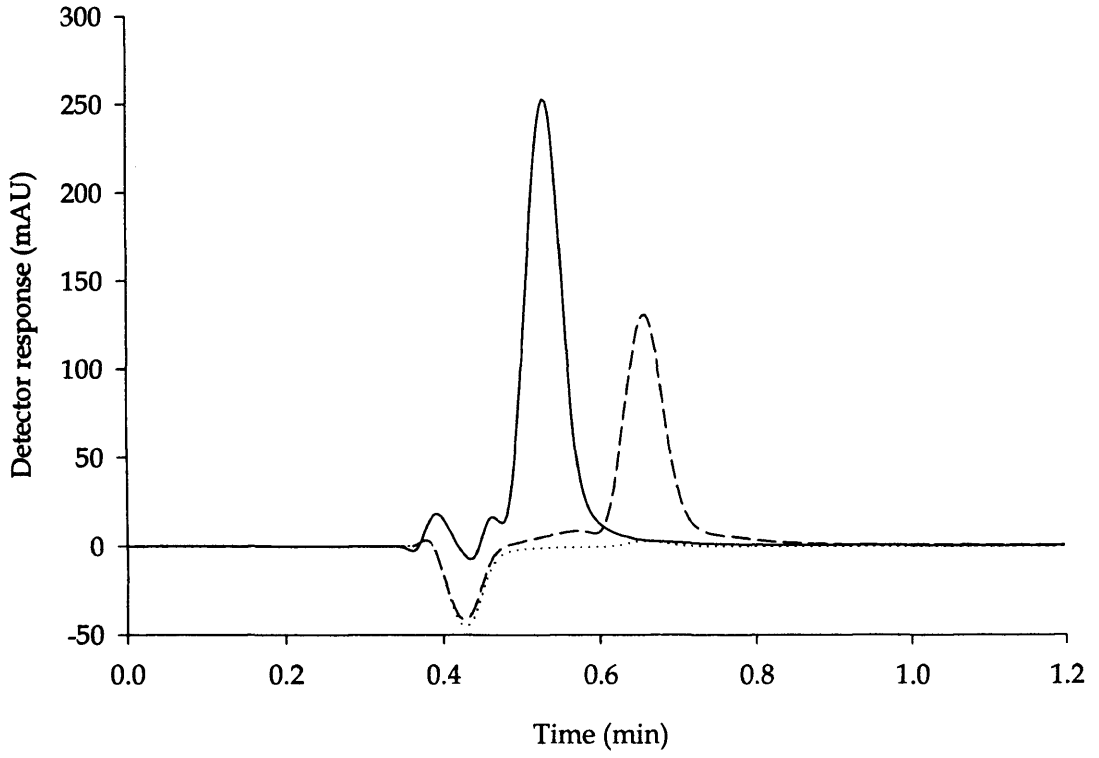


Figure 3.12. High-throughput HPLC analysis of the β-HPA and glycolaldehyde to L-erythrulose (and carbon dioxide) reaction. Chromatograms of 25mM β-HPA (solid), 25mM glycolaldehyde (dotted), and 25mM L-erythrulose (dashed) are superimposed. The retention times of β-HPA and L-erythrulose are 0.53 minutes and 0.66 minutes, respectively. The retention time for glycolaldehyde is 0.66 minutes as well, but its peak is extremely small due to poor absorbance at 210nm wavelength.

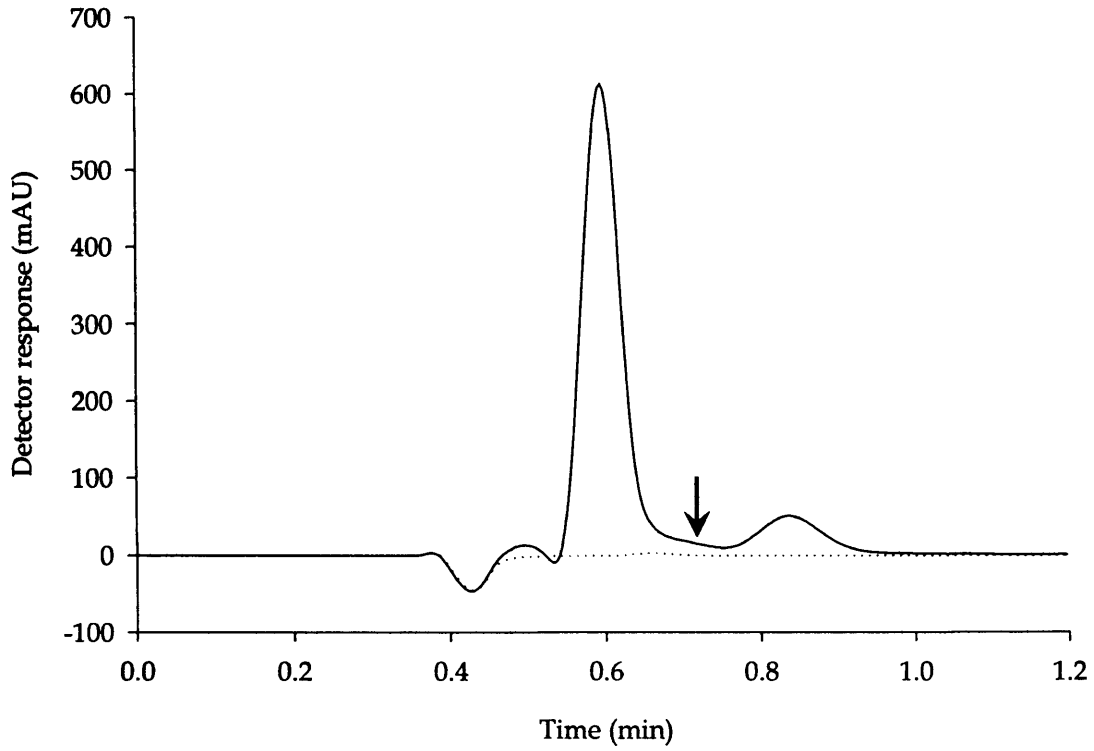


Figure 3.13. High-throughput HPLC analysis of the pyruvate and glycolaldehyde to (*S*)-3,4-dihydroxybutan-2-one (and carbon dioxide) reaction. Chromatograms of 25mM pyruvate (solid) and glycolaldehyde (dotted) are superimposed. The main peak in the pyruvate chromatogram occurs at 0.59 minutes and a smaller peak occurs at 0.84 minutes (an impurity). Glycolaldehyde does not yield an identifiable peak. (*S*)-3,4-dihydroxybutan-2-one could not be purchased; it was anticipated that its peak would occur at 0.72 minutes (arrow).

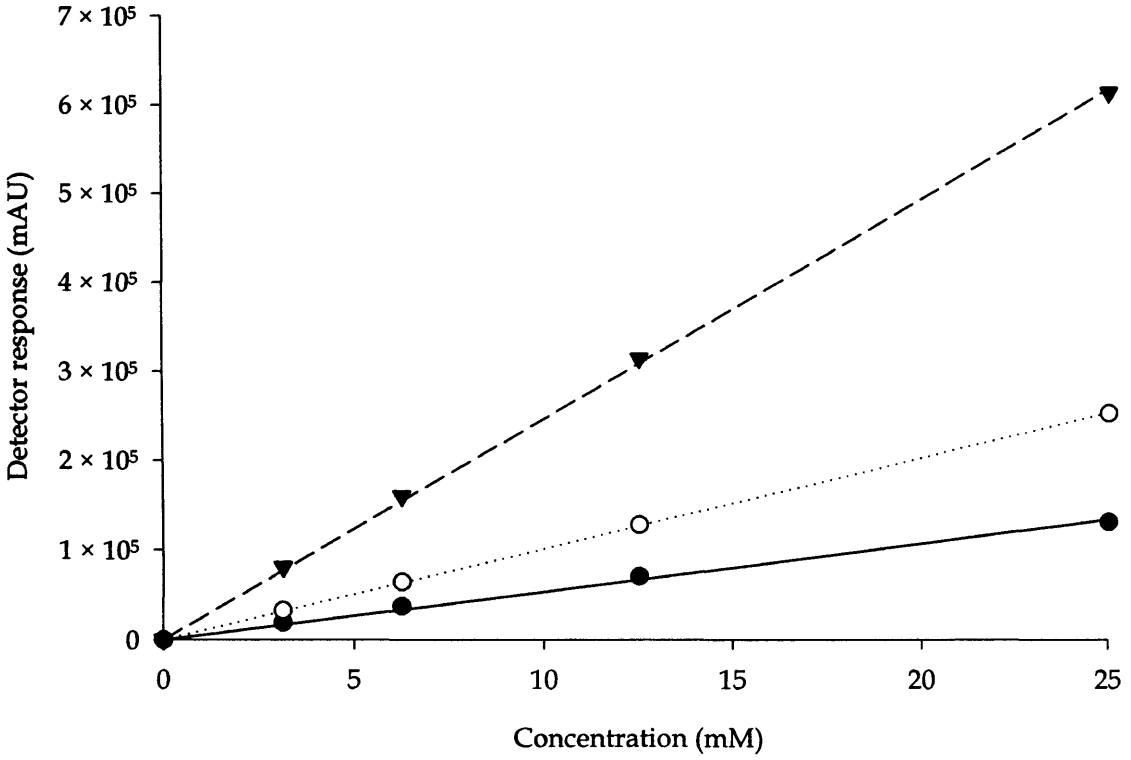


Figure 3.14. Calibration curves for transketolase reaction components separated by high-throughput HPLC. The relationships between concentration and detector response are approximately linear for L-erythrulose (●), β-HPA (○), and pyruvate (▼) with R^2 values of 0.9993, 0.9997, and 0.9993, respectively. Above 25mM the relationships become more curved.

3.4 Discussion

The goal of this project was to isolate mutants of transketolase that were capable of catalysing the transfer of a C_2 moiety from pyruvate to glycolaldehyde, yielding (*S*)-3,4-dihydroxybutan-2-one (and carbon dioxide) (Figure 1.14). The mutagenesis phase of this project would yield an array of pQR711 plasmids with mutations in the *tkt* gene (this library is designated "pQR711tkt$^{\mu}$" in this chapter). In order that members encoding transketolase variants with the desired activity could be identified, this gene library needed to be expressed. *E. coli* was chosen for this role.

Following the transformation of pQR711tkt$^{\mu}$ into *E. coli*, clones expressing active transketolase variants needed to be identified. A selection-based test could not be used as there was no obvious way of linking cell survival to the desired activity. Consequently, a screen was required. It was envisaged that this screen would comprise four steps: (1) the microwell culture of individual *E. coli* colonies; (2) lysis of the cultures; (3) incubation with the target substrates (pyruvate and glycolaldehyde); and finally (4) high-throughput analysis for the target activity.

3.4.1 Selection of host strain

The pQR711 plasmid was transformed into the *E. coli* strains BL21-Gold(DE3), JM109, and XL10-Gold by heat-shock. An agarose gel (Section 3.3.1.1) and an SDS-PAGE analysis (Section 3.3.1.2) confirmed that these transformations had been successful and that transketolase was being overexpressed in all of the

strains.

E. coli JM107 pQR711 (French and Ward, 1995) has been used in a number of experimental biotransformations as the source of transketolase. *E. coli* JM107 is not, however, an expression strain: it is a cloning strain (Table 3.3). Thus, it was expected that transforming the pQR711 plasmid into the expression strain *E. coli* BL21-Gold(DE3) (Table 3.1) would increase transketolase yield. The results of Sections 3.3.1.2 to 3.3.1.4 clearly disprove this hypothesis. *E. coli* BL21-Gold(DE3) pQR711 sonicate was found to have a transketolase activity six times less than the original JM107 pQR711 strain, adjusting for culture turbidity (Section 3.3.1.4). Another cloning strain, *E. coli* XL10-Gold pQR711, yielded the highest transketolase activity: more than double that of the original strain, adjusting for culture turbidity. *E. coli* JM109 pQR711 sonicate was found to have similar transketolase activity to JM107 pQR711 sonicate, but JM107 and JM109 are almost identical genetically (Table 3.3), so this result was not surprising.

As *E. coli* XL10-Gold pQR711 sonicate had the greatest transketolase activity out of the strains tested, adjusting for culture turbidity, it was chosen as the host for the pQR711tkt$^{\mu}$ library.[6]

3.4.2 Microwell culture

The microwell culture method described in Section 3.3.2 can be summarised as follows: (1) the wells of a 96- or 384-well plate are filled with LB Amp^{+}

[6] It was reasoned that using this host would maximise the activities of any active variant cultures.

medium (150µl or 60µl, respectively); (2) a Qpix2 robot is used to inoculate each well of this "culture plate" with colonies of *E. coli* XL10-Gold pQR711; (3) the culture plate is sealed with an identical, inverted plate and taped together; and finally (4) the assembly is taped onto a Variomag Teleshake unit and agitated at speeds of 600–1400rpm for 16 hours in a 37°C environment. Sealing the culture plate with an inverted plate proved to be an effective way of preventing cross-contamination of well cultures and also contamination by external sources. This approach also prevented evaporation of the well cultures and thus ensured uniform growth across every plate. The airspace above each culture was quite limited, but growth appeared to be aerobic. Further experiments would be needed to confirm this.

The well cultures in both 96- and 384-well plates increased in final optical density with agitation speed (600rpm to 1400rpm). This observation can be explained by the increased mixing that occurred at higher speeds: settling of the cells was reduced and oxygen transfer from the headspace was increased. The reduced mixing that occurs in smaller wells explains the low turbidities of 384-well plate cultures compared to 96-well plate cultures. It is likely that the cell density of 384-well plate cultures could be boosted by increased agitation.[7]

Based on the results of this experiment, the following conditions were chosen for the microwell culture of *E. coli* XL10-Gold pQR711tkt$^{\mu}$: 384-well plates and 1400rpm agitation. The average final OD_{600} of well cultures when

[7] 1400rpm was the upper limit for the apparatus used.

grown under these conditions was 3.41ODU. Although slightly higher turbidities (and, therefore, higher yields of transketolase) could have been achieved in 96-well plates, this advantage was offset by the increased handling required for the library (four times the number of culture plates).

3.4.3 Optimisation of cofactors

To ensure that the highest possible activities were observed in the screening reactions, the concentrations of cofactors needed to be optimised. The transketolase activity of *E. coli* XL10-Gold pQR711 sonicate was studied at various concentrations of ThDP and Mg^{2+} to determine these optima (Section 3.3.3).

Transketolase activity was found to dramatically increase in the presence of ThDP. The addition of just 0.5mM ThDP to *E. coli* XL10-Gold pQR711 sonicate had a significant effect: a 579% rise in activity when 0mM Mg^{2+} was present. Higher concentrations of ThDP were found to boost transketolase activity further, but with diminishing gains: a 16-fold increase in ThDP concentration (8mM) yielded an overall rise in activity of only 803% (at 0mM Mg^{2+}). This pattern of behaviour was observed at all concentrations of Mg^{2+}. The addition of 0.5mM Mg^{2+} to the sonicate was found to improve transketolase activity at all concentrations of ThDP, except 0mM. Again, the increases in activity levelled-off with increasing Mg^{2+} concentration: at 0.5mM ThDP, the addition of 0.5mM Mg^{2+} yielded a rise in activity of 24%, while 64-fold more Mg^{2+} (32mM) yielded only a 50% rise.

The huge rises in transketolase activity that accompanied the addition of cofactors suggest that the majority of transketolase protein in the sonicate was in the apo from (unbound to cofactors). Furthermore, it may be concluded that ThDP was much more scarce in the sonicate than divalent cations. This conclusion is supported by four observations: (a) apotransketolase requires equal molar amounts of ThDP and a divalent cation to form the holoenzyme; (b) the K_ds of ThDP and Mg^{2+} were much lower than the cofactor concentrations tested (Sprenger *et al.*, 1995);[8] (c) the addition of ThDP had a much more dramatic effect on the transketolase activity of the sonicate than the addition of Mg^{2+}; and (d) at 0mM ThDP the addition of Mg^{2+} had no effect on transketolase activity (i.e. ThDP concentration was the limiting factor). The near absence of ThDP from the sonicate can be explained by the fact that it was at physiological concentration inside the *E. coli* cells, and therefore at even lower concentration in the bulk sonicate. Divalent cations were present in the sonicate courtesy of the growth medium. Yeast extract typically contains ≤0.10% (w/w) Mg^{2+} and ≤0.05% (w/w) Ca^{2+} (Merck, 2004), equating to ≤0.206mM and ≤0.062mM, respectively, in the growth medium.

The greatest transketolase activity was observed when 8mM ThDP and 32mM Mg^{2+} were added to the reaction (12.91mkat.m^{-3}). However, ≥90% of this activity was achieved with far lower concentrations of cofactors: ≥1mM ThDP and ≥2mM Mg^{2+}.

[8] For ThDP concentrations >0.1mM the K_d of Mg^{2+} is <8.7μM. For Mg^{2+} concentrations >0.1mM the K_d of ThDP is <8μM.

It was anticipated that the screening reactions would contain microwell culture lysate and screening solution (substrates, cofactors, and buffer) in equal amounts (1:1). The concentration of mutant transketolase in each reaction would be equal to a 1.705ODU culture lysate (3.41ODU/2 = 1.705ODU): 6.82-fold greater than that used in this experiment. Theoretically, this increased transketolase concentration should not require higher cofactor concentrations to maintain 90% activity as it is still very small in comparison (micromolar versus millimolar). Nevertheless, the cofactor concentrations were scaled up for the screening reactions as a precaution (7mM ThDP and 14mM Mg^{2+}).

3.4.4 Lytic methods

To determine the best method for lysing the *E. coli* XL10-Gold pQR711tkt$^{\mu}$ microwell cultures, aliquots of a 3.41ODU *E. coli* XL10-Gold pQR711 culture were subjected to various lytic treatments and tested for transketolase activity (Section 3.3.4).

Sonication was found to be the most effective method at releasing active transketolase: the subsequent reactions had the greatest mean transketolase activity. Untreated cells had the lowest mean transketolase activity: 60% of the sonication value. The remaining methods yielded lysates with intermediate activities. The mean transketolase activity of cultures treated to a single cycle of freeze/thaw was found to be 69% of the sonication value. A second cycle increased this activity to 76%, but a third cycle yielded

no further improvement. Incubating the *E. coli* XL10-Gold pQR711 culture with 10% (v/v) BugBuster Reagent at room temperature yielded a mean transketolase activity of 66% after 30 minutes, rising to 72% after one hour. Similarly, incubating the culture with 20% (v/v) BugBuster Reagent yielded a mean transketolase activity of 68% after 30 minutes, rising to 73% after one hour. It was concluded that BugBuster Reagent released transketolase in a time-dependent manner, with greater amounts being released over time.

As sonication lysed the cells most effectively, it would have made the best choice for lysing the *E. coli* XL10-Gold pQR711tkt$^{\mu}$ library. However, the available equipment was too slow and unwieldy for high-throughput application. Incubating cells with BugBuster Reagent yielded only small increases in mean transketolase activity over that of untreated cells, so the extra effort (pipetting) that would have been required in the screening phase was not justifiable. Subjecting culture aliquots to cycles of freezing and thawing was not as labour-intensive, but did require significant amounts of time for the freezing periods. The transketolase activity yielded by a single freeze/thaw cycle was considered the best possible return for time and effort. Consequently, this method was chosen for processing the *E. coli* XL10-Gold pQR711tkt$^{\mu}$ library.

3.4.5 High-throughput assays for transketolase activity

The experiment described in Section 3.3.5.1 demonstrated a relationship between rate of A_{573} change and transketolase concentration in unbuffered

reactions containing β-HPA, glycolaldehyde, cofactors, and the pH indicator cresol red. Unfortunately, repetitions of this experiment proved that the assay was highly unreliable. A possible explanation for this behaviour is that the number of protons provided by dissolved carbon dioxide in each sample was significant compared to the number consumed by the reaction. The following experiments should be performed to understand this variability better: (a) monitor one or more reactions by A_{573} and HPLC analysis (Section 2.3.9) simultaneously; and (b) monitor the A_{573} and dissolved carbon dioxide concentration[9] of one or more reactions simultaneously. The first experiment would demonstrate whether or not transketolase-mediated condensation was actually occurring in the unbuffered samples and the second would determine whether or not dissolved carbon dioxide was contributing to changes in A_{573}. If the results of these experiments could be used to reduce the variability of the spectrophotometric assay, it would be an extremely powerful tool for screening transketolase activity.

In contrast, the novel HPLC method described in Section 3.3.5.2 proved to a highly reliable screen for transketolase activity. It successfully resolved mixtures of β-HPA (donor substrate) and L-erythrulose (product) in under 1.2 minutes.[10] Although (*S*)-3,4-dihydroxybutan-2-one could not obtained, it was anticipated that mixtures of this compound (the desired product) with pyruvate (the desired donor substrate) would be resolved in a similar manner: they have similar chemical structures to L-erythrulose and β-HPA

[9] An *in situ* probe could be used to do this.
[10] Glycolaldehyde was invisible due to a very low extinction coefficient at 210nm wavelength.

(respectively), lacking a single hydroxyl group in each case. This chemical change caused pyruvate to have a retention time 0.06 minutes greater than β-HPA; thus, it was expected that (*S*)-3,4-dihydroxybutan-2-one would have a retention time approximately 0.06 minutes greater than L-erythrulose. As the precise figure was unknown it was decided that target activity would be initially identified in the screening reactions by pyruvate depletion, rather than product formation. The relationship between pyruvate concentration and detector response (A_{210}) was linear up to 25mM, so it was decided that these screening reactions would be diluted and quenched 4-fold (100mM/4 = 25mM) before analysis.

3.4.6 Screening procedure

The procedure for screening the pQR711tkt$^{\mu}$ library is summarised in Figure 3.15. The processes highlighted in blue have been developed in this chapter to achieve two goals: (a) to maximise the yield of holotransketolase in each screening reaction (and therefore activity); and (b) to allow variants with the desired activity to be identified. Using this protocol a 384-well plate containing 380 screening reactions (plus four controls) could be screened for the target activity in 11.68 hours (4 hours for reaction and 7.68 hours for fast HPLC analysis).

