Application of *in vivo* selective methods to investigate novel activity in the enolase superfamily

Sylvain F. R. M. Royer

A thesis submitted for the Degree of

Doctor of Philosophy

University of Edinburgh

January 2005

Abstract

High-throughput screening is now widely used to isolate specific or improved activities of various enzymes. However, most of the current screening methods are based on *in vitro* procedures. The aim of this project is the development of an *in vivo* directed evolution approach to study a novel racemase enzyme.

Although *in vivo* selection in itself is not new, this project emphasises a new approach to the high throughput *in vivo* concept, selecting an enzyme activity apparently unintended by nature. The selection, which can screen a library of 10^6-10^8 mutants per assay, is complementary to *in vitro* screening, and applies new genetic tools for strain construction. Since the racemase under investigation is active on commercially important compounds, this method may also be suitable to identify and improve new industrial biocatalysts. The focus of the study is an *N*-acylamino acid racemase from *Amycolatopsis* which displays relaxed substrate specificity upon *N*-acetylated amino acids but no activity towards free amino acids. As such, this enzyme is potentially valuable to use with well established enantioselective acylases for the dynamic kinetic resolution of L- or D-amino acids. However, the racemase displays properties which limit its usefulness as a biocatalyst, such as a low substrate turnover, and inhibition of activity at substrate concentrations exceeding 50mM. The racemase also appears to have evolved serendipitously from a different activity of the enolase superfamily.

The *in vivo* selection developed in this work is based on the construction of specific auxotrophic mutants, which require expression of this racemase activity for growth. This offers a way to explore this very interesting and novel *N*-acylamino acid racemase activity of an enzyme in the enolase superfamily. The screening of libraries of variants derived from the wild type enzyme was applied, to unveil mutations which improve activity, and demonstrate the value of the *in vivo* approach. This *in vivo* selection could therefore potentially help to understand the evolution of new activities in the enolase superfamily and develop a new class of biocatalyst to expand the utility of acylase resolutions by industry.

Acknowledgements

I would first of all like to thank Professor Nick Turner for giving me the opportunity to carry out this research. Thanks also to Dr Ian Fotheringham for his valuable input, help, advice and support, and who taught me the microbiology and genetic tools I used in this research.

I am indebted to Dr Ian Fotheringham for his patience and dedication in the proofreading of this thesis.

I would also like to thank all the members of the Turner-Flitsch group I met in Edinburgh, and everyone in the laboratories I have worked in for their help and kindness; and to the technical staff at the Chemistry Department for providing a wonderful service. I would also like to thank Toni Fleming, Gareth Roberts, Nicole Kaftzik, Derek Macmillan and Kevin Bailey for their helpful discussions and advices with microbiology, protein expression and purification, chemistry and modelling. I would like to thank everyone from Ingenza for their kindness.

I would like to thank Vicente Gotor-Fernandez for his fierceness in playing pool, and "Marivi", Maria Victoria Navarro for her kindness and her passion for movies.

I would also like to thank Fabien, Sandrine, Ludovic, Nathalie and David for their support and the long discussions on the phone.

Finally, I would like to thank my Mum, Dad and Brother for their continuing enthusiasm for life and support, and believing in me through all my decisions.

"The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed."

Albert Einstein

Abbreviations

Ac	acetyl group
Amp	ampicillin
A. orientalis	*Amycolatopsis orientalis*
Å	angstrom
bp	base pair
Cml	chloramphenicol
$CaCl_2$	calcium chloride
cat	chloramphenicol acetyl transferase
$CoCl_2$	cobalt chloride
$CuSO_4$	copper sulfate
DAP	α,ε-diaminopimelic acid
dGTP	deoxyribose guanosine triphosphate
DKR	Dynamic Kinetic Resolution
DMSO	dimethyl sulfoxide
DNA	deoxyribose nucleic acid
dNTP	deoxyribose nucleoside triphosphate
E	enantiomeric ratio
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetra acetic acid
FLP	flippase
FRT	FLP recognition target
e.e.	enantiomeric excess
Et	ethyl group
g/L	gram per liter
HF	high fidelity
HPLC	high performance liquid chromatography
IPTG	isopropylthio-β-D-galactoside
Kan	kanamycin
kb	kilobase

kDa	kiloDalton
$MgCl_2$	magnesium chloride
$MnCl_2$	manganese chloride
$MgSO_4$	magnesium sulfate
$MnSO_4$	manganese sulfate
MeOH	methanol
Met	methionine
Me	methyl group
min	minute
MLE	muconate lactonising enzyme
MR	mandelate racemase
NAAAR	*N*- acylamino acid racemase
$(NH_4)_2SO_4$	ammonium sulfate
$NaCNBH_3$	sodium cyanoborohydride
$NaBH_4$	sodium borohydride
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
OSBS	*o*-succinyl benzoate synthase
PEG	polyethyleneglycol
PCR	polymerase chain reaction
Phe	phenylalanine
Ph	phenyl group
PhCHO	benzaldehyde
rpm	round per minute
SDS-PAGE	sodium dodecylsulfate-polyacrylamide gel electrophoresis
Tris	tris(hydroxymethyl)aminomethane
UV	ultra-violet

Contents

1. Introduction

Biocatalysis has been a very important and successful method for the industrial production of amino acids. In 1999 the total market for proteinogenic amino acids was about 1.5 million tons, with estimated revenue of around €3.5 billion. In respect of the market volume the most important amino acid is L-glutamate (650000 tons per year), followed by methionine (450000 tons per year), L-phenylalanine and L-aspartate (12000 tons per year each). In general, amino acids have this considerable economic importance because of their broad spectrum of industrial applications[1]. Biosynthetic routes have been important for the production of enantiomerically pure amino acids, which are valuable compounds for the agricultural, pharmaceutical[2] and food industries. For example, L-amino acids derived from large scale fermentation are widely used as food additives *e.g.* L-glutamate in the form of monosodium glutamate as a flavour enhancer, L-phenylalanine[3] in the sweetener Aspartame™, or L-lysine and L-threonine as animal feed supplements. Biotransformation using immobilised enzymes has also been successfully applied to produce amino acids such as L-aspartate and L-methionine[4-6]. Increasingly, biocatalysis is also sought to prepare unnatural amino acids such as L-*tert*-leucine[7] and *S*-phenyl-cysteine[8] for pharmaceutical use. In these cases the enantiomerically pure amino acid is produced by biotransformations such as resolution and in some cases asymmetric synthesis. D-amino acids prepared by biocatalytic methods[9] such as hydantoinase[10], are also used in semi-synthetic β–lactam antibiotics and as chiral synthons for organic synthesis, or as intermediates for cosmetics, pesticides and pharmaceuticals. The use of biocatalysis is gaining in popularity due to increased concerns regarding the impact of chemistry on the natural environment, which affects the choice for new synthetic processes. Thus, the development of new routes to chiral molecules, which are more environmentally friendly, has become an important factor for public awareness as well as for the industrial demand[11]. In addition, biocatalytic processes are often more economically and ecologically attractive, because less waste are produced which needs to be disposed of or recycled.

1.1. Routes to enantiomerically pure amino acids

Production of amino acids can be achieved using various general routes. Amino acids are isolated as components of natural protein-containing material in the extractive processes, while alternative chemical or biological methods allow the targeted synthesis of amino acids. Due to the increased and broader demand by industry, researchers have tried to apply biocatalysis more generally, to capitalise upon the fact that many enzymes possess enantiospecificity unmatched elsewhere in the synthetic chemist's repertoire. However it has been difficult to find a general and economic method, as each process has particular limitations.

1.1.1. Extraction

L-cysteine is currently the main amino acid produced using extractive processes. Keratin from hair, feathers and animal bristles is hydrolysed by treatment with strong acids. The dimeric oxidation product cystine can be filtered from the hydrolysate because of its poor solubility. Electrochemical reduction of the disulfide bridge in cystine yields cysteine[12]. Although direct access to various amino acids can be achieved by extractive methods, the formation of salts as by-products can lead to considerable environmental impact. Recent approaches, using extremophilic biocatalysts to overcome this disadvantage have already led to promising results[13, 14].

1.1.2. Fermentation

Traditional fermentation is particularly efficient at large scale especially (>1,000 tons/annum), but very specific and generally limited to natural L-amino acids. Fermentation has relied on the ability of certain microorganisms to tolerate and/or to excrete certain amino acids at high concentrations. Typically, L- and D-amino acids are toxic to cells at the high concentrations required for economical production at large scale and mechanisms of resistance are poorly characterised. Additionally,

overproduction of a particular amino acid generally requires multiple mutations related to a specific biochemical pathway which are not applicable to the production of other amino acids. Accordingly each system requires extensive individual research and development, and the method therefore is generally limited to a few proteinogenic L-amino acids[15] (glutamic acid, lysine, phenylalanine and threonine). Thus, for the production of lysine, an early fermentation process made use of a pair of *E. coli* mutants (Figure 1). Wild type *E. coli* can synthesise its own lysine from carbohydrates and ammonia, but the first mutant lacked the α,ε-diaminopimelic acid decarboxylase that normally converts α,ε-diaminopimelic acid (DAP) to lysine.

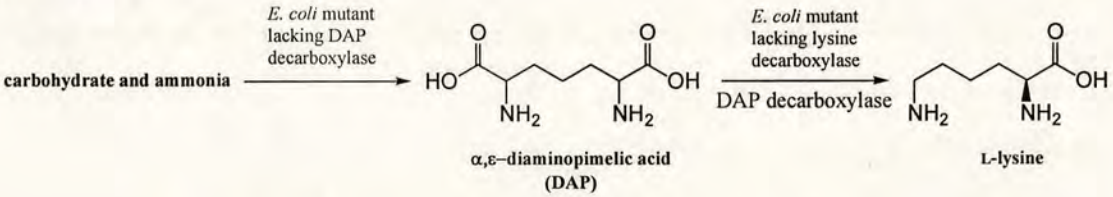


Figure 1: Production of lysine using a pair of *E. coli* mutants

After the concentration of DAP had reached a maximum in the presence of the first mutant, the first strain was removed and another *E. coli* strain was added. This mutant produced DAP decarboxylase, but lacked lysine decarboxylase, allowing lysine to accumulate.

A second method to produce lysine use a single-stage fermentation process is now generally used for the microbial synthesis of L-lysine. A mutant of *Corynebacterium glucamicum* is used during this process. Molasses is the most common carbon source, and biotin is used to suppress the excretion of glutamic acid. The final concentration of L-lysine is reported to be nearly 60 g/L, and the fermentation cycle takes between 48 and 72 hours. The yield of L-lysine from carbohydrate is nearly 40%, according to the stoichiometry (Figure 2).

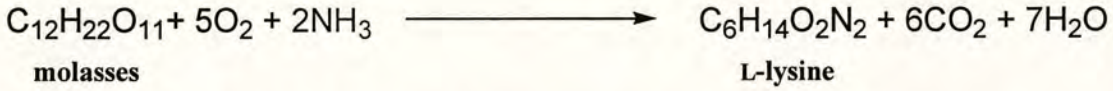


Figure 2: Production of lysine via a single-stage fermentation process

Nevertheless, thanks to a detailed understanding of amino acid biosynthesis, fermentation methods have benefited from both recombinant and classical methods of strain improvement, including genetically modified strains producing threonine[16, 17] or cysteine[18] and methods of *in vivo* selection to improve phenylalanine biosynthesis by removing phenylalanine-mediated feedback inhibition[19]. Though amino acid production by fermentation is limited in scope, these general methods and molecular tools can now be more widely applied, as discussed below.

1.1.3. Resolution

Resolution methods have been more broadly applied than fermentation approaches for amino acid production. More than 60% of all enzyme reactions conducted by industry are enantioselective or regioselective resolution processes, which utilise hydrolase enzymes such as acylases esterases, amidases or lipases[20, 21]. A particular example of esterase activity concerns α-Chymotrypsin. It is a serine protease that catalyses the hydrolysis of amide bonds of proteins at aromatic amino acids such as phenylalanine, tyrosine and tryptophan. It can also catalyse the hydrolysis of ester bonds and has been use synthetically in the enantioselective and regioselective transformation of a variety of amino acids. The enzyme is highly selective for L-amino acids and this has been exploited in the enantioselective hydrolysis of various protected racemic amino acid esters such as *N*-acetyl-DL-tryptophan methyl ester, *N*-acetyl-DL-tyrosine ethyl ester and *N*-acetyl-DL-phenylalanine methyl ester[22]. In each case, hydrolysis of the L-enantiomer occurred giving rise to the L-amino acids and the unreacted D-enantiomers, all with high optical purity.

These processes are advantageous in some aspects: they are often broadly applicable in industry due to the relatively relaxed substrate specificity of the enzymes used[20] and the enzymes are often stable and appropriate for scale-up. In the case of amino acid manufacture, specific examples of useful hydrolytic enzymes termed acylases have become known, the enzyme function is well understood and the enzymes

themselves are increasingly available at low cost. The example shown in Figure 3 is the acylase process to produce L-methionine, which is operated at large scale by Degussa.

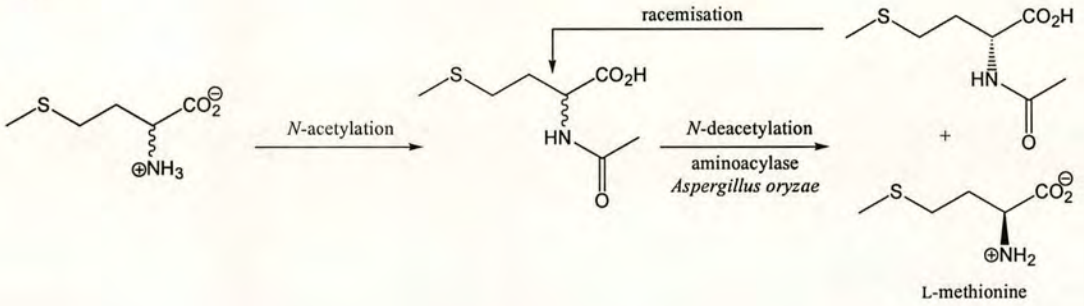


Figure 3: Resolution of *N*-acetyl methionine using aminoacylase

The racemic *N*-acylamino acid substrates used are inexpensively prepared (*N*-acylamino acids) and recovery of the products is straightforward. Nevertheless, in the L-methionine process and other resolutions employing acylase, the yield is limited to a maximum of 50% due to the remaining unreacted isomer. To exceed this yield, complex and expensive recycle processes are required, generally involving chemical steps that often use solvents, which introduce ecological problems. If the process economics cannot tolerate the recycle, the unreacted isomer is discarded, creating a further problem in waste disposal. Accordingly, people have investigated methods to increase the yield of resolution reactions, such as selective crystallisation or substrate racemisation. This has also prompted interest in the development of asymmetric approaches to the synthesis of amino acids as single enantiomers using chemistry or biocatalysis.

1.1.4. Asymmetric methods

In addition to the resolution of racemic mixtures as mentioned above, there are three main approaches to preparing optically active amino acids: chemical synthesis using compounds from the chiral pool, asymmetric synthesis, resolution of a racemic mixture as mentioned above and biotechnological methods.

1.1.4.1. Chemical methods

Although chemical asymmetric syntheses of amino acids are known and frequently applied at small scale[23, 24], there are few examples of the large scale application of this approach. The Strecker synthesis or the Bucherer method[25], a modification of the Strecker method in which ammonium carbonate takes the place of ammonia (Figure 4), are generally used in the industrial manufacture of racemic amino acids, or alternatively using cobalt or palladium as a transition metal catalyst in amidocarbonylation[26].

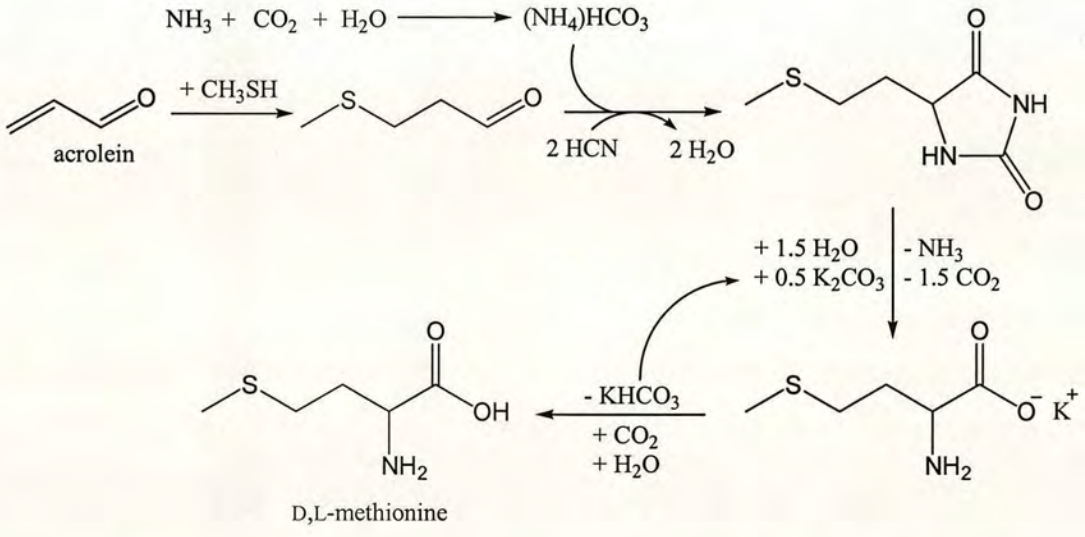


Figure 4: Methionine synthesis by the Degussa process

Even though there are few commercially viable chemical processes for the production of enantiomerically pure amino acids, racemates can still be converted readily into enantiomerically pure compounds by a number of biocatalytic methods. In the case of methionine, its principal use does not require enantiomerically pure methionine, as the nonnatural D-methionine can be metabolised by adults and animals into the biologically active L-enantiomer through a transamination reaction, and therefore could be synthesised using procedures that do not provide enantiomerically pure products.

Interesting and potentially economical routes to amino acids using asymmetric synthesis have been described. These typically employ transition metal catalysts in hydrogenation reactions. Asymmetric hydrogenation has its origins in the soluble Wilkinson catalyst modified with chiral phosphine ligands. Thus, in the early 1970s, this led to the commercialisation of levodopa by Monsanto (Figure 5).

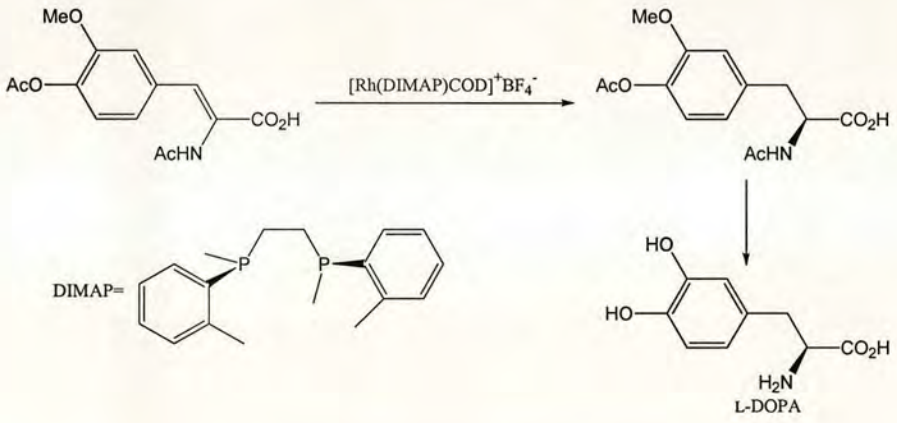


Figure 5: Monsanto process for L-DOPA

This was a landmark in industrial asymmetric synthesis[27, 28], which opened the way for other commercialised catalytic asymmetric syntheses promoted by a chiral transition metal complex, such as the Enichem synthesis of *S*-phenylalanine, using a bis-aminophosphine PNNP[29, 30]. Some more classical chemical routes have been studied for a long time and the list is not exhaustive[31, 32]. Still, original approaches were recently reported using asymmetric Strecker reactions[33], Mannich-type reactions (Lewis acid mediated addition)[34, 35], organolithium

reagents[36] or organoboranes[37, 38]. Although in some cases successful, the drawback to this approach remains that the catalysts are often very selective for substrates and therefore the method cannot be applied broadly. In addition the preparation of the catalyst is often expensive and there can be problems in recovery[39] or disposal of the heavy metals used. As a result these methods can often not be scaled up economically. More recent works[40-46] studied the combination of metal catalysts and enzymes for the synthesis of optically active compounds.

1.1.4.2. Biological asymmetric methods

Whereas living microorganisms are used as biocatalysts in the production of amino acids by fermentation, the catalysts in biotransformations have various forms, such as isolated enzymes or metabolically inactive cells. As a result, transformation of compounds that do not occur in natural metabolic pathways may be possible. Thus, these processes are the methods of choice for the production of D-amino acids and other nonnatural amino acids such as L-*tert*-leucine. Biocatalytic routes to the synthesis of amino acids as single enantiomers have been extensively studied. Although these methods are free of some of the problems faced by metal catalysts, such as the cost of expensive ligands and waste disposal, they generally suffer the same problem of limited breadth of application. This is a major problem for applied biocatalysis where the development time is often longer than a chemical route. If the application is narrow then the risk is greater in process development at large scale. For example, the reductive amination of a keto acid to the corresponding α-amino acid using a dehydrogenase biocatalyst[47, 48] is reported to be very efficient in specific cases, such as the manufacture of L-*tert*-leucine[49, 50] (Figure 6).

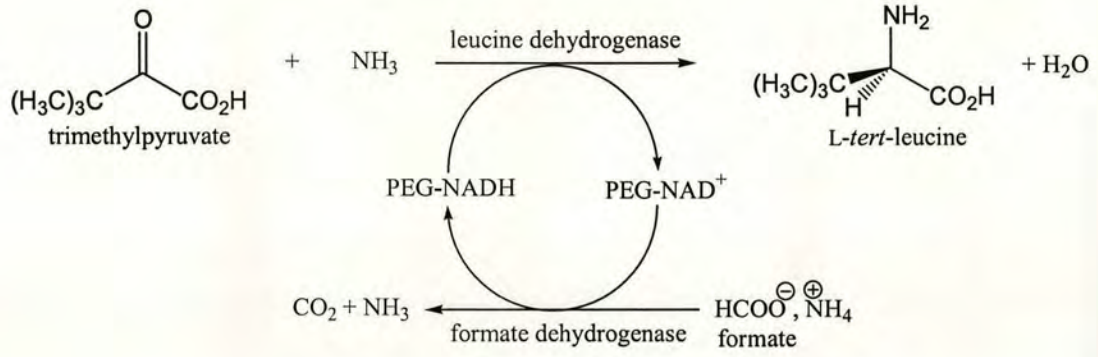


Figure 6: Enzymatic synthesis of L-*tert*-leucine

For soluble reactants and products, enzymes are preferentially immobilized in an enzyme-membrane reactor. To prevent the cofactor from escaping through the membrane, it can be enlarged with polyethyleneglycol (PEG)[51]. With a suitable process of cofactor regeneration, enzymatic cofactor-dependent reductive amination of trimethylpyruvate is the route of choice for large-scale synthesis of enantiomerically pure L-*tert*-leucine. Degussa operates an industrial process producing high-quality L-*tert*-leucine via this route. However this biocatalyst is dependent upon NADH for activity and requires a second biocatalyst and complex engineering to retain the expensive NADH[52]. As a result its application seems limited to particular, high value amino acids, as it has not been reported in any other application.

Similarly, aminotransferases[53] have been successfully scaled for the synthesis of particular amino acids, such as L-2-aminobutyric acid. This type of biocatalyst is attractive since it has a broad specificity but is compromised because it generates a keto acid by-product that has to be removed. The occurrence of this by-product suggests that the recovery of the desired amino acid product may also be complex and case dependent (Figure 7).

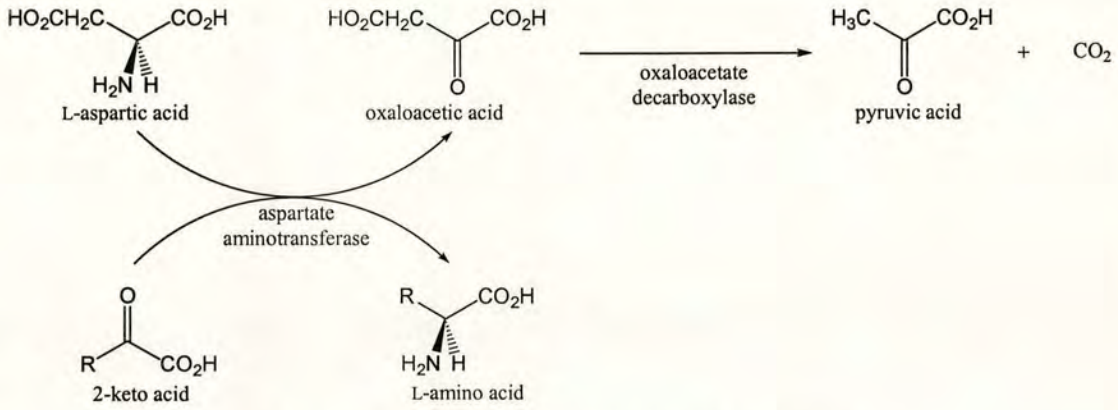


Figure 7: Synthesis of chiral amino acids with aminotrasferases

A more general solution, which would take advantage of the broad understanding and applicability of hydrolytic enzymes in kinetic resolution, would result from greater availability of dynamic kinetic resolution.

1.1.4.3. Deracemisation and dynamic kinetic resolution

Deracemisation is the conversion of a racemic mixture into a single enantiomer through a stereoinversion reaction, giving theoretical yields and enantiomeric excess (e.e.) of 100%. However, a more recent definition was proposed in 2001[54]: Deracemisation constitutes any process during which a racemate is converted into a non-racemic product in 100% theoretical yield without intermediate separation of materials. According to this definition, dynamic kinetic resolution, dynamic kinetic asymmetric transformation, dynamic thermodynamic resolution, stereoinversion, and enantioconvergent transformation of a racemate would all constitute subtypes of deracemisation. On the other hand, desymmetrisation of prochiral and *meso*-compounds and kinetic resolution of racemates (including subtypes thereof, such as sequential or parallel kinetic resolution and enantiodivergent transformations) are not considered to be deracemisation.

Dynamic kinetic resolution is a major goal[55, 56] in applied biocatalysis, as classical kinetic resolutions are limited to a maximum of 50% due to the remaining

unreacted isomer. The dynamic kinetic resolution process (Figure 8) involves a classical kinetic resolution coupled with *in situ* substrate racemisation[57].

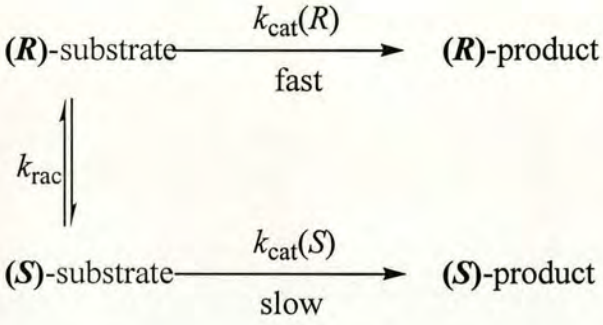


Figure 8: Dynamic kinetic resolution

Substrate racemisation in theory enables complete conversion of the starting material to the desired product, resulting in theoretical yield and e.e. of 100%. But for the dynamic kinetic resolution to be effective, the rate of racemisation (k_{rac}) must be at least as great as the rate of product formation ($k_{cat}(R)$), to reach the theoretical yield of 100%. For the maximum efficiency, product formation should be irreversible, the enantiomeric ratio (E) should be ≥ 20 (where $E = k_{cat}(R)/ k_{cat}(S)$) and k_{rac} should be approximately ten times faster than $k_{cat}(R)$[58].

Since the harsh conditions of chemical racemisation are generally incompatible *in situ* with enzymatic hydrolysis, an appealing approach is the establishment of an entirely enzymatic process in which substrate specific racemisation is achieved by means of a second enzyme which operates under conditions compatible with an enantioselective hydrolase such as the acylase biocatalyst mentioned above. An example of one such process, which has been developed commercially for amino acid production, involves hydantoinase reactions[59]: This process is efficient in certain cases, such as D-phenylglycine manufacture, although it often requires three biocatalysts acting in concert (hydantoinase, carbamoylase, racemase) (Figure 9).

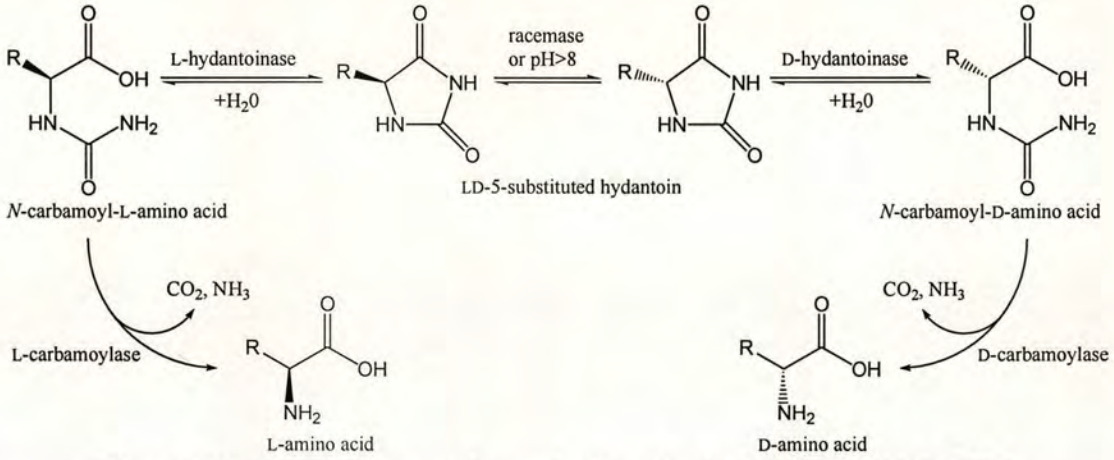


Figure 9: Hydantoinase process for production of D- (or L-) amino acids by DKR

The most versatile biocatalysts for this process currently derive from *Aspergillus*, but are not readily expressed with high efficiency and in a co-ordinated manner in a more readily manipulated and scalable system *e.g. E. coli*. This system is also currently limited to D-amino acids[60] although recent discoveries[61] and efforts[62, 63] to expand the utility of hydantoinase to L-amino acids have effectively applied the process of directed evolution[64, 65].

Despite its limitations, the hydantoinase process remains a potentially excellent and high yielding route and a good example of the potential utility of racemases in industry.

Analogous to the hydantoinase system, the prospect of developing a substrate specific racemase enzyme to use in conjunction with the well-established acylase process is therefore extremely interesting. In the case of acylase hydrolysis, the substrate racemase would require being highly active and specific for *N*-acylated compounds such as *N*-acylamino acids but to display little or no activity upon the free amino acids themselves. Ideally the enzyme should also possess a broad substrate tolerance within the family of *N*-acylamino acids (Figure 10). The development of general methods to obtain and improve such racemase enzymes is the focus of this thesis.

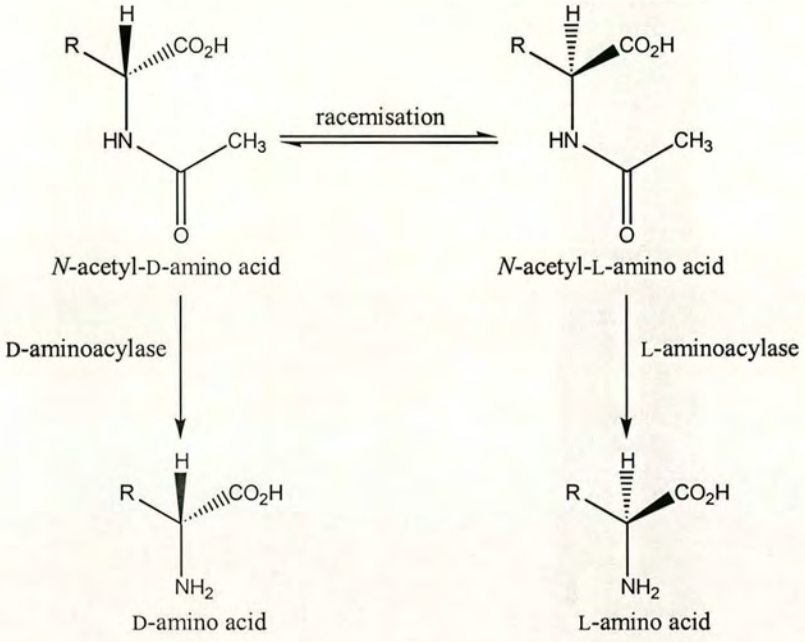


Figure 10: Reaction scheme of the acylase process

The fact that an enzyme from *Amycolatopsis species* TS-1-60, has been shown to possess a minor level of this activity[66-70] is therefore very encouraging to the overall objective. Subsequently, other enzymes with similar activity have been described in *Streptomyces atratus*[71], *Amycolatopsis azurea*[72] and *Amycolatopsis orientalis*[73].

1.2. *N*-acylamino acid racemase (NAAAR)

1.2.1. *Amycolatopsis species* TS-1-60 NAAAR

During the 1980's, work was carried out to identify a biocatalyst with *N*-acylamino acid racemase (NAAAR) activity. A screen[68-70] specifically for *N*-acetyl methionine racemase activity, employing 49,000 different strains, identified a racemase in an *Amycolatopsis* species that displays relaxed substrate specificity towards *N*-acylated amino acids but no activity towards the free amino acid equivalents (might be explained by substrate preference or the chemistry of α-proton acidity[74, 75]).

Some development work was carried out upon this enzyme to explore its potential application in the production of enantiomerically pure amino acids. Used in conjunction with an enantioselective amidohydrolase or acylase, this enzyme offers the possibility to establish a dynamic kinetic resolution process, producing a single amino acid isomer from racemates of particular acylated amino acid substrates. The enzyme shows activity upon a broad range of *N*-acylated amino acid substrates, particularly methionine and phenylalanine. The findings that *N*-acetyl-L-methionine displayed the highest activity, and that substrates with hydrophobic side-chains were preferred (methionine, phenylalanine, glutamine, tryptophan, leucine), suggested that the side-chain-binding region has a preference for binding hydrophobic moieties.

Upon further examination by a different research group[76] the enzyme was characterised, not principally as a racemase but as a dehydratase, *o*-succinyl benzoate synthase (OSBS), a member of the muconate lactonising enzyme (MLE) subgroup of the enolase superfamily of enzymes. Before describing this superfamily, it is important to define what is meant by the term superfamily. One classical definition simply describes enzymes families as homologous enzymes of >50% sequence pairwise identity, known to catalyse the same reaction on structurally similar substrates; and superfamilies as evolutionarily related proteins of <50% sequence identity[77]. But homologous proteins in superfamilies may also exhibit percent

sequence identity well below 50% and have diverged to the point that they mediate very different overall functions. A more recent definition is that the members of mechanistically diverse superfamilies catalyse different overall reactions that are related by a common partial reaction in which a chemically similar intermediate is generated and significantly stabilised by conserved functional groups and elements of secondary and tertiary structure. In the case of the enolase superfamily, various subgroups of enzymes carry out distinct reactions which all involve the initial abstraction of a proton.

The low level of racemase activity of the *Amycolatopsis species* TS-1-60 OSBS was confirmed and appears to be a by-product of the reaction mechanism when supplied with *N*-acylamino acid substrates. There is no evidence as to any selective pressure or advantage to the organism to have evolved this racemase activity.

The NAAAR is not a very efficient catalyst for the racemisation of its ''preferred'' substrate, *N*-acetyl methionine; as shown by the values of k_{cat}, K_M, and k_{cat}/K_M which are respectively 6.4 s^{-1}, 11 mM, and 590 M^{-1} s^{-1}[76]. The reason for the low activity is that NAAAR is not the natural reaction catalysed by this protein; instead, the protein is an OSBS as judged by the value of its ''superior'' kinetic constants for this reaction, 120 s^{-1}, 0.48 mM, and $2.5x10^5$ M^{-1} s^{-1}, respectively, for k_{cat}, K_M, and k_{cat}/K_M. Thus, the protein originally characterised for its activity toward racemisation of *N*-acetyl methionine has been found to show a *ca.* 1000-fold greater activity toward dehydration of 6-hydroxy-2-succinyl-2,4-cyclohexadiene carboxylate (Figure 11), which was concluded to be the physiological substrate for this enzyme.

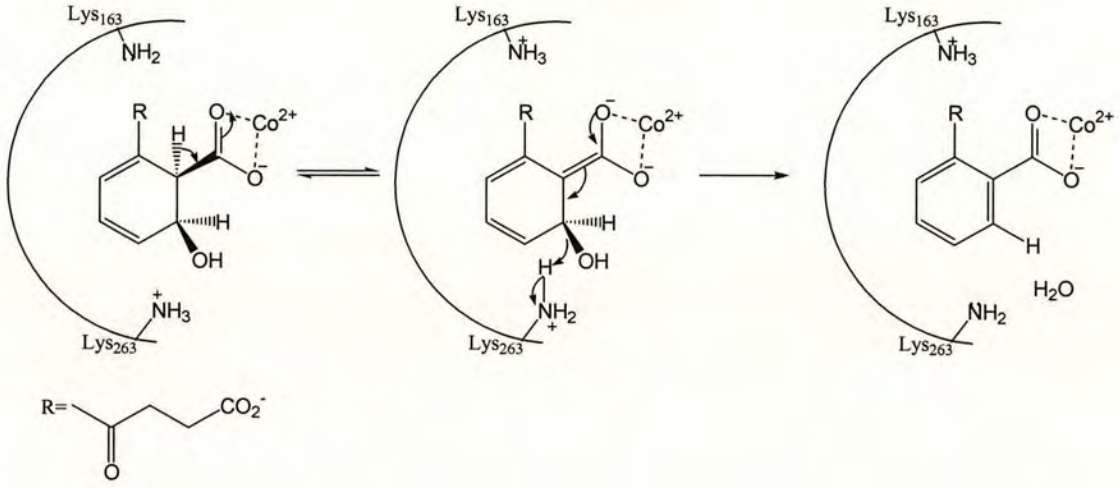


Figure 11: The elimination reaction of a substituted cyclohexadiene catalysed by the promiscuous OSBS/NAAAR

Accordingly, the NAAAR reaction catalyzed by the OSBS from *Amycolatopsis species* TS-1-60 is an example of functional promiscuity, arising because the hydrophobic substrate for the NAAAR reaction can bind in the same cavity as the substrate for the OSBS reaction and be properly positioned between Lys163 and Lys263, located at the ends the second and sixth β-strands, respectively, so that the 1,1-proton transfer reaction can occur (Figure 12[78]).

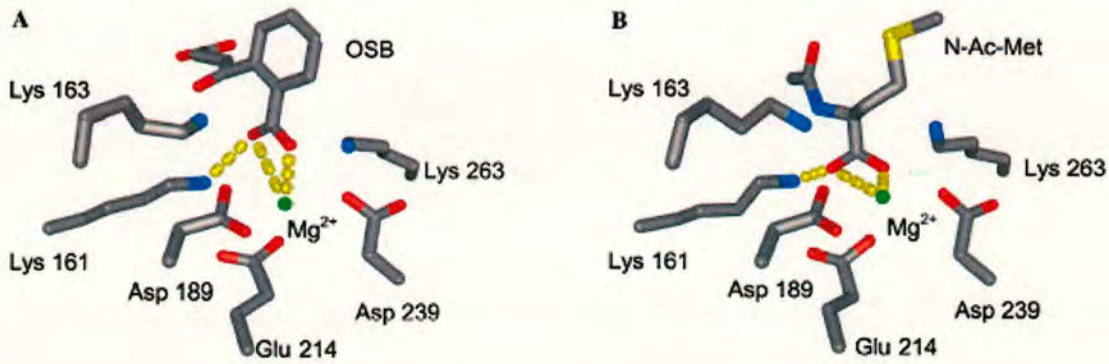


Figure 12: The active site of the promiscuous OSBS/NAAAR from *Amycolatopsis species* TS-1-60. Position of the substrate for (A): the OSBS reaction; (B): the NAAAR reaction

The mechanisms of both the natural OSBS reaction and the promiscuous NAAAR reaction are initiated by abstraction of the α-proton of the substrate to generate an

enolate anion intermediate stabilized by coordination to the essential Mg^{2+} (Figure 13).

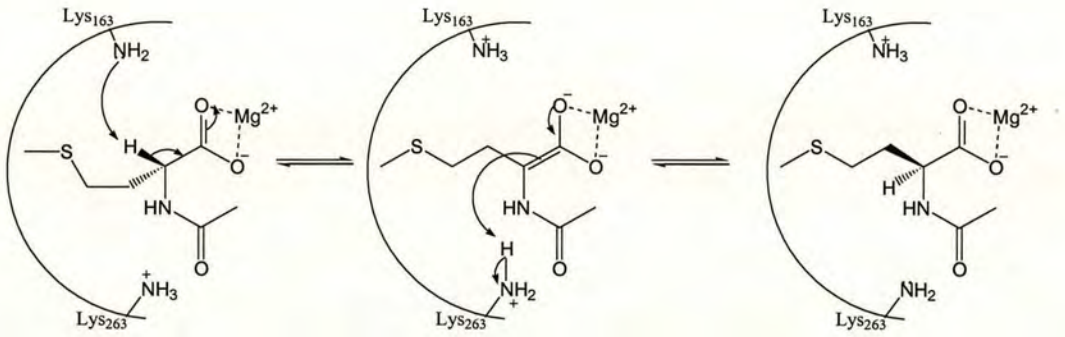


Figure 13: The mechanism of NAAAR reaction

The structure and mechanistic details of NAAAR are now well documented[75, 79], and in the case of the NAAAR expressed by *Deinococcus radiodurans*, the framework K170-D195-E220-D245-K269 was determined to catalyse 1,1-proton exchange of *N*-acylamino acids. The enzyme is constituted by four subunits enclosing the substrate (Figure 14[80]): catalytic site (C), metal-binding site (M), side-chain-binding site (S) and a flexible lid region (L).