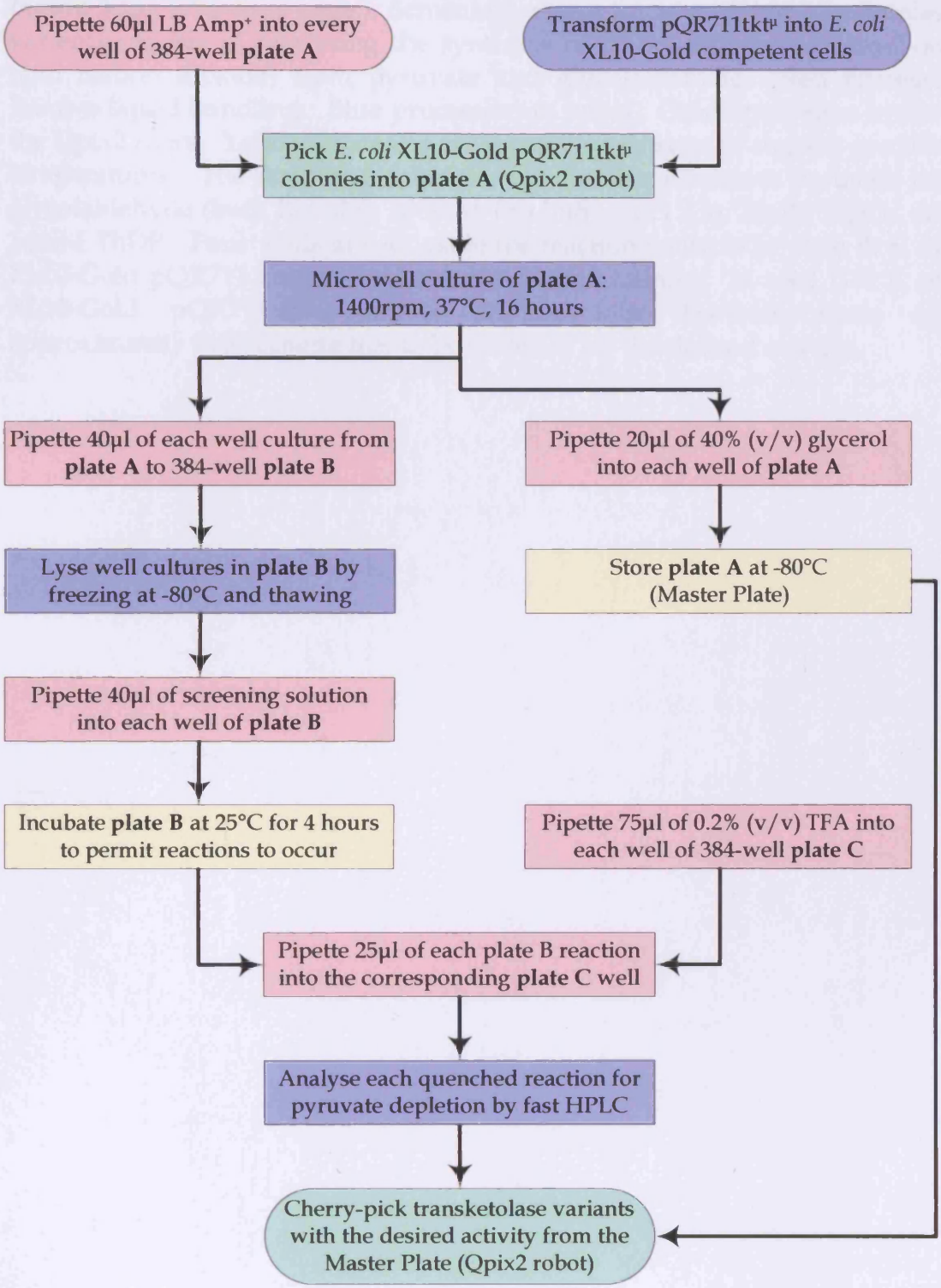

Pipette 60µl LB Amp+ into every well of 384-well plate A
Transform pQR711tktµ into E. coli XL10-Gold competent cells
Pick E. coli XL10-Gold pQR711tktµ colonies into plate A (Qpix2 robot)
Microwell culture of plate A: 1400rpm, 37°C, 16 hours
Pipette 40µl of each well culture from plate A to 384-well plate B
Pipette 20µl of 40% (v/v) glycerol into each well of plate A
Lyse well cultures in plate B by freezing at -80°C and thawing
Store plate A at -80°C (Master Plate)
Pipette 40µl of screening solution into each well of plate B
Incubate plate B at 25°C for 4 hours to permit reactions to occur
Pipette 75µl of 0.2% (v/v) TFA into each well of 384-well plate C
Pipette 25µl of each plate B reaction into the corresponding plate C well
Analyse each quenched reaction for pyruvate depletion by fast HPLC
Cherry-pick transketolase variants with the desired activity from the Master Plate (Qpix2 robot)

Figure 3.15. (Previous page.) Screening system for identifying transketolase variants capable of catalysing the synthesis of (*S*)-3,4-dihydroxybutan-2-one (and carbon dioxide) from pyruvate and glycolaldehyde. Red processes involve liquid handling. Blue processes are novel. Green processes involve the Qpix2 robot. Yellow processes are incubation or storage steps at specified temperatures. The screening solution contains the substrates pyruvate and glycolaldehyde (both 200mM), 100mM Tris buffer (pH 7.5), 28mM $MgCl_2$, and 14mM ThDP. Four wells are set aside for reaction controls in plate B: *E. coli* XL10-Gold pQR711 culture replaces the variant culture. In total, 378 *E. coli* XL10-Gold pQR711tkt$^{\mu}$ colonies require four 384-well plates and approximately 1600 pipette tips to be screened for the desired activity.

Chapter 4 - Library design

4.1 Introduction

Meshalkina and co-workers (1995) have demonstrated that 1-hydroxyethyl ThDP, unlike 1,2-dihydroxyethyl ThDP, is not a viable reaction intermediate for transketolase-mediated condensation. Rather than being transferred to the acceptor substrate, the 1-hydroxyethyl moiety is split from the ThDP molecule and released as a free aldehyde (acetaldehyde). 1-hydroxyethyl ThDP is the intermediate of the target reaction (Figures 1.13 and 4.1), so this activity had to be eliminated from transketolase for the project goal to be achieved.

The second engineering challenge was to modify transketolase so that it was capable of synthesising 1-hydroxyethyl ThDP from pyruvate and ThDP in the first place. Wild-type transketolase does not exhibit this activity, although it is observed in other ThDP-dependent enzymes such as acetolactate synthase (EC 2.2.1.6), 1-deoxy-D-xylulose-5-phosphate (DXP) synthase (EC 2.2.1.7), pyruvate decarboxylase (EC 4.1.1.1), and pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1).

It was proposed that these two refinements could be brought about by modification of the residues in the active site of transketolase that lie close to the C2-hydroxyl group of the reaction intermediate. One possible route to these changes would be through rational protein engineering of the active site using the protein sequences and/or crystal structures of acetolactate synthase, DXP synthase, pyruvate decarboxylase, and pyruvate dehydrogenase (lipoamide) as maps. However, the difficulties associated with rational

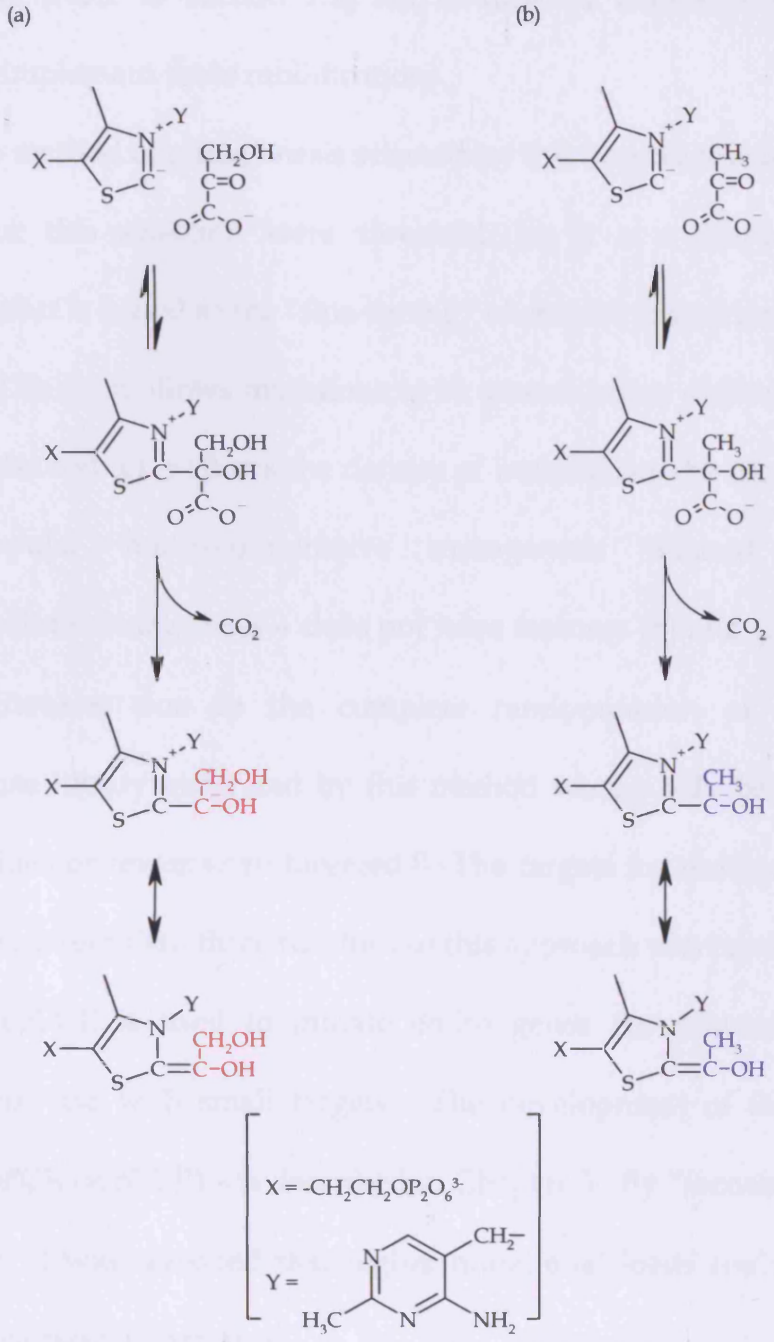


Figure 4.1. Formation of α-carbanion/enamine compounds from ThDP and (a) β-HPA or (b) pyruvate. The α-carbanion/enamine formed in (a) is 1,2-dihydroxyethyl ThDP (the 1,2-dihydroxyethyl group is coloured red) and in (b) is 1-hydroxyethyl ThDP (the 1-hydroxyethyl group is coloured blue). Mechanism (b) is observed in several ThDP-dependent enzymes including acetolactate synthase, DXP synthase, pyruvate decarboxylase, and pyruvate dehydrogenase (lipoamide).

engineering (refer to Section 1.2) led to directed evolution being chosen instead to implement these modifications.

The method of mutagenesis selected for this approach was epPCR. The reasons for this selection were threefold: (a) it is a non-recombinative technique that is suited to the "fine-tuning" of enzyme characteristics (refer to Section 1.2.2); (b) it allows mutations to be spread across defined stretches of amino acids; and (c) it allows the density of mutations to be modulated. The other popular non-recombinative mutagenesis method - random oligonucleotide mutagenesis - does not have features (b) and (c) and creates massive libraries due to the complete randomisation of codons. A transketolase library generated by this method would only be screenable if three residues or fewer were targeted.[11] The targets for mutagenesis in this project were larger than three residues so this approach was rejected.

As epPCR is used to mutate entire genes the protocol had to be modified for use with small targets. The development of this method - focused epPCR (fepPCR) - is described in Chapter 5. By "focusing" mutation in this way it was reasoned that higher mutational loads could be applied while maintaining library size.

This chapter describes the results of three activities: (a) the selection of target sites; (b) the design of primers for the fepPCR of these targets; and (c) the estimation of desired mutation rates and library sizes.

[11] The full complement of 20 naturally-occurring amino acids are available at each of the three positions so there are 8000 unique protein sequences ($20^3 = 8000$).

4.2 Methods

4.2.1 Selection of target sites

The PDB structure file 1GPU describes the structure of *S. cerevisiae* transketolase in complex with the reaction intermediate 1,2-dihydroxyethyl ThDP (Fiedler *et al.*, 2002). RasMol 2.7 (Sayle and Milner-White, 1995), a molecular-modelling package, was used to inspect this structure and locate residues in close proximity to the C2-hydroxyl group of the 1,2-dihydroxyethyl moiety.

4.2.2 Primer design

The nucleotide sequence of *tkt* (Sprenger, 1993) was used to devise primers for the PCR-based amplification of target sites A, B, and C. The software package AnnHyb 3.5 (Friard, 2004) was used to ensure that the physical properties of these primers were conducive to PCR.

4.2.3 Estimation of library size and coverage

The sizes of pQR711 libraries with mutations in target A (pQR711A$^{\mu}$), target B (pQR711B$^{\mu}$), target C (pQR711C$^{\mu}$), and all three targets (pQR711ABC$^{\mu}$) were estimated using the formulae detailed in Section 1.2.3.1. The numbers of transformants required for 90% coverage of each library were estimated using the formulae detailed in Section 1.2.3.2.

4.3 Results

4.3.1 Selection of target sites

It was reasoned that the residues closest to the C2-hydroxyl group of 1,2-dihydroxyethyl ThDP were the most likely to be responsible for: (a) splitting 1-hydroxyethyl ThDP into acetaldehyde and ThDP; and (b) not permitting the synthesis of 1-hydroxyethyl ThDP from pyruvate and ThDP in the first place. A structure of *E. coli* transketolase in complex with 1,2-dihydroxyethyl ThDP would have been ideal for identifying these residues. Such a structure did not exist, so the structure of the highly-homologous *S. cerevisiae* transketolase (see Figure 1.2) was used instead (PDB structure file 1GPU). It was anticipated that after the residues of interest had been identified in *S. cerevisiae* transketolase, the corresponding residues in *E. coli* transketolase would be located.

The residues measured to be closest to the C2-hydroxyl group of 1,2-dihydroxyethyl ThDP were, in order of proximity, His103, His 481, His69, Gly116, His30, Asp477, Leu118, and Pro117 (Section 4.2.1) (Table 4.1). Residues His69, Gly116, Pro117, and Leu118 were eliminated as targets for mutagenesis because their side-chains were observed to face away from the hydroxyl group. It was believed that modification of these residues would have little impact on activities (a) and (b). The remaining residues (His30, His103, Asp477, and His481) were chosen as the primary targets for mutagenesis. Flanking residues were incorporated into the targets to permit movement of the primary residues. The overall targets are recorded in Table

Residue	Distance (Å)	Side-chain facing towards C2-hydroxyl?
His30	4.55	Yes
His69	3.40	No
His103	2.76	Yes
Gly116	3.84	No
Pro117	5.93	No
Leu118	5.81	No
Asp477	5.61	Yes
His481	3.33	Yes

Table 4.1. Residues of transketolase in close proximity to the C2-hydroxyl group of 1,2-dihydroxyethyl ThDP. "Distance" is the distance between the oxygen atom of the C2-hydroxyl group and the closest atom of the residue. (Based on PDB structure file 1GPU.)

Residue identity	Primary/ flanking	Residue number	
		S. cerevisiae	*E. coli*
Target A			
Ser	Flanking	28	24
Gly	Flanking	29	25
His	Primary	30	26
Target B			
Gly	Flanking	102	99
His	Primary	103	100
Pro	Flanking	104	101
Target C			
Asp	Primary	477	469
Gly	Flanking	478	470
Pro	Flanking	479	471
Thr	Flanking	480	472
His	Primary	481	473

Table 4.2. Residues in targets A, B, and C. The residue identities are conserved in the *S. cerevisiae* and *E. coli* transketolases. "Primary" residues are those residues closest to the C2-hydroxyl group of 1,2-dihydroxyethyl ThDP. "Flanking" residues sit either-side of the primary residues.

4.2 and depicted in Figure 4.2a. These targets were identified in *E. coli* transketolase by studying its sequence and crystal structure (PDB structure file 1QGD) (Table 4.2 and Figure 4.2b).

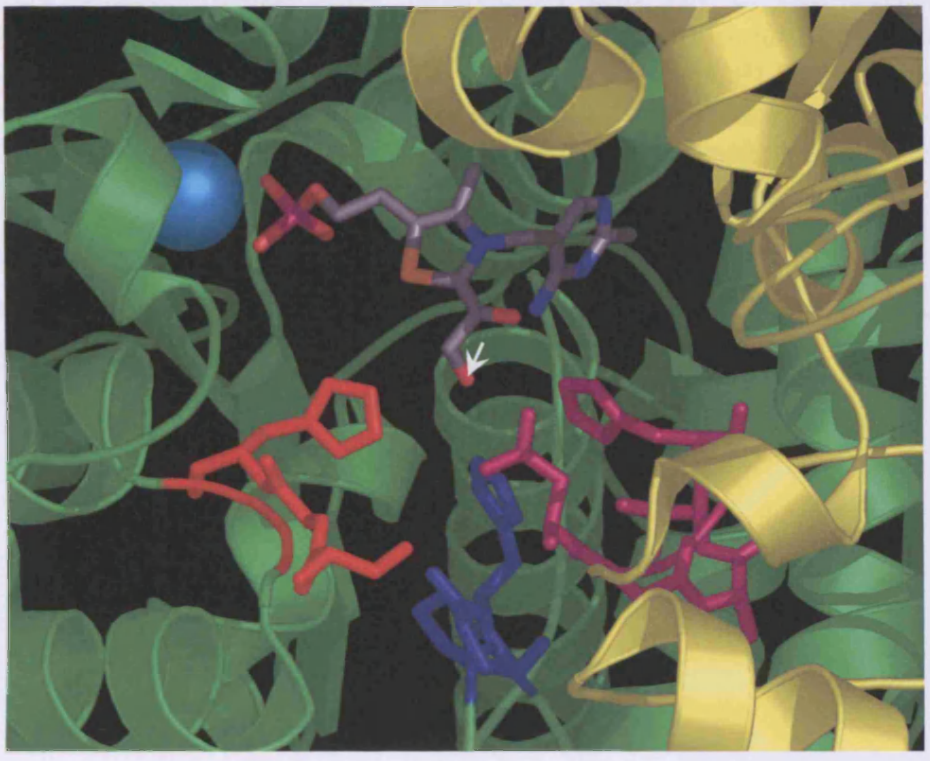
4.3.2 Primer design

Oligonucleotide primers for the fepPCR amplification of the discontinuous target sequences were designed using the *E. coli tkt* sequence (Section 4.2.2) (Figure 4.3). The primers were devised so that each tried to meet the following criteria (Innis *et al.*, 1990): (a) 17–28 bases in length; (b) 50–60% G/C content; (c) melting temperature (T_m) of 55–80°C (target was 60°C); and (d) no significant secondary structure. The sequences and physical properties of the primers are recorded in Table 4.3.

4.3.3 Estimation of library size and coverage

It was believed that a library of pQR711 plasmids with point mutations in the target sites A, B, and C would be the most likely to yield mutants of transketolase with the desired activity (Figure 1.14). It was envisaged that the construction of this library (pQR711ABC$^{\mu}$) would be followed by its transformation into *E. coli* XL10-Gold and subsequent screening (Chapter 3). FepPCR (Chapter 5) mutates bases at random, so the pQR711ABC$^{\mu}$ library would be made up of variants with varying numbers of point mutations, i.e. some would contain no point mutations while others would contain one, two, three or more. In effect, the library would be a mixture of four (or more)