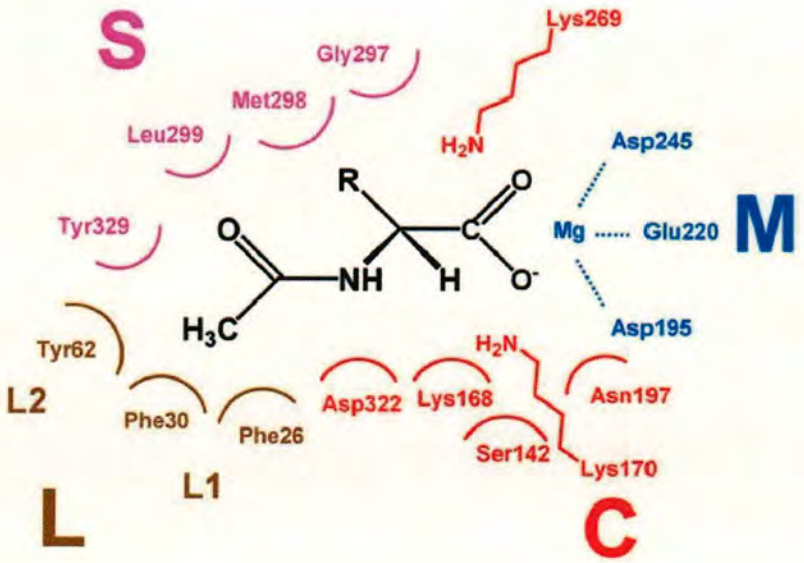


Figure 14: Schematic representation of the 4 subunits of the active site of the NAAAR from *D. radiodurans* enclosing the substrate

Although examples of the potential use of NAAAR in an applied bioprocess have been described at low substrate concentration (<50 mM) the enzyme does not appear to have been developed for commercial use. This may be partly due to the highly competitive arena for low cost amino acid production, *e.g.* fermentative production, standard resolution approaches and alternative resolution methods employing enzymes such as hydantoinase. More likely is the fact that, the enzyme has properties, which are significantly limiting for a useful biocatalyst, such as low turnover, and substrate mediated inhibition at concentrations exceeding 50 mM.

Nevertheless, there are some aspects of this enzymatic activity, which render it extremely attractive as an industrial biocatalyst. Primarily, it has very broad specificity for amino acid side chains, it is produced from a single gene (although, it is a monomer in *E. coli*, but a homo-octamer in *Amycolatopsis*), it requires a divalent metal ion for activity but no complex prosthetic group as cofactor and therefore no need for retention or recycling of a cofactor. In addition, the reaction substrates can be prepared using inexpensive and general chemistry. These properties warrant additional investigation of this enzymatic activity at both academic and applied levels, such as the potential to identify further examples of this activity or to improve the *Amycolatopsis species* TS-1-60 racemase activity and the potential to produce NAAAR in recombinant cells with acylase or amidase.

Since this activity seems to have arisen by chance, as there is no known requirement for this activity in microbial cells, then it seems reasonable to consider that perhaps additional examples of NAAAR could be found if a powerful selection or screen for this activity was devised. Additionally, the application of a powerful high throughput screen or selection could enable the artificial evolution or enhancement of this activity using methods of directed evolution[81], which are now becoming widespread and increasingly powerful, due to new methods and tools *e.g.* error prone polymerase chain reaction (PCR) and mutator strains and the increasing availability of data derived from microbial bioinformatics and genomics.

1.2.2. Enolase superfamily

The enolase superfamily[82-84] is one of the best-described enzyme superfamilies to date. This group of enzymes is related in three main ways. The first way is the common aspects of enzyme mechanism: Its members can be characterised by mediating a common fundamental chemical step, the stabilisation of enolate anions generated from abstraction of a proton α to a carboxylate (Figure 15).

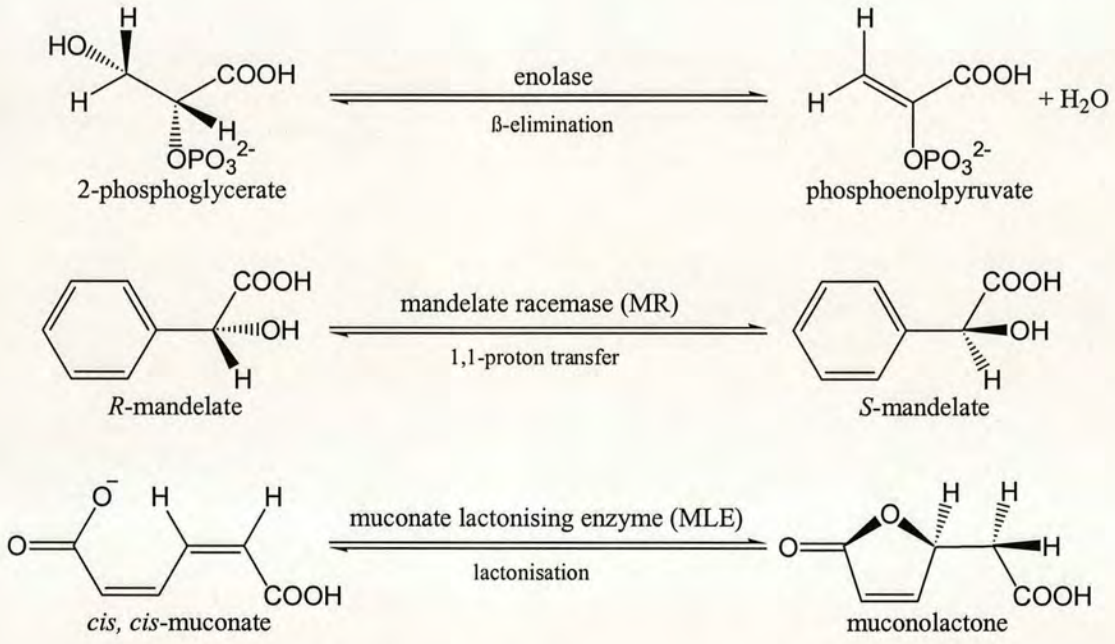


Figure 15: Characteristic reaction of the 3 subgroups of the enolase superfamily

The second way is the sequence identity/homology of its members. Indeed, even if the sequences of these enzymes are divergent, some closely conserved motifs in these sequences at specific positions are remarkable: 7 positions (Lys 164, Lys 166, Asp 195, Glu 221, Glu or Asp 247, His 297 or Lys 273, Glu 317) seem to have a very important role in the activity of this superfamily. By comparing the binding pockets among members of the enolase superfamily, it appeared differences in the side-chain-binding site and flexible lid region may contribute to their distinct substrate specificities, whereas the catalytic site and the metal-binding site are highly

conserved, suggesting they are essential for their common strategy in the formation of an enolic intermediate.

The last way is related to the 3-D structural conservation (domains) (Figure 16). The activity of each enzyme is dependent upon a divalent metal ion (Mg^{2+} or Mn^{2+}). This feature provides a critical link for understanding the interdependence of their structures and functions.

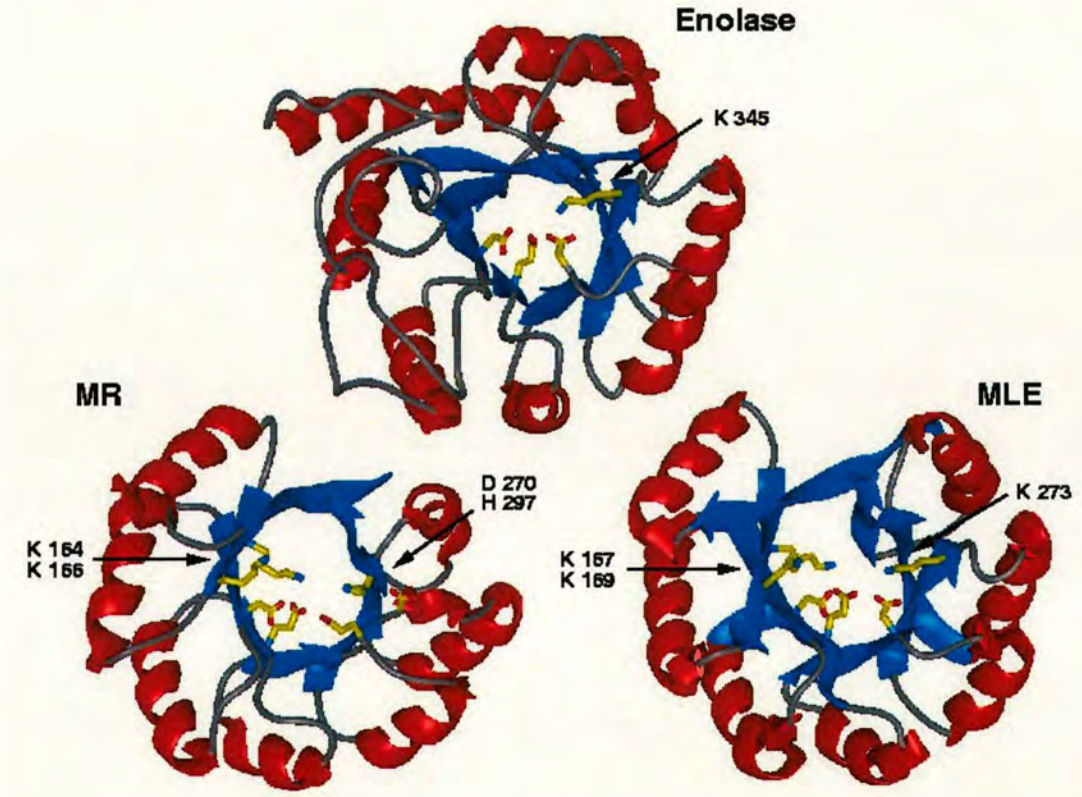


Figure 16: The highly conserved active site structure in the 3 subgroups of the enolase superfamily

The enolase superfamily is constituted by more than 100 sequences to date, closely related (homology for a few of them is shown Table 1 - page 21), which can be classified amongst these three subgroups:

- Enolase related enzymes including enolases, CPEPS, cMycBP,
- Mandelate racemase (MR),
- Muconate lactonizing enzyme group (MLE): including OSBS.

Since NAAAR is a member of this superfamily but dehydratase activity rather than racemase activity is the principal native activity of the enzyme, this suggests that given members of this superfamily may possess characteristics of the reactions carried out by different enolase superfamily. This raises the possibility that members may mutate under appropriate evolutionary conditions to acquire these different activities. A recent study which introduced single substitutions within members of the MLE group[85], one of the 3 subgroups of the enolase superfamily, and the group to which the OSBS/NAAAR belongs, appeared to support the idea that interconversion of enolase superfamily activities is possible. This indicates the functional promiscuity of these enzymes.

This raises several questions: Could directed evolution *in vitro* followed by *in vivo* selection demonstrate fluidity amongst these enzymes *i.e.* introducing novel activity to members of this family? Could selection for NAAAR activity reveal new enzymes appropriate for industry? Could novel isolates of NAAAR activity be more appropriate for industrial application than the NAAAR that is known? Could this expand the role of dynamic kinetic resolution? Could the application of such enzymes be extended from amino acids to the much broader class of chiral amines? To address these questions, a highly specific *in vivo* selection for *N*-acylamino acid racemase activity was devised and developed in this work.

Table 1: Sequences producing significant alignments with *Amycolatopsis species* TS-1-60 NAAAR gene

Homology: sequences producing significant alignments with *A. sp.* TS-1-60 NAAAR gene (Source:www.npsa-pbil.ibcp.fr using BLAST search)

STRAINS	TYPE OF ENZYMES	E-VALUE*
Amycolatopsis sp. TS-1-60	NAAAR	Use as reference
Amycolatopsis orientalis subsp. Lurida	NAAAR	0
Deinococcus radiodurans	NAAAR	3.00E-82
Aeropyrum pernix	398AA LONG HYPOTHETICAL NAAAR	4.00E-82
Deinococcus radiodurans	NAAAR	6.00E-82
Thermoplasma acidophilum	PROBABLE NAAAR	6.00E-61
Thermoplasma volcanium	NAAAR	1.00E-55
Rhodococcus opacus	CHLOROMUCONATE CYCLOISOMERASE	2.00E-30
Alcaligenes eutrophus (Ralstonia eutropha)	PUTATIVE MUCONATE CYCLOISOMERASE (EC 5.5.1.1)	1.00E-28
Pseudomonas aeruginosa	MUCONATE CYCLOISOMERASE I	1.00E-23
Alcaligenes eutrophus (Ralstonia eutropha)	CHLOROMUCONATE CYCLOISOMERASE (EC 5.5.1.7) (MUCONATE CYCLOISOMERASE II)	1.00E-23
Pseudomonas putida	PUTATIVE MUCONATE CYCLOISOMERASE I	6.00E-23
Burkholderia sp. NK8	CHLOROMUCONATE CYCLOISOMERASE (EC 5.5.1.7)	1.00E-22
Acinetobacter calcoaceticus	MUCONATE CYCLOISOMERASE I (EC 5.5.1.1) (*CIS ,CIS*- MLE I)	2.00E-22
Rhodococcus opacus	MUCONATE CYCLOISOMERASE I (EC 5.5.1.1) (*CIS, CIS*-MLE I)	2.00E-22
Alcaligenes eutrophus (Ralstonia eutropha)	CHLOROMUCONATE CYCLOISOMERASE (EC 5.5.1.7)	5.00E-22
Pseudomonas putida	MLE	1.00E-21
Pseudomonas aeruginosa	CHLOROMUCONATE CYCLOISOMERASE	1.00E-21
Pseudomonas putida	CHLOROMUCONATE CYCLOISOMERASE (EC 5.5.1.7) (MUCONATE CYCLOISOMERASE II)	2.00E-21
Pseudomonas aeruginosa	CHLOROMUCONATE CYCLOISOMERASE	2.00E-21
Pseudomonas sp. (strain P51)	CHLOROMUCONATE CYCLOISOMERASE (EC 5.5.1.7) (MUCONATE CYCLOISOMERASE II)	4.00E-21
Frateuria sp. ANA-18	MUCONATE CYCLOISOMERASE (EC 5.5.1.1)	6.00E-21
Pseudomonas putida	MUCONATE CYCLOISOMERASE I (EC 5.5.1.1) (*CIS, CIS*-MLE I)	6.00E-21
Alcaligenes eutrophus (Ralstonia eutropha)	CHLOROMUCONATE CYCLOISOMERASE	7.00E-21
Variovorax paradoxus	CHLOROMUCONATE CYCLOISOMERASE	3.00E-20
Acinetobacter lwoffii	MUCONATE CYCLOISOMERASE I 1 (EC 5.5.1.1) (*CIS, CIS*-MLE I 1) (MLE1)	4.00E-20
Acinetobacter lwoffii	MUCONATE CYCLOISOMERASE I 2 (EC 5.5.1.1) (*CIS, CIS*-MLE I 2) (MLE2)	5.00E-20
Streptomyces coelicolor	PUTATIVE ISOMERASE	8.00E-20
Rhizobium loti (Mesorhizobium loti)	MLE	2.00E-18
Deinococcus radiodurans	CHLOROMUCONATE CYCLOISOMERASE, PUTATIVE	2.00E-17
Halobacterium sp. (strain NRC-1)	CHLOROMUCONATE CYCLOISOMERASE	1.00E-15
Pseudomonas putida	MR (EC 5.1.2.2) (MR)	3.00E-14
Bacillus halodurans	MUCONATE CYCLOISOMERASE	5.00E-14
Comamonas acidovorans (Pseudomonas acidovorans)	CHLOROMUCONATE CYCLOISOMERASE (FRAGMENT)	8.00E-14
Rhodoferax sp. P230	CHLOROMUCONATE CYCLOISOMERASE (FRAGMENT)	1.00E-13
Streptomyces coelicolor	PUTATIVE RACEMASE	6.00E-12
Rhizobium loti (Mesorhizobium loti)	MR	6.00E-12
Sulfolobus solfataricus	MR /MLE RELATED PROTEIN	6.00E-12
Streptomyces coelicolor	PUTATIVE ISOMERASE	6.00E-12
Staphylococcus aureus subsp. aureus Mu50	OSBS	6.00E-12
Escherichia coli O157:H7	PUTATIVE MUCONATE CYCLOISOMERASE I	8.00E-12
Staphylococcus aureus	OSBS	1.00E-11

*The E-value is related to homology; the closer to zero, the more important is the homology between the gene from TS-1-60 and the others genes.

1.3. Approach of this project

The work described in this thesis describes the conception and testing of such a method to explore the potential to discover and to evolve more active and industrially suitable variants of this type of racemase biocatalyst. This project is valuable for different reasons: Firstly, it will help in the understanding of the fluidity/interchangeablility of genes and enzymes in the enolase superfamily. Secondly it could give rise to a valuable new family of industrial biocatalysts. Thirdly, it could illustrate the continuing value of *in vivo* selective methods to augment *in vitro* methods of directed evolution and provide new methods for biocatalyst discovery and development. It offers an *in vivo* method that benefits from earlier understanding of the value of this type of approach in fermentation routes to amino acids but takes advantage of tools recently made available.

1.3.1. Directed evolution

By observing evolution in nature, scientists have learned how to apply and benefit from natural processes and have discovered various types of enzymes that have been used, as they existed in the nature. However, the problem with the use of native enzymes is that nature did not design enzymes to operate under the arduous and demanding conditions required by industry. Now new methods and new data are changing the way that we discover enzymes. Nowadays, screening methods are much improved and can benefit from increased use of robotics, miniaturisation, high throughput analysis and better analytical methods such as mass spectra and NMR. Accordingly, the analysis of enzymatic activities and reactions are easier and more efficient. Powerful tools in chemistry (combinatorial chemistry) and in biology (such as PCR, directed evolution and strains construction) as well as bioinformatics tools[86, 87] (genomic databases) have opened a very large field of investigation in research.

The improvement of molecular biology tools as PCR, mutator strains, biological mutation and then, in the 1990s, error prone PCR[88-90] and Stemmer gene

shuffling[91] have increased the number of opportunities and rate of development of new enzymes. Thanks to genomics, it may be possible to explore a tiny fraction of the possibilities that nature, even in 4 billion years, has not.

Bioinformatics permit quicker and easier use of databases than a few years ago such that it is now possible to identify homology between genes (proteins), which could potentially be interesting to screen and thereby to identify novel enzymes for particular applications. Thus, it is becoming possible to explore the ability to convert one enzyme activity to another perhaps within a superfamily of enzymes related by mechanism and structure to further advance new enzyme discovery. Often the key is to devise a powerful screening method or a powerful *in vivo* selective approach. By choosing an appropriate selective method, it could become possible to screen, identify, and even evolve novel enzymes. This work will address this issue with particular emphasis on as the discovery of novel *N*-acylamino acid racemases.

Enzymes can be tailored for optimal performance in industrial applications by evolutionary molecular engineering[92], also called directed evolution or *in vitro* molecular evolution (Figure 17).

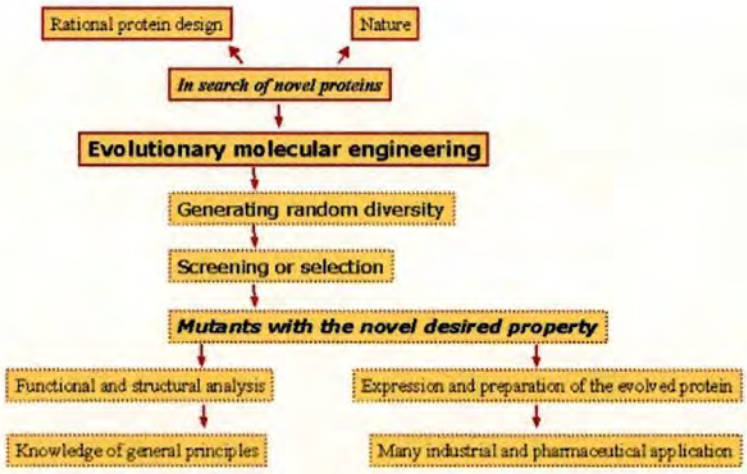


Figure 17: The main steps of the discovery and evolution of novel proteins

The evolutionary strategy is a simple adaptive process: the accumulation over multiple generations of beneficial mutations produced randomly and fixed by

selective pressures. The major steps in a typical directed enzyme evolution experiment are outlined in Figure 18.

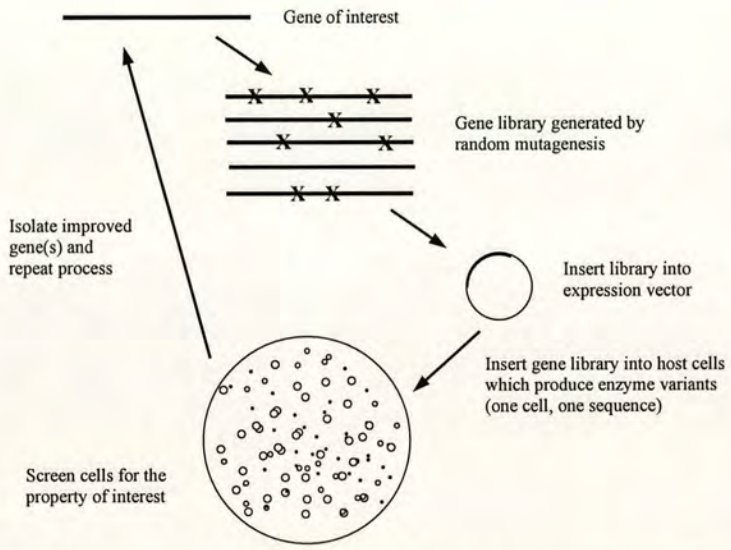


Figure 18: Experimental strategy for directed evolution using error prone PCR

The genetic diversity for evolution is created by mutagenesis and/or recombination of one or more parent sequences. These altered genes are cloned back into a plasmid for expression in a suitable host organism (bacteria or yeast). Clones expressing improved enzymes are identified in a high-throughput screen, or in some cases, by selection, and the genes encoding those improved enzymes are isolated and recycled to the next round of directed evolution. The random mutant library is typically generated using a method of random DNA mutation such as a chemical mutagen, a DNA repair deficient "mutator" organism or error prone PCR in which the point mutation frequency is controlled by varying the pH, the concentration of divalent cations, or other parameters.

As shown in Figure 19, the number of variants that could be made if every possible amino acid substitution were accessible (for example, by codon-synthesis methods) increases exponentially along the simultaneous substitution of amino acids. If point mutations are introduced randomly at the DNA level, the accessible sequence

diversity is limited. On average, only 5.7 different amino acids are accessible by single base pair changes in a codon.

The number of possible variants of a protein that can be created by introducing M substitutions simultaneously over N amino acids is $19^M [N!/(N-M)!\,M!]$

$5.7^M [N!/(N-M)!\,M!]$

	The number of possible variants
Sequence length = 320, Number of amino acids changed = 1	6 080 1 824
Sequence length = 320, Number of amino acids changed = 2	18 425 440 1 658 290
Sequence length = 320, Number of amino acids changed = 3	37 108 836 160 1 001 938 576
Sequence length = 320, Number of amino acids changed = 4	55 876 630 047 920 452 600 703 388

Figure 19: Variation of the possible variants of a protein in function of the number of amino acids changed

Furthermore, the number of clones[93] that must be screened to see a given number of unique amino acid level variants must increase to account for remaining wild-type protein sequences arising from the degeneracy of the genetic code. To have 95% confidence that a given number of amino acid level variants has been sampled in a screen or selection, the number of unique clones must be multiplied by a factor of approximately ten.

Thus the `sequence space' of possible proteins is huge beyond the imagination (20^N). Even in 4 billion years, nature has had a chance to explore but a tiny fraction of these possibilities. A laboratory exploration of this vast space of sequences and their corresponding functions must obviously be severely limited and carefully guided[94]. Because much of sequence space will be devoid of the desired function and probably even folded proteins, it is best to direct the evolution of one (or more) existing enzymes rather than look for function in random peptide libraries.

For studying wild-type protein function, the Vartanian group[95] suggested creating a single amino acid substitution per protein, which corresponds to approximately 1.5

mutations per gene. This level of mutagenesis allows to independently characterise the effects of each amino acid substitution on protein function[85, 96]. For directed protein evolution, mutagenesis rates that average 2 to 6 mutations per gene are regarded as most effective for creating mutant libraries to find proteins with enhanced activity[97, 98] or interconvert functionalities of homologous enzymes[85, 99-101]. Mutational levels beyond 6 mutations per gene usually result in the complete loss of protein activity[102]. However, there have been exceptional cases where proteins tolerate extremely high levels of mutation[95, 103].

1.3.1.1. Random mutagenesis using error prone PCR

PCR-based random mutagenesis has gained popularity over chemical methods (*e.g.*, nitrous acid, hydroxylamine, etc.) since it can produce a higher level and a larger variety of mutation[104, 105]. The procedure involves performing a PCR reaction under conditions that reduce the fidelity of nucleotide incorporation, cloning the resulting PCR fragments, and then screening the resulting library for novel mutations which affect protein activity[106, 107]. To do so, the standard procedure is altered by increasing the concentration of $MgCl_2$, adding $MnCl_2$, or changing the concentration of the 4 dNTPs[88, 90]. Using the right conditions it is possible to introduce mutations at a rate up to 7 base pairs per 1000 during a 30 cycles PCR. Kits for PCR random mutagenesis are commercially available, and optimised to offer a variety of buffer conditions for performing random mutagenesis, allowing to adjust the reaction for a desired error rate. For example, with the kit available from Clontech, the mutagenesis rate is first raised by increasing the amount of manganese in the reaction. Further increases in mutation rate are obtained by increasing the level of dGTP in the reaction, while keeping the concentration of manganese constant. Mutation rates from two to eight mutations per 1000bp can therefore be achieved.
Nevertheless, the method requires numerous manipulations (PCR reaction, digestion, ligation) and introduces some difficulties: ligation of the PCR products is critical and the mutational bias should be optimised.

1.3.1.2. Random mutagenesis using mutator strain Epicurian Coli ® XL1-Red

Stratagene constructed the XL1-Red mutator strain to generate random mutations quickly and easily. This strain[108] is deficient in three of the primary DNA repair pathways in *E. coli*: *mutS*[109], *mutD*[110] and *mutT*[111]. The *mutS* mutation causes error-prone mismatch repair, while the *mutD* mutation causes the deficiency in 3'- to 5'- exonuclease of DNA polymerase III (subunit of DNA polymerase III involved in proofreading and self-editing function, which normally removes incorrectly paired nucleotides during DNA replication). The *mutT* mutation disables the hydrolysis of 8-oxodGTP generated by highly reactive oxygen species (superoxide, hydrogen peroxide, hydroxyl radical) during normal aerobic metabolism. It can then be misincorporated during replication to opposite template adenines. The random mutation rate is approximately 5,000-fold higher than that of wild-type (mutation rate compares well to error prone PCR). The XL1-Red procedure is a straightforward method: it requires simply using the construct to transform the XL1-Red strain and propagating the transformants. After propagation in the XL1-Red strain, the mutated plasmid DNA is retrieved and used to retransform the provided XL1-Blue competent cells for analysis. Nevertheless it has some disadvantages: the rapid mutation rate affects the chromosome[112], and after prolonged growth, the subsequent colonies are probably not genetically identical to the original strain, leading the host strain to be unstable, and some mutations can cause plasmid loss that may eliminate useful enzyme variants.

1.3.2. Selective method: a new approach to the high throughput concept?

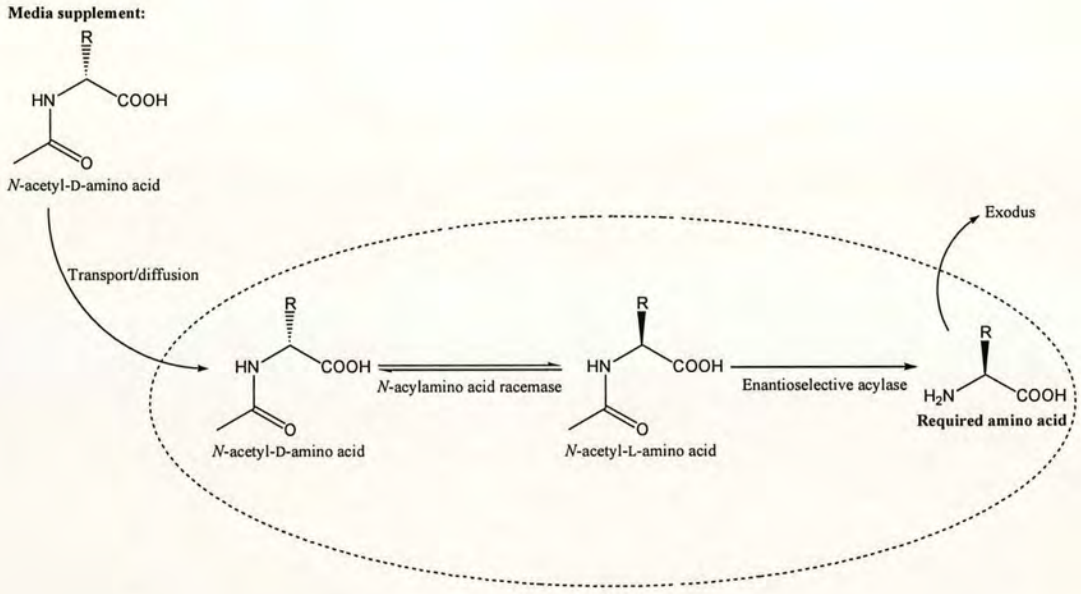


Figure 20: Proposed *in vivo* selection

The selection for NAAAR activity (Figure 20) applies specific amino acid auxotrophs so that an *N*-acetyl-D-amino acid can be supplied as the only available precursor of a deficient L-amino acid. If the cell needs a specific L-amino acid for growth and is capable of hydrolysing the corresponding *N*-acetyl-L-amino acid then maybe it can be supplemented with the corresponding *N*-acetyl-D-amino acid to select for racemisation to the corresponding *N*-acetyl-L-amino acid.

This method, if successful, could offer the possibility to readily screen a library of 10^6-10^8 mutants for *N*-acylamino acid racemase activity in 3-4 days. Using auxotrophs of *E. coli* available from public strain repositories (*e.g.* the *E. coli* Genetic Stock Center), specific racemase activities can be sought. In this case, L-phenylalanine and L-methionine auxotrophs were chosen as the test systems. The main reasons for the choice of these 2 systems reside in the fact that *N*-acetyl methionine and *N*-acetyl phenylalanine are known to be favoured substrates for the

known NAAAR from strain TS-1-60. These two L-amino acids are also relevant industrially, and are easy to access and to prepare as *N*-acetyl derivatives.

The first stage in this work is to confirm the phenotype of *E. coli* amino acid auxotrophic mutants. This is very important as it establishes the supplemental growth medium requirements of an L-amino acid auxotroph (for example *pheA* mutant requiring L-phenylalanine). This test should show the inability of the auxotroph to grow on the D-amino acid. The next step is to test for hydrolysis of an *N*-acyl-L-amino acid precursor when provided in the growth medium as the only supply of the deficient amino acid. This is to determine any necessity to supply acylase enzyme exogenously because no certainty exists concerning the ability of the strain to hydrolyse the *N*-acyl-L-amino acids. As the *N*-acyl-L-amino acid hydrolysis must not be rate limiting, one way is to introduce a high level of L-aminoacylase known in the literature[113]. In principle, this should render the cells able to efficiently use *N*-acyl-L-amino acids as a source of required free amino acids (Figure 21).

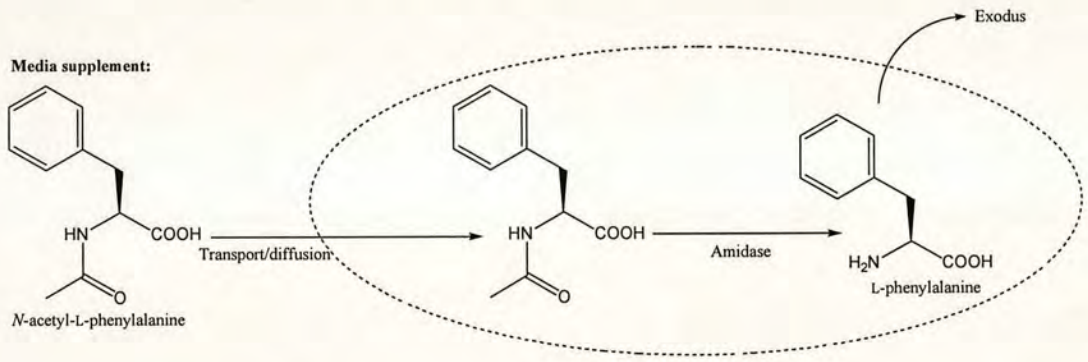


Figure 21: Test on *pheA* mutant for *N*-acyl-L-phenylalanine hydrolysis

The selection is then carried out by plating the cells on a M9/glucose minimal medium, which is supplemented with only the *N*-acetyl-D-amino acid. The theoretical basis of the selection is therefore that bacteria carrying and expressing an *N*-acetylamino acid racemase gene with the desired specificity are able to grow whereas all other bacteria plated cannot grow (Figure 22).

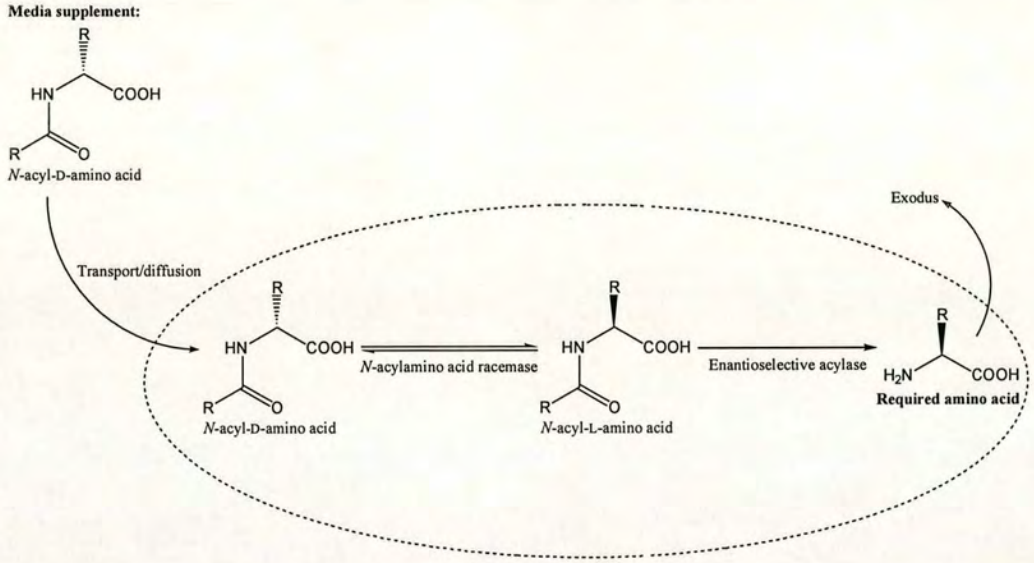


Figure 22: Theoretical screen for *N*-acylamino acid racemase

The plasmid carrying any gene whose product enables growth of the auxotrophic host is then isolated from these bacteria; and tested for NAAAR activity. Such isolates upon characterisation may then serve as templates for further mutation and selection in a process of directed evolution of NAAAR activity (Figure 23).

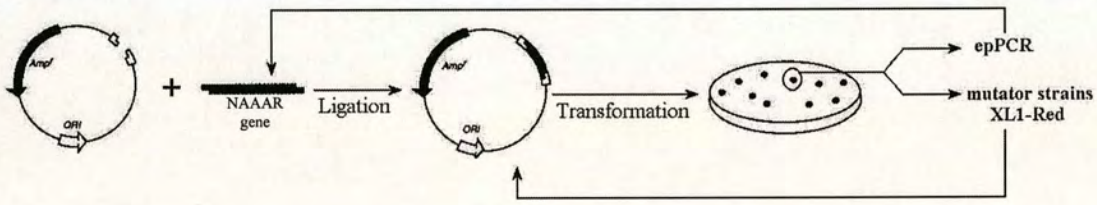


Figure 23: Methods for directed evolution of NAAAR activity

1.4. Conclusions and aims

The potential to identify novel enzymatic methods to conduct dynamic kinetic resolution processes is an exciting field. The NAAAR activity, which was shown to be a secondary function of an enzyme which is primarily a dehydratase of the enolase enzyme superfamily, is an appropriate target for studies of directed evolution. A significant number of members of this enzyme superfamily have been identified and the gene sequences published. The potential exists to identify additional NAAAR activities through sequence alignment and screening using the data in microbial gene and genome banks. The fact that one member of this family displays racemase activity indicates that others may also do so or that the activity could be selected. Because this enzyme activity produces amino acids as the product, it may be possible to devise efficient and versatile selective methods using microbial mutants deficient in specific amino acids (auxotrophs) in order to select for enzymes with *N*-acylamino acid racemase activity and to evolve altered substrate specificity or other properties in the known *Amycolatopsis species* TS-1-60 enzyme and any others identified.

Using similar approaches the potential may exist to subsequently select evolution of *N*-acylamino acid racemase activity in mutant libraries derived from this enzyme and homologues. Although chemically a more challenging reaction, this approach would also provide a graphic illustration of the fluidity of enzyme function and would have major implications for applied biocatalysis. Finally, although the earlier studies showed that the NAAAR activity of the strain TS-1-60 is not ideal for biocatalysis due to slow turnover and substrate inhibition, in common with other racemases it has fundamental features appropriate for applied biocatalysis.

In this project, *in vivo* selections will be devised and constructed which will hopefully identify variants of enolase superfamily enzymes that show novel NAAAR activity. These selection methods will be based upon the construction of mutants of *E. coli*, which eliminate an essential enzyme, thereby preventing growth on an

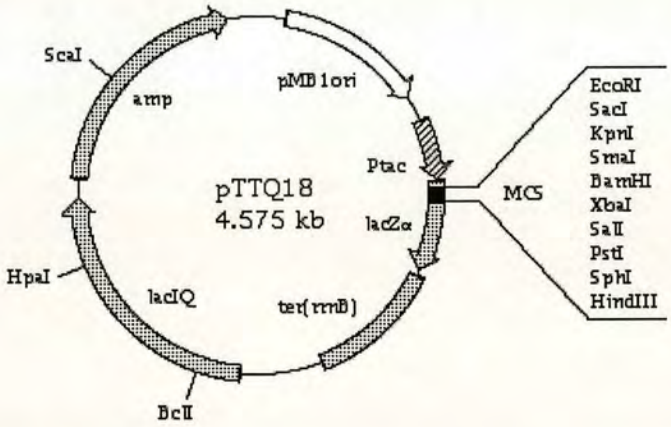
appropriate selective medium. Complementation of this deficiency will form the basis of the identification of novel activity in variants generated by random mutagenesis of enolase superfamily members. Genes encoding enzymes of the enolase superfamily will be identified, isolated and subjected to novel selective methods to determine if new activity can be introduced.

2. Results and discussion

Since NAAAR activity appears to have arisen by chance the basis of this work is to determine if this activity can be identified and/or enhanced using directed evolution. Such an approach raises the possibility that enolase superfamily members may be mutated to acquire these different activities.

2.1. Cloning various OSBS genes into pTTQ18 system

The plasmid pTTQ18[114] is a high-copy number plasmid containing ~100 copies per cell. Gene expression is driven using the tac promoter derived from the *E. coli* lac operon and controlled by induction using IPTG. The replication of the plasmid is controlled by ColE1 (pBR322) as origin of replication. The plasmid can be selected in growing cells by addition of ampicillin to the culture, as it carries an ampicillin resistance (provided by its β–lactamase (*bla*) gene).



2.1.1. Amplification of NAAAR gene from TS-1-60

The NAAAR gene was amplified from *Amycolatopsis* TS-1-60 (IFO 15079) using primers containing EcoRI and SphI restriction sites. The PCR was carried out using the entire chromosomal DNA of TS-1-60, purified using Qiagen Genomic DNA Kit and stored at 4°C, to avoid subsequent difficulties concerning precipitation due to

freezing. This gene was particularly difficult to clone due to a high level of G and C (70%). Many PCR experiments were carried out, varying the following specific parameters:

- Use of 2 different polymerases (*Taq* DNA polymerase and High Fidelity polymerase,
- Temperature of annealing and denaturation,
- Use of DMSO (2%, 5% and 10%),
- Primers and extended primers,
- Gradient of temperature,

The final conditions that reproducibly yielded a PCR product are listed in the Materials section.

The PCR product was digested with EcoRI/SphI restriction endonucleases and then ligated into EcoRI/SphI digested pTTQ18. The final construct was then used to transform Top10 *E. coli* cells. The Top10 *E. coli* strain is a very versatile host[115], improving the quality of plasmid DNA by elimination of non-specific endonuclease by endonuclease I (*end*A1) and reducing the occurrence of recombination of the gene of interest (*rec*A1). Also, unmethylated DNA propagated from non-*E. coli* sources such as PCR is transformed with higher efficiency (*hsd*RMS[116]). The transformation efficiency of this strain is 10^9 cfu/μg. Positive transformants were identified by restriction analysis and tested for expression in JM109. This plasmid was named pRS7-1 and sequenced to verify no changes were introduced from the original sequence.

2.1.2. Amplification of NAAAR gene from *Amycolatopsis orientalis* NRRL 2452

The NAAAR gene was amplified from *Amycolatopsis orientalis* (NRRL 2452) with primers containing EcoRI and SphI restriction sites. The PCR was carried out using the isolated chromosomal DNA of *A. orientalis*, purified using Qiagen Genomic DNA Kit and stored at 4°C, to avoid subsequent difficulties concerning precipitation due to freezing. The amplification was not as difficult as for the NAAAR gene from TS-1-60, due to the similarities between the sequences of the 2 genes. The same

conditions were used for the PCR reaction and successfully yielded the expected single band. The PCR product was digested with EcoRI/SphI restriction endonucleases and then ligated to EcoRI/SphI digested pTTQ18. The final construct was then used to transform *E. coli* Top 10. Positive transformants were identified by restriction analysis and used for expression in JM109. This plasmid was named pRS8-9 and sequenced to verify no changes were introduced from the original sequence.

2.1.3. Amplification of other OSBS genes

In addition to the cloning of two NAAAR genes mentioned above, two other OSBS genes, one from *Bacillus subtilis* and the other from *E. coli*, were cloned into pTTQ18, and respectively named pRS6-2, and pRS9-1. Although mapped using restriction enzymes, no further work was carried out upon these genes or their encoded enzymes due to time limitation and inconclusive data arising from growth studies.

2.2. Creation of libraries by random mutagenesis

Two different methods were used to create mutations within the cloned genes and create libraries of mutants which could be used for screening later with the *in vivo* selection. The first method, by error prone PCR, used the kit from Clontech (Diversify™ PCR random mutagenesis), whereas the second method used the mutator strain Epicurian Coli ® XL1-Red.

2.2.1. Library generation using error prone PCR

The two plasmids pRS7-1 and pRS8-9 were subjected to error-prone PCR using the manufacturer's protocol. Each plasmid was used as a DNA template and was subjected to 9 parallel PCR reactions with the same conditions but with different reaction mixtures. The 2 parameters modified in the mixture were the concentrations

of $MnSO_4$ and dGTP. As the concentration of these two reagents was increased, the intensity of the corresponding band decreased (Figure 24). This was due to the fact that amplification (processiveness of the polymerase) becomes more limited as mutations increase. Unfortunately, the error prone PCR failed on pRS8-9 despite several attempts using modified protocols, entire plasmid or fragment as DNA template. No fragment of the desired size was amplified, the electrophoresis gels showing either no band fragment or a smear.