(a) *S. cerevisiae* transketolase in complex with 1,2-dihydroxyethyl ThDP



(b) *E. coli* transketolase in complex with ThDP

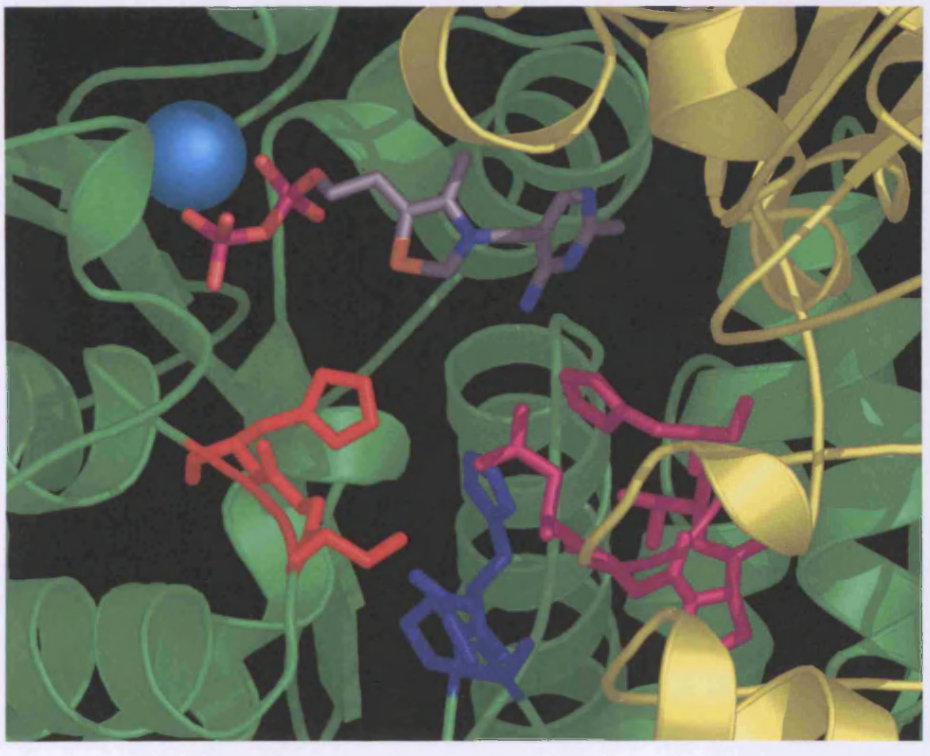


Figure 4.2. (Previous page.) The active sites of *S. cerevisiae* and *E. coli* transketolases. The two diagrams are: (a) *S. cerevisiae* transketolase interacting with 1,2-dihydroxyethyl ThDP and Ca^{2+}; and (b) *E. coli* transketolase interacting with ThDP and Ca^{2+}. In each case one monomer of the enzyme is coloured green, the other is coloured yellow, and the Ca^{2+} ion is a cyan CPK sphere. The 1,2-dihydroxyethyl ThDP and ThDP molecules are shown as "stick" models, as are the target sites (Table 4.2). Target A is coloured red, target B is coloured blue, and target C is coloured magenta. In (a) the oxygen atom of the C2-hydroxyl group of the 1,2-dihydroxyethyl moiety is identified by a white arrow. (Created from the PDB structure files 1GPU and 1QGD. The second phosphate group of 1,2-dihydroxyethyl ThDP was missing from the first file.)

```
1     ATG TCC TCA CGT AAA GAG CTT GCC AAT GCT ATT CGT GCG CTG AGC  45
1     Met Ser Ser Arg Lys Glu Leu Ala Asn Ala Ile Arg Ala Leu Ser  15
                                          Target A
46    ATG GAC GCA GTA CAG AAA GCC AAA TCC GGT CAC CCG GGG GCC CCT  90
16    Met Asp Ala Val Gln Lys Ala Lys Ser Gly His Pro Gly Ala Pro  30

91    ATG GGT ATG GCT GAC ATT GCC GAA GTC CTG TGG CGT GAT TTC CTG  135
31    Met Gly Met Ala Asp Ile Ala Glu Val Leu Trp Arg Asp Phe Leu  45

136   AAA CAC AAC CCG CAG AAT CCG TCC TGG GCT GAC CGT GAC CGC TTC  180
46    Lys His Asn Pro Gln Asn Pro Ser Trp Ala Asp Arg Asp Arg Phe  60

181   GTG CTG TCC AAC GGC CAC GGC TCC ATG CTG ATC TAC AGC CTG CTG  225
61    Val Leu Ser Asn Gly His Gly Ser Met Leu Ile Tyr Ser Leu Leu  75

226   CAC CTC ACC GGT TAC GAT CTG CCG ATG GAA GAA CTG AAA AAC TTC  270
76    His Leu Thr Gly Tyr Asp Leu Pro Met Glu Glu Leu Lys Asn Phe  90
                                          Target B
271   CGT CAG CTG CAC TCT AAA ACT CCG GGT CAC CCG GAA GTG GGT TAC  315
91    Arg Gln Leu His Ser Lys Thr Pro Gly His Pro Glu Val Gly Tyr  105

316   ACC GCT GGT GTG GAA ACC ACC ACC GGT CCG CTG GGT CAG GGT ATT  360
106   Thr Ala Gly Val Glu Thr Thr Thr Gly Pro Leu Gly Gln Gly Ile  120

361   GCC AAC GCA GTC GGT ATG GCG ATT GCA GAA AAA ACG CTG GCG GCG  405
121   Ala Asn Ala Val Gly Met Ala Ile Ala Glu Lys Thr Leu Ala Ala  135

406   CAG TTT AAC CGT CCG GGC CAC GAC ATT GTC GAC CAC TAC ACC TAC  450
136   Gln Phe Asn Arg Pro Gly His Asp Ile Val Asp His Tyr Thr Tyr  150

451   GCC TTC ATG GGC GAC GGC TGC ATG ATG GAA GGC ATC TCC CAC GAA  495
151   Ala Phe Met Gly Asp Gly Cys Met Met Glu Gly Ile Ser His Glu  165

496   GTT TGC TCT CTG GCG GGT ACG CTG AAG CTG GGT AAA CTG ATT GCA  540
166   Val Cys Ser Leu Ala Gly Thr Leu Lys Leu Gly Lys Leu Ile Ala  180

541   TTC TAC GAT GAC AAC GGT ATT TCT ATC GAT GGT CAC GTT GAA GGC  585
181   Phe Tyr Asp Asp Asn Gly Ile Ser Ile Asp Gly His Val Glu Gly  195

586   TGG TTC ACC GAC GAC ACC GCA ATG CGT TTC GAA GCT TAC GGC TGG  630
196   Trp Phe Thr Asp Asp Thr Ala Met Arg Phe Glu Ala Tyr Gly Trp  210

631   CAC GTT ATT CGC GAC ATC GAC GGT CAT GAC GCG GCA TCT ATC AAA  675
211   His Val Ile Arg Asp Ile Asp Gly His Asp Ala Ala Ser Ile Lys  225

676   CGC GCA GTA GAA GAA GCG CGC GCA GTG ACT GAC AAA CCT TCC CTG  720
226   Arg Ala Val Glu Glu Ala Arg Ala Val Thr Asp Lys Pro Ser Leu  240

721   CTG ATG TGC AAA ACC ATC ATC GGT TTC GGT TCC CCG AAC AAA GCC  765
241   Leu Met Cys Lys Thr Ile Ile Gly Phe Gly Ser Pro Asn Lys Ala  255

766   GGT ACC CAC GAC TCC CAC GGT GCG CCG CTG GGC GAC GCT GAA ATT  810
256   Gly Thr His Asp Ser His Gly Ala Pro Leu Gly Asp Ala Glu Ile  270

811   GCC CTG ACC CGC GAA CAA CTG GGC TGG AAA TAT GCG CCG TTC GAA  855
271   Ala Leu Thr Arg Glu Gln Leu Gly Trp Lys Tyr Ala Pro Phe Glu  285

856   ATC CCG TCT GAA ATC TAT GCT CAG TGG GAT GCG AAA GAA GCA GGC  900
286   Ile Pro Ser Glu Ile Tyr Ala Gln Trp Asp Ala Lys Glu Ala Gly  300

901   CAG GCG AAA GAA TCC GCA TGG AAC GAG AAA TTC GCT GCT TAC GCG  945
301   Gln Ala Lys Glu Ser Ala Trp Asn Glu Lys Phe Ala Ala Tyr Ala  315

946   AAA GCT TAT CCG CAG GAA GCC GCT GAA TTT ACC CGC CGT ATG AAA  990
316   Lys Ala Tyr Pro Gln Glu Ala Ala Glu Phe Thr Arg Arg Met Lys  330

991   GGC GAA ATG CCG TCT GAC TTC GAC GCT AAA GCG AAA GAG TTC ATC  1035
331   Gly Glu Met Pro Ser Asp Phe Asp Ala Lys Ala Lys Glu Phe Ile  345
```

```
1036  GCT AAA CTG CAG GCT AAT CCG GCG AAA ATC GCC AGC CGT AAA GCG  1080
346   Ala Lys Leu Gln Ala Asn Pro Ala Lys Ile Ala Ser Arg Lys Ala  360

1081  TCT CAG AAT GCT ATC GAA GCG TTC GGT CCG CTG TTG CCG GAA TTC  1125
361   Ser Gln Asn Ala Ile Glu Ala Phe Gly Pro Leu Leu Pro Glu Phe  375

1126  CTC GGC GGT TCT GCT GAC CTG GCG CCG TCT AAC CTG ACC CTG TGG  1170
376   Leu Gly Gly Ser Ala Asp Leu Ala Pro Ser Asn Leu Thr Leu Trp  390

1171  TCT GGT TCT AAA GCA ATC AAC GAA GAT GCT GCG GGT AAC TAC ATC  1215
391   Ser Gly Ser Lys Ala Ile Asn Glu Asp Ala Ala Gly Asn Tyr Ile  405

1216  CAC TAC GGT GTT CGC GAG TTC GGT ATG ACC GCG ATT GCT AAC GGT  1260
406   His Tyr Gly Val Arg Glu Phe Gly Met Thr Ala Ile Ala Asn Gly  420

1261  ATC TCC CTG CAC GGT GGC TTC CTG CCG TAC ACC TCC ACC TTC CTG  1305
421   Ile Ser Leu His Gly Gly Phe Leu Pro Tyr Thr Ser Thr Phe Leu  435

1306  ATG TTC GTG GAA TAC GCA CGT AAC GCC GTA CGT ATG GCT GCG CTG  1350
436   Met Phe Val Glu Tyr Ala Arg Asn Ala Val Arg Met Ala Ala Leu  450

1351  ATG AAA CAG CGT CAG GTG ATG GTT TAC ACC CAC GAC TCC ATC GGT  1395
451   Met Lys Gln Arg Gln Val Met Val Tyr Thr His Asp Ser Ile Gly  465

                    Target C
1396  CTG GGC GAA GAC GGC CCG ACT CAC CAG CCG GTT GAG CAG GTC GCT  1440
466   Leu Gly Glu Asp Gly Pro Thr His Gln Pro Val Glu Gln Val Ala  480

1441  TCT CTG CGC GTA ACC CCG AAC ATG TCT ACA TGG CGT CCG TGT GAC  1485
481   Ser Leu Arg Val Thr Pro Asn Met Ser Thr Trp Arg Pro Cys Asp  495

1486  CAG GTT GAA TCC GCG GTC GCG TGG AAA TAC GGT GTT GAG CGT CAG  1530
496   Gln Val Glu Ser Ala Val Ala Trp Lys Tyr Gly Val Glu Arg Gln  510

1531  GAC GGC CCG ACC GCA CTG ATC CTC TCC CGT CAG AAC CTG GCG CAG  1575
511   Asp Gly Pro Thr Ala Leu Ile Leu Ser Arg Gln Asn Leu Ala Gln  525

1576  CAG GAA CGA ACT GAA GAG CAA CTG GCA AAC ATC GCG CGC GGT GGT  1620
526   Gln Glu Arg Thr Glu Glu Gln Leu Ala Asn Ile Ala Arg Gly Gly  540

1621  TAT GTG CTG AAA GAC TGC GCC GGT CAG CCG GAA CTG ATT TTC ATC  1665
541   Tyr Val Leu Lys Asp Cys Ala Gly Gln Pro Glu Leu Ile Phe Ile  555

1666  GCT ACC GGT TCA GAA GTT GAA CTG GCT GTT GCT GCC TAC GAA AAA  1710
556   Ala Thr Gly Ser Glu Val Glu Leu Ala Val Ala Ala Tyr Glu Lys  570

1711  CTG ACT GCC GAA GGC GTG AAA GCG CGC GTG GTG TCC ATG TCG TCT  1755
571   Leu Thr Ala Glu Gly Val Lys Ala Arg Val Val Ser Met Ser Ser  585

1756  ACC GAC GCA TTT GAC AAG CAG GAT GCT GCT TAC CGT GAA TCC GTA  1800
586   Thr Asp Ala Phe Asp Lys Gln Asp Ala Ala Tyr Arg Glu Ser Val  600

1801  CTG CCG AAA GCG GTT ACT GCA CGC GTT GCT GTA GAA GCG GGT ATT  1845
601   Leu Pro Lys Ala Val Thr Ala Arg Val Ala Val Glu Ala Gly Ile  615

1846  GCT GAC TAC TGG TAC AAG TAT GTT GGC CTG AAC GGT GCT ATC GTC  1890
616   Ala Asp Tyr Trp Tyr Lys Tyr Val Gly Leu Asn Gly Ala Ile Val  630

1891  GGT ATG ACC ACC TTC GGT GAA TCT GCT CCG GCA GAG CTG CTG TTT  1935
631   Gly Met Thr Thr Phe Gly Glu Ser Ala Pro Ala Glu Leu Leu Phe  645

1936  GAA GAG TTC GGC TTC ACT GTT GAT AAC GTT GTT GCG AAA GCA AAA  1980
646   Glu Glu Phe Gly Phe Thr Val Asp Asn Val Val Ala Lys Ala Lys  660

1981  GAA CTG CTG TAA  1992
661   Glu Leu Leu End
```

Figure 4.3. (Previous two pages.) Nucleotide sequence of the *E. coli tkt* gene. The corresponding sequence of residues in the transketolase protein is shown below the nucleotide sequence. The three target sites are shaded red (A), blue (B), and magenta (C). Target A is residues Ser24–His26 (nucleotides 70–78). Target B is residues Gly99–Pro101 (nucleotides 295–303). Target C is residues Asp469–His473 (nucleotides 1405–1419). (Sequence data from Sprenger, 1993.)

Primer	Sequence (5′ → 3′)	Length (bases)	T_m (°C)
Target A			
tktAF	ATGGACGCAGTACAGAAAGCCAAA	24	59.0
tktAR	CCCATAGGGGCCCCCGG	17	60.6
Target B			
tktBF	GTCAGCTGCACTCTAAAACTCCG	23	57.8
tktBR	ACCAGCGGTGTAACCCACTTC	21	58.8
Target C			
tktCF	ACTCCATCGGTCTGGGCGAA	20	59.8
tktCR	CGACCTGCTCAACCGGCTG	19	60.2

Table 4.3. Forward and reverse primers for the mutagenesis of the target sites. Forward primers have an "F" suffix, while reverse primers have an "R" suffix. Each T_m was calculated for a primer concentration of 250nM and a salt concentration of 50mM.

different "sublibraries", each with a different point mutation rate and size (Patrick *et al.*, 2003). Screening a particular number of *E. coli* XL10-Gold pQR711ABC$^{\mu}$ transformants would, therefore, cover these sublibraries to varying extents. The level of coverage in each case would be dependent upon the size of the sublibrary and the proportion of the overall library that it constitutes. This section presents the calculations performed to estimate the size of each pQR711ABC$^{\mu}$ sublibrary and the corresponding number of relevant transformants required to achieve 90% coverage (Section 4.2.3).

This section also includes similar calculations for three further libraries with point mutations limited to the individual targets (pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$). It was decided that the 1pmpv sublibraries of these targets would be screened before construction of library pQR711ABC$^{\mu}$ to determine whether single mutations were capable if yielding the desired activity.

4.3.3.1 Sublibrary sizes

Using the formula in Section 1.2.3.1 the sizes of the pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$ 1pmpv sublibraries were estimated to be 18, 18, and 30 unique protein sequences, respectively (Table 4.4).

The numbers of unique protein sequences in the pQR711ABC$^{\mu}$ 1, 2, and 3 point mutations per variant (pmpv) sublibraries were estimated to be 66, 1967, and 3.53×10^4 unique protein sequences, respectively (Table 4.5). As the combined length of targets A, B, and C was short (33 bases) the probability of multiple point mutations occurring in the same codon was found to be not

Library	Target length		Number of point mutations (x)	Number of unique nucleotide sequences (V)	Number of unique protein sequences (W)
	Bases (N)	Residues ($N/3$)			
pQR711A$^{\mu}$	9	3	1	27	18
pQR711B$^{\mu}$	9	3	1	27	18
pQR711C$^{\mu}$	15	5	1	45	30

Table 4.4. Sizes of the pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$ 1pmpv sublibraries. Sublibraries with point mutation rates higher than 1pmpv were not expected to feature significantly in the overall libraries. (See Section 1.2.3.1 for the formula used to calculate these data. Note that the numbers of unique protein sequences are estimates.)

Number of point mutations (x)	Probability of mutations cooperating	Number of unique nucleotide sequences (V)	Number of unique protein sequences (W)
1	0.000	99	66
2	0.042	4752	1967
3	0.127	1.47×10^5	3.53×10^4

Table 4.5. Sizes of the pQR711ABC$^{\mu}$ 1, 2, and 3pmpv sublibraries. The overall length of the three target sites is 11 residues (33 bases). "Probability of mutations cooperating" is the probability that more than one point mutation will occur in a codon. (See section 1.2.3.1 for the formulae used to calculate these data. Note that the numbers of unique protein sequences are estimates.)

insignificant in both the 2 and 3pmpv sublibraries: 0.042 and 0.127, respectively. The calculated sizes of the 2 and 3pmpv sublibraries were, therefore, slight underestimates.

4.3.3.2 Sublibrary coverage

The numbers of transformants required for 90% coverage of the pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$ 1pmpv sublibraries were calculated from the sublibrary sizes. The results were: 42 for pQR711A$^{\mu}$, 42 for pQR711B$^{\mu}$, and 69 for pQR711C$^{\mu}$ (Figure 4.4). Note that these are numbers of transformants *withthe correct number of point mutations* (1pmpv). In order that these figures could be converted into numbers of transformants *selected at random from the library*, the sublibrary compositions of the libraries needed to be determined by sequencing (refer to Section 5.4). The multiplier in each case would be the reciprocal of the proportion of the library that the 1pmpv sublibrary constitutes.

The numbers of transformants required for 90% coverage of the pQR711ABC$^{\mu}$ sublibraries were calculated to be: 152 for the 1pmpv sublibrary, 4536 for the 2pmpv sublibrary, and 8.13×10^4 for the 3pmpv sublibrary (Figure 4.5). Again, these are numbers of transformants *with the correct numbers of point mutations.*

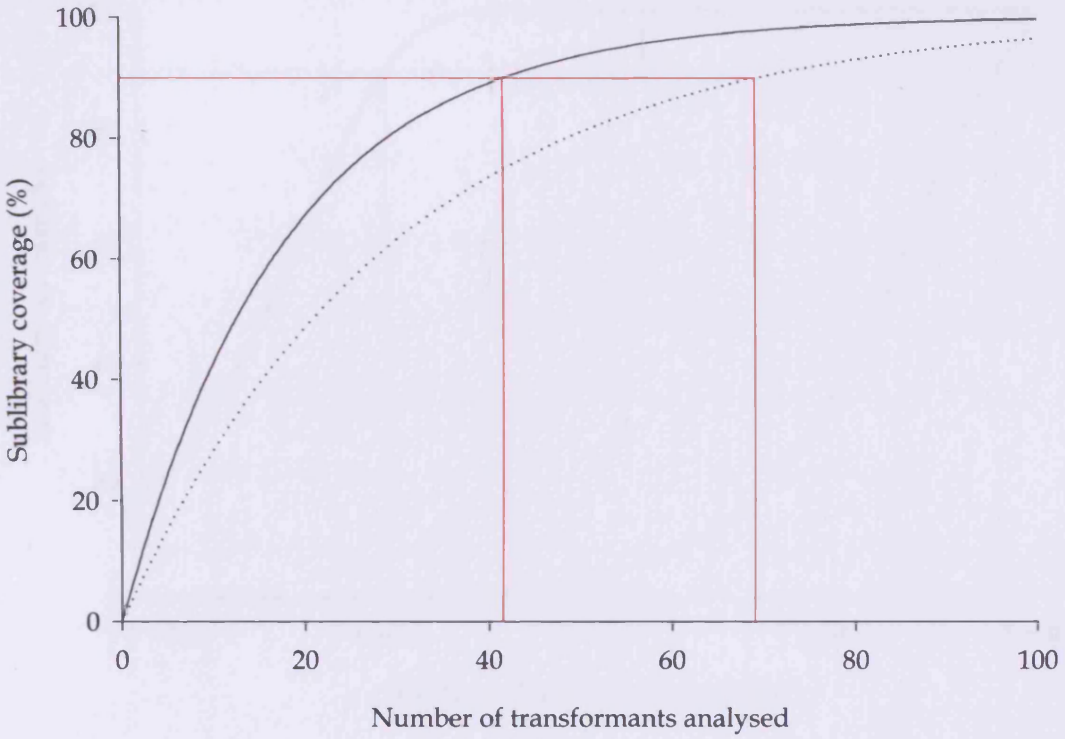


Figure 4.4. Numbers of transformants needed for 90% coverage of the pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$ 1pmpv sublibraries. These sublibraries contain approximately 19, 19, and 31 unique protein sequences, respectively. 90% coverage of each of these sublibraries requires the following numbers of relevant transformants to be analysed: (a) 42 for the 1pmpv sublibraries of pQR711A$^{\mu}$ and pQR711B$^{\mu}$ (solid); and (b) 69 for the 1pmpv sublibrary of pQR711C$^{\mu}$ (dotted). (See Section 1.2.3.2 for the formula used to calculate these curves.)

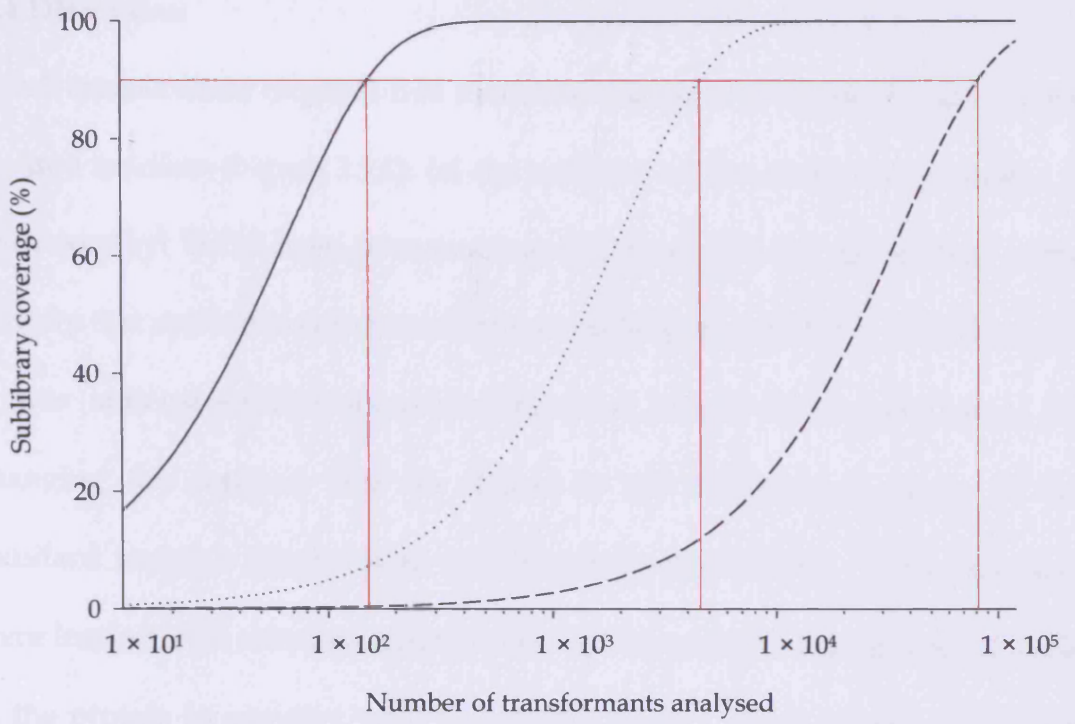


Figure 4.5. Numbers of transformants needed for 90% coverage of three pQR711ABCμ sublibraries. pQR711ABCμ sublibraries with 1, 2, and 3pmpv have approximately 68, 2121, and 3.95×10^4 unique protein sequences, respectively. 90% coverage of each of these libraries would require the following numbers of relevant transformants to be analysed: (a) 152 for the 1pmpv sublibrary (solid); (b) 4536 for the 2pmpv sublibrary (dotted); and (c) 8.13×10^4 for the 3pmpv sublibrary (dashed). (See Section 1.2.3.2 for the formula used to calculate these curves.)

4.4 Discussion

E. coli transketolase requires two modifications for it to be able to catalyse the desired reaction (Figure 1.14): (a) the addition of the ability to synthesise 1-hydroxyethyl ThDP from pyruvate and ThDP; and (b) the elimination of the activity that splits this compound into acetaldehyde and ThDP. In this project it was anticipated that these modifications would be brought about by changing the residues that lie closest to the C2-hydroxyl group of the standard reaction intermediate (1,2-dihydroxyethyl ThDP). These residues were located in *S. cerevisiae* transketolase by inspecting a crystal structure file of the protein in complex with 1,2-dihydroxyethyl ThDP (Section 4.3.1). Of the residues considered, His30, His103, Asp477, and His481 were singled out for directed evolution. These residues were identified in *E. coli* transketolase (His26, His100, Asp469, and His473) and the targets were expanded to include flanking residues. Overall, the target sequences chosen were: Ser24–His26 (target A), Gly99–Pro101 (target B), and Asp469–His473 (target C).

The method chosen to mutate these target sequences was fepPCR, a focused version of epPCR (refer to Chapter 5). Focusing mutation from the whole gene to these sequences would allow a very high mutation rate to be applied, while maintaining a reasonable library size. It was believed that this approach would yield combinations of substitutions that operated synergistically.[12]

Primer oligonucleotides were constructed to flank the target sequences

[12] These combinations might not be identified by making single amino acid substitutions because the intermediates may be unfavourable.

so that fepPCR could be performed (Section 4.3.2). Each primer was designed to have a T_m of approximately 60°C and no significant secondary structure.

Initially, three small libraries with mutations limited to either target A (pQR711A$^{\mu}$), target B (pQR711B$^{\mu}$), or target C (pQR711C$^{\mu}$) would be generated with these primers and screened for the desired activity (Figure 1.14). The numbers of relevant transformants required for 90% coverage of each 1pmpv sublibrary were estimated to be: 42 for pQR711A$^{\mu}$, 42 for pQR711B$^{\mu}$, and 69 for pQR711C$^{\mu}$ (Section 4.3.3.2). It was anticipated that all three libraries could, therefore, be screened in the space of a day (refer to Section 3.4.6).

Next, variants with multiple mutations across all three targets (pQR711ABC$^{\mu}$) would be generated and screened. The numbers of relevant transformants required for 90% coverage of each of the three smallest sublibraries were calculated to be: 152 for the 1pmpv sublibrary, 4536 for the 2pmpv sublibrary, and 8.13×10^4 for the 3pmpv sublibrary (Section 4.3.3.2). If the protocol devised in Chapter 3 was used to screen these sublibraries, 90% coverage of each would be achieved in ≥1, ≥12, and ≥212 days, respectively (380 screening reactions per day). Based on these figures 90% coverage of the 3pmpv sublibrary was rejected as an unrealistic project goal and 90% coverage of the 2pmpv sublibrary was chosen instead.

Chapter 5 - The fepPCR-QuikChange method

5.1 Introduction

Error-prone PCR (epPCR) followed by ligation (epPCR-ligation) is a popular methodology for generating plasmid libraries. A gene is amplified by PCR under conditions that reduce the fidelity of the DNAP and the mutated product is ligated into a vector. There are, however, two significant problems with this approach: (a) mutations are spread across the entire gene so the majority have little chance of affecting the enzyme's characteristics (Clackson *et al.*, 1998); and (b) ligation is extremely labour-intensive and occasionally impossible.

Another commonly-used protocol for generating plasmid libraries is random oligonucleotide mutagenesis. Oligonucleotides that encode a specific region of a plasmid-borne gene are synthesised with certain codons randomised and extended into whole plasmids using a PCR-based cloning method such as QuikChange (Stratagene Ltd.). Two drawbacks with this approach are: (a) the level of mutation in the library cannot be varied unless new oligonucleotides are designed and synthesised; and (b) extremely large libraries are generated by the randomisation of just a few codons.

The mutagenesis targets in this project were discontinuous and extremely short (Table 4.2) so epPCR-ligation was not a viable method for library construction. Random nucleotide mutagenesis-QuikChange was not suitable either because the mutation density would be much too high for 11

residues.[13]

In order that the desired libraries could be constructed a novel technique was developed: focused epPCR (fepPCR). FepPCR is the PCR-based amplification of a short, defined region in a gene under conditions that drastically lower the fidelity of the DNAP. In other words: fepPCR is epPCR of diminutive targets with higher mutation densities. The small size of the product means that it may be cloned into plasmids by QuikChange SDM, thus avoiding the gruelling process of ligation.

This chapter describes the results of subjecting the target B site in *tkt* to fepPCR and cloning the product into whole plasmids. Two different QuikChange methods were examined and compared for this cloning step: QuikChange Site-Directed Mutagenesis (SDM) and QuikChange Multi Site-Directed Mutagenesis (MSDM).

5.2 Methods

5.2.1 FepPCR amplification of target B

5.2.1.1 Amplification of target B at various concentrations of Mn^{2+}

11 50µl PCR reactions were set up containing the following (Table 5.1): 5U *Taq* DNAP (Roche Diagnostics Ltd.), *Taq* DNAP buffer (Roche Diagnostics), 50ng pQR711 DNA, 0.5µM primer tktBF (5'-GTCAGCTGCACTCTAAAACTCCG-3') (Cruachem Ltd.), 0.5µM primer tktBR (5'-ACCAGCGGTGTAACCCACTTC-3') (Cruachem Ltd.), 0.2mM each

[13] Randomising 11 residues would lead to a library size of 2.05×10^{14} unique protein sequences ($20^{11} = 2.05 \times 10^{14}$): too large for any known screening or selection system to handle.

Reagent	Volume (µl)	Final amount/conc.
ddH_2O	23	
10× *Taq* DNAP buffer	5	
$MgCl_2$	3	1.5mM
$MnCl_2$	10	0–1mM
dNTPs	1	0.2mM each
tktBF	1	0.5µM
tktBR	1	0.5µM
pQR711 DNA	5	50ng
Taq DNAP	1	5U
Total	50	

Table 5.1. The components of each fepPCR reaction. Each reaction contained a different concentration of $MnCl_2$.

Phase	Cycles	Step(s)	Temperature (°C)	Duration (min)
1	1	Initial denaturation	94	3
2	25	Denaturation	94	1
		Annealing	55	1
		Extension	72	1
3	1	Final extension	72	5
		Final hold	4	∞

Table 5.2. Program of temperature cycling used for fepPCR. (Modified from Roche Diagnostics Ltd., 2000.)

Figure 5.1. Recognition site of the restriction enzyme *BstE* II (a) in general and (b) in target B. The enzyme cleaves the sequence at the points indicated by the arrows. Target B is shaded blue.

deoxynucleoside triphosphate (dATP, dCTP, dGTP, and dTTP), and 1.5mM $MgCl_2$ in pure water. $MnCl_2$ was added to the reactions at varying concentrations (0–1mM). A control reaction was set up containing every component except *Taq* DNAP.

The reactions were submitted to a program of temperature cycling in a TechGene thermal cycler (Techne Ltd.) (Table 5.2). Following fepPCR, the samples were purified using QIAquick nucleotide removal kits (QIAGEN Ltd.). The purified samples were run on a 2.0% (w/v) agarose gel.

5.2.1.2 Mutation densities of the products from reactions B0.0 to B1.0

A "cutting solution" was prepared by adding 30µl of the restriction enzyme *BstE* II (New England Biolabs Ltd.) (Figure 5.1) and 30µl of 10× NEBuffer 3 (New England Biolabs Ltd.) to 90µl of pure water. 10µl of this solution was added to 10µl of each purified fepPCR reaction. The digestions were incubated at 37°C for two hours and then analysed on three separate polyacrylamide gels.

5.2.2 Generation of *E. coli* XL10-Gold pQR711B$^{\mu}$ libraries by two different methods

5.2.2.1 QuikChange SDM

A variation of the QuikChange SDM protocol (Stratagene Ltd., 2003a) was used. A 50µl PCR reaction was made up containing the following (see Table 5.3): 2.5U *PfuTurbo* DNAP (Stratagene Ltd.), *PfuTurbo* DNAP buffer (Stratagene Ltd.), 150ng pQR711 DNA, and 1µl of dNTP solution (Stratagene

Reagent	Volume (µl)	Final amount
ddH_2O	34	
10× *PfuTurbo* DNAP buffer	5	
dNTP mix	1	
0.7mM Mn^{2+} fepPCR product	5	250ng
pQR711 DNA	5	150ng
PfuTurbo DNAP	1	2.5U
Total	51	

Table 5.3. The components of the QuikChange SDM reaction. The concentration of dNTPs in the dNTP mix is not specified by Stratagene Ltd.

Phase	Cycles	Step(s)	Temperature (°C)	Duration (min)
1	1	Initial denaturation	95	0.5
2	18	Denaturation	95	0.5
		Annealing	55	1.0
		Extension	68	12.0
3	1	Final hold	4	∞

Table 5.4. Program of temperature cycling used for QuikChange SDM. (Modified from Stratagene Ltd., 2003a.)

Ltd.) in pure water. 250ng of DNA from the purified 0.7mM Mn^{2+} fepPCR reaction (B0.7) replaced the usual mutagenic primers. The QuikChange reaction was submitted to a program of temperature cycling in a TechGene thermal cycler (Techne Ltd.) (Table 5.4).

The QuikChange reaction was digested with 10U *Dpn* I (Stratagene Ltd.) for 2 hours at 37°C. 1µl of the digestion was transformed into competent *E. coli* XL10-Gold cells by heat-shock (Section 2.3.7). The transformed cells were spread on LB Amp^+ agar. Plasmid DNA was prepared (Section 2.3.6) from overnight cultures (Section 2.3.2) of 10 colonies and sequenced with primer TKN (Section 2.3.13).

5.2.2.2 QuikChange MSDM

A variation of the QuikChange MSDM protocol (Stratagene Ltd., 2003b) was used. A 25µl PCR reaction was made up containing the following (see Table 5.5): 2.5U QuikChange Multi enzyme blend (Stratagene Ltd.), QuikChange Multi reaction buffer (Stratagene Ltd.), 0.75µl QuikSolution, 150ng pQR711 DNA, and 1µl of dNTP solution (Stratagene Ltd.) in pure water. 100ng of DNA from the purified 0.7mM Mn^{2+} fepPCR reaction (B0.7) replaced the usual mutagenic primers. The QuikChange reaction was submitted to a program of temperature cycling in a TechGene thermal cycler (Techne Ltd.) (Table 5.6).

The QuikChange reaction was digested with 10U *Dpn* I (Stratagene Ltd.) for 2 hours at 37°C. 1.5µl of the digestion was transformed into

Reagent	Volume (µl)	Amount
ddH_2O	10.5	
10× QuikChange Multi reaction buffer	2.5	
QuikSolution	0.75	
dNTP mix	1	
0.7mM Mn^{2+} fepPCR product	5	100ng
pQR711 DNA	5	150ng
QuikChange Multi enzyme blend	1	2.5U
Total	25	

Table 5.5. The components of the QuikChange MSDM. The compositions of QuikSolution and QuikChange Multi enzyme blend are not revealed by Stratagene Ltd. The concentrations of dNTPs in the dNTP mix are also unknown.

Phase	Cycles	Step(s)	Temperature (°C)	Duration (min)
1	1	Initial denaturation	95	1
2	30	Denaturation	95	1
		Annealing	55	1
		Extension	65	12.0
3	1	Final hold	4	∞

Table 5.6. Program of temperature cycling used for QuikChange MSDM. (Modified from Stratagene Ltd., 2003b.)

competent *E. coli* XL10-Gold cells by heat-shock (Section 2.3.7). The transformed cells were spread on LB Amp^+ agar. Plasmid DNA was prepared (Section 2.3.6) from overnight cultures (Section 2.3.2) of 20 colonies and sequenced with primer TKN (Section 2.3.13).

5.3 Results

5.3.1 FepPCR amplification of target B

5.3.1.1 Amplification of target B at various concentrations of Mn^{2+}

Target B (refer to Chapter 4) was chosen as the model sequence for the development of fepPCR. 11 fepPCR reactions (B0.0 to B1.0) were set up containing the following: *Taq* DNAP, *Taq* DNAP buffer, pQR711 DNA, primer tktBF, primer tktBR, each deoxynucleoside triphosphate (dATP, dCTP, dGTP, and dTTP), and Mg^{2+} (Section 5.2.1.1). To reduce the fidelity of the DNAP, and thus achieve "error-prone" PCR, Mn^{2+} was added to the reactions (Beckman *et al.*, 1985; Cadwell *et al.*, 1995). To generate a range of mutation rates in the reactions the concentration of Mn^{2+} was increased in 0.1mM increments from 0mM in reaction B0.0 to 1mM in reaction B1.0.

The reactions were submitted to a program of temperature cycling and then purified. Agarose gel electrophoresis revealed that the 53bp fragment had been amplified in the first eight reactions (B0.0 to B0.7) with Mn^{2+} concentrations higher than 0.7mM inhibiting fepPCR (Figure 5.2).

5.3.1.2 Mutation densities of the products from reactions B0.0 to B1.0

The mutation densities of the products from fepPCR reactions B0.0 to B1.0

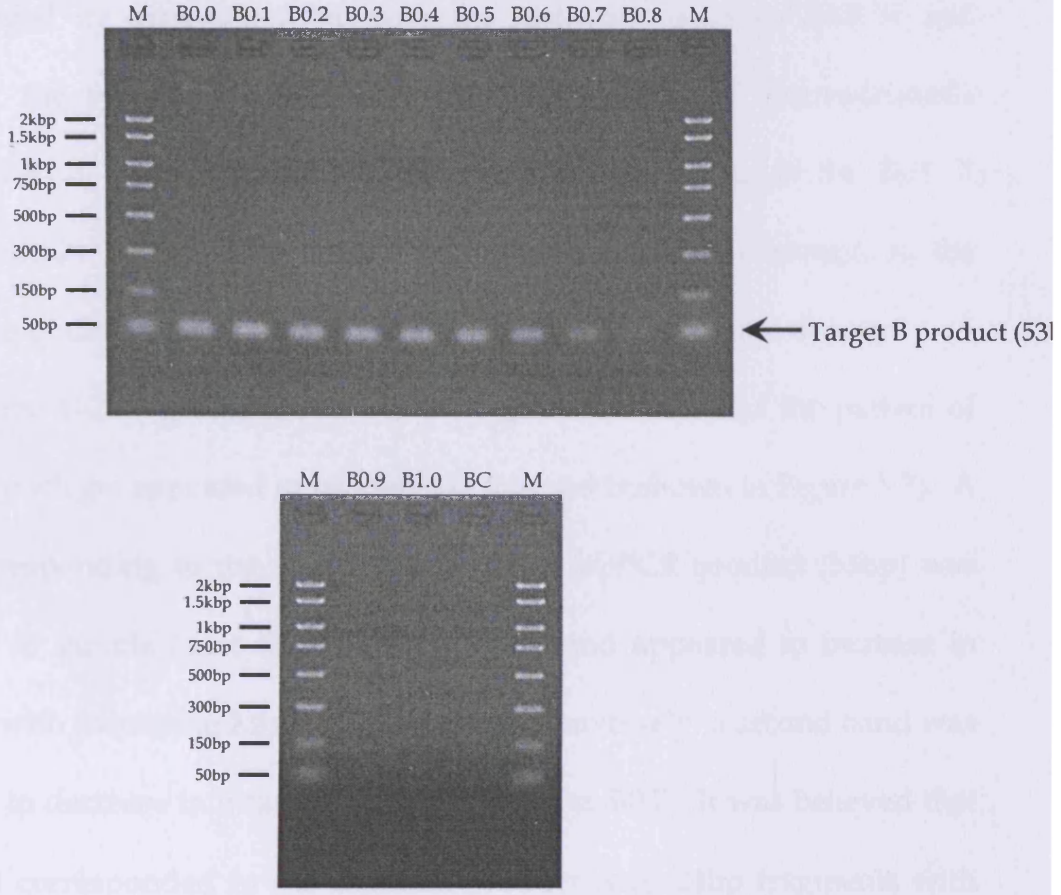


Figure 5.2. 2.0% agarose gel electrophoresis of target B fepPCR reactions. Sample lanes "B0.0" to "B1.0" contain the corresponding fepPCR reactions. Lane "BC" contains the control reaction (no *Taq* DNAP). All "M" lanes contain marker DNA (50–2000bp). The 53bp band in lanes "B0.0" to "B0.7" corresponds to the fepPCR product of target B.

were gauged by digesting them with the restriction enzyme *BstE* II and analysing the resulting fragments by polyacrylamide gel electrophoresis (Section 5.2.1.2). It was reasoned that mutations occurring in the *BstE* II recognition site in target B (Figure 5.1b) would preclude cleavage, so the proportion of undigested product would increase with mutation density.

Three 4–20% gradient TBE replicate gels were run and the pattern of bands on each gel appeared to be identical (one gel is shown in Figure 5.3). A band corresponding to the undigested target B fepPCR product (53bp) was observed in sample lanes B0.0 to B0.7. This band appeared to increase in intensity with increasing Mn^{2+} concentration. Conversely, a second band was observed to decrease in intensity from lane B0.3 to B0.7. It was believed that this band corresponded to the digested product (two 24bp fragments with staggered ends). Neither of the bands were seen when the Mn^{2+} concentration was >0.7mM because fepPCR was prevented.

The proportion of undigested product in each treated reaction (U) was determined quantitatively through the use of the software package Quantity One (Bio-Rad Laboratories). In each case the peak intensities of the undigested band and the fragment band were measured and U was calculated by dividing the former value by the sum of both values. Figure 5.4 records the U values computed for the B0.0 to B0.7 sample lanes of all three gels. Mean U was observed to jump from 0% at 0.2mM Mn^{2+} to 16.6% at 0.3mM; subsequent increments of 0.1mM yielded progressively smaller increases in mean U. Mean U reached a maximum at 0.7mM Mn^{2+} when 40.0% of the

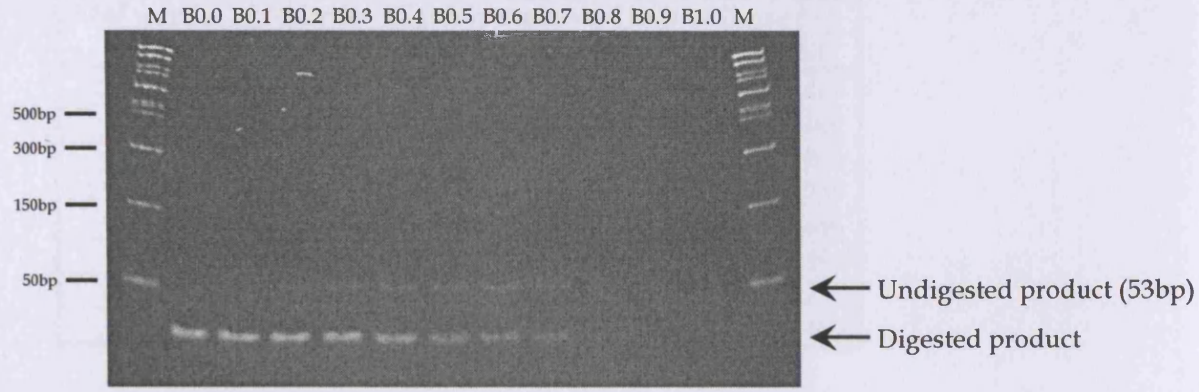


Figure 5.3. Polyacrylamide gel electrophoresis of the target B fepPCR reactions following digestion with restriction enzyme *BstE* II. Sample lanes "B0.0" to "B1.0" contain the corresponding digested fepPCR reactions. Both "M" lanes contain marker DNA (50–2000bp). The 53bp band in lanes "B0.0" to "B0.7" corresponds to the fepPCR product of target B. The second band in lanes "B0.2" to "B0.7" corresponds to the digested fragments of the target B fepPCR product.

Mn^{2+} conc. (mM)	Proportion of product mutated in *Bst*EII site (%)				
	A	B	C	Mean	±σM
0.0	0.000	0.000	0.000	0.000	0.000
0.1	0.000	0.000	0.000	0.000	0.000
0.2	0.000	0.000	0.000	0.000	0.000
0.3	15.022	17.579	17.061	16.554	0.780
0.4	24.205	25.239	26.474	25.306	0.656
0.5	31.917	35.695	32.347	33.320	1.194
0.6	36.969	39.163	37.059	37.730	0.717
0.7	42.805	39.161	38.037	40.001	1.439

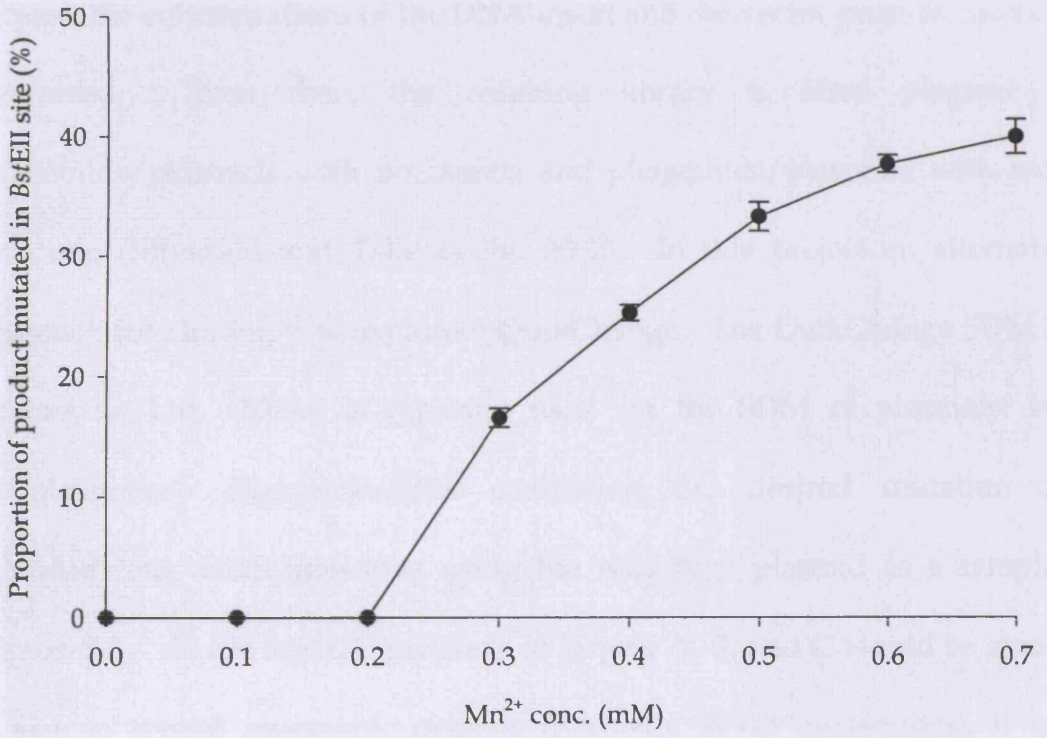


Figure 5.4. The effect of Mn^{2+} concentration on the proportion of target B fepPCR product with mutated *BstE* II recognition sites. The error bars correspond to $\pm\sigma_M$.

product had mutations in the *BstE* II recognition site.

5.3.2 Generation of *E. coli* XL10-Gold pQR711Bμ colonies by two different methods

The conventional method for cloning epPCR fragments is ligation (Miyazaki and Takenouchi, 2002). This method is, however, extremely labour-intensive because the concentrations of the DNA insert and the vector must be carefully optimized. Even then, the resulting library is often plagued by phagemids/plasmids with no inserts and phagemids/plasmids with more than one (Miyazaki and Takenouchi, 2002). In this project an alternative approach for cloning was explored: QuikChange. The QuikChange SDM kit (Stratagene Ltd., 2003a) is typically used for the SDM of plasmids: two complementary oligonucleotides containing the desired mutation are extended into whole plasmids using the wild-type plasmid as a template (Figure 5.5). As the fepPCR products of targets A, B, and C would be similar in size to typical mutagenic primers (normally 25–45 nucleotides), it was reasoned that QuikChange SDM could be used for the cloning step. This approach would be tested by constructing a pQR711Bμ library from the purified product of a target B fepPCR reaction. If it proved to be successful then libraries pQR711Aμ and pQR711Cμ would be generated in the same way. Library pQR711ABCμ would be constructed by sequential mutation of the three targets: (1) target A (pQR711Aμ); then (2) target B (pQR711ABμ); and finally (3) target C (pQR711ABCμ).

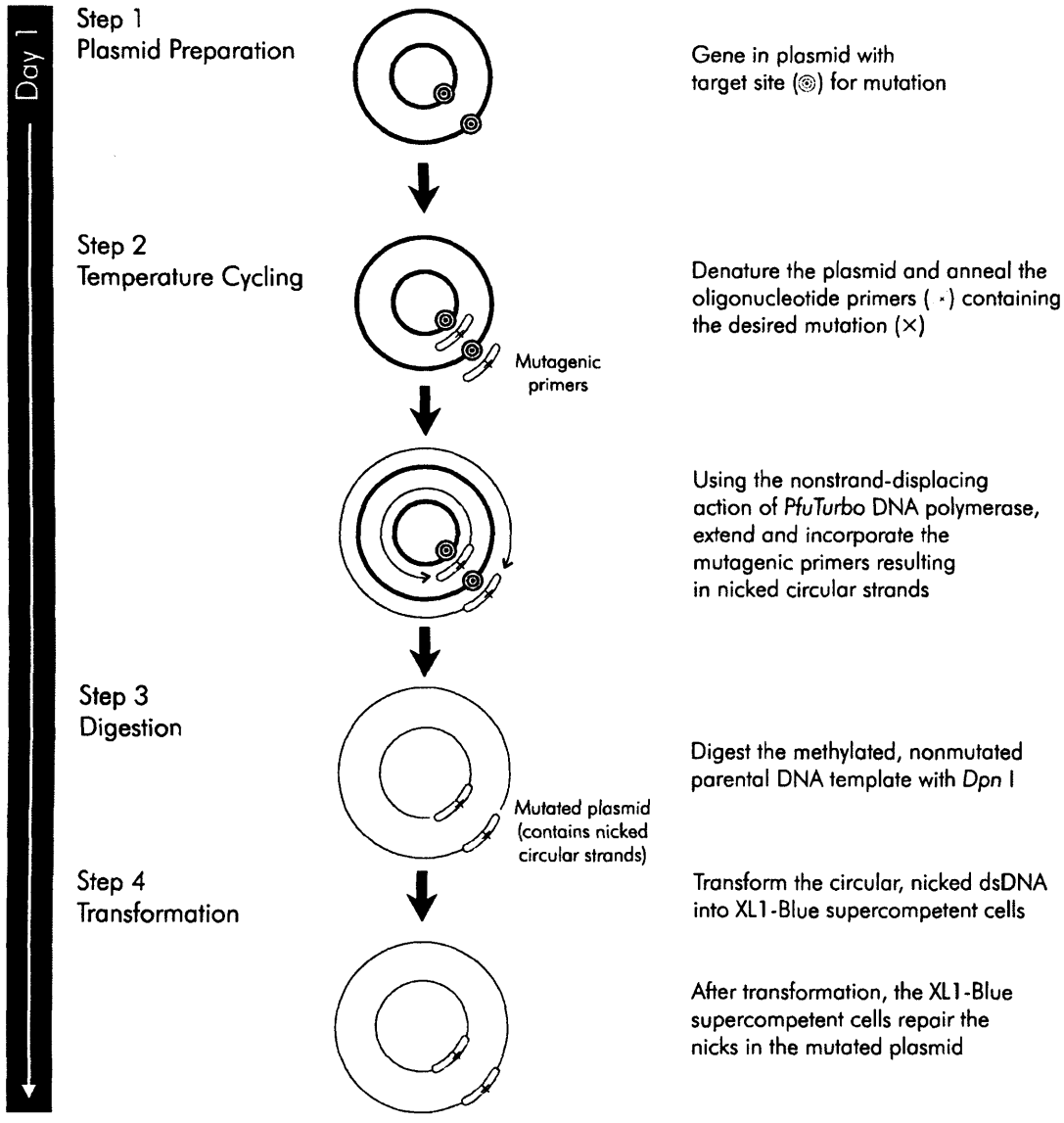


Figure 5.5. Overview of the QuikChange SDM method. (Reproduced from Stratagene Ltd., 2003a.)