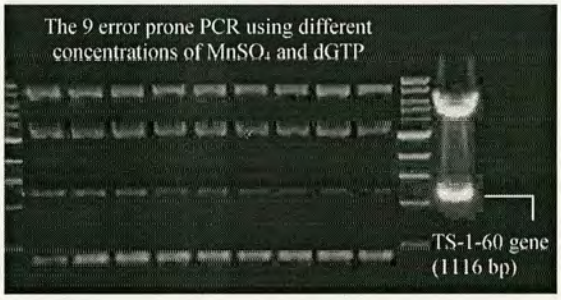


Figure 24: error prone PCR on NAAAR TS-1-60 gene

The nine mutagenesis reactions carried out on TS-1-60 racemase gene were then mixed as followed: Reactions 1 to 3, 4 to 6 and 7 to 9 mixed together to give 3 different samples with an average number of mutations per 1000 bp of 2.3, 4.3 and 7 respectively. These three samples were then digested and cloned into expression vector pTTQ18. These three libraries were purified and named pRS7-1/13, pRS7-1/46 and pRS7-1/79. The detail protocol of this method is described Materials section.

2.2.2. Libraries generated using the mutator strain Epicurian Coli ® XL1-Red

Two methods were used to produce libraries of mutated plasmids using *E. coli* XL1-Red. The initial step was common to both methods: The two plasmids were used to transform the mutator strain using the manufacturer's protocol.

2.2.2.1. First method to create libraries of mutants

The first method consisted of spreading the transformed cells on LB Agar Amp^{200}. The plates were then incubated for 24-30 hours, and then the cells recovered by resuspending the colonies into 10 ml of LB. As *E. coli* XL1-Red has a doubling time of 90-120 mins, approximately 15 generations of cells were produced. This corresponded to one cycle of mutation. The plasmids were purified for screening, while 150 µl of resuspended cells was spread on each of five freshly prepared LB Agar Amp^{200} plates for another cycle of mutation. At the end of this cycle, the plasmids were isolated and used to transform *E. coli* XL1-Red, and subjected to another cycle of mutation. Plasmids were purified and used to retransform every two cycles fresh mutator strain to avoid the rapid accumulation of mutations in the host genomic DNA, which can lead to the loss of genotype specificity and viability. The plasmid was subjected to 5 cycles of mutation to produce the library of mutants after 15, 30, 45, 60 and 75 cells generations (pRS7-1/1P to pRS7-1/5P libraries).

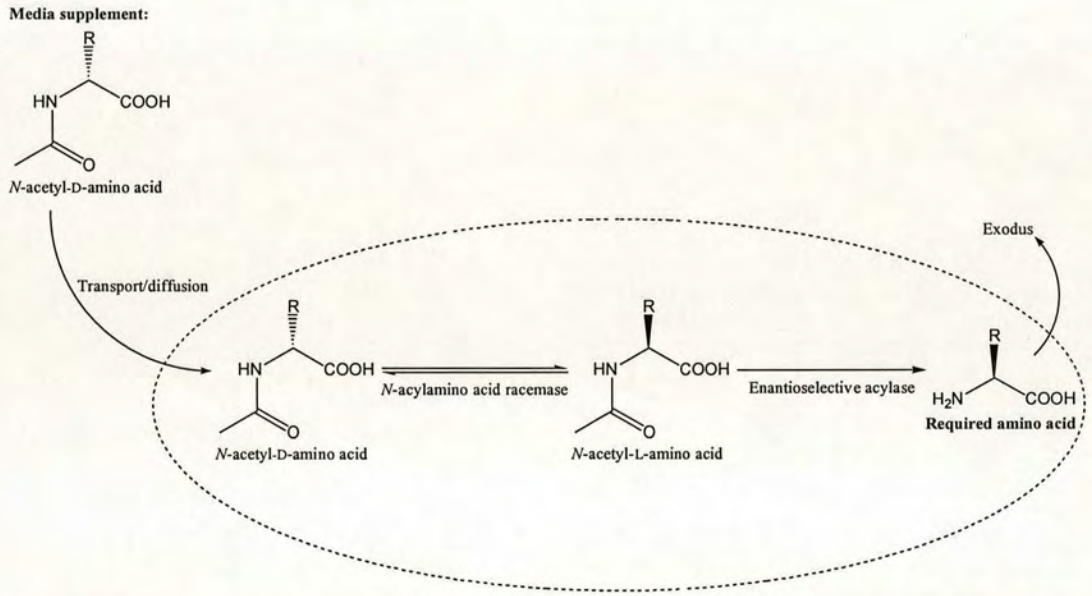
2.2.2.2. Second method to create libraries of mutants

The second method consisted of using the transformed cells to inoculate 20 ml of LB Amp^{200} and the culture was grown for 24-30 hours. This corresponded to one cycle of mutation. The plasmids were purified for screening, while 20 µl of the culture was used to inoculate LB Amp^{200}, and subjected to a second cycle of growth. The plasmids were then purified and used to transform fresh *E. coli* XL1-Red, and subjected to another cycle of mutation. The plasmids were transformed every two cycles of growth. The plasmid was subjected to 5 cycles of mutation to produce the library of mutants after 15, 30, 45, 60 and 75 cells generations. (pRS7-1/1S to pRS7-1/5S libraries).

2.3. *In vivo* selection

The central aim of the project was the development of an *in vivo* method to identify specific enzymes using a specific bacterial auxotrophic mutant. The selection for the enzyme which relieves the auxotrophy should be high throughput, enabling the screening of 10^8-10^9 independent colonies per plate.

Principles of the *in vivo* selection to relieve the auxotrophy:



2.3.1. L-phenylalanine (L-phe) as the first auxotrophy to be examined

Initially the phenylalanine system was studied, mainly because a phenylalanine auxotroph was already available from the *E. coli* Genetic Stock Center (Yale University, New Haven, USA).

2.3.1.1. Feeding studies on the L-phe auxotroph strain CGSC7421

See Appendix 11.1. for genotype of this strain.

2.3.1.1.1. Initial growth tests of the auxotrophic strain on M9/glucose medium

Initial feeding experiments were carried out to confirm the phenotype of the strain and to show that only the free L-amino acid could relieve the L-phenylalanine auxotrophy. The exact composition of the M9/glucose minimal medium is shown in the Materials section. In order for the strain to be useful and appropriate for the selection, no growth should be observed on an M9/glucose based minimal medium alone as well as when supplemented with D-phenylalanine or *N*-acetyl-D-phenylalanine, but observed when supplemented with L-phenylalanine and with *N*-acetyl-L-phenylalanine. In the latter case, this should indicate the ability of the strain to hydrolyse the acetyl group (existence of a natural acylase).

Media supplement	**Expected growth for selection**	**Actual growth**
None	No	No
L-phenylalanine	Yes	Yes
D-phenylalanine	No	Yes
N-Ac-L-phenylalanine	Yes	Yes
N-Ac-D-phenylalanine	No	Yes

Table 2: Feeding tests on the L-phe auxotroph

Experiments shown in Table 2 showed the strain to be able to grow on M9/glucose minimal medium supplemented by either *N*-acetyl-D-Phe or *N*-acetyl-L-Phe as well as D-Phe although no growth was observed on unsupplemented M9 alone.
Following this observation, an extensive literature review[117-119] revealed that the complementation of certain L-amino acid auxotrophies in *E. coli* could be relieved by supplementation of the growth medium with the corresponding D-enantiomer of the amino acid. Genetic and biochemical studies have conclusively shown this to be

attributed to a broad specificity D-amino acid dehydrogenase encoded by the *dadA* gene and induced by many D-amino acids. The action of the dehydrogenase converts the D-amino acid to the corresponding keto acid, which can then be transaminated to the L-isomer by one or more endogenous transaminases. The D-amino acid dehydrogenase is principally required for growth on racemic alanine or L-alanine as a carbon source and is non-essential to the prototrophic cell. Based on this literature review, Figure 25 shows the expected routes by which *E. coli* can use a D-amino acid to overcome an L-amino acid auxotrophy.

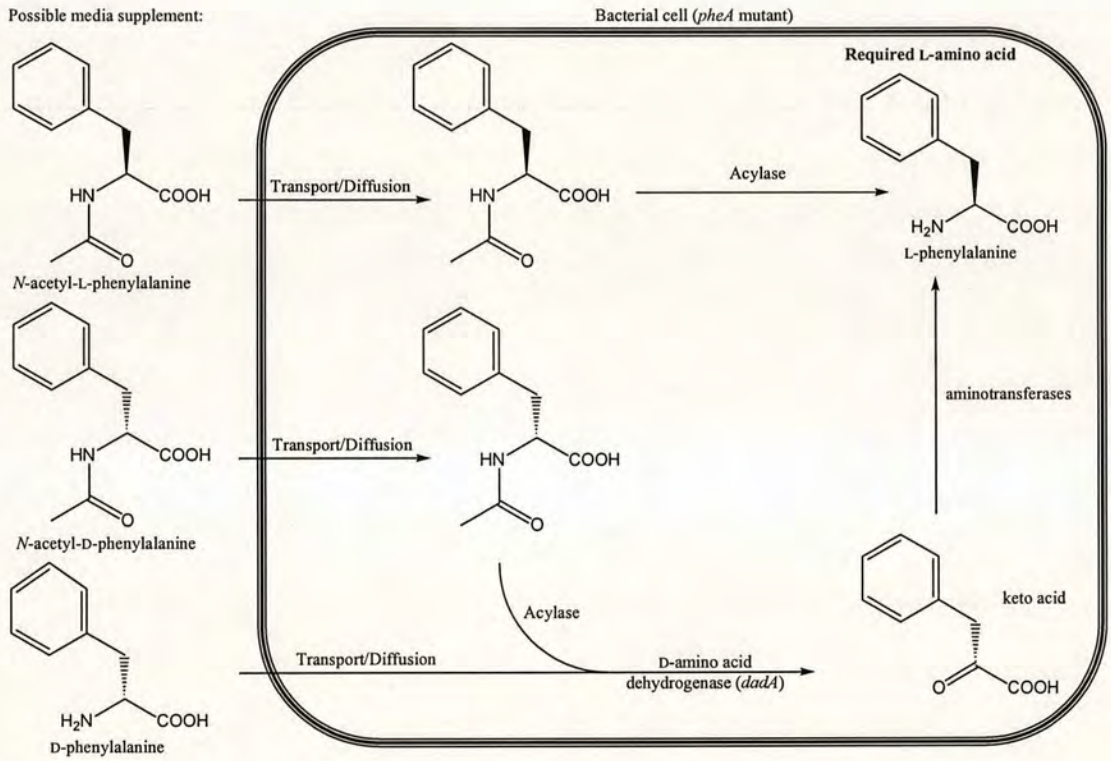


Figure 25: Modes of amino acid utilisation by *E. coli*

Accordingly, the *dadA* gene was required to be deleted in amino acid auxotrophic strains of *E. coli* to determine the effect on further development of the planned selection (as no mutants were already available with the appropriate genetic backgrounds). Deletions are the most appropriate for this type of experiment (where a powerful growth selection is applied) to avoid false positives through reversion of

the mutation. Additionally, the deletion method[120] (rapid and general) could be developed into a useful method in the lab.

The principle of this method is fairly simple (Figure 26): A PCR product is generated using primers with extensions that are homologous to regions adjacent (H1 and H2) to the gene to be inactivated and also homologous to template plasmids carrying antibiotic resistance genes (P1 and P2) that are flanked by FRT (FLP recognition target) sites. The strain to be modified is then transformed with the PCR product in the presence of a helper plasmid, which stabilises the PCR product and facilitates integration and gene replacement at the target. The deletants are selected by acquired antibiotic resistance, and the deletion is verified by PCR. The antibiotic resistance can eventually be removed if desired using an FLP helper plasmid.

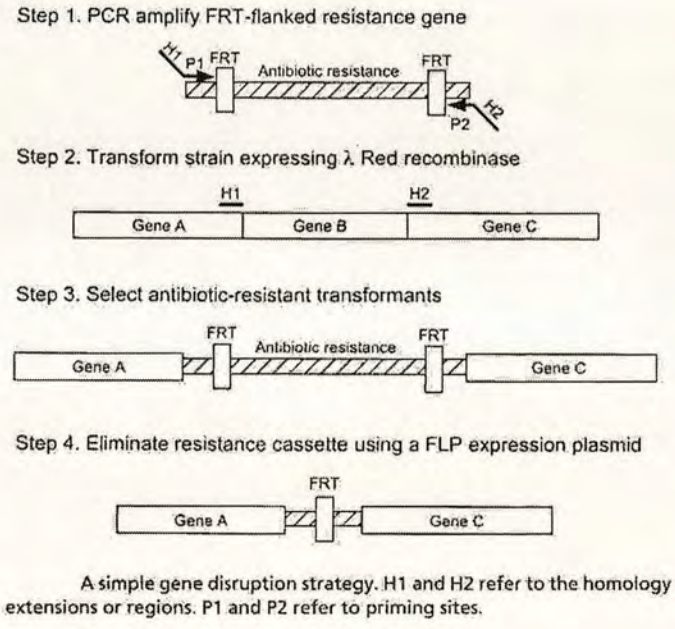


Figure 26: Gene disruption strategy

2.3.1.1.2. Deletion of the *dadAX* genes from CGSC7421 using the λ Red system.

The deletion was first created in strain BW25113 which is supplied with the disruption system. This strain is ΔaraBAD and *hsdR* which according to the authors makes the creation of deletions easier.

In *E. coli*, the *dadX* gene is involved in L-alanine catabolism. This gene encodes an alanine racemase and is situated in an operon together with the *dadA* gene, encoding D-alanine dehydrogenase. The *dadA* gene encodes one of the two subunits of D-amino acid dehydrogenase. This enzyme is bound by strongly hydrophobic interactions to the inner membrane and directly linked to the respiratory chain. The *dadA* and *dadX* genes are transcribed together from a promoter proximal to *dadA*. As the *dadA* and *dadX* genes form one transcriptional unit[121], it seemed logical to delete both genes rather than only *dadA* to avoid possible complications with membranes function and amino acid transport. The *cat* gene (chloramphenicol resistance) was introduced to select the deletant strain. The first step of the inactivation of *dadAX* was the amplification of the *cat* gene (chloramphenicol acetyl transferase-chloramphenicol resistance) by PCR, with homologous priming sequences of the gene to inactivate. The PCR product was then used along with the RED helper plasmid to transform the original strain BW25113. The antibiotic resistance replacing *dadAX* allowed the selection of transformants, in which the *dadA-dadX* operon was disrupted. The new strain was named ET1.

Experiments were then carried out to transfer the *dadAX* deletion to a phenylalanine auxotroph by P1 transduction. The strain ET1 was first infected and lysed with phage P1, and the P1 lysate was then mixed with the strain to be modified. After transduction the new strain was screened on chloramphenicol plates to select the positive mutants. Although 5 mutants were successfully isolated and tested against the deletion, the shape of the colonies were irregular, and the colonies screened on M9 with different supplements were still able to grow on M9 which indicated either one or several genes enabling the biosynthesis of L-Phe from D-Phe and *N*-Ac-D-Phe. After several attempts the conclusion was that no transductant of the correct phenotype could be obtained. At this point a change in the choice of the auxotrophic selection was therefore made. Although there were other options to construct a L-phe auxotroph (*e.g.* direct deletion of *pheA* in ET1), another system using L-methionine auxotrophy was to be considered at this time. Emphasis was switched to the methionine system as the NAAAR is described to be the most active using *N*-acetyl methionine.

2.3.2. L-methionine as the second auxotrophy to be examined

2.3.2.1. Use of CGSC6563

See Appendix 11.1. for genotype of this strain.

The advantage of using L-methionine auxotrophy is that a methionine auxotroph which also carries a *dadA* mutation was already available. But this approach also had some drawbacks. A strain containing point mutation of *metB* has the possibility to revert. This situation is not ideal for a high throughput selection due to the potential for false positives. The strain therefore needs to be frequently restreaked on M9 to rule out this possibility (Figure 27).

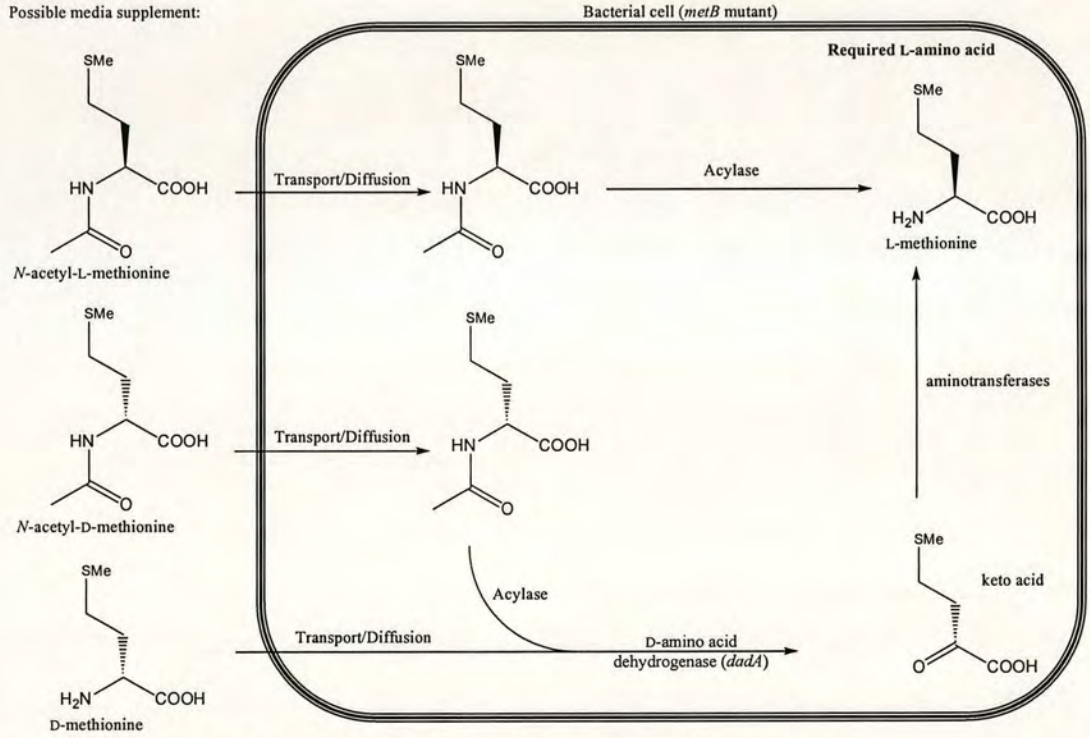


Figure 27: Modes of methionine utilisation by *E. coli*

2.3.2.1.1. Feeding studies on CGSC6563

These feeding experiments were carried out to confirm that only the free L-amino acid could relieve the methionine auxotrophy. In order for the strain to be useful and appropriate for the selection, no growth should be observed on M9 alone as well as when supplemented with D-methionine or *N*-acetyl-D-methionine, but observed when supplemented with L-methionine and eventually with *N*-acetyl-L-methionine. In the latter case, this growth should indicate the ability of the strain to hydrolyse the acetyl group (existence of a natural acylase).

After testing this strain on M9/glucose minimal medium to validate the selection, it appeared the strain was behaving as expected (Table 3). The auxotrophy of this strain was confirmed, and its ability to hydrolyse *N*-acetyl-L-methionine meant that there was no requirement to use a complementary acylase. Due to the *dadA* mutation, the strain was not able to use *N*-acetyl-D-methionine as a source of L-methionine.

Media supplement	Expected growth for selection	Actual growth
None	No	No
L-methionine	Yes	Yes
D-methionine	No	No
N-Ac-L-methionine	Yes	Yes
N-Ac-D-methionine	No	No

Table 3: Feeding assays on CGSC6563

2.3.2.1.2. Validation of the selection using 2 enzymes with NAAAR activity

In order to validate the selection it was necessary to show that although the strain was unable to grow on *N*-acetyl-D-methionine, if the NAAAR was expressed into the strain, the stain would acquire the ability to grow on *N*-acetyl-D-methionine. The strain CGSC6563 was therefore transformed with pTTQ18, pRS7-1 and pRS8-9, and two assays were carried out in parallel on the three systems on M9/glucose minimal

medium, by inducing the expression of NAAAR. The results, shown in Table 4, confirmed the selection was working.

Media supplement	pTTQ18 (control)		pRS7-1		pRS8-9	
	No induction	Induction With IPTG	No induction	Induction With IPTG	No induction	Induction With IPTG
None	No	No	No	No	No	No
L-methionine	Yes	Yes	Yes	Yes	Yes	Yes
D-methionine	No	No	No	No	No	No
N-Ac-L-methionine	Yes	Yes	Yes	Yes	Yes	Yes
N-Ac-D-methionine	No	No	No	Yes	No	Yes

Table 4: Growth observed using CGSC6563 transformed with different plasmids (red highlight the specific effect of the NAAAR activity)

When the plasmids carrying the NAAAR genes were induced, the strain was able to grow on *N*-acetyl-D-methionine, as the only source of L-methionine. As no growth of CGSC6563 was observed when only D-methionine was supplemented but growth observed when the only supplement to the M9/glucose minimal medium was *N*-acetyl-D-methionine, the most likely explanation was that the *N*-acetyl-D-methionine was being converted into *N*-acetyl-L-methionine, consistent with the expression of functional NAAAR activity.

However, the colonies growing on *N*-acetyl-D-methionine were significantly smaller than the colonies growing on L-methionine or *N*-acetyl-L-methionine. This observation could be explained if only a small amount of *N*-acetyl-D-methionine was racemised, and therefore the rate of racemisation was insufficient for normal growth. Therefore this threshold for growth provided a basis to calibrate the strain for efficient selection of mutations leading to more efficient racemisation. Thus, the next stage was to optimise the cell growth in the selection to be able to identify strongly growing colonies (important for finding mutants). Two calibrations were required to achieve this.

The first experiment was to study the tolerance and growth of the strain when subjected to different concentration of cobalt. Several groups reported the requirement of a divalent metal ion for OSBS or NAAAR catalysis. The enolate anions are generated in the first coordination sphere of the divalent metal so these can be stabilized by potentially large electrostatic interactions. Although the M9/glucose minimal medium contained Mg^{2+}, NAAAR has better affinity with Co^{2+} for activity[70, 72, 73]. However it is also known that cobalt is toxic to *E. coli* above a certain concentration[122-124].

Therefore the cobalt tolerance of the auxotroph stain had to be tested. The 2 systems, CGSC6563/pRS7-1 and CGSC6563/pRS8-9, expressing the NAAAR were plated on M9/glucose containing *N*-acetyl-D-methionine as well as IPTG to induce the expression of NAAAR. As shown in Table 5, two parameters were taken in consideration to determine the optimal concentration of cobalt. Firstly the size of the colonies which could be interpreted as being related to the increase in the availability of L-methionine, and secondly the total number of colonies for the same reason.

$CoCl_2$ concentration (mM)	0	0.1	0.15	0.2	0.25	0.3	0.35	0.5
Colonies size	-	*	*	**	**	***	***	****
Colony number	-	*	*	**	**	***	**	*

Increasing number of asterisks represents an increase in size (or number)

Table 5: Cobalt tolerance of CGSC6563

The optimal concentration of cobalt was determined experimentally to be 0.3 mM. At this concentration the growth of the auxotroph benefited in two ways (size and number) from the concentration of cobalt as it was used by the NAAAR to catalyse the reaction of racemisation of the *N*-acetyl-D-methionine. However cobalt became toxic for the cells for a concentration above 0.3 mM reflected by the decrease of the number of colonies. This could be linked to the maximal uptake of cobalt into the binding site of the enzyme.

Moderate concentration (*e.g.* 0.3 mM) seemed to have a positive effect on the growth, but had however no influence on the speed of the growth, which still was 4-5

days to obtain a large colony size. These conditions may enable the isolation of cobalt independent variants.

Thus, the second calibration was based on the *N*-acetyl-D-methionine substrate concentration. As for the first calibration with cobalt tolerance, it was considered that it might be possible to increase the rate of cell growth by increasing the concentration of substrate and thereby further calibrating the selection. The usual substrate concentration used in M9/glucose minimal medium is 50 mg/l (0.262 mM) for *N*-acetyl-D-methionine. This concentration is not sufficient for normal cell growth as shown. This observation could be related to the high K_M values, respectively 11.3 mM and 24 mM, for the NAAAR from TS-1-60 (pRS7-1) and *A. orientalis* (pRS8-9). These values were determined by *in vitro* assay with purified enzyme.

Table 6 summarises the results of the effects of the substrate concentration upon cell growth using a cobalt concentration of 0.3 mM. Several measurements were taken. The first observation was on the colony growth: the growth was faster with *N*-acetyl-D-methionine concentrations at least 5 times higher than initially studied. As the concentration of *N*-acetyl-D-methionine increased, the duration of the incubation necessary to obtain significant colonies size decreased from 5 days to 2 days. The second observation concerned the size of the colonies: the higher the concentration, the bigger the colonies. The third concerned the number of colonies which increased as did the concentration of *N*-acetyl-D-methionine. But the most interesting observation arose from comparing the results at 41.83 mM and 52.3 mM *N*-acetyl-D-methionine: fewer colonies with pRS7-1 were observed at the higher concentration, but the number and the size of the colonies carrying pRS8-9 were still increasing. It was considered that this could reflect the reported substrate inhibition of these enzymes by *N*-acetyl-D-methionine (50 mM for pRS7-1, 200 mM for pRS8-9).

Table 6: Effect of the substrate on the growth of CGSC6563 using 0.3 mM $CoCl_2$

***N*-Ac-D-methionine concentration (mM)**		0.262	0.523	1.05	2.09	3.14	4.18	5.23	10.46	20.91	31.37	41.83	52.3
Growth (days)		5	5	4	4	3	3	3	2	2	2	2	2
Colonies size		Very small	*	**	**	***	***	***	****	****	****	*****	*****
Colonies number	**pRS7-1**	few	*	*	**	**	**	***	***	***	****	****	**
	pRS8-9	few	*	*	**	**	**	***	***	***	****	****	*****
Increasing number of asterisks represents an increase in size (or number)													

The optimal conditions for colony growth were thus established by consideration of the results outlined above. As the colonies should ideally be small enough to differentiate better racemase activity through increased colony size, the optimal concentration of *N*-acetyl-D-methionine should be 15 times higher (3.14 mM) than the concentration of *N*-acetyl-L-methionine which enables similar growth (size and number) to supplementation with L-methionine. However, the optimal concentration of *N*-acetyl-D-methionine to screen libraries of mutants could then be slightly decreased from this "threshold" to perhaps 1-2 mM to identify mutations in NAAAR which generate improvements in racemase activity, depending on how stringently the selection is to be operated.

2.3.2.1.3. Screening libraries of mutants

The first libraries to be screened were the libraries created by error prone PCR (section 2.2.1). These 3 libraries were used to transform CGSC6563 to detect racemase variants with better activity and plated directly on M9/glucose minimal medium supplemented with 2 mM of *N*-acetyl-D-methionine, 0.3 mM of $CoCl_2$ and 1mM IPTG. A sample of each transformation mixture was spread on LB agar Amp^{200} as control for transformation efficiency. No colonies of significant size appeared on the plates suggesting that the libraries of mutants did not possess any particular mutations that significantly improved the racemase activity.

However, the efficiency of cell transformation by plasmid DNA (Figure 28) was found to be extremely poor (10^7-10^8 expected transformants by electroporation): the 2 first samples yielded low numbers of transformants (9000, 7500) and the last one yielded 60,000 transformants.

Figure 28: Transformation of the 3 different libraries of TS-1-60 mutants created by error prone PCR

The second libraries to be screened were those created using the XL1-Red mutator strain. The five libraries (pRS7-1/1P to pRS7-1/5P libraries) were used to transform CGSC6563 and plated directly on M9/glucose minimal medium supplemented with 2 mM of *N*-acetyl-D-methionine, 0.3 mM of $CoCl_2$ and 1 mM IPTG. A sample of each transformation mixture was spread on LB agar Amp^{200} (200μg/ml) as control for transformation efficiency. As observed with the previous libraries screened, no colonies grew on the selective plates.

The cells plated on LB agar indicated that these transformations with plasmid DNA from the 5 cycles of mutation also yielded low levels, only 10^2-10^3 colonies/μg of DNA. Ideally, as mentioned earlier, the number of transformants should be ca 10^8 colonies/μg of DNA. A possible explanation for the low transformation efficiency was the purity of the DNA. This hypothesis was refuted by using the libraries of mutants to transform different commercially available strains of XL1-Blue (chemically competent and electrocompetent cells). The same quantities of plasmid DNA (50 ng DNA from pRS7-1 libraries) yielded 10^5-10^6 transformants/μg with these cells. Although lower than expected, this was significantly higher than the transformants obtained from CGSC6563 which yielded only 10^2-10^3 transformants/μg and suggested that for some reason this strain was poorly transformable with these plasmids.

Although CGSC6563 appeared promising for the selection experiment, the concerns over the poor transformation efficiency were significant. The low level of transformation mitigated against the benefit of the very high throughput of the *in vivo* selective method. Clearly improvement of the competency of the CGSC 6563 strain was needed and hence chemically competent and electrocompetent cells were prepared from CGSC6563. Repetition of the preparation of the chemically competent and electrocompetent cells was carried out by harvesting the cells at a different OD_{600} (from 0.4 to 0.8), using fresh isolates and modified protocols. However the results were reproducible for all the transformations, yielding at most of only a few thousand transformants when CGSC6563 was used. Consequently the auxotrophic strain was deemed not to be suitable for very efficient transformation. One possible explanation was that it may be possible that other mutations in that strain affected the transformation and thus rendering it inadequate for the desired purpose, namely the screening of mutant libraries with the highest number of transformants.

Therefore it was critical to the success of this approach to find a suitable auxotroph strain, which also could be transformed by recombinant mutated plasmid DNA at high efficiency. A new strategy was planned focussing on the deletion of 2 genes (one involved in the pathway of methionine biosynthesis and the second one as described earlier: *dadAX* gene) using a host strain, which was first tested for efficient transformation to create a more appropriate auxotroph for selection. This strategy had the additional benefit that it could employ deletion mutations that could not revert.

2.3.2.2. Modified strategy to create a methionine auxotroph

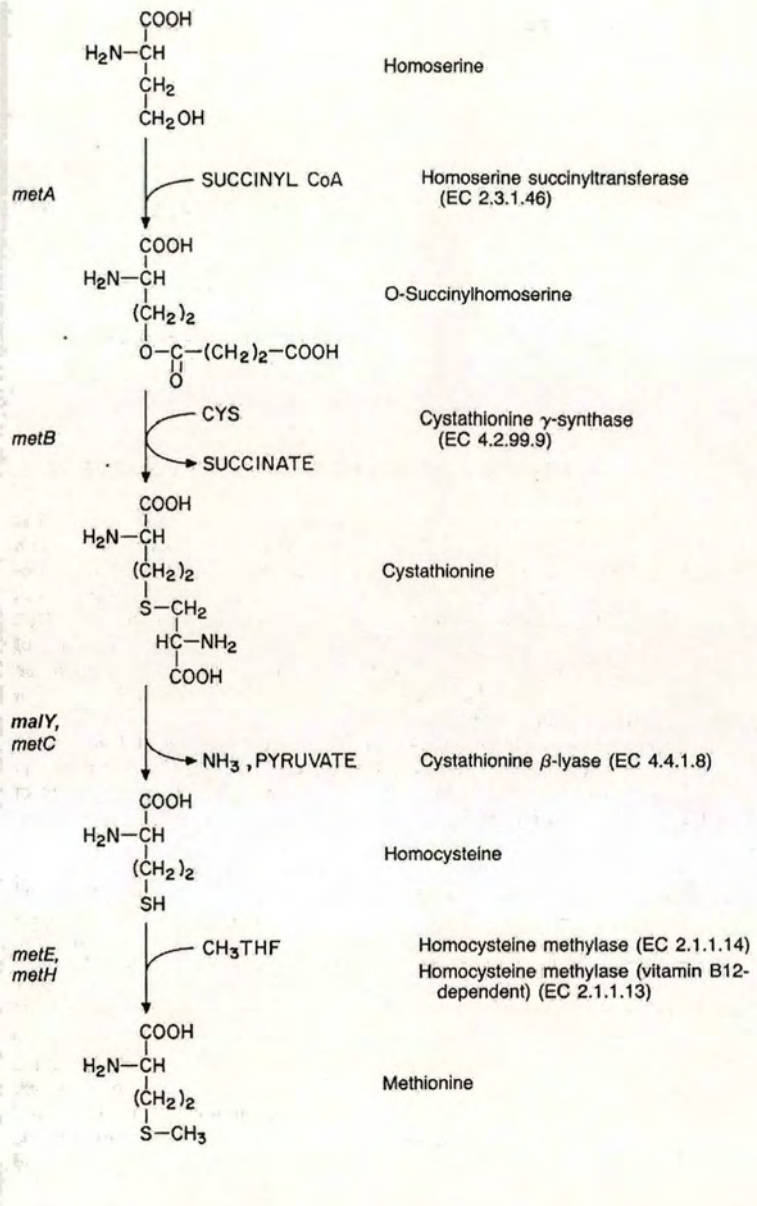


Figure 29: The pathway of methionine biosynthesis in *E. coli*

In order to create a methionine auxotroph strain, it was crucial to determine rationally which gene had to be inactivated. In *E. coli*, methionine is generated from homoserine (Figure 29) in four steps involving the participation of six different enzymes. Therefore four possibilities existed to disrupt methionine biosynthesis. The

choice of the particular step to interrupt methionine biosynthesis had to be considered carefully as it could affect the general health and fitness of the strain. Homoserine is converted into *O*-succinyl L-homoserine by homoserine succinyl transferase (encoded by the *metA* gene). The reaction of *O*-succinyl L-homoserine and L-cysteine is catalysed by *metB* (cystathionine γ-synthase) to yield succinate and cystathionine, which is then degraded into homocysteine by the action of *metC*. This reaction can also be catalysed by *malY* which is a bifunctional enzyme (repressor of maltose regulon / cystathionine β-lyase). The final step is an *S*-methyl transfer which is catalysed in two alternative ways by two enzymes, *metE* (cobalamine-independent) and/or *metH* (cobalamine-dependent), to yield methionine.

Interrupting the fourth step would therefore involve two deletions, which meant increased time as well as difficulty. Similarly, the interruption of the third step would also involve two deletions (time consuming, difficulties). Concerning the two first steps, none of the genes involved appeared to be essential for other functions in *E. coli*; therefore either of them could be deleted. However, it seemed more rational to interrupt the first step of the pathway, rather than the next one.

Therefore, for the above reasons the gene chosen for deletion in the pathway of the methionine biosynthesis was the *metA* gene.

2.3.2.2.1. Use of CGSC4474 (W3110)

See Appendix 11.1. for genotype of this strain.

The prototrophic rec^+ strain W3110 (CGSC4474) has no additional supplementary requirement for growth on M9/glucose minimal medium. The choice of this strain offered several advantages namely (i) 2 mutations only, which means a more simple growth medium, (ii) less likely to complicate the selective approach and (iii) if deletion mutations were introduced in *dadAX* and *metA*, no reversion would be possible to generate false positives. In addition, the introduction of antibiotic markers at the deleted loci would assist in reducing contamination on selective plates and make it easier to follow any changes at the deleted loci (by PCR). The idea was to use the deletion system developed by Datsenko And Wanner[120] directly on W3110

strain to inactivate the two genes involved in the biosynthesis of L-methionine (*dadA* and *metA*) and create a new auxotroph strain. However this strain is arabinose positive, and therefore is able to catabolise arabinose and use it as carbon and energy source. This was a concern as the gene deletion system uses arabinose in the induction of the RED system (pKD46), to enable gene deletions, so the function of arabinose as inducer could be compromised by its consumption as a carbon and energy source rather than a gratuitous inducer. This point is alluded to by the authors who reported this deletion system but the magnitude of the effect of using an ara^+ host is not reported.

Discussions with other groups using the method revealed problems when the strain BW25113 was not used as the host for the deletion. The system is clearly more efficient in an arabinose negative strain such as BW25113 than in arabinose positive strains. It was then decided to carry out inactivation of each gene in two stages: firstly create the deletion into BW25113 and secondly transduce the deletion into CGSC4474 (Figure 30). Attempts to introduce the mutation in CGSC4474 were unsuccessful. Alternatively, an additional deletion in *metA* could be introduced to the *dadAX* mutant of BW25113 (ET1) and used as the auxotroph strain if the transduction of the *dadAX* mutation was not successful.

The replacement of *dadA* and *metA* respectively by chloramphenicol and kanamycin resistance, would also allow more control over possible contaminants.

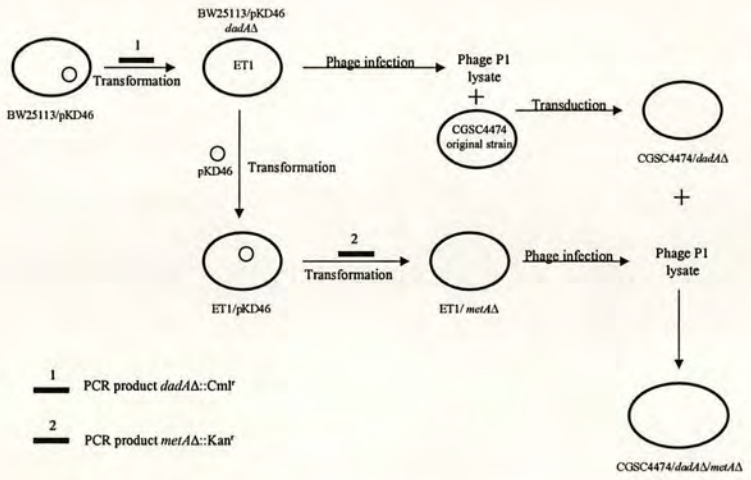


Figure 30: Auxotroph strain construction stategy

2.3.2.2.2. Use of BW25113 as the auxotrophic, selective host.

See Appendix 11.1. for genotype of this strain.

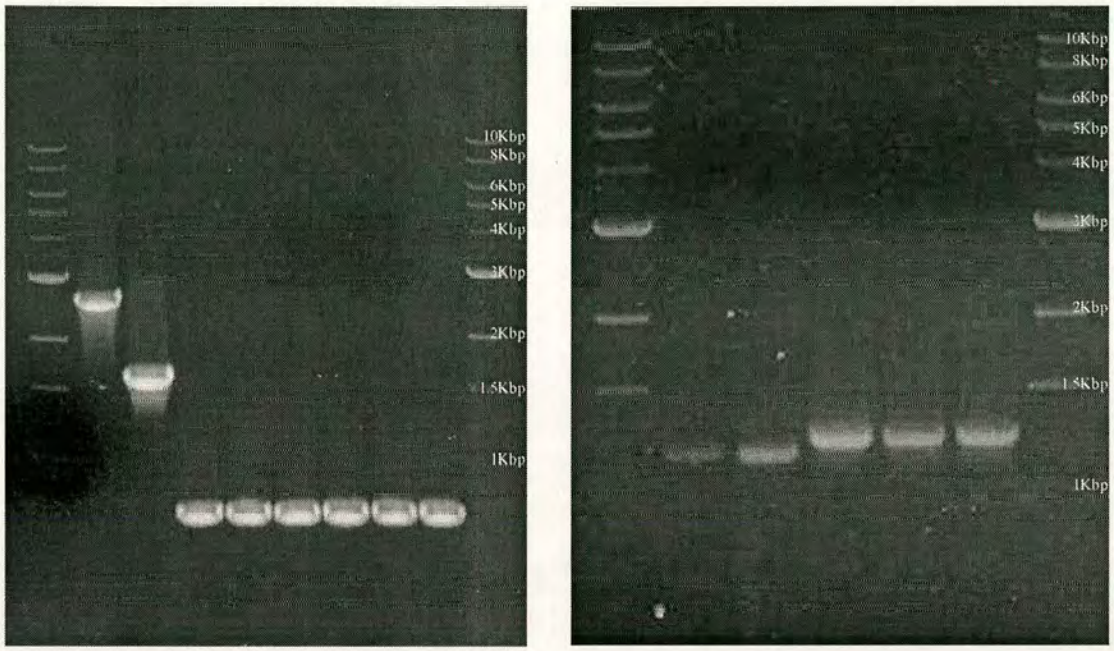
After several unsuccessful attempts to replace part of the *metA* gene, new primers were designed using longer homology extensions. However, this experiment proved to be unsuccessful. As an alternative, new primers were designed to delete part of the *metB* gene. After several unsucessful attempts to replace part of the *metB* gene with kanamycin resistance, the strategy was modified and a part of the *metB* gene was succesfully replaced with chloramphenicol resistance instead of kanamycin. Phenotypic expression (outgrowth) is less important for chloramphenicol resistance than kanamycin resistance. Kanamycin is an inhibitor of protein synthesis, which binds to the S12 protein of the 30S ribosomal subunit and thereby inhibits the initiation of synthesis of the polypeptide chain. However, when the kanamycin concentration is sufficiently low, initiation can still occur, causing extensive misincorporation of amino acids. This misincorporation happens because kanamycin alters codon-anticodon recognition, i.e. it induces misreading of the code. Thus spontaneous mutation frequencies to kanamycin resistance have been studied[125, 126], and appear to arise at high frequencies under conditions of low selective antibiotic concentrations and specific growth conditions. In contrast chloramphenicol inhibits peptidyl transferase activity of the 70S ribosome. This antibiotic is very powerful and hence more efficient than kanamycin (not only inhibits growth but also is toxic for the cells).

The *dadA-dadX* operon had previously been disrupted in the strain called ET1 (section 2.3.1.1.2). The deletion was confirmed by PCR (Figure 31, column 1) and growth on M9/glucose minimal medium (supplemented with different methionine derivatives). The *cat* gene (chloramphenicol acetyl transferase- chloramphenicol resistance) was introduced to select for the new strain. However, in order to use the chloramphenicol resistance a second time to select the disruption of the *metB* gene, the removal of the *cat* gene replacing the first deletion was required.

The strain ET1 was therefore transformed with pCP20[127] (FLP helper plasmid) and transformants were selected at 30°C. The helper plasmid shows thermal induction of FLP synthesis. pCP20 expresses the FLP recombinase, which acts on the directly repeated FRT[128] (FLP recognition target) site flanking the resistance gene. The strain was then simply cured of pCP20 by growth at 43°C, as the plasmid has a temperature-sensitive replicon. The loss of the *cat* gene was checked by PCR on 6 different transformants using flanking oligos (Figure 31, columns 1'). The new strain called SET1 was purified 4 times at 43°C to be sure that the plasmid was completely cured.