Stratagene Ltd. markets a variation of the QuikChange Site-Directed Mutagenesis (SDM) kit known as the QuikChange Multi Site-Directed Mutagenesis (MSDM) kit (Figure 5.6). It differs from QuikChange SDM in its ability to mutate up to five sites simultaneously (compared to one). The advantage of this system over the standard SDM kit was immediately clear: targets A, B, and C could be mutated simultaneously, eliminating the need for sequential mutagenesis. Initially, however, QuikChange MSDM was examined and compared to QuikChange SDM for its efficacy at generating a suitable pQR711B$^{\mu}$ library from the purified product of a target B fepPCR reaction.

5.3.2.1 QuikChange SDM

A variation of the QuikChange SDM protocol (Stratagene Ltd., 2003a) was used (Section 5.2.2.1). A PCR reaction was made up containing the following: *PfuTurbo* DNAP, *PfuTurbo* DNAP buffer, pQR711 DNA, and dNTPs in pure water. An aliquot of the purified 0.7mM Mn^{2+} fepPCR reaction (B0.7) replaced the usual mutagenic primers. The QuikChange reaction was submitted to a program of temperature cycling and then digested with the restriction enzyme *Dpn* I to remove the methylated parental DNA. The remaining plasmid library (pQR711B$^{\mu}$-S where "S" stands for SDM) was transformed into competent *E. coli* XL10-Gold cells by heat-shock. The transformed cells were spread on LB Amp^{+} agar and approximately 4900 colonies were obtained. Plasmid DNA was prepared from 10 colonies

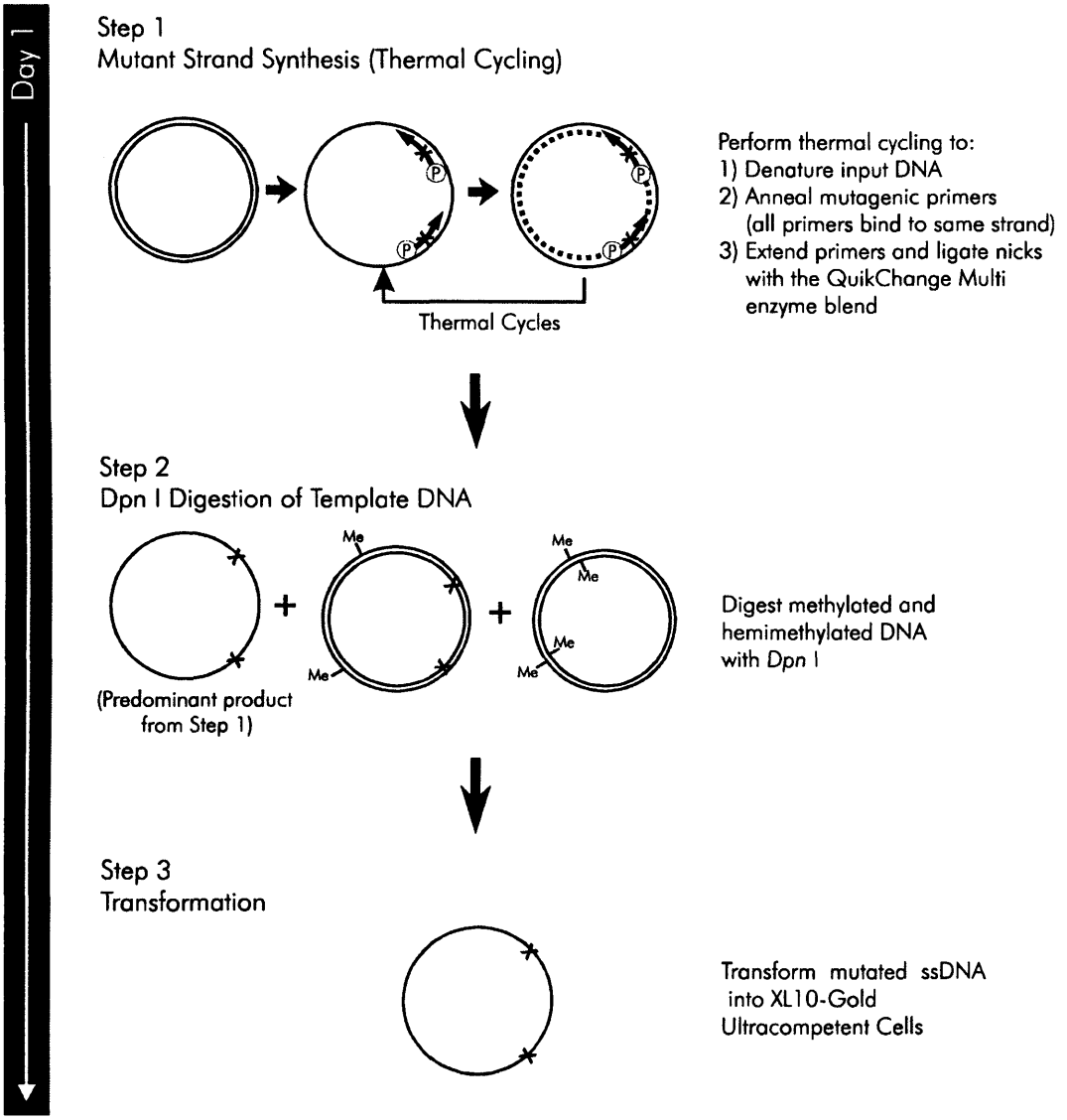


Figure 5.6. Overview of the QuikChange MSDM method. (Reproduced from Stratagene Ltd., 2003b.)

(pQR711Bμ-S1A1 to S1A10 [plate 1, wells A1 to A10]) and sequenced with primer TKN (Figure 5.7).

Five sequences were found to contain single point mutations (pQR711Bμ-S1A2, S1A4, S1A6, S1A7, and S1A10). The five remaining sequences were wild-type (pQR711Bμ-S1A1, S1A3, S1A5, S1A8, and S1A9). These data are illustrated as a pie chart in Figure 5.8. Three of the 10 sequences (30%) had mutations in the *BstE* II recognition site.

5.3.2.2 QuikChange MSDM

A variation of the QuikChange MSDM protocol (Stratagene Ltd., 2003b) was used (Section 5.2.2.2). A PCR reaction was made up containing the following: QuikChange Multi enzyme blend, QuikChange Multi reaction buffer, QuikSolution, pQR711 DNA, and dNTPs in pure water. An aliquot of the purified 0.7mM Mn^{2+} fepPCR reaction (B0.7) replaced the usual mutagenic primers. The QuikChange reaction was submitted to a program of temperature cycling and then digested with the restriction enzyme *Dpn* I to remove the methylated parental DNA. The remaining plasmid library (pQR711Bμ-M where "M" stands for MSDM) was transformed into competent *E. coli* XL10-Gold cells by heat-shock. The transformed cells were spread on LB Amp^+ agar and approximately 6400 colonies were obtained. Plasmid DNA was prepared from 20 colonies (pQR711Bμ-M1A1 to M1A20) and sequenced with primer TKN (Figure 5.9).

Four sequences were found to contain single point mutations

```
pQR711          272 GTCAGCTGCACTCTAAAACTCCGGGTCACCCGGAAGTGGGTTACACCGCTGGT 324
pQR711Bμ-S1A1   272 ..................................................... 324
pQR711Bμ-S1A2   272 ..................................A.................. 324
pQR711Bμ-S1A3   272 ..................................................... 324
pQR711Bμ-S1A4   272 ...................................T.................. 324
pQR711Bμ-S1A5   272 ..................................................... 324
pQR711Bμ-S1A6   272 ...................................T.................. 324
pQR711Bμ-S1A7   272 ..................................G.................. 324
pQR711Bμ-S1A8   272 ..................................................... 324
pQR711Bμ-S1A9   272 ..................................................... 324
pQR711Bμ-S1A10  272 ...................................T.................. 324
```

Figure 5.7. Nucleotide sequences of the target B regions in 10 *E. coli* XL10-Gold pQR711Bμ-S colonies generated by QuikChange SDM. Target B (*tkt* nucleotides +295 to +303) is shaded blue and the *BstE* II recognition site is shaded black.

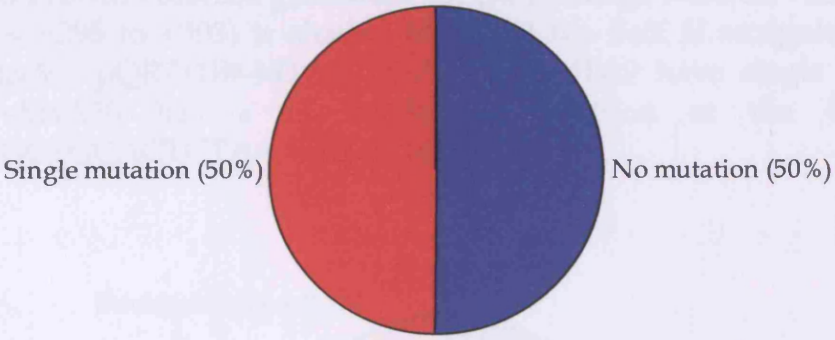


Figure 5.8. A pie chart illustrating the proportion of pQR711Bμ-S sequences with single point mutations in target B. Target B is 9 nucleotides in length.

```
pQR711          272 GTCAGCTGCACTCTAAAACTCCGGGTCACCCGGAAGTGGGTTACACCGCTGGT 324
pQR711Bμ-M1A1   272 ..........................C..-.......................... 323
pQR711Bμ-M1A2   272 .............................................................. 324
pQR711Bμ-M1A3   272 ..........................A..C..-....................... 323
pQR711Bμ-M1A4   272 .............................................................. 324
pQR711Bμ-M1A5   272 ..........................C.....A....................... 324
pQR711Bμ-M1A6   272 .............................................................. 324
pQR711Bμ-M1A7   272 .............................................................. 324
pQR711Bμ-M1A8   272 .............................................................. 324
pQR711Bμ-M1A9   272 ..........................................-..................... 323
pQR711Bμ-M1A10  272 .............................................................. 324
pQR711Bμ-M1A11  272 ..........................A...G.....G...................... 324
pQR711Bμ-M1A12  272 .............................................................. 324
pQR711Bμ-M1A13  272 .............................................................. 324
pQR711Bμ-M1A14  272 .............................................T................ 324
pQR711Bμ-M1A15  272 .............................................................. 324
pQR711Bμ-M1A16  272 ..........................A.....A....................... 324
pQR711Bμ-M1A17  272 ..........................A..A.....................  324
pQR711Bμ-M1A18  272 .............................................................. 324
pQR711Bμ-M1A19  272 .............................................................. 324
pQR711Bμ-M1A20  272 ..................◊.....A.....A....................... 349
```

Figure 5.9. Nucleotide sequences of the target B regions in 20 *E. coli* XL10-Gold pQR711Bμ-M colonies generated by QuikChange MSDM. Target B (*tkt* nucleotides +295 to +303) is shaded blue and the *BstE* II recognition site is shaded black. pQR711Bμ-M1A1, M1A3, and M1A9 have single deletions. pQR711Bμ-M1A20 has a 25 nucleotide addition at the ◊ symbol (5'-GCCAGCTGCACTCTAAAACTCCGGA-3').

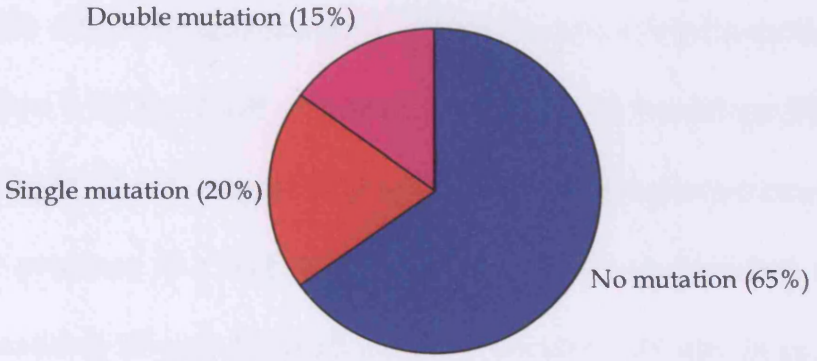


Figure 5.10. A pie chart illustrating the proportions of pQR711Bμ-M sequences with single and double point mutations in target B. Target B is 9 nucleotides in length. "No mutation (65%)" includes sequences that were frame-shifted.

(pQR711Bμ-M1A1, M1A14, M1A16, and M1A17) and three were found to contain double point mutations (pQR711Bμ-M1A3, M1A5, and M1A11) (Figure 5.10). Of these sequences, two were frame-shifted (pQR711Bμ-M1A1 and M1A3). Two other sequences were frame-shifted without any point mutations (pQR711Bμ-M1A9 and M1A20). The 11 remaining sequences were wild-type (pQR711Bμ-M1A2, M1A4, M1A6, M1A7, M1A8, M1A10, M1A12, M1A13, M1A15, M1A18, and M1A19). Six of the 20 sequences (30%) had mutations in the *BstE* II recognition site.

5.4 Discussion

5.4.1 FepPCR amplification of target B

Agarose gel electrophoresis revealed that fepPCR amplification of target B was possible when the concentration of Mn^{2+} was ≤0.7mM (reactions B0.0 to B0.7) (Section 5.3.1.1). Mn^{2+} concentrations >0.7mM (reactions B0.8 to B1.0) prevented fepPCR and detectable amounts of product were not produced.

The presence of a *BstE* II recognition site (six nucleotides) in target B (nine nucleotides) (Figure 5.1b) allowed the mutation density in each reaction product to be gauged. It was reasoned that if mutations occurred in the *BstE* II recognition site, cleavage by the restriction enzyme *BstE* II would be precluded. The proportion of undigested product would, therefore, increase with mutation density.

Aliquots of the reactions were incubated with *BstE* II for two hours (to permit complete digestion) and then analysed by polyacrylamide gel

electrophoresis (Section 5.3.1.2). The following results were obtained: (a) the proportion of fepPCR product with mutations in the *BstE* II recognition site (*U*) was nil for Mn^{2+} concentrations ≤0.2mM; and (b) above 0.2mM Mn^{2+} *U* increased in a non-linear fashion with Mn^{2+} concentration to a maximum of 40.0% at 0.7mM.

In 1985 Beckman and co-workers demonstrated that there were three mechanisms by which Mn^{2+} cations disrupt the fidelity of *E. coli* DNAP I: (a) by interacting with the DNA template; (b) by binding to single-stranded regions of the DNA template; and (c) by binding to weak accessory sites on the DNAP. Beckman observed phenomenon (a) at low concentrations of Mn^{2+} (≤0.1mM) and the more significant phenomena (b) and (c) at higher concentrations (0.5–1.5mM). These findings may be used to explain the fepPCR results: (i) the products of reactions B0.0 to B0.2 were ostensibly free of mutation because the Mn^{2+} concentrations were too low (≤0.2mM) for any phenomenon other than (a) to occur; and (ii) the products of reactions B0.3 to B0.7 were increasingly mutated because the greater Mn^{2+} concentrations (0.3–0.7mM) brought about phenomena (b) and (c). Further experimentation would, of course, be necessary to prove or disprove this hypothesis.

The relationship between the proportion of fepPCR product with mutations in the *BstE* II recognition site (*U*) and the average point mutation density of the product (*B*) is:

$$U \approx 1-(R/L)^B$$

Where R = length of the recognition site (6 nucleotides for *Bst*EII) and L = length of the target (9 nucleotides for target B). Or, after rearranging:

$$(R/L)^B \approx 1-U$$
$$B \approx \log_{(R/L)}(1-U)$$

(Note that this relationship is approximate because addition and deletion mutations also contribute to U.) This formula was used to calculate the average point mutation densities of the target B fepPCR products from their B values (Table 5.7). The product of reaction B0.7 had the highest point mutation density: 1.26pmpv. This figure was slightly higher than the target density for the target B library (1pmpv, refer to Section 4.3.3) and was consequently chosen for the construction of the pQR711B$^{\mu}$ libraries.[14]

5.4.2 Generation of *E. coli* XL10-Gold pQR711B$^{\mu}$ colonies by two different methods

QuikChange SDM and QuikChange MSDM were developed by Stratagene Ltd. to replace the labour-intensive cloning techniques typically used in SDM. Both of these methods were used to generate pQR711B$^{\mu}$ libraries from the product of fepPCR reaction B0.7; using this product instead of the usual mutagenic primers meant that libraries were generated, rather than single variants of SDM. The quality of the pQR711B$^{\mu}$ library generated by each method was determined by sequencing.

[14] It was anticipated that some loss of point mutation density would occur during the cloning step.

Mn^{2+} conc. (mM)	*U* (%)	Point mutation density
0.0	0.0	0.00
0.1	0.0	0.00
0.2	0.0	0.00
0.3	16.5	0.45
0.4	25.3	0.72
0.5	33.3	1.00
0.6	37.7	1.17
0.7	40.0	1.26

Table 5.7. The relationship between average point mutation density in target B and the Mn^{2+} concentration in target B fepPCR products. *U* is the proportion of undigested product in each reaction following treatment with the restriction enzyme *BstE* II.

50% of the pQR711Bμ-S plasmids created by QuikChange SDM were found to have single point mutations in target B (1pmpv) (Section 5.3.2.1). The remainder were completely wild-type. In comparison, 20% of the pQR711Bμ-M plasmids generated by QuikChange MSDM had single point mutations in target B (1pmpv) and 15% had double point mutations (2pmpv) (Section 5.3.2.2). Both methods were, therefore, successful at cloning the mutated fepPCR product into whole pQR711 plasmids.

QuikChange MSDM generated some plasmids (20%) that were frame-shifted due to additions or deletions, but this phenomenon was absent from the pQR711Bμ-S plasmids. Frame-shifted proteins are drastically mutated or truncated, so one fifth of the pQR711Bμ-M library was, therefore, useless. It should be stated, however, that this phenomenon has been observed in other experiments using QuikChange SDM, so its apparent absence from the pQR711Bμ-S library was probably a function of the small sample size (only 10 sequences).

40.0% of the fepPCR product had mutations in the *BstE* II recognition site of target B (refer to section 5.4.1), while only 30% of the pQR711Bμ-S and pQR711Bμ-M sequences had them (in both cases). This reduction in mutation density was probably caused by one (or more) of the following: (a) a preference of the QuikChange mechanism for less-mutated primer sequences (fepPCR product); (b) wild-type plasmids being reconstructed from trace amounts of the fepPCR primers tktBF and tktBR (instead of fepPCR product); or (c) incomplete elimination of wild-type template (pQR711) by *Dpn* I

digestion prior to transformation.

Transforming small amounts of the QuikChange reactions into competent *E. coli* XL10-Gold cells resulted in large numbers of colonies being generated: 4900 from 1µl of the QuikChange SDM reaction (total volume: 50µl) and 6400 from 1µl of the QuikChange MSDM reaction (total volume: 25µl). These figures were used to calculate the total numbers of colonies available from the entire reactions: ~2.45 × 10^5 colonies of *E. coli* XL10-Gold pQR711B$^{\mu}$-S ($4900 \times \frac{50}{1} = 2.45 \times 10^5$) and ~1.60 × 10^5 colonies of *E. coli* XL10-Gold pQR711B$^{\mu}$-M ($6400 \times \frac{25}{1} = 1.60 \times 10^5$). Such numbers were more than sufficient for any screening operation.

5.4.3 Application of the fepPCR-QuikChange method

5.4.3.1 This project

FepPCR successfully generated highly-mutated copies of a short, defined region in the *tkt* gene (target B). The point mutation rate in the product was observed to increase with Mn^{2+} concentration, reaching a maximum of 0.14 per nucleotide at 0.7mM (a ~1270-fold reduction in the fidelity of *Taq* polymerase [Tindall and Kunkel, 1988]).

QuikChange SDM and QuikChange MSDM both proved to be effective at cloning this fepPCR product into whole plasmids. Firstly, the point mutation rate of the fepPCR product was well preserved in library pQR711B$^{\mu}$-S and library pQR711B-$^{\mu}$M. Secondly, the point mutations were limited to the targeted site (*tkt* nucleotides +295 to +303) in all cases (bar sequence

pQR711Bμ-MA11). Thirdly, large numbers of *E. coli* XL10-Gold pQR711Bμ colonies were obtained from each transformation.

The 1pmpv sublibrary component of library *E. coli* XL10-Gold pQR711Bμ-S (50%) was found to be greater than that of library XL10-Gold pQR711Bμ-M (20%). To achieve 90% coverage of this sublibrary would, therefore, require the following numbers of colonies: 84 colonies of *E. coli* XL10-Gold pQR711Bμ-S ($\frac{1}{0.5} \times 42 = 84$, refer to Section 4.3.3.2) or 220 colonies of XL10-Gold pQR711Bμ-M ($\frac{1}{0.2} \times 42 = 210$). In order that processing time was kept to a minimum the pQR711Bμ-S library was chosen for screening and QuikChange SDM was selected to generate the libraries pQR711Aμ and pQR711Cμ.

QuikChange SDM was rejected as the cloning method for the construction of library pQR711ABCμ as it would have involved sequential mutagenesis of targets A, B, and C. QuikChange MSDM was selected instead as it permitted simultaneous mutagenesis of the targets (Stratagene Ltd., 2003b): a swifter and more efficient approach.

5.4.3.2 Other projects

The implementation of the fepPCR-QuikChange protocol in a directed evolution project is summarised in Figures 5.11 and 5.12. QuikChange SDM (Figure 5.11) is used for cloning when there is a single target for mutagenesis and QuikChange MSDM (Figure 5.12) is used when there are multiple targets.

The strengths of fepPCR-QuikChange as a mutagenesis method are: (a)

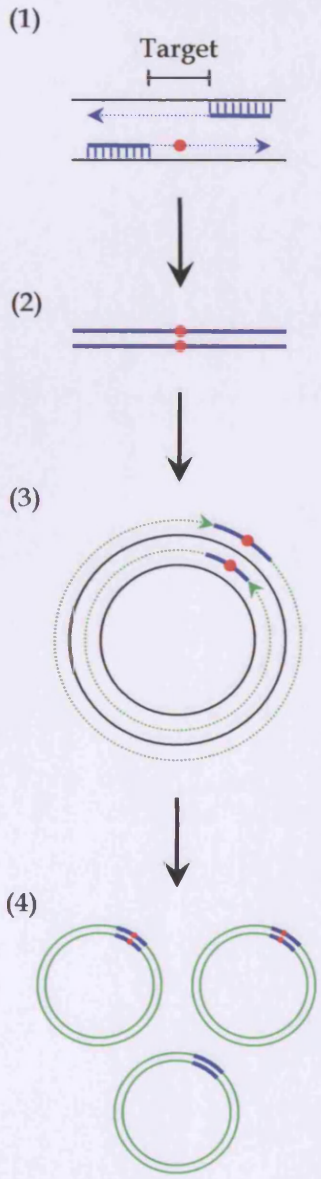


Figure 5.11. Implementation of fepPCR-QuikChange SDM in the construction of a plasmid library with a single mutagenesis target. (1) fepPCR of the target site is performed (point mutations are shown as red dots). (2) The resulting array of mutant dsDNA fragments is purified by a QIAquick nucleotide removal kit (one fragment, coloured blue, is shown here). (3) QuikChange SDM: the purified mutant fragments are extended into whole plasmids using wild-type plasmid DNA as the template (the template is coloured black and the new DNA is coloured green). (4) The methylated parental DNA is removed by digestion with the restriction enzyme *Dpn* I, leaving a library of plasmids with different mutations in the target site. Of the three plasmids shown here two are members of the 1pmpv sublibrary and one is wild-type.

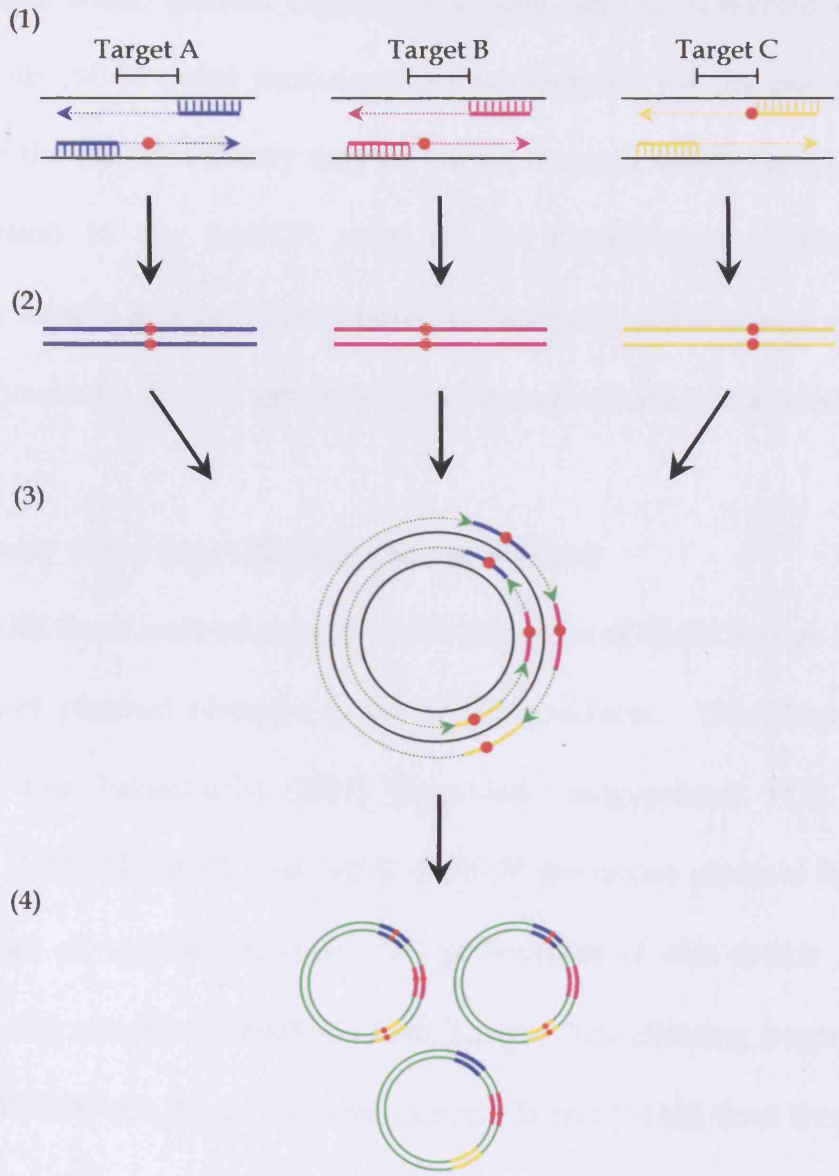


Figure 5.12. Implementation of fepPCR-QuikChange MSDM in the construction of a plasmid library with multiple mutagenesis targets. The process is identical to that of fepPCR-QuikChange SDM (Figure 5.11) except that: (a) multiple target sites (coloured blue, magenta, and yellow) are amplified by fepPCR in parallel reactions (step 1) and purified (step 2); and (b) the mutant fragments from all fepPCR reactions are incorporated into whole plasmids simultaneously by QuikChange MSDM (step 3). Of the three plasmids shown here (step 4) one is a member of the 3pmpv sublibrary, one is a member of the 2pmpv sublibrary, and one is a member of the 1pmpv sublibrary.

one or more short, defined regions of a gene may be subjected to extreme mutagenesis (≤0.14 point mutations per nucleotide); (b) the point mutation density in the plasmid library may be varied through modulation of the Mn^{2+} concentration in the fepPCR step; (c) the QuikChange cloning step is extremely simple and rapid (compared to ligation); and (d) large numbers of colonies (variants) may be generated from a single cloning reaction (≥10^5).

5.4.4 Novelty of the fepPCR-QuikChange method

Prior to 2002 there were no reports in the literature of QuikChange being used to construct plasmid libraries from epPCR products. This changed when Miyazaki and Takenouchi (2002) described "megaprimer PCR of whole plasmid" (MEGAWHOP). As MEGAWHOP generates plasmid libraries by QuikChange of epPCR products, the publication of this article had some impact on the novelty of fepPCR-QuikChange. The priming fragments used in MEGAWHOP are, however, considerably larger (~1kb) than those used in fepPCR-QuikChange (~50bp).

The fepPCR process that generates the priming fragments in the first instance currently remains a novel variation of epPCR.

Chapter 6 – Libraries and screening

6.1 Introduction

The results of the previous chapter demonstrated the need to choose the appropriate QuikChange protocol for creating a library by fepPCR-QuikChange. As libraries pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$ had single targets for mutagenesis, QuikChange SDM was selected as the cloning method in these cases. In contrast, the multiple targets of library pQR711ABC$^{\mu}$ required that QuikChange MSDM be used for its construction.

E. coli XL10-Gold pQR711B$^{\mu}$-S colonies had been created by this point so this chapter describes generation of XL10-Gold pQR711A$^{\mu}$-S, XL10-Gold pQR711C$^{\mu}$-S, and XL10-Gold pQR711ABC$^{\mu}$-M colonies. The screening procedure developed in Chapter 3 (summarised in Figure 3.15) was then used to test the transketolase variants in all four libraries for the desired activity: the conversion of pyruvate and glycolaldehyde to (*S*)-3,4-dihydroxybutan-2-one (and carbon dioxide) (Figure 1.14). The desired levels of sublibrary coverage were: (a) 90% coverage the single target 1pmpv sublibraries; and (b) 90% coverage of the pQR711ABC$^{\mu}$ 2pmpv sublibrary.

6.2 Methods

6.2.1 Preparation of *E. coli* XL10-Gold pQR711A$^{\mu}$-S, XL10-Gold pQR711C$^{\mu}$-S, and XL10-Gold pQR711ABC$^{\mu}$-M libraries

6.2.1.1 FepPCR amplification of targets A and C

FepPCR amplification of target A was performed by repeating Section 5.2.1.1 with tktAF (5'-ATGGACGCAGTACAGAAAGCCAAA-3') (Cruachem Ltd.) and tktAR (5'-CCCATAGGGGCCCCCGG-3') (Cruachem Ltd.) replacing the primers tktBF and tktBR, repectively. A smaller range of $MnCl_2$ (Mn^{2+}) concentrations was used in these reactions: 0, 0.7, 0.8, 0.9, and 1.0mM.

FepPCR amplification of target C was performed by repeating Section 5.2.1.1 with tktCF (5'-ACTCCATCGGTCTGGGCGAA-3') (Cruachem Ltd.) and tktCR (5'-CGACCTGCTCAACCGGCTG-3') (Cruachem Ltd.) replacing the primers tktBF and tktBR, repectively. The range of $MnCl_2$ (Mn^{2+}) concentrations used in the reactions was identical to that used in the fepPCR amplification of target A.

6.2.1.2 Generation of E. coli *XL10-Gold pQR711A$^{\mu}$-S and XL10-Gold pQR711C$^{\mu}$-S colonies by QuikChange SDM*

The purified 0.7mM Mn^{2+} fepPCR products of targets A (A0.7) and C (C0.7) were used in QuikChange SDM (refer to Section 5.2.2.1) to generate colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$-S and XL10-Gold pQR711C$^{\mu}$-S, respectively.

Plasmid DNA was prepared (Section 2.3.6) from overnight cultures (Section 2.3.2) of 10 *E. coli* XL10-Gold pQR711A$^{\mu}$-S colonies and sequenced with primer TKN (Section 2.3.13). Plasmid DNA was prepared (Section 2.3.6)

from overnight cultures (Section 2.3.2) of 10 *E. coli* XL10-Gold pQR711C$^{\mu}$ colonies and sequenced with primer TKC (Section 2.3.13).

6.2.1.3 Generation of E. coli *XL10-Gold pQR711ABC$^{\mu}$-M colonies by QuikChange MSDM*

The purified 0.7mM Mn^{2+} fepPCR products of targets A (A0.7), B (B0.7), and C (C0.7) were used in QuikChange MSDM (Section 5.2.2.2) to generate colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M. 100ng of each purified fepPCR product were included in the reaction. Plasmid DNA was prepared (Section 2.3.6) from overnight cultures (Section 2.3.2) of 20 colonies and sequenced with primer TKN and primer TKC (in separate reactions) (Section 2.3.13).

6.2.2 Screening of libraries pQR711A$^{\mu}$-S, pQR711B$^{\mu}$-S, pQR711C$^{\mu}$-S, and pQR711ABC$^{\mu}$-M for the desired activity

6.2.2.1 Microwell culture

A Qpix2 robot (Genetix Ltd.) was used to pick 147 colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$-S into 147 wells of a single 384-well microwell culture plate (refer to Section 3.2.2). Four wells were inoculated with *E. coli* XL10-Gold pQR711 to provide controls. The plate was incubated at 37°C and agitated at a rate of 1600rpm for 16 hours.

84 colonies of *E. coli* XL10-Gold pQR711B$^{\mu}$-S, 230 colonies of XL10-Gold pQR711C$^{\mu}$-S, and 2388 colonies of XL10-Gold pQR711ABC$^{\mu}$-M were cultured using the same procedure. Four wells in every plate were set aside for *E. coli* XL10-Gold pQR711 controls.

6.2.2.2 Incubation with target substrates and HPLC analysis

Every microwell culture plate was screened using the procedure described in Figure 3.15. HPLC chromatograms were collected for all of the screening reactions following four hours of incubation at 25°C. The pyruvate peak heights in each library's chromatograms were consolidated into a single text file using a small Java program.

6.3 Results

6.3.1 Preparation of *E. coli* XL10-Gold pQR711A$^{\mu}$-S, XL10-Gold pQR711C$^{\mu}$-S, and XL10-Gold pQR711ABC$^{\mu}$-M colonies

6.3.1.1 FepPCR amplification of targets A and C

FepPCR amplifications of targets A and C were carried out using the following pairs of primers: (a) tktAF and tktAR for target A; and (b) tktCF and tktCR for target C (Section 6.2.1.1). The following range of Mn^{2+} concentrations was utilised in both cases: 0, 0.7, 0.8, 0.9, and 1.0mM.

Agarose gel electrophoresis revealed that the 50bp target A product had been amplified in reactions A0.0 to A0.8 (Figure 6.1). Mn^{2+} concentrations higher than 0.8mM prevented fepPCR.

A separate gel revealed that the 54bp target C product had been amplified in reactions C0.0 to C0.9 (Figure 6.2). Mn^{2+} concentrations higher than 0.9mM prevented fepPCR.

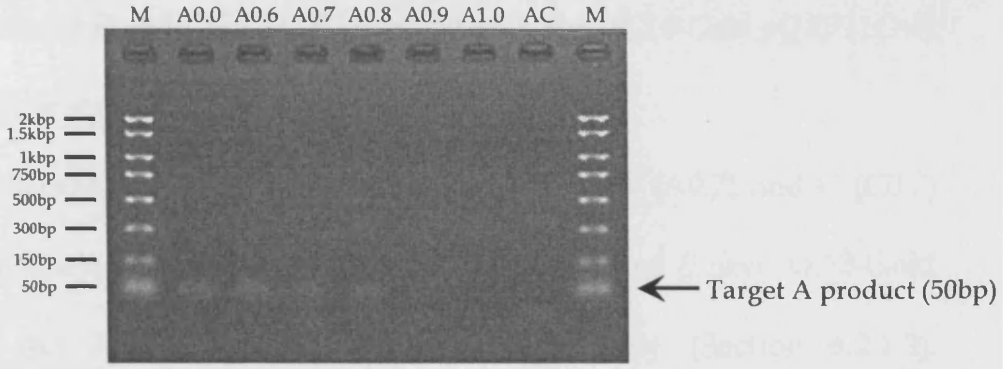


Figure 6.1. 2.0% agarose gel electrophoresis of target A fepPCR reactions. Sample lanes "A0.0" to "A1.0" contain the corresponding fepPCR reactions. Lane "AC" contains the control reaction (no *Taq* DNAP). Both "M" lanes contain marker DNA (50–2000bp). The 50bp band in lanes "A0.0" to "A0.8" corresponds to the fepPCR product of target A.

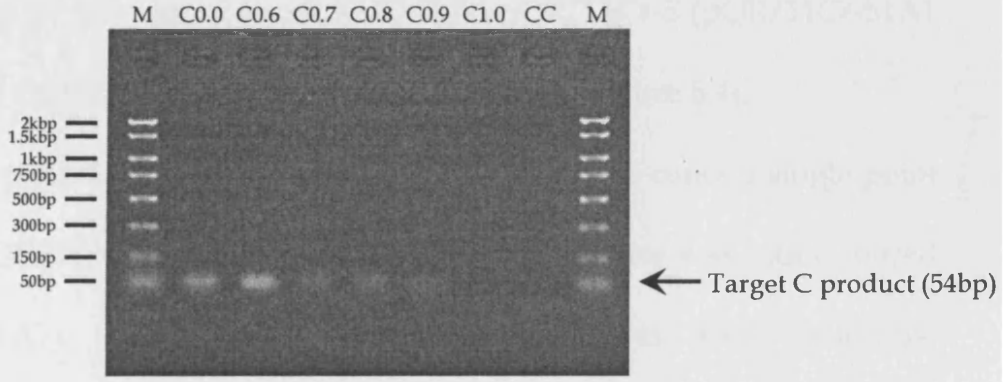


Figure 6.2. 2.0% agarose gel electrophoresis of target C fepPCR reactions. Sample lanes "C0.0" to "C1.0" contain the corresponding fepPCR reactions. Lane "CC" contains the control reaction (no *Taq* DNAP). Both "M" lanes contain marker DNA (50–2000bp). The 54bp band in lanes "C0.0" to "C0.9" corresponds to the fepPCR product of target C.

6.3.1.2 Generation of E. coli *XL10-Gold pQR711A$^{\mu}$-S and XL10-Gold pQR711C$^{\mu}$-S colonies by QuikChange SDM*

The purified 0.7mM Mn^{2+} fepPCR products of targets A (A0.7) and C (C0.7) were used in QuikChange SDM to generate colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$-S and XL10-Gold pQR711C$^{\mu}$-S, respectively (Section 6.2.1.2). Approximately 6100 colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$ and 5500 colonies of XL10-Gold pQR711C$^{\mu}$ were obtained. Plasmid DNA was prepared from 10 colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$-S (pQR711A$^{\mu}$-S1A1 to S1A10) and each was sequenced with primer TKN (Figure 6.3). Plasmid DNA was prepared from 10 colonies of *E. coli* XL10-Gold pQR711C$^{\mu}$-S (pQR711C$^{\mu}$-S1A1 to S1A10) and each was sequenced with primer TKC (Figure 6.4).

Of the pQR711A$^{\mu}$-S sequences, two were found to contain single point mutations (pQR711A$^{\mu}$-S1A2 and S1A7). A single sequence was frame-shifted (pQR711A$^{\mu}$-S1A1). The seven remaining sequences were wild-type (pQR711A$^{\mu}$-S1A3, S1A4, S1A5, S1A6, S1A8, S1A9, and S1A10).

Of the pQR711C$^{\mu}$ sequences, three were found to contain single point mutations (pQR711C$^{\mu}$-S1A1, S1A4, and S1A7). Two sequences were frame-shifted (pQR711C$^{\mu}$-S1A5 and S1A6). The five remaining sequences were wild-type (pQR711C$^{\mu}$-S1A2, S1A3, S1A8, S1A9, and S1A10).

Figure 6.3. Nucleotide sequences of the target A regions in 10 *E. coli* XL10-Gold pQR711A$^{\mu}$-S colonies generated by QuikChange SDM. Target A (nucleotides +70 to +78) is shaded red. pQR711A$^{\mu}$-S1A1 has a single deletion.

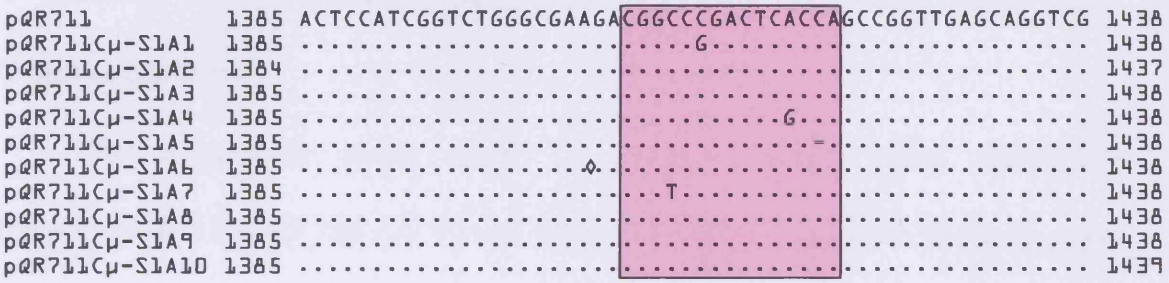


Figure 6.4. Nucleotide sequences of the target C regions in 10 *E. coli* XL10-Gold pQR711C$^{\mu}$-S colonies generated by QuikChange SDM. Target C (nucleotides +1405 to +1419) is shaded magenta. pQR711C$^{\mu}$-S1A5 has a single deletion. pQR711C$^{\mu}$-S1A6 has a 38 nucleotide addition at the ◊ symbol (5′-CTCCATCGGTCTGGGCGAACTCCATCGGTCTGGGCGAA-3′).

6.3.1.3 Generation of E. coli *XL10-Gold pQR711ABC$^{\mu}$-M colonies by QuikChange MSDM*

The purified 0.7mM Mn^{2+} fepPCR products of targets A (A0.7), B (B0.7), and C (C0.7) were used in QuikChange MSDM to generate colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M (Section 6.2.1.3). 100ng of each purified fepPCR product were included in the reaction. Approximately 4400 colonies were obtained. Plasmid DNA was prepared from 20 colonies (pQR711ABC$^{\mu}$-M1A1 to M1A20) and each was sequenced with primer TKN and primer TKC (Figures 6.5 to 6.7).

Six sequences were found to contain single point mutations (pQR711ABC$^{\mu}$-M1A5, M1A8, M1A9, M1A12, M1A19, and M1A20), four were found to contain double point mutations (pQR711ABC$^{\mu}$-M1A2, M1A10, M1A11, and M1A17), two were found to contain triple point mutations (pQR711ABC$^{\mu}$-M1A4 and M1A6), and one was found to contain a quadruple point mutation (pQR711ABC$^{\mu}$-M1A13) (Table 6.1 and Figure 6.8). Of these sequences, three were frame-shifted (pQR711ABC$^{\mu}$-M1A2, M1A10, and M1A12). Two other sequences were frame-shifted without any point mutations (pQR711ABC$^{\mu}$-M1A14 and M1A15). The five remaining sequences were wild-type (pQR711ABC$^{\mu}$-M1A1, M1A3, M1A7, M1A16, and M1A18).

Figure 6.5. Nucleotide sequences of the target A regions in 20 *E. coli* XL10-Gold pQR711ABCμ-M colonies generated by QuikChange MSDM. Target A (nucleotides +70 to +78) is shaded red.

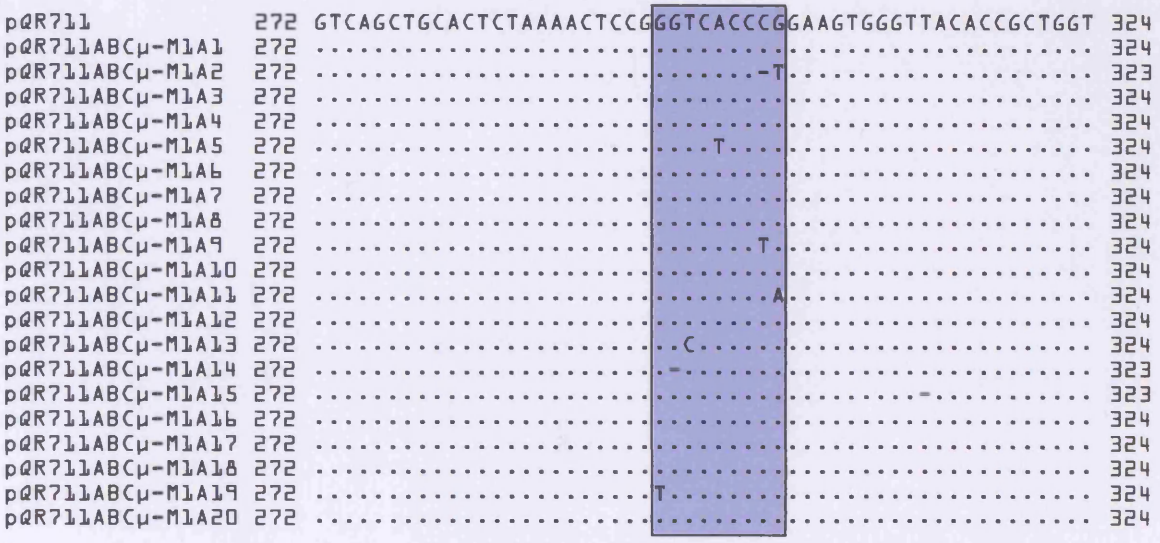


Figure 6.6. Nucleotide sequences of the target B regions in pQR711ABCμ-MA1 to MA20. Target B (nucleotides +295 to +303) is shaded blue. pQR711ABCμ-M1A2, M1A14, and M1A15 have single deletions.

Sequence	Number of target site mutations				Frame-shifted?
	A	B	C	Total	
pQR711ABCμ-M1A1	0	0	0	0	No
pQR711ABCμ-M1A2	1	1	0	2	Yes
pQR711ABCμ-M1A3	0	0	0	0	No
pQR711ABCμ-M1A4	1	0	2	3	No
pQR711ABCμ-M1A5	0	1	0	1	No
pQR711ABCμ-M1A6	1	0	2	3	No
pQR711ABCμ-M1A7	0	0	0	0	No
pQR711ABCμ-M1A8	0	0	1	1	No
pQR711ABCμ-M1A9	0	1	0	1	No
pQR711ABCμ-M1A10	0	0	2	2	Yes
pQR711ABCμ-M1A11	0	1	1	2	No
pQR711ABCμ-M1A12	0	0	1	1	Yes
pQR711ABCμ-M1A13	0	1	3	4	No
pQR711ABCμ-M1A14	0	0	0	0	Yes
pQR711ABCμ-M1A15	0	0	0	0	Yes
pQR711ABCμ-M1A16	0	0	0	0	No
pQR711ABCμ-M1A17	1	0	1	2	No
pQR711ABCμ-M1A18	0	0	0	0	No
pQR711ABCμ-M1A19	0	1	0	1	No
pQR711ABCμ-M1A20	0	0	1	1	No

Table 6.1. Summary of the numbers of point mutations occurring in targets sites A, B, and C of pQR711ABCμ-M1A1 to M1A20. Sequences which contain deletions or additions are identified in the final column of the table.

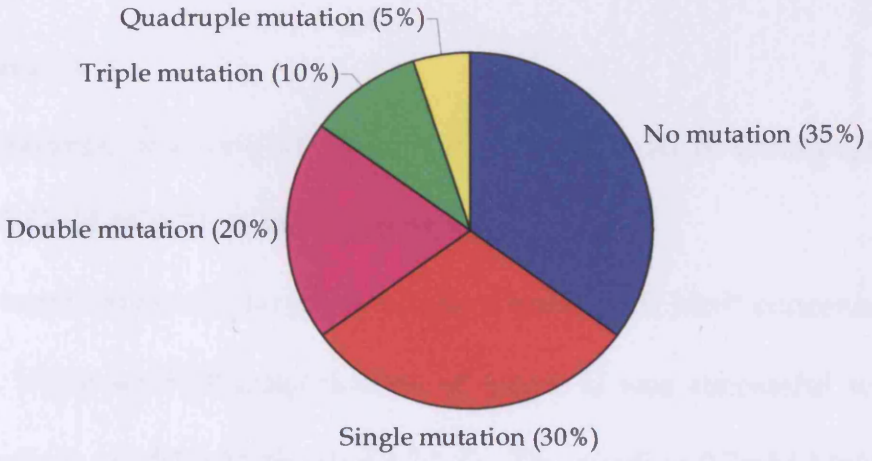


Figure 6.8. A pie chart illustrating the proportions of pQR711ABCμ-M sequences with single, double, triple, and quadruple point mutations in the overall target. The combined length of targets A, B, and C is 33 nucleotides. "No mutation (35%)" includes sequences that were frame-shifted.