The inactivation of the *metB* gene using the same antibiotic marker and method was performed on two strains: SET1 and BW25113. The deletion of the *metB* gene was successful using a homologous priming sequence longer than described in the reported method (70 bp instead of 36-50 bp recommended) but failed on previous attempts using 50bp homologous priming sequences either with kanamycin or chloramphenicol resistance genes. The strain derived from SET1 was named SET2, while the strain derived from BW25113 was called SET3. The correct disruption was checked by PCR (Figure 31, columns 21, 22 and 3). To assist selection and control of the new strain the chloramphenicol resistance was retained in the genome.
The transduction of the two disrupted genes into CGSC4474 to create an auxotroph strain was considered. However, as a result of the two disruptions introduced in BW25113, the new mutant SET2 seemed perfectly appropriate for as the auxotroph strain and therefore, it was decided to assess SET2 for auxotrophy and thus acceptability for *in vivo* selection.

First deletion: *dadAX* **Second deletion: *metB***



C 1 1' (Cml^R removed)

C: control (wild type genes) 2609 bp
1: ET1 (BW25113 with *dadAX* inactivated) 1749 bp,
1': SET1 (ET1- Cml^S) 649bp,

C C 21 22 3

C: control (wild type gene) 1161 bp
21, 22: SET2 (SET1 with *metB* inactivated)
3: SET3 (BW25113 with *metB* inactivated)
Size disrupted gene (21,22,3): 1330 bp

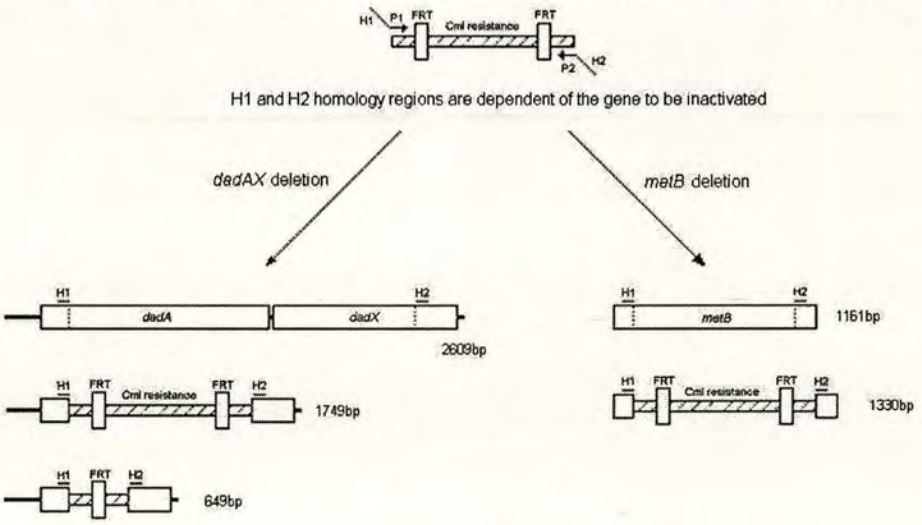


Figure 31: PCR verification of the deletions

2.3.2.2.3. Feeding studies of SET2

See Appendix 11.1. for genotype of this strain.

In order to validate the selection, the growth of the newly created strain had to be tested on M9/glucose minimal medium for auxotrophy. As expected, the strain was unable to grow if not supplemented either with L-methionine or *N*-acetyl-L-methionine due to the deletion of the 2 genes (Table 7).

Growth medium	***metB*$^+$**	***metB*$^-$**		
	SET1	**SET21**	**SET22**	**SET3**
M9	Yes	No	No	No
M9+L-methionine	Yes	Yes	Yes	Yes
M9+ *N*-acetyl-L-methionine	Yes	Yes	Yes	Yes
M9+D-methionine	Yes	No	No	Yes
M9+ *N*-acetyl-D-methionine	Yes	No	No	Slight

Table 7: Growth of the newly created strains from BW25113 observed on different media

The auxotrophy of the SET2 strain was confirmed, and its ability to hydrolyse *N*-acetyl-L-methionine meant no requirement to use a complementary acylase. Similarly, the strain SET3 (*ΔmetB::Cml*^{R}) was not prototrophic as the biosynthesis of methionine was interrupted: no growth was observed any more on M9/glucose minimal medium. However, the expression of *dadA,* which was not inactivated in this strain, facilitated deamination of D-methionine as growth was still observed on D-methionine and *N*-acetyl-D-methionine. This strain however serves as an appropriate control to demonstrate efficient transport of the D-amino acid substrates (Figure 32).

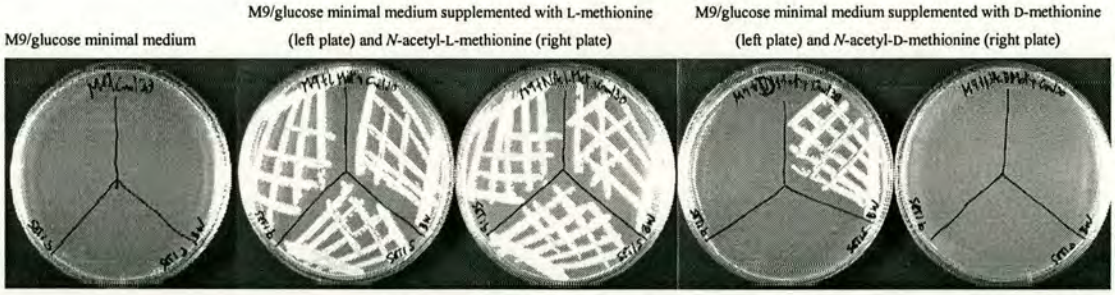


Figure 32: Growth assays on strains SET21(top left), SET22 (bottom) and SET3 (top right)

A new auxotroph strain SET2 requiring L-methionine as substrate was created using *E. coli* BW25113 (exonuclease deficient). 2 genes were knocked out: *dadAX* (D-amino acid dehydrogenase/alanine racemase) and *metB* (*O*-succinylhomoserine(thiol)-lyase), preventing the mutant strain from growing if not supplemented with L-methionine. SET2 could potentially be a good candidate for *in vivo* selection.

2.3.2.2.4. Validation of the selection using NAAARs

As the SET21 strain was proven to be a methionine auxotroph, and unable to grow on *N*-acetyl-D-methionine or D-methionine, then if the NAAAR is expressed into the strain, we would expect *N*-acetyl-D-methionine to be racemised. The strain would then be able to grow when supplemented only with *N*-acetyl-D-methionine, as *N*-acetyl-L-methionine would be generated. To test this proposal, the strain SET21was transformed with pTTQ18, pRS7-1 and pRS8-9, and two assays (with and without induction of the expression of NAAAR) were carried out in parallel on the three systems in M9/glucose minimal medium. The results, shown in Table 8, confirmed the phenotype to be as expected and appropriate for selection. When induced, the recombinant strain SET21 expressing NAAAR was able to grow on *N*-acetyl-D-methionine, consistent with the production of L-methionine. As the only supplement incorporated to the M9/glucose minimal medium was *N*-acetyl-D-methionine, the

only explanation was that the substrate was converted into *N*-acetyl-L-methionine, establishing the expression of the NAAAR activity.

Media supplement	pTTQ18 (control)		pRS7-1		pRS8-9	
	No induction	Induction With IPTG	No induction	Induction With IPTG	No induction	Induction With IPTG
None	No	No	No	No	No	No
L-methionine	Yes	Yes	Yes	Yes	Yes	Yes
D-methionine	No	No	No	No	No	No
N-Ac-L-methionine	Yes	Yes	Yes	Yes	Yes	Yes
N-Ac-D-methionine	No	No	No	Yes	No	Yes

Table 8: Growth observed using SET21 transformed with different plasmids (red highlight the specific effect of the NAAAR activity)

Thus, the next stage was to optimise the cell growth in the selection to be able to identify significant colonies (important for finding mutants). Two calibrations were required to do so.

2.3.2.2.5. Cobalt tolerance of SET21

As mentioned earlier, (section 2.3.2.1.2.) NAAAR requires a divalent metal ion for catalysis. Although the M9/glucose minimal medium contained Mg^{2+}, NAAAR has higher activity in the presence of Co^{2+} ions. However, as also noted above, cobalt is also toxic to *E. coli* at high concentrations. Therefore cobalt tolerance of the auxotroph stain had to be tested. The 2 NAAARs systems were plated on M9 supplemented with *N*-acetyl-D-methionine and IPTG to induce the expression of NAAAR.

An empirical determination of the optimum concentration of Co^{2+} was used. The experiment, based on the growth of the strain using a range of Co^{2+} concentrations from 0 to 1mM (0.05mM increments). To minimize the errors from measurements, 4 replicates of plates were used (Figures 33 and 34). The error in measuring the size is

around 0.1mm. Increasing the concentration of Co^{2+} did not have any positive effect on the number of colonies. The number of colonies was found to decrease at a constant rate (Figure 33). At a concentration of cobalt greater than 0.4 mM, less than 20% of the colonies survived.

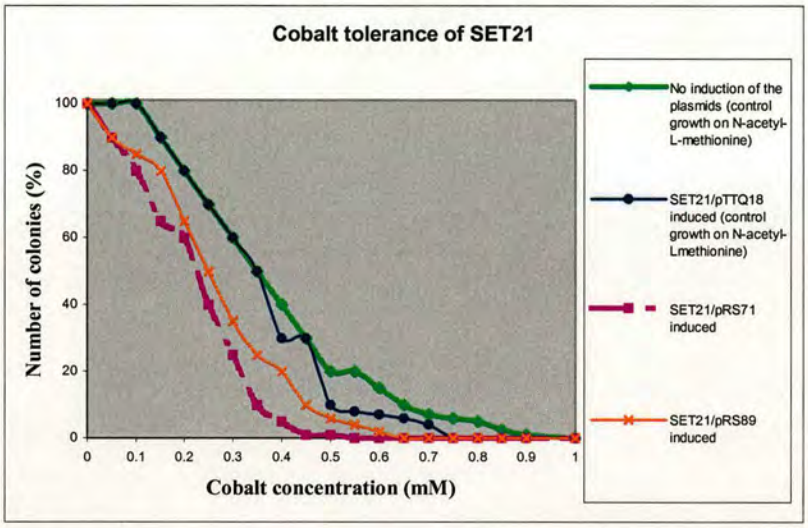


Figure 33: Effect of cobalt on the number of colonies

However, the increased concentration of cobalt had a slight positive effect on the size of the colonies (Figure 34). Thus when comparing the growth of cells bearing the cloned NAAAR genes or control plasmid (pTTQ18), at concentrations of cobalt lower than 0.3 mM, the inhibitory effect of cobalt on colony size was not as significant upon colonies expressing NAAAR as in the control. This observation might be explained by the fact cobalt was sequestered by the NAAAR enzymes. At concentrations of cobalt higher than 0.3 mM, the size of colonies became very small and the colonies were very difficult to distinguish.

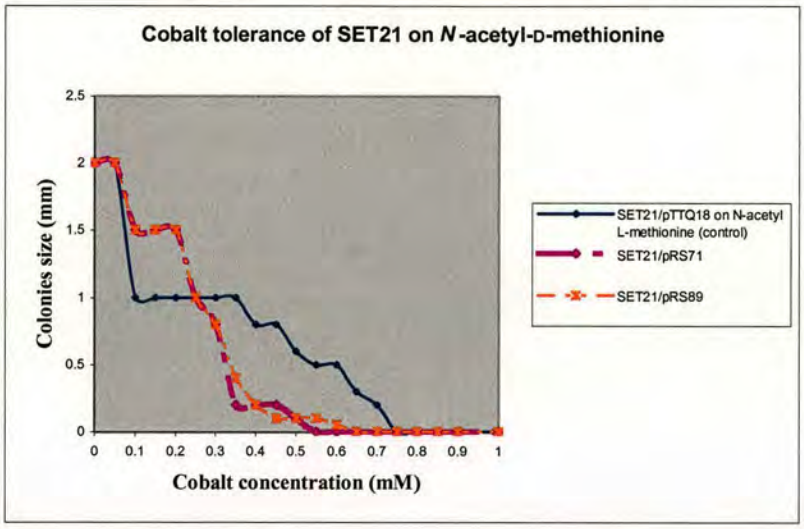


Figure 34: Effects of cobalt concentration on the size of colonies

The growth decreased rapidly after only minimal supplementation with Co^{2+} ions. The expression of the 2 racemases seemed to affect the growth/ health of SET21 (comparison of the growth when induced or not, Figure 33). This observation was supported by the growth of the control plasmid (pTTQ18), as its growth was identical to the assay with or without induction. To verify this hypothesis, growth controls were performed with the 3 systems on L-methionine (Figure 35) to compare rate of growth with the *N*-acetyl-D-methionine assays.

As suggested, the expression of the two racemases affected the health and growth of SET21. When grown on M9 + L-methionine, decrease of growth of the three systems was observed as the concentration of cobalt increased. But when growth was performed on M9 + L-methionine + IPTG, a difference of the growth was observed between the control system (pTTQ18, top left portion of the plate) and the 2 NAAAR systems (pRS7-1 and pRS8-9, respectively bottom portion and top left portion). This difference of the growth was possibly due to the fact that the genes were expressed in a high-copy vector (~100 copies per cell). As the protein was produced, the carbon source was depleted more rapidly, and the growth suffered accordingly. At a concentration of 0.25 mM Co^{2+} and above, the colony size was so small that individual colonies were difficult to distinguish.

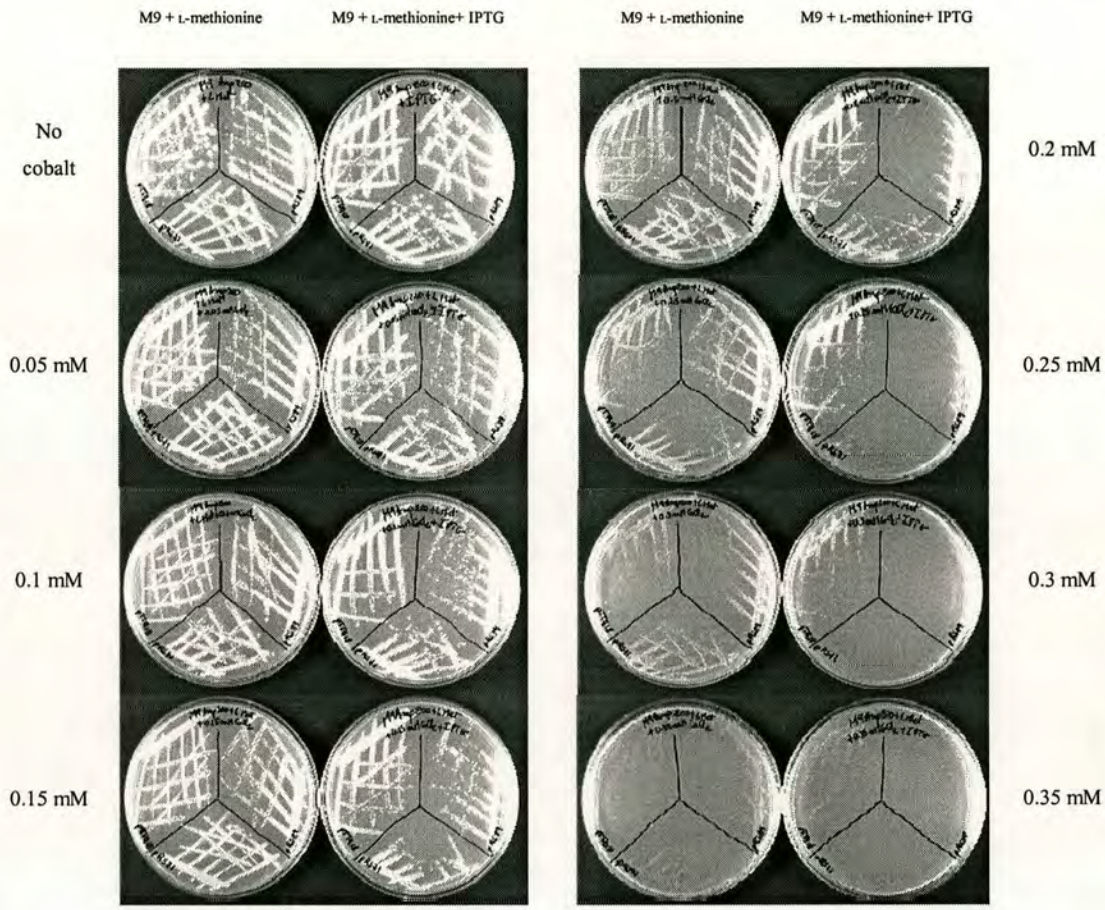


Figure 35: Cobalt tolerance of SET21 when NAAAR expressed. Top left: control using pTTQ18; bottom: using pRS7-1 (NAAAR from TS-1-60); top right: using pRS8-9 (NAAAR from A. orientalis))

Considering the fact that a concentration of 0.2 mM Co^{2+} in the assay produced colonies which were 75% of the normal size, the concentration for cobalt should not exceed 0.2 mM in the assay. Using a range of cobalt concentration from 0 to 0.2 mM therefore seemed to be most appropriate as the aim of the screening is to isolate improved enzymes (*e.g.* greater reaction rate, less cobalt-dependant, higher level of resistance to substrate inhibition).

2.3.2.2.6. Effect of the substrate concentration on SET21

Using SET21 transformants, similar results were found to the effects of substrate concentration when using transformants of CGSC6563 (Table 6). As discussed earlier, the optimal conditions were then established considering these results. As the colonies have to be small enough to differentiate better NAAAR activity, the optimal concentration of *N*-acetyl-D-methionine to screen libraries of mutants should be 15 times higher (3.14 mM) than the concentration of *N*-acetyl-L-methionine used to test the selection. This gives the same kind of growth (colony size and number) as when the medium is supplemented with L-methionine. For the screening of libraries, this optimal concentration could be slightly decreased to allow a better differentiation of improvement of activity, depending on how stringent the selection is intended to be.

2.4. Screening libraries of mutants

The 13 libraries of randomly mutated pRS7-1 plasmid DNA (sections 2.2.1, 2.2.2.1 and 2.2.2.2) were used to transform SET21 and plated directly on M9/glucose minimal medium supplemented with 2 mM of *N*-acetyl-D-methionine, 0.2 mM of $CoCl_2$ and 1 mM IPTG. A sample of each transformation mixture was spread on LB agar Amp^{200} as control for transformation efficiency. Libraries were transformed with a relatively good efficiency (10^5-10^6 transformants/μg of DNA).

After picking and replating the largest and fastest growing colonies, on identical media, only one mutant appeared to have a significantly improved growth rate. This mutant (designated 3P2) was isolated from the library called pRS7-1/3P, which was created from the third cycle of growth of the mutator strain grown on plates. The plasmid was isolated and, used to retransform fresh SET21 cells, and then plated onto M9/glucose minimal medium supplemented with 2 mM of *N*-acetyl-D-methionine, 0.2 mM of $CoCl_2$ and 1 mM IPTG to validate that the growth was due to the plasmid expression and not from the host. In principle the strain should not have reverted, although some secondary mutation could in theory have occurred. The growth was confirmed as plasmid dependent.

Mutant 3P2 was then plated on M9/glucose minimal medium supplemented with a range of concentrations of *N*-acetyl-D-methionine, ranging from 0.26 mM to 2.00 mM, to determine the threshold for growth of the 3P2 colony in comparison to the wild type NAAAR. This experiment was designed to provide an indication of the improvement in the activity of the NAAAR. Surprisingly, at the lowest concentration of *N*-acetyl-D-methionine (0.26 mM), the strain SET21/3P2 grew at the same rate as if it was supplemented with L-methionine, whereas the strain bearing the unmutated plasmid (SET21/pRS7-1) required a concentration of *N*-acetyl-D-methionine of 2 mM to show a similar rate of growth. This significantly reduced requirement for *N*-acetyl-D-methionine in comparison to the unmutated pRS7-1 strongly suggested that the NAAAR activity in the cell had been significantly increased. The concentration of *N*-acetyl-D-methionine had in fact been reduced by a factor of 7.63. This factor correlates well with data generated *in vitro*, as will be discussed later. As this mutant

was produced by random mutagenesis using XL1-Red mutator strain, the possibility arose that expression of the gene had been increased since not only the gene but the entire plasmid had been subjected to random mutations. Protein gels of whole cell cultures were then performed to determine whether the increased rate of growth of the SET21/3P2 mutant was attributable to the NAAAR enzyme being expressed at a higher level in the cell than the NAAAR of SET21/pRS7-1. The results (Figure 36) indicated no obvious increased level of production of NAAAR (The NAAAR band is expected to be ca 39 kDa, as indicated by the arrow).

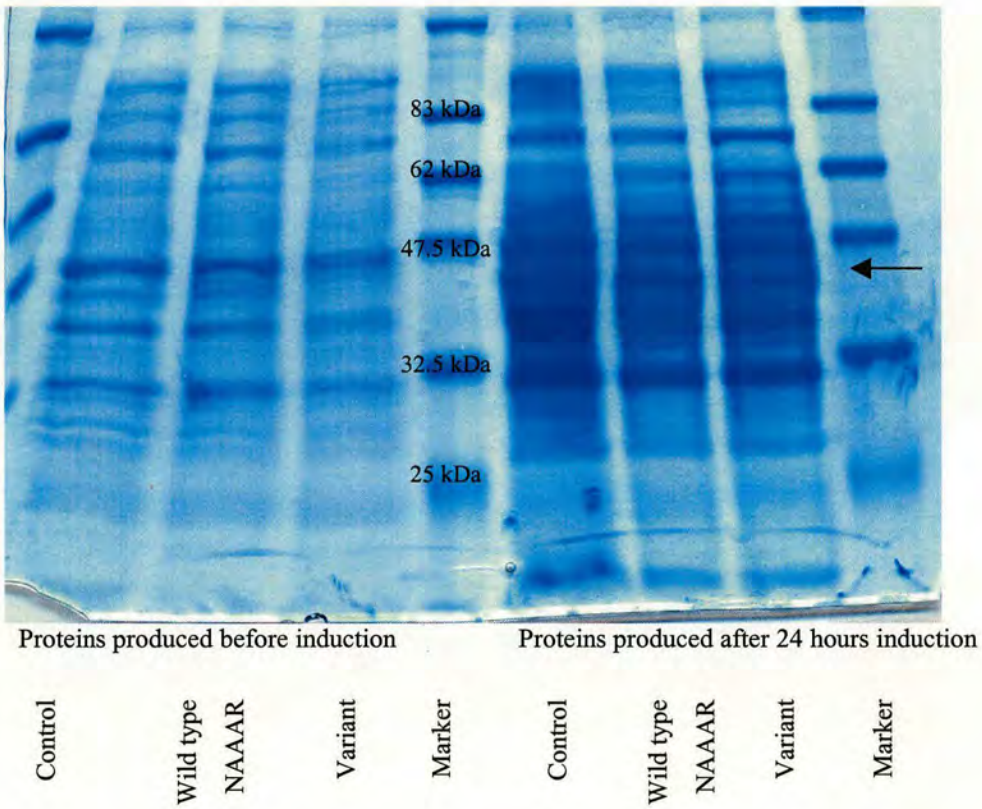


Figure 36: Protein gel of pRS7-1 and its variant 3P2

Plasmid DNA of pRS7-1 and 3P2 was prepared and complete, double stranded sequence analysis was carried out on the coding and regulatory regions of the NAAAR gene from pRS7-1 and 3P2. Alignment of the deduced amino acid sequences of the NAAAR genes from pRS7-1 and 3P2 showed that a single DNA change had been introduced ($G_{878} \rightarrow A_{878}$) in the NAAAR gene of 3P2. This change

leads to a single amino acid difference between the wild type TS-1-60 NAAAR, which has glycine (GGG) at position 293, and the 3P2 variant in which the glycine has been replaced by glutamic acid (GAG) at position 293 (See Appendix 11.5.).

2.5. Chiral HPLC assays

An efficient assay was required to assess *in vitro* dynamic kinetic resolution and to analyse the racemisation versus the hydrolysis of *N*-Ac-D-Amino Acids. Chiral high-performance liquid chromatography (HPLC) columns are used in both application development and preparative work. HPLC has proven to be useful for resolving racemates and the determination of absolute configuration of enantiomers and has become the method of choice for chiral separations[129]. Many chiral columns have been used to measure the enantiomeric composition of amino acids or *N*-acylamino acid derivatives individually. However no method has been developed to measure simultaneously the enantiomeric composition of amino acids and their *N*-acyl derivatives. Recently a Taiwanese group has investigated several HPLC chiral columns and reported[130] a suitable column to measure the enantiomeric composition of amino acids and *N*-acetyl amino acids simultaneously and directly from the enzymatic reaction solution (aqueous solution), more specifically using DL-methionine and *N*-acetyl-DL-methionine. This column, CHIROBIOTIC T™, developed by D.W. Armstrong, is commercialised by ASTEC (Advanced Separation Technologies, Whippany, USA). The column is based on covalently bond of an amphoteric macrocyclic glycopeptide (teicoplanin) to silica gel as the chiral stationary phase ligand and can be used in both the normal-phase and reversed-phase modes[131, 132]. Teicoplanin is naturally produced by the *Actinoplanes teicomycetus* mildew and contains 20 asymmetric centers surrounding four cavities. The teicoplanin chiral stationary phase appears to have excellent enantioselectivity for native amino acids, peptides, α-hydroxycarboxylic acid, and a variety of neutral analytes, including cyclic amides and amines[133]. The separation mechanism is believed to involve hydrogen bonding, dipole stacking, steric and charge interaction, π–π complexation, and inclusion.

2.5.1. Standard calibration

In order to measure the enantiomeric composition of amino acids and *N*-acetyl amino acids simultaneously, standard calibration chromatograms (Figure 37) had to be run to identify the optimal mobile phase composition. The concentration used was 10mM for all amino acids and *N*-acetyl amino acids.

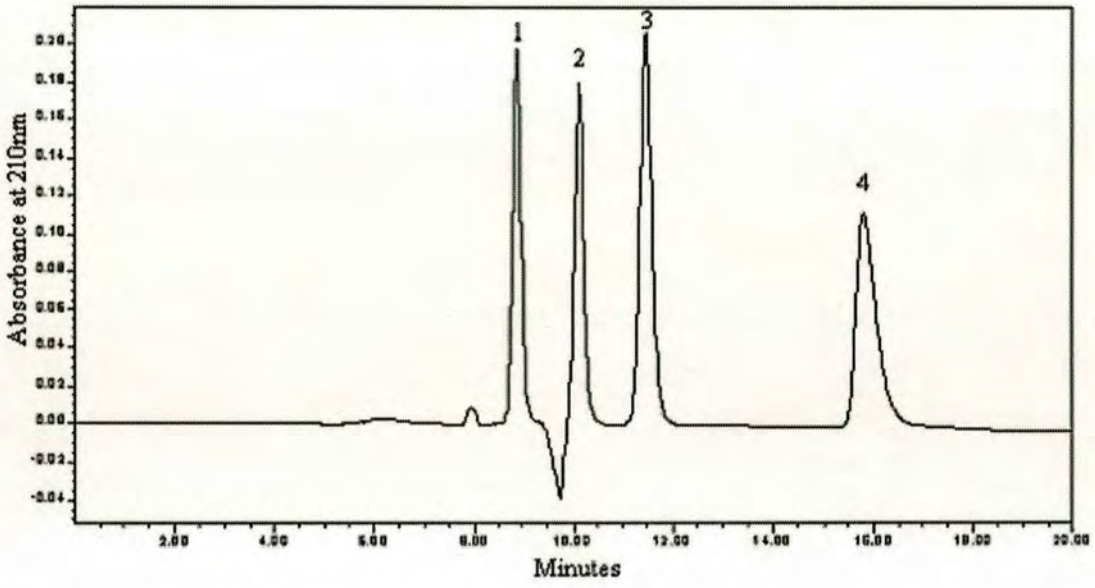


Figure 37: HPLC profile of D-, L-methionine and *N*-acetyl-D-, L-methionine
(1: L-Met, 2: D-Met, 3: *N*-acetyl-L-Met, 4: *N*-acetyl-D-Met)

2.5.2. Simultaneous analysis of *in vivo* racemisation and hydrolysis by HPLC

pRS7-1 was expressed in JM109 by induction with IPTG. The cells were recovered by centrifugation, and resuspended into an aqueous solution containing 5 mg/ml of *N*-acetyl-D-methionine (0.262 mM). An initial sample was taken and injected as reference (Figure 38). The mixture was then incubated at 37°C, and a sample was taken at intervals of 1, 2, 4 and 48 hours and injected into the column (Figure 39).

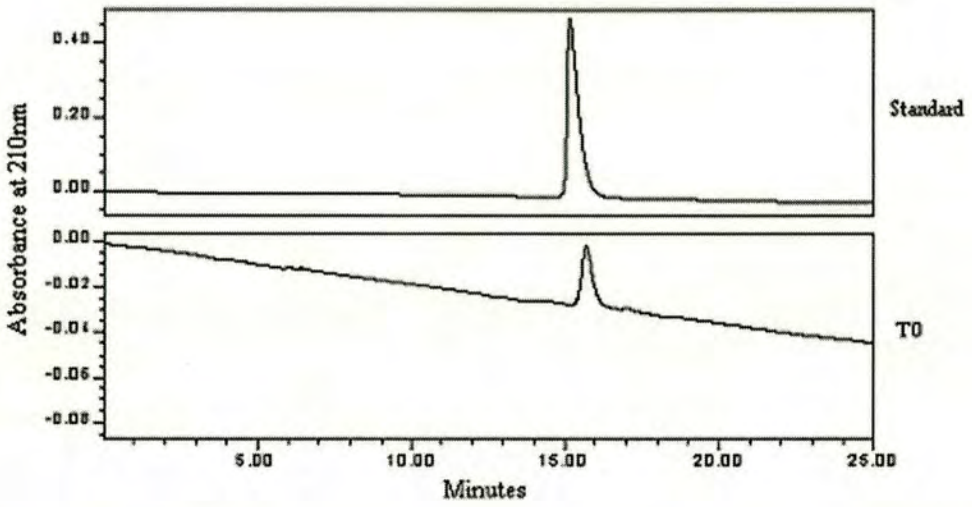


Figure 38: Initial profile of the enzymatic reaction containing 0.262mM *N*-acetyl-D-Met

T0: Start of the *in vivo* reaction; standard concentration used: 10mM

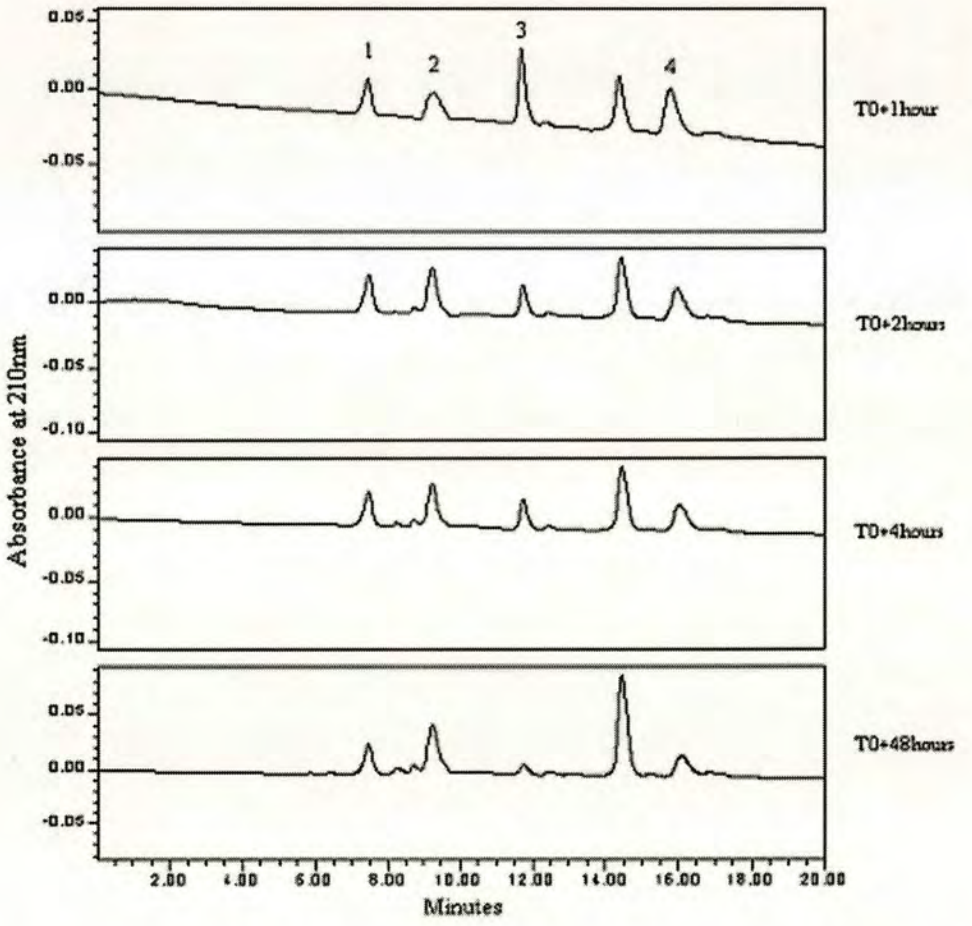


Figure 39: HPLC profile of the *in vivo* racemisation and hydrolysis of 0.262mM *N*-acetyl D-Met

The clear progress of separation of the amino acids and their derivatives allowed the various reactions to be followed in real time, namely (i) the racemisation of *N*-acetyl-D-methionine (4) into *N*-acetyl-L-methionine (3), (ii) the hydrolysis of *N*-acetyl-L-methionine (3) forming L-methionine (1), and (iii) the hydrolysis of *N*-acetyl-D-methionine (4) forming D-methionine (2). However, after 2 hours, the reaction slowed down as the cells died.

Thus, the HPLC assay appeared to be promising for assessing the activity of NAAAR semi-quantitatively. This assay is therefore complementary to the *in vivo* selection, which is more qualitative.

2.6. HPLC assay on the mutant NAAAR

The data generated *in vivo* indicated that a single mutation in the TS-1-60 gene had given rise to an NAAAR variant which enhanced the growth of *E. coli* under conditions where NAAAR activity was the rate limiting factor for growth. To support this finding, and verify that the 3P2 variant did in fact possess a higher level of NAAAR activity than its progenitor, an assay for NAAAR activity was developed *in vitro*, and was carried out using extracts of SET21/pRS7-1 and SET21/3P2. Two different experiments were carried out. The first experiment attempted to validate the fact that cobalt was absolutely required for racemisation as the enzyme is a metalloenzyme, and to test the 3P2 variant for possible reduced cobalt dependence. The second experiment was carried out to more accurately quantify the level of racemisation of *N*-acetyl-D-methionine by 3P2 and pRS7-1 and to correlate the results with those observed with the *in vivo* selection concerning an improvement of the racemase activity of 3P2.

2.6.1. Calibration of the HPLC profile

In order to identify the different peaks on the HPLC profile, standards were run under the conditions to be used for the racemisation assays as shown in Figure 40.

The separation system (mobile phase composition) used was different to the one described above in section 2.5.2.

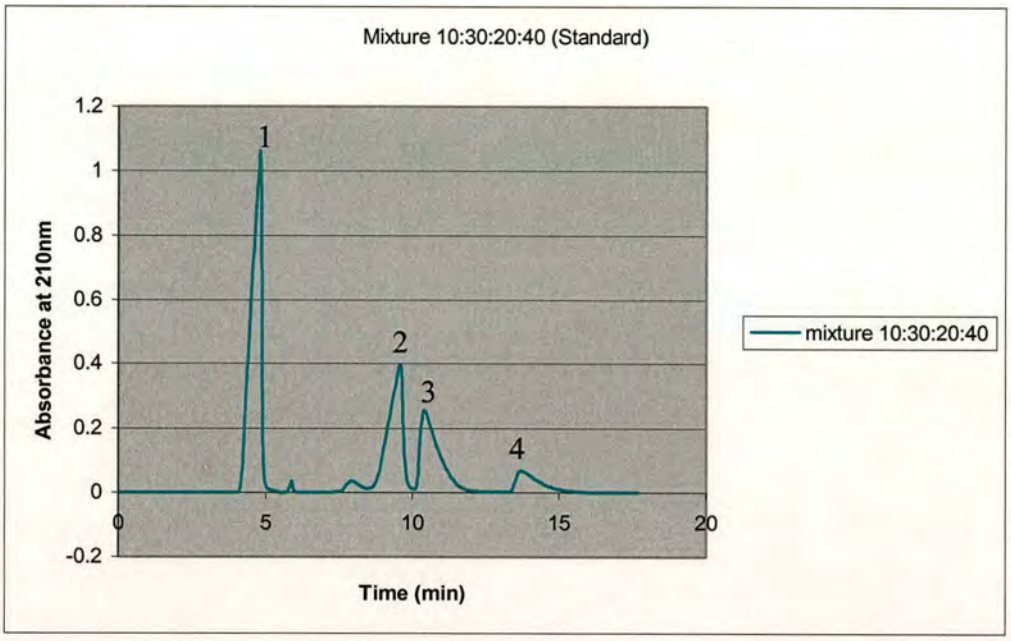


Figure 40: HPLC profile of a mixture of 10mM D-methionine(4), 30mM L-methionine(2), 20mM *N*-acetyl-D-methionine(3) and 40mM *N*-acetyl-L-methionine(1)

It was important to be able to accurately quantify the different products in the mixture to follow the progress of the racemisation reaction. The areas of the different peaks reflected the quantity of the different methionine enantiomers and derivatives. The intensity of the peaks for the enantiomers were different but it appeared that a relationship existed between the concentrations and the areas of the peaks corresponding to *N*-acetyl-L-methionine and *N*-acetyl-D-methionine. Determination of the unknown amount of one enantiomer was therefore possible using the known concentration of the other enantiomer. Experimentally, using the specific conditions described in section 9.2.2., it was found that:

$$\text{Area}_{\text{N-acetyl-L-methionine}} = 1.1475 * \text{Area}_{\text{corresponding to equivalent absorption N-acetyl-D-methionine}}$$

Therefore it was possible to determinate the conversion of *N*-acetyl-D-methionine into *N*-acetyl-L-methionine.

2.6.2. NAAAR requirement for cobalt

NAAAR has been widely described as being a metalloenzyme. It requires a metal ion, like magnesium, manganese, or zinc but gives the highest activity using cobalt[70]. It was decided to assess the activity of the 3P2 variant with cobalt using the optimal concentration of 6 mM determined experimentally by different groups. As shown in Figures 41 to 44, the assays were carried out over a period of 24 hours. For clarity, the figures show the experiments only at the beginning and after 24 hours, although samples were taken regularly.

When no cobalt was added, racemisation did not occur with the wild type racemase, as no peak was observed at 5 minutes (the retention time corresponding to *N*-acetyl-L-methionine) (Figure 41). This was an important observation as it was thought that the presence of metal ions within the cells might have been sufficient to enable some NAAAR activity.

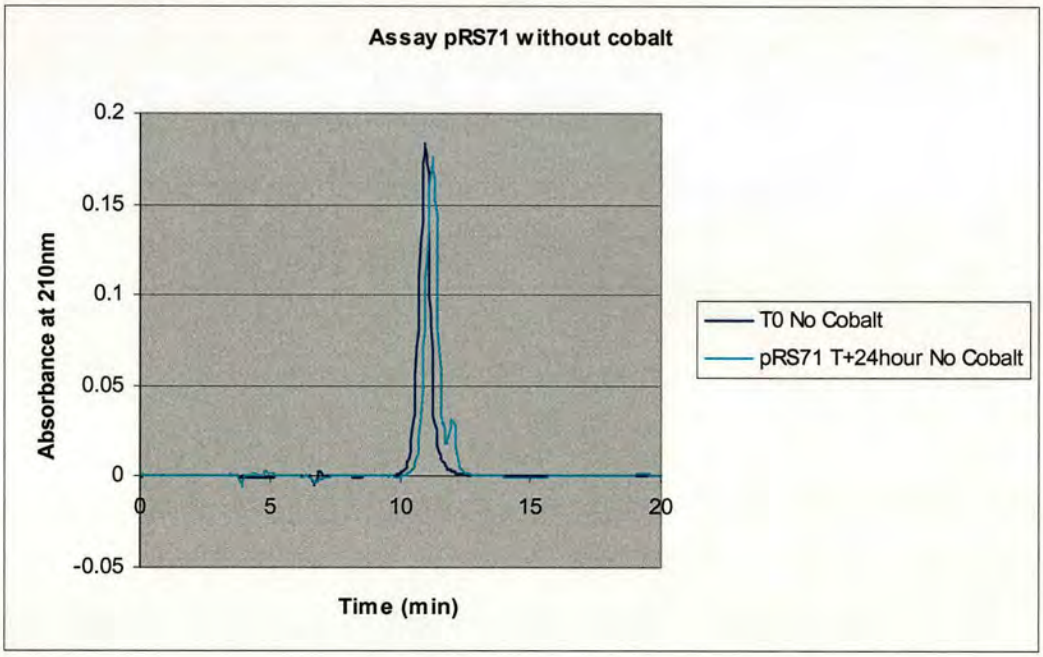


Figure 41: HPLC profile of biotransformation using 10mM *N*-acetyl-D-Met and pRS7-1 expressed in 50mg of wet weight of SET2-1 without addition of cobalt

However, as shown in Figure 42, when pRS7-1 was replaced by the 3P2 variant, the HPLC profile showed a slight degree of racemisation of 2.1% (presence of a small peak at 5 minutes). In addition a corresponding decrease in the concentration of *N*-acetyl-D-methionine was also observed. The latter could possibly be explained by hydrolysis to D-methionine since there also appeared to be a new peak at 15 minutes which corresponded to D-methionine. Approximately 50% of the *N*-acetyl-D-methionine had disappeared, which could not be explained (converted and used by some surviving cells- at this concentration of cobalt no cells could normally survived- cobalt tolerance of the strain- no growth at all for a concentration above 1 mM).