6.3.2 Screening of libraries pQR711A$^{\mu}$-S, pQR711B$^{\mu}$-S, pQR711C$^{\mu}$-S, and pQR711ABC$^{\mu}$-M for the desired activity

The following numbers of variants were successfully screened using the procedure described in Figure 3.15: (1) 147 colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$-S; then (2) 84 colonies of *E. coli* XL10-Gold pQR711B$^{\mu}$-S (generated in Chapter 5); then (3) 230 colonies of *E. coli* XL10-Gold pQR711C$^{\mu}$-S; and finally (4) 2388 colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M (Section 6.2.2). The PeakNet 5.1 HPLC-management software did not identify a peak at (or around) 0.72 minutes (the anticipated retention time of ([*S*]-3,4-dihydroxybutan-2-one) in any of the chromatograms. This finding was confirmed by ranking the chromatograms from each library in order of final pyruvate concentration (Figures 6.9 to 6.12) and examining the lowest 10% in each case by eye.

6.4 Discussion

6.4.1 Preparation of *E. coli* XL10-Gold pQR711A$^{\mu}$-S, XL10-Gold pQR711C$^{\mu}$-S, and XL10-Gold pQR711ABC$^{\mu}$-M colonies

FepPCR amplification of target A was successful with Mn^{2+} concentrations of ≤0.8mM, while fepPCR amplification of target C was successful with Mn^{2+} concentrations of ≤0.9mM (Section 6.3.1.1). The purified 0.7mM Mn^{2+} fepPCR products (A0.7 and C0.7) were chosen to generate libraries pQR711A$^{\mu}$ and pQR711C$^{\mu}$ so that the mutation rates in these libraries would be equal to that of library pQR711B$^{\mu}$. The QuikChange SDM method was selected to create these libraries so that the 1pmpv sublibrary component of each

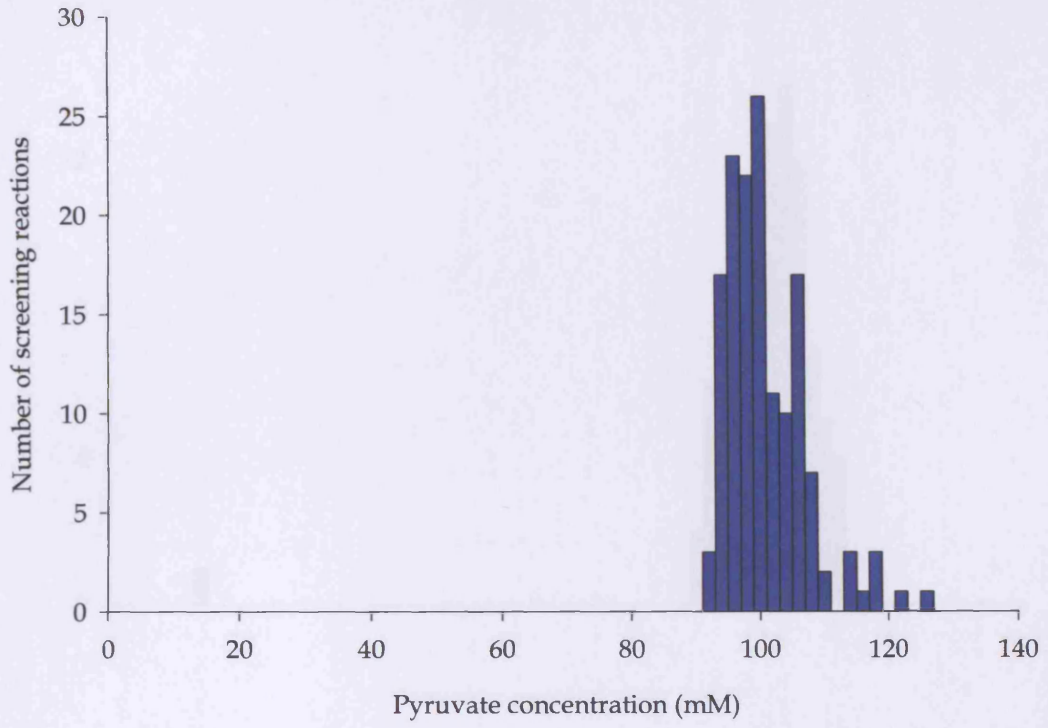


Figure 6.9. Histogram of the final pyruvate concentrations in the screening reactions of library *E. coli* XL10-Gold pQR711A$^{\mu}$-S. A total of 147 variants were screened.

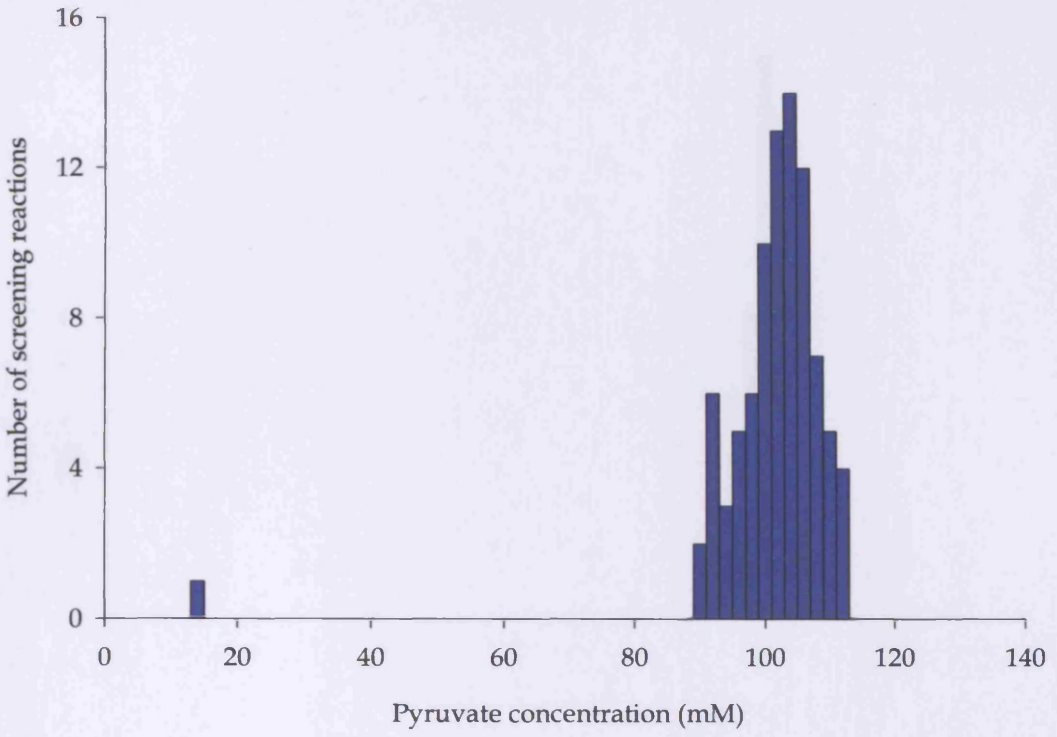


Figure 6.10. Histogram of the final pyruvate concentrations in the screening reactions of library *E. coli* XL10-Gold pQR711B$^{\mu}$-S. A total of 84 variants were screened. The chromatogram for variant pQR711B$^{\mu}$-S2B20 revealed a dramatically reduced donor substrate (pyruvate) peak (final concentration: 6.75mM), but no associated product ([*S*]-3,4-dihydroxybutan-2-one) peak.

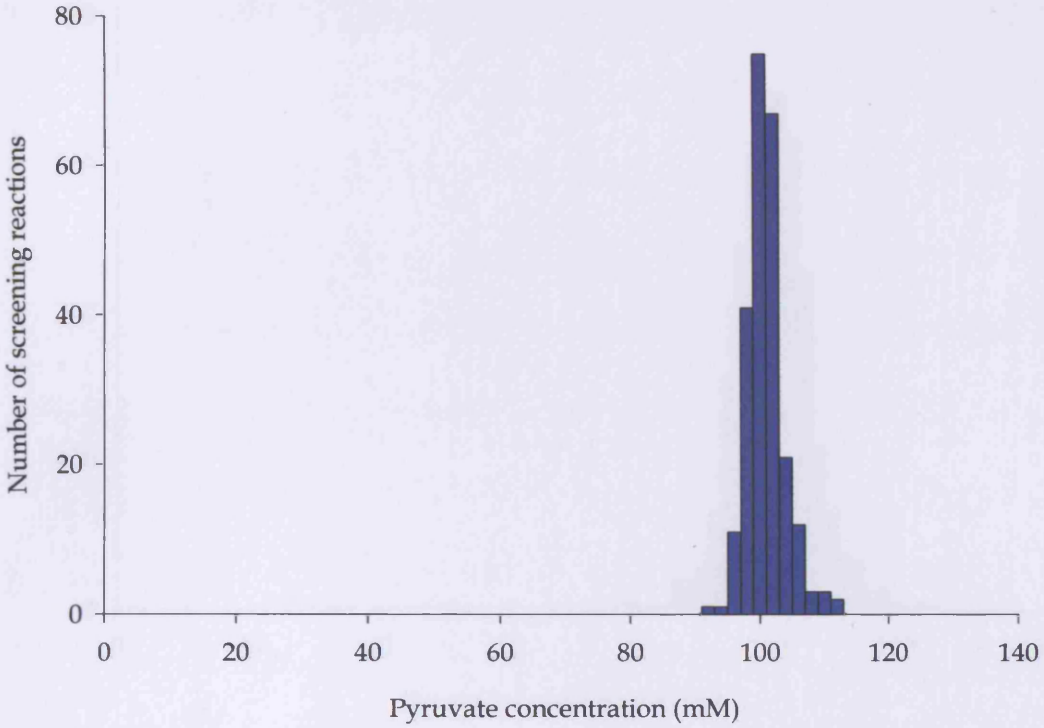


Figure 6.11. Histogram of the final pyruvate concentrations in the screening reactions of library *E. coli* XL10-Gold pQR711C$^{\mu}$-S. A total of 230 variants were screened.

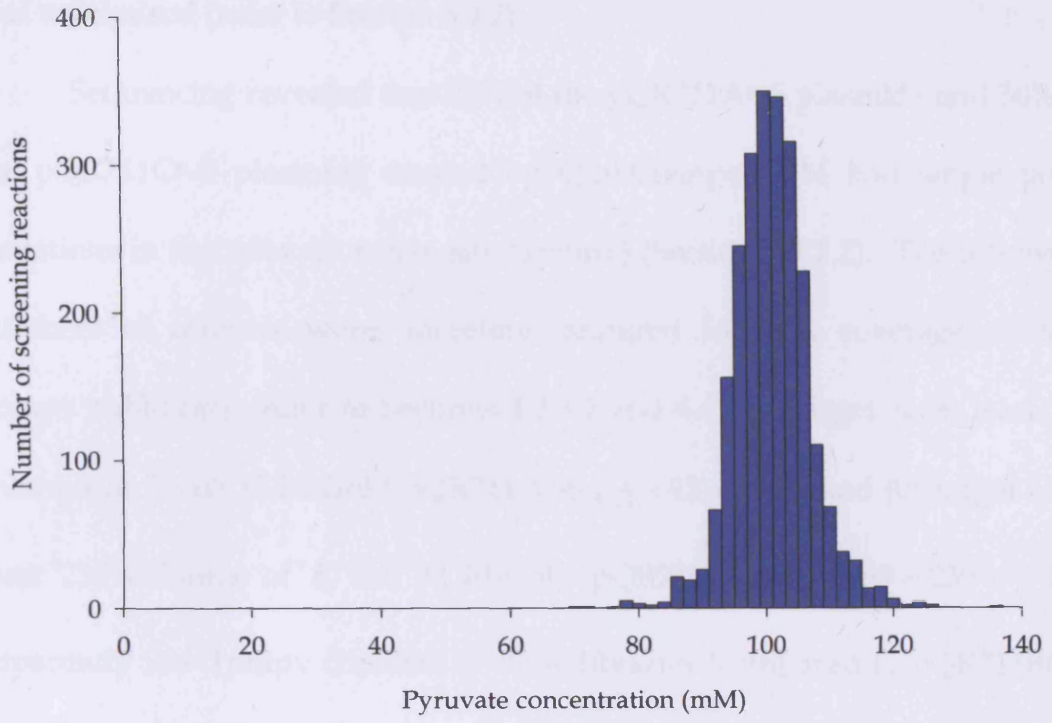


Figure 6.12. Histogram of the final pyruvate concentrations in the screening reactions of library *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M. A total of 2388 variants were screened.

was maximised (refer to Section 5.4.2).

Sequencing revealed that 20% of the pQR711A$^{\mu}$-S plasmids and 30% of the pQR711C$^{\mu}$-S plasmids created by QuikChange SDM had single point mutations in the relevant target site (1pmpv) (Section 6.3.1.2). The following numbers of colonies were, therefore, required for 90% coverage of each 1pmpv sublibrary (refer to Sections 4.3.3.2 and 4.4): (a) target A: at least 210 colonies of *E. coli* XL10-Gold pQR711A$^{\mu}$-S ($\frac{1}{0.2} \times 42 = 210$); and (b) target C: at least 230 colonies of *E. coli* XL10-Gold pQR711C$^{\mu}$-S ($\frac{1}{0.3} \times 69 = 230$). The apparently low 1pmpv contents of these libraries (compared to pQR711B$^{\mu}$-S) may have been due to one or more of the following: (a) random variation in the sample sets; (b) *Taq* DNAP having a variable misincorporation rate in 0.7mM Mn^{2+}, depending on the target sequence; or (c) the cloning step (QuikChange) being of variable efficiency, depending on the target sequence. The small size of the sample sets means that factor (a) almost certainly made a contribution; further deductions based upon these data would, therefore, be unreliable. Additional experiments that could be used to gauge the influences of factors (b) and (c) in the absence of factor (a) are described later.

The purified 0.7mM Mn^{2+} fepPCR products of targets A (A0.7), B (B0.7), and C (C0.7) were used in QuikChange MSDM to generate colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$. QuikChange MSDM was chosen for this task for two reasons: (a) it had been successful at generating a plasmid library from a fepPCR product (pQR711B$^{\mu}$-M); and (b) it was theoretically capable of incorporating the fepPCR products of targets A, B, and C into whole pQR711

plasmids simultaneously (Stratagene Ltd., 2003b). If QuikChange SDM had been chosen for this role instead, library pQR711ABC$^{\mu}$ would have had to have been generated by sequential mutagenesis of the three targets (pQR711A$^{\mu}$, then pQR711AB$^{\mu}$, and finally pQR711ABC$^{\mu}$). This process would have been much slower than QuikChange MSDM and introduced significantly more variables.

30% of the pQR711ABC$^{\mu}$-M plasmids generated by QuikChange MSDM had single point mutations (1pmpv) in the overall target (targets A, B, and C), 20% had double point mutations (2pmpv), 10% had triple point mutations (3pmpv), and 5% had quadruple point mutations (Section 6.3.1.3). The following numbers of colonies were, therefore, required for 90% coverage of the two sublibraries of interest (refer to Sections 4.3.3.2 and 4.4): (a) 1pmpv: at least 507 colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M ($\frac{1}{0.3} \times 152 = 507$); and (b) 2pmpv: at least 22680 colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M ($\frac{1}{0.2} \times 4536 = 22680$). In the 20 plasmids sequenced the overall distribution of point mutations was: four in target A, six in target B, and 14 in target C (a total of 24). In theory, an even point mutation density (point mutations per base-pair) across the three targets would have resulted in the following distribution: 6.5 in target A, 6.5 in target B, and 11 in target C. As with libraries pQR711A$^{\mu}$-S and pQR711C$^{\mu}$-S the small size of the data set precludes any meaningful discussion of variations between these two distributions.

The random variation that almost certainly impacted on the sequence analyses of libraries pQR711A$^{\mu}$-S, pQR711C$^{\mu}$-S, and pQR711ABC$^{\mu}$-M (this

Chapter) and pQR711B$^{\mu}$-S (refer to Chapter 5) calls into question the accuracy of the subsequent calculations for library coverage numbers. The levels of library coverage that are stated as being achieved in this project (refer to Section 6.4.2) are, therefore, only approximate. In future applications of fepPCR-QuikChange accuracy could be improved either by sequencing more library members or doing the following: (a) ensuring that all of the target sites contain a restriction enzyme recognition site (targets A and C did not); and (b) measuring the mutation density of each target site in the plasmid library (post-QuikChange) by PCR amplification (using the fepPCR primers) and subsequent restriction analysis (see Section 5.4.1). There are two advantages of this approach over extra sequencing: (a) it would be far less expensive; and (b) it would allow accurate determination of the mutation densities in the target sites at each stage of the process (post-fepPCR and post-QuikChange). Sequencing of library members would, however, still be required to ensure that mutation had been effectively focused to the target sites.

6.4.2 Screening of libraries pQR711A$^{\mu}$-S, pQR711B$^{\mu}$-S, pQR711C$^{\mu}$-S, and pQR711ABC$^{\mu}$-M for the desired activity

The following (approximate) levels of sublibrary coverage were achieved during the screening phase of this project: (a) ~81% of the pQR711A$^{\mu}$ 1pmpv sublibrary; (b) ~90% of the pQR711B$^{\mu}$ 1pmpv sublibrary; and (c) ~90% of the pQR711C$^{\mu}$ 1pmpv sublibrary (Section 6.3.2). Not a single HPLC chromatogram exhibited a product ([*S*]-3,4-dihydroxybutan-2-one) peak. One

variant (pQR711B$^{\mu}$-S2B20) did exhibit an appreciable drop in donor substrate (pyruvate) concentration, but attempts to repeat this observation with new cultures failed. These findings suggested that a single point mutation in the target sites was unlikely to yield a variant of *E. coli* transketolase with the desired activity (Figure 1.14).

Next, the 2pmpv sublibrary of pQR711ABC$^{\mu}$ was screened. 2388 colonies of *E. coli* XL10-Gold pQR711ABC$^{\mu}$-M provided ~22% coverage of the pQR711ABC$^{\mu}$ 2pmpv sublibrary.[15] Unfortunately, not a single HPLC chromatogram exhibited a product peak or an appreciable drop in donor substrate concentration.

If it is assumed that the screening system did not suffer from a fundamental flaw, then the following may be deduced: (a) a variant of transketolase with the desired activity was not among the members of the four libraries that were screened; (b) it is extremely unlikely that any single point mutation in the codons of residues Ser24-His26, Gly99-Pro101, or Asp469-His473 of *E. coli* transketolase would yield a variant with the desired activity; and (c) at least 22% of variants with double point mutations do not have the desired activity. As the screening system has successfully revealed transketolase activity with only one modification (β-HPA as the donor substrate), the only conceivable flaw with it is that the product (*S*)-3,4-dihydroxybutan-2-one does not yield a peak at (or around) 0.72 minutes, perhaps due to a low exctinction coefficient at 210nm wavelength or unusual

[15] The target of 22680 colonies (90% coverage) was not achieved owing to equipment failures and time constraints.

behaviour in the HPLC column. As it is almost identical in structure to L-erythrulose these possibilities seem unlikely, but the acquisition of a sample of this compound is absolutely essential for the HPLC method to be validated and the conclusions of this chapter to be confirmed.

Chapter 7 - General discussion

7.1 Overall summary of this project

Transketolase has significant potential as an industrial biocatalyst. It transfers a C_2 (1,2-dihydroxyethyl) moiety from a donor substrate to a wide range of acceptor substrates in a stereospecific manner. 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 to modify the donor substrate-specificity of *E. coli* transketolase so that it would accept pyruvate as a donor substrate. Pyruvate is vastly cheaper than β-HPA but is chemically similar and would still render the reaction irreversible.

E. coli transketolase required two modifications for it to be able to catalyse the desired reaction (Figure 1.14): (a) the addition of the ability to synthesise 1-hydroxyethyl ThDP from pyruvate and ThDP; and (b) the elimination of the activity that splits this compound into acetaldehyde and ThDP. It was hoped that these modifications could be brought about by the directed evolution of three stretches of residues that lie close to the C2-hydroxyl group of the standard reaction intermediate (1,2-dihydroxyethyl ThDP). It is this hydroxyl group that is absent from pyruvate and its (potential) reaction intermediate.

The three stretches of residues chosen for mutagenesis were: Ser24–His26 (target A), Gly99–Pro101 (target B), and Asp469–His473 (target C)

(Chapter 4). Primers were designed to flank these targets and they were amplified by PCR in the presence of 0.7mM Mn^{2+} (fepPCR) (Chapters 5 and 6). Restriction analysis of the target B product revealed that these cations had lowered the fidelity of *Taq* DNAP approximately 1270-fold: its frequency of misincorporation was 0.14 per nucleotide (Chapter 5). The fepPCR products of targets A, B, and C were simultaneously incorporated into whole pQR711 plasmids using QuikChange MSDM, thus creating the library pQR711ABC$^{\mu}$ (Chapter 6). Three additional libraries were prepared from the individual fepPCR products: libraries pQR711A$^{\mu}$, pQR711B$^{\mu}$, and pQR711C$^{\mu}$ (Chapters 5 and 6). As only single products were being cloned, QuikChange SDM was chosen to generate these libraries.

A high-throughput screen for the desired activity was developed at the same time as the libraries were constructed. The screen has the following steps: (1) transformation of the plasmid library into *E. coli* XL10-Gold competent cells; (2) microwell culture of individual colonies; (3) lysis of the cultures; (4) incubation of the lysates with cofactors and the target substrates (pyruvate and glycolaldehyde); and finally (5) high-throughput HPLC analysis measuring donor substrate (pyruvate) depletion. The first four steps were optimised to ensure the highest possible concentration of holotransketolase in each screening reaction (Chapter 3).

The sublibrary compositions of the libraries were determined by sequencing and the numbers of colonies required for 90% coverage of certain sublibraries were calculated: (a) 210 for the pQR711A$^{\mu}$ 1pmpv sublibrary; (b)

84 for the pQR711B$^{\mu}$ 1pmpv sublibrary; (c) 230 for the pQR711C$^{\mu}$ 1pmpv sublibrary; and (d) 22680 for the pQR711ABC$^{\mu}$ 2pmpv sublibrary (Chapters 5 and 6). It was anticipated that the single-target 1pmpv sublibraries would be screened first to see whether the desired activity could be obtained through a single point mutation in either target A, target B, or target C. The larger pQR711ABC$^{\mu}$ 2pmpv sublibrary - with potential synergistic alliances between mutated residues - would be screened afterwards.

The following levels of coverage were actually obtained during the screening step: (a) ~81% of the pQR711A$^{\mu}$ 1pmpv sublibrary; (b) ~90% of the pQR711B$^{\mu}$ 1pmpv sublibrary; (c) ~90% of the pQR711C$^{\mu}$ 1pmpv sublibrary; and (d) ~22% coverage of the pQR711ABC$^{\mu}$ 2pmpv sublibrary (Chapter 6). Unfortunately, not a single variant was identified that possessed the desired activity. The ramifications of this finding and possible extensions to this project are described below.