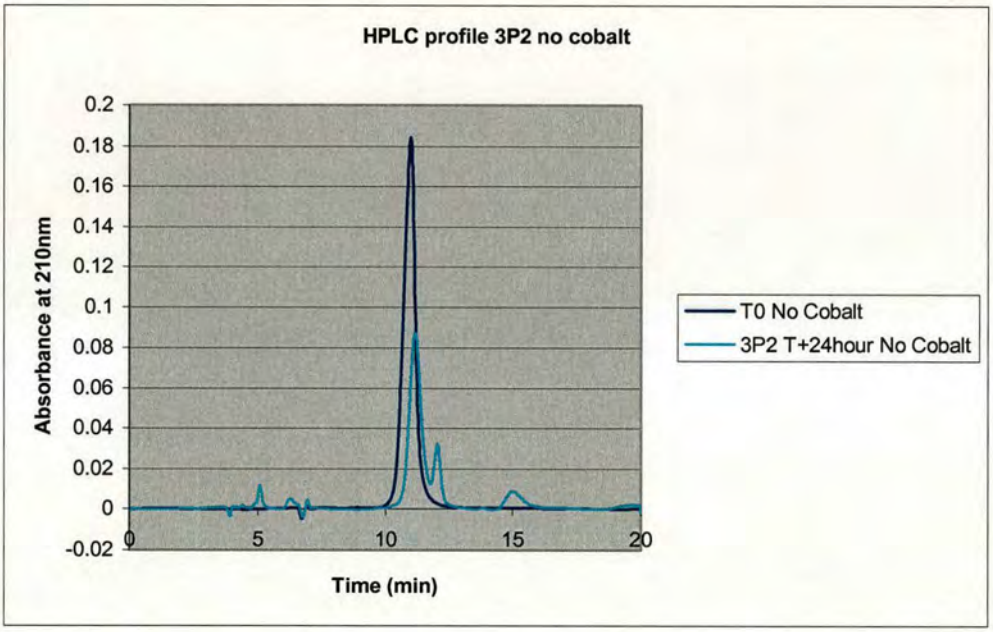


Figure 42: HPLC profile of biotransformation using 10mM *N*-acetyl-D-methionine and 3P2 expressed in 50mg of wet weight of SET2-1 without addition of cobalt of 50mg of wet weight

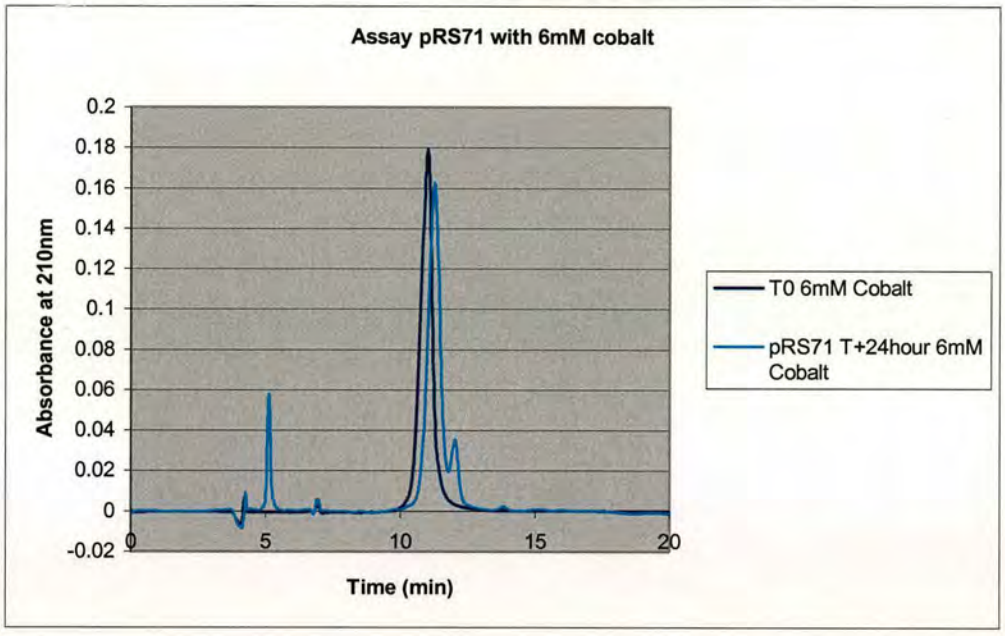


Figure 43: HPLC profile of biotransformation using 10mM *N*-acetyl-D-methionine and wild type enzyme expressed in 50mg of wet weight of SET2-1 and assayed with 6mM cobalt

When cobalt was used in the assays, both the 3P2 and the pRS7-1 wild type NAAAR enzymes showed activity. The wild type enzyme showed a lower level of activity, and racemisation after 24 hours was incomplete (Figure 43). Only 10.1% of *N*-acetyl-D-methionine was racemised (14.1% conversion if the concentration of *N*-acetyl-L-methionine was compared to the concentration of the residual *N*-acetyl-D-methionine).

Compared to the wild type enzyme, the variant enzyme showed significantly greater racemase activity, with almost complete racemisation (75.8% conversion if the concentration of *N*-acetyl-L-methionine was compared to the concentration of the residual *N*-acetyl-D-methionine, 16.6% conversion if the concentration of *N*-acetyl-L-methionine was compared to the initial concentration of *N*-acetyl-D-methionine) occurring within 24 hours (Figure 44).

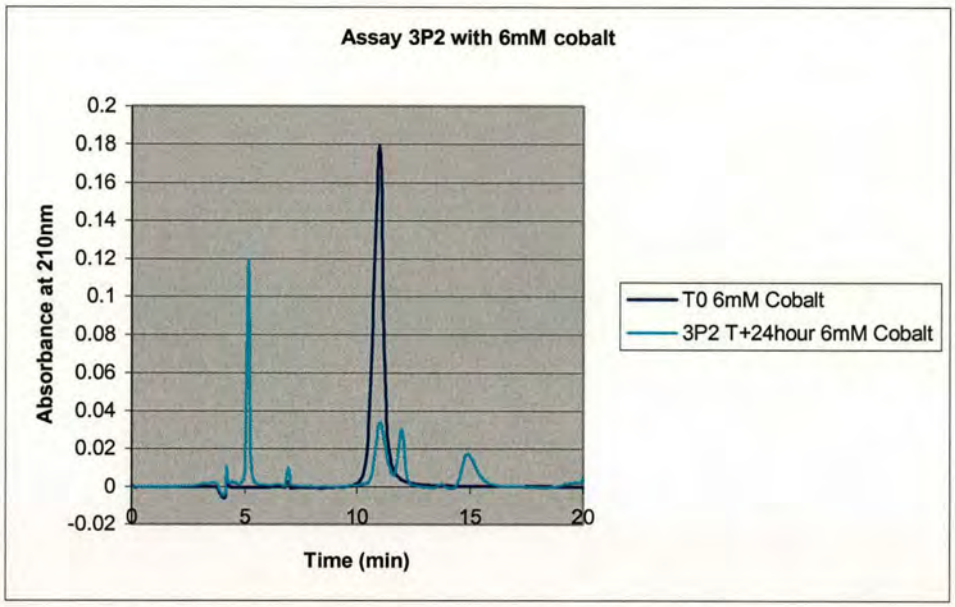


Figure 44: HPLC profile of biotransformation using 10mM *N*-acetyl-D-methionine and the variant enzyme expressed in 50mg of wet weight of SET2-1 assayed with 6mM cobalt

However, another peak appeared with similar retention time (15 mins) to that of D-methionine. This was unexpected because if this was indeed D-methionine, a similar peak due to hydrolysis might be expected with SET21/pRS7-1 since the only difference between the two strains was the single amino acid mutation of NAAAR. In addition, the hydrolysis of *N*-acetyl-L-methionine might also be expected to occur, and a peak corresponding to L-methionine should be visualised on the HPLC profile. L-methionine would not be expected to be consumed rapidly by resting cells, in the presence of a highly toxic concentration of cobalt. As this peak at 15 minutes only appeared when the 3P2 variant activity was assayed, a possible explanation could be the introduction in 3P2 of a secondary hydrolytic activity towards *N*-acetyl-D-methionine.

A more detailed time course experiment was then carried out to compare the racemase activity of the wild type TS-1-60 enzyme and the 3P2 variant. The racemisation was followed over 24 hours to assess the respective NAAAR activities. As shown in Figures 45 to 51, racemisation of *N*-acetyl-D-methionine was seen to occur much more rapidly with the 3P2 variant than the wild type enzyme. The decrease in concentration of *N*-acetyl-D-methionine is symbolised by "⇩", while the increase in concentration of *N*-acetyl-L-methionine is symbolised by "⇧".

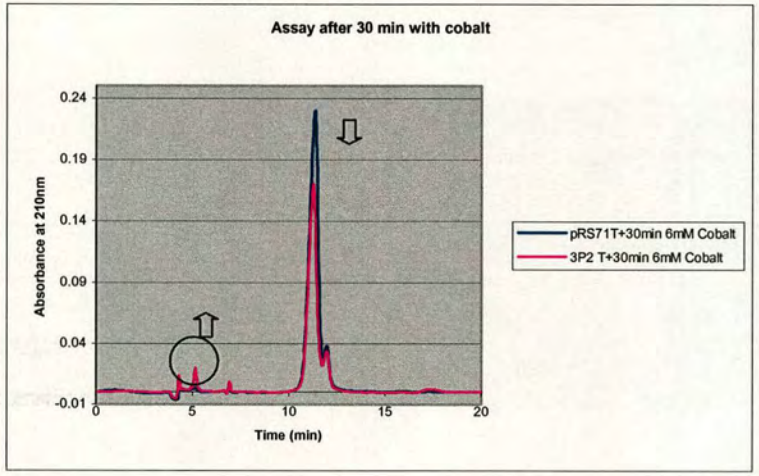


Figure 45

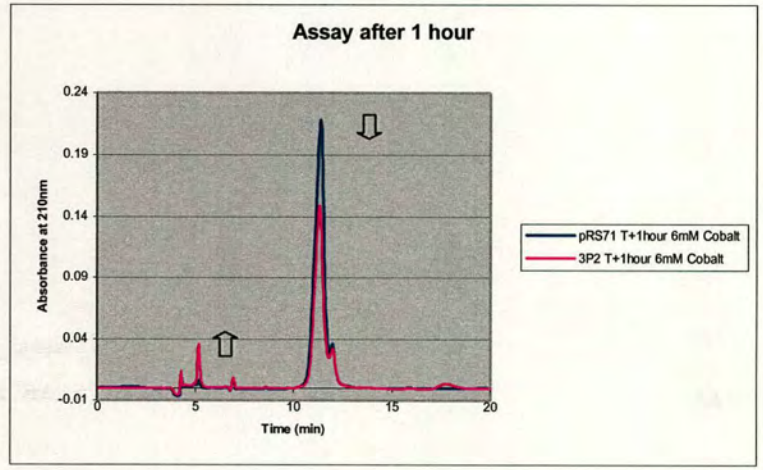


Figure 46

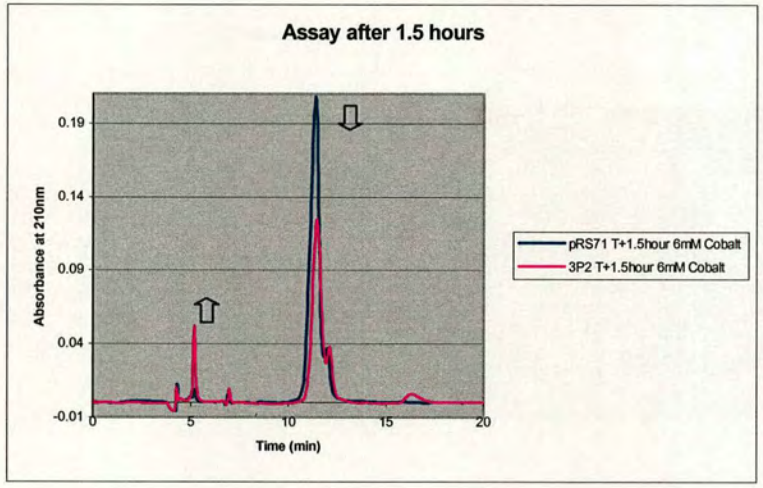


Figure 47

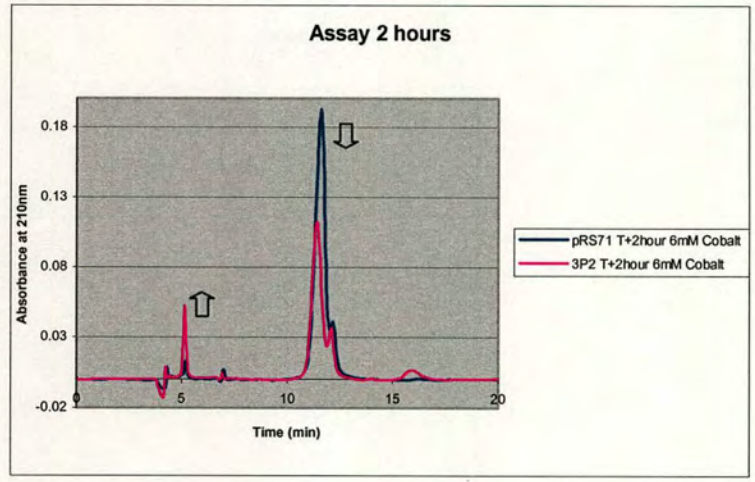


Figure 48

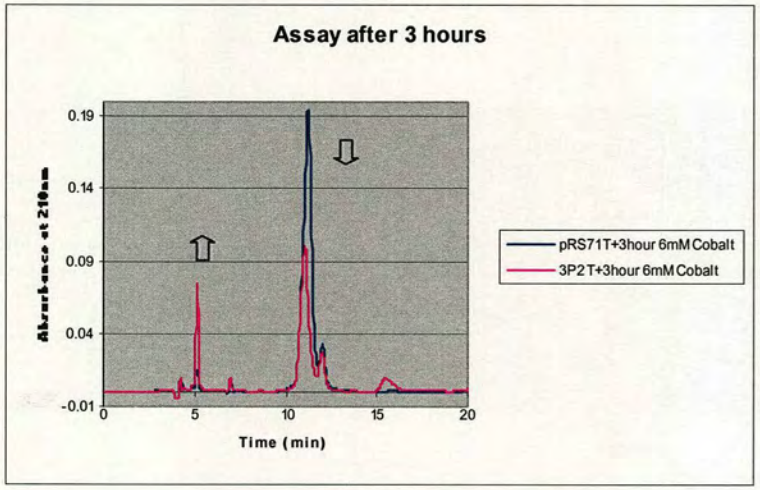


Figure 49

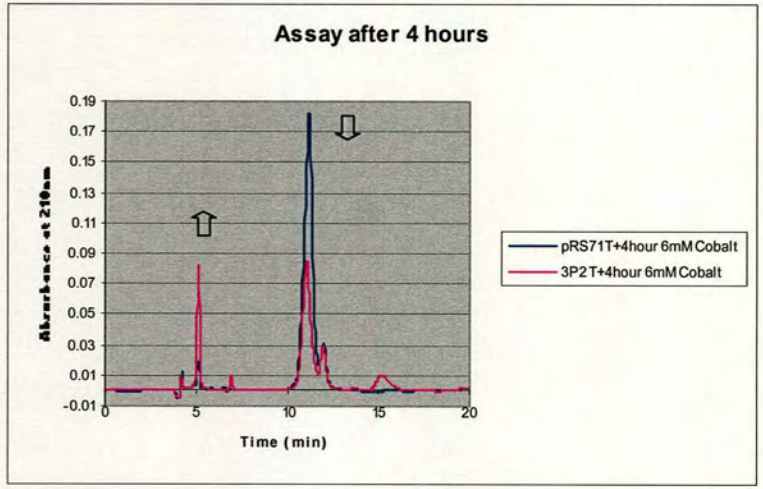


Figure 50

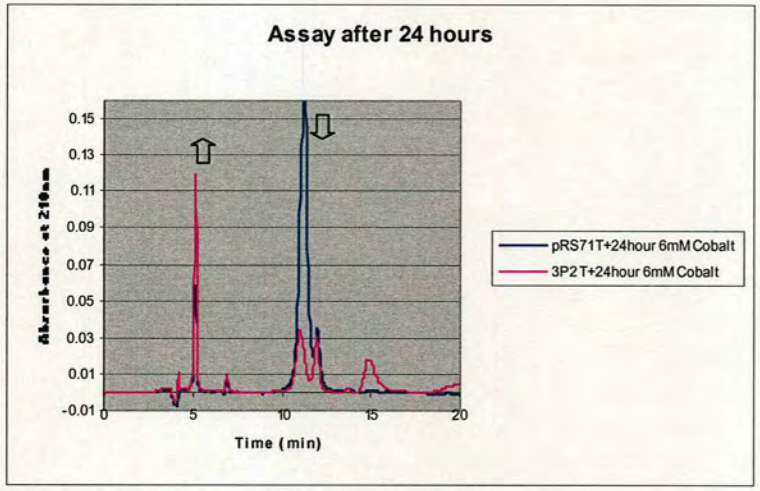


Figure 51

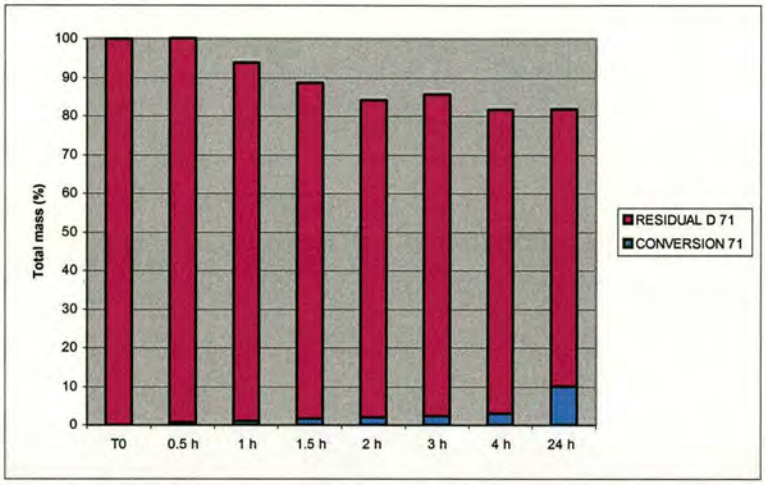


Figure 52: Mass balance of pRS71 racemisation assay

These data are summarised in Figures 52 to 55. The mass balance in the assay using the wild-type enzyme (Figure 52) showed an overall decrease of *N*-acetyl-methionine over time (loss of almost 20%), whereas ideally it should remain constant and close to the initial mass (T0). This suggested either some *N*-acetyl-D-methionine or some *N*-acetyl-L-methionine (or both) is degraded during the process, and cannot be visualised.

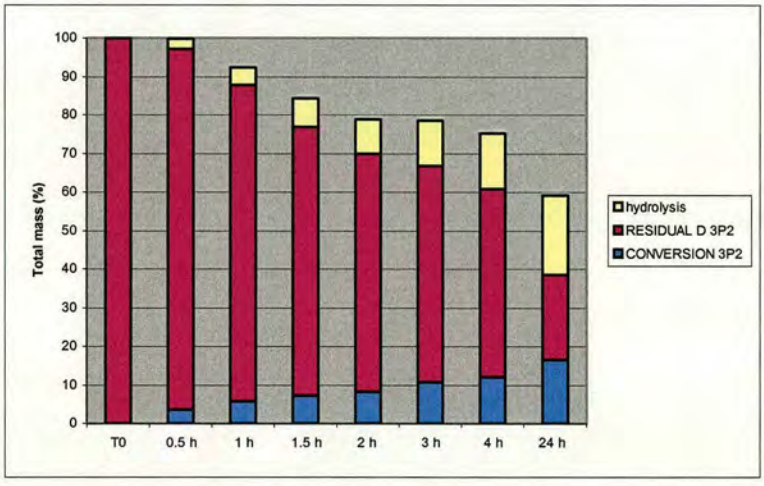


Figure 53: Mass balance of 3P2 racemisation assay

The same observation was made with the mass balance of the assay using the variant enzyme (Figure 53), but the loss of material was doubled (40%), and the appearance

of a new peak (having the same retention time as D-methionine) suggested that in addition to the unexplained loss of the *N*-acetyl-D or/and L- methionine, some of the *N*-acetyl-D-methionine was hydrolysed as well. Since the only known difference between the two strains examined, is the single point mutation in the 3P2 variant, this suggested the variant enzyme might have evolved a new activity, a D-aminoacylase. If so, then the mutation G293E increased the rate of the racemisation, but also introduced an other catalytic activity. Following this observation, an extensive literature review[134-136] revealed several articles which report aspartic acid to be closely involved in the catalytic mechanism of D-aminoacylase. Glutamic acid is also mentioned where it probably acts as a general base to activate the nucleophile water. Abstracting the proton from the water molecule, and the tightly bound zinc ion polarizes the carbonyl-oxygen bond, thus facilitating the nucleophilic attack on the amide carbon atom to form the tetrahedral intermediate. Cleavage of the carbon-nitrogen bond is assisted by the simultaneous protonation of the amide nitrogen. Similarly, in the case of the variant enzyme, the novel glutamic acid could play the same role. Interestingly, the concentration of the *N*-acetyl-L-methionine formed seemed to increase in the same proportion as that of D-methionine. This might reflect a competition between two reactions catalysed by the variant enzyme: racemisation and hydrolysis. However, even with the loss of some starting material, the concentration of *N*-acetyl-L-methionine formed by the mutant enzyme remained far more significant than that of the wild-type enzyme.

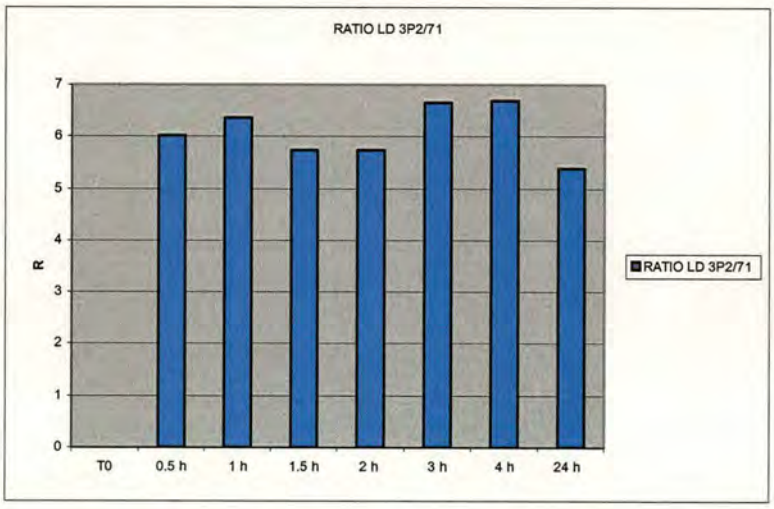


Figure 54: Comparison of the rate of racemisation between the variant and the wild type enzyme (R=ratio of the products from each reaction over time deduced from HPLC, considering the residual *N*-acetyl-D-methionine as reference)

Additionally, the comparison of the rate of racemisation as shown in Figure 54 consistently showed a 6-fold increase in the rate of racemisation by 3P2 over the wild type TS-1-60 NAAAR enzyme throughout the experiment. Interestingly, this result showed a remarkable similarity with the solid phase assay (plate assay). With the reference being the rate of colony growth on L-methionine, to obtain a similar growth rate with the wild type enzyme, a concentration of *N*-acetyl-D-methionine of 2 mM was required; whereas the variant enzyme showed the same growth for a concentration 7.63 times lower (0.262 mM- section 2.5). The solid phase assay could therefore possibly offer more quantitative information than expected about the improvement of the variant.

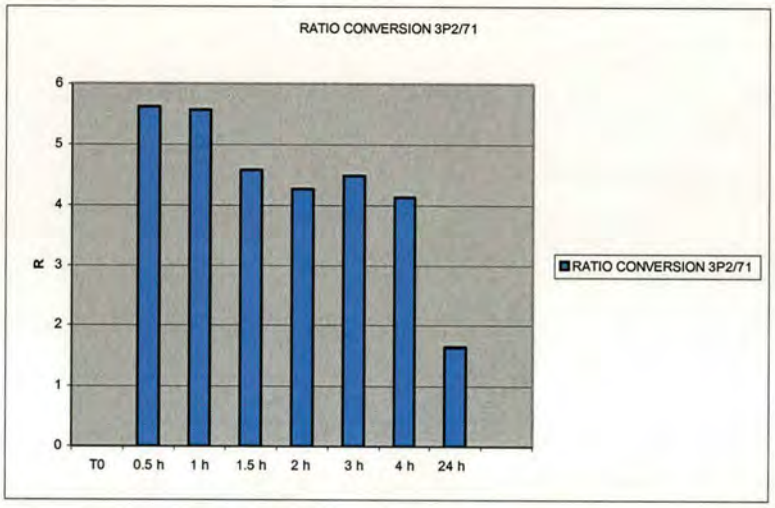


Figure 55: Comparison of the rate of racemisation between the variant and the wild type enzyme (R=ratio of the products from each reaction over time deduced from HPLC, considering the initial concentration of *N*-acetyl-D-methionine as reference)

However, if the comparison of the rate of racemisation takes into consideration the initial concentration of *N*-acetyl-D-methionine (Figure 55), the rate is no longer constant but decreases throughout the experiment. This might reflect the depletion of the substrate more important for the variant experiment: the rate of racemisation of the variant slows down as the concentrations of *N*-acetyl-L- and D-methionine equilibrate.

These experiments were repeated to assess the racemase activity of pRS7-1 and 3P2 variant using either *N*-acetyl-D-methionine or *N*-acetyl-L-methionine as starting material. The results are shown in Figures 56 and 57.

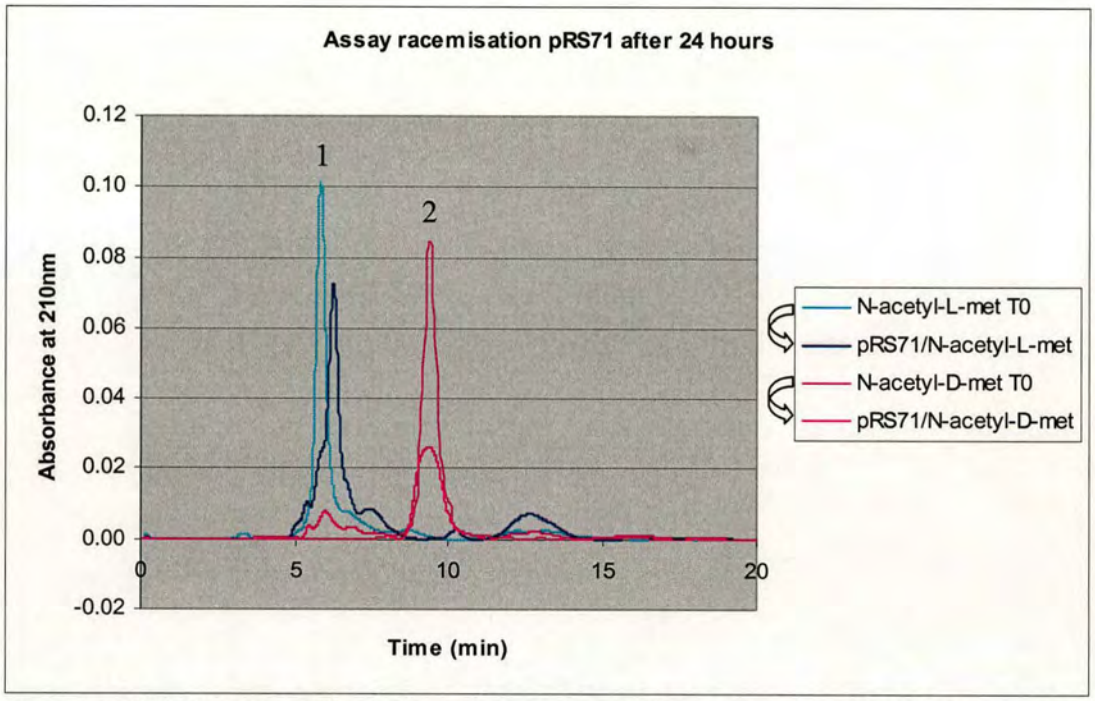


Figure 56: HPLC profile of the *in vivo* reaction using pRS71 expressed in SET2-1 assayed with 6mM cobalt with either *N*-acetyl-L- (1) or *N*-acetyl-D-met (2) as starting material

The wild type enzyme showed as expected[70, 80] a substrate preference towards *N*-acetyl-L-methionine. Conversion from *N*-acetyl-D-methionine to *N*-acetyl-L-methionine was preferred. As previously no peaks corresponding to the D-, or L-methionine were observed.

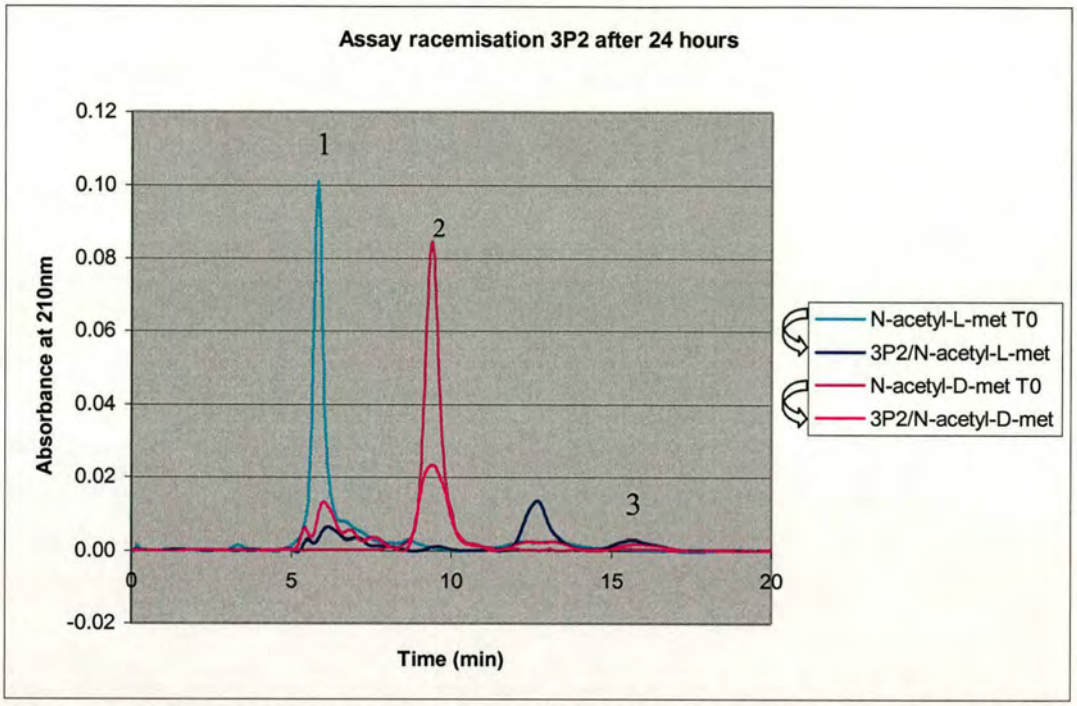


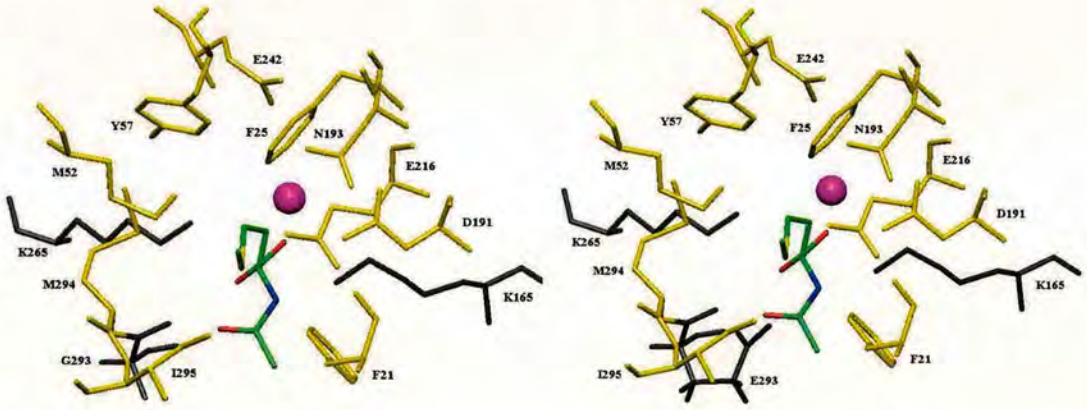
Figure 57: HPLC profile of the *in vivo* reaction using 3P2 expressed in SET2-1 assayed with 6mM cobalt with either *N*-acetyl-L- (1) or *N*-acetyl-D-met (2) as starting material

The 3P2 variant showed the same enantiopreference towards *N*-acetyl-L-methionine. However, whatever starting material was used, a peak corresponding to D-methionine (3) appeared (15 minutes), suggesting again the mutation carried by the 3P2 variant introduced a second catalytic activity.

However, in another experiment, the additional peak observed was not D-methionine, but L-methionine. Therefore these results have to be interpretated with extreme caution until further purification and full kinetic characterisation of the NAAAR activity of the 3P2 variant.

2.7. Hypothesis about the role of the mutation

High resolution crystal structures of NAAAR enzymes from *Amycolatopsis. species* TS 1-60, used in this study, and *Deinoccocus radiodurans* are now available[79, 80] and the mechanism of NAAAR activity is now well documented. From these analyses, the framework K165-D185-E216-D241-K265 has been determined to catalyse 1,1-proton exchange of *N*-acylamino acids in this enzyme. The enzyme is constituted by four subunits enclosing the substrate (Figure 14): catalytic site (C), metal-binding site (M), side-chain-binding site (S) and a flexible lid region (L). The NAAAR activity is attributed to the location of two lysine residues (K165, K265) in sufficiently close proximity to facilitate proton abstraction or donation from either face of N-acylamino acid substrates. A modelling study of the G293E mutation of 3P2 variant shows its location to be in close proximity to the substrate and the terminal amino group of each lysine residue (Figure 58).



Wild type **3P2 variant**

Figure 58: Active site residues of TS-1-60

Lysine residues at positions 165 and 265 which participate in *N*-acylamino acid racemisation are indicated in black. Glycine 293 which is mutated to glutamic acid in 3P2 is also indicated in black

The software used for the modelling was Insight II from Accelrys (http://www.accelrys.com), and the initial crystal structures used were available on the NCBI website (http://www.ncbi.nlm.nih.gov).

The mutation (**G293E**) is located in the S region, which is a hightly conserved region between NAAAR sequences (See Appendix 11.5.), introducing a long acidic chain, which could act either:

1) As a proton donor to the one of the lysine catalyst (K165 or K265) and then abstracting the α-proton of the carboxylate. The negatively charged intermediate is stabilised by the bound Co^{2+} ion, and then the second lysine catalyst donates an α-proton to yield the racemised product.

2) As a proton donor to the intermediate formed by the abstraction of the α-proton of the carboxylate. (Regenerated by one of the lysines).

These two hypotheses are very unlikely as the pH is neutral or slightly basic (*in vivo* reaction), so the glutamic acid is as its deprotonated form.

3) By abstracting the α-proton of the carboxylate

4) As a proton relay between the 2 lysines. This hypothesis could explain the observed improvement of the NAAAR activity.

5) As a general base during catalysis, abstracting a proton from a water molecule. The amide bond then undergoes a nucleophilic attack to form the tetrahedral intermediate. Cleavage of the carbon-nitrogen bond is assisted by the simultaneous protonation of the amide nitrogen. This hypothesis could explain the formation of the deacetylated amino acid.

Although based only on a model at present, this active site location would be consistent with the mutation exerting an influence on substrate proton abstraction, as shown Figure 59. A high-resolution crystal structure of the 3P2 variant would confirm the position of E293 and greatly assist the interpretation of the effect of this mutation upon catalysis.

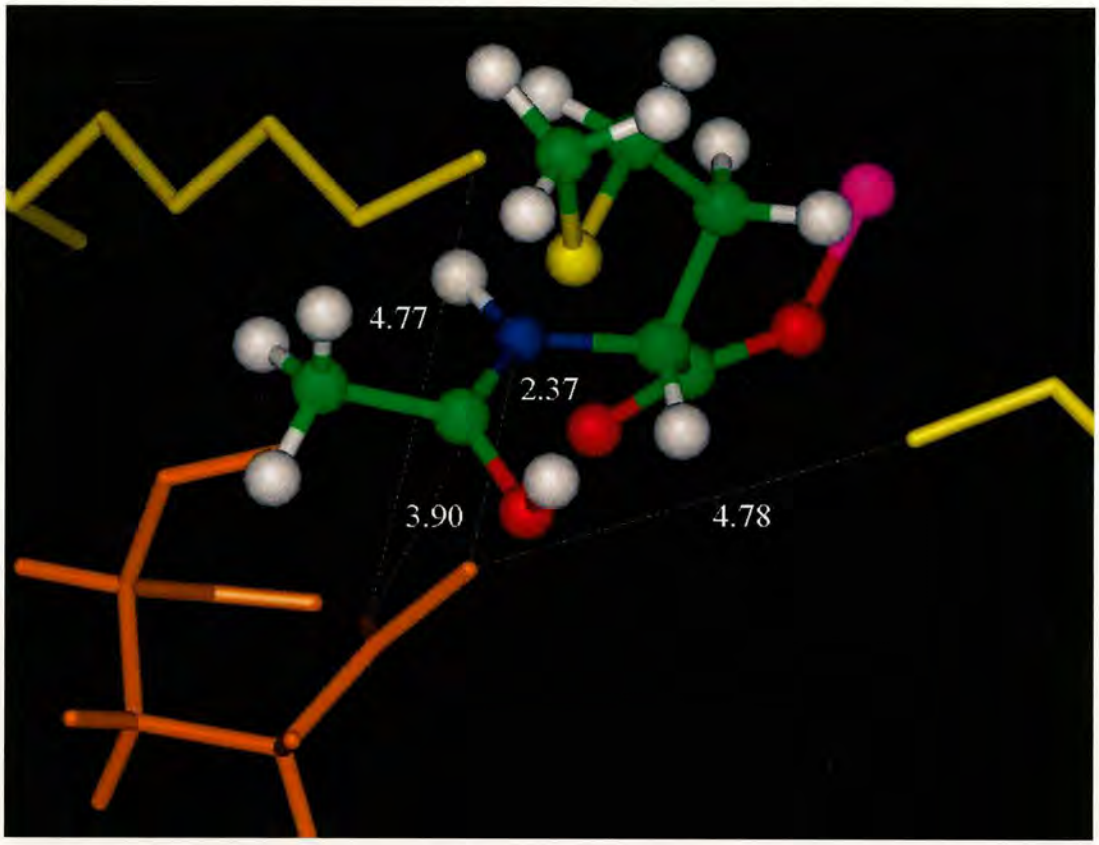


Figure 59: Model of the active site of 3P2 variant indicating the estimated distances (in Å) between substate and the different residues involved in the NAAAR activity

2.8. Conclusion

In conclusion, this study attempted to develop a new high throughput *in vivo* screening method to identify *N*-acylamino acid racemase activity in enzymes of the enolase superfamily. In addition the bacterial strain used in the screen was employed as a host in which to select random mutations which increase the NAAAR activity in specific enzymes. The principle objectives were achieved, and showed this to be an original approach to the high throughput concept, complemented with a more quantitative *in vitro* assay using HPLC. The choice to select the NAAAR from *Amycolatopsis* TS-1-60 was based upon the fact the enzyme already possessed a low level of NAAAR activity but also some properties, which were significant limiting for a useful biocatalyst, such as relatively low turnover, and substrate mediated inhibition at concentrations exceeding 50mM. The *in vivo* selective method was an

excellent way to explore this very interesting *N*-acylamino acid racemase activity reported in the enolase superfamily. The known NAAAR was used to determine its effect in restoring growth to the *E. coli* mutant and therefore validation of this selection.

Furthermore the screening of libraries of mutants derived from the wild type enzyme unveiled a mutant strain with an improved activity, which further demonstrated the value of the *in vivo* approach, as it is possible to screen easily a library of 10^6-10^8 mutants per assay. Another interesting observation about the selection was made on the correlation between substrate inhibition of the two racemases and the concentration at which the colonies appeared to decrease in number. The selection may be suitable for gathering additional useful direct information about the characteristics of the enzymes such as substrate inhibition and kinetics (general improvement of the rate of catalysis) without having to purify the enzymes. Additionally, the mutation made in the wild type enzyme of TS-1-60 is a good basis for future studies (saturation mutagenesis, purification, kinetics, additional rounds of laboratory evolution) and eventual detailed investigations of the hypotheses made about the effect of the mutation upon enzyme function (substrate inhibition, general improvement of the activity).

Additionally, the complementarity of this method and methods of rational mutagenesis proved to be useful. The work of Gerlt, Babbitt, and Wang on NAAAR have helped to correlate the structure of this enzyme (and others of the enolase superfamily) to its function. This has provided a basis to locate the 3P2 mutation and to predict its effect upon the enzyme. Similarly, the method developed in this work permits the identification of mutations which might never be considered in structure/function and rational mutagenesis approaches. The combination of the two strategies is therefore synergistic and extremely powerful for future application.

2.9. Future work

The 3P2 variant identified in this work should be fully characterised in future work to support the preliminary *in vitro* racemisation data. To conduct detailed kinetic analysis the enzyme should be first isolated and purified to homogeneity. This can perhaps best be achieved using a well developed procedure such as poly-histidine tagging (Invitrogen). This work has been initiated by sub-cloning the genes encoding the wild type TS-1-60 NAAAR and the 3P2 variant into the vector pET16b to construct in-frame poly-histidine fusions. The resulting constructs will next be characterised. The 3P2 variant will then be purified to homogeneity by Ni-chelation chromatography for full kinetic characterisation of its NAAAR activity. In parallel it is now appropriate to conduct additional random mutation upon the 3P2 variant and saturation mutagenesis at the G293E position to examine the effect of alternate amino acid substitutions at this location. This may require re-calibration of the selection using lower levels of expression of the 3P2 variant. In addition it would be extremely interesting to introduce the G293E mutation to the corresponding position (also a glycine) in the NAAAR of *A. orientalis* and other close homologues of the TS-1-60 NAAAR.

The variant unveiled by the selection may also have acquired a degree of acylase activity with high enantioseletivity towards *N*-acetyl-D-methionine. To further test this possibility, particularly with purified enzyme, the racemisation should be monitored in separate experiments which use either *N*-acetyl-D-methionine and *N*-acetyl-L-methionine as starting material. This will provide conclusive evidence for any acylase activity and could reveal the variant to be very useful industrially, as the substrate (*N*-acetyl-L-methionine) might be racemised and then hydrolysed immediately, yielding 100% D-methionine from *N*-acetyl-L-methionine without the assistance of any other enzyme- the ultimate goal of an efficient dynamic kinetic resolution. This could be the first step towards a new way of rational design of biocatalysts, which to date has generally been based on improving the activity of existing enzymes or shifting the activity towards another existing activity in the same

superfamily. The fact that it might be possible to introduce a novel activity in an enzyme which was not originally designed for such an activity appears to be very exciting, as the design may focus on how the reaction is catalysed (and the residues which are involved in a particular reaction), and may consider the superfamily aspect (general characteristics/aspects) rather than individual specificity.