7.2 Overall conclusions

The following conclusions were reached during the development of the high-throughput screen for the desired activity (Figure 1.14):

a) Of the *E. coli* strains tested, XL10-Gold was the best host for the pQR711 plasmid. After adjusting for culture turbidity, *E. coli* XL10-Gold pQR711 sonicate was found to have a transketolase activity approximately double that of JM107 pQR711 sonicate (the original strain used for transketolase-mediated condensations at UCL). *E. coli*

XL10-Gold was, therefore, chosen to host the libraries.

b) *E. coli* XL10-Gold pQR711 cultures were grown to substantial turbidities in 96- and 384-well plates. A mean final OD_{600} of 3.41ODU was achieved in 60µl cultures agitated at 1400rpm for 16 hours in a 37°C incubator.[16]

c) Of the lytic methods tested, sonication proved to be the most effective at releasing transketolase from *E. coli* XL10-Gold pQR711 culture. However, a single cycle of freezing at -80°C and thawing was judged to be the most efficient way of lysing microwell cultures in a high-throughput manner.

d) A spectrophotometric assay for transketolase activity, based on the pH change that accompanies transketolase-mediated condensation, proved to be unreliable. Further work may allow the potential of this approach to be realised.

e) An accurate, high-throughput HPLC assay for the desired activity was developed. Based on a 50mm guard column, it is capable of processing one sample in 1.2 minutes.

The following conclusions were reached during the development of fepPCR-QuikChange and the construction of transketolase libraries:

a) The residues most likely to perform the role of β-HPA/pyruvate discrimination in *E. coli* transketolase are His26, His100, Asp469, and

[16] Researchers at UCL are currently developing new methods of microwell culture that are better characterised and more effective than the method described here.

His473. These residues lie within 5.61Å of the C2-hydroxyl group of the 1,2-dihydroxyethyl ThDP reaction intermediate in the highly-homologous *S. cerevisiae* transketolase.

b) The best targets for mutagenesis were reasoned to be Ser24–His26, Gly99–Pro101, and Asp469–His473. These targets incorporated the four residues of primary importance, plus residues either side to allow flexibility or side-chain repacking.

c) FepPCR of each of the three target sites was possible in Mn^{2+} concentrations of up to 0.7–0.9mM with higher concentrations causing more point mutations. The misincorporation rate of *Taq* polymerase in 0.7mM Mn^{2+} was estimated to be 0.14 per nucleotide.

d) Overall, the fepPCR-QuikChange technique proved to be effective at creating plasmid libraries with mutation focused to specific sites. Not only were mutations limited to the target stretches, but it was possible to derive large numbers of colonies from each cloning reaction: more than 1×10^5 in every case. QuikChange MSDM was successful at assimilating the fepPCR products of multiple targets into whole plasmids simultaneously.

The following conclusions were reached during the screening step:

a) The first phase did not reveal the desired activity in the following sublibraries: (i) ~81% of the pQR711A$^{\mu}$ 1pmpv sublibrary; (ii) ~90% of the pQR711B$^{\mu}$ 1pmpv sublibrary; and (iii) ~90% of the pQR711C$^{\mu}$

1pmpv sublibrary. This suggested that a single point mutation in any of the target sites was unlikely to yield an active variant.

b) The second phase of screening (library pQR711ABC$^{\mu}$) provided complete (~100%) coverage of the 1pmpv sublibrary. This extra coverage failed to yield an active variant so it was concluded that no single point mutation in any of the target sites was capable of yielding the desired activity.

c) The second phase of screening also provided ~22% coverage of the pQR711ABC$^{\mu}$ 2pmpv sublibrary. As no active variants were identified it was concluded that one fifth of all possible pQR711ABC$^{\mu}$ double point mutation variants do not possess the desired activity.

The failure of this project to identify a single variant of transketolase with the desired activity (Figure 1.14) 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. If it is assumed that the formation of 10mM (*S*)-3,4-dihydroxybutan-2-one in a screening reaction would have been detected, then a variant yielding less product than this would have had an activity of <0.102mkat.m^{-3} in a 0.25ODU culture sonicate: 126-fold less than the activity of a wild-type *E. coli* XL10-Gold pQR711 sonicate with β-HPA. Such a variant would have been of little use as a biocatalyst, although it may have been suitable for further fine tuning by directed evolution.

It is also possible that a useful variant existed in the portion of the 2pmpv sublibrary that was not screened. However, the absence of even a low-activity variant amongst those that were screened does not encourage optimism.

A possible explanation for the dearth of variants with detectable activities is that the project goal was too ambitious for the sublibraries that were screened. The project goal was ambitious for two reasons: (a) two significant modifications to the enzyme functionality were required; and (b) the enzyme did not exhibit the desired activity to any extent to begin with.

7.3 Future work

Possible extensions to this project that might enable the achievement of the project goal (Figure 1.14) are:

a) Screening a pQR711ABC$^{\mu}$ sublibrary with a higher point mutation density, such as 3pmpv, or

b) Increasing the number of target sites or expanding the existing ones, or

c) Using random oligonucleotide mutagenesis, rather than fepPCR, to create a library that accesses all twenty amino acids at the four primary residues (His26, His100, Asp469, and His473). The total number of unique protein sequences in such a library would be 1.6×10^5 ($20^4 = 1.6 \times 10^5$).

The latter three of these extensions would require a screening system that was

capable of handling larger numbers of screening reactions than the existing one. Possible improvements to the system are:

a) Using the spectrophotometric assay for transketolase activity. This would require a massive improvement in the reliability of the assay, but it would allow multiple 384-well plates to be screened simultaneously.

b) Culturing library members in 1536-well plates. Good turbidities may be achieved if a high shake speed is used (>1400rpm); one way of attaining such a speed would be through the use of a pneumatic shaker.[17]

c) Pooling cultures before screening. If 1536-well plates were used with, for example, four cultures per screening reaction, it would allow 6144 variants to be screened per plate.

Alternatively, the ambitiousness of the goal could be reduced. There are two ostensible ways to achieve this. The first approach is to replace the starting point of the directed evolution process - wild-type transketolase - with a mutant that exhibits a small amount of the desired activity (Figure 1.14). It may be possible to generate such a mutant through rational protein engineering using the active sites of pyruvate-accepting ThDP-dependent enzymes as templates (see Section 4.1). The second approach is to target acceptor substrate-specificity instead. Such a goal would be "softer" because

[17] Techelan LLC manufactures microwell plate shakers that are capable of shake speeds in excess of 10000rpm.

a substrate with some initial (wild-type) activity could be used.

Whether a more complex library is screened for the desired activity (Figure 1.14), the ambitiousness of the project goal is reduced, or an entirely new enzyme selected for direct evolution, it would be extremely interesting to see whether fepPCR-QuikChange can deliver the step enhancements that it promises.

References

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, *Nucleic Acids Res. 28*, E88.

Arnold F.H. (1998) Design by directed evolution, *Acc. Chem. Res. 31*, 125-131.

Beckman R.A., Mildvan A.S., Loeb L.A. (1985) On the fidelity of DNA replication: manganese mutagenesis *in vitro*, *Biochemistry 24*, 5810-17.

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

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

Bolte J., Demuynck C., Samaki H. (1987) Utilization of enzymes in organic chemistry: transketolase catalyzed synthesis of ketoses, *Tetrahedron Lett. 28*, 5525-5528.

Brocklebank S., Woodley J.M., Lilly M.D. (1999) Immobilised transketolase for carbon-carbon bond synthesis: biocatalyst stability, *J. Mol. Cat. B - Enzymatic 7*, 223-231.

Cadwell R.C., Joyce G.F. (1995) Mutagenic PCR. In: Dieffenbach C.W., Dveksler G.S. (eds.) PCR primer: a laboratory manual, *Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York, USA)*.

Cedrone F., Ménez A., Quéméneur E. (2000) Tailoring new enzyme function by rational redesign, *Curr. Opin. Struct. Biol. 10*, 405-410.

Chauhan R.P., Powell L.W., Woodley J.M. (1997) Boron based separations for *in situ* recovery of L-erythrulose from transketolase-catalyzed condensation, *Biotechnol. Bioeng. 56*, 345-351.

Cheetham P. (1994) Case studies in applied biocatalysis - from ideas to products. In: Cabral J., Best. D., Boross L., Tramper H. (eds.) Applied biocatalysis, *Harwood Academic Publishers (London, UK)*.

Chen K., Arnold F. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide, *Proc. Natl. Acad. Sci. USA 90*, 5618-5622.

Clackson T., Ultsch M.H., Wells J.A., de Vos A.M. (1998) Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity, *J. Mol. Biol. 277*, 1111-1128.

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

Crameri A., Raillard S.A., Bermudez E., Stemmer W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution, *Nature 391*, 288-291.

Dalby P.A. (2003) Optimising enzyme function by directed evolution, *Curr. Opin. Struc. Biol. 13*, 500-505.

Dalmas V., Demuynck C. (1993) An efficient synthesis of sedoheptulose catalyzed by spinach transketolase, *Tetrahedron: Asymmetry 4*, 1169-1172.

de la Haba G., Leder I.G., Racker E. (1955) Crystalline transketolase from bakers yeast: isolation and properties, *J. Biol. Chem. 214*, 409-426.

Demuynck C., Bolte J., Hecquet L., Samaki H. (1990) Enzymes as reagents in organic chemistry: transketolase-catalysed synthesis of D-[1,2-$^{13}C_2$]xylulose, *Carbohyd. Res. 206*, 79-85.

Ding K.L., Ishii A., Mikami K. (1999) Super high throughput screening (SHTS) of chiral ligands and activators: Asymmetric activation of chiral diol-zinc catalysts by chiral nitrogen activators for the enantioselective addition of diethylzinc to aldehydes, *Angew. Chem. Int. Edit. 38*, 497-501.

Domagk G.F., Horecker B.L. (1965) Fructose and erythrose metabolism in *Alcaligensis faecalis*, *Arch. Biochem. Biophys. 109*, 342-349.

Draths K.M., Frost J.W. (1990) Synthesis using plasmid-based biocatalysis: plasmid assembly and 3-deoxy-D-arabino-heptulosonate production, *J. Am. Chem. Soc. 112*, 1657-1659.

Dybkaer R. (2001) Unit "katal" for catalytic activity (IUPAC Technical Report), *Pure Appl. Chem. 73*, 927-931.

Effenberger F., Null V. (1992) Enzyme-catalyzed reactions .13. a new, efficient synthesis of fagomine, *Liebigs Ann. Chem. 11*, 1211-1212.

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

Fiedler E., Thorell S., Sandalova T., Golbik R., König S., Schneider G. (2002) Snapshot of a key intermediate in enzymatic thiamin catalysis: Crystal structure of the α-carbanion of (α,β-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from Saccharomyces cerevisiae, *Proc. Natl. Acad. Sci. USA 99*, 591-595.

Fisher R.A. (1930) The genetical theory of natural selection, *Oxford University Press (Oxford, UK)*.

French C., Ward J.M. (1995) Improved production and stability of *E. coli* recombinants expressing transketolase for large scale biotransformation, *Biotechnol. Lett. 17*, 247-252.

Friard, O. (2004) AnnHyb, *http://bioinformatics.org/annhyb/*.

Gerhardt S., Echt S., Busch M., Freigang J., Auerbach G., Bader G., Martin W.F., Bacher A., Huber R., Fischer M. (2003) Structure and properties of an engineered transketolase from maize, *Plant Physiol. 132*, 1941-1949.

Giver L., Gershenson A., Freskgard P.-O., Arnold F.H. (1998) Directed evolution of a thermostable esterase, *Proc. Natl. Acad. Sci. USA 95*, 12809-12813.

Greener A., Callahan M., Jerpseth B. (1997) An efficient random mutagenesis technique using an *E. coli* mutator strain, *Mol. Biotechnol. 7*, 189-195.

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

Griffiths J.S., Cheriyan M., Corbell J.B., Pocivavsek L., Fierke C.A., Toone E.J. (2004) A bacterial selection for the directed evolution of pyruvate aldolases, *Bioorg. Med. Chem. 12*, 4067-4074.

Guérard 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.

Gyamerah M., Willetts A.J. (1997) Kinetics of overexpressed transketolase from *Escherichia coli* JM 107/pQR 700, *Enzyme Microbial. Technol. 20*, 127-134.

Hayes R.J., Bentzien J., Ary M.L., Hwang M.Y., Jacinto J.M., Vielmetter J., Kundu A., Dahiyat B.I. (2002) Combining computational and experimental screening for rapid optimization of protein properties, *Proc. Natl. Acad. Sci. USA 99*, 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.

Hecquet L., Demuynck C., Schneider G., Bolte J. (2001) Enzymatic syntheses of ketoses: study and modification of the substrate specificity of transketolase from *Saccharomyces cerevisiae*, *J. Mol. Cat. B - Enzymatic 11*, 771–776.

Henikoff S., Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks, *Proc. Natl. Acad. Sci. USA 89*, 10915–10919.

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

Heinrich P.C., Steffen H., Janser P., Wiss O. (1972) Studies on the reconstitution of apotransketolase with thiamine pyrophosphate and analogs of the coenzyme, *Eur. J. Biochem. 30*, 533–541.

Hobbs G.R., Lilly M.D., Turner N.J., Ward J.M., Willets A.J., Woodley J.M. (1993) Enzyme-catalyzed carbon-carbon bond formation - use of transketolase from *Escherichia coli*, *J. Chem. Soc. Perk. T. 1*, 165–166.

Hobbs G.R., Mitra R.K., Chauhan R.P., Woodley J.M., Lilly M.D. (1996) Enzyme-catalyzed 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.

Hult K., Berglund P. (2003) Engineered enzymes for improved organic synthesis, *Curr. Opin. Struct. Biol. 14*, 395–400.

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 Lett. 41*, 4481–4485.

Innis M.A., Gelfand D.H., Sninsky J.J., White T.J. (1990) PCR Protocols: Guide to Methods and Applications, *Academic Press (New York, USA)*.

Jander G., Norris S.R., Joshi V., Fraga M., Rugg A., Yu S., Li L., Last R.L. (2004) Application of a high-throughput HPLC-MS/MS assay to *Arabidopsis* mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality, *Plant J. 39*, 465–475.

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

Kiely M.E., Tan E.L., Wood T. (1969) The purification of transketolase from *Candida utilis*, *Can. J. Biochem. 47*, 455–460.

Kobori Y., Myles D.C., Whitesides G. (1992) Substrate specificity and carbohydrate synthesis using transketolase, *J. Org. Chem. 57*, 5899–5907.

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

Lindqvist Y., Schneider G., Ermler U., Sundström M. (1992) Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 Å resolution, *EMBO J. 11*, 2373–2379.

Littlechild J.A., Turner N.J., Hobbs G.R., Lilly M.D., Rawas R., Watson W. (1995) Crystallization and preliminary X-ray crystallographic data of *Escherichia coli* transketolase, *Acta. Cryst. D51*, 1074–1076.

Lutz S., Ostermeier M., Benkovic S.J. (2001a) Rapid generation of incremental truncation libraries for protein engineering using α-phosphothioate nucleotides, *Nucleic Acids Res. 29*, 16e.

Lutz S., Ostermeier M., Moore G.L., Maranas C.D., Benkovic S.J. (2001b) Creating multiple-crossover DNA libraries independent of sequence identity, *Proc. Natl. Acad. Sci. USA 98*, 11248–11253.

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.T., Arnold F.H. (2000) Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine, *Nat. Biotechnol. 18*, 317–320.

Merck (2004) Yeast extract, granulated, *http://service.merck.de/microbiology/tedisdata/prods/4973-1_03753_0500_9025.html*.

Meshalkina L.E., Neef H., Tjaglo M.V., Schellenberger A., Kochetov G.A. (1995) The presence of a hydroxyl group at the C-1 atom of the transketolase substrate molecule is necessary for the enzyme to perform the transferase reaction, *FEBS Lett. 375*, 220–222.

Mikami K., Angelaud R., Ding K.L., Ishii A., Tanaka A., Sawada N., Kudo K., Senda M. (2001) Asymmetric activation of chiral alkoxyzinc catalysts by chiral nitrogen activators for dialkylzinc addition to aldehydes: Super high-throughput screening of combinatorial libraries of chiral ligands and activators by HPLC-CD/UV and HPLC-OR/RIU systems, *Chem.-Eur. J. 7*, 730-737.

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

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

Mitra R.K., Woodley J.M. (1996) A useful assay for transketolase in asymmetric syntheses, *Biotechnol. Tech. 10*, 167-172.

Mitra R.K., Woodley J.M., Lilly M.D. (1998) *Escherichia coli* transketolase-catalyzed carbon-carbon bond formation: biotransformation characterization for reactor evaluation and selection, *Enzyme Microb. Tech. 22*, 64-70.

Miyazaki K., Takenouchi M. (2002) Creating random mutagenesis libraries using megaprimer PCR of whole plasmid, *BioTechniques 33*, 1033-1038.

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

Müller Y.A., Lindqvist Y., Furey W., Schulz G., Jordan F., Schneider G. (1993) The thiamin diphosphate binding fold, *Structure 1*, 95-103.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Erlich H. (1986) Specific enzymatic amplification of DNA *in vitro* - the polymerase chain-reaction, *Cold. Spring. Harb. Sym. 51*, 263-273.

Murphy D.J., Walker D.A. (1982) The properties of transketolase from photosynthetic tissue, *Planta 155*, 316-320.

Nikkola M., Lindqvist Y., Schneider G. (1994) Refined structure of transketolase form *Saccharomyces cerevisiae* at 2.0 Å resolution, *J. Mol. Biol. 238*, 387-404.

Nilsson U., Meshalkina., Lindqvist Y., Schneider G. (1997) Examination of substrate binding in thiamine diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis, *J. Biol. Chem. 272*, 1864-1869.

Nilsson U., Hecquet L., Gefflaut T., Guerard C., Schneider G. (1998) Asp477 is a determinant of the enantioselectivity in yeast transketolase, *FEBS Lett. 424*, 49–52.

Oliphant A.R., Nussbaum A.L., Struhl K. (1986) Cloning of random-sequence oligodeoxynucleotides, *Gene 44*, 177–183.

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

Paoletti F., Aldinucci D. (1986) Immunoaffinity purification of rat liver transketolase: evidence for multiple forms of the enzyme, *Arch. Biochem. Biophys. 245*, 212–219.

Patrick W.M., Firth A.E., Blackburn J.M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries, *Protein Eng. 16*, 451–457.

QIAGEN Ltd. (2001) Bench guide, 91.

Racker E. (1961) Transketolase. In: Boyer P.D., Lardy H., Myrbõck K. (eds.) The Enzymes, *Academic Press (New York, USA)*.

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

Roche Diagnostics Ltd. (2000) *Taq* DNA polymerase information sheet.

Sambrook J., Fritsch E.F., Maniatis T. (1989) Molecular Cloning: A Laboratory Manual, *Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York, USA)*.

Santoro S.W., Schultz P.G. (2002) Directed evolution of the site specificity of Cre recombinase, *Proc. Natl. Acad. Sci. USA 99*, 4185–4190.

Sayle R.A., Milner-White E.J. (1995) RasMol: Biomolecular graphics for all, *Trends Biochem. Sci. 20*, 374–376.

Schenk G., Layfield R., Candy J.M., Duggleby R.G., Nixon P.F. (1997) Molecular evolutionary analysis of the thiamine-diphosphate-dependent enzyme, transketolase, *J. Mol. Evol. 44*, 552–572.

Schneider G., Lindqvist Y. (1993) Enzymatic thiamine catalysis: mechanistic implications from the three-dimensional structure of transketolase, *Bioorg. Chem. 21*, 109–117.

Schneider G., Lindqvist Y. (1998) Crystallography and mutagenesis of transketolase: mechanistic implications for enzymatic thiamin catalysis, *Biochim. Biophys. Acta 1385*, 387-398.

Sevestre A., Hélaine V., Guyot G., Martin C., Hecquet L. (2003) A fluorogenic assay for transketolase from *Saccharomyces cerevisiae*, *Tetrahedron Lett. 44*, 827-830.

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

Shao Z., Arnold F.H. (1996) Engineering new functions and altering existing functions, *Curr. Opin. Struct. Biol. 6*,513-518.

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

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

Sieber V., Martinez C.A., Arnold F.H. (2001) Libraries of hybrid proteins from distantly related sequences, *Nat. Biotechnol. 19*, 456-460.

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

Singleton C.K., Wang J.J.L., Shan L., Martin P.R. (1996) Conserved residues are functionally distinct within transketolases of different species, *Biochemistry 35*, 15865-15869.

Sneeden J.L., Loeb L.A. (2003) Random oligonucleotide mutagenesis. In: Arnold F.H., Georgiou G. (eds.) Directed evolution library creation: methods and protocols, *Humana Press (New Jersey, USA)*.

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)*.

Sprenger G.A. (1993) Nucleotide sequence of the *Escherichia coli* K-12 transketolase (*tkt*) gene, *Biochim. Biophys. Acta 1216*, 307-310.

Sprenger G.A., Schörken 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.

Sprenger G.A., Schorken U., Wiegert T., Grolle S., de Graaf A.A., Taylor S.V., Begley T.P., Bringer-Meyer S., Sahm H. (1997) Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol, *Proc. Natl. Acad. Sci. USA 94*, 12857–12862.

Stemmer W.P.C. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling, *Nature 370*, 389–391.

Stinson S.C. (2000) Chiral drugs, *Chem. Eng. News 78*, 55–78.

Stratagene Ltd. (2003a) QuikChange® Site-Directed Mutagenesis Kit, catalogue number 200519.

Stratagene Ltd. (2003b) QuikChange® Multi Site-Directed Mutagenesis Kit, catalogue number 200514.

Sundström M., Lindqvist Y., Schneider G., Hellman U., Ronne H. (1993) Yeast *TKL1* gene encodes a transketolase that is required for efficient glycolysis and biosynthesis of aromatic amino acids, *J. Biol. Chem. 268*, 24346–24352.

Sutcliffe M.J., Scrutton N.S. (2000) Enzyme catalysis: over-the-barrier or through-the-barrier?, *Trends Biochem. Sci. 25*, 405–408.

Sutcliffe M.J., Scrutton N.S. (2002) A new conceptual framework for enzyme catalysis. Hydrogen tunnelling coupled to enzyme dynamics in flavoprotein and quinoprotein enzymes, *Eur. J. Biochem. 269*, 3096–3102.

Tindall K.R., Kunkel T.A. (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase, *Biochemistry 27*, 6008–6013.

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

Villafranca J.J., Axelrod B. (1971) Heptulose synthesis from nonphosphorylated aldoses and ketoses by spinach transketolase, *J. Biol. Chem. 246*, 3126–3131.

Voigt C.A., Mayo S.L., Arnold F.H., Wang Z.G. (2001a) Computational method to reduce the search space for directed protein evolution, *Proc. Natl. Acad. Sci. USA 98*, 3778–3783.

Voigt C.A., Mayo S.L., Arnold F.H., Wang Z.G. (2001b) Computationally focusing the directed evolution of proteins, *J. Cell Biochem. Suppl. 37*, 58–63.

Wada M., Hsu C.-C., Franke D., Mitchell M., Heine A., Wilson I.A., Wong C.-H. (2003) Directed evolution of *N*-acetylneuraminic acid aldolase to catalyze enantiomeric aldol reactions, *Bioorg. Med. Chem. 11*, 2091–2098.

Wikner C., Meshalkina L., Nilsson U., Bäckström Y., Lindqvist Y., Schneider G. (1995) His103 in yeast transketolase is required for substrate recognition and catalysis, *Eur. J. Biochem. 233*, 750–755.

Wikner C., Nilsson U., Meshalkina L., Udekwu C., Lindqvist Y., Schneider G. (1997) Identification of catalytically important residues in yeast transketolase, *Biochemistry 36*, 15643–15649.

Williams G.J., Domann S., Nelson A., Berry A. (2003) Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution, *Proc. Nat. Acad. Sci. USA 100*, 3143–3148.

Woodley J.M., Mitra R.K., Lilly M.D. (1996) Carbon-carbon bond synthesis - reactor design and operation for transketolase-catalyzed biotransformations, *Ann. Ny. Acad. Sci. 799*, 434–445.

Zhang J.-H., Dawes G., Stemmer W.P.C. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening, *Proc. Natl. Acad. Sci. USA 94*, 4504–4509.

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

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

Zimmermann 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.