Furthermore, following the successful construction of this selection, the screening of potential novel racemases could proceed and the deletion method could be applied to additional amino acids of academic and commercial interest, such as phenylalanine, leucine, and tyrosine auxotrophs, to enable the testing of this selection with alternate substrates.

Genes from various sources could also be studied. The first source could be NAAAR homologues or libraries of fragments from bacterial strains including: *Rhodoccocus*, *Streptomycetes* and *Amycolatopsis*. Introducing libraries containing chromosomal fragments of these organisms into selective strains would test the frequency of NAAAR occurrence. These homologues or fragments could be either in native forms or following mutagenesis using mutator or error-prone PCR. The homologue approach could make use of many hundreds of individual NAAAR homologues now identifiable in public genomic databases. This point illustrates how microbial genomics has provided a wealth of information and accessible genetic material for biocatalyst development. The last source could use genes encoding known enolase superfamily members: OSBS from *Bacillus subtilis*, enolase from *E. coli*, cycloisomerase from *Rhodococcus erythropolis*, racemase from *Pseudomonas putidas* and potentially others.

The approach could be extended to the selection for additional superfamily enzyme activities. Deletion mutations could also be planned in the *E. coli* OSBS and enolase genes. These would be used to select for these activities also in randomly mutated enolase superfamily enzymes and other gene sources such as libraries.
Mutations in the *E. coli* menC gene, which encodes OSBS, have been used previously to define the native activity of the TS-1-60 NAAAR. Therefore a deletion

in the gene may be more broadly applicable to identify this activity through mutation of related enzymes of the superfamily. Similarly the gene encoding the *E. coli* enolase could be deleted. Complementation of this essential function could also be explored to identify mutations in related enzymes that lead to acquisition of enolase activity.

In this way, it is hoped that some insight into the potential interconversion of these closely related enzyme activities can be gained and the potential to evolve novel activities also explored. Ideally some of these experiments will result in valuable novel biocatalysts. The methods described in this work, for example the gene deletion system, may also be developed into very general methods potentially useful for biocatalyst generation through selection. Finally this work illustrates that principles that date from very early work in strain development can still be applied in novel and powerful applications by taking advantage of recent advances in methodology and informatics.

3. Experimental: Stock Recipes (Maniatis)

Chemicals and reagents for buffers, assay mixtures and cultures medium were used as supplied from commercial sources, and were standard laboratory grade.

3.1. LB medium

25 g/L of LB Broth (Fisher Chemicals) or prepared as followed per litre:

Sodium chloride	10 g
Triptone	10 g
Yeast extract	5 g

3.2. LB agar medium

LB medium	300 ml
Agar	4.5 g

3.3. R-top agar

Tryptone	10 g
NaCl	8 g
Yeast extract	1 g
Agar	8 g
Water	to a final volume of 1 litre
After autoclaving:	
$CaCl_2$ 1M	2 ml
20% glucose	5 ml

3.4. Medium for R plates

Tryptone	10 g
NaCl	8 g
Yeast extract	1 g
Agar	12 g
Water	to a final volume of 1 litre
After autoclaving:	
$CaCl_2$ 1M	2 ml
20% glucose	5 ml

3.5. 10x M9 salts

$Na_2HPO_4.7H_2O$	128 g
KH_2PO_4	30 g
NaCl	5 g
NH_4Cl	10 g
Water	to a final volume of 1 litre

3.6. 50x TAE

Tris base	242 g
Glacial acetic acid	57.1 ml
EDTA 0.5M (pH 8.0)	100 ml
Water	to a final volume of 1 litre

3.7. MC buffer

$MgSO_4$ 0.1M

$CaCl_2$ 0.005M

4. Experimental: cloning

4.1. General techniques

E. coli strains were obtained from *E. coli* Genetic Stock Center (Yale University, New Haven, USA). *E. coli* XL1-Red and XL1-Blue competent cells were purchased from Stratagene (La Jolla, USA). *E. coli* Top10 were purchased from Invitrogen (Paisley, UK).

Expression vector pTTQ18 was provided by Dr Michael Stark (University of Dundee, UK). pET16b was purchased from Novagen (Nottingham, UK).

Oligonucleotides were purchased fom Interactiva (Ulm, Germany) and ThermoElectron (Ulm, Germany).

Diversify™ PCR random mutagenesis was purchased from BD Clontech UK (Basingstoke, UK).

Restriction enzymes and T4 ligase were purchased from New England Biolabs (Hitchin, UK). PuReTaq™ Ready-to-go™ PCR beads were purchased from Amersham Biosciences (Little Chalfont,UK).

Purification kit for plasmids, genomic DNA, PCR and gel extraction were purchased from Qiagen (Hilden, Germany). E-Gels were from Invitrogen.

Readygels (Precasted 10-20% Tris-HCl acrylamide gels) were purchased from Bio-Rad and run either on a BioRad Powerpac 200 or Powerpac 1000. EZBlue™ gel staining solution was purchased from Sigma.

PCR experiments were carried out on a Stratagene RoboCycler Gradient 96 or on an Eppendorf Mastercycler® personal (Eppendorf AG, Hamburg, Germany). Electroporation were performed using a Bio-Rad Gene Pulser II Electroporation System (Hemel Hemperstead, UK). Centrifugation were performed either with a Eppendorf 5415C or 5415R microcentrifuge or Heraeus Biofuge pico for small scale or with a Sorvall® RC 5C Plus or RC 26 Plus centrifuge (Kendro Laboratory Products Limited (Bishop's Stortford, UK) for larger scale.

OD and DNA concentration were analysed using an Eppendorf BioPhotometer.

CHIROBIOTIC T™ chiral column (Catalogue number: 12024 – 5μm 250x4.6mm) was purchased from Astec (Advanced Separative Technologies Ltd, Congleton, UK).

4.2. Construction of plasmids

4.2.1. With NAAARs from TS-1-60 and *Amycolatopsis orientalis*

4.2.1.1. Gene amplification from TS-1-60 by PCR

The gene (1105 bp) was amplified from the chromosomal DNA using PCR. The chromosomal DNA was purified using QIAGEN Genomic DNA extraction kit yielding a concentration of DNA of 450 μg/ml. Two nucleotides were designed according to the termini of the coding region, and containing unique restriction sites for subcloning into the expression vector pTTQ18 and utilised in the reaction.

The N-terminus EcoRI primer contained a unique EcoRI restriction site indicated in bold:
5'-GAC**GAATTC**GAAACTCAGCGGTGTGGAACTGCGCCGGG-3'

The C-terminus SphI primer contained a unique SphI restriction site indicated in bold:
5'-GAC**GCATGC**CTACGAACCGATCCACACCTTTGCCGTGGTC-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead with Taq DNA polymerase. One bead was mixed with 0.6 μl of each primer (from a 1 pmol/μl primer solution), 1 μl of chromosomal DNA template and 22.8 μl of water.

The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 98°C	5 mins
Denaturation	2) 95°C	1 mins
Annealing	3) 64°C	45 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	3 mins

The PCR product mixed with loading buffer was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The corresponding band (1116 bp) was purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer (elution buffer).

4.2.1.2. Gene amplification from *A. orientalis* by PCR

The gene (1105 bp) was amplified from the chromosomal DNA using PCR. The chromosomal DNA was purified using QIAGEN Genomic DNA extraction kit yielding a concentration of DNA of 470 µg/ml. Two nucleotides were designed according to the termini of the coding region, and containing unique restriction sites for subcloning into the expression vector pTTQ18 and utilised in the reaction.

The N-terminus EcoRI primer contained a unique EcoRI restriction site indicated in bold:

5'-GAC**GAATTC**GAAACTCAGCGGTGTGGAACTGCGCCGGG-3'

The C-terminus SphI primer contained a unique SphI restriction site indicated in bold:

5'-GAC**GCATGC**CTACGAACCGATCCACGCTTTCTCCGTGGTG-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead. One bead was mixed with 0.6 µl of each primer (from a 1 pmol/µl primer solution), 1 µl of chromosomal DNA template and 22.8 µl of water.

The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 98°C	5 mins
Denaturation	2) 95°C	1 mins
Annealing	3) 64°C	45 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	3 mins

The PCR product mixed with loading buffer was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The corresponding band (1116 bp) was purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer.

4.2.1.3. Construction of the plasmids pRS7-1 and pRS8-9

The amplified DNA fragments and the vector pTTQ18 were doubled digested with EcoRI and SphI. The reaction mixture were prepared as followed:

17 µl of PCR product or pTTQ18 vector DNA

2 µl of EcoRI NEBuffer

0.5 µl EcoRI

0.5 µl SphI

The correct fragment (4514 bp) issued from the doubled digestion of the vector was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The DNA was purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer (elution buffer).

The digested PCR products were purified using QIAquick PCR purification kit, and eluted in 30 µl of EB buffer.

The ligation of the PCR product and the vector (ratio concentration vector/insert=1/5) was performed using T4 DNA ligase following the supplier's

protocol. The reaction was incubated at room temperature for 30mins-1hour. The ligation was then purified with QIAquick PCR purification kit. The construct was used to transform *E. coli* Top10 electrocompetent cells. The electrocompetent cells were thawed on ice and mixed with 10 to 25 ng of construct, and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 μF and a resistance of 200 Ω. The cells were immediately resuspended in 850 μl of SOC after transformation and incubated at 37°C under agitation for 1.5 hours.

Colonies were screened on LB Agar Amp^{200}. Plasmid DNA was purified with a QIAprep spin miniprep kit, and then doubled digested with EcoRI and SphI. The fragments (4514 and 1116 bp) issued from the doubled digestion were identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain.

Additional digestions were performed on the plasmid to verify the presence of various restriction sites.

4.2.1.4. Expression of NAAAR gene in pRS7-1

Two 500 ml glass conical flasks containing 50 ml of LB Amp^{200} were inoculated with 50 μl of overnight culture. IPTG was added to one culture to a final concentration of 1mM, while the second culture served as a control. Growth under agitation at 37°C was followed by measuring OD_{600}. A 1ml sample was spun down for each measure, the pellet resuspended into 100 μl of SDS-reducing loading buffer and heated for 1 min at 100°C, and spun down to remove all debris. Samples were loaded in an acrylamide gel (10-20% Tris-HCl), and run at a constant voltage of 200 V. The gel was stained using staining reagent (EZBlue™ from Sigma) and agitated on a rocking platform until the gel was dark brown or blue. The gel was then washed with water.

4.2.2. OSBS

4.2.2.1. From *E. coli*

The *men*C gene encoding the OSBS from *E. coli* (963 bp) was amplified by PCR from a single colony. Two oligonucleotides, corresponding to the termini of the coding region, were designed to contain unique restriction sites for subcloning into pTTQ18 and utilised in the reaction.

The N-terminus EcoRI primer contained a unique EcoRI restriction site indicated in bold:

5'-GAC**GAATTC**GCGTAGCGCGCAGGTATAC-3'

The C-terminus SphI primer contained a unique SphI restriction site indicated in bold:

5'-GAC**GCATGC**TCATAACAACCGCTCCAGTGC-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead. One bead was mixed with 0.6 µl of each primer (from a 1 pmol/µl primer solution), 23.8 µl of water and a single colony taken from a fresh plate.

The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 96°C	5 mins
Denaturation	2) 95°C	40 s
Annealing	3) 50°C	40 s
Extension	4) 72°C	1 min
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	5 mins

The PCR product mixed with loading buffer was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The corresponding band was purified using a Qiagen gel extraction kit, and eluted in 30 µl of EB buffer.

The PCR product and the vector pTTQ18 were doubled digested with EcoRI and SphI. The reaction mixture were prepared as followed:

17 µl of PCR product or pTTQ18 vector DNA

2 µl of EcoRI NEBuffer

0.5 µl EcoRI

0.5 µl SphI

The correct fragment (4514 bp) issued from the doubled digestion of the vector was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The DNA was purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer.

The digested PCR product was purified using QIAquick PCR purification kit, and eluted in 30 µl of EB buffer.

The ligation of the PCR product and the vector (ratio concentration vector/insert=1/5) was performed using T4 DNA ligase following the supplier's protocol. The reaction was incubated at room temperature for 30 mins-1 hour. The ligation was then purified with QIAquick PCR purification kit. The construct was used to transform *E. coli* Top10 electrocompetent cells. The electrocompetent cells were thawed on ice and mixed with 10 to 25 ng of construct, and then transferred to a prechilled 1mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 µF and a resistance of 200 Ω. The cells were immediately resuspended in 850 µl of SOC after transformation and incubated at 37°C under agitation for 1.5 hours.

Colonies were screened on LB Agar Amp^{200}. Plasmid DNA was purified with a QIAprep spin miniprep kit, and then doubled digested with EcoRI and SphI. The fragments (4514 and 971 bp) issued from the doubled digestion were identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain.

Additional digestions were performed on the plasmid to map it and verify the presence of various restriction sites.

4.2.2.2. From *Bacillus subtilis* 168

The *menC* gene encoding the OSBS from Bacillus subtilis (1113 bp) was amplified by PCR from the chromosomal DNA. The chromosomal DNA was purified using QIAGEN Genomic DNA extraction kit yielding a concentration of DNA of 350 μg/ml. Two oligonucleotides, corresponding to the termini of the coding region, were designed to contain unique restriction sites for subcloning into pTTQ18 and utilised in the reaction.

The N-terminus SacI primer contained a unique SacI restriction site indicated in bold:

5'-GAC**GAGCTC**TATCGAAATAGAAAAAATCACGC-3'

The C-terminus SphI primer contained a unique SphI restriction site indicated in bold:

5'-GTC**GCATGC**TTAACCATGCTGTGTGAAAACATC-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead. One bead was mixed with 0.6 μl of each primer (from a 1 pmol/μl primer solution), 1 μl of chromosomal DNA template and 22.8 μl of water.

The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 96°C	5 mins
Denaturation	2) 95°C	40 s
Annealing	3) 50°C	40 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	5 mins

The PCR product mixed with loading buffer was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The corresponding band was purified using a Qiagen gel extraction kit, and eluted in 30 μl of EB buffer.

The PCR product and the vector pTTQ18 were doubled digested with SacI and SphI. The reaction mixture were prepared as followed:

17 µl of PCR product or pTTQ18 vector DNA
2 µl of NEBuffer 1
0.5 µl SacI
0.5 µl SphI

The correct fragment (4524 bp) issued from the doubled digestion of the vector was identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain. The DNA was purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer.

The digested PCR product was purified using QIAquick PCR purification kit, and eluted in 30 µl of EB buffer.

The ligation of the PCR product and the vector (ratio concentration vector/insert=1/5) was performed using T4 DNA ligase following the supplier's protocol. The reaction was incubated at room temperature for 30 mins-1 hour. The ligation was then purified with QIAquick PCR purification kit. The construct was used to transform *E. coli* Top10 electrocompetent cells. The electrocompetent cells were thawed on ice and mixed with 10 to 25 ng of construct, and then transferred to a prechilled 1mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 µF and a resistance of 200 Ω. The cells were immediately resuspended in 850 µl of SOC after transformation and incubated at 37°C under agitation for 1.5 hours.

Colonies were screened on LB Agar Amp^{200}. Plasmid DNA was purified with a QIAprep spin miniprep kit, and then doubled digested with SacI and SphI. The fragments (4524 and 1120 bp) issued from the doubled digestion were identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain.

Additional digestions were performed on the plasmid to map it and verify the presence of various restriction sites.

5. Experimental: directed evolution

5.1. Using error-prone PCR

The random mutagenesis was performed following the manufacturer's protocol (Clontech) using the same primers used to clone the gene into pTTQ18. Nine reactions were prepared, and the thermal cycling was performed using the following parameters:

Initial denaturation and activation	1) 98°C	5 mins
Denaturation	2) 95°C	1 mins
Annealing	3) 64°C	45 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	3 mins

The PCR products mixed with loading buffer were identified using TAE agarose gel electrophoresis by UV visualisation of ethidium bromide stain.

The corresponding nine bands were cut from the gel and mixed as followed: Reactions 1 to 3, 4 to 6 and 7 to 9 mixed together to give 3 different samples. The three samples were purified using a QIAquick gel extraction kit, and eluted in 30 µl of buffer.

These three samples the vector were then doubled digested with EcoRI and SphI. Fragment from each sample were ligated into expression vector pTTQ18. The ligation was performed using T4 DNA ligase following the supplier's protocol. The reaction was incubated at room temperature for 30 mins. The ligation was then purified with QIAquick PCR purification kit.

These three libraries were named pRS7-1/13, pRS7-1/46 and pRS7-1/79.

5.2. Using mutator strain

Two methods were used to produce libraries of mutated plasmids using *E. coli* XL1-Red. These methods differed from each other in the growth conditions: cultured in suspension in liquid medium or cultured on plates.

5.2.1. First method to create libraries of mutants: growth on solid medium

The plasmid was used to transform the mutator strain using the manufacturer's protocol. Then the method consisted in spreading on the transformed cells on LB Agar Amp^{200}.

The plates were incubated at 37°C for 24-30 hours, and then the cells recovered by resuspending the colonies into 10 ml of LB. As *E. coli* XL1-Red has a doubling time of 90-120 mins, approximately 15 generations of cells were produced (first cycle of mutation). The plasmids were purified (library pRS7-1/1P), while 150 µl of resuspended cells was spread on each of five freshly prepared LB Agar Amp^{200} plates. The plates were incubated at 37°C for 24-30 hours, and then the cells recovered by resuspending the colonies into 10 ml of LB (second cycle of mutation, 30 generations of cells). At the end of this cycle, the plasmids were isolated (library pRS7-1/2P) and used to transform fresh *E. coli* XL1-Red. The cells were plated on five freshly prepared LB Agar Amp^{200} plates, and incubated at 37°C for 24-30 hours. The cells were recovered by resuspending the colonies into 10 ml of LB (third cycle of mutation, 45 generations of cells). The plasmids were purified (library pRS7-1/3P), while 150 µl of resuspended cells was spread on each of five freshly prepared LB Agar Amp^{200} plates. The plates were incubated at 37°C for 24-30 hours, and then the cells recovered by resuspending the colonies into 10 ml of LB (fourth cycle of mutation, 60 generations of cells). The plasmids were isolated (library pRS7-1/4P) and used to transform fresh *E. coli* XL1-Red. The cells were plated on five freshly prepared LB Agar Amp^{200} plates, and incubated at 37°C for 24-30 hours. The cells were recovered by resuspending the colonies into 10 ml of LB (fifth cycle of

mutation, 75 generations of cells). The plasmids were purified and constituted the library pRS7-1/5P.

5.2.2. Second method to create libraries of mutants: growth in solution

The two plasmids were used to transform the mutator strain using the manufacturer's protocol. Then this method consisted in growing the transformed the cells in suspension in LB Amp^{200} under agitation (250 rpm) at 37°C.
The transformed cells were used to inoculate 20 ml of LB Amp^{200} and the culture was grown for 24-30 hours (first cycle of mutation). The plasmids were purified (library pRS7-1/1S), while 20 µl of the culture was used to inoculate LB Amp^{200}, and subjected to a second cycle of growth under agitation for 24-30 hours. The plasmids were then purified (pRS7-1/2S) and used to transform fresh *E. coli* XL1-Red. The transformed cells were used to inoculate 20 ml of LB Amp^{200} and the culture was grown for 24-30 hours (third cycle of mutation). The plasmids were purified (library pRS7-1/3S), while 20 µl of the culture was used to inoculate LB Amp^{200}, and subjected to a fourth cycle of growth under agitation for 24-30 hours. The plasmids were then purified (pRS7-1/4S) and used to transform fresh *E. coli* XL1-Red. The transformed cells were used to inoculate 20 ml of LB Amp^{200} and the culture was grown for 24-30 hours (fifth cycle of mutation). The plasmids were purified and the library called pRS7-1/5S).

6. Experimental: development of an in vivo selective method

6.1. Feeding assays on the strains: Growth on M9/glucose minimal medium plates

To 270 ml of an autoclaved mixture of water and 1.5% agar, the following volume of solutions were added successively:

$MgSO_4$ 1M	720 µl
$CaCl_2$ 1M	36 µl
20% glucose	7.2 ml
10x M9 salts	36 ml
Antibiotic if required:	300 µl of the stock solution
IPTG if required:	300 µl of 1M stock solution
Cobalt Chloride:	According to the concentration required
Supplement substrate:	3 ml of a 5 mg/ml stock solution

(Concentration of supplement substrate was adapted according the concentration needed)

When plates with different concentrations of substrate were required, 30 ml of the hot M9 agar medium was poured into sterile tubes, and the appropriate concentration of substrate added individually to each tube and then poured into a petri dish.

6.2. Creation of a methionine auxotroph strain

This method used is desribed by Datsenko & Wanner[120].

6.2.1. Deletion of *dadAX* gene

6.2.1.1. Creation of the PCR fragment to disrupt the *dadAX* loci

The *dadAX* loci (*dadA* 1660..2958 *dadX* 2968..4038) were first disrupted in *E. coli* BW25113. The following primers were used for amplification of the *cat* gene (chloramphenicol resistance-1015 bp) from pKD3:

N-terminus *dadA* primer contained the priming site of the *cat* gene indicated in bold and the homology extension of *dadA*:
5'-
AACCAGTGCCGCGAATGCCGGGCAAATCTCCCCCGGATATGCTGCACCGT
CATATGAATATCCTCCTTAG-3'

C-terminus *dadX* primer contained the priming site of the *cat* gene indicated in bold and the homology extension of *dadX*:
5'-
CCACTAAAACAGGGGTACCGGTTAAGGCGTGGCGCGGATAACCGTCGGC
G **TGTGTAGGCTGGAGCTGCTT**-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead with Taq DNA polymerase. One bead was mixed with 0.6 µl of each primer (from a 1 pmol/µl primer solution), 1 µl of plasmid DNA template (pKD3) and 22.8 µl of water.

The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 96°C	5 mins
Denaturation	2) 95°C	40 s
Annealing	3) 50°C	40 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	5 mins

The PCR product (1115 bp) was gel-purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer.

6.2.1.2. *dadAX* gene disruption

Transformants carrying the Red helper plasmid (pKD46) were grown in a 5 ml SOB culture with 5 µl Amp^{100} and 10 µl 0.5M L-arabinose at 30°C to an $OD_{600}\approx0.6$ and made electrocompetent by concentrating 100-fold and washing three times with ice cold sterile water. The electrocompetent cells were thawed on ice and mixed with 10 to 50 ng of PCR product, and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 µF and a resistance of 200 Ω. The shocked cells were immediately added to 1 ml of SOC and incubated 1 hour at 37°C under agitation. One half was then spread onto LB agar Cml^{30} to select Cml^{R} transformants. If none grew within 24 hours, the remainder was spread after standing at room temperature. After primary selection, mutants were maintained on LB agar without antibiotic. They were colony purified once nonselectively at 37°C and then tested for ampicillin sensitivity to test for the loss of the helper plasmid. If it was not lost, then a few were colony purified once at 43°C and similarly tested.

6.2.1.3. PCR verification

The following primers flanking the site of the *dadAX* operon were used to verify simultaneous loss of the native gene fragment and gain of the new mutant-specific fragment:

N-terminus checking primer for *dadAX* disruption:
5'-CCACTAATCCGCTTCTGAACG-3'

C-terminus checking primer for *dadAX* disruption:
5'-CTTGTTGTAAGCCGGATCGGAG-3'

6.2.2. Transduction of *dadAX* deletion created into BW25113 to the phenylalanine auxotroph strain CGSC7421 using phage P1

6.2.2.1. Preparation of lysates

5 ml of LB broth containing 5 mM $CaCl_2$ (250 µl of 0.1M $CaCl_2$ for 5 ml of LB broth) were inoculated with 5 µl of an overnight culture of ET1. The culture was aerated until the cells were growing exponentially and had reached a density of 2.10^8 cells/ml. $P1_{vir}$ were preadsorbed by adding 10^7 phage to 1 ml of this culture and incubating for 20 minutes in a 37°C water bath. 2.5 ml R-top agar (kept at 45°C) were then added and immediately plated onto a freshly made R plate. The plate was incubated at 37°C, face up, for 8 hours. At the end of this time the soft agar layer was scraped into a small centrifuge tube. The surface of the plate was washed with 1 ml of broth and the wash was added to the centrifuge tube. 5 drops of chloroform were added and the mixture was vortexed vigorously for 30 seconds. After standing 10 minutes at room temperature the cells debris were centrifuged down and the supernatant, which contained the P1 lysate was collected.

6.2.2.2. Transduction with P1 lysates

5 ml of fresh overnight culture of CGSC7421 were centrifuged and the pellet was resuspended in 5 ml of MC buffer. The solution was aerated at 37°C for 15 minutes. 100 µl of the suspended cells were then added to each of 5 eppendorff tubes. 100 µl of the P1 lysate were added to the first tube. 100 µl of a 10^{-1} dilution of the lysate were added to the second, 100 µl of a 10^{-2} dilution of the lysate were added to the third, 100 µl of a 10^{-3} dilution of the lysate were added to the fourth. Nothing was added to the fifth, which served as a control. A sixth tube was prepared with 100 µl of the P1 lysate, but no cells, as another control. The P1 were preadsorbed by incubation at 37°C in a water bath for 20 minutes. 200 µl of 1M sodium citrate (to prevent killing of the recipient strain; sodium citrate prevents readsorption of the P1 phage by removing Ca^{2+} ions). The entire contents of each of the six tubes were plated on LB plates containing the appropriate antibiotic (the deleted gene carrying an antibiotic resistance gene as a marker for correct deletion). The plates were incubated at 37°C for 48 hours, and then examined to determine whether P1 transduction of the deletion occurred.

6.2.2.3. PCR verification of the transduction of the disruption

The following primers flanking the site of the *dadAX* operon were used to verify simultaneous loss of the native gene fragment and gain of the new mutant-specific fragment into CGSC7421 using the previous PCR conditions:
N-terminus checking primer for *dadAX* disruption:
5'-CCACTAATCCGCTTCTGAACG-3'

C-terminus checking primer for *dadAX* disruption:
5'-CTTGTTGTAAGCCGGATCGGAG-3'

6.2.2.4. Elimination of chloramphenicol gene

Cml^R mutant (ET1) was grown in a 5 ml LB culture with 5 µl Cml^{30} at 37°C to an $OD_{600}\approx0.6$ and made electrocompetent by concentrating 100-fold and washing three times with ice cold sterile water. The electrocompetent cells were thawed on ice and mixed with 10 to 50 ng of pCP20 (FLP helper plasmid), and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 µF and a resistance of 200 Ω. The shocked cells were immediately added to 1 ml of SOC and incubated 1 hour at 30°C under agitation. 200 µl was then spread onto LB agar Amp^{200}. Amp^R transformants were selected at 30°C, after wich a few were colony purified once nonselectively at 43°C and then tested for loss of all antibiotic resistances. The majority lost the FRT-flanked chloramphenicol resistance gene and the FLP helper plasmid simultaneously.

The same primers flanking the site of the *dadAX* operon used to verify simultaneous loss of the native gene fragment and gain of the new mutant-specific were used once again to confirm elimination of the chloramphenicol resistance using the previous PCR conditions.

6.2.3. Deletion of *metA* gene

6.2.3.1. Creation of the PCR fragment to disrupt the *metA* gene

Attempts to delete *metA* (972bp) in *E. coli* BW25113 using the method described by Datsenko & Wanner[120]. The following primers were used for amplification of the *kan* gene (kanamycin resistance-1304 bp) from pKD 13:

N-terminus *metA* primer contained the priming site of the *kan* gene indicated in bold and the homology extension of *metA*:

5'-
CGTCTTTGTGATGACAACTTCTCGTGCGTCTGGTCAGGAAATTCGTCCAC
ATTCCGGGGATCCGTCGACC-3'

C-terminus *metA* primer contained the priming site of the *kan* gene indicated in bold and the homology extension of *metA*:
5'-
GTTGAGCCAGTTGGTAAACAGTAAATTACCGTGACTACGCCAGCTCGCTC
GTGTAGGCTGGAGCTGCTTCG-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead with Taq DNA polymerase. One bead was mixed with 0.6 µl of each primer (from a 1 pmol/µl primer solution), 1 µl of plasmid DNA template (pKD13) and 22.8 µl of water.
The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 96°C	5 mins
Denaturation	2) 95°C	40 s
Annealing	3) 50°C	40 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	5 mins

The PCR product (1404 bp) was gel-purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer.

6.2.3.2. *metA* gene disruption

Transformants carrying the Red helper plasmid (pKD46) were grown in a 5 ml SOB culture with 5 µl Amp^{100} and 10 µl 0.5M L-arabinose at 30°C to an $OD_{600}\approx0.6$ and made electrocompetent by concentrating 100-fold and washing three times with ice cold sterile water. The electrocompetent cells were thawed on ice and mixed with 10 to 50 ng of PCR product, and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance

of 25 µF and a resistance of 200 Ω. The shocked cells were immediately added to 1 ml of SOC and incubated 1 hour at 37°C under agitation. One half was then spread onto LB agar Kan^{30} to select Kan^{R} transformants. If none grew within 24 hours, the remainder was spread after standing at room temperature.

6.2.4. Deletion of *metB* gene

6.2.4.1. Creation of the PCR fragment to disrupt the *metB* gene

Attempts to deleted *metB* (1161 bp) in *E. coli* BW25113 by the method described by Datsenko & Wanner[120]. The following primers were used for amplification of the *kan* gene (1304 bp) from pKD 13:

N-terminus *metB* primer contained the priming site of the *kan* gene indicated in bold and the homology extension of *metB*:
5'-
CTTTACCGGATTTAATGAACCGCGCGCGCATGATTACTCGCGTCGCGGCA
GTGTAGGCTGGAGCTGCTTCG-3'

C-terminus *metB* primer contained the priming site of the *kan* gene indicated in bold and the homology extension of *metB*:
5'-
CCGGCGGCAGCACGCGCTTCTGGTGCCATGCCTGCATGTGTCATGGTTGC
ATTCCGGGGATCCGTCGACC-3'

The following primers were used for amplification of the *cat* gene (1015 bp) from pKD3:

N-terminus *metB* primer contained the priming site of the *cat* gene indicated in bold and the homology extension of *metB*:

5'-
CTTTACCGGATTTAATGAACCGCGCGCGCATGATTACTCGCGTCGCGGCA
CATATGAATATCCTCCTTAG-3'

C-terminus *metB* primer contained the priming site of the *cat* gene indicated in bold and the homology extension of *metB*:
5'-
CCGGCGGCAGCACGCGCTTCTGGTGCCATGCCTGCATGTGTCATGGTTGC
TGTGTAGGCTGGAGCTGGAGCTGCTT-3'

N-terminus *metB* extended primer contained the priming site of the *cat* gene indicated in bold and a bigger homology extension of *metB*:
5'-
CGAACAGTATGGTTGCGTTGTCCCACCGATCCATCTTTCCAGCACCTATA
ACTTTACCGGATTTAATGAA **CATATGAATATCCTCCTTAG**-3'

C-terminus *metB* extended primer contained the priming site of the *cat* gene indicated in bold and a bigger homology extension of *metB*:
5'-
GCAATTAAATCTTCGCCATCTTCAATACCGGTGGAGATACGCAGCAGCGT
CTCGGAGATCCCGGCGGCAG **TGTGTAGGCTGGAGCTGCTT**-3'

The PCR reaction was performed using a pureTaq Ready-to-go PCR bead with Taq DNA polymerase. One bead was mixed with 0.6 µl of each primer (from a 1 pmol/µl primer solution), 1 µl of plasmid DNA template (pKD3 or pKD13) and 22.8 µl of water.

The PCR reaction was performed using the following optimised cycle conditions:

Initial denaturation and activation	1) 96°C	5 mins
Denaturation	2) 95°C	40 s
Annealing	3) 50°C	40 s
Extension	4) 72°C	1 min 30s
Amplification	5) Repeat from 2) for 30 cycles	
Final extension	6) 72°C	5 mins

The PCR products (1404 bp for *kan* gene, 1115 bp for *cat* gene and 1155 bp for *cat* gene with extended homology of *metB* gene) were gel-purified using a QIAquick gel extraction kit, and eluted in 30 µl of EB buffer.

6.2.4.2. *metB* gene disruption

Transformants carrying the Red helper plasmid (pKD46) were grown in a 5 ml SOB culture with 5 µl Amp^{100} and 10 µl 0.5M L-arabinose at 30°C to an $OD_{600}\approx0.6$ and made electrocompetent by concentrating 100-fold and washing three times with ice cold sterile water. The electrocompetent cells were thawed on ice and mixed with 10 to 50 ng of PCR product, and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 µF and a resistance of 200 Ω. The shocked cells were immediately added to 1 ml of SOC and incubated 1 hour at 37°C under agitation. One half was then spread onto LB agar Kan^{30} or Cml^{30} to select Kan^R or Cml^R transformants. If none grew within 24 hours, the remainder was spread after standing at room temperature. After primary selection, mutants were maintained on LB agar without antibiotic. They were colony purified once nonselectively at 37°C and then tested for ampicillin sensitivity to test for the loss of the helper plasmid. If it was not lost, then a few were colony purified once at 43°C and similarly tested.
MetB was disrupted in *E. coli* BW25113 using the extended primers for Cml^R.

6.2.4.3. PCR verification

The following primers flanking the site of the *metB* gene were used to verify simultaneous loss of the native gene fragment and gain of the new mutant-specific fragment:

N-terminus checking primer for *metB* disruption:
5'-ATGACGCGTAAACAGGCCACC-3'
C-terminus checking primer for *metB* disruption:
5'-TTACCCCTTGTTTGCAGCCCG-3'

7. Experimental: validation of the in vivo DKR

7.1. Preparation of electrocompetent cells

A 5 ml LB (or LB Cml^{30} when SET21 was used) was inoculated with a single colony of CGSC6563 (or SET21) from a fresh plate, and grown overnight under shaking at 37°C.

5 ml LB (or LB Cml^{30} when SET21 was used) were then inoculated with 5 µl of the overnight culture and cultured under agitation at 37°C until the cells density reached an $OD_{600} \approx 0.6$.

The cells were then transferred into a pre-chilled eppendorff tube, and spun down at 13000 rpm for 4 min, and supernatant removed. The cells were then washed twice with 1 ml of ice-cold sterile water and spun down for 10 min at 13000 rpm before removing the supernatant. The cells were finally resuspended into 50 µl of ice-cold sterile water and kept on ice.

7.2. Transformation into SET21 by electroporation

The electrocompetent cells were thawed on ice and mixed with 10 to 25 ng of plasmid DNA, and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 μF and a resistance of 200 Ω. The cells were immediately resuspended in 850 μl of SOC after transformation and incubated at 37°C under agitation for 45 mins to 1 hour.
200 μl were then plated on LB Agar Amp^{200}.

8. Experimental: screening of libraries

8.1. Preparation of electrocompetent cells using CGSC6563 or SET21

See section 7.1

8.2. Transformation of libraries by electroporation

The electrocompetent cells were thawed on ice and mixed with 10 to 25 ng of plasmid DNA, and then transferred to a prechilled 1 mm gap electroporation cuvette, and subjected to a single pulse of 1.8 kV, set at a capacitance of 25 μF and a resistance of 200 Ω. The cells were immediately resuspended in 850 μl of SOC after transformation and incubated at 37°C under agitation for 45 mins to 1 hour.
The cells were then washed twice with 1ml of sterile 0.5% saline, and resuspended into 1 ml of sterile 0.5% saline. The suspension was spread on five M9 Amp^{200} plates supplemented with 2 mM *N*-acetyl-D-methionine, and 0.3 mM $CoCl_2$, and incubated at 37°C.

9. Experimental: in vitro assays of NAAAR

9.1. Preparation of cells

Two 50 ml LB Amp^{200} were inoculated with 50 µl of an overnight culture of 5 ml of cells and grown under agitation at 37°C. When $OD_{600}\approx0.6$, 50 µl 1M IPTG was added to one of the culture, and the cultures left under agitation overnight at 37°C. The cells were then spun down and washed once with sterile water. The water was carefully removed and the cells were then weighed.

9.2. HPLC assays the enzymatic reactions

The column used in these assays was the CHIROBIOTIC T™ chiral column.

9.2.1. Initial test to assess the dynamic kinetic resolution

The reaction solution (per ml) was prepared as followed:
Phosphate buffer (0.1M / pH 7.0): 100 µl (10 mM)
N-acetyl-D-methionine: 5 mg (0.262 mM)
Sterile water: 900 µl
Cells (wcw): 50 mg of cells
The reaction mixture was incubated at 37°C under agitation. 50 µl samples were taken regularly from both reaction mixtures, and spun down. The supernatant was carefully pipetted into a HPLC tube and directly analysed.

The conditions used for the first HPLC assay were as followed:
Mobile phase: 0.1% TEAA/methanol/H_20: 10/25/65 (v/v/v)
Flow rate: 0.6 ml/min
Injection volume: 5 µl
Wavelength detection: 210 nm
Duration: 20 min

9.2.2. Enzymatic assays with cobalt

The reaction solution (per ml) was prepared as followed:

Tris-HCl (0.5M / pH 7,5): 100 µl (50 mM)

N-acetyl-D-methionine (0.1M): 50 µl (5 mM)

Cobalt chloride (0.1M): 60 µl (6 mM)

Sterile water: 890 µl

50 mg wet weight of cells

The two reaction mixtures were then incubated at 37°C under agitation. 50 µl samples were taken regularly from both reaction mixtures, and spun down. The supernatant was carefully pipetted into a HPLC tube and directly analysed.

The optimised conditions used for the HPLC assay were as followed:

Mobile phase: 0.1% TEAA/methanol/H_20: 10/20/70 (v/v/v)

Flow rate: 0.3 ml/min

Injection volume: 5 µl

Wavelength detection: 210 nm

Duration: 20 min

10.Bibliography

1. Breuer, M., Ditrich, K., Habicher, T., Hauer, B., Keβeler, M., Sturmer, R., Zelinski, T., *Industrial methods for the production of optically active intermediates.* Angewandte Chemie International Edition, 2004. **43**(7): p. 788-824.
2. Bommarius, A., Schwarn, M., Drauz, K., *Biocatalysis to amino acid-based chiral pharmaceuticals - examples and perspectives.* Journal of Molecular Catalysis B: Enzymatic, 1998. **5**(1-4): p. 1-11.
3. Fotheringham, I., *Phenylalanine*, in *Encyclopedia of bioprocess technology: Fermentation, biocatalysis and bioseparation*, I. John Wiley & Sons, Editor. 1999. p. 1943-1954.
4. Vojtisek, V., Guttmann, T., Barta, M., Netrval, J., *Preparation of L-aspartic acid by means of immobilized Alcaligenes metalcaligenes cells.* Biotechnology and Bioengineering, 1986. **28**(7): p. 1072-1079.
5. Yuan, Y.-J., Wang, S-H., Song, Z-X., Gao, R-C., *Production of L-methionine by immobilized pellets of Aspergillus oryae in a pack bed reactor.* Journal of Chemical Technology and Biotechnology, 2002. **77**(5): p. 602-606.
6. Nakajima, N., Conrad, D., Sumi, H., Suzuki, K., Esaki, N., Wandrey, C., Soda, K., *Continuous conversion to optically pure L-methionine from D-enantiomer contaminated preparations by an immobilized enzyme membrane reactor.* Journal of Fermentation and Bioengineering, 1990. **70**(5): p. 322-325.
7. Bommarius, A., Schwarn, M., Stingl, K., Kottenhahn, M., Huthmacher, K., Drauz, K., *Synthesis and use of enantiomerically pure tert-leucine.* Tetrahedron: Asymmetry, 1995. **6**(12): p. 2851-2888.
8. Maier, T., *Semisynthetic production of unnatural L-α–amino acids by metabolic engineering of the cysteine-biosynthetic pathway.* Nature Biotechnology, 2003. **21**(4): p. 422-427.
9. Yagasaki, M., Ozaki, A., *Industrial biotransformations for the production of D-amino acids.* Journal of Molecular Catalysis B: Enzymatic, 1998. **4**(1-2): p. 1-11.
10. Park, J.-H., Kim, G-J., Kim, H-S., *Production of D-amino acid using whole cells of recombinant Escherichia coli with separatly and coexpressed D-hydantoinase and N-carbamoylase.* Biotechnology Progress, 2000. **16**(4): p. 564-570.
11. Van Beilen, J., Li, Z., *Enzyme technology: an overview.* Current Opinion in Biotechnology, 2002. **13**(4): p. 338-344.
12. Trevin, S., Marzouk, H., Sable, E., Lavorel, O., *Usage of poultry slaughterhouse waste products: Electrosynthesis of amino acid L-cysteine at Bretagne Chimie Fine.* Actualite Chimique, 1998. **10**: p. 52.
13. Friedrich, A., Antranikian, G., *Keratin degradation by Fervidobacterium pennavorans, a novel thermophilic anaerobic species of the order thermotogales.* Applied and Environmental Microbiology, 1996. **62**(8): p. 2875-2882.

14. Riessen, S., Antranikian, G., *Isolation of Thermoanaerobacter keratinophilus sp. nov., a novel thermophilic, anaerobic bacterium with keratinolytic activity.* Extremophiles, 2001. **5**(6): p. 399-408.
15. Ault, A., *The monosodium glutamate story: The commercial production of MSG and other amino acids.* Journal of Chemical Education, 2004. **81**(3): p. 347-355.
16. Masuda, M., Takamatsu, S., Nishimura, N., Komatsubara, S., Tosa, T., *Improvement of nitrogen supply for L-threonine production by a recombinant strain of Serratia marcescens.* Applied Biochemistry and Biotechnology, 1992. **37**(3): p. 255-265.
17. Okamoto, K., Kino, K., Ikeda, M., *Hyperproduction of L-threonine by an Escherichia coli mutant with impaired L-threonine uptake.* Bioscience, Biotechnology, and Biochemistry, 1997. **61**(11): p. 1877-1882.
18. Maier, T., Winterhalter, C., *Fermentative production of L-cysteine with a recombinant Escherichia coli strain.* 2000, Consortium Fuer Elektrochemische Industrie GmbH, Germany: WO 2001027307.
19. Nelms, J., Edwards, R. M., Warwick, J., Fotheringham, I., *Novel mutations in the pheA gene of Escherichia coli K-12 which result in highly feedback inhibition -resistant variants of chorismate mutase/prephenate dehydratase.* Applied and Environmental Microbiology, 1992. **58**(8): p. 2592-2598.
20. *Chirality in Industry: The commercial manufacture and applications of optically active compounds*, ed. A. Collins, Sheldrake, G., Crosby, J. 1994: John Wisley & Sons Ltd.
21. Reetz, M., *Lipases as practical biocatalysts.* Current Opinion in Chemical Biology, 2002. **6**(2): p. 145-150.
22. Kung, W., Cesa, M., Burrington, J., *Enantioselective hydrolysis of N-acylamino acid esters using a combined enzyme system.* 1986, Standard Oil Co., USA: EP 182517.
23. Yet, L., *Recent developments in catalytic asymmetric Strecker-type reactions.* Angewandte Chemie International Edition, 2001. **40**(5): p. 875-877.
24. Ma, J.-A., *Recent developments in the catalytic assymetric synthesis of α– and β–amino acids.* Angewandte Chemie International Edition, 2003. **42**(36): p. 4290-4299.
25. Wagner, H., Tanner, H., Liebetanz, E., Mannsfed, S., Pfeiffer, A., *Improved methionine synthesis.* 1969, Degussa AG, Germany: DE 1906405.
26. Beller, M., Eckert, M., *Amidocarbonylation-An efficient route to amino acid derivatives.* Angewandte Chemie International Edition, 2000. **39**(6): p. 1010-1027.
27. Knowles, W., *Assymetric hydrogenation.* Accounts of Chemical Research, 1983. **16**(3): p. 106-112.
28. Ojima, I., Clos, N., Bastos, C., *Recent advances in catalytic asymmetric reactions promoted by transition metal complexes.* Tetrahedron, 1989. **45**(22): p. 6901-6939.
29. Fiorini, M., Giongo, G., *Aminophosphorine-rhodium complexes as catalysts in assymetric hydrogenation. The dependence of the enantioselectivity of the structure of the chiral ligands.* Journal of Molecular Catalysis, 1979. **5**(4): p. 303-310.

30. Fiorini, M., Giongo, G., *Aminophosphorine-rhodium complexes as catalysts in assymetric hydrogenation. The influence of the chiral amino groups on the catalyst enantioselectivity.* Journal of Molecular Catalysis, 1980. **7**(3): p. 411-413.
31. Duthaler, R., *Recent developments in the stereoselective synthesis of α-amino acids.* Tetrahedron, 1994. **50**(6): p. 1539-1650.
32. Maruoka, K., Ooi, T., *Enantioselective amino acid synthesis by chiral-phase transfert catalysis.* Chemical Reviews, 2003. **103**(8): p. 3013-3028.
33. Ishitani, H., Komiyama, S., Hasegawa, Y., Kobayashi, S., *Catalytic asymmetric Strecker synthesis. Preparation of enantiomerically pure α-amino Acid derivatives from aldimines and tributyltin cyanide or achiral aldehydes, amines, and hydrogen cyanide using a chiral zirconium catalyst.* Journal of the American Chemical Society, 2000. **122**(5): p. 762-766.
34. Arend, M., *Assymetric catalytic aminoalkylations: New powerful methods for the enantioselective synthesis of amino acid derivatives, Mannich bases, and homoallylic amines.* Angewandte Chemie International Edition, 1999. **38**(19): p. 2873-2874.
35. Attrill, R., Tye, H., Cox, L., *α-amino acids synthesis via a Cu(II) chiral Lewis acid mediated addition of soft carbon nucleophiles to glycine cation equivalents.* Tetrahedron: Asymmetry, 2004. **15**(11): p. 1681-1684.
36. Kenworthy, M., Kilburn, J., Taylor, R., *Highly functionalized organolithium reagents for enantiomerically pure α-amino acid synthesis.* Organic Letters, 2004. **6**(1): p. 19-22.
37. Chen, Y., Lurain, A., Walsh, P., *Ageneral highly enantioselective method for the synthesis of D and L α-amino acids and allylic amines.* Journal of The American Chemical Society, 2002. **124**(41): p. 12225-12231.
38. Collier, P., Campbell, A., Patel, I., Raynham, T., Taylor, R., *Enantiomerically pure α-amino acid synthesis via hydroboration-Suzuki cross-coupling.* Journal of Organic Chemistry, 2002. **67**(6): p. 1802-1815.
39. Kragl, U., Dwars, T., *The development of new methods for the recycling of chiral catalysts.* Trends in Biotechnology, 2001. **19**(11): p. 442-449.
40. Dixneuf, P., Guyot, T., Ness, M., Roberts, S., *Synthesis of optically active allenes using tandem enzyme and palladium-catalysed reactions.* Chemical Communications, 1997. **21**: p. 2083-2084.
41. Kim, M.-J., Ahn, Y., Park. J., *Dynamic kinetic resolutions and asymmetric transformations by enzymes coupled with metal catalysis.* Current Opinion in Biotechnology, 2002. **13**(6): p. 578-587.
42. Reetz, M., *Directed evolution of selective enzymes and hybrid catalysts.* Tetrahedron, 2002. **58**(32): p. 6595-6602.
43. Pamies, O., Backvall, J-E., *Combined metal catalysis and biocatalysis for an efficient deracemization process.* Current Opinion in Biotechnology, 2003. **14**(4): p. 407-413.
44. Pamies, O., Backvall, J-E., *Combination of Enzymes and Metal Catalysts. A Powerful Approach in Asymmetric Catalysis.* Chemical Reviews, 2003. **103**(8): p. 3247-3261.
45. Turner, N., *Controlling Chirality.* Current Opinion in Biotechnology, 2003. **14**(4): p. 401-406.

46. Pamies, O., Backvall, J-E., *Chemoenzymatic dynamic kinetic resolution.* Trends in Biotechnology, 2004. **22**(3): p. 130-135.
47. Wandrey, C., Wichmann, R., Leuchtenberger, W., Kula, M., Bueckmann, A., *Continuous enzymic transformation of water-soluble α-keto carboxylic acids into the corresponding amino acids.* 1981, Degussa, Fed. Rep. Ger.; Gesellschaft fuer Biotechnologische Forschung m.b.H.: EP 23346.
48. Nakajima, N., Esaki,N., Soda, K., *Enzymatic conversion of racemic methionine to the L-enantiomer.* Journal of the Chemical Society, Chemical Communications, 1990. **13**: p. 947-948.
49. Schmidt, E., Bossow, B., Wichmann, R., Wandrey, C., *Enzyme membrane reactor - an alternative approach for continuous operation with enzymes.* Kemija u Industriji, 1986. **35**(2): p. 71-77.
50. Kragl, U., Vasic-Racki, D., Wandrey, C.. *Continuous production of L-tert-leucine in series of two enzyme membrane reactors. Modeling and computer simulation.* Bioprocess Engineering, 1996. **14**(6): p. 291-297.
51. Kula, M., Wandrey, C., *Continuous enzymatic transformation in an enzyme-membrane reactor with simultaneous NADH regeneration.* Methods in Enzymology, 1987. **136**: p. 9-21.
52. Kragl, U., Kruse, W., Hummel, W., Wandrey, C., *Enzyme engineering aspects of biocatalysis: Cofactor regeneration as example.* Biotechnology and Bioengineering, 1996. **52**(2): p. 309-319.
53. Fotheringham, I., Kidman, G., McArthur, B., Robinson, L., Scollar, M., *Aminotransferase-catalyzed conversion of D-amino acids to L-amino acids.* Biotechnology Progress, 1991. **7**(4): p. 380-381.
54. Faber, K., *Non-sequential processes for the transformation of a racemate into stereoisomeric product: proposal for stereochemical classification.* Chemistry: A european journal, 2001. **7**(23): p. 5004-5010.
55. Strauss, U., Felfer, U., Faber, K., *Biocatalytic transformation of racemates into chiral buiding blocks in 100% chemical yield and enantiomeric excess.* Tetrahedron: Asymmetry, 1999. **10**(1): p. 107-117.
56. Brown, S., Parker, M-C., Turner, N., *Dynamic kinetic resolution: synthesis of optically active α-amino acid derivatives.* Tetrahedron: Asymmetry, 2000. **11**(8): p. 1687-1690.
57. Schnell, B., Faber, K., Kroutil, W., *Enzymatic racemisation and its application to synthetic biotransformations.* Advanced Synthesis and Catalysis, 2003. **345**(6-7): p. 653-666.
58. Stecher, H., Faber, K., *Biocatalytic deracemization techniques. Dynamic resolutions and stereoinversions.* Synthesis, 1997. **1**: p. 1-16.
59. Altenbuchner, J., Siemann-Herzberg, M., Syldatk, C., *Hydantoinases and related enzymes as biocatalysts for the synthesis of unnatural chiral amino acids.* Current Opinion in Biotechnology, 2001. **12**(6): p. 559-563.
60. May, O., Buchholz, S., Schwarm, M., Drauz, K., Turner, R., Fotheringham, I., *D-amino acid production by genetically engineered Escherichia coli.* 2004, Degussa AG, Germany: WO 2004042047.
61. Buchanan, K., Burton, S., Dorrington,R., Matcher, G., Skepu, Z., *A novel Pseudomonas putida strain with high levels of hydantoin-converting activity,*

producing L-amino acids. Journal of Molecular Catalysis B: Enzymatic, 2001. **11**(4-6): p. 397-406.

62. Altenbuchner, J., Mattes, R., Syldatk, C., Wiese, A., Wilms, B., Bommarius, A., Tischer, W., *Whole cell catalyst comprising a hydantoinase, a racemase and a carbamoylase for the production of amino acids*. 2001, Degussa-Huels Aktiengesellschaft, Germany; Universitaet Stuttgart; Roche Diagnostics GmbH: WO 2001023582.
63. Wilms, B., Wiese, A., Syldatk, C., Mattes, R., Altenbuchner, J., *Development of an Escherichia coli whole cell biocatalyst for the production of L-amino acids.* Journal of Biotechnology, 2001. **86**(1): p. 19-30.
64. May, O., Nguyen, P., Arnold, F., *Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine.* Nature Biotechnology, 2000. **18**(3): p. 317-320.
65. Burton, S., Dorrington, R., *Hydantoin-hydrolysing enzymes for the enantioselective production of amino acids: insights and applications.* Tetrahedron: Asymmetry, 2004. **15**(18): p. 2737-2741.
66. Tokuyama, M., Hatano, K., Nakahama, K., Takahashi, T., *Cloning of the genes for industrially useful acylamino acid racemases from Actinomycetes*. 1992, Takeda Chemical Industries Ltd, Japan: EP 474965.
67. Verseck, S., Bommarius, A., Kula, M.-R., *Screening, overexpression and characterization of an N-acylamino acid racemase from Amycolatopsis orientalis subsp. lurida.* Applied Microbiology and Biotechnology, 2001. **55**(3): p. 354-361.
68. Tokuyama, S., Hatano, K., *Purification and properties of thermostable N-acylamino acid racemase from Amycolatopsis sp. TS-1-60.* Applied Microbiology and Biotechnology, 1995. **42**(6): p. 853-859.
69. Tokuyama, S., Hatano, K., *Cloning, DNA sequencing, and heterologous expression of the gene for thermostable N-acylamino acid racemase from Amycolatopsis sp. TS-1-60 in Escherichia coli.* Applied Microbiology and Biotechnology, 1995. **42**(6): p. 884-889.
70. Tokuyama, S., *Discovery and application of a new enzyme N-acylamino acid racemase.* Journal of Molecular Catalysis B: Enzymatic, 2001. **12**(1-6): p. 3-14.
71. Tokuyama, S., Miya, H., Hatano, K, Takahashi, T., *Purification and properties of a novel enzyme, N-acylamino acid racemase, from Streptomyces atratus Y-53.* Applied Microbiology and Biotechnology, 1994. **40**(6): p. 835-842.
72. Su, S.-C., Lee, C.-Y., *Cloning of N-acylamino acid racemase gene from Amycolatopsis azurea and biochemical characterization of gene product.* Enzyme and Microbial Technology, 2002. **30**(5): p. 647-655.
73. May, O., Verseck, S., Bommarius, A., Drauz, K., *Development of dynamic kinetic resolution processes for biocatalytic production of natural and nonnatural L-amino acids.* Organic Process Research & Development, 2002. **6**(4): p. 452-457.
74. Richard, J., Amyes, T., *Proton transfer at carbon.* Current Opinion in Chemical Biology, 2001. **5**(6): p. 626-633.

75. Taylor Ringia, E., Garrett, J., Thoden, J., Holden, H., Rayment, I., Gerlt, J., *Evolution of Enzymatic Activity in the Enolase Superfamily: Functional Studies of the Promiscuous o-Succinylbenzoate Synthase from Amycolatopsis.* Biochemistry, 2004. **43**(1): p. 224-229.
76. Palmer, D., Garrett, J., Sharma, V., Meganathan, R., Babbitt, P., Gerlt, J., *Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase.* Biochemistry, 1999. **38**(14): p. 4252-4258.
77. Babbitt, P., Gerlt, J., *New functions from old scaffolds: How nature reengineers enzymes for new functions.* Advances in Protein Chemistry, 2000. **55**: p. 1-28.
78. Gerlt, J., Babbitt, P., Rayment,I., *Divergent evolution in the enolase superfamily: the interplay of mechanism and specicity.* Archives of Biochemistry and Biophysics, 2005. **433**(1): p. 59-70.
79. Thoden, J., Taylor Ringia, E., Garrett, J., Gerlt, J., Holden, H., Rayment, I., *Evolution of Enzymatic Activity in the Enolase Superfamily: Structural Studies of the Promiscuous o-Succinylbenzoate Synthase from Amycolatopsis.* Biochemistry, 2004. **43**(19): p. 5716-5727.
80. Wang, W.-C., Chiu, W.-C., Hsu, S.-K., Wu, C.-L., Chen, C.-Y., Liu, J.-S. ,Hsu, W.-H., *Structural Basis for Catalytic Racemization and Substrate Specificity of an N-Acylamino Acid Racemase Homologue from Deinococcus radiodurans.* Journal of Molecular Biology, 2004. **342**(1): p. 155-169.
81. Turner, N., *Directed evolution of enzymes for applied biocatalysis.* Trends in Biotechnology, 2003. **21**(11): p. 474-478.
82. Babbitt, P.C., Hasson, M., Wedekind, J., Palmer, D., Barrett, W.,Reed, G., Rayment, I., Ringe, D., Kenyon, G., Gerlt, J., *The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids.* Biochemistry, 1996. **35**(51): p. 16489-16501.
83. Gerlt, J., Babbitt, P., *Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis.* Current Opinion in Chemical Biology, 1998. **2**(5): p. 607-612.
84. Gerlt, J., Raushel, F., *Evolution of function in $(\beta/\alpha)_8$–barrel enzymes.* Current Opinion in Chemical Biology, 2003. 7(2): p. 252-264.
85. Schmidt, D., Mundorff, E., Dojka, M., Bermudez, E., Ness, J., Govindarajan, S., Babbitt, P., Minshull, J., Gerlt, J., *Evolutionary Potential of $(\beta/\alpha)_8$-Barrels: Functional Promiscuity Produced by Single Substitutions in the Enolase Superfamily.* Biochemistry, 2003. **42**(28): p. 8387-8393.
86. Matthiessen, M.W., *Affordable biocomputing for everyone: using the internet, freeware and open-source software.* Trends in Biochemical Sciences, 2002. **27**(11): p. 586-588.
87. Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Research, 1997. **25**(17): p. 3389-3402.
88. Leung, D., Chen, E., Goeddel, D., *A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction.* Technique, 1989. **1**: p. 11-15.

89. Eckert, K., Kunkel, T., *DNA polymerase fidelity and the polymerase chain reaction.* PCR Methods and Applications, 1991. **1**(1): p. 17-24.
90. Cadwell, R., Joyce, G., *Mutagenic PCR.* PCR Methods and Applications, 1994. **3**(6): p. S136-S140.
91. Stemmer, W., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. **370**(6488): p. 389-391.
92. Zhao, H., Chockalingam, K., Chen, Z., *Directed evolution of enzymes and pathways for industrial biocatalysis.* Current Opinion in Biotechnology, 2002. **13**(2): p. 104-110.
93. Kuchner, O., Arnold, F., *Directed evolution of enzymes catalysts.* Trends in Biotechnology, 1997. **15**(12): p. 523-530.
94. Arnold, F., *Design by directed evolution.* Accounts of Chemical Research, 1998. **31**(3): p. 125-131.
95. Vartanian, J.-P., Henry, M., Wain-Hobson, S., *Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions.* Nucleic Acids Research, 1996. **24**(14): p. 2627-2631.
96. Seah, S., Britton, K., Rice, D., Asano, Y., Engel, P., *Single amino acid substitution in Bacillus sphaericus phenylalanine dehydrogenase dramatically increases its discrimination between phenylalanine and tyrosine substrates.* Biochemistry, 2002. **41**(38): p. 11390-11397.
97. Shafikhani, S., Siegel, R., Ferrari, E., Schellenberger, V., *Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization.* BioTechniques, 1997. **23**(2): p. 304-310.
98. Bocola, M., Otte, N., Jaeger, K.-E., Reetz, M., Thiel, W., *Learning from Directed Evolution: Theoretical Investigations into Cooperative Mutations in Lipase Enantioselectivity.* ChemBioChem, 2004. **5**(2): p. 214-223.
99. Broun, P., Shanklin, J., Whittle, E., Somerville, C., *Catalytic plasticity of fatty acid modificationenzymes underlying chemical diversity of plant lipids.* Science, 1998. **282**(5392): p. 1315-1317.
100. Xiang, H., Luo, L., Taylor, K., Dunaway-Mariano, D., *Interchange of catalytic activity within the 2-enoyl-coenzyme A hydratase/isomerase superfamily based on a common active site template.* Biochemistry, 1999. **38**(24): p. 7638-7652.
101. Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., Tawfik, D., *The 'evolvability' of promiscuous protein functions.* Nature Genetics, 2005. **37**(1): p. 73-76.
102. Ueguchi, C., Suzuki, T., Yoshida, T., Tanaka, K-I., Mizuno, T., *Systematic mutational analysis revealing the functional domain organisation of Escherichia coli nucloid protein H-NS.* Journal of Molecular Biology, 1996. **263**(2): p. 149-162.
103. Zha, D., Wilensek, S., Hermes, M., Jaegerb, K.-E., Reetz, M., *Complete reversal of enantioselectivity of an enzyme-catalyzed reaction by directed evolution.* Chemical Communications, 2001. **24**: p. 2664-2665.
104. Fromant, M., Blanquet, S., Plateau, P., *Directed random mutagenesis of gene-sized DNA fragments using polymerase chain reaction.* Analytical biochemistry, 1995. **224**(1): p. 347-353.

105. Ueda, H., Nakanishi, M., Suzuki, E., Nagamune, T., *Enhancement of mutation frequency with nucleotide triphosphate analogs in PCR random mutagenesis.* Journal of Fermentation and Bioengineering, 1995. **79**(3): p. 303-305.
106. Wan, L., Twitchett, M., Eltis, L., Mauk, A., Smith, M., *In vitro evolution of horse heart myoglobin to increase peroxidase activity.* Proceedings of the Natural Academy of Sciences of the United States of America, 1998. **95**(22): p. 12825-12831.
107. Melnikov, A., Youngman, P., *Random mutagenesis by recombinational capture of PCR products in Bacillus subtilis and Acinetobacter calcoaceticus.* Nucleic Acids Research, 1999. **27**(4): p. 1056-1062.
108. Greener, A., Callahan, M., Jerpseth, B., *An efficient random mutagenesis technique using an E. coli mutator strain.* Methods in Molecular Biology, 1996. **57**: p. 375-385.
109. Glickman, B., Radman, M., *Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction.* Proceedings of the Natural Academy of Sciences of the United States of America, 1980. **77**(2): p. 1063-1067.
110. Scheuermann, R., Tam, S., Burgers, P.,Lu, C., Echols, H., *Identification of the ε–subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: Afidelity subunit for DNA replication.* Proceedings of the Natural Academy of Sciences of the United States of America, 1983. **80**(23): p. 7085-7089.
111. Cox, E.C., *Bacterial mutator genes and the control of spontaneous mutation.* Annual Review of Genetics, 1976. **10**: p. 135-136.
112. Funchain, P., Yeung, A., Stewart, J., Lin, R., Slupska, M., Miller, J., *The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness.* Genetics, 2000. **154**(3): p. 959-970.
113. Sakanyan, V., Desmarez, L., Legrain, C., Charlier, D., Mett, I., Kochikyan, A., Savchenko, A., Boyen, A., Falmagne, P., Pierard, A., Glansdorff, N., *Gene cloning, sequence analysis, purification and characterization of a thermostable aminoacylase from Bacillus stearothermophilus.* Applied and Environmental Microbiology, 1993. **59**(1): p. 3878-3888.
114. Stark, M., *Multicopy expression vectors carrying the lac represser gene for regulated high-level expression of genes in Escherichia coli.* Gene, 1987. **51**(2-3): p. 255-267.
115. Grant, S., Jessee, J., Bloom, F., Hanahan, D., *Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants.* Proceedings of the Natural Academy of Sciences of the United States of America, 1990. **87**(12): p. 4645-4649.
116. Janscak, P., Sandmeier, U., Bickle, T., *Single amino acid substitutions in the HsdR subunit of the type IB restriction enzyme EcoAI uncouple the DNA translocation and DNA cleavage activities of the enzyme.* Nucleic Acids Research, 1999. **27**(13): p. 2638-2643.
117. Wild, J., Klopotowski, T., *D-amino acid dehydrogenase of Escherichia coli K12: positive selection of mutants defective in enzyme activity and*

localisation of the structural gene. Molecular and General Genetics, 1981. **181**(3): p. 373-378.

118. Kuhn, J., Somerville, R., *Uptake and utilization of aromatic D-amino acids in Escherichia coli K12.* Biochimica et Biophysica Acta, 1974. **332**(2): p. 298-312.
119. Franklin, F., Venables, W., *Biochemical, Genetic, and regulatory studies of alanine catabolism in Escherichia coli K12.* Molecular and General Genetics, 1976. **149**(2): p. 229-237.
120. Datsenko, K.A., Wanner, Barry L., *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the natural academy of sciences of the United States of America, 2000. **97**(12): p. 6640-6645.
121. Lobocka, M., Hennig, J., Wild, J., Klopotowski, T., *Organisation and expression of the Escherichia coli K-12 dad operon encoding the smaller subunit of D- amino acid dehydrogenase and the catabolic alanine racemase.* Journal of Bacteriology, 1994. **176**(5): p. 1500-1510.
122. Hassen, A., Saidi, N., Cherif, M., Boudabous, A., *Resistance of environmental bacteria to heavy metals.* Bioresource Technology, 1998. **64**(1): p. 7-15.
123. Geslin, C., Llanos, J., Prieur, D., Jeanthon, C., *The manganese and iron superoxide dismutases protect Escherichia coli from heavy metal toxicity.* Research in Microbiology, 2001. **152**(10): p. 901-905.
124. Kansoh, A., El-Shafei, H., *Multiple heavy metal tolerance in some fungal and bacterial strains.* African Journal of Mycology and Biotechnology, 1998. **6**(3): p. 31-40.
125. Kashiwagi, K., Tsuhako, M., Sakata, K., Saisho, T., Igarashi, A., Da Costa, S., Igarashi, K., *Relation between spontaneous aminoglycoside resistance in Escherichia coli and a decrease in oligopeptide binding protein.* Journal of Bacteriology, 1998. **180**(20): p. 5484-5488.
126. Monteiro, A., Ferreira, R., Padilla, G., Ferreira, L., Costa, S., *Environmental and genetic factors affecting mutability to aminoglycoside antibiotics among Escherichia coli K12 strains.* Genetics and Molecular Biology, 2003. **26**(2): p. 221-227.
127. Cherepanov, P., Wackernagel, W., *Gene disruption in Escherichia coli: Tc^R and Km^R cassettes with the option of Flp-catalysed excision of the antibiotic-resistance determinant.* Gene, 1995. **158**(1): p. 9-14.
128. Martinez-Morales, F., Borges, A., Martinez, A., Shanmugam, K., Ingram, L., *Chromosomal integration of heterologous DNA in Escherichia coli with precise removal of markers and replicons used during construction.* Journal of Bacteriology, 1999. **181**(22): p. 7143-7148.
129. McIninch, J., Geiser, F., Prickett, K., May, S., *Determination of the absolute configuration of α–hydroglycine derivatives by enzymatic conversion and chiral high-performance liquid chromatography.* Journal of Chromatography A, 1998. **828**(1-2): p. 191-198.
130. Yu, Y.-P., Wu, S.-H., *Simultaneous analysis of enantiomeric composition of amino acids and N-acetyl-amino acids by enantioselective chromatography.* Chirality, 2001. **13**(5): p. 231-235.

131. Armstrong, D.W., Zukowski, J., *Direct enantiomeric resolution of monoterpene hydrocarbons via reversed-phase high-performance liquid chromatography with an α–cyclodextrin bonded stationary phase.* Journal of Chromatography A, 1994. **666**(1-2): p. 445-448.
132. Armstrong, D.W., Rundlett, K., Reid, G., *Use of a macrocyclic antibiotic, Rifamycin B, and indirect detection for the resolution of racemic amino alcohol by CE.* Analytical Chemistry, 1994. **66**(10): p. 1690-1695.
133. Armstrong, D.W., Liu, Y., Ekborgott, K., *A covalently bonded teicoplanin chiral stationary phase for HPLC enantioseparations.* Chirality, 1995. **7**(6): p. 474-497.
134. Liaw, S.-H., Chen, S-J., Ko, T-P., Hsu, C-S., Chen, C-J., Wang, A., Tsai, Y-C., *Crystal Structure of D-Aminoacylase from Alcaligenes faecalis DA1: a novel subset of amidohydrolases and insights into the enzyme mechanism.* Journal of Biological Chemistry, 2003. **278**(7): p. 4957-4962.
135. Lai, W.-L., Chou, L-Y., Ting, C-Y., Kirby, R., Tsai, Y-C., Wang, A., Liaw, S-H., *The functional role of the binuclear metal center in D-aminoacylase.* Journal of Biological Chemistry, 2004. **279**(14): p. 13962-13967.
136. Lindner, H., Lunin, V., Alary, A.,Hecker, R., Cygler, M., Menard, R., *Essential roles of zinc ligation and enzyme dimerization for catalysis in the aminoacylase-1/M20 family.* Journal of Biological Chemistry, 2003. **278**(45): p. 44496-44504.

11. Appendices

11.1. Genotypes of *E. coli* strains used

Phe auxotroph strain CGSC7421 (CAG12128) : λ^-, *pheA18::Tn10, rph-1*

Met auxotroph strain CGSC6563: *araD139,* Δ*(argF-lac)169*, λ^-, *flb-5301,* Δ*dadA247, trpB202, fruA25, relA1, rpsL150(str^R), metB185, deoC1*

Strain CGSC4474 (W3110): λ^-, *IN(rrnD-rrnE), rph-1*

Strain BW25113: Δ(*araD-araB)567,* Δ*lacZ4787(::rrnB-4), lacIp-4000(lacI^Q)*, λ^-, *rpoS396*(Am*), rph-1*, Δ*(rhaD-rhaB)568, rrnB-4, hsdR514*

Strain ET1: BW25113, Δ*dadAX1959::cat^R*; strain SET1: ET1-cat^S
Strain SET2: SET1, Δ*metB920::cat^R*
Strain SET3: BW25113, Δ*metB920::cat^R*

11.2. NAAARs gene sequences

Source: NCBI (http://www.ncbi.nlm.nih.gov)

11.2.1. Gene cloned from *Amycolatopsis sp.* TS-1-60 (from IFO 15079)

The sequence in bold corresponds to the priming site used for the PCR reaction. The DNA sequence encoding the acylamino acid racemase starts from base 62 to 1168.

```
1   GAATTCCCCG GGTGACCGGC TTCGACCGAG CCGGCTTTTA CGTGATCTCC AAGGAGGAGC
61  AGTGAAACTC AGCGGTGTGG AACTGCGCCG GGTGCAGATG CCGCTCGTCG CCCCGTTCCG
121 GACTTCGTTC GGCACCCAGT CGGTCCGCGA GCTCTTGCTG CTGCGCGCGG TCACGCCGGC
181 CGGCGAGGGC TGGGGCGAAT GCGTGACGAT GGCCGGTCCG CTGTACTCGT CGGAGTACAA
241 CGACGGCGCG GAACACGTGC TGCGGCACTA CTTGATCCCG GCGCTGCTGG CCGCGGAAGA
```

```
301  CATCACCGCG GCGAAGGTGA CGCCGCTGCT GGCCAAGTTC AAGGGCCACC GGATGGCCAA
361  GGGCGCGCTG GAGATGGCCG TGCTCGACGC CGAACTCCGC GCGCACGAGA GGTCGTTCGC
421  CGCCGAACTC GGATCGGTGC GCGATTCTGT GCCGTGCGGC GTTTCGGTCG GGATCATGGA
481  CACCATCCCG CAACTGCTCG ACGTCGTGGG CGGATACCTC GACGAGGGTT ACGTGCGGAT
541  CAAGCTGAAG ATCGAACCCG GCTGGGACGT CGAGCCGGTG CGCGCGGTCC GCGAGCGCTT
601  CGGCGACGAC GTGCTGCTGC AGGTCGACGC GAACACCGCC TACACCCTCG GCGACGCGCC
661  GCAGCTGGCC CGGCTCGACC CGTTCGGCCT GCTGCTGATC GAGCAGCCGC TGGAAGAGGA
721  GGACGTGCTC GGCCACGCCG AACTGGCCCG CCGGATCCAG ACACCGATCT GCCTCGACGA
781  GTCGATCGTG TCGGCGCGCG CGGCGGCGGA CGCCATCAAG CTGGGCGCGG TCCAAATCGT
841  GAACATCAAA CCGGGCCGCG TCGGCGGGTA CCTGGAAGCG CGGCGGGTGC ACGACGTGTG
901  CGCGGCGCAC GGGATCCCGG TGTGGTGCGG CGGGATGATC GAGACCGGCC TCGGCCGGGC
961  GGCGAACGTC GCGCTGGCCT CGCTGCCGAA CTTCACCCTG CCCGGCGACA CCTCGGCGTC
1021 GGACCGGTTC TACAAAACCG ACATCACCGA GCCGTTCGTG CTCTCCGGCG GCCACCTCCC
1081 GGTGCCGACC GGACCGGGCC TCGGCGTGGC GCCGATTCCG GAGCTGCTGG ACGAGGTGAC
1141 CACGGCAAAG GTGTGGATCG GTTCGTAGCC CGCTACGAAT TCCGGAGGTA GATTTGGTCG
1201 GATCGGACCA GCCGGTCCGC ACGAGGCCGG ATCTACCTTC GGGGGGTGCT GACACCGGTG
1261 CCGAGCAAAC CGCACACGAG TCTGGGACGC GTCCTCGAAG CTCTCGGGGA CGTGCTCCTC
1321 GAGCCGGTCG CCGTCGGCGC GACACGCGGC GGCAGCTCGG CGGGGTGGTG ATTCACGACC
1381 CGCACGACGA CGCGGAATTC
```

Native protein sequence (translated from the sequence above):

```
V K L S G V E L R R V Q M P L V A P F R T S F G T Q S V R E L L L L
R A V T P A G E G W G E C V T M A G P L Y S S E Y N D G A E H V L R
H Y L I P A L L A A E D I T A A K V T P L L A K F K G H R M A K G A
L E M A V L D A E L R A H E R S F A A E L G S V R D S V P C G V S V
G I M D T I P Q L L D V V G G Y L D E G Y V R I K L K I E P G W D V
E P V R A V R E R F G D D V L L Q V D A N T A Y T L G D A P Q L A R
L D P F G L L L I E Q P L E E E D V L G H A E L A R R I Q T P I C L
D E S I V S A R A A A D A I K L G A V Q I V N I K P G R V G G Y L E
A R R V H D V C A A H G I P V W C G G M I E T G L G R A A N V A L A
S L P N F T L P G D T S A S D R F Y K T D I T E P F V L S G G H L P
V P T G P G L G V A P I P E L L D E V T T A K V W I G S
```

11.2.2. Gene cloned from *Amycolatopsis orientalis* (from NRRL 2452)

The sequence in bold corresponds to the priming site used for the PCR reaction. The DNA sequence encoding the acylamino acid racemase starts from base 191 to 1297.

```
1    GGATCCGCGG ACGGGCCGAC CGTGCTCGTC GCGGTCCCGC CCGACATCGA ACGGTTGCGC
61   CGCACCGACC CCGGTCGCGG CGACGCGTGG CGCGTCGGCT GCGCGAAGTC CTCGGCGGAC
121  TGATGGCGGA CAACGCCAAG GTCGCCGGTT TCGATCGTGC CGGCTGGTAC GTGATCTCCA
181  AGGAGCAGTC GTGAAACTCA GCGGTGTGGA ACTGCGCCGG GTCCGGATGC CGCTCGTGGC
241  CCCGTTCCGG ACGTCGTTCG GGACGCAGTC CGAGCGGGAA TTGCTGCTGG TCCGCGCGGT
301  GACCCCGGCG GGCGAGGGCT GGGGCGAATG TGTCGCGATG GAGGCGCCGC TCTACTCGTC
361  GGAGTACAAC GACGCCGCCG AGCACGTGCT GCGGAACCAT CTGATCCCCG CACTGCTGGC
421  GGCCGAGGAC GTGACCGCGC ACAAGGTGAC GCCGTTGCTG GCGAAGTTCA AGGGCCACCG
481  GATGGCGAAG GGCGCGCTGG AGATGGCGGT CCTCGACGCC GAACTCCGCG CGCATGACCG
541  GTCCTTCGCG GCCGAGCTGG GGTCCACTCG CGACTCCGTG GCCTGCGGGG TCTCGGTCGG
601  GATCATGGAC TCGATCCCGC ACCTGCTCGA CGTCGTCGGC GGCTACCTCG ACGAGGGCTA
661  CGTCCGGATC AAGCTGAAGA TCGAGCCCGG CTGGGACGTC GAGCCGGTCC GGCAGGTGCG
721  TGAGCGCTTC GGTGACGACG TGCTGCTGCA GGTCGACGCG AACACCGCGT ACACGCTGGG
781  CGACGCGCCC CTGCTGTCCC GGCTCGACCC GTTCGACCTG CTGCTGATCG AGCAGCCGCT
841  CGAAGAAGAG GACGTGCTCG GCCACGCCGA GCTGGCCAAG CGGATCCGGA CGCCGATCTG
901  CCTCGACGAG TCGATCGTCT CGGCCAAGGC CGCCGCGGAC GCGATCAAGC TCGGCGCCTG
961  CCAGATCGTC AACATCAAAC CGGGCCGGGT CGGCGGATAC CTCGAAGCCC GCCGGGTGCA
1021 CGACGTCTGC GCGGCACACG GGATCGCGGT GTGGTGCGGC GGGATGATCG AGACCGGGCT
1081 CGGCCGGGCG GCCAACGTCG CACTGGCCTC GCTGCCCGGC TTCACGCTGC CGGGGGACAC
1141 CTCGGCGTCC GGCCGGTTCT ATCGCACCGA CATCACCGAG CCGTTCGTGC TGGACGCCGG
1201 GCATCTGCCG GTGCCGACCG GGCCGGGCCT CGGGGTGACT CCGATTCCGG ATCTTCTGGA
1261 CGAGGTCACC ACGGAGAAAG CGTGGATCGG TTCGTAGCTT TCTACGAATT C
```

11.3. Sequences of the clones of NAAARs

11.3.1. NAAAR gene sequence and in-frame translation from pRS7-1 (wild type gene from TS-1-60)

```
      M  N  S  K  L  S  G  V  E  L  R  R  V  Q  M  P  L ·
1     ATGAATTCGA AACTCAGCGG TGTGGAACTG CGCCGGGTGC AGATGCCGCT

      . V  A  P  F  R  T  S  F  G  T  Q  S  V  R  E  L  L ·
51    CGTCGCCCCG TTCCGGACTT CGTTCGGCAC CCAGTCGGTC CGCGAGCTCT

      .. L  L  R  A  V  T  P  A  G  E  G  W  G  E  C  V
101   TGCTGCTGCG CGCGGTCACG CCGGCCGGCG AGGGCTGGGG CGAATGCGTG

      T  M  A  G  P  L  Y  S  S  E  Y  N  D  G  A  E  H ·
151   ACGATGGCCG GTCCGCTGTA CTCGTCGGAG TACAACGACG GCGCGGAACA

      . V  L  R  H  Y  L  I  P  A  L  L  A  A  E  D  I  T ·
201   CGTGCTGCGG CACTACTTGA TCCCGGCGCT GCTGGCCGCG GAAGACATCA

      .. A  A  K  V  T  P  L  L  A  K  F  K  G  H  R  M
251   CCGCGGCGAA GGTGACGCCG CTGCTGGCCA AGTTCAAGGG CCACCGGATG

      A  K  G  A  L  E  M  A  V  L  D  A  E  L  R  A  H ·
301   GCCAAGGGCG CGCTGGAGAT GGCCGTGCTC GACGCCGAAC TCCGCGCGCA

      . E  R  S  F  A  A  E  L  G  S  V  R  D  S  V  P  C ·
351   CGAGAGGTCG TTCGCCGCCG AACTCGGATC GGTGCGCGAT TCTGTGCCGT

      .. G  V  S  V  G  I  M  D  T  I  P  Q  L  L  D  V
401   GCGGCGTTTC GGTCGGGATC ATGGACACCA TCCCGCAACT GCTCGACGTC

      V  G  G  Y  L  D  E  G  Y  V  R  I  K  L  K  I  E ·
451   GTGGGCGGAT ACCTCGACGA GGGTTACGTG CGGATCAAGC TGAAGATCGA

      . P  G  W  D  V  E  P  V  R  A  V  R  E  R  F  G  D ·
501   ACCCGGCTGG GACGTCGAGC CGGTGCGCGC GGTCCGCGAG CGCTTCGGCG

      .. D  V  L  L  Q  V  D  A  N  T  A  Y  T  L  G  D
551   ACGACGTGCT GCTGCAGGTC GACGCGAACA CCGCCTACAC CCTCGGCGAC

      A  P  Q  L  A  R  L  D  P  F  G  L  L  L  I  E  Q ·
601   GCGCCGCAGC TGGCCCGGCT CGACCCGTTC GGCCTGCTGC TGATCGAGCA

      . P  L  E  E  E  D  V  L  G  H  A  E  L  A  R  R  I ·
651   GCCGCTGGAA GAGGAGGACG TGCTCGGCCA CGCCGAACTG GCCCGCCGGA

      .. Q  T  P  I  C  L  D  E  S  I  V  S  A  R  A  A
701   TCCAGACACC GATCTGCCTC GACGAGTCGA TCGTGTCGGC GCGCGCGGCG

      A  D  A  I  K  L  G  A  V  Q  I  V  N  I  K  P  G ·
751   GCGGACGCCA TCAAGCTGGG CGCGGTCCAA ATCGTGAACA TCAAACCGGG
```

```
       . R  V  G   G  Y  L  E   A  R  R   V  H  D   V  C  A  A ·
 801   CCGCGTCGGC GGGTACCTGG AAGCGCGGCG GGTGCACGAC GTGTGCGCGG

       .. H  G  I   P  V  W   C  G  G  M   I  E  T   G  L  G
 851   CGCACGGGAT CCCGGTGTGG TGCGGCGGGA TGATCGAGAC CGGCCTCGGC

        R  A  A  N   V  A  L   A  S  L   P  N  F  T   L  P  G ·
 901   CGGGCGGCGA ACGTCGCGCT GGCCTCGCTG CCGAACTTCA CCCTGCCCGG

       . D  T  S   A  S  D  R   F  Y  K   T  D  I   T  E  P  F ·
 951   CGACACCTCG GCGTCGGACC GGTTCTACAA AACCGACATC ACCGAGCCGT

       .. V  L  S   G  G  H   L  P  V  P   T  G  P   G  L  G
1001   TCGTGCTCTC CGGCGGCCAC CTCCCGGTGC CGACCGGACC GGGCCTCGGC

        V  A  P  I   P  E  L   L  D  E   V  T  T  A   K  V  W ·
1051   GTGGCGCCGA TTCCGGAGCT GCTGGACGAG GTGACCACGG CAAAGGTGTG

       . I  G  S   *
1101   GATCGGTTCG TAG
```

11.3.2. Gene sequence from pRS8-9 (wild type gene *A. orientalis*)

```
1    ATGAATTCGA AACTCAGCGG TGTGGAACTG CGCCGGGTCC GGATGCCGCT CGTGGCCCCG
61   TTCCGGACGT CGTTCGGAAC GCAGTCGGAA CGGGAACTGC TGCTCGTCCG CGCGGTGACC
121  CCGGCGGGCG AGGGCTGGGG CGAATGTGTT GCGATGGAGG CGCCGCTCTA CTCGTCGGAG
181  TACAACGACG CCGCCGAACA CGTGCTGCGC ACCCACCTGA TCCCCGCCCT GCTGGCCGCG
241  GAGGACGTGA CCGCGTACAA GGTGACGCCG TTGCTGGCGA AGTTCAAAGG GCACCGGATG
301  GCGAAGGGGG CGCTGGAGAT GGCGGTCCTC GACGCAGAAC TCCGCGCGCA CGAGCGGTCC
361  TTTGCCGCCG AGCTCGGGTC CACACGCGAT TCCGTGGCCT GCGGGGTCTC GGTCGGGATC
421  ATGGACTCGA TCCCGCAGCT GCTCGACGTC GTCGGCGGCT ACCTCGACGA AGGCTATGTC
481  CGGATCAAAC TGAAGATCGA GCCCGGCTGG GACGTCGAGC CGGTCCGGCA GGTGCGTGAG
541  CGCTTCGGTG ACGACGTGCT GCTCCAGGTC GACGCGAACA CCGCGTACAC CCTGGGCGAC
601  GCGCCCCTGC TGTCCAGGCT CGACGCCTTC GATCTGCTGC TGATCGAGCA GCCGCTCGAA
661  GAGGAGGACG TGCTCGGCCA CGCCGAACTG GCGAAGCGGA TCCGGACGCC GATCTGCCTC
721  GACGAGTCGA TCGTGTCCGC CAAGGCCGCC GCGGACGCGA TCAAGCTGGG CGCGTGCCAG
781  ATCGTCAACA TCAAACCGGG CCGCGTCGGC GGCTACCTCG AAGCACGCCG GGTGCACGAC
841  GTCTGCGCGG CGCACGGGGT GGCGGTGTGG TGCGGCGGGA TGATCGAGAC CGGCCTCGGC
901  CGCGCGGCCA ACGTCGCGCT GGCCTCGCTG CCCGGTTTCA CGCTGCCGGG GGACACCTCG
961  GCCTCCGGCC GGTTCTACCG CACGGACATC ACCGAGCCGT TCGTGCTGGA CGCCGGGCAC
1021 CTGCCGGTGC CGACGGGACC GGGACTGGGC GTGACTCCGA TTCCGGAACT CCTGGACGAG
1081 GTCACCACGG AGAAAGCGTG GATCGGTTCG TAG
```

11.3.3. Gene and corresponding NAAAR protein sequence from the variant 3P2.

The base highlighted in red corresponds to the base mutation in the variant 3P2, whilst the amino acid in bold corresponds to the mutation G293E (glycine->glutamic acid) made in the protein sequence of this variant.

```
      M  N  S  K  L  S  G  V  E  L  R  R  V  Q  M  P  L  .
1     ATGAATTCGA AACTCAGCGG TGTGGAACTG CGCCGGGTGC AGATGCCGCT

      .  V  A  P  F  R  T  S  F  G  T  Q  S  V  R  E  L  L  .
51    CGTCGCCCCG TTCCGGACTT CGTTCGGCAC CCAGTCGGTC CGCGAGCTCT

      .  .  L  L  R  A  V  T  P  A  G  E  G  W  G  E  C  V
101   TGCTGCTGCG CGCGGTCACG CCGGCCGGCG AGGGCTGGGG CGAATGCGTG

      T  M  A  G  P  L  Y  S  S  E  Y  N  D  G  A  E  H  .
151   ACGATGGCCG GTCCGCTGTA CTCGTCGGAG TACAACGACG GCGCGGAACA

      .  V  L  R  H  Y  L  I  P  A  L  L  A  A  E  D  I  T  .
201   CGTGCTGCGG CACTACTTGA TCCCGGCGCT GCTGGCCGCG GAAGACATCA

      .  .  A  A  K  V  T  P  L  L  A  K  F  K  G  H  R  M
251   CCGCGGCGAA GGTGACGCCG CTGCTGGCCA AGTTCAAGGG CCACCGGATG

      A  K  G  A  L  E  M  A  V  L  D  A  E  L  R  A  H  .
301   GCCAAGGGCG CGCTGGAGAT GGCCGTGCTC GACGCCGAAC TCCGCGCGCA

      .  E  R  S  F  A  A  E  L  G  S  V  R  D  S  V  P  C  .
351   CGAGAGGTCG TTCGCCGCCG AACTCGGATC GGTGCGCGAT TCTGTGCCGT

      .  .  G  V  S  V  G  I  M  D  T  I  P  Q  L  L  D  V
401   GCGGCGTTTC GGTCGGGATC ATGGACACCA TCCCGCAACT GCTCGACGTC

      V  G  G  Y  L  D  E  G  Y  V  R  I  K  L  K  I  E  .
451   GTGGGCGGAT ACCTCGACGA GGGTTACGTG CGGATCAAGC TGAAGATCGA

      .  P  G  W  D  V  E  P  V  R  A  V  R  E  R  F  G  D  .
501   ACCCGGCTGG GACGTCGAGC CGGTGCGCGC GGTCCGCGAG CGCTTCGGCG

      .  .  D  V  L  L  Q  V  D  A  N  T  A  Y  T  L  G  D
551   ACGACGTGCT GCTGCAGGTC GACGCGAACA CCGCCTACAC CCTCGGCGAC

      A  P  Q  L  A  R  L  D  P  F  G  L  L  L  I  E  Q  .
601   GCGCCGCAGC TGGCCCGGCT CGACCCGTTC GGCCTGCTGC TGATCGAGCA

      .  P  L  E  E  E  D  V  L  G  H  A  E  L  A  R  R  I  .
651   GCCGCTGGAA GAGGAGGACG TGCTCGGCCA CGCCGAACTG GCCCGCCGGA

      .  .  Q  T  P  I  C  L  D  E  S  I  V  S  A  R  A  A
701   TCCAGACACC GATCTGCCTC GACGAGTCGA TCGTGTCGGC GCGCGCGGCG
      A  D  A  I  K  L  G  A  V  Q  I  V  N  I  K  P  G  .
751   GCGGACGCCA TCAAGCTGGG CGCGGTCCAA ATCGTGAACA TCAAACCGGG
```

```
      . R  V  G   G  Y  L  E   A  R  R   V  H  D   V  C  A  A ·
801   CCGCGTCGGC GGGTACCTGG AAGCGCGGCG GGTGCACGAC GTGTGCGCGG

      .. H  G  I   P  V  W   C  G  E  M   I  E  T   G  L  G
851   CGCACGGGAT CCCGGTGTGG TGCGGCGAGA TGATCGAGAC CGGCCTCGGC

       R  A  A  N   V  A  L   A  S  L   P  N  F  T   L  P  G ·
901   CGGGCGGCGA ACGTCGCGCT GGCCTCGCTG CCGAACTTCA CCCTGCCCGG

      . D  T  S   A  S  D  R   F  Y  K   T  D  I   T  E  P  F ·
951   CGACACCTCG GCGTCGGACC GGTTCTACAA AACCGACATC ACCGAGCCGT

      .. V  L  S   G  G  H   L  P  V  P   T  G  P   G  L  G
1001  TCGTGCTCTC CGGCGGCCAC CTCCCGGTGC CGACCGGACC GGGCCTCGGC

       V  A  P  I   P  E  L   L  D  E   V  T  T  A   K  V  W ·
1051  GTGGCGCCGA TTCCGGAGCT GCTGGACGAG GTGACCACGG CAAAGGTGTG

      . I  G  S  *
1101  GATCGGTTCG TAG
```

11.4. Targeted genes for deletion to create methionine auxotroph

11.4.1. *dadA-dadX* genes (dadAX)

The *dadAX* genes were knockout, deletion replaced by Cml^R and then Cml^R removed from genome of new *E. coli* strain ET1 to yield SET1. The sequences highlighted in red are the homologous regions of the primers used to replace the fragment in between by chloramphenicol resistance. The sequences in bold are the priming sequences used to check the deletion.

dadA 1660..2958
dadX 2968..4038

```
1441 TTATCAGAGT GCCACTAATC CGCTTCTGAA CGGAATTTTA TGCTGGATAA AAAGGGCGTT
1501 CAGCAGGAGA TACTAAAGAC GCCATATTGC CGCAGAGTCA GGGAGATGTG ACCAGCTTCA
1561 CCATAAAAAA GCCGCATGTT GAATAATATT TTCAACTGAG TTATCAAGAT GTGATTAGAT
1621 TATTATTCTT TTACTGTATC TACCGTTATC GGAGTGGCTA TGCGAGTTGT CATACTGGGA
1681 AGTGGTGTGG TAGGCGTTGC CAGCGCCTGG TACTTAAATC AGGCAGGACA TGAGGTCACC
1741 GTCATTGATC GGGAGCCGGG GGCCGCCCTG GAAACCAGTG CCGCGAATGC CGGGCAAATC
1801 TCCCCCGGAT ATGCTGCACC GTGGGCGGCA CCAGGTGTGC CTTTAAAAGC GATTAAATGG
1861 ATGTTCCAGC GCCATGCGCC GCTGGCGGTT CGTCTCGACG GTACGCAGTT CCAGTTGAAA
1921 TGGATGTGGC AAATGTTACG TAACTGCGAC ACCAGCCACT ACATGGAAAA CAAAGGGCGG
1981 ATGGTGCGTC TGGCGGAATA CAGCCGTGAT TGCCTGAAAG CATTGCGCGC CGAAACCAAT
2041 ATTCAGTATG AAGGGCGTCA GGGTGGGACG CTGCAACTGT TCCGTACCGA ACAACAGTAT
2101 GAAAATGCGA CCCGCGATAT CGCCGTGCTG GAAGATGCCG GCGTACCGTA TCAGCTGCTG
2161 GAATCCAGCC GCCTGGCGGA AGTGGAGCCC GCGCTGGCAG AAGTGGCGCA CAAACTGACG
2221 GGCGGCCTGC AGTTACCCAA TGATGAAACC GGAGACTGTC AGCTATTTAC CCAGAATCTG
2281 GCGCGGATGG CGGAGCAGGC GGGGGTTAAA TTCCGCTTTA ATACGCCCGT TGACCAACTG
2341 CTTTGCGACG GCGAGCAAAT CTACGGCGTG AAGTGTGGCG ATGAAGTGAT TAAGGCCGAT
2401 GCGTATGTGA TGGCGTTTGG TTCTTACTCG ACGGCGATGC TCAAAGGCAT TGTTGATATT
2461 CCGGTGTATC CGCTGAAAGG CTACTCGCTG ACCATTCCAA TTGCGCAGGA AGATGGTGCG
2521 CCGGTATCCA CCATTCTTGA TGAAACCTAC AAAATCGCCA TTACCCGTTT CGATAACCGC
2581 ATTCGTGTTG GCGGAATGGC GGAGATTGTT GGTTTTAATA CCGAGCTGTT GCAACCGCGT
2641 CGTGAAACGC TGGAGATGGT GGTTCGCGAT CTCTATCCAC GCGGCGGTCA TGTCGAGCAG
2701 GCGACTTTCT GGACTGGTCT GCGCCCGATG ACGCCAGACG GCACGCCGGT TGTCGGGCGT
2761 ACACGCTTTA AAAATCTGTG GCTGAATACC GGTCACGGCA CGCTCGGCTG GACGATGGCT
```

2821 TGCGGTTCCG GTCAGTTGTT AAGCGATCTG CTCTCTGGTC GCACGCCAGC GATCCCATAT
2881 GAGGATCTAA GCGTAGCGCG CTACAGCCGT GGATTTACGC CATCACGTCC GGGCCATTTA
2941 CATGGCGCAC ACAGCTAAGG AAACGAGATG ACCCGTCCGA TACAGGCCAG CCTCGATCTG
3001 CAGGCATTAA AACAGAATCT GTCCATTGTC CGCCAGGCCG CGACGCACGC GCGCGTCTGG
3061 TCGGTGGTAA AAGCGAACGC TTACGGGCAT GGTATTGAGC GTATCTGGAG CGCGATCGGG
3121 GCCACCGATG GCTTTGCATT GCTTAACCTG GAAGAGGCAA TAACGTTACG TGAGCGCGGC
3181 TGGAAAGGAC CGATCCTGAT GCTGGAAGGA TTTTTCCATG CTCAGGATCT GGAGATTTAT
3241 GACCAGCACC GCCTGACCAC CTGCGTACAC AGCAACTGGC AGCTCAAAGC ACTGCAAAAT
3301 GCGCGGCTAA AAGCACCGTT GGATATTTAT CTTAAAGTGA ACAGTGGGAT GAATCGGTTG
3361 GGCTTCCAGC CCGATCGCGT GCTTACCGTC TGGCAGCAGT TGCGGGCAAT GGCGAATGTT
3421 GGCGAAATGA CCCTGATGTC GCATTTTGCC GAAGCGGAAC ATCCTGATGG AATTTCCGGC
3481 CGGATGGCGC GTATTGAGCA GGCGGCGGAG GGGCTGGAGT GTCGGCGTTC GTTGTCCAAT
3541 TCGGCGGCGA CTCTGTGGCA CCCGGAAGCG CATTTTGACT GGGTTCGGCC TGGCATTATT
3601 TTGTATGGCC GTTCGCCGTC CGGTCAGTGG CGTGATATCG CCAATACCGG ATTACGTCCG
3661 GTGATGACGC TAAGCAGTGA GATTATTGGT GTCCAGACGC TAAAAGCGGG CGAGCGTGTG
3721 GGCTACGGCG GTCGCTATAC TGCGCGCGAT GAACAGCGAA TCGGCATTGT CGCCGCAGGG
3781 TA**CGCCGACG GTTATCCGCG CCACGCCTTA ACCGGTACCC CTGTTTTAGT GG**ACGGCGTG
3841 CGCACCATGA CGGTGGGGAC CGTCTCGATG GATATGCTAG CGGTCGATTT AACGCCTTGC
3901 CCGCAGGCGG GTATTGGTAC GCCGGTTGAG CTGTGGGGCA AGGAGATCAA AATTGATGAT
3961 GTCGCCGCCG CTGCCGGAAC GGTGGGCTAT GAGTTGATGT GCGCGCTGGC CGTACGCGTC
4021 CCGGTTGTGA CGGTGTAA**CT TGTTGTAAGC CGGATCGGAG** GCAACGTCTT CTGGGTGCAA

11.4.2. *metB* gene

The sequences highlighted in red are the homologous regions of the primers used to replace the fragment in between by chloramphenicol resistance. The sequences in bold are the priming sequences used to check the deletion.

```
1    ATGACGCGTA AACAGGCCAC CATCGCAGTG CGTAGCGGGT TAAATGACGA CGAACAGTAT
61   GGTTGCGTTG TCCCACCGAT CCATCTTTCC AGCACCTATA ACTTTACCGG ATTTAATGAA
121  CCGCGCGCGC ATGATTACTC GCGTCGCGGC AACCCAACGC GCGATGTGGT TCAGCGTGCG
181  CTGGCAGAAC TGGAAGGTGG TGCTGGTGCA GTACTTACTA ATACCGGCAT GTCCGCGATT
241  CACCTGGTAA CGACCGTCTT TTTGAAACCT GGCGATCTGC TGGTTGCGCC GCACGACTGC
301  TACGGCGGTA GCTATCGCCT GTTCGACAGT CTGGCGAAAC GCGGTTGCTA TCGCGTGTTG
361  TTTGTTGATC AAGGCGATGA ACAGGCATTA CGGGCAGCGC TGGCAGAAAA ACCCAAACTG
421  GTACTGGTAG AAAGCCCAAG TAATCCATTG TTACGCGTCG TGGATATTGC GAAAATCTGC
481  CATCTGGCAA GGGAAGTCGG GGCGGTGAGC GTGGTGGATA ACACCTTCTT AAGCCCGGCA
541  TTACAAAATC CGCTGGCATT AGGTGCCGAT CTGGTGTTGC ATTCATGCAC GAAATATCTG
601  AACGGTCACT CAGACGTAGT GGCCGGCGTG GTGATTGCTA AAGACCCGGA CGTTGTCACT
661  GAACTGGCCT GGTGGGCAAA CAATATTGGC GTGACGGGCG GCGCGTTTGA CAGCTATCTG
721  CTGCTACGTG GGTTGCGAAC GCTGGTGCCG CGTATGGAGC TGGCGCAGCG CAACGCGCAG
781  GCGATTGTGA AATACCTGCA AACCCAGCCG TTGGTGAAAA AACTGTATCA CCCGTCGTTG
841  CCGGAAAATC AGGGGCATGA AATTGCCGCG CGCCAGCAAA AAGGCTTTGG CGCAATGTTG
901  AGTTTTGAAC TGGATGGCGA TGAGCAGACG CTGCGTCGTT TCCTGGGCGG GCTGTCGTTG
961  TTTACGCTGG CGGAATCATT AGGGGGAGTG GAAAGTTTAA TCTCTCACGC CGCAACCATG
1021 ACACATGCAG GCATGGCACC AGAAGCGCGT GCTGCCGCCG GGATCTCCGA GACGCTGCTG
1081 CGTATCTCCA CCGGTATTGA AGATGGCGAA GATTTAATTG CCGACCTGGA AAATGGCTTC
1141 CGGGCTGCAA ACAAGGGGTA A
```

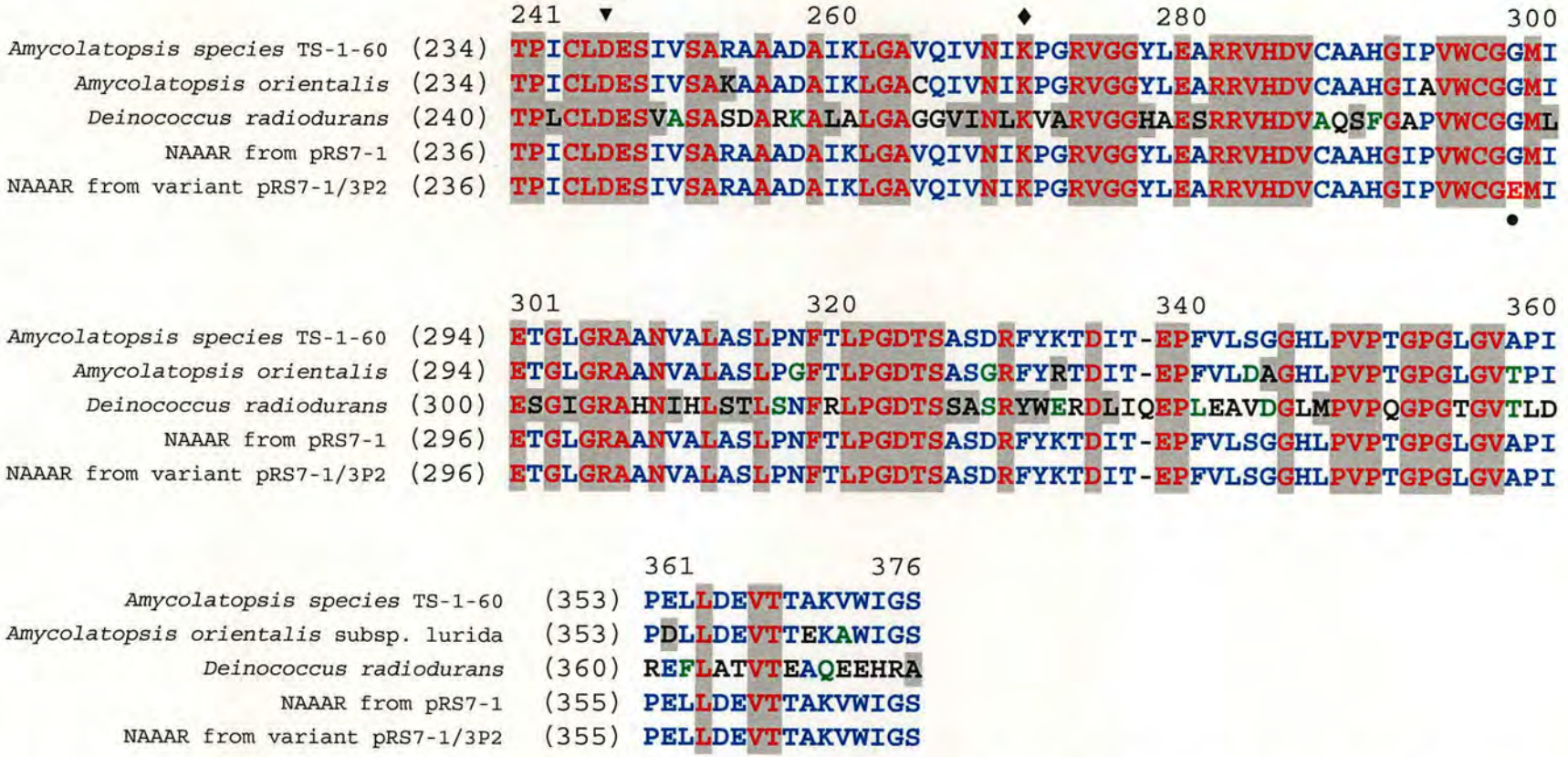
11.5. Amino acid alignment of NAAAR enzymes

```
                                    1                   20                  40                  60
Amycolatopsis species TS-1-60  (1)  -------MKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGEGWGECVTMAGP
    Amycolatopsis orientalis   (1)  -------MKLSGVELRRVRMPLVAPFRTSFGTQSERELLLVRAVTPAGEGWGECVAMEAP
    Deinococcus radiodurans    (1)  MAHTGRMFKIEAAEIVVARLPLKFRFETSFGVQTHKVVPLLILHGEGVQGVAEGTMEARP
                NAAAR from pRS7-1  (1)  -----MNSKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGEGWGECVTMAGP
NAAAR from variant pRS7-1/3P2  (1)  -----MNSKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGEGWGECVTMAGP

                                   61                  80                  100                 120
Amycolatopsis species TS-1-60 (54)  LYSSEYNDGAEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELR
    Amycolatopsis orientalis  (54)  LYSSEYNDAAEHVLRNHLIPALLAAEDVTAHKVTPLLAKFKGHRMAKGALEMAVLDAELR
    Deinococcus radiodurans   (61)  MYREETIAGALDLLRGTFLPAILGQTFANPEAVSDALGSYRGNRMARAMVEMAAWDLWAR
                NAAAR from pRS7-1 (56)  LYSSEYNDGAEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELR
NAAAR from variant pRS7-1/3P2 (56)  LYSSEYNDGAEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELR

                                  121                 140                 160          ◆      180
Amycolatopsis species TS-1-60 (114)  AHERSFAAELGSVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPV
    Amycolatopsis orientalis  (114)  AHDRSFAAELGSTRDSVACGVSVGIMDSIPHLLDVVGGYLDEGYVRIKLKIEPGWDVEPV
    Deinococcus radiodurans   (121)  TLGVPLGTLLGGHKEQVEVGVSLGIQADEQATVDLVRRHVEQGYRRIKLKIKPGWDVQPV
                NAAAR from pRS7-1 (116)  AHERSFAAELGSVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPV
NAAAR from variant pRS7-1/3P2 (116)  AHERSFAAELGSVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPV

                                  181              ▼ 200                 220▼                 240
Amycolatopsis species TS-1-60 (174)  RAVRERFGDDVLLQVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQ
    Amycolatopsis orientalis  (174)  RQVRERFGDDVLLQVDANTAYTLGDAPLLSRLDPFDLLLIEQPLEEEDVLGHAELAKRIR
    Deinococcus radiodurans   (181)  RATREAFP-DIRLTVDANSAYTLADAGRLRQLDEYDLTYIEQPLAWDDLVDHAELARRIR
                NAAAR from pRS7-1 (176)  RAVRERFGDDVLLQVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQ
NAAAR from variant pRS7-1/3P2 (176)  RAVRERFGDDVLLQVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQ
```



Comparison of the amino acid sequences of *N*-acylamino acid racemase from *A. sp.* TS-1-60, *A. orientalis* and *D. radiodurans*, and the ones expressed from pRS7-1, and its variant from pRS7-1/3P2. Sequence alignment by Vector NTI suite 6.0.

The following colour representations were used to characterise the different amino acids: **Identical residue**; **Conservative residue**; **Block of similar residue**; **Weakly similar residue**; **Non-similar residue**.
The acid/base catalysis sites are indicated by "♦"; the metal ion binding residues are indicated by "▼". The position of the variant pRS7-1/3P2 mutation was highlighted in red and indicated by "●".

D-Amino Acid Tolerant Hosts for D-Hydantoinase Whole Cell Biocatalysts

By *Robert J. Turner, John Aikens, Sylvain Royer, Louis DeFilippi, Abigail Yap, Denise Holzle, Neil Somers,* and *Ian G. Fotheringham**

Whole cell biocatalysts which enable the concerted use of D-hydantoinase, D-carbamoylase, and racemase enzymes are valuable for the production of D-amino acids. However, *Escherichia coli* host strains used for this purpose efficiently degrade D-amino acids. This work demonstrates that D-amino acid degradation occurs largely through the concerted action of D-amino acid dehydrogenase, encoded by the *dadA* gene, and D-serine dehydratase, encoded by the *dsdA* gene. Deletion mutants of *E. coli* which lack these activities were constructed and compared against wild type strains in D-amino acid degradation. An *E. coli dadA* mutant reduced the degradation of D-methionine by one third, D-phenylalanine by two-thirds, and D-2-aminobutyric acid nearly completely. Even though the *dadA* mutant had no effect on D-serine degradation, a *dadA dsdA* double mutant of *E. coli* additionally reduced degradation of D-serine, as well as D-phenylalanine, almost entirely. These strains are appropriate hosts for whole cell biosynthesis of D-amino acids using general approaches such as the hydantoinase system.

1 Introduction

The hydantoinase/carbamoylase process is well established as an efficient means to produce D-amino acids by the dynamic kinetic resolution of D/L 5-monosubstituted hydantoins [1,2]. D-amino acids are now in increasing demand as valuable intermediates in the development of new pharmaceuticals including peptidomimetic drugs and recent advances in directed evolution and genomic approaches seem likely to expand the utility of the hydantoinase process in the manufacture of both D- and L-amino acids [3]. The efficiency and enantioselectivity of the process requires the concerted use of a hydantoinase biocatalyst with a carbamoylase and, depending on the substrate, a racemase biocatalyst. It is therefore important to the overall process economics to produce these biocatalysts efficiently. Direct application of *Escherichia coli* whole cell biocatalysts which produce each of the required enzymes from cloned genes has been shown to be extremely suitable for this purpose [4]. However, *E. coli* has been previously shown to be capable of D-amino acid degradation through the action of endogenous oxidoreductase enzymes which can efficiently catabolise D-amino acids to their corresponding keto acids [5–8]. In particular the D-amino acid dehydrogenase (D-AAO) encoded by the *E. coli dadA* gene, encodes a dehydrogenase with broad substrate specificity for D-amino acids. This ability to degrade D-amino acids, compromises the usefulness and breadth of application of an otherwise efficient and scalable industrial route to produce D-amino acids. We report here, the construction and application of *E. coli* mutants deleted in the *dadAX* operon and also in the *dsdA* gene, encoding D-serine dehydratase and demonstrate that the use of such mutants as host cells can significant reduce the catabolism of D-amino acids in metabolically active whole cells of *E. coli*.

2 Materials and Methods

Strains

Strains designated CGSC were obtained from the Coli Genetic Stock Center (Yale University, CT). The following strains were used in this work. *E. coli* CGSC7177 (*metA28,* λ^-), CGSC6563 [*metB185, araD139,* Δ*(argF-lac)169,* λ^-, *flb-5301,* Δ*dadA247, trpB202, fruA25, relA1, rpsL150(*str^R*), deoC1*], BW25113 [Δ(*araD-araB*)*567*, ΔlacZ 4787 (::ρρνB-4) (Δ*lacIp-4000*(*lacI*^{q}), λ^-, *rpoS396*(Am), *rph-1*, Δ(*rhaD-rhaB*)*-568, rrnB-4, hsdR514*)] [9], ET2 (BW25113, Δ*dadAX::cat*^{R}), ET3 (ET2, Δ*dadAX:: cat*^{S}), ET4 (ET3, Δ*dsdA::kan*^{R})Plasmids.

Reagents

General reagents used in this work such as media, substrates, and antibiotics were obtained from Sigma-Aldrich (St. Louis, MO). Enzymes for restriction and ligation of DNA were obtained from NEB (Beverly, MA) and used according to standard protocols [10].

Oligonucleotides

Oligonucleotides used to amplify the chloramphenicol resistance marker from plasmid KD3 [9], used in the deletion of the *dadAX* genes of BW25113 were as follows:

5′-AACCAGTGCCGCGAATGCCGGGCAAATCTCC-CCCGGATATGCTGCACCGTCATATGAATATCCTCC-TTAG-3′ and

[*] I. Fotheringham, author to whom correspondence should be addressed, e-mail: embra1@mac.com), R. J. Turner, J. Aikens, S. Royer, L. DeFilippi, A. Yap, D. Holzle, N. Somers, deCODE Genetics, 2501 Davey Road, Woodridge IL 60517, USA.

 DOI: 10.1002/elsc.200402165

5′-CCACTAAAACAGGGGTACCGGTAGGCGCGTG-GCGCGGATAACCGTCGGCGGTGTAGGCTGGAGC-TGCTTCG-3′

Oligonucleotides used to confirm the deletion of *dadAX* in BW25113 were as follows:

5′-CCACTAATCCGCTTCTGAACG-3′ and

5′-CTCCGATCCGGCTTACAACAAG-3′

Oligonucleotides used to amplify the kanamycin resistance marker from plasmid KD13 [9], used in the deletion of the *dsdA* gene of ET3 were as follows:

5′-GCGGGCACATTCCTGCTGTCATTTATCATCTA-AGCGCAAAGAGACGTACTGTGTAGGCTGGAGCT-GCTTC-3′ and

5′-GCAGCATCGCTCACCCAGGGAAAGGATTGCG-ATGCTGCGTTGAAACGTTAATGGGAATTAGCCAT-GGTCC-3′

Oligonucleotides used to confirm the deletion of *dsdA* in ET3 were as follows:

5′-GCACTATCGTTACGGACTCG-3′ and

5′-GCGAAAGGCGTTGGGTGATG-3′

Gene Deletions

Deletion of the *dadAX* and *dsdA* genes was carried out according to the protocol of Wanner *et al.* [9] in which antibiotic resistance markers from the vectors pKD3 (chloramphenicol resistance) and pKD13 (kanamycin resistance) were amplified by PCR primers which provide flanking sequences homologous to the target genes. Fragments of DNA capable of expressing antibiotic resistance markers were thus used to replace fragments of target genes by homologous recombination and could be subsequently removed by the use of a temperature sensitive helper plasmid pCP20.

Microbial Culturing and DNA Manipulation

General molecular biology procedures were carried out using standard protocols and enzymes used according to manufacturers' recommendations [10]. Strains were cultured on LB medium for routine cloning and preparation of cells for assay. Antibiotics, ampicillin (100 μg/mL), kanamycin (50 μg/mL), and chloramphenicol (25 μg/mL) were supplemented as appropriate. Amplification of DNA by PCR was carried out using the HF PCR kit from Clontech (Palo Alto, CA). Preparation and manipulation of DNA fragments was carried out using QIAGEN (Valencia, CA) midi and mini molecular biology kits. Transformation of *E. coli* strains was carried out by electroporation using a BioRad GenePulser II with Pulse Controller.

3 Amino Acid Degradation Assay

Assays to study the degradation of the amino acids D-methionine, D-2-aminobutyric acid, D-phenylalanine, or D-serine by *E. coli* cells were carried out using freshly cultured whole cells of BW25113, ET3, or ET4. Shake flasks containing 200 mL of LB were inoculated from 5 mL overnight LB cultures of BW25113, ET3, or ET4 and grown to an $O.D._{600}$ of 0.8. Cells were then harvested by centrifugation at 5,000 G for 20 minutes, washed in saline and used for assay of amino acid degradation at a concentration of 100 mg/mL wet cell weight. In addition to the whole cells, the assay mixture contained 50 mM amino acid substrate in 50 mM phosphate buffer at pH 8.0. Assays were carried out at 30 °C for 10 hrs or 24 hrs, as indicated, with gentle agitation. Analysis of amino acid degradation was determined from the absorbance of residual amino acids following derivatization by the AccQ-Tag method and HPLC (Waters Corporation, MA).

3.1 Results

An initial qualitative study was undertaken to examine the effect of a deletion mutation in *dadA* upon D-methionine degradation by resting cells of *E. coli* K12. For this purpose two strains obtained from the CGSC were compared. The HPLC profiles of the incubation of D-methionine with cells of CSSC7177 (wild type *dadA*) or CGSC6563 (Δ*dadA247*) are shown in Fig. 1. Significant loss of D-methionine and appearance of its corresponding keto acid, 4-methylthio-2-oxobutyric acid, was apparent in the presence of CSSC7177 cells with no significant conversion evident in the presence of cells of strain CGSC6563. The appearance of 4-methylthio-2-oxobutyric acid strongly suggested a significant role for the *dadA* gene product in this result. However, these strains were not isogenic and contained additional mutations, including one in the methionine biosynthetic pathway, which could also account for this observation. Therefore, we next constructed strains of *E. coli* which differed only in the deletion of the *dadAX* region.

The *E. coli* K12 strain BW25113 was chosen as the host for gene deletion studies due to its *ara*$^-$ genotype which rendered the strain particularly appropriate to the particular gene deletion protocol used. Using this approach, the *dadAX* operon of BW25113 was replaced almost entirely by the chloramphenicol (cml) resistance marker derived from pKD3. The resulting strain, ET2, (BW25113 Δ*dadAX::cml*^{r}) was then cured of the cml resistance marker using the plasmid pCP20 to yield the cml^s strain ET3 (BW25113 Δ*dadAX*). Confirmation that gene fragments had recombined at the corresponding genomic targets was obtained from PCR fragments generated by primers which flanked the *dadAX* regions of homologous recombination in the ET2, ET3, and ET4. These showed that the 2.6 kb fragment of the wild type *dadAX* operon of BW25113 had been replaced in ET2 by a 1.7 kb fragment corresponding to the deletion mutation containing the cml resistance gene. Following removal of the cml marker from ET3, the 1.7 kb fragment was replaced by the expected 0.7 kb PCR fragment.

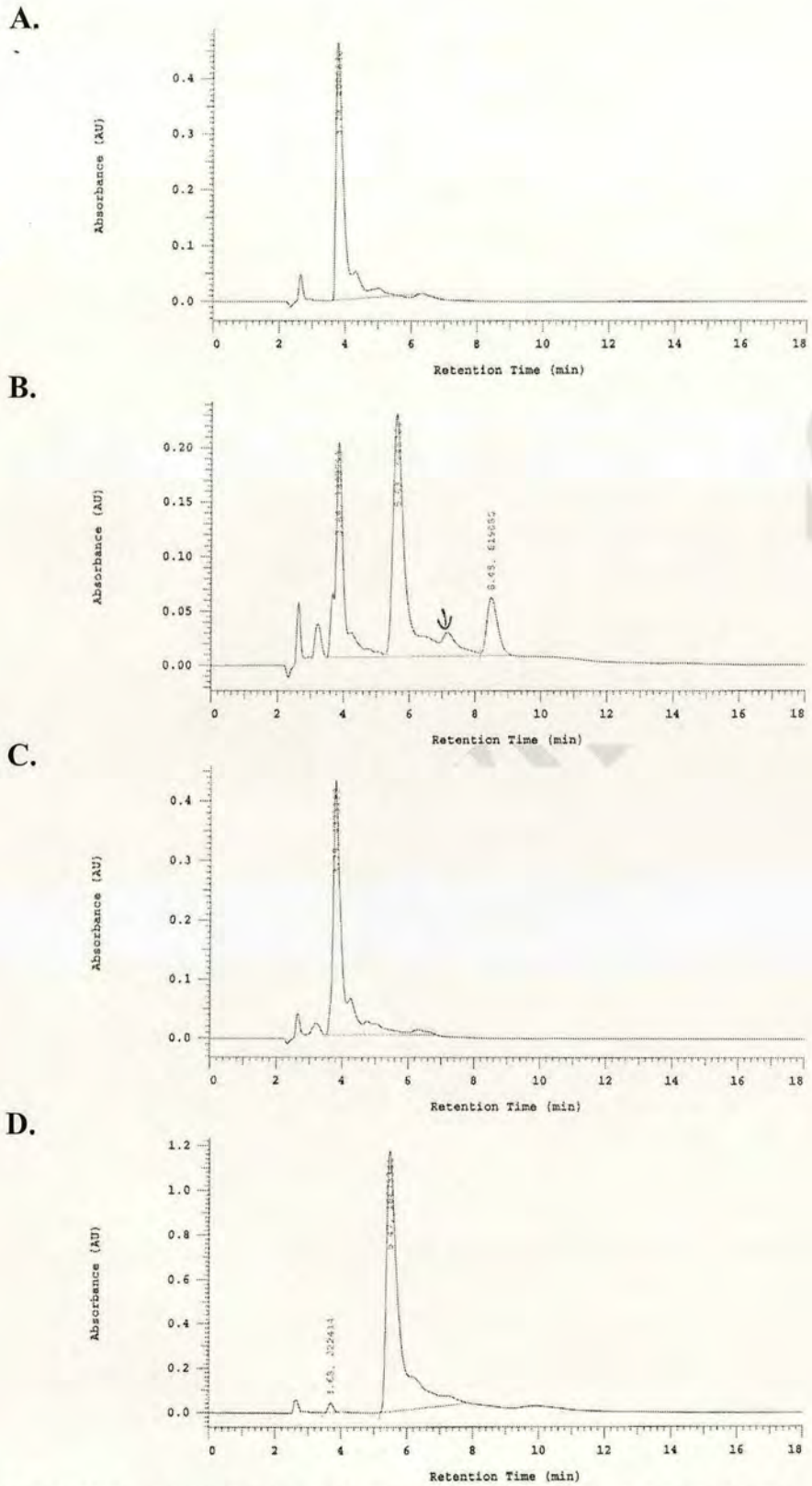


Figure 1. HPLC profile of the degradation of D-methionine by resting cells of *E. coli* following 24 hr incubation at 30 °C. A) D-methionine in the absence of cells. B) D-methionine incubated with 100 mg/mL wcw CGSC7177 (wild type *dadA*). C) D-methionine incubated with 100 mg/mL wcw CGSC6563 (Δ*dadAX*). D) α-Keto acid standard. In each assay, D-methionine was present at 50 mM in 50 mM phosphate buffer pH 8.0.

Similarly, the 1.6 kb fragment of the wild type *dsdA* operon of ET3 was shown to have been replaced in ET4 by a 1.9 kb fragment corresponding to the deletion mutation containing the kanamycin resistance gene.

Strains BW25113 and ET3 were then cultured for D-amino acid degradation analysis. Degradation of D-amino acid substrates following 10 hr incubating of either strain with D-methionine, D-phenylalanine, D-2-aminobutyric acid, or D-serine is shown in Tab. 1. The results indicate that the elimination of D-amino acid dehydrogenase by deletion of the *dadAX* operon results in moderately reduced degradation of D-methionine by whole cells of *E. coli*. The effect of the *dadAX* deletion is however considerably greater upon D-phenylalanine and D-2-aminobutyric acid where degradation of D-phenylalanine is reduced to 11.6 % from 37.4 % and degradation of D-2-aminobutyric acid is completely eliminated. In ET3, the complete degradation of D-serine remains unaffected by mutation of *dadAX*.

The *dsdA* gene of strain ET3 was then replaced by the kanamycin (kan) resistance marker of pKD13 using a similar approach to the deletion of *dadAX* operon. The resulting strain, ET4 (BW25113 Δ*dadAX*, Δ*dsdA::kan*^{r}) was then cultured in LB medium as before with BW25113 and ET3 and harvested for amino acid degradation assays. The results of incubation of ET4 with D-methionine, D-phenylalanine, D-2-aminobutyric acid, and D-serine are shown in Tab. 2 in comparison to the earlier data generated using BW25113 and ET3. The results show that deletion of both the *dadAX* operon and the *dsdA* gene effectively abolishes the degradation of D-phenylalanine and D-2-aminobutyric acid in ET4 and in addition, mutation of *dsdA* reduces degradation of D-serine from100 % to 7 % in 10 hrs, indicating that D-serine dehydratase is the major D-serine catabolic activity in *E. coli* K12.

Table 1. Effect of *dadA* mutation upon degradation of D-amino acids by *E.coli*. Values represent percentage loss of D-amino acid from initial concentrations present in the assay. Assay mixture contained cells at 100 mg/mL wcw, 50 mM amino acid substrate in 50 mM phosphate buffer at pH 8.0. Assays were carried out at 30 °C for 10 hrs.

Strain	D-Met	D-Phe	D-Abu	D-Ser
BW25113	–13.8	–37.4	–34.4	–100.0
ET3	–8.7	–11.6	2.5	–100.0

Table 2. Effect of *dadA* and *dsdA* mutation upon degradation of D-amino acids by *E.coli*. Values represent percentage loss of D-amino acid from initial concentrations present in the assay. Assay mixture contained cells at 100 mg/mL wcw, 50 mM amino acid substrate in 50 mM phosphate buffer at pH 8.0. Assays were carried out at 30 °C for 10 hrs.

Strain	D-Met	D-Phe	D-Abu	D-Ser
BW25113	–13.8	–37.4	–34.4	–100.0
ET3	–8.7	–11.6	2.5	–100.0
ET4	–9.0	2.3	–1.6	–7.0

3.2 Discussion

Deletion mutations in the *dadA* and *dsdA* genes encoding D-amino acid dehydrogenase and D-serine dehydratase, respectively, were constructed to investigate their effect upon catabolism of D-amino acids by resting whole cells of *E. coli*. The studies showed that degradation of commercially important D-amino acids such as D-methionine, D-phenylalanine, D-2-aminobutyric acid, and D-serine by *E. coli* can be significantly reduced or eliminated through these mutations. The *dadX* gene, encoding alanine racemase, was also deleted as part of the *dadAX* operon, but is not likely to be implicated in D-amino acid catabolism. Under the assay conditions used, enzymatic degradation of particular D-amino acids by *E. coli* cells with wild type *dadA* and *dsdA* genes, occurs at significantly different rates. While D-methionine is slowly oxidized to 4-methylthio-2-oxobutyric acid, D-phenylalanine, D-2-aminobutyric acid are more rapidly degraded and D-serine is completely eliminated after 10 hrs. This is an important consideration if *E. coli* is to be considered as an appropriate host organism for the manufacture of a variety of D-amino acids using a broadly applicable bioprocess such as the hydantoinase method. The demonstration that deletion mutants of *E. coli* such as the ET3 and ET4 strains described in this work do not significantly degrade a range of D-amino acids, significantly increases the potential to use *E. coli* host strains such as these in applied bioprocesses.

Acknowledgments

The authors would like to acknowledge Degussa AG and deCODE Genetics Inc. for the provision of financial support and excellent research facilities for this work. Also the scientific support and contributions from Dr. Oliver May and Dr. Stefan Verseck of Degussa are gratefully acknowledged.

Received: August 17, 2004 [CET 2165]

References

[1] C. Syldatk, et al., Microbial and enzymatic production of L-amino acids from DL-5-monosubstituted hydantoins, in *Biocatalytic Production of Amino Acids and Derivatives* (Eds: J. D. Rozzell, F. Wagner), Hanser, New York **1992**, p. 131–176.

[2] C. Syldatk et al., Microbial and enzymatic production of D-amino acids from DL-5-monosubstituted hydantoins, in *Biocatalytic Production of Amino Acids and Derivatives* (Eds: J. D. Rozzell, F. Wagner), Hanser, New York **1992**, p. 75–127.

[3] O. May, P. T. Nguyen, F.. Arnold, Inverting Enantioselectivity by Directed Evolution of Hydantoinase for Improved Production of L-Methionine, *Nat. Biotechnol.* **2000**, *18*, 317–320.

[4] O. May et al., Development of Dynamic Kinetic Resolution Processes for Biocatalytic Production of Natural and Nonnatural L-Amino Acids, *Org. Process Res. Dev.* **2002**, *6*, 452–457.

[5] J. Kuhn, R. L. Somerville, Uptake and utilization of aromatic D-amino acids in *Escherichia coli* K12, *BBA* **1974**, *332 (2)*, 298–312.

[6] J. Wild, T. Klopotowski, D-Amino acid dehydrogenase of *Escherichia coli* K12: Positive selection of mutants defective in enzyme activity and localiization of the structural gene, *Mol. Gen. Genet.* **1981**, *181*, 373–378.

[7] J. Wild, B. Obrepalska, Regulation of expression of the dadA gene encoding D-amino acid dehydrogenase in *Escherichia coli*: Analysis of dadA-lac fusions and direction of dadA transcription, *Mol. Gen. Genet.* **1982**, *186*, 405–410.

[8] J. Wild et al., D-Amino acid dehydrogenase: The enzyme of the first step of D-histidine and D-methionine racemization in Salmonella typhimurium, *Molec. Gen. Genet.* **1974**, *128*, 131–146.

[9] K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *PNAS* **2000**, *97 (12)*, 6640–6645.

[10] T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor NY **1982**.

